A Natural Introduction to Linear Algebra

Sebastian Corry
Grinnell College

January 16, 2025

CONTENTS

PREFACE

I wrote the bulk of these notes in the summer of 2024. As the title suggests, they aim to present the basic theory of vector spaces naturally, which is to say without making choices. By "choice," I mean "choice of basis": matrices are not mentioned until the final chapter, at which point the arguments are straightforward calculations. The reason for this approach is that many concepts—*e.g.*, the adjugate—appear as opaque formulas when presented matricially, whereas their coordinate-free characterizations are straightforward and insightful.

Referring to Paul Halmos and himself, Irving Kaplansky wrote "we share a philosophy about linear algebra: we think basis-free, we write basis-free, but when the chips are down we close the office door and compute with matrices like fury" [4]. In these notes, we think coordinate-free, we develop theory coordinate-free, and we only bring in matrices once the foundations have been laid for the chips to fall on. Chapter 1 introduces our basic objects, vector spaces, and some related concepts, such as dimension. Chapter 2 focuses on linear transformations and their structure (*e.g.*, rank-nullity), uses quotient spaces and direct sums to introduce universal properties, and concludes with dual spaces and projections. Chapter 3 discusses similarity invariants such as the trace and determinant and concludes with a brief introduction to spectral theory. Chapter 4 is about inner product spaces and includes a proof of the spectral theorem. Finally, Chapter 5 introduces matrices and Gaussian elimination and provides several methods for computing determinants.

The prerequisites for these notes are minimal. Ideas from calculus are mentioned in some examples, though the bright reader could get by without these. Prior experience with proofs is a must. The reader may benefit from previous experience with abstract algebra (groups, universal properties), but such background is not formally necessary.

I have done my best to cite the sources from which I learned the coordinate-free characterizations reproduced here, though a certain amount must be attributed to "folklore" and, of course, I do not claim any of the ideas as my own.

1

SPACE & POSITION

We begin by defining position. In the simplest case, consider the Cartesian plane $\mathbb{R}^2$, whose elements are the pairs (a, b) with real a and b. We typically picture this as a plane with axes drawn perpendicular to each other. Within this picture, the pair (a, b) may be visualized as the point obtained by moving a units along the x axis, then b along the y axis; that is, it describes a position.

There is some ambiguity in this description, as it assumes that we agree on the orientation of our axes. This can be avoided if we instead write $(a, b) = a\mathbf{x} + b\mathbf{y}$ where, for the time being, the sum is taken formally. The objects $\mathbf{x}$ and $\mathbf{y}$ could be interpreted as directions in the plane, so $\mathbf{x}$ represents an arrow pointing along the x axis and likewise for $\mathbf{y}$. If I imagine a different set of axes than you, I might write $a\mathbf{x}' + b\mathbf{y}$, where $\mathbf{x}'$ lies along the first quadrant diagonal with respect to your axes. Clearly, $\mathbf{x} \neq \mathbf{x}'$, so $a\mathbf{x} + b\mathbf{y}$ and $a\mathbf{x}' + b\mathbf{y}$ represent different positions.

Sums of the form $\sum a_k \mathbf{x}_k$ are known as *linear combinations*. In this chapter, we will devise an algebraic structure that encompasses such sums, thereby encoding a (homogeneous) notion of position.

§1.1. FIELDS

Our current goal is to make sense of the symbols involved in a linear combination $\sum a_k \mathbf{x}_k$. We will begin by defining the coefficients. For the sake of simplicity, consider a linear combination $a\mathbf{x}$. In the $\mathbb{R}^2$ example, we interpret this as the point at a distance a along the x axis. More generally, the symbol $\mathbf{x}$ specifies a direction and a unit of length in some space, and $a\mathbf{x}$ represents the point obtained by travelling a times along $\mathbf{x}$. Since we can travel in the direction $\mathbf{x}$, we ought to also be able to travel in the opposite direction, which might be denoted $-\mathbf{x} = (-1)\mathbf{x}$. Similarly, if we travel a times along $\mathbf{x}$, then another b times further, we have moved $a + b$ times along it in total. Lastly, if we can move $a \neq 0$ times along $\mathbf{x}$, we ought to be able to move $1/a$ times along it.

If we insist that the coefficients in a linear combination satisfy these properties, we find that they form the following structure:

DEFINITION 1.1. A *field* is a set K together with commutative and associative operations $+, (\cdot) : K \times K \to K$ satisfying

1. If $a, b \in K$, then $a + b, ab \in K$,

2. There is an element $0 \in K$ such that $0 + a = a$ for all $a \in K$,

3. There is an element $1 \in K \setminus \{0\}$ such that $1a = a$ for all $a \in K$,

4. If $a \in K$, then there is an element $-a \in K$ such that $a + (-a) = a - a = 0$,

5. If $a \in K \setminus \{0\}$, then there is an element $a^{-1} \in K$ such that $aa^{-1} = 1$,

6. If $a, b, c \in K$, then $(a + b)c = ac + bc$.

We note that fields have most of the nice algebraic properties one would expect.

PROPOSITION 1.2. Let K be a field. Then, for all $a, b \in K$,

1. $0 \cdot a = 0$,

2. $(-1)a = -a$,

3. If $ab = 0$, then $a = 0$ or $b = 0$.

Proof. **1.** We have

$$0 \cdot a + 0 = 0a = (0 + 0)a = (0 \cdot a) + (0 \cdot a),$$

so adding $-(0 \cdot a)$ to both sides yields $0 \cdot a = 0$.

2. By 1.,

$$a + (-a) = 0 = 0 \cdot a = (1 - 1)a = a + (-1)a,$$

and cancelling yields the result.

3. Suppose $ab = 0$ and $a \neq 0$. Then

$$ab = 0 = a \cdot 0,$$

and we may multiply both sides by a^{-1}, obtaining the identity $b = 0$. Taking the contrapositive of this yields the other case.

□

EXAMPLE 1. The number systems $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields with respect to their usual notions of addition and multiplication. For most purposes, it suffices to imagine our coefficients as living in the fields $\mathbb{R}$ or $\mathbb{C}$. ◇

EXAMPLE 2. Consider the set of integers $\mathbb{Z}$. Under its usual notions of addition and multiplication, $\mathbb{Z}$ is not quite a field (spaces with structures similar to $\mathbb{Z}$ are known as *rings*). In particular, it is missing multiplicative inverses. However, we can use $\mathbb{Z}$ to construct a field.

Let $p \in \mathbb{Z}$ be prime. For $a, b \in \mathbb{Z}$, write $a \sim b$ if p divides $a - b$. Then $\sim$ is an equivalence relation. Let $\mathbb{Z}/p\mathbb{Z}$ denote $\mathbb{Z}/\sim$ under the operations of addition and multiplication defined by $[a] + [b] = [a + b]$ and $[a][b] = [ab]$. To show that these are well-defined, suppose $a \sim a'$ and $b \sim b'$. Then p divides $a - a'$ and $b - b'$, so p divides

$$(a+b)-(a'+b')=(a-a')+(b-b'),$$

and $(a+b)\sim(a'+b')$. Similarly, p divides

$$ab-a'b=(a-a')b \text{ and } a'b-a'b'=a'(b-b'),$$

so $ab\sim a'b\sim a'b'$, hence $ab\sim a'b'$.

The set $\mathbb{Z}/p\mathbb{Z}$ is a field under these operations. In particular, it contains multiplicative inverses: if $[a]\in\mathbb{Z}/p\mathbb{Z}\setminus\{[0]\}$, then we may choose some representative $r\in[a]$ such that $1\leqslant r<p$. Indeed, since $a\in\mathbb{Z}$, we may consider the Euclidean division $a=qp+r$ by p with $0\leqslant r<p$, and p divides $a-r$, so $a\sim r$. Since $0\notin[a]$, $r\neq 0$. Given such a representative, we have that r and p are coprime, so there are integers s and t such that $sr+tp=1$. Thus, $sr-1$ is divisible by p, and $[s][r]=[1]$, *i.e.* $[s]=[r]^{-1}$. $\diamond$

This example highlights one way in which general fields may differ from $\mathbb{R}$ or $\mathbb{C}$. Let p be prime, so $\mathbb{Z}/p\mathbb{Z}$ is a field. Then $\sum_{k=1}^{p}[1]=[p]=[0]$. That is, in fields like $\mathbb{Z}/p\mathbb{Z}$, we can add 1 to itself some finite (positive) number of times and obtain 0. For any field K, we refer to the smallest non-zero number c such that $\sum_{k=1}^{c}1=0$ as the *characteristic* of K. If no such c exists, we say K is characteristic 0.

§1.2. VECTOR SPACES

At this point, we would do well to introduce some terminology. Given a linear combination $\sum a_k\mathbf{x}_k$, we will refer to the a_k's as *scalars* and the $\mathbf{x}_k$'s as *vectors*. In the previous section, we characterized scalars as elements of a field. Similarly, in this section, we will define a vector as an element of a certain algebraic structure.

Suppose we have two linear combinations $\sum a_k\mathbf{x}_k$ and $\sum b_k\mathbf{y}_k$. We can form the sum

$$\gamma\left(\sum a_k\mathbf{x}_k\right)+\delta\left(\sum b_k\mathbf{y}_k\right),$$

which is itself a linear combination in which $\sum a_k\mathbf{x}_k$ and $\sum b_k\mathbf{y}_k$ act as vectors. That is, a linear combination of vectors is itself a vector. This hints at the algebraic structure at hand:

DEFINITION 1.3. A *vector space over a field* K is a set V, together with a commutative and associative operation $+ : V\times V\to V$ and an action $(\cdot) : K\times V\to V$ of K on V satisfying

1. There is an element $0\in V$ such that $0+x=x$ for all $x\in V$,
2. For each $x\in V$, there is some element $-x\in V$ such that $x+(-x)=x-x=0$,
3. If $1\in K$ is the multiplicative identity of K, then $1x=x$ for all $x\in V$,
4. For all $\alpha,\beta\in K$ and $x\in V$, $\alpha(\beta x)=(\alpha\beta)x$,
5. For all $\alpha,\beta\in K$ and $x,y\in V$, $(\alpha+\beta)x=\alpha x+\beta x$ and $\alpha(x+y)=\alpha x+\alpha y$.

We refer to the elements of V as *vectors* and those of K as *scalars*. Moving forward, we will denote vectors with Latin letters and scalars with Greek letters.

As before, vector spaces have predictable algebraic properties:

PROPOSITION 1.4. Let V be a vector space over a field K. Then, for all $\alpha, \beta \in K$ and $x \in V$,

1. $0x = 0 = \alpha 0$,
2. $(-1)x = -x$,
3. If $\alpha x = 0$, then $\alpha = 0$ or $x = 0$,
4. If $\alpha x = \beta x$ and $x \neq 0$, then $\alpha = \beta$.

Proof. **1.** As before,

$$0x + 0 = 0x = (0+0)x = 0x + 0x,$$

and we may cancel. Similarly,

$$\alpha 0 + 0 = \alpha 0 = \alpha(0+0) = \alpha 0 + \alpha 0.$$

2. By 1.,

$$x + (-x) = 0 = (1-1)x = x + (-1)x,$$

and adding $(-x)$ to both sides yields the desired equality.

3. Suppose $\alpha x = 0$. If $\alpha \neq 0$, then we may proceed as in the proof of **Proposition 1.2**. Taking the contrapositive of this covers the other case.

4. Suppose $\alpha x = \beta x$. Then $(\alpha - \beta)x = 0$. If $x \neq 0$, then by 3., $\alpha - \beta = 0$.

□

This discussion began with geometric considerations, retreated to formal sums, and returns now to geometry. The rest of this section is devoted to examples of vector spaces, the goal being two-fold: first, to show that we have successfully captured our old notion of space and position; second, to illuminate the spatial structure hiding in non-obvious places.

EXAMPLE 3. Let K be a field, and consider K^n as a vector space with addition and scaling defined by

$$(x_k)_{k \in [n]} + (y_k)_{k \in [n]} = (x_k + y_k)_{k \in [n]}$$

and

$$\alpha (x_k)_{k \in [n]} = (\alpha x_k)_{k \in [n]}.$$

Letting $K = \mathbb{R}$ and $n = 2$, we recover the Cartesian plane $\mathbb{R}^2$. This also gives us an explicit connection between coordinates and linear combinations:

$$(\alpha, \beta) = \alpha(1, 0) + \beta(0, 1).$$

More generally, suppose $(\alpha_k)_{k\in[n]}$ is a vector in K^n. For the sake of succintness, we introduce the *Kronecker delta function* $\delta : \mathbb{Z}_{\geqslant 0} \times \mathbb{Z}_{\geqslant 0} \to \{0, 1\}$, which is defined by the property $\delta(i, j) = \delta_{ij} = 1$ if and only if $i = j$. Then

$$(\alpha_k)_{k\in[n]} = \sum_{k=1}^{n} \alpha_k (\delta_{kj})_{j\in[n]}.$$

Visually, we may interpret this as K^n having a standard choice of axes, with each $(\delta_{kj})_{j\in[n]}$ being the unit vector lying along the positive kth axis. We typically refer to $(\delta_{kj})_{j\in[n]}$ with the symbol e_k. ◇

EXAMPLE 4. Fix some $n \geqslant 0$, and consider a polynomial (with real coefficients) $p(t) = \sum_{k=0}^{n} \alpha_k t^k$, so p has degree $\leqslant n$. Observe that p has the form of a linear combination. More generally, the set $\mathbb{R}_n[t]$ of real polynomials with degree $\leqslant n$ is a vector space over $\mathbb{R}$, with the usual notions of polynomial addition and scaling. The take-away here is that polynomials may be thought of as points in space.

This geometric interpretation is clearest when we let $n = 0$, in which case the elements of $\mathbb{R}_0[t]$ are precisely the contant polynomials. In other words, this is the space $\mathbb{R}^1$, the real number line.

A more interesting case occurs when $n = 1$: vectors in $\mathbb{R}_1[t]$ are the polynomials of the form $\alpha t + \beta$. While we have yet to make this notion precise, the space $\mathbb{R}_1[t]$ is, in some sense, two-dimensional. That is, it is structurally similar to $\mathbb{R}^2$, in that each $\alpha t + \beta \in \mathbb{R}_1[t]$ is determined by some pair $(\alpha, \beta) \in \mathbb{R}^2$.

Note that we could just as easily have defined the space $\mathbb{C}_n[t]$ of complex polynomials of degree $\leqslant n$ or, more generally, $K_n[t]$ for any field K. ◇

§1.3. DIMENSION

Throughout this section, fix a vector space V (over a field K).

We introduced linear combinations in order to unambiguously describe position, thereby solving the problem (in coordinates) of the same symbol denoting two different points. The dual problem is this: fixing some set of vectors, when do two different linear combinations denote the same position? Ideally, never.

To make things more precise, fix vectors $x_1, \ldots, x_n$. We want to determine under what conditions the linear combinations of these vectors are unique. That is, under what conditions

$$\sum_{k=1}^{n} \alpha_k x_k = \sum_{k=1}^{n} \beta_k x_k \tag{1.1}$$

implies $\alpha_k = \beta_k$ for each k. Rearranging (1.1) yields

$$\sum_{k=1}^{n} (\alpha_k - \beta_k) x_k = 0.$$

Letting $\gamma_k = \alpha_k - \beta_k$, the uniqueness of linear combinations may be restated in the following

DEFINITION 1.5. A collection of vectors $\{x_i\}$ is *linearly independent* if, for any scalars $\gamma_1, \ldots, \gamma_n$,

$$\sum_{k=1}^{n} \gamma_k x_k = 0$$

implies $\gamma_k = 0$ for each k. We say $\{x_1, \ldots, x_n\}$ is *linearly dependent* if it is not linearly independent.

Observe that the empty set is vacuously linearly independent. We may verify the preceding discussion easily.

PROPOSITION 1.6. The vectors $x_1, \ldots, x_n$ are linearly independent if and only if

$$\sum_{k=1}^{n} \alpha_k x_k = \sum_{k=1}^{n} \beta_k x_k$$

implies $\alpha_k = \beta_k$ for all k.

Proof. **($\Longrightarrow$)** Rearranging the equality as before, we get

$$\sum_{k=1}^{n} (\alpha_k - \beta_k) x_k = 0.$$

By linear independence, it follows that, for each k, $\alpha_k - \beta_k = 0$, and we are done.

($\Longleftarrow$) Observe that $\sum_{k=1}^{n} 0 x_k = 0$. Therefore, if

$$\sum_{k=1}^{n} \alpha_k x_k = 0,$$

then, by uniqueness, $\alpha_k = 0$ for each k. $\square$

We may consider linear independence as a notion of irredundance. If a set is linearly independent, then none of its vectors may be built out of the others, as this would yield a non-trivial vanishing combination. Under this interpretation, linear dependence becomes redundancy. To be precise,

PROPOSITION 1.7. A collection of non-zero vectors $x_1, \ldots, x_n$ is linearly dependent if and only if there is some $i \geqslant 2$ such that x_i is a linear combination of the previous vectors $x_1, \ldots, x_{i-1}$.

Proof. **($\Longrightarrow$)** Choose the least i such that $\{x_1, \ldots, x_i\}$ is linearly dependent. Since any singleton $\{x\}$ is linearly independent so long as x is non-zero, $i \geqslant 2$. We may choose scalars $\alpha_1, \ldots, \alpha_i$, not all zero, such that

$$\sum_{k=1}^{i} \alpha_k x_k = 0.$$

Note that $\alpha_i \neq 0$, for otherwise we would have a dependence relation among $x_1, \ldots, x_{i-1}$, contradicting the minimality of i. Rearranging, we get

$$x_i = \sum_{k=1}^{i-1} \frac{-\alpha_k}{\alpha_i} x_k.$$

($\Longleftarrow$) We have scalars $\alpha_1, \ldots, \alpha_{i-1}$ such that

$$x_i = \sum_{k=1}^{i-1} \alpha_k x_k,$$

so

$$\sum_{k=1}^{i-1} \alpha_k x_k - x_i = 0.$$

It follows that $\{x_1, \ldots, x_n\} \supseteq \{x_1, \ldots, x_i\}$ is linearly dependent. □

EXAMPLE 5. For a geometrically insightful example of linear independence, consider the plane $\mathbb{R}^2$. If two vectors (a_1, a_2) and (b_1, b_2) are linearly dependent, then there are scalars α and β, not both zero, such that

$$\alpha(a_1, a_2) + \beta(b_1, b_2) = 0. \tag{1.2}$$

Without loss of generality, suppose β is non-zero. Then (1.2) becomes

$$(b_1, b_2) = \frac{-\alpha}{\beta}(a_1, a_2),$$

so (a_1, a_2) and (b_1, b_2) lie on a common line through the origin. Therefore, any two planar points that are not collectively collinear with the origin are linearly independent. The most obvious example is the pair of vectors $(1, 0)$ and $(0, 1)$, corresponding to points on the x and y axes respectively. ◇

Having solved the problem of uniqueness, the natural follow-up is that of existence. For this, we introduce a useful construction.

DEFINITION 1.8. The *span* of a set $\{x_1, \ldots, x_n\}$, written $\operatorname{span}(x_1, \ldots, x_n)$, is the set of all linear combinations of $x_1, \ldots, x_n$. We say that the vectors $x_1, \ldots, x_n$ span V if $\operatorname{span}(x_1, \ldots, x_n) = V$.

EXAMPLE 6. Consider the space of complex polynomials (of any degree) $\mathbb{C}[t]$. This space is not spanned by any finite set of vectors. Indeed, suppose we have $p_1, \ldots, p_n \in \mathbb{C}[t]$ and, without loss of generality, suppose that $d = \deg(p_n)$ is maximal among this collection. Then any $p \in \operatorname{span}(p_1, \ldots, p_n)$ has degree at most d. However, the vector $tp_n \in \mathbb{C}[t]$ has degree strictly greater than d, hence $tp_n \notin \operatorname{span}(p_1, \ldots, p_n)$.

Despite this, we may find interesting sets by looking at spans of vectors in $\mathbb{C}[t]$. For example, the elements of $\operatorname{span}(1, t, t^2)$ are the linear combinations of the form $\alpha_0 + \alpha_1 t + \alpha_2 t^2$. That is, $\operatorname{span}(1, t, t^2) = \mathbb{C}_2[t]$. Note that this set is itself a complex vector space; we will come back to this point later. ◇

There is an important relationship between linear independence and span. Intuitively, if Y spans V, then any linearly independent set can be built from Y. Furthermore, Y may be redundant, *i.e.* larger than it needs to be, and X is irredundant so, in some sense, minimal. This suggests, though certainly does not prove, that Y is larger than X.

THEOREM 1.9 (Steinitz Exchange Lemma)**.** If X and Y are sets such that X is linearly independent and Y is finite and spanning, then $|X| \leqslant |Y|$, and there is some $Z \subseteq Y$ with $|Z| = |Y| - |X|$ such that $X \cup Z$ is spanning.

Proof. Suppose $X = \{x_1, x_2, \dots\}$ and $Y = \{y_1, \dots, y_n\}$. Consider the set

$$S_1 = \{x_1, y_1, \dots, y_n\}.$$

Since $x_1 \in V = \operatorname{span}(y_1, \dots, y_n)$, S_1 is linearly dependent, and some y_i is a linear combination of the previous vectors $x_1, y_1, \dots, y_{i-1}$. Re-enumerating the y_i's as needed, we may, without loss of generality, assume that $i = n$. Then $y_n \in \operatorname{span}(S_1 \setminus \{y_n\})$, so $S_1 \setminus \{y_n\}$ spans V. Therefore, the set

$$S_2 = \{x_2\} \cup (S_1 \setminus \{y_n\}) = \{x_2, x_1, y_1, \dots, y_{n-1}\}$$

is linearly dependent. Continuing in this manner, and re-enumerating Y as necessary, we obtain a sequence of linearly dependent spanning sets (S_k) given by

$$S_k = \{x_k\} \cup (S_{k-1} \setminus \{y_{n-k+2}\}) = \{x_k, \dots, x_1, y_1, \dots, y_{n-k+1}\}$$

for $k \geqslant 2$. Observe that this sequence continues so long as we have exhausted neither X nor Y. Suppose $|X| > |Y|$. Then Y will be exhausted before X and, in particular, the final set in our sequence will be

$$S_{n+1} = \{x_{n+1}, \dots, x_1\} \subseteq X.$$

However, each S_k is linearly dependent, so this would imply that X is linearly dependent, a contradiction. It follows that $|X| \leqslant |Y|$ so, in particular, $X = \{x_1, \dots, x_m\}$ for some $m \geqslant 1$. Then the last set in our sequence is

$$S_m = \{x_m, \dots, x_1, y_1, \dots, y_{n-m+1}\},$$

which is a spanning set containing X. It is also linearly dependent, so there is some y_i which may be removed without altering the span. We have generally assumed $y_i = y_{n-m+1}$. Removing this element yields the spanning set

$$\{x_m, \dots, x_1, y_1, \dots, y_{n-m}\} = X \cup Z,$$

with $Z = \{y_1, \dots, y_{n-m}\}$, so $|Z| = n - m = |Y| - |X|$. □

We are now ready to introduce and characterize the key objects of this section.

DEFINITION 1.10. A collection of vectors is a *basis* if it is linearly dependent and spans V.

COROLLARY 1.11 (Dimension Theorem)**.** If V has a finite basis, then any two bases of V have the same cardinality.

Proof. Let X be a finite basis of V, and suppose Y is any other basis. Then Y is linearly independent and X is spanning, so $|Y| \leqslant |X|$, *i.e.* Y is finite. Additionally, X is linearly independent and Y is spanning, so $|X| \leqslant |Y|$, and the result follows. □

The reader should note that not every vector space has a finite basis, as shown in **Example 7**. While the dimension theorem does generalize to such spaces, these notes will focus on the finite case. Unless otherwise stated, all vector spaces are assumed to have finite bases or, equivalently, to have finite spanning sets.

COROLLARY 1.12. Every linearly independent set X can be extended to a basis.

Proof. Choose a basis B of V. There is some $Z \subseteq B$ with $|Z| = |B| - |X|$ such that $X \cup Z$ spans V. Clearly, $|X \cup Z| \leqslant |B|$, hence $X \cup Z$ is linearly independent, since it spans V and B is linearly independent. □

COROLLARY 1.13. If B is a basis and X is either linearly independent or spanning, then $|X| = |B|$ implies X is a basis.

Proof. We will first consider the case that X is linearly independent. If X is not a basis, then it may be extended to a basis Y with $|Y| > |X| = |B|$, a contradiction. Alternatively, if X is spanning and not a basis, then there is some linearly independent $Y \subsetneq X$ that spans V, which also contradicts the dimension theorem. □

Clearly, bases are both rigid and ubiquitous objects. The cardinality of a basis tells us a lot about the structure of its vector space; we give this number a name:

DEFINITION 1.14. The *dimension* of a vector space V, denoted $\dim V$, is the cardinality of any basis of V.

To see that this definition recaptures our intuitive notion of dimension, consider the following examples:

EXAMPLE 7. In **Example 4**, we defined a set $\{e_1, \ldots, e_n\}$ of vectors in $\mathbb{R}^n$ by

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0), \\ e_2 &= (0, 1, 0, \ldots, 0), \\ e_3 &= (0, 0, 1, \ldots, 0), \\ &\vdots \\ e_n &= (0, 0, 0, \ldots, 1). \end{aligned}$$

A brief consideration shows that this set is a basis for $\mathbb{R}^n$, which we refer to as the *standard basis*. From this, we see that $\mathbb{R}^n$ is n-dimensional, which aligns with our intuition. We've seen that the standard basis corresponds to our usual choice of axes. More generally, choosing a basis in a vector space is essentially the same as choosing axes. ◇

EXAMPLE 8. Considering the space $\mathbb{C}_n[t]$ of complex polynomials of degree $\leqslant n$, we see that the set $\{1, x, x^2, \ldots, x^n\}$ is a basis and has cardinality $n+1$. That is, $\dim \mathbb{C}_n[t] = n + 1$. ◇

§1.4. ISOMORPHISMS

We've seen that some vector spaces have similar structure. For example, recall that the vectors in $\mathbb{R}_1[t]$ are of the form $\alpha t + \beta$. This representation is unique, as can be seen from the fact that $\{1, t\}$ is a basis. Therefore, we may represent each of these vectors as a pair (α, β), which lives in the space $\mathbb{R}^2$. Furthermore, if we have vectors $\alpha t + \beta$ and $\gamma t + \delta$, then their sum is $(\alpha + \gamma)t + (\beta + \delta)$, and this equality holds between their representatives in $\mathbb{R}^2$: $(\alpha, \beta) + (\gamma, \delta) = (\alpha + \gamma, \beta + \delta)$. This suggests that the structure of $\mathbb{R}_2[t]$ is in some sense the same as that of $\mathbb{R}^2$, in that we may equivalently take linear combinations in one or the other.

DEFINITION 1.15. If V and W are vector spaces over a common field K, a mapping $\varphi : V \to W$ is an *isomorphism* if it is invertible and preserves linear combinations, *i.e.*

$$\varphi(\alpha x + \beta y) = \alpha\varphi(x) + \beta\varphi(y). \tag{1.3}$$

We say V is *isomorphic* to W, written $V \cong W$, if there is an isomorphism $V \to W$.

Functions $\varphi : V \to W$ satisfying (1.3) are called *linear*.

PROPOSITION 1.16. The relation $\cong$ is reflexive, symmetric, and transitive.

Proof. **(Reflexivity)** For any vector space V, the identity $\mathbb{1}_V$ is an isomorphism from V to itself.

(Symmetry) Suppose we have an isomorphism $\varphi : V \to W$. Then we have $\varphi^{-1} : W \to V$ satisfying

$$\varphi^{-1}(\alpha x + \beta y) - \alpha\varphi^{-1}(x) - \beta\varphi^{-1}(y) \overset{\varphi}{\longmapsto} \alpha x + \beta y - \alpha x - \beta y = 0.$$

Since φ is injective and $\varphi(0) = 0$, the left-hand side vanishes, so

$$\varphi^{-1}(\alpha x + \beta y) = \alpha\varphi^{-1}(x) + \beta\varphi^{-1}(y),$$

hence φ^{-1} is an isomorphism.

(Transitivity) If $\varphi : U \to V$ and $\psi : V \to W$ are isomorphisms, then $\psi \circ \varphi : U \to W$ is as well. □

However, $\cong$ is not quite an equivalence relation simply because the class of vector spaces over K is not a set.

We've seen that $\mathbb{R}_2[t] \cong \mathbb{R}^2$ and, in particular, the map $\alpha t + \beta \mapsto (\alpha, \beta)$ is an isomorphism. A brief reflection shows that the map

$$\sum_{k=0}^{n} \alpha_k t^k \longmapsto \sum_{k=1}^{n+1} \alpha_k e_{k+1}$$

is an isomorphism, so $\mathbb{R}_n[t] \cong \mathbb{R}^{n+1}$ for each $n \geqslant 1$. We construct this isomorphism by having it send each t_k—$0 \leqslant k \leqslant n$—to e_{k+1} and demanding linearity. That is, it emerges from the choice of basis $\{1, t, \ldots, t^n\}$. This suggests a way of constructing isomorphisms in general.

PROPOSITION 1.17. If V is an n-dimensional vector space over a field K, then $V \cong K^n$.

Proof. Choose a basis $\{x_1, \ldots, x_n\}$ of V. Let $\varphi : V \to K^n$ be defined by

$$\sum_{k=1}^{n} \alpha_k x_k \overset{\varphi}{\longmapsto} \sum_{k=1}^{n} \alpha_k e_k.$$

Clearly, φ is linear, so we need only show that it is invertible. Letting $\psi : K^n \to V$ be defined by

$$\sum_{k=1}^{n} \alpha_k e_k \overset{\psi}{\longmapsto} \sum_{k=1}^{n} \alpha_k x_k,$$

we see that $\varphi \circ \psi = \mathbb{1}_{K^n}$ and $\psi \circ \varphi = \mathbb{1}_V$, so $\psi = \varphi^{-1}$. □

Consequently, two vector spaces V and W are isomorphic if and only if they have the same dimension. Why then should we study vector spaces apart from K^n?

There are a couple of reasons. First and foremost, we want to study vector spaces as they arise in the structures of sets such as $\mathbb{R}_n[t]$, and developing the theory in a coordinate-free manner allows us to do this more fluidly. Secondly, coordinate-free arguments can be more insightful, as proofs in K^n are often obfuscated by tedious computations and formal manipulations.

That being said, coordinate representations are often helpful. In Chapter 5, we will develop computational tools based in K^n, which are essential when using linear algebra in the wild. Additionally, since all n-dimensional spaces are all isomorphic to K^n, it is reasonable to visualize the familiar coordinate spaces when thinking about vector spaces.

§1.5. SUBSPACES

Consider a real n-dimensional space V. Choose some integer $m < n$, and let $x_1, \ldots, x_m$ be an arbitrary collection of vectors in V. By assumption, there is some linearly independent set (*e.g.*, a basis) of cardinality strictly greater than m, so $P = \operatorname{span}(x_1, \ldots, x_m) \neq V$. Nonetheless, this set has some nice algebraic properties. Indeed, $0 \in P$ and, for all $\sum \xi_k x_k, \sum \eta_k x_k \in P$,

$$\alpha \left(\sum_{k=1}^{m} \xi_k x_k \right) + \beta \left(\sum_{k=1}^{m} \eta_k x_k \right) = \sum_{k=1}^{m} (\alpha \xi_k + \beta \eta_k) x_k \in P,$$

so P is itself a real vector space. This raises the question of what it looks like relative to V. For the sake of simplicity, suppose $\{x_1, \ldots, x_m\}$ is linearly independent. We may extend this set to a basis $\{x_1, \ldots, x_m, x_{m+1}, \ldots, x_n\}$ of V and choose the isomorphism $V \to \mathbb{R}^n$ that sends each x_k to e_k. If $m = 1$, P is sent to the line through the origin and e_1. If $m = 2$, it's sent to the plane containing e_1, e_2, and the origin.

To summarize, spans of vectors are vector spaces themselves; they look like lines, planes, etc. containing the origin. Naturally, such geometric objects are of interest. However, as discussed in the previous section, we want coordinate-free definitions, and defining our analog of lines in terms of spans would go against this. Instead, we offer the following basis-free formulation:

DEFINITION 1.18. If V is a vector space over K and $P \subseteq V$, then P is a *subspace* of V if

1. P is non-empty,

2. For all $x, y \in P$ and $\alpha, \beta \in K$, $\alpha x + \beta y \in P$.

Since P is non-empty, $0 = 0x \in P$ and, for each $x \in P$, $-x = (-1)x \in P$. Therefore, $P \subseteq V$ is a subspace if it has a vector space structure induced by V via restriction.

We've seen that the span of a collection of vectors is a subspace. The converse relies on the following obvious but important fact.

PROPOSITION 1.19. If P is a subspace of a finite-dimensional vector space V, then P is finite-dimensional.

Proof. Suppose P is infinite-dimensional, so every finite linearly independent subset of P is not spanning. We will construct a linearly independent set of size m for each $m \in \mathbb{Z}_{\geqslant 0}$. For $m = 0$, the empty set $\varnothing \subseteq P$ is linearly independent. Suppose we have a linearly independent set $I = \{x_1, \ldots, x_m\} \subseteq P$ for some $m \geqslant 0$. Since P is infinite-dimensional, I does not span P, and we may choose some non-zero vector $x \in P \setminus \operatorname{span}(I)$. If

$$\sum_{k=1}^{m} \alpha_k x_k + \beta x = 0$$

then $\beta = 0$, else

$$x = \sum_{k=1}^{m} \frac{\alpha_k}{\beta} x_k \in \operatorname{span}(I),$$

a contradiction. It follows from the linear independence of I that each α_k vanishes, so $I \cup \{x\} = \{x_1, \ldots, x_m, x\} \subseteq P$ is a linearly independent set of size $m+1$. Since each linearly independent subset of P is also an independent subset of V, the exchange lemma implies V is infinite-dimensional. □

COROLLARY 1.20. If P is a subspace of a finite-dimensional vector space V, then $\dim P \leqslant \dim V$.

Proof. Choose a basis $X_P = \{x_1, \ldots, x_m\}$ of P. Then $X_P \subseteq V$ is linearly independent, so $\dim P = |X_P| \leqslant \dim V$. □

We turn our attention now to the construction of new subspaces from old. Intuitively, the intersection of two subspaces should correspond to their intersection as hyperplanes.

PROPOSITION 1.21. If $A, B \subseteq V$ are subspaces, then $A \cap B$ is a subspace.

Proof. We have that $0 \in A, B$, so $0 \in A \cap B$, *i.e.* $A \cap B$ is non-empty. If $x, y \in A \cap B$, then $x, y \in A, B$. Since A and B are subspaces, $\alpha x + \beta y \in A, B$, hence $\alpha x + \beta y \in A \cap B$. □

Unfortunately, the union of subspaces is not generally a subspace. To see this, consider the plane $\mathbb{R}^2$ and let A and B be distinct lines through the origin: their union is not a point, line, or plane. However, we have the next best thing:

DEFINITION 1.22. Let A and B be subspaces of a vector space V. Then the *sum* of A and B, denoted $A + B$, is the set of all vectors of the form $a + b$ for $a \in A$ and $b \in B$.

PROPOSITION 1.23. If $A, B \subseteq V$ are subspaces, then $A + B$ is a subspace as well.

Proof. Since $0 \in A, B$,

$$0 + 0 = 0 \in A + B.$$

For all $a+b, a'+b' \in A+B$—$a, a' \in A$ and $b, b' \in B$—

$$\alpha(a+b)+\beta(a'+b') = (\alpha a+\beta a')+(\alpha b+\beta b') \in A+B,$$

since $\alpha a+\beta a' \in A$ and $\alpha b+\beta b' \in B$. □

We can often decompose vector spaces into sums of subspaces. However, decompositions are almost never unique. Indeed, each partition of any given spanning set yields another decomposition. That being said, some of these decompositions are nicer than others. Note that any sum $V = A+B$ induces a decomposition of vectors $x = a+b$. Under what conditions is this unique?

PROPOSITION 1.24. Let $V = A+B$. Each $x \in V$ can be written uniquely in the form $x = a+b$—$a \in A$ and $b \in B$—if and only if $A \cap B = \{0\}$.

Proof. **($\Longrightarrow$)** If there is non-zero $y \in A \cap B$, then $y = 0+y = y+0$, so y cannot be expressed uniquely in the form $y = a+b$. Taking the contrapositive yields the implication.

($\Longleftarrow$) Choose bases $\{a_1, \ldots, a_m\}$ and $\{b_1, \ldots, b_n\}$ of A and B respectively. Since $V = A+B$, these sets jointly span V. Suppose

$$\sum_{k=1}^{m} \alpha_k a_k + \sum_{k=1}^{n} \beta_k b_k = 0.$$

Then

$$\sum_{k=1}^{m} \alpha_k a_k = -\sum_{k=1}^{n} \beta_k b_k,$$

so $\sum \alpha_k a_k \in B$. However, $A \cap B = \{0\}$, so $\sum \alpha_k a_k = 0$. It follows from the linear independence of $\{a_i\}$ and $\{b_i\}$ that the α_k's and β_k's all vanish, hence $\{a_1, \ldots, a_m, b_1, \ldots, b_n\}$ is a basis of V. Therefore, each $x \in V$ can be written uniquely in the form

$$x = \sum_{k=1}^{m} \alpha_k a_k + \sum_{k=1}^{n} \beta_k b_k,$$

and we are done. □

We refer to subspaces A and B as *disjoint* if $A \cap B = \{0\}$. A corollary of this proof is that $\dim(A+B) = \dim A + \dim B$ when A and B are disjoint. We have a stronger result of this form:

THEOREM 1.25 (Inclusion-Exclusion for Subspaces)**.** If A and B are subspaces, then

$$\dim(A+B) = \dim A + \dim B - \dim(A \cap B).$$

Proof. Choose a basis $X = \{x_1, \ldots, x_p\}$ of $A \cap B$. Extend this to bases

$$X_A = \{x_1, \ldots, x_p, a_1, \ldots, a_m\} \text{ and } X_B = \{x_1, \ldots, x_p, b_1, \ldots, b_n\}$$

of A and B respectively. Clearly

$$X_A \cup X_B = \{x_1, \ldots, x_p, a_1, \ldots, a_m, b_1, \ldots, b_n\}$$

spans $A + B$. We will show that it is also linearly independent, hence a basis. Suppose

$$\sum_{k=1}^{p} \xi_k x_k + \sum_{k=1}^{m} \alpha_k a_k + \sum_{k=1}^{n} \beta_k b_k = 0. \tag{1.4}$$

Then

$$\sum_{k=1}^{p} \xi_k x_k + \sum_{k=1}^{m} \alpha_k a_k = -\sum_{k=1}^{n} \beta_k b_k \in A \cap B,$$

so there are scalars η_k such that

$$\sum_{k=1}^{p} \xi_k x_k + \sum_{k=1}^{m} \alpha_k a_k = \sum_{k=1}^{p} \eta_k x_k,$$

hence

$$\sum_{k=1}^{p} (\xi_k - \eta_k) x_k + \sum_{k=1}^{m} \alpha_k x_k = 0.$$

Since X_A is linearly independent, each of these coefficients vanishes—in particular, each $\alpha_k = 0$—so (1.4) becomes

$$\sum_{k=1}^{p} \xi_k x_k + \sum_{k=1}^{n} \beta_k b_k = 0.$$

By the linear independence of X_B, each ξ_k and β_k vanishes, so all of the coefficients in (1.4) vanish, *i.e.* $X_A \cup X_B$ is a basis of $A + B$. By the regular inclusion-exclusion principle,

$$|X_A \cup X_B| = |X_A| + |X_B| - |X_A \cap X_B| = |X_A| + |X_B| - |X|,$$

from which the result follows immediately. □

EXAMPLE 9. Consider a vector space V with basis $\{x_1, \ldots, x_n\}$. For each $k \in [n]$, let $S_k = \operatorname{span}(x_k)$. Clearly, $V = \sum_{k=1}^{n} S_k$. Since $x_1, \ldots, x_n$ are linearly independent, the subspaces S_k are pairwise disjoint, and each $x \in V$ admits a unique decomposition $x = \sum_{k=1}^{n} x'_k$ with each $x'_k \in S_k$. Since each x'_k is in the span of x_k, $x'_k = \alpha_k x_k$ for some unique α_k, so this decomposition becomes $x = \sum_{k=1}^{n} \alpha_k x_k$. That is, we have restated our characterization of bases in terms of one-dimensional subspaces, *i.e.* axes. ◇

2

TRANSFORMATIONS

The theory of vector spaces is greatly enhanced by the consideration of structure-preserving maps between them. Isomorphisms are particularly nice instances of these. In this chapter, we study a more general class of transformations and prove several results regarding their structure.

§2.1. LINEAR TRANSFORMATIONS

DEFINITION 2.1. Let V and W be vector spaces over a field K. A *linear transformation* is a function $T : V \to W$ satisfying the following identity:

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y). \tag{2.1}$$

If $V = W$, T is a *linear operator*.

We often omit parenthesis and composition symbols when working with linear transformations, so $T(x)$ and $S \circ T$ become Tx and ST. By a quick induction, (2.1) is equivalent to

$$T \sum_{k=1}^{n} \alpha_k x_k = \sum_{k=1}^{n} \alpha_k T x_k.$$

That is, linear transformations preserve linear combinations.

Brief reflection shows that linear combinations and compositions of linear transformations are themselves linear. In other words, the set $\mathscr{L}(V, W)$ of linear transformations $V \to W$ is itself a vector space. In the case $V = W$, we have an additional (non-commutative) multiplicative structure given by composition. To save time, we write $\mathscr{L}(V)$ for the space of operators $\mathscr{L}(V, V)$.

Part of the reason for studying linear transformations is that they are extremely easy to characterize. Suppose $T : V \to W$ is linear, and let $\{x_1, \ldots, x_n\}$ be a basis of V. Then, for all $\sum \alpha_k x_k \in V$,

$$T \sum_{k=1}^{n} \alpha_k x_k = \sum_{k=1}^{n} \alpha_k T x_k.$$

That is, the behavior of T is completely determined by its behavior on any given basis. This is easiest to see in the case $V = W$. Consider V with some fixed basis. Then T

may be viewed as moving each of the basis vectors of V (morally, each of the axes) and transforming every other point relative to these. Formally,

PROPOSITION 2.2. If V and W are vector spaces, if $\{x_1, \ldots, x_n\}$ is a basis in V, and if $y_1, \ldots, y_n$ are any (not necessarily distinct) n vectors in W, then there is a unique linear transformation $T : V \to W$ such that $Tx_k = y_k$ for all k.

Proof. We may take each vector in V in the form, uniquely-determined, $\sum \alpha_k x_k$. Define $T : V \to W$ by

$$T \sum_{k=1}^{n} \alpha_k x_k = \sum_{k=1}^{n} \alpha_k y_k.$$

Then T is linear, and $Tx_k = y_k$ for all k. To show that T is unique, suppose we have $S \in \mathscr{L}(V, W)$ with the same property. Then $(S - T)x_k = y_k - y_k = 0$ so, by linearity,

$$(S - T) \sum_{k=1}^{n} \alpha_k x_k = \sum_{k=1}^{n} \alpha_k (S - T) x_k = 0$$

for all $\sum \alpha_k x_k \in V$, *i.e.* $S - T = 0$. □

COROLLARY 2.3. If V and W are n and m-dimensional, then $\mathscr{L}(V, W)$ is nm-dimensional.

Proof. Let $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ be bases of V and W respectively. Then, for each $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$, let B_{ij} be the unique linear transformation defined by $B_{ij} x_k = \delta_{ik} y_j$ for all k. Suppose

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \xi_{ij} B_{ij} = 0.$$

Then, for each $1 \leqslant k \leqslant n$,

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{m} \xi_{ij} B_{ij} x_k = \sum_{i=1}^{n} \sum_{j=1}^{m} \xi_{ij} \delta_{ik} y_j = \sum_{j=1}^{m} \xi_{kj} y_j.$$

By the linear independence of Y, each ξ_{ij} vanishes, hence the B_{ij}'s are linearly independent. To show that they form a basis, let $T \in \mathscr{L}(V, W)$ be arbitrary. Being a linear transformation, T is uniquely determined by its values on X, so it is sufficient to show that there is some linear combination of $\{B_{ij}\}$ that agrees with T on this basis. For each x_k, there are scalars $\alpha_{k1}, \ldots, \alpha_{km}$ such that

$$Tx_k = \sum_{j=1}^{m} \alpha_{kj} y_j.$$

Consider the linear transformation given by

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} B_{ij}.$$

Then, for each x_k,

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_{ij}B_{ij}x_k = \sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_{ij}\delta_{ik}y_j = \sum_{j=1}^{m}\alpha_{kj}y_j = Tx_k.$$

We have shown that $\{B_{ij}\}$ is spanning, hence a basis. Since $|\{B_{ij}\}| = nm$, we are done. □

Before moving on, we ought to consider some examples. The reader should take a moment to verify the linearity of each.

EXAMPLE 10. For any vector space V, we have the *identity operator* $\mathbb{1}_V$ that sends each $x \in V$ to itself. Similarly, between any two vector spaces V and W we have the *zero operator* 0, which sends every vector to 0. A brief reflection shows that the zero operator is the additive identity of the space $\mathscr{L}(V, W)$. ◇

EXAMPLE 11. The map from $\mathbb{C}_1[t]$ to $\mathbb{C}_2[t]$ given by left-multiplication by t, *i.e.* $p(t) \mapsto tp(t)$, is linear. Choosing the basis $\{1, t\}$ of $\mathbb{C}_1[t]$, we may equivalently define this as the unique linear transformation that sends 1 to t and t to t^2. Note that the image of this tranformation in $\mathbb{C}_2[t]$ is the 2-dimensional subspace of polynomials with constant term zero. ◇

EXAMPLE 12. Let $V = A + B$, where A and B are disjoint subspaces. Then each $x \in V$ admits a unique decomposition $x = a + b$, with $a \in A$ and $b \in B$. Let $E : V \to V$ be the linear operator defined by $E(a + b) = a$, so E acts like the identity on A and the zero operator on B. This transformation is an example of a *projection*. ◇

EXAMPLE 13. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, then the *derivative of f at $p \in \mathbb{R}^n$*, denoted df_p, is the unique linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ satisfying

$$\lim_{h\to 0}\frac{|f(p+h) - f(p) - df_p(h)|}{|h|} = 0.$$

That is, df_p provides the best linear approximation to f near p. Furthermore, the function $d(\cdot)_p$ from the space of functions $\mathbb{R}^n \to \mathbb{R}^m$ differentiable at p to $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ is itself linear. ◇

EXAMPLE 14. Let $C^0[0, 1]$ denote the vector space of continuous functions $[0, 1] \to \mathbb{R}$. The *Volterra operator* is the function $V : C^0[0, 1] \to C^0[0, 1]$ defined by

$$(Vf)(x) = \int_0^x f(t)dt$$

for all $f \in C^0[0, 1]$ and $x \in [0, 1]$. As the name suggests, V is a linear operator. ◇

§2.2. AND THEIR STRUCTURE

For this section, fix a linear transformation $T : V \to W$. The structure of T is best studied in terms of subspaces associated to it. But first, we need to understand how subspaces behave under T.

As a brief review of terminology, the *image of A under T* is the set

$$T(A) = \{Tx : x \in A\},$$

and the *preimage of B under T* is the set

$$T^{-1}(B) = \{x \in V : Tx \in B\}.$$

PROPOSITION 2.4. The image or preimage of a subspace (of V or W respectively) under T is a subspace.

Proof. Suppose $A \subseteq V$ is a subspace. Clearly, $0 \in T(A)$, since $T0 = 0$. If $Tx, Ty \in T(A)$, then $\alpha x + \beta y \in A$, so

$$\alpha Tx + \beta Ty = T(\alpha x + \beta y) \in T(A).$$

Similarly, suppose $B \subseteq W$ is a subspace. Since $T0 = 0 \in B$, $0 \in T^{-1}(B)$. If $x, y \in T^{-1}(B)$, then

$$T(\alpha x + \beta y) = \alpha Tx + \beta Ty \in B,$$

so $\alpha x + \beta y \in T^{-1}(B)$. □

We want to quantify the injectivity and surjectivity of T using vector spaces. Starting with the latter, by definition, T is surjective if and only if $T(V) = W$. More generally, we might say that the larger $T(V)$ is, the more surjective T is. From the result above, we know that $T(V)$ is a subspace of W. This lends a natural notion of the size of $T(V)$:

DEFINITION 2.5. The *image* of T is the subspace $\operatorname{im}(T) = T(V)$. The *rank* of T, written $\operatorname{rank}(T)$, is the dimension of its image.

Quantifying injectivity is a little trickier. We may reframe the problem as follows: how can we vary x so that Tx remains constant? Note that $Tx = Ty$ if and only if $T(x - y) = 0$. Therefore, for any $x \in V$, $T(x + h) = Tx$ if and only if $Th = 0$; the space of all such h is the preimage of the zero vector.

DEFINITION 2.6. The *kernel* of T is the subspace $\ker(T) = T^{-1}(\{0\})$. The *nullity* of T, denoted $\operatorname{nullity}(T)$, is the dimension of its kernel.

Before proving any results, we reconsider some examples from the previous section.

EXAMPLE 15. Consider the projection E from **Example 12**. We saw that it acts like the identity on A and the zero operator on B. From this, we infer that $\operatorname{im}(E) = A$ and $\ker(E) = B$. ◇

EXAMPLE 16. Let V denote the Volterra operator from **Example 14**. Suppose $Vf = 0$, *i.e.* $(Vf)(x) = 0$ for all $x \in [0, 1]$. Since f is continuous, Vf is differentiable with $(Vf)'(x) = f(x)$ for all $x \in (0, 1)$. And since Vf is constant, $f(x) = (Vf)'(x) = 0$ for all $x \in (0, 1)$. Lastly, since f is continuous, $f(0) = f(1) = 0$, for otherwise we could choose some ε-neighborhood of 0 (resp. 1) in which f is non-zero, a contradiction. It follows that $f = 0$, so $\ker(Vf) = \{0\}$, *i.e.* V is injective.

Now, were $C^0[0, 1]$ finite-dimensional, V would also be surjective for reasons that will be discussed later. However, $C^0[0, 1]$ is not finite-dimensional, and V is not surjective since every function in its image is differentiable on the open interval $(0, 1)$, a property not shared by all continuous functions. ◇

We will use our new tools to provide another characterization of isomorphisms. Our main result on the structure of linear transformations, rank-nullity, appears in the next section.

PROPOSITION 2.7. A linear transformation $T : V \to W$ is an isomorphism if and only if it sends bases to bases.

Proof. **($\Longrightarrow$)** By definition, T is invertible, so its image is W, and its kernel is trivial. Let $X = \{x_1, \ldots, x_n\}$ be a basis of V. Then each element of $\mathrm{im}(T)$ is of the form

$$T \sum_{k=1}^{n} \alpha_k x_k = \sum_{k=1}^{n} \alpha_k T x_k,$$

i.e. W is spanned by the vectors $Tx_1, \ldots, Tx_n$. Furthermore, these vectors are linearly independent. Indeed, if

$$\sum_{k=1}^{n} \alpha_k T x_k = T \sum_{k=1}^{n} \alpha_k x_k = 0,$$

then $\sum \alpha_k x_k \in \ker(T)$, which is trivial, so

$$\sum_{k=1}^{n} \alpha_k x_k = 0.$$

By the linear independence of X, each α_k vanishes, so $\{Tx_1, \ldots, Tx_n\}$ is a basis of W.

($\Longleftarrow$) Choose a basis $X = \{x_1, \ldots, x_n\}$ of V, so $T(X) = \{Tx_1, \ldots, Tx_n\}$ is a basis of W. Clearly, $\mathrm{im}(T) = W$, so we need only show that T has a trivial kernel. Suppose

$$0 = T \sum_{k=1}^{n} \alpha_k x_k = \sum_{k=1}^{n} \alpha_k T x_k.$$

Then, by the linear independence of $T(X)$, each α_k vanishes, hence $\sum \alpha_k x_k = 0$. It follows that $\ker(T) = \{0\}$. $\square$

§2.3. QUOTIENT SPACES

We've seen that the nice linear transformations are those with trivial kernels, *i.e.* the injective ones. They are nice because they preserve linear independence, so we may find unique representations of the vectors in their image. In this section, we will construct spaces through which we can naturally factor general transformations into injective ones.

Consider a linear transformation $T : V \to W$ and a subspace $M \subseteq \ker(T)$. We want to cut M out of T's kernel. We can do this by taking V and collapsing M to the origin. Formally, this looks like a space $Q(M)$ together with a surjective linear map $\pi_M : V \to Q(M)$, called the *canonical projection*, satisfying $\ker(\pi_M) = M$. We want to extend T naturally to a transformation $\widetilde{T} : Q(M) \to W$. That is, we demand that $Q(M)$ and π_M be such that, for every $T \in \mathscr{L}(V, W)$ with $\ker(T) \supseteq M$, the following diagram commutes:

$$\begin{tikzcd} Q(M) \arrow[dr, "\widetilde{T}"] & \\ V \arrow[u, "\pi_M"] \arrow[r, "T"'] & W \end{tikzcd} . \tag{2.2}$$

This is called the *universal property of quotients*, and it defines $Q(M)$ up to isomorphism. Indeed, if we have $(Q(M), \pi_M)$ and $(R(M), \tau_M)$ satisfying (2.2), then we may factor π_M through $R(M)$ and τ_M through $Q(M)$, obtaining the commutative diagram

$$\begin{tikzcd} Q(M) \arrow[r, "\widetilde{\tau}_M"] & R(M) \arrow[d, "\widetilde{\pi}_M"] \\ V \arrow[u, "\pi_M"] \arrow[ur, "\tau_M"'] \arrow[r, "\pi_M"'] & Q(M) \end{tikzcd} .$$

Therefore $\pi_M = \widetilde{\pi}_M \circ \widetilde{\tau}_M \circ \pi_M$. However, $\pi_M = \mathbb{1}_{Q(M)} \circ \pi_M$ so, by uniqueness, $\widetilde{\pi}_M \circ \widetilde{\tau}_M = \mathbb{1}_{Q(M)}$. Swapping $(Q(M), \pi_M)$ and $(R(M), \tau_M)$ in the diagram yields $\widetilde{\tau}_M \circ \widetilde{\pi}_M = \mathbb{1}_{R(M)}$, hence $\widetilde{\pi}_M$ and $\widetilde{\tau}_M$ are inverses, and $Q(M)$ is isomorphic to $R(M)$.

Having shown that the universal property provides a complete definition of quotients, we will now construct one. If we take a subspace M and collapse it to the origin, every other vector should be identified with its displacement from M. In the case of the Cartesian plane, if M is some line through the origin, we are interested in the vector space formed by the set of lines parallel to M.

More generally, fix a vector space V and a subspace M. Two vectors x and y lie in a hyperplane parallel to M if one can be obtained from the other by translating by some element of M, *i.e.* if $x = y + m$ or, equivalently, if $x - y \in M$. Write $x \sim y$ when $x - y \in M$. Then $\sim$ is an equivalence relation on V. Note that for each $y \in [x]_\sim$, $y - x = m \in M$ and $y = x + m$. That is, $[x]_\sim = x + M = \{x + m : m \in M\}$. The set $x + M$ is called the *coset* associated with x mod M.

DEFINITION 2.8. Let V be a vector space and M a subspace of V. The *quotient space of V mod M*, denoted V/M, is the vector space of $\sim$ equivalence classes of V with addition and scaling defined by

$$\alpha(x + M) + \beta(y + M) = (\alpha x + \beta y) + M.$$

The *canonical projection* $\pi_M : V \to V/M$ is given by $\pi_M(x) = x + M$.

If $x + M = x' + M$ and $y + M = y' + M$, *i.e.* if $x - x'$ and $y - y'$ are vectors in M, then

$$(\alpha x + \beta y) - (\alpha x' + \beta y') = \alpha(x - x') + \beta(y - y') \in M,$$

so

$$(\alpha x + \beta y) + M = (\alpha x' + \beta y') + M.$$

It follows that π_M is linear, and the operations given above are well-defined.

We will now show that the construction V/M satisfies the universal property of quotients.

THEOREM 2.9 (Factorization Theorem)**.** For every linear transformation $T : V \to W$ with $\ker(T) \supseteq M$, there is a unique transformation $T/M : V/M \to W$ such that $T = T/M \circ \pi_M$.

Proof. Let $T : V \to W$ be any linear transformation satisfying $\ker(T) \supseteq M$. Note that if $x - y \in M$, then $Tx - Ty = T(x - y) = 0$. Therefore, we may define a transformation $T/M : V/M \to W$ by $T/M(x + M) = Tx$. Then

$$\alpha(x + M) + \beta(y + M) \xmapsto{T/M} \alpha Tx + \beta Ty = \alpha T/M(x + M) + \beta T/M(y + M),$$

so T/M is linear. Clearly, $T = T/M \circ \pi_M$, so we need only show that T/M is the unique transformation satisfying this identity. Suppose we have $\widetilde{T} : V/M \to W$ such that $T = \widetilde{T} \circ \pi_M$. Then

$$\widetilde{T}(x + M) = (\widetilde{T} \circ \pi_M)x = Tx = T/M(x + M)$$

for all $x + M \in V/M$, so $\widetilde{T} = T/M$. $\square$

We introduced quotient spaces as a means of making kernels trivial. We are now ready to show that they can.

COROLLARY 2.10 (First Isomorphism Theorem)**.** If $T : V \to W$ is linear, then the transformation $x + \ker(T) \mapsto Tx$ is an isomorphism $V/\ker(T) \to \operatorname{im}(T)$.

Proof. Let $M = \ker(T)$, and consider the transformation T/M. If $x + M \in \ker(T/M)$, then $Tx = 0$, so $x \in M$. Thus, $x + M = 0 + M$, and the $\ker(T/M)$ is trivial. Clearly, $\operatorname{im}(T/M) = \operatorname{im}(T)$, so T/M is an isomorphism, and we are done. $\square$

This result also gives us an easy expression for the dimension of quotient spaces.

COROLLARY 2.11. If V is a vector space, M a subspace, then $\dim V/M = \dim V - \dim M$.

Proof. Choose a basis $\{x_1, \ldots, x_m\}$ of M, and extend it to a basis $\{x_1, \ldots, x_n\}$ of V. Let $T : V \to V$ be the linear transformation defined by $Tx_k = 0$ for $1 \leqslant k \leqslant m$ and $Tx_k = x_k$ for $m < k \leqslant n$. Then $\ker(T) = M$ and $\operatorname{im}(T) = \operatorname{span}(x_{m+1}, \ldots, x_n)$ so, by the first isomorphism theorem, $\dim V/M = \dim(\operatorname{im}(T)) = n - m$. $\square$

We are now ready to prove our main result regarding the structure of linear transformations.

THEOREM 2.12 (Rank-Nullity)**.** If $T \in \mathscr{L}(V, W)$, then

$$\dim V = \operatorname{rank}(T) + \operatorname{nullity}(T).$$

Proof. By the first isomorphism theorem, $\operatorname{im}(T) \cong V/\ker(T)$ so, taking the dimension of both sides, we have $\operatorname{rank}(T) = \dim V - \operatorname{nullity}(T)$. $\square$

When constructing V/M, we visualized it as the set of hyperplanes parallel to M. In particular, let's consider the case where $V = \mathbb{R}^2$ and M is some line through the origin. If N is some complementary subspace (so $M \cap N = \{0\}$ and $M + N = V$), then N is a line through the origin distinct from M. Consequently, it is not parallel to M nor to any of the lines parallel to M, *i.e.* the elements of V/M. As such, it intersects each coset of M exactly once. This yields a one-to-one correspondence; in fact, it yields an isomorphism.

PROPOSITION 2.13. If V is a vector space and M a subspace, then any subspace complementary to M is isomorphic to V/M.

Proof. Let N be any complement of M. Let $\pi_M|_N : N \to V/M$ be the restriction of π_M to N. Since $\ker(\pi_M) = M$, and since M and N are disjoint, the restriction $\pi_M|_N : N \to V/M$ has trivial kernel and is injective. We have $V = M + N$ so, for each $x + M \in V/M$, $x = m + n$ for some $m \in M$ and $n \in N$. Then $\pi_M|_N(n) = n + M = n + m + M = x + M$, hence $\pi_M|_N$ is surjective, and we are done. □

§2.4. DIRECT SUMS

In §**1.5**, we introduced sums of subspaces, which allow us to glue together subspaces of a vector space. In general, we want to be able to combine spaces simply. The trouble is that, given vector spaces M and N, the expression $M + N$ only makes sense if we have an ambient space V containing both M and N as subspaces. In this section, we provide a general construction for the sum of arbitrary families of vector spaces (over a common field K).

We will start by characterizing what a sum of vector spaces ought to be. As with quotients, the definition will be by universal property. Consider vector spaces X and Y and linear transformations $T_X : X \to W$ and $T_Y : Y \to W$. Rather than thinking about gluing X and Y together, consider combining T_X and T_Y. If $T : C(X,Y) \to W$ is obtained by gluing together T_X and T_Y, then we should be able to appropriately restrict T to recover each of these transformations. That is, we want to embed X and Y into $C(X,Y)$ such that T acts like its constituent transformations on the corresponding embedded spaces. Then $C(X,Y)$ contains subspaces that are naturally identified with X and Y, *i.e.* it serves as an ambient space for the sum $X + Y$. By demanding that T is uniquely determined by T_X and T_Y, we are saying that it is determined by its behavior on X and Y, so these subspaces together encode all of $C(X,Y)$. For that reason, it is reasonable to call $C(X,Y)$ the sum of X and Y.

Formally, given vector space X and Y, a vector space $C(X,Y)$ together with embeddings $\iota_X : X \to C(X,Y)$ and $\iota_Y : Y \to C(X,Y)$—the *canonical inclusions*—satisfies the *universal property of direct sums* if, for any $T_X \in \mathscr{L}(X,W)$ and $T_Y \in \mathscr{L}(Y,W)$, there is a unique linear transformation $T : C(X,Y) \to W$ such that the following diagram commutes:

$$\begin{tikzcd} & W & \\ X \arrow[ur, "T_X"] \arrow[r, "\iota_X"'] & C(X,Y) \arrow[u, "T"] & Y \arrow[ul, "T_Y"'] \arrow[l, "\iota_Y"] \end{tikzcd} . \tag{2.3}$$

A similar argument as with the universal property of quotients shows that (2.3) defines $C(X,Y)$ up to isomorphism. Constructing $C(X,Y)$ is easy:

DEFINITION 2.14. If X and Y are vector spaces over a field K, their *direct sum*, denoted $X \oplus Y$, is the vector space consisting of the set $X \times Y$ under the operations defined by

$$\alpha(x_1, y_1) + \beta(x_2, y_2) = (\alpha x_1 + \beta x_2, \alpha y_1 + \beta y_2).$$

The *canonical inclusions* are given by $x \xmapsto{\iota_X} (x, 0)$ and $y \xmapsto{\iota_Y} (0, y)$.

The reader will verify that $X \oplus Y$ is a vector space.

THEOREM 2.15. If $A : X \to W$ and $B : Y \to W$ are linear transformations, there is a unique transformation $A \oplus B : X \oplus Y \to W$ such that $(A \oplus B) \circ \iota_X = A$ and $(A \oplus B) \circ \iota_Y = B$. That is, $X \oplus Y$ satisfies the universal property of sums.

Proof. Suppose we have linear $T : X \oplus Y \to W$ such that $T \circ \iota_X = A$ and $T \circ \iota_Y = B$. Then $T(x, 0) = (T \circ \iota_X)x = Ax$ and $T(0, y) = (T \circ \iota_Y)y = By$ for all $x \in X$ and $y \in Y$. By linearity, $T(x, y) = T(x, 0) + T(0, y) = Ax + By$ for all $(x, y) \in X \oplus Y$. We define $A \oplus B : X \oplus Y \to W,\ (x, y) \mapsto Ax + By$, and we are done. $\square$

When $A : X \to X$ and $B : Y \to Y$ are linear operators, we write

$$A \oplus B = (\iota_X \circ A) \oplus (\iota_Y \circ B),$$

so $(A \oplus B)(x + y) = (Ax, By)$.

As a generalization of sums of subspaces, the dimension of direct sums is unsurprising.

PROPOSITION 2.16. $\dim(X \oplus Y) = \dim X + \dim Y$.

Proof. Note that $\iota_X(X)$ and $\iota_Y(Y)$ are disjoint, and $\iota_X(X) + \iota_Y(Y) = X \oplus Y$. Since ι_X and ι_Y are injective, $\dim(\iota_X(X)) = \dim X$ and $\dim(\iota_Y(Y)) = \dim Y$. Thus,

$$\dim(X \oplus Y) = \dim\left(\iota_X(X) + \iota_Y(Y)\right) = \dim X + \dim Y.$$

$\square$

In this proof, we saw that X and Y may be viewed as disjoint subspaces of $X \oplus Y$. The converse also holds: if M and N are disjoint subspaces of some vector space V, then $M + N$ is naturally isomorphic to $M \oplus N$. This follows from the fact that each vector in $M + N$ can be represented uniquely in the form $m + n$, with $m \in M$ and $n \in N$, and that the correspondence $m + n \mapsto (m, n)$ is linear. On account of this, we write $M + N$ as $M \oplus N$ when M and N are disjoint. While this does overload notation, such an ambiguity is acceptable up to canonical isomorphism.

The direct sum operation is commutative, *i.e.* $X \oplus Y \cong Y \oplus X$, as can been seen by verifying that $(x, y) \mapsto (y, x)$ is an isomorphism. Similarly,

$$(X \oplus Y) \oplus Z \cong X \oplus (Y \oplus Z),$$

with the isomorphism given by $((x, y), z) \mapsto (x, (y, z))$. Consequently, we may unambiguously refer to direct sums of finite families of vector spaces. The universal property generalizes as well: Fix a family of vector spaces $(X_k)_{k \in [n]}$. For any family of linear transformations $(T_k : X_k \to W)_{k \in [n]}$, there is a unique linear transformation $T : \bigoplus_{k=1}^n X_k \to W$ such that the following diagram commutes for each k:

$$\begin{tikzcd} & W \\ X_k \arrow[ur, "T_k"] \arrow[r, "\iota_k"'] & \bigoplus_{k=1}^n X_k \arrow[u, "T"'] \end{tikzcd},$$

where ι_k is the canonical inclusion of X_k.

We can further generalize to arbitrary families, with the only additional detail being that the underlying set of $\bigoplus_{k \in I} X_k$ is the set of all $(x_k)_{k \in I} \in \prod_{k \in I} X_k$ such that $x_k = 0$ for all but finitely many k. However, the study of such sums is beyond the scope of these notes.

Some remark is due on the connection between direct sums and quotients. The space $X \oplus Y$ is, intuitively-speaking, the smallest vector space containing X and Y as complementary subspaces. Consequently, if we quotient out one of these constituents, we should be left with the other.

PROPOSITION 2.17. The vector space $(X \oplus Y)/Y$ is isomorphic to X.

Proof. Consider the linear projection $\varphi : X \oplus Y \to X, (x, y) \mapsto x$. Clearly, $\ker(\varphi) = Y$ and $\operatorname{im}(\varphi) = X$ so, by the first isomorphism theorem, $\varphi/Y : (X \oplus Y)/Y \to X$ is an isomorphism. □

§2.5. DUAL SPACES

Consider the real line $\mathbb{R}^1$. How can we impose a measurement system on this space? Well, visualizing $\mathbb{R}^1$ as the typical real number line, we can imagine laying a two-way infinite ruler on it, with evenly-spaced positive ticks extending in one direction, negatives in the other, and with the 0 mark lying on top of the 0 vector.

Let $\psi : \mathbb{R}^1 \to \mathbb{R}$ be the function sending each vector to its value on the ruler. We can say a few things about ψ. In the first place, the ruler's value varies evenly, so adding a vector h to $x \in \mathbb{R}^1$ displaces its value under ψ by $\psi(h)$, *i.e.* $\psi(x + h) = \psi(x) + \psi(h)$. By similar reasoning, scaling x scales its value under ψ proportionally: $\psi(\alpha x) = \alpha\psi(x)$. It follows that ψ is linear. That is, we may think of the rulers we can place on $\mathbb{R}^1$ as corresponding to linear maps $\mathbb{R}^1 \to \mathbb{R}$ and, in doing so, we see that they form a vector space.

Generalizing this, we might say that the systems of measurement on a space V over a field K (and here we mean the algebraically-nice systems) correspond to the linear transformations $V \to K$. On account of this analogy, we give such transformations a name:

DEFINITION 2.18. A *linear functional* on a vector space V (over a field K) is a linear transformation $V \to K$. The *dual space* of V is the vector space of linear functionals $V' = \mathscr{L}(V, K)$.

In the starting example, each linear functional in $(\mathbb{R}^1)'$ is uniquely determined by its value at any fixed non-zero vector x_1. Consequently, the linear functional φ defined by $\varphi(x_1) = 1$ constitutes a basis of $(\mathbb{R}^1)'$. More generally, we know that a linear functional, being a transformation, is determined by its value on a basis. This allows us to push bases from a space to its dual.

PROPOSITION 2.19. If $X = \{x_1, \ldots, x_n\}$ is a basis of V, then there is a unique basis, called the *dual basis* of X, $X' = \{x^1, \ldots, x^n\}$ of V' such that $x^i(x_j) = \delta_{ij}$ for all $1 \leqslant i, j \leqslant n$. Consequently, $\dim V' = \dim V$.

Proof. Define each x^i by $x^i(x_j) = \delta_{ij}$ for each $x_j \in X$. Suppose

$$\sum_{k=1}^{n} \alpha_k x^k = 0.$$

Then, for each $1 \leqslant i \leqslant n$,

$$0 = \sum_{k=1}^{n} \alpha_k x^k(x_i) = \sum_{k=1}^{n} \alpha_k \delta_{ki} = \alpha_i,$$

so $x^1, \ldots, x^n$ are linearly independent. Furthermore, for each $\varphi \in V'$, let

$$\psi = \sum_{k=1}^{n} \varphi(x_k) x^k.$$

Then

$$\psi(x_i) = \sum_{k=1}^{n} \varphi(x_k) x^k(x_i) = \sum_{k=1}^{n} \varphi(x_k) \delta_{ki} = \varphi(x_i)$$

for each $x_i \in X$. Since linear transformations are uniquely determined by their values on $x_1, \ldots, x_n$, it follows that $\psi = \varphi$, so $x^1, \ldots, x^n$ spans V'. □

Unfortunately, V is not *naturally* isomorphic to V': in order to construct an isomorphism $V \to V'$, we must choose a basis. There is, however, a natural isomorphism between V and the double dual V''. The idea of the correspondence is that we can measure linear functionals by looking at their value at a point. It is a nice property of finite-dimensional vector spaces that this is a complete characterization.

THEOREM 2.20 (Reflexivity)**.** To each $\mu \in V''$, there corresponds a unique $x \in V$ such that $\mu(\varphi) = \varphi(x)$ for all $\varphi \in V'$. The function $\mu \mapsto x$ is an isomorphism between V'' and V.

Proof. Choose a basis $\{x_1, \ldots, x_n\}$ of V, and let $\{x^1, \ldots, x^n\}$ be the corresponding dual basis in V'. Let $\mu \in V''$ be arbitrary. For existence, let

$$x = \sum_{k=1}^{n} \mu(x^k) x_k.$$

Then, for any $\varphi = \sum \alpha_k x^k \in V'$,

$$\mu(\varphi) = \sum_{k=1}^{n} \alpha_k \mu(x^k)$$

and

$$\varphi(x) = \sum_{k=1}^{n} \sum_{i=1}^{n} \alpha_k \mu(x^i) x^k(x_i) = \sum_{k=1}^{n} \sum_{i=1}^{n} \alpha_k \mu(x^i) \delta_{ki} = \sum_{k=1}^{n} \alpha_k \mu(x^k),$$

so $\mu(\varphi) = \varphi(x)$. For uniqueness, suppose we have y such that, for all $\varphi \in V'$, $\mu(\varphi) = \varphi(y)$. Then $\varphi(x) - \varphi(y) = \varphi(x - y) = 0$ for all $\varphi \in V'$. If $x - y$ were non-zero, then we could choose φ such that $\varphi(x - y) \neq 0$. Therefore, $x = y$.

Consider the map $T : V'' \to V$ such that $\mu(\varphi) = \varphi(T\mu)$ for all $\varphi \in V'$. Then T is linear, since, for any $\mu, \eta \in V''$,

$$(\alpha\mu + \beta\eta)(\varphi) = \alpha\mu(\varphi) + \beta\eta(\varphi) = \varphi(\alpha T\mu + \beta T\eta).$$

Furthermore, T is injective, since $T\mu = T\eta$ implies

$$\mu(\varphi) = \varphi(T\mu) = \varphi(T\eta) = \eta(\varphi)$$

for all $\varphi \in V'$. Lastly, T is surjective as, for each $x \in V$, we may define $\mu_x \in V''$ by $\mu_x(\varphi) = \varphi(x)$; clearly $T\mu_x = x$. □

In light of this result, we identify V with V''.

We turn our attention now to subspaces of dual spaces. In particular, we want to find a natural correspondence between subspaces of V and subspaces of V'. Fix a subspace M. The naïve approach is to embed M' in V'. The problem with this is a lack of canonicity.

Try sending each $\varphi|_M \in M'$ to $\varphi \in V'$ such that φ acts like $\varphi|_M$ on M and vanishes everywhere else. The only way to do this so that φ is linear is to choose a decomposition $V = M \oplus N$ and define $\varphi = \varphi|_M \oplus 0_N$. The problem is that the resulting φ is dependent on our choice of N, of which there are generally many. Indeed, if we have $V = M \oplus N_1 = M \oplus N_2$ with $N_1 \neq N_2$, then we can choose a non-zero vector $x \in N_1 \setminus N_2$. By assumption $x = m + n_2$ for some $m \in M$ and $n_2 \in N_2$. Furthermore, since $x \notin N_2$, $m \neq 0$. Choose $\varphi|_M \in M'$ such that $\varphi|_M(m) \neq 0$, and let $\varphi_1 = \varphi|_M \oplus 0_{N_1}$ and $\varphi_2 = \varphi|_M \oplus 0_{N_2}$. Since $x \in N_1$, $\varphi_1(x) = 0$. However, $\varphi_2(x) = \varphi|_M(m) \neq 0$. Thus, $\varphi_1 \neq \varphi_2$, and our proposed embedding of M' is dependent on an arbitrary choice.

Perhaps a different approach is needed. Rather than considering functionals that are interesting on M, let's consider ones that are interesting away from M.

DEFINITION 2.21. Let M be a subspace of a vector space V. The *annihilator* of M is the subspace $M^0 = \{\varphi \in V' : \varphi(m) = 0 \text{ for all } m \in M\}$.

Reflexivity carries over to annihilators. The following result holds under the identification $V'' = V$.

PROPOSITION 2.22. If M is a subspace of V, then $M^{00} = M$.

Proof. If $y \in M^{00}$, then $\varphi(y) = y(\varphi) = 0$ for all $\varphi \in M^0$, so $y \in M$. Indeed, if $y \notin M$, then we could choose a linear functional that vanishes on M, so is in M^0, but does not vanish at y. Conversely, if $x \in M$, then $x(\varphi) = \varphi(x) = 0$ for all $\varphi \in M^0$, so $x \in M^{00}$. □

Consequently, the map $M \mapsto M^0$ provides a contravariant correspondence between subspaces of V and subspaces of V'.

The space M^0 looks like V' ignoring M. As V/M is obtained by forcing M to be trivial, the next result follows readily from its universal property:

PROPOSITION 2.23. If M is an m-dimensional subspace of an n-dimensional vector space V, then $(V/M)' \cong M^0$. Consequently, $\dim M^0 = n - m$.

Proof. If $\varphi \in M^0$, then $M \subseteq \ker(\varphi)$, so we may define a map $M^0 \to (V/M)'$, $\varphi \mapsto \varphi/M$. Let π_M denote the canonical projection onto M. For all $\varphi, \psi \in M^0$ and all $x + M \in V/M$,

$$(\alpha\varphi/M + \beta\psi/M)(x + M) = (\alpha\varphi x + \beta\psi x) + M = (\alpha\varphi + \beta\psi)/M(x + M),$$

so $\alpha\varphi/M + \beta\psi/M = (\alpha\varphi + \beta\psi)/M$, and the map $\varphi \mapsto \varphi/M$ is linear. If $\varphi/M = \psi/M$, then $\varphi = \varphi/M \circ \pi_M = \psi/M \circ \pi_M = \psi$, so our transformation is injective. Furthermore, for each $\xi \in (V/M)'$, let $\varphi = \xi \circ \pi_M$. Then $\varphi/M = \xi$, so our transformation is surjective, hence an isomorphism. □

Lastly, we can lift transformations between spaces to transformations between duals.

DEFINITION 2.24. If $A : V \to W$ is linear, then the *adjoint* of A is the linear transformation $A' : W' \to V'$ defined by $A'\varphi(x) = \varphi(Ax)$.

The map $A \mapsto A'$ has some nice algebraic properties.

PROPOSITION 2.25. Let $A, B \in \mathscr{L}(V, W)$. Then

1. In the case $V = W$, $0'_V = 0_{V'}$ and $\mathbb{1}'_V = \mathbb{1}_{V'}$.

2. $(\alpha A + \beta B)' = \alpha A' + \beta B'$,

3. $(AB)' = B'A'$,

4. If A is invertible, then so is A', and $(A^{-1})' = (A')^{-1}$,

5. $A'' = A$.

Proof. For any $\varphi \in W'$ and any $x \in V$,

1. $0'_V\varphi(x) = \varphi(0) = 0$ and $\mathbb{1}'_V\varphi(x) = \varphi(x)$.

2. $(\alpha A + \beta B)'\varphi(x) = \alpha\varphi(Ax) + \beta\varphi(Bx) = (\alpha A' + \beta B')\varphi(x)$.

3. $(AB)'\varphi(x) = \varphi(ABx) = (A'\varphi)(Bx) = (B'A')\varphi(x)$.

4. By **1.** and **3.**, $(AA^{-1})' = (A^{-1})'A' = \mathbb{1}_{W'}$ and $(A^{-1}A)' = A'(A^{-1})' = \mathbb{1}_{V'}$.

5. By reflexivity, $A''x(\varphi) = \varphi(A''x)$. Then $\varphi(A''x) = A'\varphi(x) = \varphi(Ax)$ for all $x \in V$ and $\varphi \in V'$, so $A'' = A$.

□

To conclude, we will prove an important structure theorem regarding T and T':

THEOREM 2.26. If $T : V \to W$ is linear, then $\ker(T') = \operatorname{im}(T)^0$. Consequently, $\operatorname{rank}(T) = \operatorname{rank}(T')$.

Proof. By definition, $\varphi \in \ker(T')$ if and only if $T'\varphi(x) = \varphi(Tx) = 0$ for all $x \in V$. That is, $\varphi \in \ker(T')$ if and only if $\varphi(y) = 0$ for all $y \in \operatorname{im}(T)$, *i.e.* $\varphi \in \operatorname{im}(T)^0$. Taking the dimension of both sides and applying rank nullity, we have

$$\dim W - \operatorname{rank}(T') = \dim W - \operatorname{rank}(T),$$

so $\operatorname{rank}(T') = \operatorname{rank}(T)$. □

§2.6. PROJECTIONS

In this section, we consider a class of non-invertible transformations that emerge from direct sum decompositions. For our discussion, fix a vector space $V = M \oplus N$.

DEFINITION 2.27. The *projection onto M along N* is the linear operator $P_{M,N} \in \mathscr{L}(V)$ defined by $P_{M,N}(m+n) = m$.

The action of $P_{M,N}$ is best visualized in the 3-dimensional case, where M is a plane and N is a line. Then $P_{M,N}(x)$ is the shadow on M cast by a point x on account of a spotlight directed along N. If x lies on M, then it is its own shadow. This observation lends a complete characterization of projections:

PROPOSITION 2.28. A linear operator $P : V \to V$ is a projection if and only if $P^2 = P$.

Proof. **($\Longrightarrow$)** If $P = P_{A,B}$ for some A and B satisfying $A \oplus B = V$, then $P^2(a+b) = P(a) = a = P(a+b)$ for all $a + b \in V$, so $P^2 = P$.

($\Longleftarrow$) If $x = Py \in \operatorname{im}(P) \cap \ker(P)$, then $0 = Px = P^2y = Py = x$. Thus $\operatorname{im}(P)$ and $\ker(P)$ are disjoint. By inclusion-exclusion and rank-nullity,

$$\dim(\operatorname{im}(P) \oplus \ker(P)) = \operatorname{rank}(P) + \operatorname{nullity}(P) = \dim V,$$

so $V = \operatorname{im}(P) \oplus \ker(P)$. For all $Py + x \in V$ with $Py \in \operatorname{im}(P)$ and $x \in \ker(P)$,

$$P(Py + x) = P^2y = Py.$$

Therefore, P is the projection onto $\operatorname{im}(P)$ along $\ker(P)$. □

In this proof, we saw that P projects onto its image along its kernel. It's an easy exercise to show that this holds in general; that is, $\operatorname{im}(P_{M,N}) = M$ and $\ker(P_{M,N}) = N$. It is a particularly nice property of projections that their images and kernels are disjoint; this does not hold for transformations in general.

PROPOSITION 2.29. If $V = M \oplus N$, then $P'_{M,N} = P_{N^0,M^0}$.

Proof. Let $\varphi \in V'$ be arbitrary. To begin,

$$P'_{M,N}\varphi(x) = \varphi(P_{M,N}(x)) = \varphi(P^2_{M,N}(x)) = (P'_{M,N})^2\varphi(x)$$

for all $x \in V$, so $P'_{M,N} = (P'_{M,N})^2$ is a projection.

As to what $P'_{M,N}$ projects onto, note that

$$P'_{M,N}\varphi(n) = \varphi(P_{M,N}n) = \varphi(0) = 0$$

for all $n \in N$, so $\operatorname{im}(P'_{M,N}) \subseteq N^0$. Conversely, if $\psi \in N^0$, then

$$P'_{M,N}\psi(m+n) = \psi(m) = \psi(m+n)$$

for all $m + n \in V$, so $\psi = P'_{M,N}\psi \in \operatorname{im}(P'_{M,N})$. Thus, $\operatorname{im}(P'_{M,N}) = N^0$. As for the kernel, if $\varphi \in M^0$, then

$$P'_{M,N}\varphi(m+n) = \varphi(m) = 0$$

for all $m+n \in V$, so $P'_{M,N}\varphi = 0$, hence $M^0 \subseteq \ker(P'_{M,N})$. Conversely, if $P'_{M,N}\psi = 0$, then $\psi(m) = 0$ for all $m \in M$, so $\psi \in M^0$. It follows that $\ker(P'_{M,N}) = M^0$, so $P'_{M,N} = P_{N^0,M^0}$. □

COROLLARY 2.30. If $V = M \oplus N$, then $V' = M^0 \oplus N^0$.

Proof. Let P be the projection onto M along N, so $P' = P_{N^0,M^0}$. Since P' is a projection, $\operatorname{im}(P') = N^0$ and $\ker(P') = M^0$ are disjoint. Furthermore, by rank-nullity,

$$\dim(\operatorname{im}(P') \oplus \ker(P')) = \operatorname{rank}(P') + \operatorname{nullity}(P') = \dim V',$$

hence $V' = \operatorname{im}(P') \oplus \ker(P')$. □

3

INVARIANTS OF OPERATORS

If two spaces are essentially the same, intuition says we can transform them in essentially the same ways. Consider vector space V and W. Any isomorphism $T : V \to W$ induces an isomorphism $\mathscr{L}(V) \to \mathscr{L}(W)$ given by $A \mapsto TAT^{-1}$. Since $TABT^{-1} = (TAT^{-1})(TBT^{-1})$, this isomorphism preserves composition. This backs our intuition: isomorphic vector spaces are acted on by the same operators.

However, this sameness should be taken with a grain of salt; $\mathscr{L}(V)$ and $\mathscr{L}(W)$ can be the same in at least as many different ways as there are isomorphisms between the spaces they act on. This raises the question: given operators $A \in \mathscr{L}(V)$ and $B \in \mathscr{L}(W)$, under what conditions is there an isomorphism $T : V \to W$ that identifies the two, *i.e.* $A = TBT^{-1}$? If there is such an isomorphism, we say that A and B are *similar*.

Not all operators are similar. For example, operators of different ranks are not. In general, it is difficult to prove similarity. On the other hand, there are myriad ways to show dissimilarity. These take the form of *similarity invariants*, which are objects $S(A)$ associated to operators A with the property that $S(TAT^{-1}) = S(A)$. Rank is one example and, in this chapter, we construct others. Each measures a different aspect of a linear operator, and is interesting in its own right.

Much of this chapter is laying groundwork. In Chapter 5, we will develop methods for computing similarity invariants.

§3.1. MULTILINEAR MAPS

In general, we can study linear operators by looking at how they transform collections of vectors. For example, an operator is invertible if and only if it preserves bases. As such, in search of means of measuring operators, we consider functions that eat collections of vectors.

DEFINITION 3.1. Let $V_1, \dots, V_m$ and W be vector spaces over a field K. A function $\omega : \prod_{k=1}^{m} V_k \to W$ is *m-linear* if, fixing $m-1$ of its arguments, the map

$$V_k \to W,\ x \mapsto \omega(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_m)$$

is linear for each k. In other words, ω is m-linear if it is linear in each of its m components.

When $W = K$, we call ω an m-linear *form*, or m-form. When $m = 2$, we say ω is *bilinear*. Clearly, 1-linear maps are linear transformations, and 1-forms are linear functionals.

EXAMPLE 17. The *evaluation map* is the bilinear form $V' \times V \to K$ given by $(\varphi, x) = \varphi(x)$. More generally, given V and W, the mapping $\mathscr{L}(V, W) \times V \to W$ given by $(A, x) = Ax$ is bilinear. $\diamond$

EXAMPLE 18. For any m linear functionals $\varphi_1, \ldots, \varphi_m \in V'$, the map

$$V^m \to K,\ (x_1, \ldots, x_m) \mapsto \prod_{k=1}^{m} \varphi_k(x_k)$$

is an m-linear form. In general, if $\omega : \prod_{k=1}^{p} U_k \to K$ and $\eta : \prod_{k=1}^{q} V_k \to K$ are p and q-forms, then the map

$$\prod_{k=1}^{p} U_k \times \prod_{k=1}^{q} V_k \to K,\ (x_1, \ldots, x_p, y_1, \ldots, y_q) \mapsto \omega(x_1, \ldots, x_p) \cdot \eta(y_1, \ldots, y_q)$$

is a $(p+q)$-form. $\diamond$

We should note that m-linear maps don't eat collections, but rather *ordered* collections of vectors. Geometrically, this means that m-linear maps are sensitive to orientation. To study this further, we need to understand permutations.

DEFINITION 3.2. A *permutation* of a set X is a bijection of X onto itself. The set of all permutations of X, denoted S_X, is the *symmetric group* of X. When $X = \{1, \ldots, n\}$, we write $S_X = S_n$.

Given any two permutation $\pi, \tau \in S_X$, their composition $\pi\tau = \pi \circ \tau \in S_X$. For any $\sigma \in S_X$, $\sigma 1_X = \sigma = 1_X \sigma$, where $1_X \in S_X$ is the identity map $x \mapsto x$; σ has an inverse function $\sigma^{-1} \in S_X$ satisfying $\sigma\sigma^{-1} = 1_X = \sigma^{-1}\sigma$.

For any $a, b \in X$, $a \neq b$, the *transposition* of a and b, denoted $(a\ b)$, is the permutation that sends a to b, b to a, and any other c to itself. In other words, $(a\ b)$ swaps a and b and does nothing else.

The following results illuminate the role of transpositions in the structure of S_n.

PROPOSITION 3.3. Every permutation in S_n is the product of transpositions.

Proof. The proof will go by induction on n. For $n = 1$, there is only one permutation, the identity, which is the empty product of transpositions. Suppose that the claim holds for some $n - 1 \geqslant 1$. If $\sigma \in S_n$ fixes n, then it admits a restriction $\sigma|_{n-1} \in S_{n-1}$, which is the product of transpositions, hence σ is the product of the same transpositions in S_n. Alternatively, if $\sigma(n) \neq n$, then $(\sigma(n)\ n)\sigma(n) = n$, and the same argument shows $(\sigma(n)\ n)\sigma$ is the product of transpositions; composing on the left by the involution $(\sigma(n)\ n)$ yields an expression of σ as a product of transpositions. $\square$

PROPOSITION 3.4. If $\pi_1, \ldots, \pi_p$ and $\tau_1, \ldots, \tau_q$ are transpositions in S_n with

$$\prod_{k=1}^{p} \pi_k = \prod_{k=1}^{q} \tau_k,$$

then $p \equiv q \pmod 2$.

Proof. Consider the polynomial

$$f(t_1, \ldots, t_n) = \prod_{1 \leqslant i < j \leqslant n} (t_i - t_j) \in \mathbb{C}[t_1, \ldots, t_n].$$

For any $\sigma \in S_n$, define $\sigma f(t_1, \ldots, t_n) = f(t_{\sigma(1)}, \ldots, t_{\sigma(n)})$. For each factor $(t_i - t_j)$ of f, $(t_{\sigma(i)} - t_{\sigma(j)})$ is a scalar multiple of another factor. Therefore, $\sigma f = \alpha f$ for some $\alpha \in \mathbb{C}$. For any transposition $(i\ j)$, we have $(i\ j)f = -f$, so

$$\left(\prod_{k=1}^{p} \pi_k \right) f = (-1)^p f = (-1)^q f = \left(\prod_{k=1}^{q} \tau_k \right) f.$$

The result follows from the equality $(-1)^p = (-1)^q$. □

DEFINITION 3.5. The *sign* of a permutation $\sigma \in S_n$ is the unique scalar $\operatorname{sgn}(\sigma)$ satisfying

$$\sigma \prod_{i<j} (t_i - t_j) = \operatorname{sgn}(\sigma) \prod_{i<j} (t_i - t_j).$$

The previous two results show that the $\operatorname{sgn}(\sigma) = \pm 1$. The sign of a permutation is positive if and only if it can be written as the product of an even number of transpositions.

The symmetric group S_m acts on multilinear maps in a natural way: if $\omega : V^m \to W$ is m-linear, define $\sigma\omega : V^m \to W$ by

$$\sigma\omega(x_1, \ldots, x_m) = \omega(x_{\sigma(1)}, \ldots, x_{\sigma(m)}).$$

Brief reflection shows $\sigma\omega$ is m-linear.

An m-linear map $\omega : V^m \to W$ is *symmetric* if $\sigma\omega = \omega$ for all $\sigma \in S_m$. These are the m-linear maps that ignore orientation. For a more interesting example, ω is *skew-symmetric* if $\sigma\omega = \operatorname{sgn}(\sigma)\omega$ for all $\sigma \in S_m$ or, equivalently, $(i\ j)\omega = -\omega$ for all transpositions $(i\ j)$. Skew-symmetric maps treat orientation in a geometrically-meaningful way: reversing the orientation of their arguments reverses their output.

§3.2. TENSOR PRODUCTS

Multilinear maps are, in general, difficult to work with. Linear transformations, on the other hand, are rather easy. This suggests a strategy for studying multilinear maps: turn them into linear ones. This can be done by factoring through a vector space, just as we factored non-invertible transformations through quotients to obtain invertible ones. We start in the bilinear case.

Let U and V be vector spaces over K. We want a vector space $T(U, V)$, together with a bilinear map $t : U \times V \to T(U, V)$ such that, for any bilinear map $\omega : U \times V \to W$, there is a unique linear map $\widetilde{\omega} : T(U, V) \to W$ that makes the following diagram commute:

$$\begin{tikzcd} T(U,V) \arrow[dr, "\widetilde{\omega}"] & \\ U \times V \arrow[u, "t"] \arrow[r, "\omega"'] & W \end{tikzcd} .$$

This is called the *universal property of tensor products* and, like other universal properties, it characterizes $T(U,V)$ up to isomorphism. While we provide a construction below, the reader is encouraged to think of tensor products primarily in terms of their universal property.

Consider a vector space $F = F(U \times V)$ that has $U \times V$ as a basis. This is the *free vector space* over $U \times V$; the construction is uninspiring, hence omitted. Let M be the subspace spanned by all vectors in F of the form

$$(\alpha u_1 + \beta u_2, v_1) - \alpha(u_1, v_1) - \beta(u_2, v_1),$$
$$(u_1, \alpha v_1 + \beta v_2) - \alpha(u_1, v_1) - \beta(u_1, v_2).$$

DEFINITION 3.6. The *tensor product* of U and V is the vector space $U \otimes V = F/M$. The *canonical map* is the restriction $\otimes = \pi_M|_{U\times V} : U \times V \to U \otimes V$ of the canonical projection.

Brief reflection shows that $\otimes$ is bilinear. We write $\otimes(x,y) = x \otimes y$, and refer to the elements of $\mathrm{im}(\otimes)$ as *elementary tensors*. Since F is spanned by $U \times V$ and π_M is surjective, $U \otimes V$ is spanned by elementary tensors.

THEOREM 3.7 (Universal Property of Tensor Products)**.** If $\omega : U \times V \to W$ is bilinear, there is a unique linear transformation $\widetilde{\omega} : U \otimes V \to W$ such that $\omega = \widetilde{\omega} \circ \otimes$.

Proof. Since $U \times V$ is a basis in F, ω extends uniquely to a linear transformation $T_\omega : F \to W$. Then

$$(\alpha u_1 + \beta u_2, v_1) - \alpha(u_1, v_1) - \beta(u_2, v_1) \xmapsto{T_\omega} \omega(\alpha u_1 + \beta u_2, v_1) - \alpha\omega(u_1, v_1) - \beta\omega(u_2, v_1) = 0,$$

and similarly

$$T_\omega((u_1, \alpha v_1 + \beta v_2) - \alpha(u_1, v_1) - \beta(u_1, v_2)) = 0.$$

Consequently, $M \subseteq \ker(T_\omega)$. Let $\widetilde{\omega} = T_\omega/M : U \otimes V \to W$. Then

$$\widetilde{\omega} \circ \otimes = (\widetilde{\omega} \circ \pi_M)|_{U\times V} = T_\omega|_{U\times V} = \omega.$$

As for uniqueness, if we have a linear transformation $A : U \otimes V \to W$ such that $A \circ \otimes = \omega$, then A and $\widetilde{\omega}$ agree on all elementary tensors $x \otimes y$. Since these span $U \otimes V$, it follows that $A = \widetilde{\omega}$. □

To generalize to the m-linear case, and to show the tensor product in action, we prove the following:

PROPOSITION 3.8. For any vector spaces X, Y, and Z, $(X \otimes Y) \otimes Z \cong X \otimes (Y \otimes Z)$.

Proof. For each $x \in X$, let $\omega_x : Y \times Z \to (X \otimes Y) \otimes Z, (y,z) \mapsto (x \otimes y) \otimes z$. Then ω_x is bilinear, so it defines a unique linear transformation $\widetilde{\omega}_x : Y \otimes Z \to (X \otimes Y) \otimes Z$, which satisfies $\widetilde{\omega}_x(y \otimes z) = (x \otimes y) \otimes z$. Let $\Omega : X \times (Y \otimes Z) \to (X \otimes Y) \otimes Z$ be defined by

$$\Omega\left(x, \sum y_k \otimes z_k\right) = \widetilde{\omega}_x\left(\sum y_k \otimes z_k\right) = \sum (x \otimes y_k) \otimes z_k.$$

Then Ω is bilinear, so we have $\widetilde{\Omega} : X \otimes (Y \otimes Z) \to (X \otimes Y) \otimes Z$ sending $x \otimes (y \otimes z)$ to $(x \otimes y) \otimes z$.

Performing a symmetric construction, we have a linear transformation $\widetilde{\Gamma} : (X \otimes Y) \otimes Z \to X \otimes (Y \otimes Z)$ such that $\widetilde{\Gamma}((x \otimes y) \otimes z) = x \otimes (y \otimes z)$. Note that, since each element of $Y \otimes Z$ can be written as a linear combination of elementary tensors $y_k \otimes z_k$, each element of $X \otimes (Y \otimes Z)$ is a linear combination of vectors of the form

$$x_i \otimes \left(\sum y_k \otimes z_k\right) = \sum x_i \otimes (y_k \otimes z_k).$$

Thus $X \otimes (Y \otimes Z)$ is spanned by vectors of the form $x \otimes (y \otimes z)$. Similarly, $(X \otimes Y) \otimes Z$ is spanned by those of the form $(x \otimes y) \otimes z$. By their definitions,

$$(x \otimes y) \otimes z \xmapsto{\widetilde{\Gamma}} x \otimes (y \otimes z) \xmapsto{\widetilde{\Omega}} (x \otimes y) \otimes z$$

and

$$x \otimes (y \otimes z) \xmapsto{\widetilde{\Omega}} (x \otimes y) \otimes z \xmapsto{\widetilde{\Gamma}} x \otimes (y \otimes z).$$

Therefore $\widetilde{\Omega} = \widetilde{\Gamma}^{-1}$, and we are done. $\square$

Identifying $(X \otimes Y) \otimes Z$ with $X \otimes (Y \otimes Z)$, we may unambiguously refer to the tensor product $\bigotimes_{k=1}^m V_k$ of vector spaces $V_1, \ldots, V_m$. We have the m-linear canonical map $\otimes : \prod_{k=1}^m V_k \to \bigotimes_{k=1}^m V_k$, for which we write $\otimes(x_1, \ldots, x_m) = x_1 \otimes \cdots \otimes x_m$. As before, the elements of $\operatorname{im}(\otimes)$ are called *elementary tensors*; the proof above shows that they span.

The universal property generalizes as well: for each m-linear map $\omega : \prod_{k=1}^m V_k \to W$, there is a unique linear transformation $\widetilde{\omega} : \bigotimes_{k=1}^m V_k \to W$ such that $\omega = \widetilde{\omega} \circ \otimes$.

We conclude this section with an expression for the dimension of tensor products.

PROPOSITION 3.9. If $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ are bases for U and V respectively, then $\{x_i \otimes y_j : i \in [n], j \in [m]\}$ is a basis of $U \otimes V$. Consequently, $\dim(U \otimes V) = \dim(U) \cdot \dim(V)$.

Proof. Let W be an nm-dimensional vector space with basis $\{w_{ij} : i \in [n], j \in [m]\}$. Define $T : U \times V \to W$ by

$$T\left(\sum_{i=1}^n \alpha_i x_i, \sum_{j=1}^m \beta_j y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j w_{ij}$$

so, in particular, $T(x_i, y_j) = w_{ij}$ for all i, j. Brief reflection shows T is bilinear, so it lifts uniquely to a linear transformation $\widetilde{T} : U \otimes V \to W$ that sends $x_i \otimes y_j$ to w_{ij}. Since $\{w_{ij}\}$ is linearly independent, so is $\{x_i \otimes y_j\}$.

For any elementary tensor $x \otimes y \in U \otimes V$, we can expand x and y as linear combinations of $\{x_i\}$ and $\{y_j\}$, yielding

$$x \otimes y = \sum_{i=1}^{n} \alpha_i (x_i \otimes y) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j (x_i \otimes y_j).$$

Thus, $V = \text{span}(\text{im}(\otimes)) \subseteq \text{span}(\{x_i \otimes y_j\}) \subseteq V$, so $\{x_i \otimes y_j\}$ spans V. $\square$

§3.3. THE TRACE

Let $A : V \to V$ be a rank 1 operator. Choosing any non-zero vector $y = Ax \in \text{im}(A)$, we have $\text{im}(A) = \text{span}(y)$. After extending $\{x\}$ to a basis $\{x, x_2, \dots, x_n\}$ of V, let $\varphi_A \in V'$ be the linear functional defined by $\varphi_A(x) = 1$ and $\varphi_A(x_k) = 0$ for all x_k. Then, for all $\alpha x + \sum_{k=2}^{n} \alpha_k x_k \in V$,

$$A\left(\alpha x + \sum_{k=2}^{n} \alpha_k x_k\right) = \alpha A x = \varphi_A\left(\alpha x + \sum_{k=2}^{n} \alpha_k x_k\right) y,$$

so $A = \varphi_A(\cdot)y = (x \mapsto \varphi_A(x)y)$. We can feed A to itself, so to speak, by evaluating $\varphi_A(y)$. As we will show in this section, the scalar $\varphi_A(y)$ is independent of our choice of y, *i.e.* $A = \psi(\cdot)z$ implies $\psi(z) = \varphi_A(y)$. Furthermore, we can extend the map $A \mapsto \varphi_A(y)$ to all linear operators, and this extension is similarity-invariant.

To begin, we need to better understand the decomposition $A = \varphi_A(\cdot)y$. The key observation is that the map $(\varphi, x) \mapsto \varphi(\cdot)x$ is bilinear. From here, the essential result follows readily.

THEOREM 3.10. The linear transformation $T : V' \otimes W \to \mathscr{L}(V, W)$ defined by $T(\varphi \otimes w) = \varphi(\cdot)w$ is an isomorphism.

Proof. To show that T is well-defined, note that the map $(\varphi, x) \mapsto \varphi(\cdot)x$ is bilinear; factoring it through $V' \otimes W$ yields T. If $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$ are bases of V and W, and if $\{x^1, \dots, x^n\}$ is the dual basis in V', then $\{x^i \otimes y_j : i \in [n], j \in [m]\}$ is a basis of $V' \otimes W$. For all i, j and each x_k,

$$T(x^i \otimes y_j)(x_k) = x^i(x_k) y_j = \delta_{ik} y_j.$$

That is $T(x^i \otimes y_j)$ is just the transformation B_{ij} from the proof of **Corollary 2.3.**. In that proof, we showed that $\{B_{ij}\} = T(\{x^i \otimes y_j\})$ is a basis of $\mathscr{L}(V, W)$. It follows that T is an isomorphism. $\square$

The discussion at the beginning of this section shows that elementary tensors in $V' \otimes W$ correspond precisely to rank 1 transformations $V \to W$.

In the case $V = W$, we have the bilinear evaluation map $V' \times V \to K, (\varphi, x) \mapsto \varphi(x)$. Lifting this to $V' \otimes V$ yields a unique linear transformation $\varepsilon : V' \otimes V \to K$ satisfying $\varepsilon(\varphi \otimes x) = \varphi(x)$. If $\varphi(\cdot)y = A = \psi(\cdot)z$, then $\varphi \otimes y = T^{-1}(A) = \psi \otimes z$, and we have

$$\varphi(y) = \varepsilon(\varphi \otimes y) = \varepsilon(\psi \otimes z) = \psi(z).$$

This completes the proof of our first claim.

For the second, note that ε is defined on all of $V' \otimes V \cong \mathscr{L}(V)$. For any linear operator $A \in \mathscr{L}(V)$, the *trace* of A is the scalar $\operatorname{tr}(A) = (\varepsilon \circ T^{-1})(A)$. In general, the trace is a linear functional $\operatorname{tr} : \mathscr{L}(V) \to K$. Since the trace is the canonical extension of the rank 1 "self-evaluation" map, we can think of $\operatorname{tr}(A)$ as the evaluation of A at itself.

We conclude by showing that the trace is a similarity invariant.

PROPOSITION 3.11. If $A : V \to W$ and $B : W \to V$ are linear transformations, then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof. If $T^{-1}(A) = \sum \alpha_k(\varphi_k \otimes x_k)$, then $A = \sum \alpha_k \varphi_k(\cdot)x_k$. Composing on either side by B, we obtain the identities

$$AB = \sum \alpha_k(\varphi_k B)(\cdot)x_k,$$
$$BA = \sum \alpha_k \varphi_k(\cdot)Bx_k.$$

Then

$$AB \xmapsto{T^{-1}} \sum \alpha_k(\varphi_k B \otimes x_k) \xmapsto{\varepsilon} \sum \alpha_k \varphi_k(Bx_k) = \operatorname{tr}(BA)$$

□

THEOREM 3.12. If $A \in \mathscr{L}(V)$, and if $Q : V \to W$ is an isomorphism, then $\operatorname{tr}(QAQ^{-1}) = \operatorname{tr}(A)$.

Proof. By the associativity of composition,

$$\operatorname{tr}(QAQ^{-1}) = \operatorname{tr}((QA)Q^{-1}) = \operatorname{tr}(Q^{-1}(QA)) = \operatorname{tr}(A).$$

□

§3.4. EXTERIOR POWERS

Recall that a linear operator is invertible if and only if it preserves bases, if and only if it preserves linear independence. Consideration of multilinear maps and, in particular, bilinear maps led us to the trace. We now consider multilinear maps that depend on the linear independence of their arguments, in hope of finding an invariant that encodes the invertibility of an operator.

DEFINITION 3.13. An m-linear map $\omega : V^m \to W$ is *alternating* if $\omega(x_1, \ldots, x_m) = 0$ whenever $x_i = x_j$ for some $i \neq j$.

That is, ω is alternating if it vanishes whenever its arguments are obviously linearly dependent. As it turns out, obvious is sufficient:

PROPOSITION 3.14. Let $\omega : V^m \to W$ be an alternating m-linear map. If we have linearly dependent vectors $x_1, \ldots, x_m \in V$, then $\omega(x_1, \ldots, x_m) = 0$.

Proof. There is some $k \geqslant 2$ such that $x_k = \sum_{i=1}^{k-1} \alpha_i x_i$. Then

$$\omega(x_1, \dots, x_k, \dots, x_m) = \sum_{i=1}^{k-1} \alpha_i \omega(x_1, \dots, x_i, \dots, x_i, \dots, x_m) = 0.$$

□

Alternating maps also encode information about orientation:

PROPOSITION 3.15. If $\omega : V^m \to W$ is alternating, then it is skew-symmetric.

Proof. It is sufficient to show that $\tau\omega = -\omega$ for all transpositions τ. Let $(i\ j) \in S_m$ be arbitrary. For any $x_1, \dots, x_m \in V$,

$$\begin{aligned} 0 &= \omega(x_1, \dots, x_{i-1}, x_i + x_j, x_{i+1}, \dots, x_{j-1}, x_i + x_j, x_{j+1}, \dots, x_m) \\ &= \omega(x_1, \dots, x_i, \dots, x_j, \dots, x_m) + \omega(x_1, \dots, x_j, \dots, x_i, \dots, x_m) \\ &= \omega(x_1, \dots, x_m) + (i\ j)\omega(x_1, \dots, x_m), \end{aligned}$$

hence $(i\ j)\omega(x_1, \dots, x_m) = -\omega(x_1, \dots, x_m)$. □

At this point, it's natural to pass to the tensor product. Denote the m-fold tensor product of V with itself with the symbol $V^{\otimes m} = \bigotimes_{k=1}^m V$. For each linear transformation $\widetilde{\omega} : V^{\otimes m} \to W$ corresponding to an alternating m-linear map $\omega : V^m \to W$, we have

$$\widetilde{\omega}(x_1 \otimes \cdots \otimes x_m) = \omega(x_1, \dots, x_m) = 0$$

whenever $x_i = x_j$ for some $i \neq j$. Therefore, $E \subseteq \ker(\widetilde{\omega})$, where E is the subspace generated by all elementary tensors $x_1 \otimes \cdots \otimes x_m$ with $x_i = x_j$ for some $i \neq j$. In fact, ω is alternating if and only if the kernel of $\widetilde{\omega}$ contains E. To bring alternating maps into focus, we quotient by E.

DEFINITION 3.16. The *mth exterior power* of V is the vector space $\bigwedge^m V = V^{\otimes m}/E$. The *canonical map* is the composition $\wedge = \pi_E \circ \otimes : V^m \to \bigwedge^m V$. By convention, we write $\bigwedge^1 V = V$.

The reader will verify that $\wedge$ is alternating. As with the tensor product, we write $\wedge(x_1, \dots, x_m) = x_1 \wedge \cdots \wedge x_m$. The elements of $\mathrm{im}(\wedge)$ are called *m-blades*. Since $\mathrm{im}(\otimes)$ spans $V^{\otimes m}$, and since π_E is surjective, $\bigwedge^m V$ is spanned by m-blades.

Exterior powers specialize the universal property of tensor products.

THEOREM 3.17 (Universal Property of Exterior Products)**.** If $\omega : V^m \to W$ is an alternating m-linear map, then there is a unique linear transformation $\widehat{\omega} : \bigwedge^m V \to W$ such that $\omega = \widehat{\omega} \circ \wedge$.

Proof. We know that ω defines a linear transformation $\widetilde{\omega} : V^{\otimes m} \to W$ with kernel containing E. Let $\widehat{\omega} = \widetilde{\omega}/E : \bigwedge^m V \to W$. Then $\widehat{\omega} \circ \wedge = (\widehat{\omega} \circ \pi_E) \circ \otimes = \widetilde{\omega} \circ \otimes = \omega$. If we have another linear transformation $A : \bigwedge^m V \to W$ such that $A \circ \wedge = \omega = \widehat{\omega} \circ \wedge$, then A and $\widehat{\omega}$ agree on all m-blades, which span $\bigwedge^m V$, hence $A = \widehat{\omega}$. □

PROPOSITION 3.18. For $x_1, \dots, x_n \in V$, $x_1 \wedge \cdots \wedge x_n = 0$ if and only if the x_i's are linearly dependent.

Proof. The backwards implication follows immediately from **Proposition 3.14.**. Suppose $x_1 \wedge \cdots \wedge x_n \neq 0$ and $\sum_{k=1}^{n} \alpha_k x_k = 0$. For each $1 \leqslant i \leqslant n$, let

$$\widehat{x_i} = x_1 \wedge \cdots \wedge x_{i-1} \wedge x_{i+1} \wedge \cdots \wedge x_n.$$

Then

$$0 = \widehat{x_i} \wedge \left(\sum_{k=1}^{n} \alpha_k x_k \right) = \sum_{k=1}^{n} \alpha_k \widehat{x_i} \wedge x_k = \sum_{k=1}^{n} \alpha_k \varepsilon_k \delta_{ki} (x_1 \wedge \cdots \wedge x_n) = \alpha_i \varepsilon_i (x_1 \wedge \cdots \wedge x_n)$$

where ε_k is the sign of the permutation rearranging $\widehat{x_i} \wedge x_k$ to be in the proper order. Then the penultimate equality follows from skew-symmetry. Since $\varepsilon_i = \pm 1$ and given n-blade is non-zero, this yields $\alpha_i = 0$, so the x_i's are linearly independent. $\square$

It's reasonable to ask what $\bigwedge^m V$ looks like. Let V be a 3-dimensional real vector space, and let $m = 2$. We know $x \wedge y = 0$ if and only if x and y are linearly dependent. Therefore, the non-zero 2-blades roughly correspond to 2-dimensional subspaces of V. More precisely, note that

$$(x + \alpha y) \wedge y = x \wedge y + \alpha (y \wedge y) = x \wedge y,$$

so shearing x along y leaves $x \wedge y$ invariant. If we think of the parallelogram with adjacent sides x and y, we see that $x \wedge y = z \wedge w$ whenever the parallelogram with sides z and w is coplanar with that of x and y and has the same (signed) area. With this in mind, we may interpret $x \wedge y$ as the plane spanned by x and y, tagged by the signed area of their corresponding parallelogram. Under this interpretation, the 2-blades in $\bigwedge^2 V$ are the weighted 2-dimensional subspaces of V, together with a single point corresponding to 0. Morally speaking, $\bigwedge^m V$ is the space of linear combinations of weighted m-dimensional subspaces of V.

We conclude this section with the dimension of $\bigwedge^m V$. Being a quotient of a finite-dimensional vector space, it must be finite-dimensional itself. Ideally, we would be able to construct a basis of $\bigwedge^m V$ from a basis of V. What would this basis look like? Well, it should look like all of the (suitably weighted) m-dimensional subspaces that can be built from that basis, which correspond to all of the subsets of size m. Therefore, if V is n-dimensional, then $\bigwedge^m V$ should be $\binom{n}{m}$-dimensional. This turns out to be true, though the proof is somewhat more technical.

THEOREM 3.19. If $\{x_1, \ldots, x_n\}$ is a basis of V, then the set of all $x_{i_1} \wedge \cdots x_{i_m}$, where $i_1 < \cdots < i_m$, is a basis of $\bigwedge^m V$. Consequently, $\dim \bigwedge^m V = \binom{n}{m}$.

Before presenting the proof, we introduce some handy notation: given a tuple $I = (i_1, \ldots, i_m) \in [n]^m$, let

$$\begin{aligned} x_I &= (x_{i_1}, \ldots, x_{i_m}), \\ x_{\otimes I} &= x_{i_1} \otimes \cdots \otimes x_{i_m}, \\ x_{\wedge I} &= x_{i_1} \wedge \cdots \wedge x_{i_m}. \end{aligned}$$

We say that I is *increasing* if $i_1 < \cdots < i_m$.

Proof. We want to show that $\{x_{\wedge I} : I \in [n]^m \text{ is increasing}\}$ is a basis of $\bigwedge^m V$. Proving that it spans V is straightfoward: it is sufficient to show that it spans all m-blades. If $y_1 \wedge \cdots \wedge y_m$ is any m-blade, then we can write each y_k as a linear combination of $\{x_1, \ldots, x_m\}$. Expanding by multilinearity, we get a sum of the form

$$y_1 \wedge \cdots \wedge y_m = \sum \alpha_J x_{\wedge J},$$

where J ranges over $[n]^m$. All summands corresponding to tuples J for which $j_p = j_q$ for some $p \neq q$ vanish. For each of the remaining terms $\alpha_J x_{\wedge J}$, there is a unique permutation $\sigma_J \in S_m$ such that $\sigma_J J = (j_{\sigma_J(1)}, \ldots, j_{\sigma_J(m)})$ is increasing. By skew-symmetry,

$$x_{\wedge \sigma_J J} = \sigma_J x_{\wedge J} = \operatorname{sgn}(\sigma_J) x_{\wedge J},$$

so $x_{\wedge J} = \operatorname{sgn}(\sigma_J) x_{\wedge \sigma_J J}$. Then

$$y_1 \wedge \cdots \wedge y_m = \sum \alpha_J \operatorname{sgn}(\sigma_J) x_{\wedge \sigma_J J}$$

is a linear combination of $\{x_{\wedge I}\}$, hence $\{x_{\wedge I}\}$ spans $\bigwedge^m V$.

Let W be a vector space containing $\{I : I \in [n]^m \text{ is increasing}\}$ as a basis (*e.g.*, the space of linear combinations of this set). Let $\kappa : [n]^m \to \{0, 1\}$ be defined by $\kappa(j_1, \ldots, j_m) = 0$ if and only if $j_p = j_q$ for some $p \neq q$. By **Proposition 3.9.**, the set $\{x_{\otimes J} : J \in [n]^m\}$ is a basis of $V^{\otimes m}$, so we may define a linear transformation $\widetilde{\omega} : V^{\otimes m} \to W$ by

$$\widetilde{\omega}(x_{\otimes J}) = \kappa(J) \operatorname{sgn}(\sigma) \sigma J,$$

where $\sigma \in S_m$ is the unique permutation such that σJ is increasing. Composing on the right by $\otimes$ yields an m-linear map $\omega = \widetilde{\omega} \circ \otimes : V^m \to W$. Brief reflection shows ω is alternating, so there is a linear transformation $\widehat{\omega} : \bigwedge^m V \to W$ with $\widehat{\omega} \circ \wedge = \omega$. For each $x_{\wedge I}$,

$$\widehat{\omega}(x_{\wedge I}) = \widetilde{\omega}(x_{\otimes I}) = I.$$

Therefore, if $\sum \alpha_I x_{\wedge I} = 0$, then

$$0 = \widetilde{\omega} \sum \alpha_I x_{\wedge I} = \sum \alpha_I I,$$

so each $\alpha_I = 0$ by the linear independence of $\{I\}$ in W. It follows that $\{x_{\wedge I}\}$ is linearly independent, hence a basis.

Each increasing tuple $I \in [n]^m$ corresponds to a subset of $[n]$ of size m, and vice versa. It follows that $\dim \bigwedge^m V = |\{x_{\wedge I}\}| = \binom{n}{m}$. $\square$

§3.5. THE DETERMINANT

Let $A : V \to W$ be a linear transformation. The map

$$V^m \to \bigwedge^m W, \ (x_1, \ldots, x_m) \mapsto Ax_1 \wedge \cdots \wedge Ax_m$$

is m-linear and alternating. Applying the universal property of exterior powers, we obtain a linear transformation $A^{\wedge m} : \bigwedge^m V \to \bigwedge^m W$, the *mth exterior power* of A, which uniquely satisfies

$$A^{\wedge m}(x_1 \wedge \cdots \wedge x_m) = Ax_1 \wedge \cdots \wedge Ax_m.$$

Clearly, $\mathbb{1}_V^{\wedge m} = \mathbb{1}_{\bigwedge^m V}$. For any transformations $A : V \to W$ and $B : U \to V$,

$$(AB)^{\wedge m}(x_1 \wedge \cdots \wedge x_m) = ABx_1 \wedge \cdots \wedge ABx_m = A^{\wedge m}B^{\wedge m}(x_1 \wedge \cdots \wedge x_m),$$

so $(AB)^{\wedge m} = A^{\wedge m}B^{\wedge m}$. By the previous two lines, if A is invertible, then so is $A^{\wedge m}$, and $(A^{\wedge m})^{-1} = (A^{-1})^{\wedge m}$.

By **Theorem 3.17.**, if V is n-dimensional, then $\bigwedge^n V$ is 1-dimensional. This is geometrically obvious: since there is only one n-dimensional subspace of V, any two n-blades differ only in weight, *i.e.* by a scalar. As such, each operator $B \in \mathscr{L}(\bigwedge^n V)$ acts by scaling. In particular, for each $A \in \mathscr{L}(V)$, there is a unique scalar δ such that

$$A^{\wedge m}(x_1 \wedge \cdots \wedge x_n) = Ax_1 \wedge \cdots \wedge Ax_n = \delta(x_1 \wedge \cdots \wedge x_n)$$

for all $x_1 \wedge \cdots \wedge x_n \in \bigwedge^n V$. This scalar is the *determinant* of A, typically denoted by $\det(A) = \delta$.

Each $x_1 \wedge \cdots \wedge x_n$ corresponds to the volume of the parallelotope with adjacent sides $x_1, \ldots, x_n$. The action of A on V sends this parallelotope to another with sides $Ax_1, \ldots, Ax_n$ and volume corresponding to

$$Ax_1 \wedge \cdots \wedge Ax_n = \det(A)(x_1 \wedge \cdots \wedge x_n).$$

Therefore, the determinant of A is the factor by which A scales signed volume in V. For example, operators with determinant 1 preserve volume and orientation, whereas those with determinant -1 preserve volume and reverse orientation.

Consider a linear operator $A \in \mathscr{L}(V)$ with $\det(A) = 0$. We know that $y_1 \wedge \cdots \wedge y_n = 0$ if and only if the vectors $y_1, \ldots, y_n$ are linearly dependent, if and only if they do not form a basis of V. Thus, each non-zero $x_1 \wedge \cdots \wedge x_n$ corresponds to a basis $\{x_1, \ldots, x_n\}$ of V and vice versa, so

$$Ax_1 \wedge \cdots \wedge Ax_n = 0(x_1 \wedge \cdots \wedge x_n) = 0$$

implies $\{Ax_1, \ldots, Ax_n\}$ is not a basis of V, and A does not send bases to bases. Therefore, A is not invertible. Conversely, if A is not invertible, then

$$Ax_1 \wedge \cdots \wedge Ax_n = 0 = 0(x_1 \wedge \cdots \wedge x_n)$$

for some non-zero n-blade $x_1 \wedge \cdots \wedge x_n$, hence $\det(A) = 0$. We have shown

THEOREM 3.20. A linear operator $A \in \mathscr{L}(V)$ is invertible if and only if $\det(A) \neq 0$.

Unlike the trace, the determinant is not linear: $\det(A + B) \neq \det(A) + \det(B)$. However, it is multiplicative:

PROPOSITION 3.21. If $A, B \in \mathscr{L}(V)$, then $\det(AB) = \det(A) \cdot \det(B)$.

Proof. The proof is a straightfoward computation:

$$ABx_1 \wedge \cdots \wedge ABx_n = \det(A)(Bx_1 \wedge \cdots \wedge Bx_n) = \det(A)\det(B)(x_1 \wedge \cdots \wedge x_n)$$

□

COROLLARY 3.22. If $A \in \mathscr{L}(V)$ is invertible, then $\det(A^{-1}) = \det(A)^{-1}$.

Proof. This follows immediately from $\det(AA^{-1}) = \det(\mathbb{1}) = 1$. □

Recall that for linear operators $A : X \to X$ and $B : Y \to Y$, we write $A \oplus B$ for the linear operator $(\iota_X \circ A) \oplus (\iota_Y \circ B) : X \oplus Y \to X \oplus Y$.

PROPOSITION 3.23. If $A \in \mathscr{L}(X)$ and $B \in \mathscr{L}(Y)$, then $\det(A \oplus B) = \det(A)\det(B)$.

Proof. Choose bases $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$ of X and Y respectively. Then the union $\{x_1, \ldots, x_m, y_1, \ldots, y_n\}$ is a basis of $X \oplus Y$, hence

$$x_1 \wedge \cdots \wedge x_m \wedge y_1 \wedge \cdots \wedge y_n \neq 0.$$

Then

$$\begin{aligned}\det(A \oplus B)(x_1 \wedge \cdots \wedge x_m \wedge y_1 \wedge \cdots \wedge y_n) &= Ax_1 \wedge \cdots \wedge Ax_m \wedge By_1 \wedge \cdots \wedge By_n \\ &= \det(A)(x_1 \wedge \cdots \wedge x_m) \wedge \det(B)(y_1 \wedge \cdots \wedge y_n) \\ &= \det(A)\det(B)(x_1 \wedge \cdots \wedge x_m \wedge y_1 \wedge \cdots \wedge y_n),\end{aligned}$$

and the result follows. □

The reader is doubtless unsurprised to learn that the determinant is a similarity invariant.

THEOREM 3.24. If $A \in \mathscr{L}(V)$ and $T : V \to W$ is an isomorphism, then $\det(TAT^{-1}) = \det(A)$.

Proof. For any n-blade $x_1 \wedge \cdots \wedge x_n \in \bigwedge^n V$,

$$(TAT^{-1})^{\wedge n}(x_1 \wedge \cdots \wedge x_n) = T^{\wedge n}A^{\wedge n}(T^{-1})^{\wedge n}(x_1 \wedge \cdots \wedge x_n) = \det(A)(x_1 \wedge \cdots \wedge x_n).$$

□

Given a scalar δ, the operator $\delta\mathbb{1}$ scales its underlying space by δ, and an operator A with $\det A = \delta$ scales volume by δ. The raises the question: is there a relationship between the transformations A and $\det(A)\mathbb{1}$? In particular, is there a way to decompose $\det(A)\mathbb{1}$ into a product of A with another operator?

For each $x \in V$, the map

$$V^{n-1} \to \bigwedge^n V, \ (x_2, \dots, x_n) \mapsto x \wedge x_2 \wedge \cdots \wedge x_n$$

is $(n-1)$-linear and alternating, hence defines a unique linear map $\varphi_x : \bigwedge^{n-1} V \to \bigwedge^n V$ by the universal property of exterior powers. A quick computation shows the assignment $\varphi : V \to \mathscr{L}(\bigwedge^{n-1} V, \bigwedge^n V), \ x \mapsto \varphi_x$ is linear. Suppose $\varphi_x = 0$ for some $x \neq 0$. Since x is non-zero, we can extend it to a basis $\{x, x_2, \dots, x_n\}$ of V. Then

$$\varphi_x(x_2 \wedge \cdots \wedge x_n) = x \wedge x_2 \wedge \cdots \wedge x_n \neq 0,$$

a contradiction. Thus the kernel of φ is trivial, and φ is injective. By rank-nullity, it is also surjective, hence an isomorphism.

Since $\bigwedge^n V$ is 1-dimensional, $\mathscr{L}(\bigwedge^{n-1} V, \bigwedge^n V)$ is isomorphic to $(\bigwedge^{n-1} V)'$, so we may view each φ_x as a linear functional. Continuing with this analogy, each operator $B \in \mathscr{L}(\bigwedge^{n-1} V)$ defines an adjoint operator B' on $\mathscr{L}(\bigwedge^{n-1} V, \bigwedge^n V)$ by $B'\varphi_x = \varphi_x B$.

The *adjugate* of $A \in \mathscr{L}(V)$ is the linear operator $\operatorname{adj}(A) = \varphi^{-1}(A^{\wedge n-1})'\varphi$. In other words, $\operatorname{adj}(A)$ is defined by

$$(\operatorname{adj}(A)x) \wedge x_2 \wedge \cdots \wedge x_n = x \wedge Ax_2 \wedge \cdots \wedge Ax_n.$$

Note that $\operatorname{adj}(A)$ is similar to $(A^{\wedge n-1})'$. Returning to the determinant, we have

$$(\operatorname{adj}(A)Ax_1) \wedge x_2 \wedge \cdots \wedge x_n = Ax_1 \wedge \cdots \wedge Ax_n = \det(A)x_1 \wedge \cdots \wedge x_n.$$

Then $\varphi_{\operatorname{adj}(A)Ax} = \varphi_{\det(A)x}$, so $\operatorname{adj}(A)Ax = \det(A)x$ for all $x \in V$. We have shown

THEOREM 3.25. For any linear operator A, $\operatorname{adj}(A)A = \det(A)\mathbb{1}$.

This yields the following identity for the inverse.

COROLLARY 3.26. If $A \in \mathscr{L}(V)$ is invertible, then $A^{-1} = \det(A)^{-1}\operatorname{adj}(A)$.

§3.6. EIGENVALUES

In this section, we study the behavior of an operator on those vectors that it treats especially simply. The reader may reasonably ask why we didn't introduce the resulting invariant earlier. While we very well could have, the theory is greatly enhanced by the use of the determinant.

DEFINITION 3.27. Let $A : V \to V$ be a linear operator. A scalar λ is a *eigenvalue* of A if there is a non-zero vector x, called a *λ-eigenvector*, such that $Ax = \lambda x$. The set of all eigenvalues of A is called the *spectrum* of A.

The requirement that eigenvectors be non-zero ensures the following equivalence: A is not invertible if and only if 0 is an eigenvalue of A.

Clearly, $x \neq 0$ is a λ-eigenvector of A if and only if $(\lambda\mathbb{1} - A)x = 0$. Consequently, the subspace $E_\lambda = \ker(\lambda\mathbb{1} - A)$, the *$\lambda$-eigenspace*, is the set of all λ-eigenvectors of A together with 0. The *geometric multiplicity* of λ, denoted g_λ, is the dimension of E_λ.

We know that λ is an eigenvalue of A if and only if $\lambda\mathbb{1} - A$ has non-trivial kernel, if and only if $\lambda\mathbb{1} - A$ is not invertible. Therefore, the eigenvalues of A are precisely the scalars t satisfying $\det(t\mathbb{1} - A) = 0$. Let $n = \dim V$, and let $x_1 \wedge \cdots \wedge x_n$ be any non-zero n-blade. By definition, we have

$$\det(t\mathbb{1} - A)(x_1 \wedge \cdots \wedge x_n) = (tx_1 - Ax_1) \wedge \cdots \wedge (tx_n - Ax_n).$$

Expanding the right-hand side, we obtain a sum of n-blades of the form

$$(-1)^{n-k} t^k (y_1 \wedge \cdots \wedge y_n),$$

where each $y_i = x_j$ or Ax_j for some j. Since $\bigwedge^n V$ is 1-dimensional, each $y_1 \wedge \cdots \wedge y_n = \alpha_k(x_1 \wedge \cdots \wedge x_n)$ for some α_k. Then

$$\det(t\mathbb{1} - A)(x_1 \wedge \cdots \wedge x_n) = \left(\sum_{k=0}^{n} (-1)^{n-k} \alpha_k t^k \right) (x_1 \wedge \cdots \wedge x_n),$$

so $\det(t\mathbb{1} - A)$ is a polynomial, called the *characteristic polynomial* of A and denoted $\chi_A(t)$. Brief reflection shows χ_A has leading term t^n, so it is degree n and monic, and has constant term $(-1)^n \det(A)$.

By definition, λ is an eigenvalue of A if and only if $\chi_A(\lambda) = 0$, in which case $(t - \lambda)$ divides χ_A. The *algebraic multiplicity* of λ, denoted m_λ, is the multiplicity of λ as a root of χ_A, *i.e.* the largest integer such that $(t - \lambda)^{m_\lambda}$ divides χ_A.

We are now ready to show that the spectral theory of an operator constitutes a similarity invariant.

THEOREM 3.28. If $A \in \mathscr{L}(V)$ and $T : V \to W$ is an isomorphism, then $\chi_{TAT^{-1}} = \chi_A$. Consequently, the eigenvalues of A agree with those of TAT^{-1}, as do the corresponding algebraic and geometric multiplicities.

Proof. For our first claim, we have

$$\chi_{TAT^{-1}}(t) = \det(t\mathbb{1} - TAT^{-1}) = \det(T(t\mathbb{1} - A)T^{-1}) = \det(t\mathbb{1} - A) = \chi_A(t).$$

It follows immediately that the eigenvalues and corresponding algebraic multiplicities of A and TAT^{-1} agree. As for the geometric multiplicities, for each eigenvalue λ of A,

$$\lambda\mathbb{1} - TAT^{-1} = T(\lambda\mathbb{1} - A)T^{-1},$$

so $\lambda\mathbb{1} - TAT^{-1}$ and $\lambda\mathbb{1} - A$ are similar; it follows that

$$\text{nullity}(\lambda\mathbb{1} - TAT^{-1}) = \text{nullity}(\lambda\mathbb{1} - A),$$

and we are done. □

The following gives us a geometric tool for working with the characteristic polynomial:

PROPOSITION 3.29. If $V = V_1 \oplus V_2$, then for any $A \in \mathscr{L}(V_1)$ and $B \in \mathscr{L}(V_2)$, $\chi_{A\oplus B} = \chi_A \cdot \chi_B$.

Proof. For any $v_1 + v_2 \in V$,

$$(t\mathbb{1}_V - A \oplus B)(v_1 + v_2) = (t\mathbb{1} - A)v_1 + (t\mathbb{1} - B)v_2 = ((t\mathbb{1}_{V_1} - A) \oplus (t\mathbb{1}_{V_2} - B))(v_1 + v_2),$$

so $t\mathbb{1}_V - A \oplus B = (t\mathbb{1}_{V_1} - A) \oplus (t\mathbb{1}_{V_2} - B)$. By **Proposition 3.22.**,

$$\chi_{A\oplus B}(t) = \det(t\mathbb{1}_V - A \oplus B) = \det(t\mathbb{1}_{V_1} - A)\det(t\mathbb{1}_{V_2} - B) = \chi_A(t)\chi_B(t).$$

□

In particularly nice cases, operators admit bases of eigenvectors. Such operators are called *diagonalizable*; the name comes from Chapter 5. Note that eigenspaces corresponding to distinct eigenvalues are disjoint. Indeed, if $\lambda_1 x = Ax = \lambda_2 x$, then $(\lambda_1 - \lambda_2)x = 0$, so either $\lambda_1 = \lambda_2$ or $x = 0$. Therefore, $A : V \to V$ is diagonalizable if and only if $V = \bigoplus_\lambda E_\lambda$, if and only if $\sum_\lambda g_\lambda = \dim V$.

Suppose $A \in \mathscr{L}(V)$ is diagonalizable. Clearly, $A(E_\lambda) \subseteq E_\lambda$, so A admits a restriction $A_\lambda = A|_{E_\lambda} \in \mathscr{L}(E_\lambda)$ for each eigenvalue λ. Since A is diagonalizable, $V = \bigoplus_\lambda E_\lambda$, so $A = \bigoplus_\lambda A_\lambda$. It follows that $\chi_A = \prod_\lambda \chi_{A_\lambda}$, where λ ranges over the distinct eigenvalues of A. Each A_λ has only one eigenvalue, λ, so $\chi_{A_\lambda}(t) = 0$ if and only if $t = \lambda$. Since the degree of χ_{A_λ} matches the dimension of the space A_λ acts on, $\chi_{A_\lambda}(t) = (t - \lambda)^{g_\lambda}$. It follows that

$$\chi_A(t) = \prod_\lambda (t - \lambda)^{g_\lambda}.$$

That is, when A is diagonalizable, χ_A factors completely and algebraic and geometric multiplicity agree. The proof of the converse is obvious.

§3.7. THE JORDAN DECOMPOSITION

Understanding the spectral theory of an operator amounts to understanding its characteristic polynomial. This is easiest to do when it factors into linear terms. A field K is *algebraically closed* if every non-constant polynomial $p \in K[t]$ is the product of linear terms. Equivalently, K is algebraically closed if every non-constant polynomial with coefficients in K has a root in K. For example, $\mathbb{C}$ is algebraically closed. Throughout this section, we assume our vector spaces are over algebraically closed fields.

Note: Under this assumption, every linear operator (on a space of dimension $\geqslant 1$) has an eigenvalue.

When $A : V \to V$ is diagonalizable, we have the decomposition $V = \bigoplus_\lambda E_\lambda$ and the corresponding identity $A = \bigoplus_\lambda A|_{E_\lambda}$. We seek to generalize this decomposition to all operators. The key insight is that $(\lambda\mathbb{1} - A|_{E_\lambda}) = 0$ for each eigenvalue λ, and that this property completely characterizes the decomposition. This suggests the following strategy: decompose A into a direct sum of operators A_λ such that each $(\lambda\mathbb{1} - A_\lambda)$ is almost zero, by which we mean:

DEFINITION 3.30. A linear operator $B : V \to V$ is *nilpotent* if $B^m = 0$ for some $m \geqslant 0$. The *degree* of B is the least integer q such that $B^q = 0$.

PROPOSITION 3.31. The degree of a nilpotent operator on a vector space V is at most $\dim V$.

Proof. Let $B \in \mathscr{L}(V)$ be nilpotent of degree q. For each $k \geqslant 0$, let $Z_k = \ker(B^k)$, and consider the chain

$$Z_0 \subseteq Z_1 \subseteq Z_2 \subseteq \cdots .$$

Clearly, $Z_q = V$. If $Z_k = Z_{k+1}$, then, for all $x \in Z_{k+2}$,

$$B^{k+2}x = B^{k+1}Bx = 0,$$

so $Bx \in Z_{k+1} = Z_k$, hence $B^kBx = B^{k+1}x = 0$, and $Z_{k+1} = Z_{k+2}$. Consequently, if $Z_k = Z_{k+1}$ for some $k \leqslant q$, then $Z_k = Z_{k+(q-k)} = Z_q = V$. It follows that $B^k = 0$, so, by the minimality of q, $k = q$ is minimal such that $Z_k = Z_{k+1}$. Thus, the inclusions $Z_{k-1} \subsetneq Z_k$ are proper for all $k \leqslant q$. For each $1 \leqslant k \leqslant q$, choose $x_k \in Z_k \setminus Z_{k-1}$. Then $\{x_1, \ldots, x_q\}$ is linearly independent, so $q \leqslant \dim V$. □

Let $A : V \to V$ be a linear operator. For each eigenvalue λ of A, the *generalized λ-eigenspace* is the subspace $G_\lambda = \ker(\lambda\mathbb{1} - A)^n$ of V, where $n = \dim V$. Clearly, $E_\lambda \subseteq G_\lambda$. Note that A commutes with $(\lambda\mathbb{1} - A)$ and, by extension, $(\lambda\mathbb{1} - A)^n$ as well. It follows that $A(G_\lambda) \subseteq G_\lambda$, so A admits a restriction $A|_{G_\lambda} \in \mathscr{L}(G_\lambda)$. By the definition of G_λ, $(\lambda\mathbb{1} - A|_{G_\lambda})^n = 0$, so $(\lambda\mathbb{1} - A|_{G_\lambda})$ is nilpotent. Therefore, we need only show $V = \bigoplus_\lambda G_\lambda$.

LEMMA 3.32. If λ is an eigenvalue of a linear operator $A : V \to V$, then

$$G_\lambda = \{x \in V : (\lambda\mathbb{1} - V)^k x = 0 \text{ for some } k \geqslant 0\}.$$

Proof. The inclusion $G_\lambda \subseteq \{x \in V : (\lambda\mathbb{1} - V)^k x = 0 \text{ for some } k \geqslant 0\}$ is trivial. Suppose $(\lambda\mathbb{1} - V)^k x = 0$ for some $k \geqslant 0$. Consider the subspace

$$S = \operatorname{span}(x, (\lambda\mathbb{1} - A)x, (\lambda\mathbb{1} - A)^2 x, \ldots).$$

Clearly $(\lambda\mathbb{1} - A)(S) \subseteq S$, so we have the restriction $B = (\lambda\mathbb{1} - A)|_S \in \mathscr{L}(S)$. Since $B^k x = (\lambda\mathbb{1} - A)^k x = 0$, B is nilpotent. By **Proposition 3.30.**, it has degree $q \leqslant n$, where $n = \dim V$. Therefore,

$$(\lambda\mathbb{1} - A)^n x = (\lambda\mathbb{1} - A)^{n-q}(\lambda\mathbb{1} - A)^q x = (\lambda\mathbb{1} - A)^{n-q}B^q x = 0,$$

so $x \in G_\lambda$. □

LEMMA 3.33. If λ_1 and λ_2 are distinct eigenvalues of a linear operator $A : V \to V$, then

1. G_{λ_1} and G_{λ_2} are disjoint,
2. $(\lambda_1\mathbb{1} - A)(G_{\lambda_2}) = G_{\lambda_2}$.

Proof. First of all, $(\lambda_1\mathbb{1} - A)(G_{\lambda_2}) \subseteq G_{\lambda_2}$. Indeed, for each $y \in G_{\lambda_2}$, we have $\lambda_1 y, Ay \in G_{\lambda_2}$, so $\lambda_1 y - Ay \in G_{\lambda_2}$.

Assume, for the sake of contradiction, that there is a non-zero vector $x \in G_{\lambda_1} \cap G_{\lambda_2}$. Let $p \geqslant 1$ be the least integer satisfying $(\lambda_1\mathbb{1} - A)^p x = 0$. Then $y = (\lambda_1 - A)^{p-1}x$ is a λ_1-eigenvector. Since $y \in G_{\lambda_2}$, there is a least integer $q \geqslant 1$ such that $(\lambda_2\mathbb{1} - A)^q y = 0$, so $(\lambda_2\mathbb{1} - A)^{q-1}y$ is a λ_2-eigenvector. However,

$$A(\lambda_2\mathbb{1} - A)^{q-1}y = (\lambda_2\mathbb{1} - A)^{q-1}Ay = \lambda_1(\lambda_2\mathbb{1} - A)^{q-1}y,$$

so it is also a λ_1-eigenvector, and we thereby obtain a contradiction. This completes the proof of **1.**.

As for **2.**, consider the restriction $C = (\lambda_1\mathbb{1} - A)|_{G_{\lambda_2}} \in \mathscr{L}(G_{\lambda_2})$. By **1.**, G_{λ_2} contains no λ_1-eigenvectors, so the kernel of C is trivial. Since C is a linear operator, it follows from rank-nullity that is is surjective, and we are done. $\square$

THEOREM 3.34. Let A be a linear operator on an n-dimensional vector space V. If $\lambda_1, \ldots, \lambda_d$ are the distinct eigenvalues of A, then $V = \bigoplus_{k=1}^{d} G_{\lambda_k}$.

Proof. The proof will go by induction on d. If $d = 0$, then A has no eigenvalues, so V is 0-dimensional, and we are done. Suppose the claim holds for some $d - 1 \geqslant 0$, and suppose $A \in \mathscr{L}(V)$ has d distinct eigenvalues $\lambda_1, \ldots, \lambda_d$. Let $I = \operatorname{im}(\lambda_d\mathbb{1} - A)^n$. We first show that $V = G_{\lambda_d} \oplus I$. Suppose $y \in G_{\lambda_d} \cap I$. Then $y = (\lambda_d\mathbb{1} - A)^n x$ for some $x \in V$, so

$$(\lambda_d\mathbb{1} - A)^{2n}x = (\lambda_d\mathbb{1} - A)^n y = 0.$$

By **Lemma 3.31.**, $x \in G_{\lambda_d}$, hence

$$y = (\lambda_d\mathbb{1} - A)^n x = 0,$$

and $G_{\lambda_d} \cap I = \{0\}$. By rank-nullity,

$$\dim(G_{\lambda_d} \oplus I) = \dim G_{\lambda_d} + \dim I = \operatorname{nullity}(\lambda_d\mathbb{1} - A)^n + \operatorname{rank}(\lambda_d\mathbb{1} - A)^n = n,$$

hence $G_{\lambda_d} \oplus I = V$.

For each $(\lambda_d\mathbb{1} - A)^n x \in I$,

$$A(\lambda_d\mathbb{1} - A)^n x = (\lambda_d\mathbb{1} - A)^n Ax \in I,$$

hence $A(I) \subseteq I$. Consider the restrictions $A|_{G_{\lambda_d}} \in \mathscr{L}(G_{\lambda_d})$ and $A|_I \in \mathscr{L}(I)$, which satisfy $A|_{G_{\lambda_d}} \oplus A|_I = A$. By **Proposition 3.28.**, $\chi_A(t) = \chi_G(t)\chi_I(t)$, where $\chi_G = \chi_{A|_{G_{\lambda_d}}}$ and $\chi_I = \chi_{A|_I}$. Since $E_{\lambda_d} \subseteq G_{\lambda_d}$, I contains no λ_d-eigenvectors, hence λ_d is not an eigenvalue of $A|_I$. Furthermore, since G_{λ_d} is disjoint from the other generalized eigenspaces, $\chi_G(t)$ vanishes if and only if $t = \lambda_d$. Therefore, $A|_I$ has eigenvalues $\lambda_1, \ldots, \lambda_{d-1}$. By **Lemma 3.32.**, $G_{\lambda_k} = (\lambda_d\mathbb{1} - A)^n(G_{\lambda_k}) \subseteq I$ for $1 \leqslant k \leqslant d-1$. In other words, the generalized eigenspaces of $A|_I$ agree with those of A apart from G_{λ_d}. By the inductive hypothesis, $I = \bigoplus_{k=1}^{d-1} G_{\lambda_k}$, and the result follows. $\square$

We can write each $A \in \mathscr{L}(V)$ as the direct sum $A = \bigoplus_\lambda A|_{G_\lambda}$, where λ ranges over the distinct eigenvalues of A. This is the *Jordan decomposition* of A.

COROLLARY 3.35. The dimension of G_λ is the algebraic multiplicity of λ.

Proof. Letting m_λ denote the algebraic multiplicity of λ, we have

$$\prod_\lambda (t-\lambda)^{m_\lambda} = \chi_A(t) = \prod_\lambda \chi_{A|_{G_\lambda}}(t).$$

Since each restriction $A|_{G_\lambda}$ has only one eigenvalue, λ, we have $\chi_{A|_{G_\lambda}}(t) = (t-\lambda)^{m_\lambda}$. The degree of $\chi_{A|_{G_\lambda}}$ is the dimension of G_λ, and the result follows. $\square$

COROLLARY 3.36. If λ is an eigenvalue of A, then $g_\lambda \leqslant m_\lambda$.

Proof. This follows immediately from the inclusion $E_\lambda \subseteq G_\lambda$. $\square$

We conclude with a powerful application of the Jordan decomposition.

THEOREM 3.37 (Cayley-Hamilton)**.** Every linear operator is annihilated by its characteristic polynomial.

Proof. A brief computation shows for any $A \in \mathscr{L}(X)$, $B \in \mathscr{L}(Y)$, and $p(t) \in K[t]$, we have the identity $p(A \oplus B) = p(A) \oplus p(B)$.

Let $A : V \to V$ be a linear operator. For each eigenvalue λ, we know that the restriction $(\lambda\mathbb{1} - A|_{G_\lambda}) \in \mathscr{L}(G_\lambda)$ is nilpotent. Since $\dim G_\lambda = m_\lambda$, **Proposition 3.30.** implies $(\lambda\mathbb{1} - A|_{G_\lambda})^{m_\lambda} = 0$. Using the Jordan decomposition of A, we have

$$\chi_A(A) = \bigoplus_\lambda \chi_A(A|_{G_\lambda}) = \bigoplus_\lambda \prod_{\lambda'} (A|_{G_\lambda} - \lambda'\mathbb{1})^{m_{\lambda'}} = \bigoplus_\lambda \prod_{\lambda'} (-1)^{m_{\lambda'}} (\lambda'\mathbb{1} - A|_{G_\lambda})^{m_{\lambda'}} = 0.$$

$\square$

§3.8. EXTENSION OF SCALARS

Now, algebraic closure is an awfully strong thing to assume. Granted, the results of the previous section hold under the somewhat weaker assumption that χ_A factors completely or, equivalently, $\sum_\lambda m_\lambda = \dim V$. However, this is still a fairly strong condition, and it is worth exploring ways to extend the results of the previous section to spaces over arbitrary fields.

Given a field E, a subset $K \subseteq E$ is a *subfield* if

1. $0, 1 \in K$,
2. $a+b, ab \in K$ for all $a, b \in K$,
3. $-a \in K$ for each $a \in K$
4. $b^{-1} \in K$ for each non-zero $b \in K$.

A *field extension* of K is a field E containing K as a subfield. For example, $\mathbb{C}$ is an extension of $\mathbb{R}$, which is, in turn, an extension of $\mathbb{Q}$.

Note that an extension E of K is naturally a vector space over K. Therefore, given a vector space V over K, we can form the tensor product $E \otimes_K V$. For each $\alpha \in E$, the map $(\beta, x) \mapsto (\alpha\beta) \otimes x$ is bilinear, and there is a unique linear map sending $\beta \otimes x$ to $(\alpha\beta) \otimes x$. As such, we can turn $E \otimes V$ to a vector space over E by defining $\alpha(\beta \otimes x) = (\alpha\beta) \otimes x$ for all $\alpha \in E$. This new vector space is called the *extension of scalars* of V to E and denoted by V_E. It comes with a *canonical inclusion* $\iota_V : V \to V_E, x \mapsto 1 \otimes x$, which is clearly injective, and a universal property.

As a brief bit of terminology, we say that a linear transformation between vector spaces over K is *K-linear*.

THEOREM 3.38 (Universal Property of Extension of Scalars)**.** Let K be a field, and let E be an extension of K. If $T : V \to W$ is K-linear, then there is a unique E-linear map $T_E : V_E \to W_E$ such that $\iota_W \circ T = T_E \circ \iota_V$, *i.e.* the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\iota_V"'] & W \arrow[d, "\iota_W"] \\ V_E \arrow[r, "T_E"] & W_E \end{tikzcd}.$$

Proof. The map $E \times V \to E \otimes W, (\alpha, x) \mapsto \alpha \otimes Tx$ is bilinear, so we have a K-linear transformation $T_E = \widetilde{T} : E \otimes V \to E \otimes W$ by $T_E(\alpha \otimes x) = \alpha \otimes Tx$. Viewing it as a map $V_E \to W_E$, we see

$$T_E(\alpha \otimes x) = \alpha \otimes Tx = \alpha(1 \otimes Tx) = \alpha T_E(1 \otimes x),$$

so T_E is, in fact, E-linear. By its definition,

$$(T_E \circ \iota_V)(x) = T_E(1 \otimes x) = 1 \otimes Tx = (\iota_W \circ T)(x)$$

for all $x \in V$. This completes the proof of existence. Uniqueness follows from the fact that $\operatorname{im}(\iota_V)$ spans V_E. □

The space V_E is a good analog to V in the following sense:

PROPOSITION 3.39. If $\{x_1, \dots, x_n\}$ is a basis of V, then $\{1 \otimes x_1, \dots, 1 \otimes x_n\}$ is a basis of V_E. Consequently, $\dim_K V = \dim_E V_E$.

Proof. For any $\alpha \otimes x \in V_E$, $x = \sum_{k=1}^n \beta_k x_k$, so

$$\alpha \otimes x = \sum_{k=1}^n \alpha\beta_k(1 \otimes x_k),$$

and $\{1 \otimes x_k\}$ spans elementary tensors, hence it spans V_E as well. Let W be a vector space over E with basis $\{w_1, \dots, w_n\}$. The map $T : E \times V \to W$ defined by

$$T\left(\alpha, \sum_{k=1}^n \beta_k x_k\right) = \sum_{k=1}^n \alpha\beta_k w_k$$

is bilinear, so we have a K-linear map $\widetilde{T} : E \otimes V \to W$ satisfying $\widetilde{T}(1 \otimes x_k) = w_k$ for each w_k. For each $\alpha \in E$ and $x = \sum_{k=1}^n \beta_k x_k \in V$,

$$\widetilde{T}(\alpha \otimes x) = \sum_{k=1}^{n} \alpha\beta_k w_k = \alpha \sum_{k=1}^{n} 1 \cdot \beta_k w_k = \alpha\widetilde{T}(1 \otimes x),$$

and $\widetilde{T}$ is E-linear. Since $\widetilde{T}$ sends $\{1 \otimes x_k\}$ to the linearly independent set $\{w_k\}$, $\{1 \otimes x_k\}$ is itself linearly independent. $\square$

To generalize the results of the previous section, we need to understand the relationship between the determinant and extension of scalars.

PROPOSITION 3.40. If V is an n-dimensional vector space over K, if E is an extension of K, and if $A : V \to V$ is a K-linear operator, then $\det A_E = \det A$.

Proof. Consider the map $\varphi : (V_E)^m \to (\bigwedge^m V)_E$ defined on elementary tensors by

$$\varphi(\alpha_1 \otimes x_1, \cdots, \alpha_n \otimes x_m) = \left(\prod_{k=1}^{m} \alpha_k\right) \otimes (x_1 \wedge \cdots \wedge x_m)$$

and extended to V_E by demanding multilinearity. Clearly, φ is alternating, so there is a unique E-linear transformation $\widehat{\varphi} : \bigwedge^m(V_E) \to (\bigwedge^m V)_E$ satisfying

$$\widehat{\varphi}((\alpha_1 \otimes x_1) \wedge \cdots \wedge (\alpha_m \otimes x_m)) = \left(\prod_{k=1}^{m} \alpha_k\right) \otimes (x_1 \wedge \cdots \wedge x_m).$$

We know that $\{(1 \otimes x)_{\wedge I} : I \in [n]^m \text{ is increasing}\}$ is a basis of $\bigwedge^m(V_E)$ and, similarly, $\{1 \otimes (x_{\wedge I}) : I \in [n]^m \text{ is increasing}\}$ is a basis of $(\bigwedge^m V)_E$. A quick computation shows $\widehat{\varphi}$ sends the former to the latter, and it is therefore an isomorphism.

A brief reflection shows $(A_E)^{\wedge n} = \widehat{\varphi}^{-1}(A^{\wedge n})_E\widehat{\varphi}$. Scaling appropriately, it is sufficient to consider n-blades of the form $1 \otimes x_1 \wedge \cdots \wedge 1 \otimes x_n \in \bigwedge^n(V_E)$. Then

$$\begin{aligned}(A_E)^{\wedge n}(1 \otimes x_1 \wedge \cdots \wedge 1 \otimes x_n) &= \widehat{\varphi}^{-1}(A^{\wedge n})_E\widehat{\varphi}(1 \otimes x_1 \wedge \cdots \wedge 1 \otimes x_n)\\ &= \widehat{\varphi}^{-1}(A^{\wedge n})_E(1 \otimes (x_1 \wedge \cdots \wedge x_n))\\ &= \widehat{\varphi}^{-1}(1 \otimes (Ax_1 \wedge \cdots \wedge Ax_n))\\ &= \det(A)\widehat{\varphi}^{-1}(1 \otimes (x_1 \wedge \cdots \wedge x_n))\\ &= \det(A)(1 \otimes x_1 \wedge \cdots \wedge 1 \otimes x_n),\end{aligned}$$

and the result follows. $\square$

COROLLARY 3.41. If $A : V \to V$ is a K-linear operator, E an extension of K, then $\chi_{A_E} = \chi_A$.

Proof. For all $1 \otimes x \in V_E$,

$$(t\mathbb{1} - A)_E(1 \otimes x) = 1 \otimes (tx - Ax) = t(1 \otimes x) - A_E(1 \otimes x) = (t\mathbb{1} - A_E)(1 \otimes x),$$

so $(t\mathbb{1} - A)_E = (t\mathbb{1} - A_E)$. Therefore,

$$\chi_{A_E}(t) = \det(t\mathbb{1} - A_E) = \det(t\mathbb{1} - A) = \chi_A(t).$$

$\square$

From here, it's easy to extend the results of **§1.7.**. We borrow from algebra the fact that every field K has an algebraically-closed extension. This extension is unique (up to isomorphism); we refer to it as the *algebraic closure* of K, denoted $\overline{K}$.

To begin, note that every eigenvalue of A is an eigenvalue of A_E. Indeed, if x is a λ-eigenvector of A, then $A_E(1 \otimes x) = 1 \otimes \lambda x = \lambda(1 \otimes x)$, so $1 \otimes x$ is a λ-eigenvector of A_E. More generally, we see that $(\lambda \mathbb{1} - A_E) = (\lambda \mathbb{1} - A)_E$, and $\alpha \otimes x \in V_E$ is a λ-eigenvector of A_E if and only if x is a λ-eigenvector of A. Consequently, the λ-eigenspace of A_E is $(E_\lambda)_E$, where $E_\lambda = \ker(\lambda \mathbb{1} - A)$. By a similar argument, the generalized λ-eigenspace of A_E is precisely $(G_\lambda)_E$. It follows that A and A_E have the same algebraic and geometric multiplicities for λ. In the case $E = \overline{K}$, **Corollary 3.35.** implies the former bounds the latter in all cases.

As for the Cayley-Hamilton theorem, a quick computation shows for any polynomial $p(t)$ with coefficients in K, $p(A_E) = p(A)_E$. We know $\chi_A = \chi_{A^{\overline{K}}}$ annihilates $A_{\overline{K}}$. Therefore,

$$\chi_A(A_{\overline{K}})(1 \otimes x) = \chi_A(A)_{\overline{K}}(1 \otimes x) = 1 \otimes \chi_A(A)x = 0$$

for all $x \in V$, so $\chi_A(A) = 0$.

4

Space & Angle

Our development so far has relied on the notion of position alone. In this chapter, we introduce notions of length and angle, thereby obtaining essentially Euclidean spaces. Throughout, we assume our vector spaces are real or complex.

§4.1. Inner Products

We want to endow a vector space V with some algebraic gadget that gives it geometric structure. Let's begin by considering the case $V = \mathbb{R}^2$, which we know has natural notions of distance and angle. The idea is to find a bilinear form $\langle \cdot, \cdot \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ that encodes angle in such a way that the map $x \mapsto \langle x, x \rangle$ encodes distance. To preserve canonicity, we use unsigned angles, *i.e.* the angle between x and y is equal to the angle between y and x. From this, we see that $\langle \cdot, \cdot \rangle$ should be symmetric.

Denote by $\|x\|$ the length of the vector x, or, equivalently, the distance between the point x and 0. Suppose we have a symmetric bilinear form $\langle \cdot, \cdot \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ such that $\langle x, x \rangle = \|x\|^2$ for all $x \in \mathbb{R}^2$. Then, for any vectors x and y, we have

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle.$$

Notice that $\|x - y\|$ is the distance between x and y, so we have the law of cosines

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\|y\| \cos \theta_{x,y},$$

where $\theta_{x,y}$ is the angle between x and y. Therefore,

$$\langle x, y \rangle = \|x\|\|y\| \cos \theta_{x,y}. \tag{4.1}$$

As such, $\langle \cdot, \cdot \rangle$, should it exist, uniquely characterizes length and angle in $\mathbb{R}^2$.

On the matter of existence, we see that the map $(x, y) \mapsto x_1 y_1 + x_2 y_2$ is symmetric and bilinear. Noting that the standard basis vectors e_1 and e_2 correspond to vectors on perpendicular axes, we have the Pythagorean theorem:

$$\|x\|^2 = x_1^2 + x_2^2,$$

so the proposed map is indeed the desired one.

From here, we want to describe the class of bilinear forms that encode (Euclidean) geometric structure. Fix a real vector space V. A *inner product* on V is a symmetric bilinear form $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that the map $x \mapsto \langle x, x \rangle$ is *positive-definite*, *i.e.* $\langle x, x \rangle \geqslant 0$, with equality holding if and only if $x = 0$. Each inner product induces a norm $\|\cdot\| : V \to \mathbb{R}$ defined by

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

As before, the distance between x and y is given by $\|x - y\|$, and here we see that the requirement of positive-definiteness is equivalent to the observation that x and y are at distance 0 from each other if and only if $x = y$. Working by analogy to (4.1), we also have a notion of angle given by

$$\theta_{x,y} = \cos^{-1} \frac{\langle x, y \rangle}{\|x\| \|y\|} \in [0, \pi],$$

which is defined for all non-zero vectors $x, y \in V$. From these definitions, the law of cosines is trivially true:

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle = \|x\|^2 + \|y\|^2 - 2\|x\| \|y\| \cos \theta_{x,y},$$

so the geometry in V is essentially Euclidean.

As we saw in the previous chapter, complex spaces are especially nice to work in. Unfortunately, real inner products do not immediately generalize to the complex case. Indeed, if V is a complex vector space and $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ a symmetric bilinear form, then

$$\langle ix, ix \rangle = i\langle x, ix \rangle = -\langle x, x \rangle,$$

so the map $x \mapsto \langle x, x \rangle$ is not positive-definite, except in the trivial case. Fortunately, this can be easily remedied by weakening symmetry.

DEFINITION 4.1. Let V be a complex vector space. A *inner product* on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ satisfying the following conditions:

1. For each y, the map $x \mapsto \langle x, y \rangle$ is linear, *i.e.* $\langle \cdot, \cdot \rangle$ is linear in its first component,

2. For all $x, y \in V$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$,

3. The map $x \mapsto \langle x, x \rangle$ is positive-definite.

A vector space V together with an inner product $\langle \cdot, \cdot \rangle$ is called an *inner product space*.

Replacing $\mathbb{C}$ with $\mathbb{R}$, the reader will see that this definition recovers the original for real vector spaces. Maps satisfying conditions **1.** and **2.** are called *Hermitian*. The reader will verify the identity

$$\langle x, \alpha y + \beta z \rangle = \overline{\alpha} \langle x, y \rangle + \overline{\beta} \langle x, z \rangle,$$

which implies

$$\langle ix, ix \rangle = |i|^2 \langle x, x \rangle = \langle x, x \rangle.$$

As before, inner products induce a notion of length, given by the norm $\|x\| = \sqrt{\langle x, x \rangle}$.

EXAMPLE 19. The reader will verify that $\mathbb{C}^n$ has an inner product given by

$$\langle x, y \rangle = \sum_{k=1}^{n} x_k \overline{y_k}.$$

This is the *dot product* on $\mathbb{C}^n$; we usually write $x \cdot y = \langle x, y \rangle$. Similarly for $\mathbb{R}^n$, we have

$$x \cdot y = \sum_{k=1}^{n} x_k y_k.$$

◇

EXAMPLE 20. From algebra, we know that any non-constant polynomial of degree n has at most n roots. Consequently, for any $p \in \mathbb{C}_n[t]$, $p = 0$ if and only if $p(k) = 0$ for all $0 \leqslant k \leqslant n$. We may therefore define an inner product on $\mathbb{C}_n[t]$ by

$$\langle p, q \rangle = \sum_{k=0}^{n} p(k) \overline{q(k)}.$$

◇

EXAMPLE 21. Consider the space $C[0,1]$ of continuous functions $[0,1] \to \mathbb{R}$. We have a symmetric bilinear form given by

$$\langle f, g \rangle = \int_0^1 f(x) g(x) dx.$$

For all $f \in C[0,1]$,

$$\langle f, f \rangle = \int_0^1 (f(x))^2 dx \geqslant \int_0^1 0 dx = 0,$$

so $\langle \cdot, \cdot \rangle$ is positive. As for definiteness, if $f(p) \neq 0$ for some $p \in [0,1]$, then $f(p) \neq 0$ for some $p \in (0,1)$, and there is $\varepsilon > 0$ such that $f(x) \neq 0$ for all $x \in [p - \varepsilon, p + \varepsilon]$. Then

$$\langle f, f \rangle = \int_0^1 (f(x))^2 dx \geqslant \int_{p-\varepsilon}^{p+\varepsilon} (f(x))^2 dx \geqslant 2\varepsilon \cdot \inf f([p - \varepsilon, p + \varepsilon]) > 0.$$

◇

We conclude by proving some basis properties of inner products.

PROPOSITION 4.2. Let V be a real or complex inner product space. For $x \in V$, $\langle x, y \rangle = 0$ for all $y \in V$ if and only if $x = 0$.

Proof. **($\Longrightarrow$)** In particular, we have $\langle x, x \rangle = 0$, so $x = 0$ by positive-definiteness.
($\Longleftarrow$) For any $y \in V$, $\langle 0, y \rangle = 0 \langle 0, y \rangle = 0$. □

COROLLARY 4.3. If $\langle x_1, y \rangle = \langle x_2, y \rangle$ for all $y \in V$, then $x_1 = x_2$.

§4.2. The Riesz Representation Theorem

Fix an inner product space V. In **§2.5.**, we introduced V' as the "space of rulers" on V. We saw that V and V' are not naturally isomorphic, *i.e.* to obtain an isomorphism $V \to V'$, we must choose a basis or, in other words, choose a measurement system. However, the inner product on V provides measurements. This suggests the presence of a natural correspondence between V and V'.

Theorem 4.4 (Riesz Representation)**.** For each $\varphi \in V'$, there is a unique vector $\varphi^* \in V$ such that $\varphi(x) = \langle x, \varphi^* \rangle$ for all $x \in V$. Furthermore, the correspondence $\varphi \mapsto \varphi^*$ is conjugate-linear and bijective, *i.e.* an *anti-isomorphism*.

Proof. For each $y \in V$, denote by y^* the linear functional defined by $y^*(x) = \langle x, y \rangle$. This yields a conjugate linear map $T : V \to V', y \mapsto y^*$. Note that if $y^* = z^*$, then

$$\langle x, y \rangle = \langle x, z \rangle \implies \langle x, y - z \rangle = 0$$

for all $x \in V$, so $y - z = 0$. It follows that T is injective. As for surjectivity, let $\{x_1, \ldots, x_n\}$ be a basis of V, and suppose

$$0 = \sum_{k=1}^{n} \alpha_k x_k^* = T \sum_{k=1}^{n} \overline{\alpha_k} x_k.$$

Then

$$\sum_{k=1}^{n} \overline{\alpha_k} x_k = 0,$$

and linear independence implies $\alpha_k = \overline{\alpha_k} = 0$ for all k. It follows that $\{x_1^*, \ldots, x_n^*\}$ is a basis of V'. For each $\sum \alpha_k x_k^* \in V'$,

$$\sum_{k=1}^{n} \alpha_k x_k^* = T \sum_{k=1}^{n} \overline{\alpha_k} x_k \in \operatorname{im}(T),$$

so T is surjective, and we are done. □

When V is a real vector space, **Theorem 4.4.** yields a natural isomorphism. However, on account of the technicalities of the complex case, we avoid identifying the two. Instead, we refer to φ^* as the *Riesz vector* of $\varphi \in V'$ and, symmetrically, to y^* as the *Riesz functional* of $y \in V$. Brief reflection yields the identities $\varphi^{**} = \varphi$ and $y^{**} = y$.

Using Riesz vectors, we can give V' a canonical inner product: for $\varphi, \psi \in V'$, we define $\langle \varphi, \psi \rangle = \overline{\langle \varphi^*, \psi^* \rangle} = \langle \psi^*, \varphi^* \rangle$. Then

$$\langle \alpha\varphi_1 + \beta\varphi_2, \psi \rangle = \langle \psi^*, \overline{\alpha}\varphi_1^* + \overline{\beta}\varphi_2^* \rangle = \alpha\langle \varphi_1, \psi \rangle + \beta\langle \varphi_2, \psi \rangle,$$
$$\langle \varphi, \psi \rangle = \langle \psi^*, \varphi^* \rangle = \overline{\langle \varphi^*, \psi^* \rangle} = \overline{\langle \psi, \varphi \rangle},$$

and positive-definiteness follows from the identity $\langle \varphi, \varphi \rangle = \langle \varphi^*, \varphi^* \rangle$. The resulting inner product space is denoted by the symbol V^*.

For $M \subseteq V$ and $N \subseteq V^*$, let

$$M^* = \{x^* : x \in M\} \subseteq V^* \text{ and } N^* = \{\varphi^* : \varphi \in N\} \subseteq V.$$

Brief reflection shows $M^{**} = M$ and $N^{**} = N$, and that if M and N are subspaces, then so are M^* and N^*.

Recall that for each linear operator $A : V \to V$, we have the adjoint operator $A' : V' \to V'$ defined by $A'\varphi(x) = \varphi(Ax)$ for all $x \in V$. In terms of Riesz vectors, this becomes

$$\langle x, (A'\varphi)^* \rangle = A'\varphi(x) = \varphi(Ax) = \langle Ax, \varphi^* \rangle. \tag{4.2}$$

Ideally, these symbols would distribute is some way. Let $T : V' \to V, \varphi \mapsto \varphi^*$ be our canonical correspondence. Then (4.2) becomes

$$\langle Ax, \varphi^* \rangle = \langle x, TA'T^{-1}\varphi^* \rangle.$$

The *Hermitian adjoint* of $A : V \to V$ is the operator $A^* = TA'T^{-1} : V \to V$, which is equivalently defined by

$$\langle Ax, y \rangle = \langle x, A^*y \rangle.$$

We have the following basic properties, which should be familiar from **§2.5.**.

PROPOSITION 4.5. Let $A, B \in \mathscr{L}(V)$. Then

1. $0^* = 0$ and $\mathbb{1}^* = \mathbb{1}$,
2. $(\alpha A + \beta B)^* = \overline{\alpha} A^* + \overline{\beta} B^*$,
3. $(AB)^* = B^* A^*$,
4. If A is invertible, then so is A^*, and $(A^{-1})^* = (A^*)^{-1}$,
5. $A^{**} = A$.

Proof. For all $x, y \in V$, each result follows from **Proposition 4.2.** by the following:

1. $\langle x, 0^* y \rangle = 0$ and $\langle x, \mathbb{1}^* y \rangle = \langle x, y \rangle$.
2. $\langle x, (\alpha A + \beta B)^* y \rangle = \alpha \langle x, A^* y \rangle + \beta \langle x, B^* y \rangle = \langle x, (\overline{\alpha} A^* + \overline{\beta} B^*) y \rangle$.
3. $\langle x, (AB)^* y \rangle = \langle ABx, y \rangle = \langle x, B^* A^* y \rangle$.
4. $\langle A^{**} x, y \rangle = \langle x, A^* y \rangle = \langle Ax, y \rangle$.

□

We also have $(Ay)^* = y^* A^*$ and $(A^* \varphi^*)^* = \varphi A$. Viewing y as a linear transformation $K \to V$, these identities agree with the usual contravariance of adjoints.

We conclude with a result relating Hermitian and algebraic adjoints.

PROPOSITION 4.6. Let $A : V \to V$ be a linear operator. Then

1. $\mathrm{im}(A^*) = (\mathrm{im}(A'))^*$,

2. $\ker(A^*) = (\ker(A'))^*$.

Proof. **1.** If $A^*y \in \mathrm{im}(A^*)$, then

$$(A^*y)^*(x) = \langle x, A^*y\rangle = \langle Ax, y\rangle = y^*(Ax) = A'y^*(x)$$

for all $x \in V$, so $A^*y = (A'y^*)^* \in (\mathrm{im}(A'))^*$. Conversely, if $(A'\varphi)^* \in (\mathrm{im}(A'))^*$, then

$$\langle x, (A'\varphi)^*\rangle = A'\varphi(x) = \varphi(Ax) = \langle Ax, \varphi^*\rangle = \langle x, A^*\varphi^*\rangle$$

for all $x \in V$, so $(A'\varphi)^* = A^*\varphi^* \in \mathrm{im}(A^*)$, and we are done.

2. If $y \in \ker(A^*)$, then

$$A'y^*(x) = y^*(Ax) = \langle Ax, y\rangle = \langle x, A^*y\rangle = 0$$

for all $x \in V$, so $y^* \in \ker(A')$, and $y = y^{**} \in (\ker(A'))^*$. Conversely, if $\varphi^* \in (\ker(A'))^*$, then for all $x \in V$,

$$\langle x, A^*\varphi^*\rangle = \langle Ax, \varphi^*\rangle = \varphi(Ax) = A'\varphi(x) = 0,$$

so $A^*\varphi^* = 0$ and $\varphi^* \in \ker(A^*)$.

□

§4.3. Orthogonality

Recall that the angle between two non-zero vectors x and y in a real inner product space V is given by

$$\theta_{x,y} = \cos^{-1}\frac{\langle x, y\rangle}{\|x\|\|y\|}. \tag{4.3}$$

In this section, we show that this is well-defined.

We say that x and y are *orthogonal* if $\langle x, y\rangle = 0$. In light of the identity above, this corresponds to x and y being perpendicular to each other. A collection of vectors $\{x_1, \ldots, x_m\}$ is *orthonormal* if $\langle x_i, x_j\rangle = \delta_{ij}$ for all $1 \leqslant i, j \leqslant m$. In other words, $\{x_1, \ldots, x_m\}$ is an orthonormal collection if its elements are pairwise orthogonal and of unit length.

Intuitively, orthonormal collections are linearly independent. This is easy to verify: if $\{x_1, \ldots, x_m\}$ is an orthonormal collections, and if $\sum_{k=1}^m \alpha_k x_k = 0$, then, for each x_i,

$$0 = \langle \sum_{k=1}^m \alpha_k x_k, x_i\rangle = \sum_{k=1}^m \alpha_k \delta_{ki} = \alpha_i.$$

It follows that linear combinations of orthonormal collections are unique.

THEOREM 4.7 (Bessel's Inequality)**.** If $\{x_1, \ldots, x_m\}$ is orthonormal, then

$$\sum_{k=1}^{m} |\langle x, x_k \rangle|^2 \leqslant \|x\|^2$$

for all $x \in V$.

Proof. Let $y = x - \sum_{k=1}^{m} \langle x, x_k \rangle x_k$. Then

$$\begin{aligned}
0 \leqslant \|y\|^2 &= \langle x - \sum_{k=1}^{m} \langle x, x_k \rangle x_k, x - \sum_{k=1}^{m} \langle x, x_k \rangle x_k \rangle \\
&= \|x\|^2 - 2 \sum_{k=1}^{m} |\langle x, x_k \rangle|^2 + \sum_{i=1}^{m} \sum_{j=1}^{m} \langle x, x_i \rangle \overline{\langle x, x_j \rangle} \delta_{ij} \\
&= \|x\|^2 - 2 \sum_{k=1}^{m} |\langle x, x_k \rangle|^2 + \sum_{k=1}^{m} |\langle x, x_k \rangle|^2 \\
&= \|x\|^2 - \sum_{k=1}^{m} |\langle x, x_k \rangle|^2.
\end{aligned}$$

□

COROLLARY 4.8 (Parseval's Identity)**.** If x is in the span of orthonormal vectors $x_1, \ldots, x_m$, then

$$x = \sum_{k=1}^{m} \langle x, x_k \rangle x_k.$$

Proof. We have $x = \sum_{k=1}^{m} \alpha_k x_k$ for some scalars α_k. Then, for each $1 \leqslant i \leqslant m$,

$$\langle x, x_i \rangle = \sum_{k=1}^{m} \alpha_k \langle x_k, x_i \rangle = \sum_{k=1}^{m} \alpha_k \delta_{ki} = \alpha_i,$$

and the result follows. □

In light of this result, we ought to consider orthonormal bases. Bases are maximal linearly independent sets and, as we've seen, orthonormality is a special kind of linear independence. Intuitively, maximal orthonormal collections are bases. We say an orthonormal collection is *complete* if it is not contained in a strictly larger such set.

PROPOSITION 4.9. Every complete orthonormal set is a basis.

Proof. If $\{x_1, \ldots, x_n\}$ is orthonormal, then it is linearly independent, so we need only show that complete implies spanning. If we have a vector $x \notin \operatorname{span}(x_1, \ldots, x_n)$, then the vector $y = x - \sum_{k=1}^{n} \langle x, x_k \rangle x_k$ is non-zero and orthogonal to each x_i. Then $\{x_1, \ldots, x_n, y/\|y\|\}$ is a strictly larger orthonormal set; the result follows from the contrapositive. □

The existence of orthonormal bases follows almost immediately. If $x \in V$ is non-zero, then $\{x/\|x\|\}$ is orthonormal, and since the size of orthonormal sets is bounded by the dimension of V, $\{x/\|x\|\}$ is contained in a complete orthonormal set. More generally, we can extend every orthonormal set to an orthonormal basis.

We provide an alternative (constructive) proof of the existence of orthonormal bases, known as the *Gram-Schmidt process*. Let $\{x_1, \ldots, x_n\}$ be any basis of V, and let $y_1 = x_1/\|x_1\|$, so $\{y_1\}$ is orthonormal. For each $2 \leqslant i \leqslant n$, let

$$y_i' = x_i - \sum_{k=1}^{i-1} \langle x_i, y_k \rangle y_k.$$

Then y_i' is orthogonal to each y_k, so the set $\{y_1, \ldots, y_{i-1}, y_i = y_i'/\|y_i'\|\}$ is orthonormal. The resulting collection $\{y_1, \ldots, y_n\}$ is an orthonormal basis of V.

We can use the results of this section to prove some key properties of inner products and norms.

THEOREM 4.10 (Cauchy-Schwarz Inequality)**.** For all $x, y \in V$,

$$|\langle x, y \rangle| \leqslant \|x\| \|y\|.$$

Proof. If $y = 0$, then the argument is trivial. Suppose, then, that $y \neq 0$, in which case

$$|\langle x, y/\|y\| \rangle|^2 \leqslant \|x\|^2,$$

by Bessel's inequality. Rearranging, this becomes

$$|\langle x, y \rangle|^2 \leqslant \|x\|^2 \|y\|^2,$$

and the result follows. □

It follows that (4.3) is well-defined.

COROLLARY 4.11 (Triangle Inequality)**.** For all $x, y \in V$,

$$\|x + y\| \leqslant \|x\| + \|y\|.$$

Proof. We compute

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

and the result follows. □

Note that the distance between points x and y is given by $\|x - y\|$. The triangle inequality can be rephrased as follows: if $x, y, z \in V$, then

$$\|x - z\| \leqslant \|x - y\| + \|y - z\|.$$

In other words, detours never decrease distance.

We now turn our attention to subspaces. Two subspaces X and Y are *orthogonal* if $\langle x, y\rangle = 0$ for all $x \in X$ and $y \in Y$. Clearly, orthogonal subspaces are disjoint. Indeed, if $x \in X \cap Y$, then $\langle x, y\rangle = 0$ for all $y \in Y$. In particular, $\langle x, x\rangle = 0$, so $x = 0$ by positive-definiteness.

If $M \subseteq V$, the *orthogonal complement* of M is the set

$$M^\perp = \{x \in V : \langle x, m\rangle = 0 \text{ for all } m \in M\}.$$

If $x, y \in M^\perp$, then

$$\langle \alpha x + \beta y, m\rangle = \alpha\langle x, m\rangle + \beta\langle y, m\rangle = 0$$

for all $m \in M$, so $\alpha x + \beta y \in M^\perp$. That is, $M^\perp$ is always a subspace of V; in fact, it is the largest subspace orthogonal to M.

THEOREM 4.12 (Projection Theorem)**.** If M is a subspace of V, then $V = M \oplus M^\perp$, and $M^{\perp\perp} = M$.

Proof. Since M and $M^\perp$ are orthogonal subspaces, they are disjoint. Choose an orthonormal basis $\{x_1, \ldots, x_m\}$ of M, and extend it to an orthonormal basis $\{x_1, \ldots, x_n\}$ of V. Then the vectors $x_{m+1}, \ldots, x_n$ are orthogonal to M, hence $\{x_{m+1}, \ldots, x_n\} \subseteq M^\perp$, and $V = M \oplus M^\perp$.

As for the second claim, M is orthogonal to $M^\perp$, so $M \subseteq M^{\perp\perp}$. We have

$$M \oplus M^\perp = V = M^{\perp\perp} \oplus M^\perp,$$

so $\dim M = \dim M^{\perp\perp}$, hence $M = M^{\perp\perp}$. □

Consequently, each subspace of V yields a canonical direct sum decomposition.

DEFINITION 4.13. The *perpendicular projection* onto M is the projection $E_M = P_{M,M^\perp}$ onto M along $M^\perp$.

PROPOSITION 4.14. A linear operator $E : V \to V$ is a perpendicular projection if and only if $E = E^2 = E^*$.

Proof. **($\Longrightarrow$)** In the first place, E^* is a projection, since $(E^*)^2 = (E^2)^* = E^*$. Brief reflection shows that if X is a subspace of V, then $(X^0)^* = X^\perp$. By **Proposition 2.29.**, E' projects onto $(M^\perp)^0$ along M^0, *i.e.* $\text{im}(E') = (M^\perp)^0$ and $\ker(E') = M^0$. By **Proposition 4.6.**,

$$\text{im}(E^*) = ((M^\perp)^0)^* = M^{\perp\perp} = M \text{ and } \ker(E^*) = (M^0)^* = M^\perp,$$

so $E^* = P_{M,M^\perp} = E$.

($\Longleftarrow$) If $E = E^2 = E^*$, then E is a projection, so we need only show that $\text{im}(E)$ is orthogonal to $\ker(E)$. Let $Ex \in \text{im}(E)$ and $y \in \ker(E)$ be arbitrary. Then

$$\langle Ex, y\rangle = \langle x, E^* y\rangle = \langle x, Ey\rangle = 0,$$

and we are done. □

Perpendicular projections are our first example of *self-adjoint* operators, *i.e.* transformations $A : V \to V$ satisfying $A^* = A$. We will see later on that they are the building blocks of all such operators.

§4.4. Unitary Operators

Let V be an inner product space. Linear operators $V \to V$ do not, in general, preserve the geometric structure of V. In this section, we study the operators that do.

Definition 4.15. A linear operator $U : V \to V$ is *unitary* if $\langle Ux, Uy \rangle = \langle x, y \rangle$ for all $x, y \in V$.

Proposition 4.16. Let $U : V \to V$ be a linear operator. The following conditions are equivalent:

1. U is unitary,
2. $U^* = U^{-1}$,
3. $\|Ux\| = \|x\|$ for all $x \in V$.

Proof. **(1. $\implies$ 2.)** Let $\{x_1, \ldots, x_n\}$ be an orthonormal basis of V. Then

$$\langle Ux_i, Ux_j \rangle = \langle x_i, x_j \rangle = \delta_{ij}$$

for all $1 \leqslant i, j \leqslant n$, so $\{Ux_1, \ldots, Ux_n\}$ is an orthonormal basis of V, and U is invertible. Furthermore, for each x and all $y \in V$,

$$\langle U^*Ux, y \rangle = \langle Ux, Uy \rangle = \langle x, y \rangle,$$

hence $U^*U = \mathbb{1}$. It follows that $U^* = U^{-1}$.

(2. $\implies$ 3.) For each $x \in V$,

$$\|Ux\|^2 = \langle Ux, Ux \rangle = \langle U^*Ux, x \rangle = \langle x, x \rangle = \|x\|^2.$$

(3. $\implies$ 1.) Note that

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\mathrm{Re}(\langle x, y \rangle)$$

and

$$\|ix + y\|^2 = \|x\|^2 + \|y\|^2 - 2\mathrm{Im}(\langle x, y \rangle)$$

for all $x, y \in V$. We may then compute

$$\|Ux + Uy\|^2 = \|x\|^2 + \|y\|^2 + 2\mathrm{Re}(\langle Ux, Uy \rangle),$$
$$\|Uix + Uy\|^2 = \|x\|^2 + \|y\|^2 - 2\mathrm{Im}(\langle Ux, Uy \rangle).$$

Equating with the identities above yields

$$\mathrm{Re}(\langle Ux, Uy\rangle) = \mathrm{Re}(\langle x, y\rangle) \text{ and } \mathrm{Im}(\langle Ux, Uy\rangle) = \mathrm{Im}(\langle x, y\rangle),$$

and the result follows. $\square$

PROPOSITION 4.17. If λ is an eigenvalue of a unitary operator $U : V \to V$, then $|\lambda| = 1$.

Proof. Let x be an eigenvector of λ. Then

$$\|x\| = \|Ux\| = \|\lambda x\| = |\lambda|\|x\|,$$

which implies $|\lambda| = 1$. $\square$

COROLLARY 4.18. If $U : V \to V$ is a unitary operator, then $|\det(U)| = 1$.

Proof. Since the determinant and characteristic polynomial of U is unchanged by extending the scalars of V, we may assume V is a complex vector space. Then

$$\chi_U(t) = \prod_{\lambda}(t - \lambda)^{m_\lambda} \tag{4.4}$$

has constant term $(-1)^{\dim V}\det(U)$. Evaluating at $t = 0$ and rearranging yields the identity

$$\det(U) = (-1)^{\dim V}\prod_{\lambda}(-1)^{m_\lambda}\lambda^{m_\lambda},$$

where λ ranges over the distinct eigenvalues of U, so each has absolute value 1. Therefore, $|\det(U)| = 1$. $\square$

When $\det(U) = 1$, we say U is *special unitary*. Special unitary operators preserve length, angle, volume, and orientation. That is, geometrically, they act by rotation.

§4.5. THE SPECTRAL THEOREM

As we saw in the previous section, spectral theory has geometric significance in inner product spaces. In this section, we prove the main result of this chapter, a refinement of the Jordan decomposition.

In Chapter 3, we saw that operators over algebraically-closed fields have especially nice spectra. The fundamental theorem of algebra states that $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$. If V is a real vector space, we refer to the extension of scalars $V_{\mathbb{C}}$ as the *complexification* of V. We can write each element of $V_{\mathbb{C}}$ in the form

$$\sum(\alpha_k + i\beta_k) \otimes x_k = 1 \otimes \left(\sum \alpha_k x_k\right) + i \otimes \left(\sum \beta_k x_k\right).$$

In other words, each vector is of the form $1 \otimes x + i \otimes y$, with $x, y \in V$. We often omit the tensor products, in which case $V_{\mathbb{C}}$ is the space of formal combinations $V + iV$, with complex scaling defined by

$$(\alpha + i\beta)(x + iy) = (\alpha x - \beta y) + i(\beta x + \alpha y).$$

If V is an inner product space, then $V_{\mathbb{C}}$ has a natural inner product given by

$$\langle x+iy, z+iw\rangle_{\mathbb{C}} = \langle x,z\rangle + \langle y,w\rangle + i(\langle y,z\rangle - \langle x,w\rangle).$$

In particular,

$$\|x+iy\|_{\mathbb{C}} = \sqrt{\|x\|^2 + \|y\|^2}.$$

PROPOSITION 4.19. If V is a real inner product space, and if $A \in \mathscr{L}(V)$, then $(A_{\mathbb{C}})^* = (A^*)_{\mathbb{C}}$.

Proof. For all $x+iy, z+iw \in V_{\mathbb{C}}$, we may easily compute

$$\begin{aligned}\langle x+iy, (A_{\mathbb{C}})^*(z+iw)\rangle_{\mathbb{C}} &= \langle A_{\mathbb{C}}(x+iy), z+iw\rangle_{\mathbb{C}} = \langle Ax+iAy, z+iw\rangle_{\mathbb{C}} \\ &= \langle x+iy, A^*z+iA^*w\rangle_{\mathbb{C}} = \langle x+iy, (A^*)_{\mathbb{C}}(z+iw)\rangle_{\mathbb{C}}.\end{aligned}$$

□

It follows that A is self-adjoint (resp. unitary) if and only if $A_{\mathbb{C}}$ is.

With extension of scalars extended to inner products, we turn our attention to the main argument. To start, fix an inner product space V.

LEMMA 4.20. If $A \in \mathscr{L}(V)$ is self-adjoint, then all of the eigenvalues of A are real. In fact, all of the roots of χ_A in $\mathbb{C}$ are real.

Proof. Since $A_{\mathbb{C}}$ is self-adjoint, and since $\chi_{A_{\mathbb{C}}} = \chi_A$, it is sufficient to consider the complex case, where the eigenvalues of A coincide with the roots of χ_A. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A with eigenvector x. Then

$$\lambda\|x\|^2 = \langle Ax, x\rangle = \langle x, Ax\rangle = \overline{\lambda}\|x\|^2,$$

so $\lambda = \overline{\lambda}$. □

LEMMA 4.21. If $A : V \to V$ is a linear operator, and if $M \subseteq V$ is a subspace with $A(M) \subseteq M$, then $A^*(M^\perp) \subseteq M^\perp$.

Proof. For each $y \in M^\perp$ and all $x \in M$,

$$\langle x, A^*y\rangle = \langle Ax, y\rangle = 0,$$

so $A^*y \in M^\perp$. □

LEMMA 4.22. If $A : V \to V$ is self-adjoint, then eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof. Let λ_1, λ_2 be distinct eigenvalues of A with respective eigenvectors x and y. Then

$$\lambda_1\langle x,y\rangle = \langle Ax, y\rangle = \langle x, Ay\rangle = \lambda_2\langle x,y\rangle,$$

which yields

$$(\lambda_1 - \lambda_2)\langle x,y\rangle = 0.$$

Since $\lambda_1 \neq \lambda_2$, this implies $\langle x, y \rangle = 0$. □

THEOREM 4.23 (Spectral Theorem for Self-Adjoint Operators)**.** Let $A : V \to V$ be a self-adjoint operator with distinct eigenvalues $\lambda_1, \ldots, \lambda_d$. Then A is diagonalizable, and there are pairwise orthogonal projections $E_1, \ldots, E_d$ such that $A = \sum_{k=1}^{d} \lambda_k E_k$.

Proof. For the first claim, it is sufficient to show that the algebraic and geometric multiplicities of each λ_k agree. Let λ be an eigenvalue of A, and consider the λ-eigenspace E_λ. Clearly, $A(E_\lambda) \subseteq E_\lambda$, so $A(E_\lambda^\perp) \subseteq E_\lambda^\perp$ by **Lemma 4.22.**. We have the restrictions $B = A|_{E_\lambda}$ and $B_\perp = A|_{E_\lambda^\perp}$, so $A = B \oplus B^\perp$. Then λ is the only eigenvalue of B, and $\chi_B(t) = (\lambda - t)^{g_\lambda}$. Furthermore, λ is not an eigenvalue of $B_\perp$, so $(\lambda - t)$ does not divide $\chi_{B_\perp}(t)$, and the identity

$$\chi_A(t) = \chi_B(t) \cdot \chi_{B_\perp}(t) = (\lambda - t)^{g_\lambda} (\chi_{B_\perp})(t)$$

tells us that the g_λ is the algebraic multiplicity of λ.

For each $1 \leqslant k \leqslant d$, let E_k denote the perpendicular projection onto the λ_k-eigenspace E_{λ_k}. Then $E_1, \ldots, E_d$ are pairwise orthogonal by **Lemma 4.23.**. Since A is diagonalizable, $V = \bigoplus_{k=1}^{d} E_{\lambda_k}$, so we may take each element of V as a linear combination $x = \sum_{k=1}^{d} x_k$, where each x_k is a λ_k-eigenvector. Then

$$Ax = \sum_{k=1}^{d} Ax_k = \sum_{k=1}^{d} \lambda_k x_k = \sum_{k=1}^{d} \lambda_k E_k x,$$

so $A = \sum_{k=1}^{d} \lambda_k E_k$. □

The decomposition $A = \sum \lambda_k E_k$ is the *spectral form* of A.

When our underlying space is real, the above result is an equivalence: each operator of the form $\sum \alpha_k E_k$, for real α_k and pairwise orthogonal E_k, is self-adjoint. In the complex case, however, the spectral form corresponds to a more general class of operators. Note that

$$\left(\sum \lambda_k E_k\right) \left(\sum \lambda_k E_k\right)^* = \left(\sum \lambda_k E_k\right) \left(\sum \overline{\lambda_k} E_k\right) = \sum |\lambda_k|^2 E_k = \left(\sum \lambda_k E_k\right)^* \left(\sum \lambda_k E_k\right).$$

In general, an operator $A : V \to V$ with the property $AA^* = A^*A$ is called *normal*. Important examples include self-adjoint and unitary operators.

LEMMA 4.24. If $A : V \to V$ is normal, then λ is an eigenvalue of A if and only if $\overline{\lambda}$ is an eigenvalue of A^*.

Proof. Observe that $(\lambda \mathbb{1} - A)^* = \overline{\lambda} \mathbb{1} - A^*$, and

$$(\lambda \mathbb{1} - A)(\lambda \mathbb{1} - A)^* = |\lambda|^2 \mathbb{1} - \overline{\lambda} A - \lambda A^* + AA^* = (\lambda \mathbb{1} - A)^* (\lambda \mathbb{1} - A).$$

Note that, for any normal operator N,

$$\|Nx\|^2 = \langle Nx, Nx \rangle = \langle N^*x, N^*x \rangle = \|N^*x\|^2$$

for all $x \in V$. Consequently,

$$\|\lambda x - Ax\| = \|\overline{\lambda} x - A^* x\|,$$

and $x \in V$ is a λ-eigenvector of A if and only if it is an $\overline{\lambda}$-eigenvector of A^*. $\square$

COROLLARY 4.25. If $A : V \to V$ is normal, and if λ is an eigenvector of A, then $E_\lambda = E_{\overline{\lambda}}$.

LEMMA 4.26. If $A : V \to V$ is normal, then eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof. Let λ_1 and λ_2 be distinct eigenvalues of A, with corresponding eigenvectors x and y. Then

$$\lambda_1 \langle x, y \rangle = \langle Ax, y \rangle = \langle x, A^* y \rangle = \lambda_2 \langle x, y \rangle,$$

and the result follows as before. $\square$

THEOREM 4.27 (Spectral Theorem for Normal Operators)**.** Let V be a complex vector space, and $A : V \to V$ a normal operator with distinct eigenvalues $\lambda_1, \dots, \lambda_d$. Then A is diagonalizable, and there are pairwise orthogonal projections $E_1, \dots, E_d$ such that $A = \sum_{k=1}^{d} \lambda_k E_k$.

Proof. From here, the argument is essentially the same as for the self-adjoint case. $\square$

5

COMPUTATIONS & COORDINATES

Way back in **§1.4.**, we noted that every n-dimensional vector space over K is isomorphic to K^n. We have so far avoided coordinates. However, in this chapter, we make full use of them to develop computational tools. Many of the results follow readily from our previous work.

§5.1. MATRICES

Fix a field K. Coordinates spaces K^n come with standard choices of bases $\{e_1, \dots, e_n\}$ defined by $e_k = (\delta_{ik})_{i \in [n]}$. The utility of this basis comes from notational simplicity. For example, we have the identity

$$(\alpha_k)_{k \in [n]} = \sum_{k=1}^{n} \alpha_k e_k.$$

We can study linear transformations $T : K^n \to K^m$ by looking at where they send the standard basis. The *matrix* of T is the family $[T] = (\alpha_{ij})_{i \in [m], j \in [n]}$ of scalars defined by

$$(\alpha_{ik})_{i \in [m]} = Te_k$$

for all k.

This definition becomes clearer when we write (α_{ij}) as an array:

$$(\alpha_{ij}) = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix}.$$

Each vector in K^n is naturally a linear transformation $K \to K^n$. Taking matrices yields the identity

$$[(\alpha_1, \dots, \alpha_n)] = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Moving forward, we will interchangeably describe vectors in K^n as columns, like above, or as tuples, like before. With the change of notation, we see that the matrix of a linear transformation T is an array whose columns are the vectors $Te_1, \ldots, Te_n$.

Now, from this definition, we see that $[\xi A + \eta B] = \xi[A] + \eta[B]$ for all linear transformations $A, B : K^n \to K^m$ and scalars $\xi, \eta \in K$. Clearly, $[A] = 0$ implies $A = 0$. Write $M_{m\times n}(K)$ for the space of matrices with m rows and n columns, *i.e.* families of scalars $(\alpha_{ij})_{i\in[m],j\in[n]}$, under component-wise addition and scaling. Clearly, this space is mn-dimensional, so taking matrices yields an isomorphism $\mathscr{L}(K^n, K^m) \to M_{m\times n}(K)$.

We have additional structure in the form of a bilinear map

$$\mathscr{L}(K^p, K^m) \times \mathscr{L}(K^n, K^p) \to \mathscr{L}(K^n, K^m),\ (A, B) \mapsto AB,$$

which appears as a product structure when $n = p = m$. Through the above isomorphism, this induces a bilinear map

$$M_{m\times p}(K) \times M_{p\times n}(K) \to M_{m\times n}(K),\ ([A],[B]) \mapsto [AB].$$

When A is invertible as a linear transformation, $[A]$ has an inverse with respect to this product given by $[A]^{-1} = [A^{-1}]$.

To see what this product looks like, first consider the case $[A] = (\alpha_{ij}) \in M_{m\times n}(K)$ and $B = (\beta_k) \in K^n$. Then

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^n \alpha_{1k}\beta_k \\ \vdots \\ \sum_{k=1}^n \alpha_{mk}\beta_k \end{pmatrix}.$$

More generally, if $[A] = (\alpha_{ij}) \in M_{m\times p}(K)$ and $[B] = (\beta_{ij}) \in M_{p\times n}(K)$, this map looks like

$$\begin{aligned} \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1p} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mp} \end{pmatrix} \begin{pmatrix} \beta_{11} & \cdots & \beta_{1n} \\ \vdots & \ddots & \vdots \\ \beta_{p1} & \cdots & \beta_{pn} \end{pmatrix} &= \begin{pmatrix} [A]\begin{pmatrix} \beta_{11} \\ \vdots \\ \beta_{p1} \end{pmatrix} & \cdots & [A]\begin{pmatrix} \beta_{1n} \\ \vdots \\ \beta_{pn} \end{pmatrix} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{k=1}^p \alpha_{1k}\beta_{k1} & \cdots & \sum_{k=1}^p \alpha_{1k}\beta_{kn} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^p \alpha_{mk}\beta_{k1} & \cdots & \sum_{k=1}^p \alpha_{mk}\beta_{kn} \end{pmatrix} \end{aligned}$$

In other words, $(\alpha_{ij})(\beta_{ij}) = (\gamma_{ij})$, where $\gamma_{ij} = \sum_{k=1}^p \alpha_{ik}\beta_{kj}$ for all $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$. The author promises that this mess of indices pays off in practice.

Recall from **§1.4.** that, to each vector space V with basis $X = \{x_1, \ldots, x_n\}$, there is an isomorphism $\varphi_X : V \to K^n$ defined by

$$\sum_{k=1}^n \alpha_k x_k \xmapsto{\varphi_X} (\alpha_1, \ldots, \alpha_n),$$

so, in particular, $\varphi_X(x_k) = e_k$ for all k. This is the *coordinate isomorphism* corresponding to X.

DEFINITION 5.1. Let $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ be bases of V and W respectively. The *matrix* of a linear transformation $T : V \to W$ with respect to X and Y is the family of scalars $[T]_X^Y = [\varphi_Y T \varphi_X^{-1}]$. Equivalently, $[T]_X^Y = (\alpha_{ij})$ is the set of scalars satisfying $Tx_k = \sum_{i=1}^m \alpha_{ik} y_i$.

For the sake of succinctness, when $V = W$ and $X = Y$, we write $[T]_X^X = [T]_X$. If X and Y are implicit, they may be omitted altogether.

PROPOSITION 5.2. If $A : V \to W$ and $B : U \to V$ are linear transformations, then $[A]_Y^Z [B]_X^Y = [AB]_X^Z$.

Proof. We have

$$[A]_Y^Z [B]_X^Y = [\varphi_Z A \varphi_Y^{-1}][\varphi_Y B \varphi_X^{-1}] = [\varphi_Z AB \varphi_X^{-1}] = [AB]_X^Z.$$

□

As an immediate consequence of this, we see that $[A]_{X'}^{Y'} = [\mathbb{1}_V]_Y^{Y'} [A]_X^Y [\mathbb{1}_V]_{X'}^X$ for linear operators $A : V \to V$ and bases X, X', Y, Y' of V. The matrix $[\mathbb{1}_V]_{X'}^X = [\varphi_X \varphi_{X'}^{-1}]$ is the *change of basis matrix* from X' to X, and it sends the coordinates of x with respect to X' to those with respect to X. Note that the columns of $[\mathbb{1}_V]_{X'}^X$ consist of the coordinates in X of each basis vector in X'.

A word is needed on inverses of matrices. If $A : V \to W$ is some linear transformation, then the inverse of the matrix $[A]_X^Y$ is given by $[A^{-1}]_Y^X$. For example, the inverse of the change of basis matrix $[\mathbb{1}_V]_X^Y$ is $[\mathbb{1}_V]_Y^X$.

§5.2. GAUSSIAN ELIMINATION

Turning our attention to more practical matters, much of the utility of matrices comes from a classical interpretation. Much of linear algebra amounts to studying the spaces of solutions of equations of the form $Ax = b$. If $[A]_X^Y = (\alpha_{ij})$, $\varphi_X(x) = (x_1, \ldots, x_n)$, and $\varphi_Y(b) = (b_1, \ldots, b_m)$, this equality looks like

$$\begin{pmatrix} \sum_{k=1}^n \alpha_{1k} x_k \\ \vdots \\ \sum_{k=1}^n \alpha_{mk} x_k \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

In other words, $Ax = b$ corresponds to a system of m linear equations in n variables. The benefit of this perspective is that it tells us that computational linear algebra really boils down to solving systems of linear equations.

Consider 3×3 case below:

$$\begin{aligned} \alpha_{11}x + \alpha_{12}y + \alpha_{13}z &= b_1 \\ \alpha_{22}y + \alpha_{23}z &= b_2 \\ \alpha_{33}z &= b_3. \end{aligned}$$

Assuming non-vanishing coefficients as needed, we can solve for x, y, and z by *back substitution*:

$$
\begin{aligned}
z &= \frac{1}{\alpha_{33}} b_3, \\
y &= \frac{1}{\alpha_{22}} b_2 - \frac{\alpha_{23}}{\alpha_{22}} z = \frac{1}{\alpha_{22}} b_2 - \frac{\alpha_{23}}{\alpha_{22}\alpha_{33}} b_3, \\
x &= \frac{1}{\alpha_{11}} b_1 - \frac{\alpha_{12}}{\alpha_{11}} y - \frac{\alpha_{13}}{\alpha_{11}} z = \frac{1}{\alpha_{11}} b_1 - \frac{\alpha_{12}}{\alpha_{11}\alpha_{22}} b_2 - \frac{\alpha_{12}\alpha_{23} - \alpha_{13}\alpha_{22}}{\alpha_{11}\alpha_{22}\alpha_{33}} b_3.
\end{aligned}
$$

While somewhat tedious, this is a straightfoward computation. As such, it is a reasonable strategy to solve linear systems by reducing them to ones of this form. Equivalently, we want to reduce general matrices to *upper triangular* ones, *i.e.* matrices $(\alpha_{ij}) \in M_{m\times n}(K)$ satisfying $\alpha_{ij} = 0$ whenever $i > j$.

A linear transformation $E : K^n \to K^n$ is *elementary* if there is some $1 \leqslant i \leqslant n$ and $j \neq i$ such that $Ee_i = e_i + \xi e_j$ and $Ee_k = e_k$ for all $k \neq i$. It's easy to see that every elementary tranformation $E : e_i \mapsto e_i + \xi e_j$ has an elementary inverse given by $E^{-1} : e_i \mapsto e_i - \xi e_j$.

If we have some matrix $[A] = (\alpha_{ij}) \in M_{m\times n}$ and elementary transformation $E : K^m \to K^m$, $e_i \mapsto e_i + \xi e_j$, then

$$
[E][A] = \begin{pmatrix} | & & | \\ EAe_1 & \cdots & EAe_n \\ | & & | \end{pmatrix},
$$

so the jth row of this matrix has pth component corresponding the jth component of EAe_p. We may compute

$$
EAe_p = \sum_{k=1}^{m} \alpha_{kp} E e_k = \begin{pmatrix} \alpha_{1p} \\ \vdots \\ \alpha_{jp} + \xi\alpha_{ip} \\ \vdots \\ \alpha_{mp} \end{pmatrix}.
$$

That is, left multiplication by $[E]$ corresponds to adding ξ times the ith row of a matrix to its jth row. Brief reflection shows $[E] = (\gamma_{pq})$ has the following form: $\gamma_{kk} = 1$ for $1 \leqslant k \leqslant m$, $\alpha_{ji} = \xi$, and all other components vanish.

Each permutation $\sigma \in S_n$ gives rise to a *permutation transformation* $P : K^n \to K^n$ defined by $Pe_k = e_{\sigma(k)}$. Clearly, all permutation transformations are invertible; if P is the transformation corresponding to $\sigma \in S_n$, then P^{-1} is the transformation corresponding to σ^{-1}. Brief reflection shows left multiplication by $[P]$ rearranges the rows of a matrix according to the corresponding permutation.

We say two matrices $[A]$ and $[B]$ are *elementarily equivalent* if there is some product of elementary and permutation transformations E such that $[E][A] = [B]$.

THEOREM 5.3 (Gaussian Elimination)**.** Every square matrix is elementarily equivalent to an upper triangular matrix.

Proof. The proof will go by induction on the number of rows. For $n = 1$, the argument is trivial. Suppose, then, that we have some $n \geqslant 1$ such that all $n \times n$ matrices are elementarily equivalent to upper triangular ones, and let $[A] = (\alpha_{ij}) \in M_{n+1\times n+1}(K)$

be arbitrary. Let $[A]_{ij}$ denote the *cofactor matrix* obtained by removing the ith row and jth column of $[A]$.

We first address the case that $\alpha_{11} \neq 0$. There is an elementary transformation E_1 that subtracts a suitable multiple of row 1 from row $(n+1)$ such that $\alpha_{(n+1)1} = 0$. The matrix $[E_1 A]_{(n+1)(n+1)}$ is $n \times n$, so there is some product of elementary and permutation transformations $E_2 : K^{n+1} \to K^{n+1}$, which fixes e_{n+1} and thus admits a suitable restriction to K^n, such that $[E_2][E_1 A]_{(n+1)(n+1)}$ is upper triangular. Similarly, there is another such product $E_3 : K^{n+1} \to K^{n+1}$ such that $[E_3][E_2 E_1 A]_{11}$ is upper triangular. Then $[E_3 E_2 E_1][A]$ is upper triangular.

Alternatively, suppose $\alpha_{11} = 0$. If $\alpha_{(n+1)1} = 0$, then the argument above still applies. However, if $\alpha_{(n+1)1} \neq 0$, then we have a problem. This can be resolved by swapping rows 1 and $(n+1)$, and we are done. $\square$

COROLLARY 5.4. Every matrix is elementarily equivalent to an upper triangular matrix.

Proof. Consider a matrix $[A] \in M_{m\times n}(K)$. If $m > n$, so $[A]$ has more rows than columns, then it is sufficient to make the submatrix consisting of the bottom-most n rows upper triangular; the existence of such a reduction follows from the above. Similarly, if $m < n$, so $[A]$ has more columns than rows, we need only make the submatrix consisting of the left-most m columns upper triangular, and this again follows. $\square$

Now, since the transformations E such that $[E][A]$ is upper triangular are invertible, this reduction is very useful for computing bases for the image and kernel of A. This is due to the identities $\mathrm{im}(EA) = E(\mathrm{im}(A))$ and $\ker(EA) = \ker(A)$. For example, if $\{x_1, \dots, x_n\}$ is a basis of $\mathrm{im}(EA)$, the set $\{E^{-1}x_1, \dots, E^{-1}x_n\}$ is a basis of $\mathrm{im}(A)$.

EXAMPLE 22 (Gaussian Elimination)**.** Consider the system

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{5.1}$$

the solution space of which is the kernel of a transformation with the above matrix. We employ the *augmented matrix* notation

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 2 & 3 & 1 & 0 \\ 3 & 1 & 2 & 0 \end{array}\right).$$

We can apply row operations to this matrix directly, which correspond to multiplying both sides of (5.1) by elementary transformations. In practice, this looks like:

$$
\begin{aligned}
&\left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 2 & 3 & 1 & 0 \\ 3 & 1 & 2 & 0 \end{array}\right) \\
\to &\left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 2 & 3 & 1 & 0 \\ 0 & -5 & -7 & 0 \end{array}\right) \text{ (Subtract 3 times the first row from the third)} \\
\to &\left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -1 & -5 & 0 \\ 0 & -5 & -7 & 0 \end{array}\right) \text{ (Subtract 2 times the first row from the second)} \\
\to &\left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -1 & -5 & 0 \\ 0 & 0 & 18 & 0 \end{array}\right) \text{ (Subtract 5 times the second row from the third).}
\end{aligned}
$$

Thus

$$
\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}
$$

if and only if

$$
\begin{aligned}
x + 2y + 3z &= 0 \\
-y - 5z &= 0 \\
18z &= 0,
\end{aligned}
$$

from which we infer $z = y = x = 0$, so our kernel is trivial. ◇

§5.3. COMPUTING INVARIANTS

Throughout this section, fix a linear operator $A : V \to V$.

In **§3.3.**, we introduced the trace as the canonical evaluation $V^* \otimes V \to K$. It is not at all clear from this definition how we can compute $\mathrm{tr}(A)$. We begin this section by expressing the trace as a function of matrices.

THEOREM 5.5. If $X = \{x_1, \ldots, x_n\}$ is a basis of V and $(\alpha_{ij}) = [A]_X$, then

$$
\mathrm{tr}(A) = \sum_{k=1}^{n} \alpha_{kk}.
$$

Proof. For each $1 \leqslant i \leqslant n$, let $\varphi_k : V \to K$ be the linear functional defined by $\varphi_i(x_k) = \delta_{ik}$. Then

$$
Ax_k = \sum_{i=1}^{n} \varphi_i(x_k) A x_k,
$$

so

$$A\sum_{k=1}^{n}\xi_k x_k = \sum_{k=1}^{n}\xi_k A x_k = \sum_{k=1}^{n}\sum_{i=1}^{n}\xi_k\varphi_i(x_k)Ax_i = \sum_{i=1}^{n}\varphi_i\left(\sum_{k=1}^{n}\xi_k x_k\right)Ax_i,$$

and $A=\sum_{k=1}^{n}\varphi_k(\cdot)Ax_k$. Then

$$\operatorname{tr}(A) = \sum_{k=1}^{n}\varphi_k(Ax_k) = \sum_{k=1}^{n}\sum_{i=1}^{n}\alpha_{ik}\varphi_k(x_i) = \sum_{k=1}^{n}\alpha_{kk}.$$

□

Strikingly, this tells us that the sum of the diagonal of any matrix of A is equal to that of any other matrix of A.

Turning our attention to the determinant, we have the following basic identity:

THEOREM 5.6. For any basis $X=\{x_1,\ldots,x_n\}$, $(\alpha_{ij})=[A]_X$ implies

$$\det(A) = \sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\prod_{k=1}^{n}\alpha_{\sigma(k)k}.$$

Proof. We have

$$\det(A)x_1\wedge\cdots\wedge x_n = Ax_1\wedge\cdots\wedge Ax_n = \left(\sum_{k=1}^{n}\alpha_{k1}x_k\right)\wedge\cdots\wedge\left(\sum_{k=1}^{n}\alpha_{kn}x_k\right)$$

Expanding this, we obtain a sum, whose non-zero summands are of the form

$$\left(\prod_{k=1}^{n}\alpha_{\sigma(k)k}\right)x_{\wedge\sigma I},$$

where $I=(1,\ldots,n)$, so

$$x_{\wedge\sigma I} = x_{\sigma(1)}\wedge\cdots\wedge x_{\sigma(n)} = \operatorname{sgn}(\sigma)x_1\wedge\cdots\wedge x_n = \operatorname{sgn}(\sigma)x_{\wedge I}.$$

Conversely, for each $\sigma\in S_n$, there is a corresponding summand of this form, and we have

$$\det(A)x_{\wedge I} = \left(\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\prod_{k=1}^{n}\alpha_{\sigma(k)k}\right)x_{\wedge I}.$$

□

Of course, computing the determinant directly this way is wildly impractical: there are $n!$ summands, each of which is unpleasant to compute. Fortunately, we can use Gaussian elimination on account of the following results:

PROPOSITION 5.7. If $(\alpha_{ij})=[A]_X$ is upper triangular, then $\det(A)=\prod_{k=1}^{n}\alpha_{kk}$.

Proof. For any non-identity $\sigma \in S_n$, there is some $j \in [n]$ such that $\sigma(j) > j$. Indeed, if $\sigma(j) \leqslant j$ for all j, then, in particular, $\sigma(1) = 1$, so σ admits a restriction to $\{2, \dots, n\}$. Induction then shows $\sigma(j) = j$ for all j, so $\sigma = 1_n$. Consequently, if $(\alpha_{ij}) = [A]_X$ is upper triangular, then

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} \alpha_{\sigma(i)i} = \prod_{k=1}^{n} \alpha_{kk}.$$

□

COROLLARY 5.8. Elementary transformations have determinant 1.

PROPOSITION 5.9. If P is the permutation transformation corresponding to $\sigma \in S_n$, then $\det(P) = \operatorname{sgn}(\sigma)$.

Proof. We have

$$\det(P) e_1 \wedge \cdots \wedge e_n = e_{\sigma(1)} \wedge \cdots \wedge e_{\sigma(n)} = \operatorname{sgn}(\sigma) e_1 \wedge \cdots \wedge e_n.$$

□

Alternatively, we can use the following to compute determinants recursively:

PROPOSITION 5.10. Let $A : V \to V$ be a linear operator, let $X = \{x_1, \dots, x_n\}$ be a basis of V, and let $[A] = (\alpha_{ij}) = [A]_X$. For all i, j, let $[A]_{ij}$ denote the cofactor matrix obtained by removing the ith row and jth column of $[A]$. Then $[\operatorname{adj}(A)]_X = ((-1)^{i+j} \det [A]_{ji})$.

Proof. Recall the definition of the adjugate from **§3.5.**:

$$x \mapsto x \wedge (-) \mapsto A'(x \wedge (-)) = x \wedge A^{\wedge(n-1)}(-) \mapsto \operatorname{adj}(A)x.$$

We want to show that

$$\operatorname{adj}(A) x_k = \sum_{i=1}^{n} (-1)^{i+k} \det [A]_{ki} x_i$$

for each x_k. Pulling back to $\mathscr{L}(\bigwedge^{n-1} V, \bigwedge^{n} V)$, this is equivalent to the identity

$$x_k \wedge A^{\wedge(n-1)}(-) = \sum_{i=1}^{n} (-1)^{i+k} \det [A]_{ki} x_i \wedge (-). \tag{5.2}$$

By **Theorem 3.19.**, the set $\{\widehat{x_j} : 1 \leqslant j \leqslant n\}$ is a basis for $\bigwedge^{n-1} V$, where

$$\widehat{x_j} = x_1 \wedge \cdots \wedge x_{j-1} \wedge x_{j+1} \wedge \cdots \wedge x_n.$$

Therefore, we can prove (5.2) by showing that it holds when evaluated at each $\widehat{x_j}$.

Fix k and j. Starting with the right-hand side, we have

$$\begin{aligned}\sum_{i=1}^{n} (-1)^{i+k} \det [A]_{ki} (x_i \wedge \widehat{x_j}) &= (-1)^{j+k} \det [A]_{kj} (x_j \wedge \widehat{x_j}) \\ &= (-1)^{j+k} (-1)^{j-1} \det [A]_{kj} (x_1 \wedge \cdots \wedge x_n) \\ &= (-1)^{k-1} \det [A]_{kj} x_{\wedge I},\end{aligned}$$

where $I = (1, \ldots, n)$. As for the left-hand side,

$$x_k \wedge A^{\wedge(n-1)} \widehat{x_j} = x_k \wedge Ax_1 \wedge \cdots \wedge Ax_{j-1} \wedge Ax_{j+1} \wedge \cdots \wedge Ax_n.$$

Using the matrix $[A]_X$, this is equal to

$$x_k \wedge \left(\sum_{i=1}^{n} \alpha_{i1} x_i \right) \wedge \cdots \wedge \left(\sum_{i=1}^{n} \alpha_{i(j-1)} x_i \right) \wedge \left(\sum_{i=1}^{n} \alpha_{i(j+1)} x_i \right) \wedge \cdots \wedge \left(\sum_{i=1}^{n} \alpha_{in} x_i \right).$$

Expanding, we get a sum whose non-zero terms are of the form

$$\left(\prod_{i=1, i \neq j,k}^{n} \alpha_{\sigma(i)i} \right) x_{\wedge \sigma J},$$

where σ is any permutation of $[n]$ fixing k and

$$J = (k, 1, \ldots, k-1, k+1, \ldots, n).$$

Therefore,

$$x_k \wedge A^{\wedge(n-1)} \widehat{x_j} = \sum_{\sigma : k \mapsto k} \operatorname{sgn}(\sigma) \prod_{i=1, i \neq j,k}^{n} \alpha_{\sigma(i)i} x_{\wedge J}.$$

But this coefficient is just $\det[A]_{kj}$, so

$$x_k \wedge A^{\wedge(n-1)} \widehat{x_j} = \det[A]_{kj} x_{\wedge J} = (-1)^{k-1} \det[A]_{kj} x_{\wedge I},$$

with the last equality following from skew-symmetry. This establishes the result. □

THEOREM 5.11 (Laplace Expansion)**.** Let $(\alpha_{ij}) = [A]_X$. For each $1 \leqslant j \leqslant n$,

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} \alpha_{ij} \det [A]_{ij}.$$

Proof. By **Theorem 3.24.**, $\operatorname{adj}(A)A = \det(A)\mathbb{1}$. Let $(\gamma_{ij}) = [\operatorname{adj}(A)]_X [A]_X$, so

$$\gamma_{ij} = \sum_{k=1}^{n} (-1)^{i+k} \det [A]_{ki} \alpha_{kj}.$$

We also have $(\gamma_{ij}) = \det(A)[\mathbb{1}]_X$. Clearly, the matrix $[\mathbb{1}]_X$ has all 1s on its diagonal with all other components vanishing. Therefore, $\gamma_{jj} = \det(A)$ for all $1 \leqslant j \leqslant n$. Equating with the above, we have

$$\det(A) = \gamma_{jj} = \sum_{i=1}^{n} (-1)^{i+j} \alpha_{ij} \det[A]_{ij}.$$

□

BIBLIOGRAPHY

[1] Sheldon Axler. *Linear Algebra Done Right*. 4th edition. Springer Cham, 2023.

[2] Pierre Antoine Grillet. *Abstract Algebra*. 2nd edition. Springer New York, NY, 2007.

[3] Paul R. Halmos. *Finite-Dimensional Vector Spaces*. 1st edition. Springer New York, NY, 1974.

[4] Irving Kaplansky. "Reminiscences". In: *PAUL HALMOS Celebrating 50 Years of Mathematics*. Ed. by John H. Ewing and F.W. Gehring. 1st edition. Springer New York, NY, 1991.

[5] Serge Lang. *Algebra*. 3rd edition. Springer New York, NY, 2005.

[6] Steven Roman. *Advanced Linear Algebra*. 3rd edition. Springer New York, NY, 2007.

[7] Walter Rudin. *Principles of Mathematical Analysis*. 3rd edition. McGraw Hill, 1976.

[8] Jerry Shurman. *Calculus and Analysis in Euclidean Space*. 1st edition. Springer Cham, 2016.

[9] Jerry Shurman. *Multilinear Algebra: The Exterior Product*.

[10] Michael Spivak. *A Comprehensive Introduction to Differential Geometry, Volume 1*. 3rd edition. Publish or Perish, 1999.

[11] Michael Spivak. *Calculus on Manifolds*. Addison-Wesley Publishing Company, 1965.

[12] Sergei Treil. *Linear Algebra Done Wrong*. 2017.