MATH 416

Abstract Linear Algebra

Semester: Fall 2025

Professor: Theshani Nuradha

Contents

1 Fields and Vector Spaces

1.1 An introduction to abstract linear algebra

Linear algebra at the introductory level is usually presented through matrices and vectors in $\mathbb{R}^n$: equations are solved by row reduction, transformations are studied by writing down their matrices, and the geometry is essentially that of the plane and three-dimensional space. This works well, but it ties our intuition to a single class of objects. Abstract linear algebra strips the matrix scaffolding away and asks instead: what is the essential structure that makes a set behave like a space of vectors?

The answer is captured by two abstract objects. A *field* provides the scalars we use to scale vectors — it tells us what counts as a number. A *vector space* is then a set of objects that can be added together and scaled by elements of the field, in a way that obeys the same algebraic rules we are used to in $\mathbb{R}^n$. Once these definitions are in place, polynomials, functions, sequences, and geometric vectors all become instances of the same structure, and theorems we prove for one apply automatically to all of them.

1.2 Fields

Before we can talk about vectors we need to fix what the scalars are. Intuitively, scalars should be things we can add, subtract, multiply, and divide (except by zero), and these operations should behave the way they do for ordinary real numbers. The notion of a *field* formalises exactly this.

Definition 1.1 (Field)**.** A field $(\mathbb{F}, 0, 1, +, \cdot)$ is a set $\mathbb{F}$ containing two distinct distinguished elements 0 and 1, together with two binary operations called addition and multiplication,

$$+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}, \qquad \cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}, \tag{1.1}$$

which are closed in $\mathbb{F}$ and satisfy the following axioms.

1. Commutativity: for every $a, b \in \mathbb{F}$,

$$a + b = b + a, \qquad a \cdot b = b \cdot a. \tag{1.2}$$

2. Associativity: for every $a, b, c \in \mathbb{F}$,

$$a + (b + c) = (a + b) + c, \qquad a \cdot (b \cdot c) = (a \cdot b) \cdot c. \tag{1.3}$$

3. Additive and multiplicative identities: for every $a \in \mathbb{F}$,

$$a + 0 = a, \qquad a \cdot 1 = a. \tag{1.4}$$

4. Additive inverse: for every $a \in \mathbb{F}$ there exists $b \in \mathbb{F}$ such that

$$a + b = 0. \tag{1.5}$$

5. Multiplicative inverse: for every $a \in \mathbb{F}$ with $a \neq 0$ there exists a unique $b \in \mathbb{F}$ such that

$$a \cdot b = 1. \tag{1.6}$$

6. Distributivity: for every $a, b, c \in \mathbb{F}$,

$$a(b + c) = ab + ac. \tag{1.7}$$

Each axiom captures a single algebraic fact that we use without thinking when we manipulate real numbers. The closure conditions guarantee that adding or multiplying never takes us outside the field; commutativity and associativity say the order of operations doesn't matter; the identity and inverse axioms tell us we can always undo addition and (away from 0) undo multiplication; and distributivity is what links the two operations together.

The most familiar example of a field is the real numbers $\mathbb{R}$, with the usual addition and multiplication. The element $0 \in \mathbb{R}$ is just the number zero, and $1 \in \mathbb{R}$ is the number one. The additive inverse of a is $-a$, and the multiplicative inverse of $a \neq 0$ is $1/a$. The rational numbers $\mathbb{Q}$ also form a field. The integers $\mathbb{Z}$, however, do not: the multiplicative inverse axiom fails because $\frac{1}{2} \notin \mathbb{Z}$.

1.2.1 The complex numbers

The complex numbers form another fundamental example of a field. They are introduced to enlarge $\mathbb{R}$ so that every polynomial has a root — in particular, so that the equation $x^2 = -1$ has a solution. We define

$$\mathbb{C} = \{a + bi \,|\, a, b \in \mathbb{R}\}, \qquad i^2 = -1. \tag{1.8}$$

The distinguished elements of $\mathbb{C}$ are

$$0 = 0 + 0i, \qquad 1 = 1 + 0i. \tag{1.9}$$

Addition and multiplication on $\mathbb{C}$ are defined to extend the corresponding operations on $\mathbb{R}$. For $z_1 = a + bi$ and $z_2 = c + di$,

$$z_1 + z_2 = (a + c) + (b + d)i, \tag{1.10}$$

$$z_1 \cdot z_2 = (ac - bd) + (ad + bc)i. \tag{1.11}$$

Multiplication is what we get by treating i as a formal symbol, expanding $(a+bi)(c+di)$ via distributivity, and collecting terms using $i^2 = -1$.

The conjugate of $z = a + bi$ is the complex number obtained by reflecting z across the real axis,

$$\overline{z} = a - bi. \tag{1.12}$$

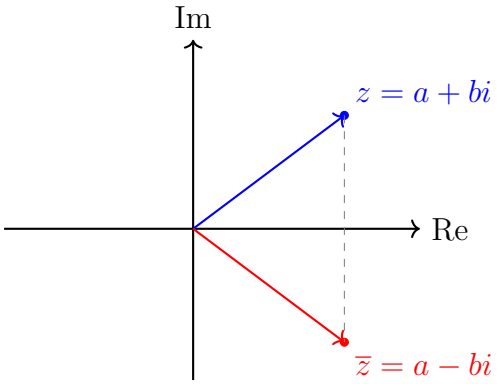

Figure 1: A complex number and its conjugate, viewed as points in the plane.

We now check that $\mathbb{C}$ really is a field. The axioms involving only addition or only multiplication of real and imaginary parts (commutativity, associativity, distributivity) follow from the corresponding facts in $\mathbb{R}$ by direct expansion, so we focus on the axioms that involve the structure of i in a non-trivial way: the existence of identities and inverses.

Proposition 1.1. $\mathbb{C}$ together with $0, 1, +, \cdot$ defined above forms a field.

Proof. Let $z = a + bi \in \mathbb{C}$. We address each non-trivial axiom in turn.

Additive identity. Adding $0 = 0 + 0i$ to z gives

$$z + 0 = (a + bi) + (0 + 0i) = (a + 0) + (b + 0)i = a + bi = z,$$

so 0 acts as an additive identity.

Multiplicative identity. Multiplying by $1 = 1+0i$ and using the multiplication formula gives

$$z \cdot 1 = (a + bi)(1 + 0i) = (a \cdot 1 - b \cdot 0) + (a \cdot 0 + b \cdot 1)i = a + bi = z.$$

Additive inverse. We seek $w = c + di$ such that $z + w = 0$. Comparing real and imaginary parts of

$$(a + c) + (b + d)i = 0 + 0i$$

yields $c = -a$ and $d = -b$, so $w = -a - bi$. This shows the additive inverse exists.

Multiplicative inverse. Suppose $z \neq 0$, equivalently a and b are not both zero. We seek $w \in \mathbb{C}$ with $zw = 1$. The standard trick for finding a multiplicative inverse in $\mathbb{C}$ is to

multiply numerator and denominator by the conjugate $\overline{z}$, which produces a real number in the denominator:

$$\frac{1}{a+bi} = \frac{1}{a+bi} \cdot \frac{a-bi}{a-bi} = \frac{a-bi}{a^2+b^2}.$$

Since $z \neq 0$ we have $a^2 + b^2 > 0$, so this is well-defined, and the inverse is

$$w = \frac{a}{a^2+b^2} + \frac{-b}{a^2+b^2} i.$$

A direct check confirms $zw = 1$.

Since every axiom is satisfied, $\mathbb{C}$ is a field. ■

From now on we use the symbol $\mathbb{F}$ to denote either $\mathbb{R}$ or $\mathbb{C}$ (or, occasionally, an arbitrary field). The whole development of linear algebra below is largely independent of which choice we make, although a few results — notably the existence of eigenvalues — depend on the field.

1.2.2 Scalars and exponents

For a field $\mathbb{F}$, any element of $\mathbb{F}$ is referred to as a scalar. For $a \in \mathbb{F}$ and a positive integer m we define repeated multiplication

$$a^m = \underbrace{a \cdot a \cdots a}_{m \text{ times}}. \tag{1.13}$$

By closure of multiplication, $a^m \in \mathbb{F}$. The familiar identities

$$(a^m)^n = a^{mn}, \tag{1.14}$$

$$(ab)^m = a^m b^m \qquad \forall a, b \in \mathbb{F} \tag{1.15}$$

follow by repeatedly applying associativity and commutativity.

1.3 Lists

Before defining the standard n-dimensional spaces, we need a small piece of bookkeeping: an ordered n-tuple of elements. We call this a list.

Definition 1.2 (List)**.** For a non-negative integer n, a list of length n is an ordered collection of n elements (which may be numbers, other lists, functions, or any objects whatsoever).

Two lists are equal if and only if they have the same length and the same entries in the same order. A list of length n is denoted

$$(x_1, x_2, \ldots, x_n), \tag{1.16}$$

where x_i is the ith entry. Lists must have finite length. The empty list is the unique list of length 0, written ().

Lists differ from sets in two important ways. First, the order of the entries matters: $(1, 2)$ and $(2, 1)$ are different lists but the same set. Second, repetitions matter: $(1, 1)$ has length two and is different from (1), although as sets they are both $\{1\}$.

1.4 The space $\mathbb{F}^n$

Definition 1.3 ($\mathbb{F}^n$)**.** For a field $\mathbb{F}$ and a positive integer n we define

$$\mathbb{F}^n = \{(x_1, \ldots, x_n) \,|\, x_i \in \mathbb{F} \text{ for } 1 \leq i \leq n\}\,. \tag{1.17}$$

$\mathbb{F}^n$ is the set of all lists of length n whose entries lie in $\mathbb{F}$. With $\mathbb{F} = \mathbb{R}$ we recover the geometric spaces familiar from calculus,

$$\mathbb{R}^2 = \{(x, y) \,|\, x, y \in \mathbb{R}\}\,, \tag{1.18}$$

$$\mathbb{R}^3 = \{(x, y, z) \,|\, x, y, z \in \mathbb{R}\}\,, \tag{1.19}$$

which we visualise as the plane and three-dimensional space respectively. Equivalently, $\mathbb{F}^n$ is the Cartesian product of $\mathbb{F}$ with itself n times,

$$\mathbb{F}^n = \underbrace{\mathbb{F} \times \mathbb{F} \times \cdots \times \mathbb{F}}_{n \text{ times}}. \tag{1.20}$$

1.4.1 Addition and the zero element in $\mathbb{F}^n$

Addition in $\mathbb{F}^n$ is defined componentwise: we add two lists by adding their entries one slot at a time. Explicitly,

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n). \tag{1.21}$$

Each component sum $x_i + y_i$ is the addition operation of $\mathbb{F}$. We pick out a distinguished zero element of $\mathbb{F}^n$, namely the list whose every entry is $0 \in \mathbb{F}$,

$$0 = \underbrace{(0, \ldots, 0)}_{n \text{ entries}}. \tag{1.22}$$

Notice this is a slight notational overload — the symbol "0" now refers to a list, not the scalar zero, but context will always make clear which is meant.

Remark. The set $\mathbb{F}^n$ is itself a field only when $n = 1$. For $n \geq 2$, the multiplicative inverse axiom fails: the element $(1, 0, \ldots, 0)$ multiplied componentwise by any other list yields a list whose second entry is zero, so it cannot equal $(1, 1, \ldots, 1)$. Instead $\mathbb{F}^n$ will be the prototypical example of the structure we develop next: a vector space.

1.5 Vector spaces

We now arrive at the central definition of the course. A vector space is, roughly, a set on which we can do addition and scalar multiplication, and on which these operations behave the way they do in $\mathbb{F}^n$. The point of stating the axioms abstractly is that many objects which look superficially very different — ordered tuples, polynomials, matrices, continuous functions — all satisfy the same rules, so any theorem we prove about an

abstract vector space applies simultaneously to all of them.

Definition 1.4 (Vector space)**.** A vector space V over a field $\mathbb{F}$ is a set V equipped with two operations: addition $+ : V \times V \to V$ and scalar multiplication $\cdot : \mathbb{F} \times V \to V$, satisfying the following axioms.

1. Commutativity of addition: $u + v = v + u$ for all $u, v \in V$.
2. Associativity: for all $u, v, w \in V$ and $a, b \in \mathbb{F}$,

$$(u + v) + w = u + (v + w), \qquad (ab) \cdot v = a \cdot (bv). \tag{1.23}$$

3. Additive identity: there exists $0 \in V$ with $v + 0 = v$ for all $v \in V$.
4. Additive inverse: for every $u \in V$ there exists $w \in V$ with $u + w = 0$.
5. Multiplicative identity: $1 \cdot v = v$ for every $v \in V$, where $1 \in \mathbb{F}$.
6. Distributivity: for all $u, v \in V$ and $a, b \in \mathbb{F}$,

$$a(u + v) = au + av, \qquad (a + b)v = av + bv. \tag{1.24}$$

Elements of a vector space are called vectors. The simplest vector space is the zero space $\{0\}$, which has only the zero vector. The empty set $\emptyset$ cannot be a vector space, because the additive identity axiom requires at least one element. A vector space defined over $\mathbb{F} = \mathbb{R}$ is called a real vector space; one defined over $\mathbb{F} = \mathbb{C}$ is a complex vector space.

Notice that the addition symbol on the left and right of the commutativity axiom both refer to addition in V, even though scalar multiplication uses the multiplication of $\mathbb{F}$. Similarly, the 0 in the additive identity axiom is the zero vector of V, not the scalar zero. Keeping track of where each operation lives is a habit worth cultivating early.

1.5.1 $\mathbb{F}^n$ as a vector space

Scalar multiplication in $\mathbb{F}^n$ is defined componentwise, just as addition was: for $\lambda \in \mathbb{F}$,

$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n). \tag{1.25}$$

Together with the componentwise addition defined above, $\mathbb{F}^n$ satisfies all six vector space axioms. The verification is mechanical: every axiom reduces to the corresponding axiom of $\mathbb{F}$ applied to each slot of the list. In particular $\mathbb{R}^n$ and $\mathbb{C}^n$ are vector spaces, and for $n \in \{1, 2, 3\}$ we recover the geometric arrows-and-points intuition that motivates the term "vector".

1.5.2 Polynomials as a vector space

Definition 1.5 ($\mathcal{P}_d(\mathbb{F})$)**.** $\mathcal{P}_d(\mathbb{F})$ denotes the set of polynomials of degree at most d with coefficients from $\mathbb{F}$:

$$\mathcal{P}_d(\mathbb{F}) = \left\{ a_0 + a_1 x + \cdots + a_d x^d \;\middle|\; a_i \in \mathbb{F} \text{ for } 0 \le i \le d \right\}. \tag{1.26}$$

We define addition of polynomials by adding coefficients of like powers,

$$(a_0 + \cdots + a_d x^d) + (b_0 + \cdots + b_d x^d) = (a_0 + b_0) + \cdots + (a_d + b_d) x^d, \tag{1.27}$$

and scalar multiplication by scaling each coefficient,

$$\lambda(a_0 + \cdots + a_d x^d) = (\lambda a_0) + \cdots + (\lambda a_d) x^d. \tag{1.28}$$

The zero vector is the polynomial whose every coefficient is zero, and the additive inverse of a polynomial is obtained by negating each coefficient. With these operations $\mathcal{P}_d(\mathbb{F})$ satisfies all six vector space axioms.

This is a useful example because polynomials feel quite different from tuples in $\mathbb{F}^n$ —

they are functions of a variable x — yet they have exactly the same algebraic structure. In particular the notion of "length" or "magnitude" has no analogue here. This is one reason we define vectors abstractly rather than as objects with magnitude and direction: the latter description is geometric and does not survive the move to function spaces.

1.5.3 Function spaces

We can generalise polynomial spaces further. For any non-empty set S and any field $\mathbb{F}$, the collection of functions from S to $\mathbb{F}$ also forms a vector space.

Definition 1.6 ($\mathbb{F}^S$)**.** For a non-empty set S and a field $\mathbb{F}$,

$$\mathbb{F}^S = \{f : S \to \mathbb{F} \mid f \text{ is a function}\}. \tag{1.29}$$

Addition and scalar multiplication are defined pointwise:

$$(f+g)(x) = f(x) + g(x) \qquad \forall x \in S, \tag{1.30}$$

$$(\lambda f)(x) = \lambda f(x) \qquad \forall x \in S,\ \lambda \in \mathbb{F}. \tag{1.31}$$

The zero vector is the constant function 0, and the additive inverse of f is the function $-f$ defined by $(-f)(x) = -f(x)$. With these operations $\mathbb{F}^S$ is a vector space. The notation $\mathbb{F}^S$ is suggestive: it generalises the case $S = \{1, 2, \ldots, n\}$, where a function is precisely a list of length n, recovering $\mathbb{F}^n$.

1.5.4 Properties of vector spaces

From the axioms alone, several useful facts follow. We prove them once and for all so that we can use them in any vector space without rederivation.

Proposition 1.2 (Uniqueness of the additive identity)**.** Every vector space has a unique additive identity.

Proof. Let V be a vector space, and suppose both 0 and $0'$ are elements of V that satisfy the additive identity axiom. We will show $0 = 0'$.

Because 0 is an additive identity, adding it to any vector leaves that vector unchanged. Applying this with $v = 0'$ gives

$$0' + 0 = 0'.$$

On the other hand, because $0'$ is also an additive identity, applying its defining property with $v = 0$ gives

$$0 + 0' = 0.$$

Combining these with the commutativity of addition, we have

$$0' = 0' + 0 = 0 + 0' = 0,$$

so the two supposedly distinct identities are equal. ■

Proposition 1.3 (Uniqueness of additive inverses)**.** Every element in a vector space has a unique additive inverse, which we denote $-v$.

Proof. Let V be a vector space and let $v \in V$. Suppose u and w are both additive inverses of v, meaning $v + u = 0$ and $v + w = 0$. We aim to show $u = w$.

The strategy is to start from w and rewrite it using the additive identity, then insert $v + u$ in place of zero. We have

$$w = w + 0 = w + (v + u) = (w + v) + u = 0 + u = u,$$

where the second equality uses $v + u = 0$, the third uses associativity of addition, the fourth uses $w + v = 0$, and the last uses the additive identity property.

Thus the two inverses must be equal. Once uniqueness is established, we are entitled to give the inverse a name; we write the additive inverse of v as $-v$, and define subtraction by $u - v = u + (-v)$. ■

Proposition 1.4. For every vector space V over $\mathbb{F}$ and every $v \in V$, $0 \cdot v = 0$, where the 0 on the left is the additive identity of $\mathbb{F}$ and the 0 on the right is the zero vector of V.

Proof. Let $v \in V$. Using the fact that $0 = 0 + 0$ in $\mathbb{F}$, together with the distributive axiom of the vector space, we obtain

$$0 \cdot v = (0 + 0) \cdot v = 0 \cdot v + 0 \cdot v.$$

We now have an equation that says $0 \cdot v$ equals itself plus itself. Adding the additive inverse $-(0 \cdot v)$ of $0 \cdot v$ to both sides eliminates one copy from each side and leaves

$$0 = 0 \cdot v,$$

as claimed. ■

Proposition 1.5. For every vector space V over $\mathbb{F}$ and every $a \in \mathbb{F}$, we have $a \cdot 0 = 0$, where both zeros are the zero vector of V.

Proof. The argument runs in parallel to the previous proposition, but now exploiting that the zero vector is unchanged by addition with itself. We start from $0 = 0 + 0$ in V, then apply the distributive axiom in the form $a(u + v) = au + av$ with $u = v = 0$:

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

Adding $-(a \cdot 0)$ to both sides cancels one occurrence from each side and yields $0 = a \cdot 0$. ■

Proposition 1.6. For every v in a vector space V over $\mathbb{F}$, $(-1) \cdot v = -v$.

Proof. Recall that $-v$ is by definition the unique additive inverse of v. So if we can show that $(-1) \cdot v$ also satisfies the additive inverse property of v, uniqueness will force $(-1) \cdot v = -v$.

We compute the sum $v + (-1) \cdot v$. Using the multiplicative identity axiom $1 \cdot v = v$, then the distributive axiom in the form $(a + b)v = av + bv$, and finally the previous proposition,

$$v + (-1) \cdot v = 1 \cdot v + (-1) \cdot v = (1 + (-1)) \cdot v = 0 \cdot v = 0.$$

Therefore $(-1) \cdot v$ is an additive inverse of v, and by uniqueness it must equal $-v$. ■

1.6 Subspaces

Many of the most important sets we encounter in linear algebra are vector spaces sitting inside larger vector spaces. The set of polynomials with no constant term, the set of vectors in $\mathbb{R}^3$ orthogonal to a fixed vector, the kernel of a linear map — all of these are subsets of a larger vector space that happen to be closed under the same addition and scalar multiplication, and therefore form vector spaces in their own right. We give them a name.

Definition 1.7 (Subspace)**.** A subset U of a vector space V is a subspace of V, written $U \leq V$, if U is itself a vector space under the addition and scalar multiplication inherited from V.

Verifying that a subset is a subspace by directly checking all six axioms would be tedious. Fortunately most of the axioms (commutativity, associativity, distributivity, the multiplicative identity property) are inherited automatically from the parent space V — they hold for all elements of V, so in particular they hold for all elements of U. The only conditions we genuinely need to verify are that U contains the zero vector and is closed under the two operations.

Proposition 1.7 (Subspace test)**.** A subset U of a vector space V is a subspace of V if and only if the following three conditions hold.

1. Additive identity: $0 \in U$.

2. Closure under addition: for every $u, w \in U$, $u + w \in U$.

3. Closure under scalar multiplication: for every $\lambda \in \mathbb{F}$ and $u \in U$, $\lambda u \in U$.

Proof. We prove both directions.

Forward direction. Suppose U is a subspace, so U is itself a vector space. Then U has an additive identity, which by uniqueness must coincide with the additive identity of V, so $0 \in U$. Closure under both operations is required by the very definition of a vector space, so conditions (2) and (3) are satisfied.

Reverse direction. Conversely, suppose $U \subseteq V$ satisfies (1), (2), and (3). We need to verify all six vector space axioms for U. Commutativity, associativity, distributivity, and the multiplicative identity axiom hold for every element of V, and elements of U are in particular elements of V, so these axioms are inherited automatically. The additive identity is in U by (1), and (2) and (3) ensure closure. The only remaining axiom is the existence of additive inverses: for $u \in U$, we must show $-u \in U$. By the previous section, $-u = (-1) \cdot u$, and since $-1 \in \mathbb{F}$, condition (3) gives $(-1) \cdot u \in U$. Thus U is a vector space, and so a subspace of V. ■

Example 0

Fix $b \in \mathbb{F}$. Determine for which values of b the set

$$U = \left\{(x_1, x_2, x_3, x_4) \in \mathbb{F}^4 \,\middle|\, x_3 = 5x_4 + b\right\}$$

is a subspace of $\mathbb{F}^4$.

Solution

We claim U is a subspace if and only if $b = 0$.

Suppose first that $b = 0$. We check all three conditions of the subspace test.

- The zero vector is $(0, 0, 0, 0)$. Its third entry is 0, and $5 \cdot 0 + 0 = 0$, so the zero vector lies in U.
- For closure under addition, take $u, w \in U$ with $u_3 = 5u_4$ and $w_3 = 5w_4$. Then

$$(u + w)_3 = u_3 + w_3 = 5u_4 + 5w_4 = 5(u_4 + w_4) = 5(u + w)_4,$$

  so $u + w \in U$.
- For closure under scalar multiplication, $(\lambda u)_3 = \lambda u_3 = \lambda \cdot 5u_4 = 5(\lambda u_4) = 5(\lambda u)_4$, so $\lambda u \in U$.

All three conditions hold, so U is a subspace.

Suppose now that $b \neq 0$. Then the zero vector fails to lie in U since $0 \neq 5 \cdot 0 + b$.

By condition (1) of the subspace test, U is not a subspace.

Geometrically, the case $b = 0$ describes a hyperplane through the origin in $\mathbb{F}^4$, while $b \neq 0$ describes a hyperplane that misses the origin. Hyperplanes through the origin are subspaces; affine hyperplanes parallel to them are not.

1.7 Sums and direct sums of subspaces

Given two subspaces U_1 and U_2 of V, their union $U_1 \cup U_2$ is generally not a subspace — if we take a vector from each, their sum need not lie in either. The right way to combine subspaces is to take the smallest subspace containing both, which is the set of all vectors that can be written as $u_1 + u_2$ with $u_i \in U_i$.

Definition 1.8 (Sum of subspaces)**.** Let V be a vector space and let $U_1, \ldots, U_m$ be subspaces of V. Their sum is

$$\sum_{i=1}^{m} U_i = U_1 + \cdots + U_m = \{u_1 + \cdots + u_m \,|\, u_i \in U_i \text{ for } 1 \leq i \leq m\}. \tag{1.32}$$

Each summand u_i is taken from its own subspace, and the addition uses the addition of V (which makes sense because all the U_i sit inside the common space V).

Example 1

Let $U = \{(x, 0, 0) \,|\, x \in \mathbb{F}\}$ and $W = \{(0, y, 0) \,|\, y \in \mathbb{F}\}$ be subspaces of $\mathbb{F}^3$. Their sum is

$$U + W = \{(x, y, 0) \,|\, x, y \in \mathbb{F}\}.$$

When $\mathbb{F} = \mathbb{R}$, this is exactly the xy-plane in $\mathbb{R}^3$.

Theorem 1.1. If $U_1, \ldots, U_m$ are subspaces of V, then $U_1 + \cdots + U_m$ is the smallest subspace of V containing each of $U_1, \ldots, U_m$.

Proof. Let $S = U_1 + \cdots + U_m$. There are two things to show: that S is a subspace, and that any subspace containing each U_i must contain S.

S is a subspace. We verify the three conditions of the subspace test. First, taking $u_i = 0 \in U_i$ for each i gives $0 + \cdots + 0 = 0 \in S$. Next, suppose $x = \sum u_i$ and $y = \sum v_i$ with $u_i, v_i \in U_i$. Then

$$x + y = \sum_i (u_i + v_i),$$

and each $u_i + v_i$ lies in U_i because each U_i is closed under addition. Therefore $x + y \in S$. Similarly, for $\lambda \in \mathbb{F}$,

$$\lambda x = \sum_i \lambda u_i \in S$$

because each U_i is closed under scalar multiplication.

S contains each U_i. For any $u \in U_j$, we can write $u = 0 + \cdots + u + \cdots + 0$, putting u in the jth slot and zeros elsewhere; this is an element of S.

S is the smallest such subspace. Suppose T is any subspace of V containing each U_i. Then for any $x = u_1 + \cdots + u_m \in S$, every u_i lies in $U_i \subseteq T$, and T is closed under addition, so $x \in T$. Hence $S \subseteq T$.

We have shown S is a subspace containing each U_i, and any subspace containing each U_i contains S. So S is the smallest subspace containing the U_i. ■

1.7.1 Direct sums

When we write a vector $v \in U_1 + \cdots + U_m$ as a sum $u_1 + \cdots + u_m$, this decomposition need not be unique — there may be many ways to split the same vector across the summands. Decompositions that are unique deserve a special name and play a central role throughout the rest of the course.

Definition 1.9 (Direct sum)**.** Let $U_1, \ldots, U_m$ be subspaces of V. The sum $U_1 + \cdots + U_m$ is called a direct sum, written $U_1 \oplus \cdots \oplus U_m$, if every element of $U_1 + \cdots + U_m$ can be written in exactly one way as $u_1 + \cdots + u_m$ with $u_i \in U_i$.

Direct sums are easier to work with because each component of a vector is uniquely determined; this lets us define operations slot-by-slot. The next theorem gives a more practical criterion for checking whether a sum is direct: rather than verifying uniqueness of every decomposition, we only need to check uniqueness of the decomposition of 0.

Theorem 1.2 (Direct sum criterion)**.** Let $U_1, \ldots, U_m$ be subspaces of V. The sum $U_1 + \cdots + U_m$ is direct if and only if the only way to write 0 as a sum $u_1 + \cdots + u_m$ with $u_i \in U_i$ is the trivial one with $u_1 = \cdots = u_m = 0$.

Proof. We prove the two directions separately.

Forward direction. Suppose the sum is direct. Then every element has a unique decomposition; in particular 0 does. The decomposition $0 = 0 + 0 + \cdots + 0$ with $u_i = 0$ is one such, and uniqueness rules out any other.

Reverse direction. Suppose the trivial representation of 0 is the only one. Take any $v \in U_1 + \cdots + U_m$ with two decompositions

$$v = u_1 + \cdots + u_m = v_1 + \cdots + v_m,$$

where $u_i, v_i \in U_i$. We want to conclude $u_i = v_i$ for each i. Subtracting the two decompositions gives

$$0 = (u_1 - v_1) + (u_2 - v_2) + \cdots + (u_m - v_m).$$

For each i, the difference $u_i - v_i$ lies in U_i because U_i is closed under addition and additive inverses. So we have written 0 as a sum with one term in each U_i. By hypothesis, every such sum is the trivial one, and therefore $u_i - v_i = 0$, that is, $u_i = v_i$. The decomposition of v is unique, so the sum is direct. ■

For two subspaces, the criterion takes a particularly clean form: their sum is direct if and only if they intersect trivially.

Theorem 1.3 (Two-subspace criterion)**.** Let U and W be subspaces of V. Then $U + W$ is a direct sum if and only if $U \cap W = \{0\}$.

Proof. Forward direction. Suppose $U + W$ is a direct sum. Take any $v \in U \cap W$. Then we can write $0 = v + (-v)$ with $v \in U$ and $-v \in W$ (since W is closed under negation, $-v \in W$ because $v \in W$). By the direct sum criterion, both summands must vanish, so $v = 0$. Hence $U \cap W \subseteq \{0\}$, and the reverse inclusion is automatic, giving $U \cap W = \{0\}$.

Reverse direction. Suppose $U \cap W = \{0\}$, and consider any decomposition of zero, $0 = u + w$ with $u \in U$ and $w \in W$. Rearranging gives $u = -w$. The right side $-w$ lies in W (closure under negation), and the left side u lies in U, so $u \in U \cap W = \{0\}$. Therefore $u = 0$, and consequently $w = -u = 0$ as well. The only decomposition of 0 is the trivial one, so by the direct sum criterion, $U + W$ is a direct sum. ■

Remark. The two-subspace criterion does *not* extend to three or more subspaces. Pairwise trivial intersections $U_i \cap U_j = \{0\}$ for $i \neq j$ are not enough to make the sum direct. For an explicit example in $\mathbb{R}^2$, take three distinct lines through the origin: any two intersect only at the origin, but their sum is not direct because any vector in $\mathbb{R}^2$ can be written as a sum of vectors on the three lines in many different ways. The general criterion (uniqueness of the decomposition of 0) is the right one.

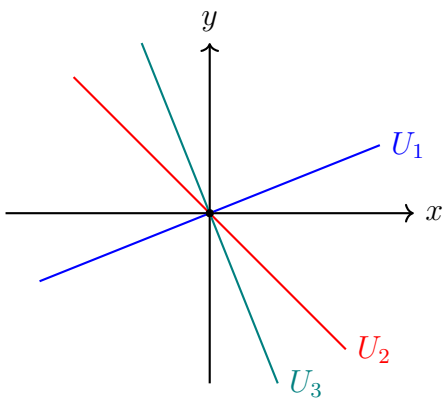

Figure 2: Three distinct lines in $\mathbb{R}^2$ that pairwise intersect only at the origin but whose sum is not direct.

2 Finite-Dimensional Vector Spaces

The previous chapter introduced vector spaces in full generality. Such generality is powerful, but to do anything concrete we need a way to describe vectors in terms of a fixed list of building blocks. Two notions make this possible: *spanning* says the building blocks are enough to reach every vector, while *linear independence* says the building blocks carry no redundancy. A list with both properties is a basis, and the length of any basis is an intrinsic property of the space called its dimension. Together these notions are the central organising tools of finite-dimensional linear algebra.

2.1 Linear combinations and span

Definition 2.1 (Linear combination)**.** Let V be a vector space over a field $\mathbb{F}$, and let $(v_1, \dots, v_m)$ be a list of vectors in V. A linear combination of this list is any vector $w \in V$ of the form

$$w = a_1 v_1 + a_2 v_2 + \cdots + a_m v_m = \sum_{i=1}^{m} a_i v_i, \qquad a_i \in \mathbb{F}. \tag{2.1}$$

Closure of V under addition and scalar multiplication guarantees that any linear combination of vectors in V is itself a vector in V. Linear combinations are the only operations we have at our disposal, so understanding which vectors can be reached as linear combinations of a given list is fundamental to everything that follows.

Example 2

Let $v_1 = (1, 0)$, $v_2 = (0, 1)$, and $v_3 = (3, 4)$ in $\mathbb{R}^2$. Then v_3 is a linear combination of v_1 and v_2, since

$$v_3 = 3v_1 + 4v_2.$$

Collecting all linear combinations of a list yields a subset of V called the span of the list.

Definition 2.2 (Span)**.** The span of a list $(v_1, \ldots, v_m)$ is the set of all linear combinations of its elements,

$$\operatorname{span}(v_1, \ldots, v_m) = \{a_1 v_1 + \cdots + a_m v_m \,|\, a_i \in \mathbb{F} \text{ for } 1 \leq i \leq m\}. \tag{2.2}$$

By convention, the span of the empty list is $\{0\}$.

If the span of a list equals all of V, we say the list spans V, and we call the list a spanning list of V. The convention $\operatorname{span}() = \{0\}$ makes the empty list "span" the zero space and is consistent with treating an empty sum as zero.

Proposition 2.1. The span of a list $(v_1, \ldots, v_m)$ in V is the smallest subspace of V containing each v_i.

Proof. Let $S = \operatorname{span}(v_1, \ldots, v_m)$. We must check three things: S is a subspace, S contains each v_i, and S is contained in any other subspace with the same property.

S is a subspace. We apply the subspace test. Setting all coefficients to zero gives the zero vector, so $0 \in S$. If $x = \sum a_i v_i$ and $y = \sum b_i v_i$ lie in S, then

$$x + y = \sum_i (a_i + b_i) v_i \in S, \qquad \lambda x = \sum_i (\lambda a_i) v_i \in S,$$

so S is closed under addition and scalar multiplication.

S contains each v_j. Setting the jth coefficient to 1 and the rest to 0 gives $v_j \in S$.

S is the smallest such subspace. Let U be any subspace of V containing each v_i. Any linear combination $\sum a_i v_i$ is formed from elements of U by scalar multiplication and addition, and since U is closed under both operations, the combination lies in U. Therefore $S \subseteq U$. ∎

Once we can talk about spans, we can ask whether finitely many vectors are enough to generate a whole space. If so, we say the space is finite-dimensional.

Definition 2.3 (Finite-dimensional)**.** A vector space V is finite-dimensional if there exists a finite list of vectors that spans V. Otherwise V is infinite-dimensional.

Example 3

$\mathbb{F}^n$ is finite-dimensional. It is spanned by the so-called standard basis $(e_1, \ldots, e_n)$, where e_i is the list with 1 in slot i and 0 in every other slot. The space $\mathcal{P}_d(\mathbb{F})$ of polynomials of degree at most d is finite-dimensional, spanned by $(1, x, x^2, \ldots, x^d)$. By contrast, the space $\mathcal{P}(\mathbb{F})$ of *all* polynomials with coefficients in $\mathbb{F}$ is not finite-dimensional: any finite list of polynomials has a maximum degree d, and a polynomial of degree larger than d cannot be written as a linear combination of polynomials of degree at most d.

2.2 Linear independence

Spanning captures one desirable property of a list: we want enough vectors to reach everywhere. Linear independence captures the complementary property: we want no redundancy. A list is redundant if one of its vectors can be expressed as a linear combination of the others; removing such a vector does not shrink the span.

Definition 2.4 (Linearly independent)**.** A list $(v_1, \ldots, v_m)$ of vectors in V is linearly independent if the only choice of scalars $a_1, \ldots, a_m \in \mathbb{F}$ for which

$$a_1 v_1 + \cdots + a_m v_m = 0 \tag{2.3}$$

holds is $a_1 = \cdots = a_m = 0$. A list that fails this condition is linearly dependent. The empty list is, by convention, linearly independent.

Linear independence admits an equivalent formulation that is often more useful. A list $(v_1, \ldots, v_m)$ is linearly independent if and only if every vector in $\operatorname{span}(v_1, \ldots, v_m)$ can be written as a linear combination $\sum a_i v_i$ in exactly one way. The two formulations agree: if two distinct combinations $\sum a_i v_i$ and $\sum b_i v_i$ gave the same vector, their difference $\sum (a_i - b_i) v_i$ would be a nontrivial expression for zero, contradicting independence;

conversely, a nontrivial expression for zero immediately produces two different decompositions of the zero vector.

The next lemma is the workhorse of finite-dimensional linear algebra. It says that any linearly dependent list with a nonzero leading vector contains a later vector already in the span of the earlier ones, so that vector can be removed without changing the span.

Lemma 2.1 (Linear dependence lemma)**.** Let $(v_1, \dots, v_m)$ be a linearly dependent list in V with $v_1 \neq 0$. Then there exists an index $j \in \{2, \dots, m\}$ such that:

1. $v_j \in \operatorname{span}(v_1, \dots, v_{j-1})$, and
2. removing v_j from the list leaves the span unchanged.

Proof. We handle the two parts in order.

Part (1). Since the list is linearly dependent, there exist scalars $a_1, \dots, a_m$ not all zero, with

$$a_1 v_1 + a_2 v_2 + \cdots + a_m v_m = 0.$$

Let j be the largest index with $a_j \neq 0$. We claim $j \geq 2$. Indeed, if $j = 1$, then $a_2 = \cdots = a_m = 0$ and the dependence becomes $a_1 v_1 = 0$ with $a_1 \neq 0$, forcing $v_1 = 0$. This contradicts our assumption, so $j \geq 2$.

By the choice of j, all coefficients beyond index j vanish, and the dependence becomes

$$a_1 v_1 + \cdots + a_j v_j = 0.$$

Since $a_j \neq 0$ we may divide and solve for v_j,

$$v_j = -\frac{a_1}{a_j} v_1 - \frac{a_2}{a_j} v_2 - \cdots - \frac{a_{j-1}}{a_j} v_{j-1},$$

which exhibits v_j as a linear combination of $v_1, \dots, v_{j-1}$, i.e. $v_j \in \operatorname{span}(v_1, \dots, v_{j-1})$.

Part (2). Let $v \in \operatorname{span}(v_1, \dots, v_m)$, so $v = \sum_{i=1}^{m} c_i v_i$ for some scalars c_i. Substituting

the expression for v_j from part (1) into this combination,

$$v = \sum_{i \neq j} c_i v_i + c_j \left(-\frac{a_1}{a_j} v_1 - \cdots - \frac{a_{j-1}}{a_j} v_{j-1} \right),$$

which is a linear combination of the list with v_j removed. So every vector in the original span already lies in the span of the reduced list, and the reverse inclusion is automatic, giving equality. ∎

Iterating the linear dependence lemma produces a fundamental quantitative comparison between spanning lists and linearly independent lists. The content of the theorem is intuitive: a spanning list must be at least as long as any linearly independent list, because each independent vector "uses up" one slot in the spanning list.

Theorem 2.2 (Length of linearly independent list $\leq$ length of spanning list)**.** In a finite-dimensional vector space V, every linearly independent list has length less than or equal to the length of every spanning list.

Proof. Suppose $(u_1, \ldots, u_m)$ is linearly independent in V and $(w_1, \ldots, w_n)$ spans V. We will produce a sequence of spanning lists of length n that incorporate $u_1, \ldots, u_m$ one at a time, deducing $m \leq n$ from the fact that we never run out of w's to remove.

Step 1. Since $(w_1, \ldots, w_n)$ spans V, the vector u_1 is a linear combination of the w_i. Prepending u_1 gives

$$(u_1, w_1, \ldots, w_n),$$

a list of length $n + 1$ which still spans V (we have only added a vector) and is linearly dependent (because u_1 is in the span of the rest). The first vector u_1 is nonzero since it lies in a linearly independent list, so the linear dependence lemma applies. The offending vector cannot be u_1 itself (it is the very first entry), so it must be one of the w_i. Remove it to obtain a list B_1 of length n that still spans V and begins with u_1.

Step k. Suppose after $k-1$ steps we have produced a list B_{k-1} of length n that spans V and begins with $u_1, \ldots, u_{k-1}$. Insert u_k at position k. The resulting list has length $n+1$, still spans V, and is linearly dependent (since B_{k-1} already spanned V, the inserted

u_k must be a combination of the other entries). By the linear dependence lemma, some vector in this list lies in the span of the preceding ones. This vector cannot be any of $u_1, \ldots, u_k$, because the u's are linearly independent and so $u_k \notin \operatorname{span}(u_1, \ldots, u_{k-1})$ and similarly for the earlier u's. Hence the offender is one of the surviving w's, which we remove to produce B_k of length n.

Each step adds one u and removes one w, preserving the length at n. The procedure can only continue as long as there are w's left to remove. After m steps we have inserted all of $u_1, \ldots, u_m$, which required removing m of the original w's; this is possible only if $n \geq m$. ■

2.3 Bases

Spanning and linear independence describe two complementary properties of a list. A list with both properties is a most efficient description of V: enough vectors to reach everywhere, but no slack. Such a list is a basis.

Definition 2.5 (Basis)**.** A basis of V is a list of vectors in V that is both linearly independent and spans V.

Example 4

The list $(e_1, \ldots, e_n)$ is the standard basis of $\mathbb{F}^n$. The list $(1, x, x^2, \ldots, x^d)$ is the standard basis of $\mathcal{P}_d(\mathbb{F})$.

Combining the two equivalent formulations of linear independence and spanning, we obtain a single clean characterisation of a basis.

Proposition 2.2 (Bases give unique linear combinations)**.** A list $(v_1, \ldots, v_n)$ is a basis of V if and only if every vector $v \in V$ can be written in exactly one way as

$$v = a_1 v_1 + \cdots + a_n v_n, \qquad a_i \in \mathbb{F}. \tag{2.4}$$

Proof. Forward direction. Suppose $(v_1, \ldots, v_n)$ is a basis. Because the list spans V, every $v \in V$ has at least one such expression. For uniqueness, suppose $\sum a_i v_i = \sum b_i v_i$.

Subtracting gives $\sum(a_i - b_i)v_i = 0$, and linear independence forces $a_i - b_i = 0$ for every i, i.e. $a_i = b_i$.

Reverse direction. Suppose every $v \in V$ has a unique representation as a linear combination of the list. Existence of a representation for every v means the list spans V. To establish linear independence, consider $\sum a_i v_i = 0$. The zero vector certainly admits the trivial decomposition $\sum 0 \cdot v_i = 0$, so by uniqueness any other decomposition of zero must coincide with this one, forcing $a_i = 0$ for every i. ■

We now prove two existence results that are mirror images of each other. Any spanning list contains a basis (trim away redundant vectors from the top), and any linearly independent list extends to a basis (build up from the bottom).

Theorem 2.3 (Spanning lists contain a basis)**.** Every spanning list of a vector space V can be reduced to a basis of V.

Proof. Let $(v_1, \ldots, v_n)$ be a spanning list of V. We process the list in order and discard any vector that turns out to be redundant.

- If $v_1 = 0$, delete it. The zero vector is never needed.
- For each i from 2 to n: if $v_i \in \text{span}(v_1, \ldots, v_{i-1})$ of the surviving earlier vectors, delete v_i.

At every step, the deleted vector lies in the span of the surviving earlier ones, so the span of the surviving list does not change. Hence the surviving list still spans V.

We claim the surviving list is linearly independent. If not, the linear dependence lemma gives some surviving vector v_i which lies in the span of the surviving earlier ones. But this is exactly the condition under which the deletion step would have removed v_i, contradicting the fact that v_i survived. Hence the surviving list is both linearly independent and spans V, making it a basis. ■

Corollary 2.3.1. Every finite-dimensional vector space has a basis.

Proof. By definition, V has a finite spanning list. Apply the previous theorem to extract a basis. ■

Theorem 2.4 (Linearly independent lists extend to a basis)**.** Every linearly independent list in a finite-dimensional vector space can be extended to a basis.

Proof. Let $(u_1, \ldots, u_m)$ be linearly independent in V. Choose any spanning list $(w_1, \ldots, w_n)$ of V, which exists because V is finite-dimensional. Form the concatenation

$$(u_1, \ldots, u_m, w_1, \ldots, w_n),$$

which still spans V since the w_i already did. Apply the spanning-list reduction procedure to extract a basis.

The key observation is that none of the u_i will be deleted during the reduction. When the procedure examines u_i, the surviving earlier vectors are $u_1, \ldots, u_{i-1}$, and u_i is not in their span because $(u_1, \ldots, u_m)$ is linearly independent. So u_i survives the deletion test and remains in the list. Therefore the resulting basis contains $u_1, \ldots, u_m$ as a prefix. ■

A direct application: any subspace of a finite-dimensional vector space has a complement — a second subspace whose direct sum with it reconstructs the ambient space.

Proposition 2.3 (Existence of a complementary subspace)**.** If V is finite-dimensional and U is a subspace of V, then there exists a subspace W of V such that $V = U \oplus W$.

Proof. Choose a basis $(u_1, \ldots, u_m)$ of U, which exists because subspaces of finite-dimensional spaces are themselves finite-dimensional (we will establish this below). Extend it to a basis $(u_1, \ldots, u_m, w_1, \ldots, w_k)$ of V, and set $W = \operatorname{span}(w_1, \ldots, w_k)$.

Every $v \in V$ has a unique basis expansion $v = \sum a_i u_i + \sum b_j w_j$, with the first sum in U and the second in W. Hence $V = U + W$. To see that the sum is direct, the two-subspace criterion says it suffices to check $U \cap W = \{0\}$. Take $v \in U \cap W$. Then v is expressible both as a linear combination $\sum a_i u_i$ of the U-basis (because $v \in U$) and as $\sum b_j w_j$ of the W-basis (because $v \in W$), so

$$\sum a_i u_i - \sum b_j w_j = 0.$$

Linear independence of the full basis of V forces $a_i = b_j = 0$ for all i, j, whence $v = 0$. ■

2.4 Dimension

We now arrive at one of the structural pillars of finite-dimensional linear algebra: the length of a basis is an intrinsic property of the space, not an artefact of which basis we chose.

Proposition 2.4 (Bases have the same length)**.** Any two bases of a finite-dimensional vector space have the same length.

Proof. Let B_1 and B_2 be bases of V. Both are linearly independent, and both span V. The comparison theorem of the previous section says the length of any linearly independent list is at most the length of any spanning list; applying this in two ways,

- B_1 is linearly independent and B_2 spans, so $|B_1| \leq |B_2|$.
- B_2 is linearly independent and B_1 spans, so $|B_2| \leq |B_1|$.

The two inequalities force $|B_1| = |B_2|$. ∎

Definition 2.6 (Dimension)**.** The dimension of a finite-dimensional vector space V, denoted $\dim V$, is the length of any basis of V.

Example 5

$\dim \mathbb{F}^n = n$, since the standard basis has length n. $\dim \mathcal{P}_d(\mathbb{F}) = d+1$, since the basis $(1, x, x^2, \ldots, x^d)$ has length $d+1$; note the off-by-one from including the constant polynomial. The space $\mathcal{P}(\mathbb{F})$ of all polynomials is infinite-dimensional.

We now collect the main structural facts about how dimension interacts with subspaces and sums.

Proposition 2.5. Every subspace of a finite-dimensional vector space is finite-dimensional.

Proof. Suppose V is finite-dimensional and $U \leq V$. We construct a basis of U by greedy extension.

If $U = \{0\}$, the empty list is a basis of U and we are done. Otherwise pick any nonzero $u_1 \in U$; the singleton (u_1) is linearly independent.

Inductively, suppose we have constructed a linearly independent list $(u_1, \dots, u_{k-1})$ in U. If $\operatorname{span}(u_1, \dots, u_{k-1}) = U$, this list is a basis of U and we stop. Otherwise pick $u_k \in U \setminus \operatorname{span}(u_1, \dots, u_{k-1})$. We claim $(u_1, \dots, u_k)$ is still linearly independent. Any nontrivial dependence $\sum_{i=1}^{k} a_i u_i = 0$ has some nonzero coefficient. If $a_k = 0$, the dependence reduces to one among $u_1, \dots, u_{k-1}$, contradicting their independence. If $a_k \neq 0$, we can solve for u_k as a linear combination of $u_1, \dots, u_{k-1}$, contradicting our choice $u_k \notin \operatorname{span}(u_1, \dots, u_{k-1})$.

This process must terminate. Every list constructed in the process is linearly independent in V, and the comparison theorem bounds the length of such a list by the length of any spanning list of V, which is finite. When the process terminates, we have a linearly independent list that spans U, that is, a basis of U. ∎

Corollary 2.4.1. If U is a subspace of a finite-dimensional vector space V, then $\dim U \leq \dim V$, with equality if and only if $U = V$.

Proof. Let B be a basis of U. As a linearly independent list in V, it can be extended to a basis of V. This extended basis has length $\dim V$ and contains B as a sublist, so $\dim U = |B| \leq \dim V$.

Suppose $\dim U = \dim V$. Then B is already a list of length $\dim V$ that is linearly independent in V. Extending it to a basis of V cannot add any vectors without exceeding length $\dim V$, so B is already a basis of V. Hence $U = \operatorname{span}(B) = V$. ∎

Theorem 2.5 (Dimension of a sum)**.** If U_1 and U_2 are finite-dimensional subspaces of V, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2). \tag{2.5}$$

Proof. The strategy is to build a basis of $U_1 + U_2$ by starting with a basis of the intersection $U_1 \cap U_2$ and extending in two directions.

Choose a basis $(v_1, \dots, v_m)$ of $U_1 \cap U_2$, so $\dim(U_1 \cap U_2) = m$. Extend this to a basis

$$(v_1, \dots, v_m, u_1, \dots, u_j)$$

of U_1, giving $\dim U_1 = m + j$. Independently extend the same starting basis to a basis

$$(v_1, \ldots, v_m, w_1, \ldots, w_k)$$

of U_2, giving $\dim U_2 = m + k$.

We claim the combined list

$$(v_1, \ldots, v_m, u_1, \ldots, u_j, w_1, \ldots, w_k)$$

is a basis of $U_1 + U_2$.

Spanning. Any element of $U_1 + U_2$ is a sum $x + y$ with $x \in U_1$ and $y \in U_2$. Expressing x in the U_1-basis puts x in $\operatorname{span}(v_i, u_i)$, and similarly for y; hence $x + y$ is a linear combination of the combined list.

Linear independence. Suppose

$$\sum a_i v_i + \sum b_i u_i + \sum c_i w_i = 0.$$

Rearrange to

$$\sum c_i w_i = -\sum a_i v_i - \sum b_i u_i.$$

The left side lies in U_2 (each $w_i \in U_2$), and the right side lies in U_1 (each $v_i, u_i \in U_1$). So both sides lie in $U_1 \cap U_2$, and we can write $\sum c_i w_i = \sum d_i v_i$ for some scalars d_i. This rearranges to

$$\sum d_i v_i - \sum c_i w_i = 0,$$

a dependence relation among the basis $(v_1, \ldots, v_m, w_1, \ldots, w_k)$ of U_2. Linear independence of this basis forces all $c_i = 0$ (and all $d_i = 0$).

With all $c_i = 0$, the original relation reduces to $\sum a_i v_i + \sum b_i u_i = 0$, a dependence among the basis of U_1. Linear independence of that basis forces $a_i = b_i = 0$.

Hence the combined list is a basis. Counting its length,

$$\dim(U_1 + U_2) = m + j + k = (m + j) + (m + k) - m = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2). \quad \blacksquare$$

Corollary 2.5.1. For finite-dimensional subspaces U_1, U_2 of V, the sum $U_1 + U_2$ is direct if and only if $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$, equivalently if and only if $U_1 \cap U_2 = \{0\}$.

Proof. By the dimension formula, $\dim(U_1+U_2) = \dim U_1+\dim U_2$ if and only if $\dim(U_1 \cap U_2) = 0$, i.e. $U_1 \cap U_2 = \{0\}$. The two-subspace direct sum criterion of the previous chapter equates the latter condition with the sum being direct. ■

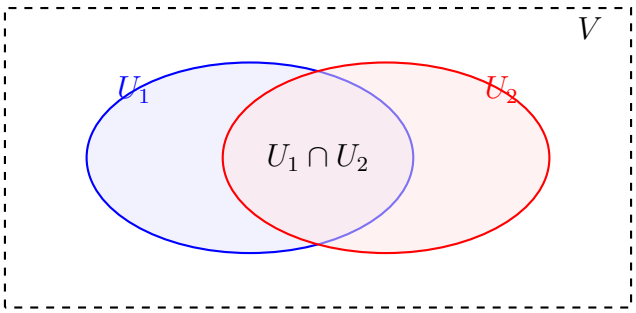

Figure 3: The dimension formula $\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$ subtracts off the double-counted intersection, in the same spirit as inclusion-exclusion for sets.

3 Linear Maps

Vector spaces on their own are static objects: we have a set, an addition, and a scalar multiplication, but no way to move between different spaces or to reshape a space in an interesting way. Linear maps are the morphisms of the theory — they are the functions between vector spaces that respect the vector space structure. Much of the rest of the course is about understanding linear maps from V to V (operators) and, through the matrix representation, concrete calculations with them.

3.1 Linear maps

Definition 3.1 (Linear map)**.** Let V and W be vector spaces over a common field $\mathbb{F}$. A function $T : V \to W$ is a linear map (or a linear transformation) if it satisfies two conditions.

1. Additivity: $T(u + v) = T(u) + T(v)$ for all $u, v \in V$.
2. Homogeneity: $T(\lambda u) = \lambda T(u)$ for all $\lambda \in \mathbb{F}$ and $u \in V$.

The two conditions can be combined into the single statement that T preserves linear combinations: $T(\lambda u + \mu v) = \lambda T(u) + \mu T(v)$. This is the defining feature of linearity and is the property on which every subsequent calculation will rest.

It is worth noting that the addition and scalar multiplication on the two sides of each condition live in different spaces: the left-hand side uses the operations of V, the right-hand side those of W. A linear map connects these two structures in a way that makes the algebra on both sides compatible.

Example 6

The function $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (x, y)$ is linear: it simply discards the third coordinate. Adding two inputs and discarding the third coordinate gives the same result as discarding each input's third coordinate first and then adding.

Similarly for scaling.

Example 7

[Zero and identity maps] The zero map sends every vector in V to the zero vector in W. The identity map on V, denoted I or I_V, sends every vector to itself. Both are linear by direct inspection.

Example 8

[Differentiation and integration] On the space $\mathcal{P}_d(\mathbb{R})$ of polynomials of degree at most d, the differentiation map $D : \mathcal{P}_d(\mathbb{R}) \to \mathcal{P}_{d-1}(\mathbb{R})$ defined by $D(p) = p'$ is linear, because differentiation satisfies $(p+q)' = p' + q'$ and $(\lambda p)' = \lambda p'$. Likewise the integration map $S : \mathcal{P}_d(\mathbb{R}) \to \mathcal{P}_{d+1}(\mathbb{R})$ defined by $S(p)(x) = \int_0^x p(t)\,dt$ is linear. Linear algebra thus sits underneath a lot of calculus: the derivative and integral are linear operators.

Definition 3.2 ($\mathcal{L}(V,W)$ and $\mathcal{L}(V)$)**.** The set of all linear maps from V to W is denoted $\mathcal{L}(V,W)$. When $V = W$, a linear map $V \to V$ is called an operator on V, and the set of all such maps is denoted $\mathcal{L}(V)$.

3.1.1 Basic properties

Every linear map preserves the zero vector. This is the first consequence of the axioms and, though nearly trivial, it is used constantly.

Proposition 3.1. Any linear map $T \in \mathcal{L}(V,W)$ sends 0_V to 0_W.

Proof. Using the additive identity property of the zero vector in V and then the additivity of T,

$$T(0_V) = T(0_V + 0_V) = T(0_V) + T(0_V).$$

We now have an equation expressing $T(0_V)$ as a sum of two copies of itself. Adding the additive inverse $-T(0_V)$ to both sides eliminates one copy from each side and gives $0_W = T(0_V)$. ∎

3.1.2 Linear maps are determined by their action on a basis

A linear map on a finite-dimensional space has, on the face of it, uncountably many values — one per input vector. The next theorem says almost the opposite: the full map is recorded by finitely many values, namely its action on a basis. This is why matrix representations work.

Theorem 3.1 (Linear extension from a basis)**.** Let V, W be vector spaces over $\mathbb{F}$, with V finite-dimensional and basis $(v_1, \ldots, v_n)$. For any choice of vectors $(w_1, \ldots, w_n)$ in W, there exists a unique linear map $T \in \mathcal{L}(V, W)$ such that $T(v_i) = w_i$ for every i.

Proof. We prove existence and uniqueness separately.

Existence. Every $v \in V$ has a unique expansion $v = a_1 v_1 + \cdots + a_n v_n$ in the basis. Define

$$T(v) = a_1 w_1 + a_2 w_2 + \cdots + a_n w_n.$$

This is well-defined precisely because the coefficients a_i are uniquely determined by v. For additivity: if $v = \sum a_i v_i$ and $v' = \sum b_i v_i$, then $v + v' = \sum (a_i + b_i) v_i$, and

$$T(v + v') = \sum (a_i + b_i) w_i = \sum a_i w_i + \sum b_i w_i = T(v) + T(v').$$

Homogeneity follows similarly: $T(\lambda v) = \sum \lambda a_i w_i = \lambda \sum a_i w_i = \lambda T(v)$. Finally, applied to $v = v_j$ (which has $a_j = 1$ and all other coefficients zero), the definition gives $T(v_j) = w_j$, as required.

Uniqueness. Suppose T and T' are two linear maps in $\mathcal{L}(V, W)$ which agree on the basis, meaning $T(v_i) = T'(v_i) = w_i$ for every i. For any $v = \sum a_i v_i$, applying additivity and homogeneity separately to both maps gives

$$T(v) = \sum a_i T(v_i) = \sum a_i T'(v_i) = T'(v),$$

so T and T' agree on every input. ∎

3.1.3 Algebra of linear maps

We can add linear maps and multiply them by scalars, and the result is again a linear map. In this way $\mathcal{L}(V, W)$ becomes itself a vector space.

Definition 3.3 (Sum and scalar multiple of linear maps)**.** For $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$, the sum $S + T$ and scalar multiple λT are the functions defined pointwise by

$$(S + T)(v) = S(v) + T(v), \qquad (\lambda T)(v) = \lambda T(v). \tag{3.1}$$

Proposition 3.2. With these operations, $\mathcal{L}(V, W)$ is a vector space.

Proof. We must check that the sum and scalar multiple of linear maps are themselves linear, and that the vector space axioms hold. Linearity of the sum: for $u, v \in V$ and $\lambda \in \mathbb{F}$,

$$\begin{aligned}(S + T)(u + v) &= S(u + v) + T(u + v) = S(u) + S(v) + T(u) + T(v) \\ &= (S(u) + T(u)) + (S(v) + T(v)) = (S + T)(u) + (S + T)(v),\end{aligned}$$

using the linearity of S and T. A similar check handles homogeneity and the scalar multiple λT.

The vector space axioms (commutativity, associativity, distributivity, and so on) for $\mathcal{L}(V, W)$ reduce to the corresponding axioms in W, applied pointwise. The zero vector of $\mathcal{L}(V, W)$ is the zero map, and the additive inverse of T is the map $-T$ defined by $(-T)(v) = -T(v)$. Since all axioms reduce to pointwise checks which hold in W, $\mathcal{L}(V, W)$ is a vector space. ■

In addition to this vector space structure, linear maps can be composed whenever the domain and codomain match up, producing another linear map.

Definition 3.4 (Product of linear maps)**.** For $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, X)$, the

product $ST \in \mathcal{L}(V, X)$ is the composition

$$(ST)(v) = S(T(v)). \tag{3.2}$$

This is really just function composition with an unusual name; we write ST instead of $S \circ T$ because it foreshadows matrix multiplication, which is exactly composition of linear maps in disguise. For the product to make sense, the domain of S must contain the range of T.

The product of linear maps is itself linear: if we trace through $ST(\lambda u + \mu v)$, the outer S and inner T each preserve linear combinations, so the composition does too.

Proposition 3.3 (Algebraic properties of the product)**.** For linear maps with compatible domains and codomains:

1. Associativity: $(T_1T_2)T_3 = T_1(T_2T_3)$.
2. Identity: $TI_V = I_WT = T$ for $T \in \mathcal{L}(V, W)$.
3. Distributivity: $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$.

Each of these is a straightforward pointwise check; the proofs amount to unpacking the definitions of composition, sum, and scalar multiple.

Remark. The product of linear maps is *not* commutative in general. The example most often given in this course: let T, S be operators on $\mathcal{P}(\mathbb{R})$ with $T(f) = f'$ and $S(f)(x) = x^2 f(x)$. By the product rule,

$$(TS)(f) = (x^2 f)' = 2xf + x^2 f',$$

whereas

$$(ST)(f) = x^2 f'.$$

The two differ by $2xf$, which is not generally zero. So $TS \neq ST$.

3.2 Null space and range

Given a linear map T, two subspaces jump out naturally: the set of inputs that get killed by T, and the set of outputs T can reach. These are called the null space and range, and they measure the extent to which T fails to be one-to-one and onto respectively.

Definition 3.5 (Null space)**.** The null space of $T \in \mathcal{L}(V, W)$ is

$$\operatorname{null}(T) = \{v \in V \mid T(v) = 0\}. \tag{3.3}$$

Definition 3.6 (Range)**.** The range of $T \in \mathcal{L}(V, W)$ is

$$\operatorname{range}(T) = \{T(v) \mid v \in V\}. \tag{3.4}$$

The null space lives inside V, the range inside W. Both are subspaces, as we now verify.

Proposition 3.4. For $T \in \mathcal{L}(V, W)$, $\operatorname{null}(T)$ is a subspace of V and $\operatorname{range}(T)$ is a subspace of W.

Proof. We apply the subspace test to each.

Null space. Since $T(0) = 0$, we have $0 \in \operatorname{null}(T)$. If $u, v \in \operatorname{null}(T)$, then $T(u) = T(v) = 0$, so

$$T(u + v) = T(u) + T(v) = 0 + 0 = 0,$$

meaning $u + v \in \operatorname{null}(T)$. Similarly $T(\lambda u) = \lambda T(u) = \lambda \cdot 0 = 0$, so $\lambda u \in \operatorname{null}(T)$. All three conditions of the subspace test hold.

Range. The zero vector 0_W is in the range because $T(0_V) = 0_W$. If $w_1, w_2 \in \operatorname{range}(T)$, there exist $v_1, v_2 \in V$ with $T(v_i) = w_i$; then $w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2) \in \operatorname{range}(T)$. Likewise $\lambda w_1 = \lambda T(v_1) = T(\lambda v_1) \in \operatorname{range}(T)$. The subspace test is satisfied.

■

3.2.1 Injectivity and surjectivity

Definition 3.7 (Injective)**.** $T : V \to W$ is injective (or one-to-one) if $T(u) = T(v)$ implies $u = v$ for all $u, v \in V$.

Definition 3.8 (Surjective)**.** $T : V \to W$ is surjective (or onto) if $\text{range}(T) = W$, that is, if every $w \in W$ equals $T(v)$ for some $v \in V$.

For general functions, injectivity requires checking every pair of inputs. For linear maps, it can be checked by examining a single subspace: the null space.

Proposition 3.5 (Injectivity test)**.** A linear map $T \in \mathcal{L}(V, W)$ is injective if and only if $\text{null}(T) = \{0\}$.

Proof. Forward direction. Assume T is injective. If $v \in \text{null}(T)$, then $T(v) = 0 = T(0)$, and injectivity forces $v = 0$. Hence the only element of the null space is 0.

Reverse direction. Assume $\text{null}(T) = \{0\}$. Suppose $T(u) = T(v)$. By linearity, $T(u - v) = T(u) - T(v) = 0$, so $u - v \in \text{null}(T) = \{0\}$. Therefore $u - v = 0$, i.e. $u = v$. ■

This test is what makes linear algebra tractable: injectivity is a condition about one special subspace, not about all pairs of vectors.

3.3 The fundamental theorem of linear maps

We now prove perhaps the most important structural result of finite-dimensional linear algebra: the dimensions of the null space and the range add up to the dimension of the domain. In pictures, a linear map takes the input space, collapses the null space to zero, and faithfully copies what is left onto the range.

Theorem 3.2 (Fundamental theorem of linear maps)**.** Let V be a finite-dimensional vector space and $T \in \mathcal{L}(V, W)$. Then $\text{range}(T)$ is finite-dimensional and

$$\dim V = \dim \text{null}(T) + \dim \text{range}(T). \tag{3.5}$$

Proof. The strategy is to build a basis of V in two layers: first a basis for $\text{null}(T)$, then additional vectors that extend it to a basis of the whole space. We will then show that T sends the additional vectors to a basis of $\text{range}(T)$.

Let $(u_1, \ldots, u_m)$ be a basis of $\text{null}(T)$, so $\dim \text{null}(T) = m$. Since $\text{null}(T) \leq V$ and V is finite-dimensional, such a basis exists and extends to a basis

$$(u_1, \ldots, u_m, v_1, \ldots, v_n)$$

of V. Thus $\dim V = m + n$.

We claim $(T(v_1), \ldots, T(v_n))$ is a basis of $\text{range}(T)$; once this is established, $\dim \text{range}(T) = n$ and the equation $\dim V = m + n$ is the desired formula.

Spanning. Any vector in $\text{range}(T)$ has the form $T(v)$ for some $v \in V$. Expand v in the full basis, $v = \sum a_i u_i + \sum b_j v_j$. Apply T to both sides; using linearity and $T(u_i) = 0$,

$$T(v) = \sum a_i T(u_i) + \sum b_j T(v_j) = \sum b_j T(v_j).$$

So any element of $\text{range}(T)$ is a linear combination of the $T(v_j)$.

Linear independence. Suppose $\sum c_j T(v_j) = 0$. Pull the linear combination inside T: $T\left(\sum c_j v_j\right) = 0$, so $\sum c_j v_j \in \text{null}(T)$. Expand this element in the null space basis: $\sum c_j v_j = \sum d_i u_i$, which we rearrange to

$$\sum c_j v_j - \sum d_i u_i = 0.$$

This is a dependence relation among the full basis of V, so all coefficients must vanish. In particular each $c_j = 0$, giving linear independence of $(T(v_1), \ldots, T(v_n))$.

Both spanning and independence hold, so $(T(v_1), \ldots, T(v_n))$ is a basis of $\text{range}(T)$. ∎

Two useful consequences follow immediately. In words: a map into a smaller space cannot be injective, and a map out of a smaller space cannot be surjective.

Corollary 3.2.1. Suppose V is finite-dimensional and $\dim V > \dim W$. Then no linear map $V \to W$ is injective.

Proof. For any $T \in \mathcal{L}(V, W)$, the fundamental theorem gives $\dim \operatorname{null}(T) = \dim V - \dim \operatorname{range}(T)$. Since $\operatorname{range}(T) \leq W$, we have $\dim \operatorname{range}(T) \leq \dim W$, and so

$$\dim \operatorname{null}(T) \geq \dim V - \dim W > 0.$$

Thus $\operatorname{null}(T) \neq \{0\}$, and by the injectivity test, T is not injective. ■

Corollary 3.2.2. Suppose V is finite-dimensional and $\dim V < \dim W$. Then no linear map $V \to W$ is surjective.

Proof. The fundamental theorem gives $\dim \operatorname{range}(T) = \dim V - \dim \operatorname{null}(T) \leq \dim V < \dim W$, so $\operatorname{range}(T) \neq W$. ■

3.4 Invertibility and isomorphisms

Finally we treat the question of when a linear map has a linear inverse. If it does, the two spaces involved are "structurally identical" from the standpoint of linear algebra: every property we care about carries across.

Definition 3.9 (Invertible)**.** A linear map $T \in \mathcal{L}(V, W)$ is invertible if there exists $S \in \mathcal{L}(W, V)$ with $ST = I_V$ and $TS = I_W$. Such an S is called the inverse of T and is denoted T^{-1}.

The inverse, when it exists, is unique: if S_1 and S_2 are both inverses of T, then we can sandwich one between T and the other,

$$S_1 = S_1 I_W = S_1(TS_2) = (S_1 T)S_2 = I_V S_2 = S_2,$$

where each equality uses one of the identity or inverse properties. So T^{-1} is well-defined.

Theorem 3.3 (Invertibility is equivalent to bijectivity)**.** A linear map $T \in \mathcal{L}(V, W)$ is invertible if and only if it is both injective and surjective.

Proof. Forward direction. Suppose T has an inverse S. If $T(u) = T(v)$, apply S to both sides: $u = ST(u) = ST(v) = v$, so T is injective. For any $w \in W$, $T(S(w)) = w$ exhibits w as a value of T, so T is surjective.

Reverse direction. Suppose T is bijective. For each $w \in W$ there exists a unique $v \in V$ with $T(v) = w$ (existence from surjectivity, uniqueness from injectivity); define $S(w) = v$. By construction $ST = I_V$ and $TS = I_W$. It remains to check that S is linear.

For additivity: let $w_1, w_2 \in W$ with $v_i = S(w_i)$, so $T(v_i) = w_i$. Then $T(v_1 + v_2) = w_1 + w_2$, so by the definition of S, $S(w_1 + w_2) = v_1 + v_2 = S(w_1) + S(w_2)$. For homogeneity: $T(\lambda v_1) = \lambda T(v_1) = \lambda w_1$, so $S(\lambda w_1) = \lambda v_1 = \lambda S(w_1)$. Thus S is linear. ■

In finite dimensions and when the two spaces have the same dimension, injectivity and surjectivity are equivalent, so checking either one suffices.

Theorem 3.4 (Injective $\iff$ surjective in same finite dimension)**.** Let V, W be finite-dimensional vector spaces with $\dim V = \dim W$, and let $T \in \mathcal{L}(V, W)$. Then the following are equivalent:

(a) T is invertible.

(b) T is injective.

(c) T is surjective.

Proof. The fundamental theorem of linear maps tells us $\dim V = \dim \operatorname{null}(T) + \dim \operatorname{range}(T)$.

Assume (b), that T is injective. Then $\operatorname{null}(T) = \{0\}$, so $\dim \operatorname{null}(T) = 0$, and hence $\dim \operatorname{range}(T) = \dim V = \dim W$. Since $\operatorname{range}(T)$ is a subspace of W with the same dimension as W, we have $\operatorname{range}(T) = W$, which is (c).

Assume (c), that T is surjective. Then $\operatorname{range}(T) = W$, so $\dim \operatorname{range}(T) = \dim W = \dim V$, and the dimension formula gives $\dim \operatorname{null}(T) = 0$, so $\operatorname{null}(T) = \{0\}$ and T is injective, which is (b).

Finally, (b) and (c) together imply bijectivity and hence by the previous theorem invertibility, which is (a); and invertibility clearly implies both (b) and (c). ■

This is one of the most useful technical theorems in the subject. In practice, to check that an operator on a finite-dimensional space is invertible, it is almost always easier to check that its null space is trivial than to explicitly invert it.

Definition 3.10 (Isomorphism)**.** An isomorphism is an invertible linear map. Two vector spaces V and W are called isomorphic if there exists an isomorphism from V to W.

Theorem 3.5. Two finite-dimensional vector spaces over the same field are isomorphic if and only if they have the same dimension.

Proof. Forward direction. If $T : V \to W$ is an isomorphism, the fundamental theorem with $\text{null}(T) = \{0\}$ gives $\dim V = \dim \text{range}(T)$; and since T is surjective, $\text{range}(T) = W$, so $\dim V = \dim W$.

Reverse direction. Suppose $\dim V = \dim W = n$. Choose bases $(v_1, \ldots, v_n)$ of V and $(w_1, \ldots, w_n)$ of W. By the linear extension theorem, there is a unique linear map $T : V \to W$ with $T(v_i) = w_i$ for each i. This map is injective, because $T(\sum a_i v_i) = \sum a_i w_i = 0$ forces all $a_i = 0$ (the w_i are linearly independent). By the previous theorem, T is invertible. ■

Corollary 3.5.1. Every n-dimensional vector space over $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.

This is a structural statement about finite-dimensional vector spaces: up to isomorphism there is essentially only one vector space of each dimension. All the concrete examples — $\mathbb{F}^n$, $\mathcal{P}_d(\mathbb{F})$, spaces of matrices, and so on — are different incarnations of the same abstract object, connected by basis-choice isomorphisms. The matrix representations developed in the next chapter exploit exactly this correspondence.

4 Matrices and Change of Basis

Matrices are the concrete arithmetical incarnation of linear maps between finite-dimensional spaces. Once we fix bases for the domain and codomain, every linear map is encoded by a rectangular array of scalars, and composition of maps becomes matrix multiplication. This translation is what makes the theory computable: we can write down a matrix, perform arithmetic with it, and read off information about the underlying map. The two parts of this chapter are the translation itself (how a linear map becomes a matrix) and the study of how that translation depends on the chosen bases.

4.1 Matrices

Definition 4.1 (Matrix)**.** For positive integers m, n and a field $\mathbb{F}$, an $m \times n$ matrix over $\mathbb{F}$ is a rectangular array of elements of $\mathbb{F}$,

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{pmatrix}. \tag{4.1}$$

The entry in row i and column j is denoted $A_{i,j}$. The set of all $m \times n$ matrices over $\mathbb{F}$ is denoted $\mathbb{F}^{m\times n}$.

Two matrices are equal if they have the same dimensions and the same entry in every position. Addition and scalar multiplication of matrices are defined entrywise:

$$(A + B)_{i,j} = A_{i,j} + B_{i,j}, \tag{4.2}$$

$$(\lambda A)_{i,j} = \lambda A_{i,j}. \tag{4.3}$$

These operations make $\mathbb{F}^{m\times n}$ a vector space of dimension mn. A basis is given by the matrices $E_{i,j}$ that have a 1 in position (i, j) and 0 everywhere else, which is why the

dimension is the total number of entries.

4.1.1 Matrix multiplication

Of the operations on matrices, multiplication is the one that looks most peculiar at first sight. Its form is dictated by the goal of making matrix multiplication correspond to composition of linear maps, which we will verify shortly.

Definition 4.2 (Matrix product)**.** For $A \in \mathbb{F}^{m\times n}$ and $C \in \mathbb{F}^{n\times p}$, their product $AC \in \mathbb{F}^{m\times p}$ has entries

$$(AC)_{i,k} = \sum_{j=1}^{n} A_{i,j}C_{j,k}. \tag{4.4}$$

The (i, k) entry of the product is the "dot product" of row i of A and column k of C. For the product to be defined, the number of columns of A must equal the number of rows of C; otherwise the sums above would not match up.

Matrix multiplication is associative and distributes over addition; both can be verified by direct entry-wise computation. It is not commutative: even when both AB and BA are defined (which requires both A and B to be square of the same size), the two products need not be equal.

A useful perspective is to think of matrix multiplication column-by-column. Writing $A_{\cdot,k}$ for the kth column of A, the formula says

$$(AC)_{\cdot,k} = A \cdot C_{\cdot,k}, \tag{4.5}$$

that is, the kth column of the product is obtained by multiplying A by the kth column of C. If we go a step further and write A in terms of its columns $a_1, \ldots, a_n$, then the kth column of AC equals

$$\sum_{j=1}^{n} C_{j,k}\, a_j, \tag{4.6}$$

a linear combination of the columns of A whose coefficients come from the kth column of

C. This view of matrix multiplication — as producing linear combinations of the columns of the left factor — is a recurring tool and comes up repeatedly in the rest of the chapter.

4.1.2 The transpose

Definition 4.3 (Transpose)**.** The transpose of $A \in \mathbb{F}^{m\times n}$, denoted A^T, is the $n \times m$ matrix obtained by interchanging rows and columns:

$$(A^T)_{k,j} = A_{j,k}. \tag{4.7}$$

Example 9

$$A = \begin{pmatrix} 3 & 1 & 8 \\ 2 & 9 & 7 \\ 4 & 2 & 0 \end{pmatrix} \implies A^T = \begin{pmatrix} 3 & 2 & 4 \\ 1 & 9 & 2 \\ 8 & 7 & 0 \end{pmatrix}.$$

Proposition 4.1 (Properties of the transpose)**.** For $A, B \in \mathbb{F}^{m\times n}$, $C \in \mathbb{F}^{n\times p}$, and $\lambda \in \mathbb{F}$,

$$(A+B)^T = A^T + B^T, \tag{4.8}$$

$$(\lambda A)^T = \lambda A^T, \tag{4.9}$$

$$(AC)^T = C^T A^T. \tag{4.10}$$

The first two are immediate from the definitions; the third reverses the order because when we swap rows and columns, the role of left and right factor in the sum defining matrix multiplication is interchanged.

4.1.3 Rank and column-row factorization

Definition 4.4 (Column and row rank)**.** The column rank of $A \in \mathbb{F}^{m\times n}$ is the dimension of the span of the columns of A viewed as vectors in $\mathbb{F}^m$. The row rank is

the dimension of the span of the rows of A viewed as vectors in $\mathbb{F}^n$.

It is not obvious a priori that column rank and row rank should agree. The next theorem establishes this via an explicit factorisation of A that exposes its rank.

Theorem 4.1 (Column-row factorization)**.** Let $A \in \mathbb{F}^{m\times n}$ have column rank $c \geq 1$. Then there exist matrices $B \in \mathbb{F}^{m\times c}$ and $R \in \mathbb{F}^{c\times n}$ such that

$$A = BR. \tag{4.11}$$

Proof. The columns of A are vectors in $\mathbb{F}^m$, and their span is c-dimensional by hypothesis. Choose c columns of A that form a basis for this span — this is possible because any spanning list of a c-dimensional space contains a basis of length c. Arrange these c columns as the columns of a matrix $B \in \mathbb{F}^{m\times c}$.

Each column of A lies in the span of the columns of B. Writing the kth column of A as a linear combination of columns of B,

$$A_{\cdot,k} = \sum_{j=1}^{c} R_{j,k} B_{\cdot,j},$$

collects the coefficients into a $c \times n$ matrix R. The displayed equation is exactly the column-by-column formula for the product BR, so $A = BR$. ∎

Corollary 4.1.1 (Row rank equals column rank)**.** For any matrix A, the row rank equals the column rank. This common value is called the rank of A and denoted $\operatorname{rank}(A)$.

Proof. Let c be the column rank of A, and write $A = BR$ as in the theorem. Each row of A is a linear combination of the rows of R (this is the row-by-row dual of the column formula), so the row space of A is contained in the row space of R. But R has only c rows, so its row space has dimension at most c. Therefore the row rank of A is at most c.

Applying the same argument to A^T (whose column rank is the row rank of A) gives the reverse inequality. Hence row rank equals column rank. ∎

4.2 Matrix of a linear map

We now formalise the correspondence between linear maps and matrices. The key insight is that a linear map on a finite-dimensional space is determined by its values on a basis, and each of those values can be recorded as a list of coordinates in another basis. Packaging all those lists into columns gives a matrix.

Definition 4.5 (Matrix of a linear map)**.** Let V and W be finite-dimensional vector spaces with bases $\mathcal{B}_V = (v_1, \dots, v_n)$ and $\mathcal{B}_W = (w_1, \dots, w_m)$, and let $T \in \mathcal{L}(V, W)$. The matrix of T with respect to these bases is the $m \times n$ matrix $\mathcal{M}(T)$ whose (i, k) entry $a_{i,k}$ is determined by the expansion

$$T(v_k) = a_{1,k} w_1 + a_{2,k} w_2 + \cdots + a_{m,k} w_m. \tag{4.12}$$

In other words, the kth column of $\mathcal{M}(T)$ records the coefficients of $T(v_k)$ in the basis of W.

When the bases need to be made explicit (which becomes important when we change basis), we write $\mathcal{M}(T, \mathcal{B}_V, \mathcal{B}_W)$. For an operator $T \in \mathcal{L}(V)$ we usually use the same basis for both domain and codomain and abbreviate to $\mathcal{M}(T, \mathcal{B})$.

Example 10

Let $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ be defined by $T(x, y) = (x + 3y, 2x, 4x - 5y)$. Using the standard bases on $\mathbb{R}^2$ and $\mathbb{R}^3$ respectively,

$$T(1, 0) = (1, 2, 4),$$
$$T(0, 1) = (3, 0, -5).$$

These become the two columns of the matrix:

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 3 \\ 2 & 0 \\ 4 & -5 \end{pmatrix}.$$

The matrix construction respects the linear structure on $\mathcal{L}(V, W)$ and the product structure, in the natural way.

Proposition 4.2 (Matrix of sum and scalar multiple)**.** For $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$, using any fixed bases,

$$\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T), \qquad \mathcal{M}(\lambda T) = \lambda \mathcal{M}(T). \tag{4.13}$$

The proof is a direct comparison of columns: the kth column of $\mathcal{M}(S + T)$ records the coefficients of $(S + T)(v_k) = S(v_k) + T(v_k)$ in the basis of W, which is just the sum of the columns of $\mathcal{M}(S)$ and $\mathcal{M}(T)$.

The deeper fact is that the matrix of a composition is the product of the matrices. This is what retroactively justifies the unusual-looking formula for matrix multiplication.

Proposition 4.3 (Matrix of a product)**.** For $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, X)$ with chosen bases on all three spaces,

$$\mathcal{M}(ST) = \mathcal{M}(S) \cdot \mathcal{M}(T). \tag{4.14}$$

Proof. Fix bases (v_k) for V, (w_j) for W, and (x_i) for X. Let $A = \mathcal{M}(T)$ and $B = \mathcal{M}(S)$, so by definition

$$T(v_k) = \sum_j A_{j,k}\, w_j, \qquad S(w_j) = \sum_i B_{i,j}\, x_i.$$

To compute $\mathcal{M}(ST)$, we need the expansion of $(ST)(v_k)$ in the basis of X. Applying S

to the expansion of $T(v_k)$, using linearity, and then substituting the expansion of $S(w_j)$,

$$\begin{aligned}(ST)(v_k) = S(T(v_k)) &= S\left(\sum_j A_{j,k}\, w_j\right) = \sum_j A_{j,k}\, S(w_j) \\ &= \sum_j A_{j,k}\left(\sum_i B_{i,j}\, x_i\right) = \sum_i \left(\sum_j B_{i,j} A_{j,k}\right) x_i.\end{aligned}$$

The coefficient of x_i in the last line is $\sum_j B_{i,j}A_{j,k}$, which is the (i,k) entry of the matrix product $BA = \mathcal{M}(S)\mathcal{M}(T)$. Hence $\mathcal{M}(ST)_{i,k} = (\mathcal{M}(S)\mathcal{M}(T))_{i,k}$, proving $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$. ■

4.2.1 Matrix of a vector

The same idea of recording an object by its coordinates in a basis applies to individual vectors. The result is a column vector.

Definition 4.6 (Matrix of a vector)**.** If $v \in V$ has basis expansion $v = b_1v_1 + \cdots + b_nv_n$, the matrix of v with respect to this basis is the column vector

$$\mathcal{M}(v) = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{F}^{n\times 1}. \tag{4.15}$$

The map $v \mapsto \mathcal{M}(v)$ is, once a basis is fixed, an isomorphism from V to $\mathbb{F}^{n\times 1}$. Every n-dimensional vector space looks like $\mathbb{F}^n$ once we put coordinates on it.

The following theorem is the cleanest statement of the correspondence between linear maps and matrices: applying a linear map to a vector is the same as multiplying the corresponding matrix by the corresponding column vector.

Theorem 4.2 (Matrix-vector correspondence)**.** Let $T \in \mathcal{L}(V,W)$ with bases chosen on both spaces. For every $v \in V$,

$$\mathcal{M}(T(v)) = \mathcal{M}(T) \cdot \mathcal{M}(v). \tag{4.16}$$

Proof. Write $v = \sum_k b_k v_k$ in the basis of V, so $\mathcal{M}(v)$ has entries b_k. Applying T and using linearity,

$$T(v) = \sum_k b_k T(v_k) = \sum_k b_k \sum_i A_{i,k} w_i = \sum_i \left(\sum_k A_{i,k} b_k \right) w_i,$$

where $A = \mathcal{M}(T)$. The coefficient of w_i is the ith entry of the column vector $A\mathcal{M}(v)$, so $\mathcal{M}(T(v)) = A\mathcal{M}(v)$. ∎

This theorem is the reason matrix multiplication was defined the way it was. The definition is not arbitrary: it is forced by the requirement that matrix arithmetic reproduce the action of linear maps.

4.3 Invertibility of matrices

Definition 4.7 (Invertible matrix)**.** A square matrix $A \in \mathbb{F}^{n \times n}$ is invertible if there exists a matrix $B \in \mathbb{F}^{n \times n}$ with $AB = BA = I_n$. When such B exists it is unique, and we write $B = A^{-1}$.

Invertibility of a matrix matches invertibility of the corresponding operator.

Proposition 4.4. An operator $T \in \mathcal{L}(V)$ is invertible if and only if $\mathcal{M}(T)$ is invertible, with respect to any (equivalently, every) basis.

Proof. Suppose T is invertible with inverse S. Then $ST = TS = I$, and applying the matrix-of-a-product formula gives $\mathcal{M}(S)\mathcal{M}(T) = \mathcal{M}(T)\mathcal{M}(S) = \mathcal{M}(I) = I_n$. So $\mathcal{M}(T)$ is invertible, with inverse $\mathcal{M}(S)$.

Conversely, suppose $\mathcal{M}(T)$ is invertible with inverse B. By the linear extension theorem, there is a unique operator $S \in \mathcal{L}(V)$ whose matrix with respect to the chosen basis is B. Then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T) = B\mathcal{M}(T) = I = \mathcal{M}(I)$, and since the map from operators to matrices is injective (a matrix determines an operator uniquely, given a basis), we conclude $ST = I$. Similarly $TS = I$. Hence T is invertible. ∎

For operators on finite-dimensional spaces, one-sided inverses suffice — the second direction comes for free.

Theorem 4.3 (One-sided inverse suffices in finite dimensions)**.** Let V be finite-dimensional and $T \in \mathcal{L}(V)$. If $S \in \mathcal{L}(V)$ satisfies $ST = I$ or $TS = I$, then T is invertible and $S = T^{-1}$.

Proof. Suppose $ST = I$. To show T is injective: if $T(v) = 0$, apply S to get $v = S(T(v)) = S(0) = 0$, so $\text{null}(T) = \{0\}$. By the theorem that injective implies invertible in equal finite dimension, T is invertible.

To identify S as the inverse, we insert the inverse pair $TT^{-1} = I$,

$$S = SI = S(TT^{-1}) = (ST)T^{-1} = IT^{-1} = T^{-1}.$$

The case $TS = I$ is handled by applying the same argument with the roles of S and T swapped in the appropriate places, together with surjectivity of T in place of injectivity. ∎

4.4 Change of basis

So far we have fixed bases once and for all. But any finite-dimensional space has many bases, and the matrix of a given operator depends on which one we pick. In this section we ask how the matrix of an operator transforms when we swap one basis for another. The answer is the conjugation formula $A = C^{-1}BC$, which is one of the most important identities in the subject.

Definition 4.8 (Change-of-basis matrix)**.** Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be two bases of V. The change-of-basis matrix from (u_i) to (v_i) is the matrix of the identity operator with respect to these two bases,

$$C = \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)). \tag{4.17}$$

Its kth column records the coefficients of u_k expanded in the v-basis,

$$u_k = \sum_{i=1}^{n} C_{i,k} v_i. \tag{4.18}$$

Because I is its own inverse, the change-of-basis matrices in opposite directions are inverses of each other.

Proposition 4.5. The change-of-basis matrix $C = \mathcal{M}(I, (u_i), (v_i))$ is invertible, with inverse $\mathcal{M}(I, (v_i), (u_i))$.

Proof. Using the matrix-of-a-product formula with intermediate bases,

$$\mathcal{M}(I, (u_i), (u_i)) = \mathcal{M}(I, (v_i), (u_i)) \cdot \mathcal{M}(I, (u_i), (v_i)).$$

The left-hand side is the matrix of the identity operator in a single basis, which is I_n. So the product of the two change-of-basis matrices is I_n, and they are inverses. ■

Example 11

In $V = \mathbb{R}^2$, let $\mathcal{U} = ((1,0),(0,1))$ be the standard basis and $\mathcal{V} = ((4,2),(5,3))$ another basis. Because the $\mathcal{V}$-vectors are given in standard coordinates, the change-of-basis matrix from $\mathcal{V}$ to $\mathcal{U}$ is just these two vectors read off as columns,

$$\mathcal{M}(I, \mathcal{V}, \mathcal{U}) = \begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix}.$$

The reverse change-of-basis matrix (which expresses the standard basis vectors in the $\mathcal{V}$-basis) is the inverse,

$$\mathcal{M}(I, \mathcal{U}, \mathcal{V}) = \begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix}^{-1} = \frac{1}{4 \cdot 3 - 5 \cdot 2} \begin{pmatrix} 3 & -5 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 3/2 & -5/2 \\ -1 & 2 \end{pmatrix}.$$

Theorem 4.4 (Change-of-basis formula)**.** Let $T \in \mathcal{L}(V)$, and let (u_i) and (v_i) be two

bases of V. Set

$$A = \mathcal{M}(T, (u_i)), \qquad B = \mathcal{M}(T, (v_i)), \qquad C = \mathcal{M}(I, (u_i), (v_i)). \tag{4.19}$$

Then

$$A = C^{-1}BC. \tag{4.20}$$

Proof. We apply the matrix-of-a-product formula with carefully chosen intermediate bases. The identity $T = I \circ T \circ I$ expresses T as a composition of three operators; taking the matrix with respect to the (u_i)-basis on both sides and inserting the (v_i)-basis on the intermediate spaces gives

$$\mathcal{M}(T, (u_i), (u_i)) = \mathcal{M}(I, (v_i), (u_i)) \cdot \mathcal{M}(T, (v_i), (v_i)) \cdot \mathcal{M}(I, (u_i), (v_i)).$$

The left-hand side is A. The three factors on the right are, respectively, C^{-1} (since $C = \mathcal{M}(I, (u_i), (v_i))$), B, and C. Thus $A = C^{-1}BC$. ∎

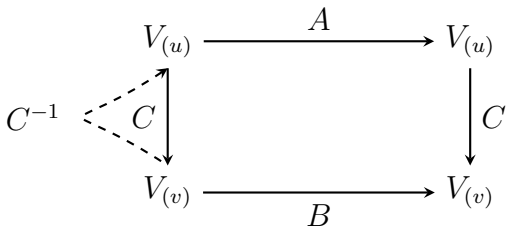

Figure 4: The change-of-basis formula $A = C^{-1}BC$. Going around the diagram: start in the (u)-coordinates, translate to the (v)-coordinates (C), apply T there (B), translate back (C^{-1}).

Definition 4.9 (Similar matrices)**.** Two matrices $A, B \in \mathbb{F}^{n\times n}$ are similar if there exists an invertible $C \in \mathbb{F}^{n\times n}$ such that $A = C^{-1}BC$.

Similarity is an equivalence relation on $\mathbb{F}^{n\times n}$, and the equivalence classes are exactly the sets of matrices that represent the same operator in different bases. Many properties of an operator (trace, determinant, characteristic polynomial, eigenvalues, rank) can be read off any matrix representing it, and are therefore invariants of the similarity class. Finding

a particularly simple matrix in a similarity class — diagonal when possible, Jordan form when not — is one of the central problems of the second half of the course.

5 Eigenvalues and Invariant Subspaces

An operator on a finite-dimensional vector space is a complicated object: it can be described by an $n \times n$ matrix, but this description changes when we change the basis, and the raw matrix tells us very little about how the operator behaves on vectors. A productive strategy is to look for subspaces on which the operator acts especially simply, and the simplest subspaces of all are the one-dimensional ones on which the operator acts as multiplication by a scalar. This leads directly to the notion of eigenvalue and eigenvector, which underpin almost everything that comes later: diagonalisability, the spectral theorem, Jordan form, the determinant as a product of eigenvalues, and more.

5.1 Invariant subspaces

Definition 5.1 (Invariant subspace)**.** Let $T \in \mathcal{L}(V)$ and let U be a subspace of V. We say U is invariant under T, or T-invariant, if $T(u) \in U$ for every $u \in U$. Equivalently, $T(U) \subseteq U$.

If a subspace U is T-invariant, then the operator T, though defined on all of V, can be restricted to act from U to U. We denote this restriction $T\big|_U$. Invariant subspaces are the natural places to "cut V apart" for the study of T: if we can decompose V into smaller invariant subspaces, we have reduced the problem of understanding T to the problem of understanding its restriction to each piece.

Example 12

For any $T \in \mathcal{L}(V)$, the subspaces $\{0\}$, V, $\operatorname{null}(T)$, and $\operatorname{range}(T)$ are all invariant. The first two are automatic: T sends $\{0\}$ to itself and sends V into V. For $\operatorname{null}(T)$: if $u \in \operatorname{null}(T)$ then $T(u) = 0 \in \operatorname{null}(T)$. For $\operatorname{range}(T)$: if $u = T(v)$ for some $v \in V$, then $T(u) = T(T(v))$ is itself in the range.

5.1.1 One-dimensional invariant subspaces

The simplest invariant subspaces are the one-dimensional ones. Suppose $U = \operatorname{span}(v)$ for some nonzero $v \in V$; so U consists of all scalar multiples of v. For U to be invariant under T, we need $T(v) \in U$, which is to say $T(v) = \lambda v$ for some scalar $\lambda \in \mathbb{F}$. This equation is the defining relation of an eigenvector, and it motivates the whole of the following theory.

5.2 Eigenvalues and eigenvectors

Definition 5.2 (Eigenvalue)**.** Let $T \in \mathcal{L}(V)$. A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of T if there exists a nonzero vector $v \in V$ with

$$T(v) = \lambda v. \tag{5.1}$$

Definition 5.3 (Eigenvector)**.** Let $T \in \mathcal{L}(V)$ and λ be an eigenvalue of T. A vector $v \in V$ is an eigenvector of T corresponding to λ if $v \neq 0$ and $T(v) = \lambda v$.

The requirement $v \neq 0$ is crucial. Without it, every scalar λ would be an eigenvalue via the trivial equation $T(0) = \lambda \cdot 0$, which says nothing interesting. The zero vector is deliberately excluded from the definition of an eigenvector.

Geometrically, an eigenvector of T is a nonzero vector whose direction is preserved by T: it gets stretched (or compressed, or reflected) by a factor λ, but it does not rotate off its own line. For an operator on $\mathbb{R}^2$ or $\mathbb{R}^3$, eigenvectors are therefore the "axes" along which the operator acts as pure scaling.

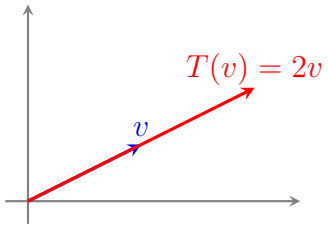

Figure 5: An eigenvector v and its image $T(v)$. The image lies on the same line through the origin as v itself; the eigenvalue $\lambda = 2$ records the stretch factor.

Example 13

Define $T \in \mathcal{L}(\mathbb{R}^3)$ by $T(x, y, z) = (7x + 3z, 3x + 6y + 9z, -6y)$. We check whether $(3, 1, -1)$ is an eigenvector:

$$T(3, 1, -1) = (21 - 3, 9 + 6 - 9, -6) = (18, 6, -6) = 6 \cdot (3, 1, -1).$$

So $(3, 1, -1)$ is indeed an eigenvector of T with eigenvalue 6.

Eigenvalues admit several equivalent characterisations. The most useful one phrases the existence of an eigenvector in terms of the failure of invertibility of a related operator.

Proposition 5.1 (Characterisations of eigenvalues)**.** Let $T \in \mathcal{L}(V)$ with V finite-dimensional, and let $\lambda \in \mathbb{F}$. The following are equivalent.

1. λ is an eigenvalue of T.
2. $T - \lambda I$ is not injective.
3. $T - \lambda I$ is not surjective.
4. $T - \lambda I$ is not invertible.

Proof. We begin by unfolding the definition of eigenvalue. By definition, λ is an eigenvalue of T if and only if there exists $v \neq 0$ with $T(v) = \lambda v$, equivalently $(T - \lambda I)(v) = 0$. This last statement says $\operatorname{null}(T - \lambda I)$ contains a nonzero vector, which is precisely the failure of injectivity of $T - \lambda I$. Hence (1) $\iff$ (2).

In a finite-dimensional space with equal-dimensional domain and codomain (which holds here since $T - \lambda I$ is an operator on V), injectivity, surjectivity, and invertibility are equivalent, as established in the previous chapter. So (2), (3), and (4) are all equivalent. ∎

The next theorem is one of the cornerstones of the subject. It says that eigenvectors corresponding to distinct eigenvalues are automatically independent — no two eigenlines

ever coincide or become redundant.

Theorem 5.1 (Eigenvectors with distinct eigenvalues are linearly independent)**.** Let $T \in \mathcal{L}(V)$ and let $v_1, \ldots, v_m$ be eigenvectors of T corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$. Then $(v_1, \ldots, v_m)$ is linearly independent.

Proof. We argue by contradiction, using the minimality of a counterexample to force a shorter counterexample, which is absurd.

Suppose for contradiction that some finite list of eigenvectors with distinct eigenvalues is linearly dependent. Among all such dependent lists, choose one of minimum length, say $(v_1, \ldots, v_m)$ with corresponding distinct eigenvalues $\lambda_1, \ldots, \lambda_m$. Any list of length one consisting of a single eigenvector is linearly independent (eigenvectors are nonzero), so $m \geq 2$, and by minimality the truncated list $(v_1, \ldots, v_{m-1})$ is linearly independent.

Because the full list is dependent, there exist scalars $a_1, \ldots, a_m$, not all zero, with

$$a_1 v_1 + a_2 v_2 + \cdots + a_m v_m = 0. \tag{$\star$}$$

We claim $a_m \neq 0$. If it were zero, $(\star)$ would be a nontrivial dependence among $v_1, \ldots, v_{m-1}$, contradicting the independence of the truncated list. So $a_m \neq 0$, and correspondingly not all of $a_1, \ldots, a_{m-1}$ can be zero (otherwise $(\star)$ would reduce to $a_m v_m = 0$ forcing $v_m = 0$, which contradicts v_m being an eigenvector).

Now apply the operator $T - \lambda_m I$ to both sides of $(\star)$. Using $(T - \lambda_m I)v_i = (\lambda_i - \lambda_m)v_i$,

$$a_1(\lambda_1 - \lambda_m)v_1 + a_2(\lambda_2 - \lambda_m)v_2 + \cdots + a_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1} = 0.$$

The v_m term has dropped out because $(\lambda_m - \lambda_m) = 0$. Since the λ_i are distinct, $\lambda_i - \lambda_m \neq 0$ for $i < m$, and so each $a_i(\lambda_i - \lambda_m)$ is a nonzero multiple of a_i. We noted above that not all a_i for $i < m$ are zero, so the displayed equation is a nontrivial dependence among $v_1, \ldots, v_{m-1}$, contradicting the independence of the truncated list.

The assumption of a dependent list of eigenvectors with distinct eigenvalues has led to a contradiction, so no such list exists. ∎

Corollary 5.1.1. An operator on an n-dimensional vector space has at most n distinct eigenvalues.

If there were more than n, the corresponding eigenvectors would form a linearly independent list of length greater than $\dim V$, which is impossible.

5.3 Polynomials applied to operators

Because $\mathcal{L}(V)$ has composition as a product, we can apply polynomials to an operator in the same way we apply them to a scalar.

Definition 5.4 (Polynomial applied to an operator)**.** For a polynomial $p(z) = a_0 + a_1 z + \cdots + a_m z^m$ and an operator $T \in \mathcal{L}(V)$, define

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m, \tag{5.2}$$

where $T^k = T \circ T \circ \cdots \circ T$ is the k-fold composition of T with itself, with the convention $T^0 = I$.

Some basic arithmetic properties are inherited from the corresponding facts for polynomials.

Proposition 5.2. For any polynomials $p, q \in \mathcal{P}(\mathbb{F})$ and operator $T \in \mathcal{L}(V)$,

1. $(pq)(T) = p(T)q(T)$,
2. $p(T)q(T) = q(T)p(T)$.

Proof. For (1), expand p and q as formal polynomials and multiply them out, noting that $T^i T^j = T^{i+j}$. The resulting expansion of $(pq)(T)$ matches the product $p(T)q(T)$ term by term.

For (2), the polynomial pq equals qp as polynomials, so by (1), $p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T)$. ∎

Remark. Though operator multiplication is not commutative in general, two polynomials in the *same* operator always commute. The commutativity comes from the fact that T commutes with itself, which is all that enters the calculation.

5.4 Existence of eigenvalues on complex spaces

Over the complex numbers, every polynomial factors completely into linear factors. This is the fundamental theorem of algebra, which we record here as a black box.

Theorem 5.2 (Fundamental theorem of algebra)**.** Let $p(z) \in \mathcal{P}_m(\mathbb{C})$ be a non-constant polynomial of degree m with leading coefficient $a_m \neq 0$. Then p has a unique factorisation (up to ordering)

$$p(z) = a_m(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_m), \tag{5.3}$$

with $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$. The λ_i are called the roots of p; they are precisely the solutions to $p(\lambda_i) = 0$.

The proof is an analytic result — we refer any standard complex analysis text. The key consequence we need is that the fundamental theorem lets us factor polynomials applied to operators into products of linear factors, and linear factors $(T - \lambda I)$ are exactly what detect eigenvalues. This gives a non-constructive existence result that is far stronger than anything available over $\mathbb{R}$.

Theorem 5.3 (Existence of eigenvalues over $\mathbb{C}$)**.** Every operator on a finite-dimensional, non-zero complex vector space has an eigenvalue.

Proof. Let V be a non-zero finite-dimensional complex vector space with $\dim V = n$, and let $T \in \mathcal{L}(V)$. The plan is to produce a polynomial p with $p(T)u = 0$ for some nonzero u, then use the factorisation of p to conclude that one of the linear factors $(T - \lambda I)$ annihilates some vector.

Pick any nonzero $u \in V$ and consider the list of $n + 1$ vectors

$$(u, Tu, T^2u, \ldots, T^nu).$$

This list lives in the n-dimensional space V, so it must be linearly dependent. Hence there exist scalars $a_0, \ldots, a_n \in \mathbb{C}$, not all zero, with

$$a_0 u + a_1 T u + a_2 T^2 u + \cdots + a_n T^n u = 0.$$

Because $u \neq 0$, some a_i with $i \geq 1$ must be nonzero (if only a_0 were nonzero the equation would read $a_0 u = 0$, forcing $a_0 = 0$). Let m be the largest index with $a_m \neq 0$, so $1 \leq m \leq n$, and define the polynomial

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m.$$

By the fundamental theorem of algebra, $p(z) = a_m(z-\lambda_1)\cdots(z-\lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$. Substituting T for z and applying both sides to u,

$$0 = p(T)u = a_m(T - \lambda_1 I)(T - \lambda_2 I)\cdots(T - \lambda_m I)u.$$

Since $a_m \neq 0$, the product of the $(T - \lambda_j I)$ applied to u is zero. Hence at least one of these operators maps some nonzero vector to zero. Specifically, applying the operators right-to-left, let k be the smallest index such that

$$(T - \lambda_k I)\left((T - \lambda_{k+1} I)\cdots(T - \lambda_m I)u\right) = 0$$

while the vector in parentheses is nonzero. (Such k exists, since the overall composition annihilates u but by choice of k the inner product is nonzero.) This inner vector is then an eigenvector of T with eigenvalue λ_k. ■

Remark. This theorem fails for real vector spaces. The rotation $T \in \mathcal{L}(\mathbb{R}^2)$ defined by $T(x, y) = (-y, x)$ has no real eigenvalue: geometrically it rotates every nonzero vector by $90°$, so no nonzero vector is preserved up to scaling. The underlying issue is that $\mathbb{R}$ is not algebraically closed — the polynomial $z^2 + 1$ has no real roots — whereas $\mathbb{C}$ is.

5.5 Upper-triangular matrices

Not every operator can be diagonalised, but on a complex vector space, every operator can at least be upper-triangularised. We develop this result now, because it will underlie many subsequent theorems (including Schur's theorem and the generalised eigenspace decomposition).

Definition 5.5 (Upper-triangular matrix)**.** A square matrix $A \in \mathbb{F}^{n\times n}$ is upper-triangular if $A_{i,j} = 0$ whenever $i > j$. Equivalently, all entries below the main diagonal vanish.

There is a clean basis-free interpretation: a matrix is upper-triangular in a given basis precisely when the successive spans of the basis vectors are invariant subspaces.

Proposition 5.3 (Characterisations of upper-triangular)**.** Let $T \in \mathcal{L}(V)$ and let $(v_1, \dots, v_n)$ be a basis of V. The following are equivalent.

1. $\mathcal{M}(T, (v_i))$ is upper-triangular.
2. $T(v_k) \in \operatorname{span}(v_1, \dots, v_k)$ for each $k = 1, \dots, n$.
3. $\operatorname{span}(v_1, \dots, v_k)$ is T-invariant for each k.

The equivalences follow by unpacking the definition of the matrix: the kth column of $\mathcal{M}(T)$ records the coefficients of $T(v_k)$ in the basis, and upper-triangularity of the matrix says the coefficients of $v_{k+1}, \dots, v_n$ in that column vanish, which is exactly (2). Statement (3) is slightly stronger-looking but follows from (2) by closing under addition.

The central theorem of this section is that operators on finite-dimensional complex spaces always admit an upper-triangular representation.

Theorem 5.4 (Schur's theorem over $\mathbb{C}$)**.** Every operator on a finite-dimensional complex vector space has an upper-triangular matrix with respect to some basis.

Proof. We proceed by strong induction on $n = \dim V$.

Base case $n = 1$. Every 1×1 matrix is trivially upper-triangular.

Inductive step. Assume the theorem holds for all complex vector spaces of dimension strictly less than n. Let $\dim V = n$ and $T \in \mathcal{L}(V)$. By the existence theorem above, T has an eigenvalue $\lambda \in \mathbb{C}$. Set

$$U = \operatorname{range}(T - \lambda I).$$

Because $T - \lambda I$ is not surjective (the existence of an eigenvalue is equivalent to its non-invertibility), $\dim U < \dim V$.

U is T-invariant. For $u \in U$, write $T(u) = (T - \lambda I)(u) + \lambda u$. The first term on the right is in U because it is an image of $T - \lambda I$, which has range U. The second term is in U because U is a subspace (closed under scalar multiplication) and $u \in U$. Their sum is therefore in U, so $T(u) \in U$.

Apply induction to $T\big|_U$. Since $\dim U < n$, the induction hypothesis applies to the restriction $T\big|_U : U \to U$. There is a basis $(u_1, \ldots, u_k)$ of U for which $\mathcal{M}(T\big|_U)$ is upper-triangular, equivalently $T(u_j) \in \operatorname{span}(u_1, \ldots, u_j)$ for each j.

Extend to a basis of V. Extend $(u_1, \ldots, u_k)$ to a basis $(u_1, \ldots, u_k, v_1, \ldots, v_\ell)$ of V. We claim the matrix of T with respect to this basis is upper-triangular, which amounts to showing T applied to each basis vector lies in the span of the basis vectors up to and including it.

For the u-part, the induction hypothesis already gives $T(u_j) \in \operatorname{span}(u_1, \ldots, u_j) \subseteq \operatorname{span}(u_1, \ldots, u_j, v_1, \ldots, v_\ell)$. For the v-part, write $T(v_j) = (T - \lambda I)(v_j) + \lambda v_j$. The first term is in $\operatorname{range}(T - \lambda I) = U = \operatorname{span}(u_1, \ldots, u_k)$. The second term λv_j is in $\operatorname{span}(v_j)$. So

$$T(v_j) \in \operatorname{span}(u_1, \ldots, u_k) + \operatorname{span}(v_j) \subseteq \operatorname{span}(u_1, \ldots, u_k, v_1, \ldots, v_j).$$

This is exactly the condition for upper-triangularity in the full basis. ∎

Upper-triangular matrices make eigenvalues especially easy to read off.

Proposition 5.4 (Eigenvalues from upper-triangular form)**.** If $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis, the eigenvalues of T are exactly

the diagonal entries of this matrix.

Proof. Write the matrix as A with diagonal entries $\lambda_1, \ldots, \lambda_n$. We use the characterisation "λ is an eigenvalue if and only if $T - \lambda I$ is not invertible", which for a square matrix is the same as the matrix of $T - \lambda I$ being singular.

The matrix of $T - \lambda I$ in this basis is $A - \lambda I_n$, another upper-triangular matrix whose diagonal entries are $\lambda_1 - \lambda, \ldots, \lambda_n - \lambda$.

If $\lambda = \lambda_i$ for some i, the (i, i) diagonal entry of $A - \lambda I_n$ is zero. We will see in the determinant chapter that an upper-triangular matrix with a zero on the diagonal is singular; alternatively, we can argue directly by showing that its columns are linearly dependent (the ith column has its first i entries determined by the upper-triangular structure and is a linear combination of earlier columns). So λ is an eigenvalue.

Conversely, if $\lambda \neq \lambda_i$ for any i, all diagonal entries of $A - \lambda I_n$ are nonzero. Such a matrix is invertible: its triangular structure, combined with nonzero diagonal, allows back-substitution to solve every system $(A - \lambda I_n)x = y$. So $T - \lambda I$ is invertible, and λ is not an eigenvalue. ■

5.6 Eigenspaces and diagonalisability

For each eigenvalue λ, the set of eigenvectors corresponding to λ, together with the zero vector, is a subspace.

Definition 5.6 (Eigenspace)**.** Let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The eigenspace of T corresponding to λ is

$$E(\lambda, T) = \{v \in V \mid T(v) = \lambda v\} = \text{null}(T - \lambda I). \tag{5.4}$$

The two descriptions coincide because $T(v) = \lambda v$ is the same equation as $(T - \lambda I)(v) = 0$. Since null spaces are always subspaces, so is $E(\lambda, T)$. It consists of all eigenvectors corresponding to λ together with 0. If λ is not an eigenvalue, $E(\lambda, T) = \{0\}$.

The next proposition strengthens the theorem that eigenvectors with distinct eigenval-

ues are linearly independent, by upgrading the statement from individual eigenvectors to whole eigenspaces.

Proposition 5.5 (The sum of eigenspaces is direct)**.** If $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of $T \in \mathcal{L}(V)$, then the sum $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum.

Proof. By the direct sum criterion, we need to show that the only way to write zero as $v_1 + \cdots + v_m$ with $v_i \in E(\lambda_i, T)$ is with every $v_i = 0$.

Suppose $v_1 + \cdots + v_m = 0$ with $v_i \in E(\lambda_i, T)$. Consider the sublist of nonzero v_i's; call them $v_{i_1}, \ldots, v_{i_k}$. Each is an eigenvector of T with eigenvalue λ_{i_j}, and the λ_{i_j} are distinct (being a subset of the distinct λ_i). By the theorem that eigenvectors with distinct eigenvalues are linearly independent, the list $(v_{i_1}, \ldots, v_{i_k})$ is linearly independent. But their sum is zero, so $k = 0$, meaning the sublist is empty and all $v_i = 0$. ■

Corollary 5.4.1. $\sum_{i=1}^{m} \dim E(\lambda_i, T) \leq \dim V$ for any list of distinct eigenvalues of T.

This is because the sum of the eigenspaces is a direct sum inside V, so the dimensions add and cannot exceed $\dim V$.

Definition 5.7 (Diagonalisable)**.** An operator $T \in \mathcal{L}(V)$ is diagonalisable if there exists a basis of V with respect to which $\mathcal{M}(T)$ is diagonal.

A diagonal matrix represents the simplest possible operator: each basis vector is scaled independently. Diagonalisability is therefore a highly desirable property when it holds. The next theorem gathers the equivalent ways to characterise it.

Theorem 5.5 (Characterisations of diagonalisability)**.** Let $T \in \mathcal{L}(V)$ with V finite-dimensional, and let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. The following are equivalent.

(a) T is diagonalisable.

(b) V has a basis consisting of eigenvectors of T.

(c) $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.

(d) $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

Proof. We prove a circle of implications.

(a) $\iff$ (b). The matrix of T with respect to $(v_1, \ldots, v_n)$ is diagonal if and only if $T(v_k)$ equals a scalar times v_k for each k, which is the definition of each v_k being an eigenvector.

(b) $\Rightarrow$ (c). Suppose V has a basis $(v_1, \ldots, v_n)$ of eigenvectors of T. Every element of V is a linear combination of basis vectors, and each basis vector lies in one of the eigenspaces. Grouping the terms of the linear combination by eigenvalue shows $V \subseteq E(\lambda_1, T) + \cdots + E(\lambda_m, T)$. The reverse containment is trivial, and the sum is direct by the proposition above.

(c) $\Rightarrow$ (d). Dimension is additive across direct sums.

(d) $\Rightarrow$ (b). For each eigenvalue λ_i, choose a basis of $E(\lambda_i, T)$. Concatenate these lists into a single list $\mathcal{L}$ of length $\sum \dim E(\lambda_i, T) = \dim V$.

To show $\mathcal{L}$ is a basis, it suffices to show it is linearly independent; a linearly independent list of length $\dim V$ is automatically a basis. So suppose we have a linear combination of $\mathcal{L}$ equal to zero. Group the terms by eigenvalue: for each i, let u_i be the sum of the terms in $\mathcal{L}$ coming from $E(\lambda_i, T)$. Then $u_i \in E(\lambda_i, T)$, and $u_1 + \cdots + u_m = 0$.

By the direct-sum property of the sum of eigenspaces, each $u_i = 0$. But u_i is a linear combination of our chosen basis of $E(\lambda_i, T)$, and a linearly independent list can only combine to zero trivially. Hence the coefficients of the original linear combination all vanish, and $\mathcal{L}$ is linearly independent. ∎

Corollary 5.5.1 (Distinct eigenvalues $\Rightarrow$ diagonalisable)**.** If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalisable.

Proof. Each eigenspace is nonzero (it contains an eigenvector), so $\dim E(\lambda_i, T) \geq 1$ for each of the $\dim V$ eigenvalues. Therefore $\sum \dim E(\lambda_i, T) \geq \dim V$. Combined with the inequality $\sum \dim E(\lambda_i, T) \leq \dim V$ of the previous corollary, we have equality, and condition (d) of the theorem holds. ∎

Remark. The converse of this corollary is false. The identity operator I on any vector

space is diagonal in every basis (and thus diagonalisable), but it has only one eigenvalue, $\lambda = 1$. More generally, diagonalisability allows repeated eigenvalues as long as each eigenspace is large enough.

6 Inner Product Spaces

Vector spaces as we have developed them so far have two operations: addition and scalar multiplication. These are enough to talk about linear combinations, bases, dimensions, linear maps, and eigenvalues, but they do not give us a way to measure length or angle. An inner product adds exactly that: it is a scalar-valued pairing of two vectors that generalises the dot product in $\mathbb{R}^n$, and from it we recover norms, orthogonality, and projections. Inner product spaces are the setting for the last third of the subject — in particular for the spectral theorem, which decomposes a well-behaved operator into mutually orthogonal scalings.

6.1 Inner products

In $\mathbb{R}^n$ we already have a natural length function,

$$\|(x_1, \ldots, x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2},$$

and a natural pairing, the dot product $\sum x_i y_i$. Length can be recovered from the dot product via $\|x\|^2 = x \cdot x$, so the dot product is the more fundamental object. The general definition of an inner product is designed to axiomatise the essential properties of the dot product in a way that survives the move to complex vector spaces and abstract vector spaces.

Definition 6.1 (Inner product on a real or complex vector space)**.** Let V be a vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. An inner product on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ satisfying, for all $u, v, w \in V$ and $\lambda \in \mathbb{F}$:

1. Positivity: $\langle v, v \rangle \geq 0$ (note that $\langle v, v \rangle$ lies in $\mathbb{F}$; when $\mathbb{F} = \mathbb{C}$, it is always in fact a real number by conjugate symmetry below).

2. Definiteness: $\langle v, v \rangle = 0$ if and only if $v = 0$.

3. Additivity in the first slot: $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$.

4. Homogeneity in the first slot: $\langle \lambda u, v\rangle = \lambda \langle u, v\rangle$.

5. Conjugate symmetry: $\langle u, v\rangle = \overline{\langle v, u\rangle}$.

A vector space equipped with an inner product is an inner product space.

Linearity in the first slot is part of the axioms, but linearity in the second slot is not. What does hold in the second slot is *conjugate-linearity*: additivity carries over, but scalars pull out with a complex conjugate,

$$\langle u, \lambda v\rangle = \overline{\lambda} \langle u, v\rangle .$$

This follows from combining homogeneity in the first slot with conjugate symmetry,

$$\langle u, \lambda v\rangle = \overline{\langle \lambda v, u\rangle} = \overline{\lambda \langle v, u\rangle} = \overline{\lambda}\, \overline{\langle v, u\rangle} = \overline{\lambda} \langle u, v\rangle .$$

When $\mathbb{F} = \mathbb{R}$, every complex conjugate is trivial and conjugate symmetry reduces to ordinary symmetry, making the inner product fully linear in each slot. The conjugate twist matters only in the complex case, but that is the case we most need for physics and applications.

Example 14

[Euclidean inner product on $\mathbb{F}^n$] The standard inner product on $\mathbb{F}^n$ is

$$\langle (u_1, \ldots, u_n), (v_1, \ldots, v_n)\rangle = u_1\overline{v_1} + u_2\overline{v_2} + \cdots + u_n\overline{v_n} = \sum_{i=1}^{n} u_i\overline{v_i}. \tag{6.1}$$

On $\mathbb{R}^n$ the conjugate is trivial and this reduces to the familiar dot product $\sum u_i v_i$.

Example 15

[L^2 inner product on polynomials] For $p, q \in \mathcal{P}(\mathbb{R})$ define

$$\langle p, q\rangle = \int_0^1 p(x)q(x)\, dx.$$

Positivity holds because $p(x)^2 \geq 0$ pointwise and the integral of a non-negative function is non-negative; definiteness follows from the continuity of p (if p is not identically zero it is nonzero on an interval, contributing a positive integral). Linearity and symmetry are immediate from properties of the integral.

6.1.1 Norm

The axioms of an inner product ensure $\langle v, v\rangle$ is a non-negative real, so we can take its square root to define a length function.

Definition 6.2 (Norm)**.** Let V be an inner product space. The norm of $v \in V$ is

$$\|v\| = \sqrt{\langle v, v\rangle}. \tag{6.2}$$

This is well-defined by positivity, and vanishes only when $v = 0$ by definiteness. The norm satisfies three basic properties, corresponding directly to the axioms of the inner product:

$$\|v\| \geq 0, \tag{6.3}$$

$$\|v\| = 0 \iff v = 0, \tag{6.4}$$

$$\|\lambda v\| = |\lambda| \, \|v\| \, . \tag{6.5}$$

The third property is a quick consequence of the homogeneity axioms. Over $\mathbb{C}$, the calculation reads

$$\|\lambda v\|^2 = \langle \lambda v, \lambda v\rangle = \lambda\overline{\lambda} \, \langle v, v\rangle = |\lambda|^2 \, \|v\|^2 \, ,$$

and taking square roots gives $\|\lambda v\| = |\lambda| \, \|v\|$, using the positivity of both sides to pin down the sign.

6.2 Orthogonality

The second angle-related notion provided by an inner product is perpendicularity.

Definition 6.3 (Orthogonal)**.** Two vectors $u, v \in V$ are orthogonal if $\langle u, v\rangle = 0$. We write $u \perp v$.

The zero vector is orthogonal to every vector, which is a convenient convention. Orthogonality has immediate geometric content when $\mathbb{F} = \mathbb{R}$: the inner product $\langle u, v\rangle = \|u\|\,\|v\|\cos\theta$ is zero exactly when $\cos\theta = 0$, which is when the two vectors are perpendicular in the ordinary geometric sense.

Proposition 6.1 (Pythagorean theorem)**.** If $u \perp v$, then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.

Proof. Expand the squared norm using bilinearity (with conjugation in the second slot),

$$\|u+v\|^2 = \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle\,.$$

The two cross-terms $\langle u, v\rangle$ and $\langle v, u\rangle$ are conjugates of each other; since $\langle u, v\rangle = 0$ by assumption, both vanish. What remains is $\langle u,u\rangle + \langle v,v\rangle = \|u\|^2 + \|v\|^2$. ∎

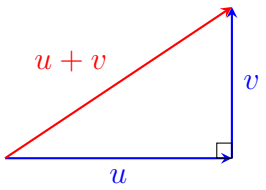

Figure 6: The Pythagorean theorem for orthogonal vectors.

The central inequality of inner product spaces is the Cauchy-Schwarz inequality. It bounds the size of $\langle u, v\rangle$ in terms of the norms of u and v, and it is the tool that converts many inner-product manipulations into genuine estimates.

Theorem 6.1 (Cauchy-Schwarz inequality)**.** For any u, v in an inner product space,

$$|\langle u, v\rangle| \le \|u\|\,\|v\|\,, \tag{6.6}$$

with equality if and only if one of u, v is a scalar multiple of the other.

Proof. If $v = 0$, both sides are zero and the inequality holds trivially (with equality, and v is a zero multiple of u). Assume $v \neq 0$.

The strategy is to split u into its component along v and a component orthogonal to v, and then apply the Pythagorean theorem. Set

$$p = \frac{\langle u, v\rangle}{\|v\|^2} v, \qquad w = u - p,$$

so that $u = p + w$. We claim $w \perp v$:

$$\langle w, v\rangle = \langle u - p, v\rangle = \langle u, v\rangle - \langle p, v\rangle = \langle u, v\rangle - \frac{\langle u, v\rangle}{\|v\|^2} \langle v, v\rangle = \langle u, v\rangle - \langle u, v\rangle = 0.$$

Since p is a scalar multiple of v and w is orthogonal to v, p and w are themselves orthogonal, and by the Pythagorean theorem,

$$\|u\|^2 = \|p + w\|^2 = \|p\|^2 + \|w\|^2 \geq \|p\|^2 .$$

Compute $\|p\|^2$ directly:

$$\|p\|^2 = \left| \frac{\langle u, v\rangle}{\|v\|^2} \right|^2 \|v\|^2 = \frac{|\langle u, v\rangle|^2}{\|v\|^2}.$$

Combining the two displayed equations,

$$\|u\|^2 \geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}.$$

Multiplying through by $\|v\|^2$ (which is positive) and taking square roots yields $\|u\| \, \|v\| \geq |\langle u, v\rangle|$, which is the inequality.

Equality holds in the Pythagorean step if and only if $w = 0$, which says $u = p$: precisely that u is a scalar multiple of v. ■

Cauchy-Schwarz immediately gives the triangle inequality for the norm.

Theorem 6.2 (Triangle inequality)**.** For any u, v in an inner product space,

$$\|u + v\| \leq \|u\| + \|v\| , \tag{6.7}$$

with equality if and only if one of u, v is a non-negative real multiple of the other.

Proof. Expand the squared norm as in the Pythagorean theorem, but now without the orthogonality assumption:

$$\|u+v\|^2 = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle = \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}\,\langle u,v\rangle\,,$$

using that $\langle u,v\rangle + \langle v,u\rangle = \langle u,v\rangle + \overline{\langle u,v\rangle} = 2\,\mathrm{Re}\,\langle u,v\rangle$.

The real part of any complex number is bounded by its absolute value, so $\mathrm{Re}\,\langle u,v\rangle \le |\langle u,v\rangle|$. Combined with Cauchy-Schwarz,

$$\|u+v\|^2 \le \|u\|^2 + \|v\|^2 + 2\,|\langle u,v\rangle| \le \|u\|^2 + \|v\|^2 + 2\,\|u\|\,\|v\| = (\|u\| + \|v\|)^2.$$

Taking square roots gives the triangle inequality. Equality in the first step requires $\langle u,v\rangle$ to be real and non-negative; equality in Cauchy-Schwarz requires u and v to be scalar multiples of each other; together, these say the scalar is non-negative real. ∎

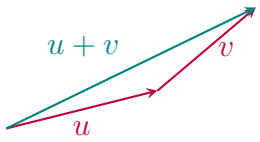

Figure 7: The triangle inequality: the direct path from the tail of u to the head of v (length $\|u+v\|$) is no longer than the bent path going through the tip of u (length $\|u\| + \|v\|$).

A final identity worth noting: the parallelogram law, which characterises norms that come from an inner product.

Proposition 6.2 (Parallelogram law)**.** For any u, v in an inner product space,

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2). \tag{6.8}$$

6.3 Orthonormal bases

Bases interact especially well with an inner product when the basis vectors are pairwise orthogonal and of unit length. Such a basis is called orthonormal, and many computations

simplify dramatically in such a basis.

Definition 6.4 (Orthonormal list)**.** A list $(e_1, \dots, e_m)$ in an inner product space V is orthonormal if

$$\langle e_i, e_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j, \end{cases} \tag{6.9}$$

that is, the vectors are pairwise orthogonal and each has norm 1.

Orthonormality is already a strong condition: it forces linear independence.

Proposition 6.3. Every orthonormal list is linearly independent.

Proof. Suppose $\sum_i a_i e_i = 0$. Take the inner product of both sides with e_j for an arbitrary index j. Using linearity in the first slot and the orthonormality relations,

$$0 = \left\langle \sum_i a_i e_i, e_j \right\rangle = \sum_i a_i \langle e_i, e_j \rangle = a_j.$$

Since j was arbitrary, every coefficient vanishes. ∎

Definition 6.5 (Orthonormal basis)**.** An orthonormal basis (abbreviated ONB) of an inner product space V is an orthonormal list that is also a basis.

The efficiency of an orthonormal basis shows up in the way coefficients in the basis expansion are computed. Instead of solving a linear system, they are read off directly by taking inner products.

Proposition 6.4 (Formula with respect to an ONB)**.** Let $(e_1, \dots, e_n)$ be an ONB of an inner product space V. For every $v \in V$,

$$v = \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \cdots + \langle v, e_n \rangle e_n, \tag{6.10}$$

and consequently

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 . \tag{6.11}$$

Proof. Since $(e_1, \ldots, e_n)$ is a basis, we can write $v = \sum_i a_i e_i$ for some scalars a_i. Taking the inner product with e_j and using orthonormality,

$$\langle v, e_j\rangle = \left\langle \sum_i a_i e_i, e_j \right\rangle = \sum_i a_i \langle e_i, e_j\rangle = a_j,$$

so $a_j = \langle v, e_j\rangle$, giving the claimed expansion.

For the norm identity, compute

$$\|v\|^2 = \langle v, v\rangle = \left\langle \sum_i a_i e_i, \sum_j a_j e_j \right\rangle = \sum_{i,j} a_i \overline{a_j} \delta_{ij} = \sum_i |a_i|^2 ,$$

using conjugate-linearity in the second slot. ∎

6.4 Gram-Schmidt procedure

Existence of orthonormal bases is not obvious — the standard basis of $\mathbb{F}^n$ is orthonormal, but in an abstract inner product space we are not handed such a gift. The Gram-Schmidt procedure constructs an orthonormal basis from any linearly independent list, by subtracting off successive projections and normalising.

Theorem 6.3 (Gram-Schmidt)**.** If $(v_1, \ldots, v_m)$ is a linearly independent list in an inner product space V, then there exists an orthonormal list $(e_1, \ldots, e_m)$ such that

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k) \qquad \text{for every } k = 1, \ldots, m. \tag{6.12}$$

Proof. We construct the e_i inductively.

Base case $k = 1$. Since the list is linearly independent, $v_1 \neq 0$, so $\|v_1\| > 0$ and we can set

$$e_1 = \frac{v_1}{\|v_1\|}.$$

Then $\|e_1\| = 1$, and $\operatorname{span}(v_1) = \operatorname{span}(e_1)$ because each is a nonzero scalar multiple of the

other.

Inductive step. Suppose $(e_1, \dots, e_{k-1})$ has been constructed and is orthonormal, with the same span as $(v_1, \dots, v_{k-1})$. Define the "orthogonalised" version of v_k by subtracting off its projections onto each e_i:

$$f_k = v_k - \langle v_k, e_1 \rangle e_1 - \langle v_k, e_2 \rangle e_2 - \cdots - \langle v_k, e_{k-1} \rangle e_{k-1}.$$

Then $f_k \neq 0$: if f_k were zero, v_k would be a linear combination of $e_1, \dots, e_{k-1}$, hence of $v_1, \dots, v_{k-1}$ (same span), contradicting linear independence.

We check $f_k \perp e_j$ for every $j < k$ by direct calculation. Using linearity in the first slot and orthonormality,

$$\langle f_k, e_j \rangle = \langle v_k, e_j \rangle - \sum_{i=1}^{k-1} \langle v_k, e_i \rangle \langle e_i, e_j \rangle = \langle v_k, e_j \rangle - \langle v_k, e_j \rangle = 0.$$

Finally, normalise: set $e_k = f_k / \|f_k\|$. Then $(e_1, \dots, e_k)$ is orthonormal. For the span equality, note that e_k is a linear combination of v_k and $e_1, \dots, e_{k-1}$, which are in $\operatorname{span}(v_1, \dots, v_k)$; conversely, v_k is a linear combination of f_k and the e_i, so of the e-list up to k. ■

Corollary 6.3.1. Every finite-dimensional inner product space has an orthonormal basis.

Proof. Take any basis of V and run Gram-Schmidt; the result is an orthonormal list of the same length as the basis, hence also a basis. ■

Corollary 6.3.2. Every orthonormal list in a finite-dimensional inner product space can be extended to an orthonormal basis.

Proof. An orthonormal list is linearly independent, so it can be extended to a basis. Running Gram-Schmidt on the extended list does nothing to the first vectors (they are already orthonormal, and the procedure leaves such vectors fixed), and produces an orthonormal basis that still contains the original orthonormal list. ■

6.5 Orthogonal complement and projection

The final major construction of this chapter is the orthogonal complement: given a subspace U, the set of vectors orthogonal to every vector in U is itself a subspace, and V splits as the direct sum of U and its complement.

Definition 6.6 (Orthogonal complement)**.** Let U be a subset of an inner product space V. The orthogonal complement of U is

$$U^{\perp} = \{v \in V \mid \langle v, u\rangle = 0 \text{ for all } u \in U\}. \tag{6.13}$$

Proposition 6.5. $U^{\perp}$ is a subspace of V.

The three conditions of the subspace test are immediate: the zero vector is orthogonal to everything; if $v, w \in U^{\perp}$ then $\langle v + w, u\rangle = \langle v, u\rangle + \langle w, u\rangle = 0$ for every $u \in U$; and scalar multiplication pulls out.

Theorem 6.4 (Orthogonal decomposition)**.** Let U be a subspace of a finite-dimensional inner product space V. Then

$$V = U \oplus U^{\perp}, \tag{6.14}$$

and consequently $\dim U^{\perp} = \dim V - \dim U$.

Proof. We first show $V = U + U^{\perp}$ by explicitly decomposing an arbitrary vector. Choose an orthonormal basis $(e_1, \ldots, e_m)$ of U (possible because U is itself a finite-dimensional inner product space). For any $v \in V$, define

$$u = \langle v, e_1\rangle\, e_1 + \cdots + \langle v, e_m\rangle\, e_m, \qquad w = v - u.$$

By construction $u \in U$ (it is a linear combination of the basis of U). We claim $w \in U^{\perp}$. For each basis vector e_j of U,

$$\langle w, e_j\rangle = \langle v, e_j\rangle - \langle u, e_j\rangle = \langle v, e_j\rangle - \langle v, e_j\rangle = 0,$$

where we used the ONB formula to evaluate $\langle u, e_j \rangle$. So w is orthogonal to each e_j, and by linearity w is orthogonal to every linear combination of them, i.e. every element of $U = \operatorname{span}(e_1, \ldots, e_m)$. Hence $w \in U^\perp$, and $v = u + w$ gives $V = U + U^\perp$.

For directness, suppose $v \in U \cap U^\perp$. Then $v \perp v$, so $\langle v, v \rangle = 0$, and by definiteness $v = 0$. By the two-subspace criterion, $U + U^\perp$ is a direct sum.

Dimension: $\dim V = \dim U + \dim U^\perp$, so $\dim U^\perp = \dim V - \dim U$. ■

Theorem 6.5. $(U^\perp)^\perp = U$ for any subspace U of a finite-dimensional inner product space.

Proof. We prove the two inclusions separately.

$U \subseteq (U^\perp)^\perp$. Let $u \in U$. To show $u \in (U^\perp)^\perp$, we must verify that u is orthogonal to every vector in $U^\perp$. Take any $w \in U^\perp$; by the definition of $U^\perp$, $\langle w, u \rangle = 0$. By conjugate symmetry, $\langle u, w \rangle = \overline{\langle w, u \rangle} = 0$. So u is orthogonal to every $w \in U^\perp$, meaning $u \in (U^\perp)^\perp$.

$(U^\perp)^\perp \subseteq U$. Let $v \in (U^\perp)^\perp$. By the orthogonal decomposition applied to U, we can write $v = u + w$ with $u \in U$ and $w \in U^\perp$. Our goal is to show $w = 0$, which will give $v = u \in U$.

Compute $\langle w, w \rangle$. Using $v = u + w$ and the linearity of the inner product,

$$\langle w, w \rangle = \langle v - u, w \rangle = \langle v, w \rangle - \langle u, w \rangle .$$

The first term $\langle v, w \rangle$ is zero because $v \in (U^\perp)^\perp$ and $w \in U^\perp$. The second term $\langle u, w \rangle$ is zero because $u \in U$ and $w \in U^\perp$. So $\langle w, w \rangle = 0$, which by definiteness forces $w = 0$. ■

The direct-sum decomposition $V = U \oplus U^\perp$ makes precise the idea of "the component of v in U" — there is a unique such component.

Definition 6.7 (Orthogonal projection)**.** Let U be a subspace of a finite-dimensional inner product space V. The orthogonal projection onto U, denoted $P_U \in \mathcal{L}(V)$, is the map defined as follows: write $v = u + w$ with $u \in U$ and $w \in U^\perp$ (which is possible and unique by the orthogonal decomposition), and set $P_U(v) = u$.

Proposition 6.6 (Properties of the orthogonal projection)**.** Let U be a subspace of a finite-dimensional inner product space V, and let P_U be the orthogonal projection onto U. Then

1. P_U is linear.
2. $\text{range}(P_U) = U$ and $\text{null}(P_U) = U^\perp$.
3. $P_U^2 = P_U$ (projection is idempotent).
4. $I - P_U = P_{U^\perp}$.
5. For every $v \in V$ and $u \in U$, $\|v - P_U v\| \leq \|v - u\|$. In other words, $P_U v$ is the point in U closest to v.

Proof of (5). Items (1)–(4) are straightforward verifications from the definition. For the closest-point property (5), write

$$v - u = (v - P_U v) + (P_U v - u).$$

The first summand $v - P_U v$ lies in $U^\perp$ by construction of the decomposition. The second summand $P_U v - u$ is the difference of two elements of U, hence itself in U. These two summands are therefore orthogonal, and by the Pythagorean theorem,

$$\|v - u\|^2 = \|v - P_U v\|^2 + \|P_U v - u\|^2 \geq \|v - P_U v\|^2 .$$

Taking square roots gives the desired inequality, with equality exactly when $P_U v - u = 0$, i.e. $u = P_U v$. ■

Property (5) is the fundamental theorem of least-squares approximation: among all vectors in U, the orthogonal projection is the one that minimises distance to v. This is used in practice in regression, data-fitting, and signal processing, where one seeks the best approximation of a given vector by a vector in a fixed subspace.

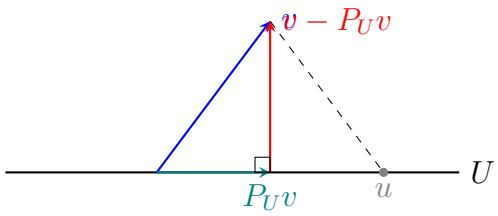

Figure 8: The projection $P_U v$ is the point of U closest to v: the dashed line to any other point $u \in U$ is longer than the perpendicular drop.

7 Operators on Inner Product Spaces

Once a vector space carries an inner product, linear maps acquire new kinds of structure that were invisible in the purely algebraic setting. Every linear map between inner product spaces has an *adjoint*, a companion map on the other side of the inner product. Operators that coincide with their adjoint, commute with it, or satisfy various related compatibility conditions turn out to have remarkably clean spectral descriptions: they can be diagonalised by orthonormal bases. The centrepiece of this chapter is the spectral theorem, which captures this phenomenon exactly.

7.1 The adjoint

Suppose we have a linear map $T : V \to W$ between inner product spaces. Given any fixed vector $w \in W$, we can combine T with w to obtain a map $V \to \mathbb{F}$ defined by $v \mapsto \langle T(v), w\rangle_W$. This map is linear in v, because both T and the inner product are linear in the first argument. A linear map from a finite-dimensional inner product space to the scalar field must take the form $v \mapsto \langle v, u\rangle$ for some unique u — this is the Riesz representation theorem, which is equivalent (in finite dimensions) to the ability to expand vectors in an orthonormal basis. The adjoint gathers these representative vectors as w varies over W.

Definition 7.1 (Adjoint)**.** Let V and W be finite-dimensional inner product spaces over $\mathbb{F}$, and let $T \in \mathcal{L}(V, W)$. The adjoint of T is the function $T^* : W \to V$ uniquely determined by the relation

$$\langle T(v), w\rangle_W = \langle v, T^*(w)\rangle_V \qquad \forall v \in V,\ w \in W. \tag{7.1}$$

The defining relation says: to move T from one side of the inner product to the other, we replace it by T^*. This is the direct analogue of integration by parts in analysis, where derivatives move across an integral by adjointness.

Example 16

Let $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ be given by $T(x_1, x_2, x_3) = (x_2 + 3x_3,\ 2x_1)$. Find $T^*(y_1, y_2)$.

Solution

We compute the left-hand side of the defining relation directly:

$$\langle T(x), y\rangle_{\mathbb{R}^2} = (x_2 + 3x_3)y_1 + 2x_1y_2.$$

Our goal is to rewrite this expression as an inner product in $\mathbb{R}^3$ of x with some vector depending on y. Collecting coefficients of x_1, x_2, x_3 yields

$$2x_1y_2 + x_2y_1 + 3x_3y_1 = \langle (x_1, x_2, x_3), (2y_2,\ y_1,\ 3y_1)\rangle_{\mathbb{R}^3}\,.$$

Reading off, $T^*(y_1, y_2) = (2y_2,\ y_1,\ 3y_1)$.

We should verify that the adjoint is itself a linear map, so that the notation $T^* \in \mathcal{L}(W, V)$ is legitimate.

Theorem 7.1 (The adjoint is linear)**.** If $T \in \mathcal{L}(V, W)$ then $T^* \in \mathcal{L}(W, V)$.

Proof. We must verify additivity and homogeneity of T^*.

Additivity. Fix $w_1, w_2 \in W$. For every $v \in V$,

$$\langle v, T^*(w_1 + w_2)\rangle = \langle T(v), w_1 + w_2\rangle = \langle T(v), w_1\rangle + \langle T(v), w_2\rangle = \langle v, T^*(w_1)\rangle + \langle v, T^*(w_2)\rangle\,,$$

where the first equality is the defining relation for T^* applied with the vector $w_1 + w_2$, the second is additivity of the inner product in the second slot, and the third is the defining relation applied with w_1 and w_2 separately. Using additivity in the second slot on the final expression,

$$\langle v, T^*(w_1 + w_2)\rangle = \langle v, T^*(w_1) + T^*(w_2)\rangle\,.$$

This holds for every v; the only vector with the same inner product against all v is itself (a standard consequence of definiteness), so $T^*(w_1 + w_2) = T^*(w_1) + T^*(w_2)$.

Homogeneity. For $\lambda \in \mathbb{F}$ and $w \in W$, using conjugate-linearity of the inner product in the second slot,

$$\langle v, T^*(\lambda w)\rangle = \langle T(v), \lambda w\rangle = \overline{\lambda}\, \langle T(v), w\rangle = \overline{\lambda}\, \langle v, T^*(w)\rangle = \langle v, \lambda T^*(w)\rangle\,.$$

Since this holds for every v, $T^*(\lambda w) = \lambda T^*(w)$. ■

Several algebraic properties follow from the defining relation essentially by inspection. We state them but omit the routine verifications, each of which consists of inserting the definition and juggling conjugates.

Proposition 7.1 (Algebra of adjoints)**.** For linear maps S, T between inner product spaces and scalars $\lambda \in \mathbb{F}$:

1. $(S+T)^* = S^* + T^*$.
2. $(\lambda T)^* = \overline{\lambda}\, T^*$.
3. $(T^*)^* = T$ (double adjoint is the identity).
4. $I^* = I$.
5. $(ST)^* = T^*\, S^*$ whenever the product makes sense — note the reversed order.

The reversal in (5) is the same reversal one sees in the inverse of a product, $(ST)^{-1} = T^{-1}S^{-1}$, and comes from moving the operators across the inner product one at a time.

7.1.1 Null space and range of the adjoint

The adjoint carries a great deal of information about the null space and range of the original map. The key identities below say that the null space and range of T and T^* are connected by orthogonal complements.

Theorem 7.2 (Null space and range of the adjoint)**.** For $T \in \mathcal{L}(V, W)$ with V and W

finite-dimensional,

$$\operatorname{null}(T^*) = \operatorname{range}(T)^\perp, \tag{7.2}$$

$$\operatorname{range}(T) = \operatorname{null}(T^*)^\perp, \tag{7.3}$$

$$\operatorname{null}(T) = \operatorname{range}(T^*)^\perp, \tag{7.4}$$

$$\operatorname{range}(T^*) = \operatorname{null}(T)^\perp. \tag{7.5}$$

Proof. It suffices to prove the first identity; the others follow from it by applying orthogonal complements (and using $(U^\perp)^\perp = U$ from the previous chapter) and by replacing T with T^*.

Fix $w \in W$. The following chain of equivalences unpacks the meaning of $w \in \operatorname{null}(T^*)$:

$$\begin{aligned}
w \in \operatorname{null}(T^*) &\iff T^*(w) = 0 \\
&\iff \langle v, T^*(w)\rangle = 0 \text{ for every } v \in V \text{ (definiteness)} \\
&\iff \langle T(v), w\rangle = 0 \text{ for every } v \in V \text{ (defining relation)} \\
&\iff w \perp T(v) \text{ for every } v \in V \\
&\iff w \in \operatorname{range}(T)^\perp.
\end{aligned}$$

The equivalence $T^*(w) = 0 \iff \langle v, T^*(w)\rangle = 0$ for every v uses that the zero vector is the only vector orthogonal to everything. ∎

7.2 The conjugate transpose

The matrix of the adjoint is particularly clean when both domain and codomain are equipped with orthonormal bases.

Definition 7.2 (Conjugate transpose)**.** The conjugate transpose of a matrix $A \in \mathbb{F}^{m\times n}$ is the matrix $A^* \in \mathbb{F}^{n\times m}$ obtained by transposing and taking the complex conjugate of every entry,

$$(A^*)_{j,k} = \overline{A_{k,j}}. \tag{7.6}$$

Example 17

$$A = \begin{pmatrix} 2 & 3 & 7+i \\ 2i & 8 & 10 \end{pmatrix} \implies A^* = \begin{pmatrix} 2 & -2i \\ 3 & 8 \\ 7-i & 10 \end{pmatrix}.$$

Over $\mathbb{R}$, the conjugate is trivial and the conjugate transpose reduces to the ordinary transpose A^T.

Theorem 7.3 (Matrix of the adjoint)**.** Let $T \in \mathcal{L}(V, W)$, and let $(e_1, \ldots, e_n)$ and $(f_1, \ldots, f_m)$ be orthonormal bases of V and W respectively. Then

$$\mathcal{M}(T^*, (f_j), (e_i)) = \mathcal{M}(T, (e_i), (f_j))^*. \tag{7.7}$$

Proof. Write $A = \mathcal{M}(T, (e_i), (f_j))$ and $B = \mathcal{M}(T^*, (f_j), (e_i))$. We need to show $B_{k,j} = \overline{A_{j,k}}$ for all j, k.

The ith column of A contains the coefficients of $T(e_k)$ when expanded in the basis (f_j). Because the basis is orthonormal, these coefficients are recovered by taking inner products with the f_j:

$$A_{j,k} = \langle T(e_k), f_j \rangle_W .$$

By the same reasoning applied to T^*,

$$B_{k,j} = \langle T^*(f_j), e_k \rangle_V .$$

Applying conjugate symmetry to the second expression and then the defining relation of the adjoint,

$$B_{k,j} = \overline{\langle e_k, T^*(f_j) \rangle_V} = \overline{\langle T(e_k), f_j \rangle_W} = \overline{A_{j,k}}.$$

So $B = A^*$. ∎

Remark. Orthonormality of both bases is essential. For non-orthonormal bases, the matrix of the adjoint is not simply the conjugate transpose of the matrix of T; the off-diagonal

inner products of the basis vectors get mixed in.

7.3 Self-adjoint and normal operators

We now specialise to operators $T \in \mathcal{L}(V)$ acting on a single inner product space V. Here T and T^* both live in $\mathcal{L}(V)$, and it becomes meaningful to compare them directly.

Definition 7.3 (Self-adjoint)**.** An operator $T \in \mathcal{L}(V)$ is self-adjoint (also called Hermitian) if $T = T^*$.

Definition 7.4 (Normal)**.** An operator $T \in \mathcal{L}(V)$ is normal if $TT^* = T^*T$, that is, if T commutes with its adjoint.

Every self-adjoint operator is automatically normal, since $TT^* = TT = TT = T^*T$ when $T^* = T$. Normal operators form a strictly larger class that is nonetheless well-behaved enough to support a spectral theorem over $\mathbb{C}$.

Proposition 7.2 (Eigenvalues of self-adjoint operators are real)**.** If $T \in \mathcal{L}(V)$ is self-adjoint and λ is an eigenvalue of T, then $\lambda \in \mathbb{R}$.

Proof. Let $v \neq 0$ be an eigenvector with $T(v) = \lambda v$. Consider the scalar $\langle T(v), v\rangle$. On the one hand,

$$\langle T(v), v\rangle = \langle \lambda v, v\rangle = \lambda \|v\|^2 .$$

On the other hand, using self-adjointness and then the defining relation,

$$\langle T(v), v\rangle = \langle v, T^*(v)\rangle = \langle v, T(v)\rangle = \langle v, \lambda v\rangle = \overline{\lambda} \|v\|^2 ,$$

where conjugate-linearity in the second slot produced the $\overline{\lambda}$. Comparing the two expressions,

$$\lambda \|v\|^2 = \overline{\lambda} \|v\|^2 .$$

Since $v \neq 0$, we have $\|v\|^2 > 0$, so we can divide to get $\lambda = \overline{\lambda}$. A scalar equal to its own complex conjugate is real, so $\lambda \in \mathbb{R}$. ∎

The next technical result is needed for the proof of the complex spectral theorem. Over $\mathbb{C}$, an operator is determined by the inner products $\langle T(v), v\rangle$ as v ranges over V.

Proposition 7.3 (Self-adjoint operators with zero numerical range vanish)**.** Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$ satisfies $\langle T(v), v\rangle = 0$ for every $v \in V$. Then $T = 0$.

Proof. We will use the identity $\langle T(v), v\rangle = 0$ for several different choices of v and combine the resulting equations to extract $\langle T(u), w\rangle$ for arbitrary u, w.

Expand $\langle T(u+w), u+w\rangle$ by linearity:

$$\langle T(u+w), u+w\rangle = \langle T(u), u\rangle + \langle T(u), w\rangle + \langle T(w), u\rangle + \langle T(w), w\rangle .$$

By hypothesis, the left side is zero, and so are the two end terms on the right. Hence

$$\langle T(u), w\rangle + \langle T(w), u\rangle = 0. \tag{$\star$}$$

Now apply the hypothesis to $u + iw$ and expand similarly. Using conjugate-linearity in the second slot,

$$0 = \langle T(u+iw), u+iw\rangle = \langle T(u), u\rangle + \bar{i}\,\langle T(u), w\rangle + i\,\langle T(w), u\rangle + |i|^2\,\langle T(w), w\rangle ,$$

which simplifies to

$$-i\,\langle T(u), w\rangle + i\,\langle T(w), u\rangle = 0. \tag{$\star\star$}$$

Divide $(\star\star)$ by i, giving $-\langle T(u), w\rangle + \langle T(w), u\rangle = 0$, then add to $(\star)$. The $\langle T(w), u\rangle$ terms sum, the $\langle T(u), w\rangle$ terms cancel in a convenient way, and rearranging gives $\langle T(u), w\rangle = 0$ for every u, w. Fixing u and letting w range, the vector $T(u)$ is orthogonal to every w, so $T(u) = 0$. Since this holds for every u, $T = 0$. ∎

Remark. The complex structure of $\mathbb{C}$ is essential. Over $\mathbb{R}$ the claim fails: the $90°$ rotation $T \in \mathcal{L}(\mathbb{R}^2)$ with $T(x, y) = (-y, x)$ satisfies $\langle T(v), v\rangle = 0$ for every v (rotation by $90°$ sends every vector to an orthogonal one), but is certainly not the zero operator.

Normal operators have a clean geometric characterisation: they are precisely those for which T and T^* produce vectors of the same length.

Theorem 7.4 (Characterization of normal operators)**.** An operator $T \in \mathcal{L}(V)$ is normal if and only if $\|T(v)\| = \|T^*(v)\|$ for every $v \in V$.

Proof. The key observation is that $T^*T - TT^*$ is self-adjoint: its adjoint is $(T^*T)^* - (TT^*)^* = T^*T - TT^*$, using properties of the adjoint. Therefore, over $\mathbb{C}$, this operator is zero if and only if $\langle (T^*T - TT^*)v, v\rangle = 0$ for every v (by the previous proposition). Over $\mathbb{R}$ a parallel statement holds for self-adjoint operators.

We now run the equivalences:

$$\begin{aligned} T \text{ normal} &\iff T^*T - TT^* = 0 \\ &\iff \langle (T^*T - TT^*)v, v\rangle = 0 \text{ for every } v \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \text{ for every } v \\ &\iff \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle \text{ for every } v \\ &\iff \|Tv\|^2 = \|T^*v\|^2 \text{ for every } v. \end{aligned}$$

In the fourth step we moved T^* across the inner product on the left and T on the right, using the defining relation for the adjoint. Taking square roots on the last line completes the equivalence. ∎

Corollary 7.4.1 (Eigenvectors of T and T^* coincide for normal operators)**.** If $T \in \mathcal{L}(V)$ is normal and $T(v) = \lambda v$, then $T^*(v) = \overline{\lambda}\, v$.

Proof. Consider the operator $T - \lambda I$. Its adjoint is $T^* - \overline{\lambda}\, I$, and a direct check shows it commutes with $T - \lambda I$ whenever T does with T^*; hence $T - \lambda I$ is itself normal. By the characterisation,

$$\|(T - \lambda I)(v)\| = \left\|(T^* - \overline{\lambda}\, I)(v)\right\|.$$

The left side is $\|T(v) - \lambda v\| = 0$ by assumption, so the right side is zero too, giving $T^*(v) = \overline{\lambda}\, v$. ∎

Corollary 7.4.2 (Orthogonality of eigenvectors for normal operators)**.** If $T \in \mathcal{L}(V)$ is normal and v_1, v_2 are eigenvectors with distinct eigenvalues $\lambda_1 \neq \lambda_2$, then $\langle v_1, v_2 \rangle = 0$.

Proof. We compute $\langle T(v_1), v_2 \rangle$ in two ways. Starting from $T(v_1) = \lambda_1 v_1$,

$$\langle T(v_1), v_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \lambda_1 \langle v_1, v_2 \rangle .$$

Using the defining relation of the adjoint and the previous corollary, $T^*(v_2) = \overline{\lambda_2} v_2$, so

$$\langle T(v_1), v_2 \rangle = \langle v_1, T^*(v_2) \rangle = \left\langle v_1, \overline{\lambda_2} v_2 \right\rangle = \lambda_2 \langle v_1, v_2 \rangle ,$$

where the conjugate on $\overline{\lambda_2}$ pulled back out as λ_2 under conjugate-linearity in the second slot. Comparing,

$$\lambda_1 \langle v_1, v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle \implies (\lambda_1 - \lambda_2) \langle v_1, v_2 \rangle = 0.$$

Since $\lambda_1 \neq \lambda_2$, the factor $\lambda_1 - \lambda_2$ is nonzero, forcing $\langle v_1, v_2 \rangle = 0$. ■

7.4 The spectral theorems

The spectral theorems are the climax of the theory developed so far. They say that, under a modest structural hypothesis, an operator can be diagonalised by an *orthonormal* basis of eigenvectors — a much stronger condition than mere diagonalisability, because the eigenvectors are not only independent but mutually perpendicular.

Theorem 7.5 (Complex spectral theorem)**.** Let V be a finite-dimensional complex inner product space and $T \in \mathcal{L}(V)$. The following are equivalent:

1. T is normal.
2. V has an orthonormal basis consisting of eigenvectors of T.
3. T has a diagonal matrix representation with respect to some orthonormal basis of V.

Proof. We prove the implications (2) $\iff$ (3), (2) $\Rightarrow$ (1), and (1) $\Rightarrow$ (3), which suffice for the three-way equivalence.

(2) $\iff$ (3). An orthonormal basis of eigenvectors diagonalises T: if $T(e_k) = \lambda_k e_k$ then $\mathcal{M}(T)$ has the λ_k on the diagonal and zeros elsewhere. Conversely, a diagonal matrix in an orthonormal basis says $T(e_k)$ is a scalar multiple of e_k for every k.

(2) $\Rightarrow$ (1). If $\mathcal{M}(T)$ is diagonal in an orthonormal basis, then by the theorem on matrices of adjoints $\mathcal{M}(T^*) = \overline{\mathcal{M}(T)}$, which is also diagonal. Diagonal matrices commute with each other because each pair of diagonal entries multiplies componentwise in either order, so $TT^* = T^*T$, i.e. T is normal.

(1) $\Rightarrow$ (3). We invoke Schur's theorem on V: there exists an orthonormal basis (obtained by applying Gram-Schmidt to the basis supplied by Schur) in which $\mathcal{M}(T)$ is upper-triangular. Let the ONB be $(e_1, \dots, e_n)$ and $A = \mathcal{M}(T)$. We claim that normality forces A to in fact be diagonal.

Compute the norm of $T(e_1)$. Since $T(e_1) = A_{1,1}e_1 + A_{2,1}e_2 + \cdots + A_{n,1}e_n$ and upper-triangularity sets $A_{j,1} = 0$ for $j > 1$, only the first term survives: $T(e_1) = A_{1,1}e_1$, so

$$\|Te_1\|^2 = |A_{1,1}|^2 .$$

On the other hand, $T^*(e_1)$ expanded in the ONB has coefficients $(T^*(e_1))_k = \langle T^*(e_1), e_k\rangle = \overline{\langle e_1, T(e_k)\rangle} = \overline{\sum_i A_{i,k} \langle e_1, e_i\rangle} = \overline{A_{1,k}}$. Hence

$$\|T^*e_1\|^2 = \sum_{k=1}^{n} |A_{1,k}|^2 = |A_{1,1}|^2 + |A_{1,2}|^2 + \cdots + |A_{1,n}|^2 .$$

Normality says these two norms are equal, which forces $A_{1,j} = 0$ for every $j > 1$: the first row of A, above the diagonal, is entirely zero.

Repeating this argument with e_2 (using upper-triangularity of A together with the information just obtained about the first row) shows the second row is zero off the diagonal, and so on. By induction all off-diagonal entries vanish, and A is diagonal. ∎

Over $\mathbb{R}$, normality alone is not enough: as already noted, rotations are normal but have no real eigenvalues. The right hypothesis is self-adjointness.

Theorem 7.6 (Real spectral theorem)**.** Let V be a finite-dimensional real inner product space and $T \in \mathcal{L}(V)$. The following are equivalent:

1. T is self-adjoint.
2. V has an orthonormal basis consisting of eigenvectors of T.
3. T has a diagonal matrix representation with respect to some orthonormal basis of V.

Sketch. The implications (2) $\iff$ (3) and (2) $\Rightarrow$ (1) go through as in the complex case: a diagonal real matrix equals its own transpose, hence its own adjoint.

The hard direction is (1) $\Rightarrow$ (2). The key lemma is that a self-adjoint operator on a nonzero real inner product space always has a (real) eigenvalue. Over $\mathbb{C}$ this was automatic from the fundamental theorem of algebra; over $\mathbb{R}$ one has to do more work, essentially controlling the discriminants of the quadratic factors that appear when one factors the characteristic polynomial over $\mathbb{R}$.

Once eigenvalues exist, the argument runs by induction on dimension. Pick an eigenvector e_1 with unit length. The subspace $\operatorname{span}(e_1)^{\perp}$ is T-invariant (this is where self-adjointness is crucial: if $v \perp e_1$ then $\langle T(v), e_1\rangle = \langle v, T(e_1)\rangle = \langle v, \lambda e_1\rangle = 0$). Restrict T to $\operatorname{span}(e_1)^{\perp}$, which is still self-adjoint, and apply the inductive hypothesis. ∎

7.5 Positive operators and isometries

Two special classes of operators deserve to be named. Positive operators generalise the idea of a non-negative scalar; isometries generalise the idea of a rigid motion.

Definition 7.5 (Positive operator)**.** An operator $T \in \mathcal{L}(V)$ is positive (also called positive semi-definite) if T is self-adjoint and $\langle T(v), v\rangle \geq 0$ for every $v \in V$.

The requirement that $\langle T(v), v\rangle$ is always a non-negative real number automatically forces $\langle T(v), v\rangle$ to be real, which over $\mathbb{C}$ is a genuine constraint. Positive operators admit several equivalent characterisations.

Theorem 7.7 (Characterisations of positive operators)**.** For $T \in \mathcal{L}(V)$ with V a finite-dimensional inner product space, the following are equivalent:

1. T is positive.
2. T is self-adjoint and all eigenvalues of T are non-negative.
3. T has a positive square root, i.e. an operator $R \in \mathcal{L}(V)$ with $R^2 = T$ and R positive.
4. $T = S^*S$ for some $S \in \mathcal{L}(V)$.

The equivalence follows by combining the spectral theorem (so that T can be diagonalised in an ONB) with routine manipulations of the scalar eigenvalues. A full proof runs several paragraphs and is left as an exercise.

Definition 7.6 (Isometry)**.** An operator $S \in \mathcal{L}(V)$ is an isometry if $\|S(v)\| = \|v\|$ for every $v \in V$. Over $\mathbb{R}$, isometries are called orthogonal operators; over $\mathbb{C}$, unitary operators.

Isometries are precisely the linear maps that preserve lengths, which is the defining feature of a rigid motion through the origin: rotations and reflections in $\mathbb{R}^n$, and their complex analogues.

Theorem 7.8 (Characterisations of isometries)**.** For $S \in \mathcal{L}(V)$ with V finite-dimensional, the following are equivalent:

(a) S is an isometry.

(b) S preserves inner products: $\langle Su, Sv\rangle = \langle u, v\rangle$ for every $u, v \in V$.

(c) $S^*S = I$.

(d) S sends every orthonormal list to an orthonormal list.

(e) S sends some orthonormal basis to an orthonormal basis.

Sketch of selected implications. (c) $\Rightarrow$ (d): If (e_i) is orthonormal, $\langle Se_i, Se_j\rangle = \langle e_i, S^*Se_j\rangle = \langle e_i, e_j\rangle = \delta_{ij}$.

(d) $\Rightarrow$ (a): Any $v = \sum a_i e_i$ satisfies $Sv = \sum a_i Se_i$. Since the Se_i are orthonormal, $\|Sv\|^2 = \sum |a_i|^2 = \|v\|^2$. ■

The condition $S^*S = I$ is especially useful: it says in matrix terms that the columns of the matrix of S (in an ONB) are orthonormal — a condition easily checked by hand.

7.6 Singular values and the singular value decomposition

The spectral theorem requires the operator to live on a single inner product space and satisfy a commutation condition with its adjoint. A natural question is whether anything of the sort can be said for a *general* linear map $T \in \mathcal{L}(V, W)$ between two different inner product spaces. The answer is yes, and it involves the operator T^*T, which is always self-adjoint and positive regardless of what T is.

Definition 7.7 (Singular values)**.** Let $T \in \mathcal{L}(V, W)$ with V, W finite-dimensional inner product spaces. The singular values of T are the non-negative square roots of the eigenvalues of the positive operator T^*T, listed in decreasing order, each repeated according to the dimension of its eigenspace in T^*T.

The operator T^*T is well-behaved: it is self-adjoint since $(T^*T)^* = T^*(T^*)^* = T^*T$, and it is positive since $\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2 \geq 0$. The spectral theorem applies, so T^*T has a real, non-negative eigenvalue spectrum, and the singular values of T are well-defined.

Theorem 7.9 (Characterisation by singular values)**.** For $T \in \mathcal{L}(V, W)$ with V, W finite-dimensional:

(a) T is injective if and only if 0 is not a singular value of T.

(b) The number of positive singular values of T, counted with multiplicity, equals $\dim \operatorname{range}(T)$.

(c) T is surjective if and only if the number of positive singular values of T equals $\dim W$.

(d) T is an isometry if and only if all singular values of T equal 1.

Proof of (b). The positive operator T^*T can be diagonalised in an orthonormal basis by the spectral theorem. The number of positive eigenvalues of T^*T counted with multiplicity equals the dimension of $\operatorname{range}(T^*T)$, since the kernel of T^*T is the sum of eigenspaces for eigenvalue 0 and the rest of V splits off orthogonally. Now $\operatorname{range}(T^*T) \subseteq \operatorname{range}(T^*)$ trivially, and $\dim \operatorname{range}(T^*) = \dim \operatorname{range}(T)$ follows from general rank identities (they are equal because $\dim V = \dim \operatorname{null}(T) + \dim \operatorname{range}(T) = \dim \operatorname{null}(T^*T) + \dim \operatorname{range}(T^*T)$ and $\operatorname{null}(T) = \operatorname{null}(T^*T)$ by the identity $\langle T^*Tv, v\rangle = \|Tv\|^2$). ∎

Remark. If $T \in \mathcal{L}(V)$ is self-adjoint over $\mathbb{C}$, its eigenvalues $\lambda_1, \ldots, \lambda_n$ are real, and one checks directly that the singular values of T are $|\lambda_1|, \ldots, |\lambda_n|$ arranged in decreasing order. So singular values generalise the notion of eigenvalue magnitude to arbitrary linear maps.

The singular value decomposition writes T explicitly in terms of its singular values and two associated orthonormal lists.

Theorem 7.10 (Singular value decomposition)**.** Let $T \in \mathcal{L}(V, W)$ have positive singular values $s_1, \ldots, s_r$ (so $r = \operatorname{rank}(T)$). Then there exist orthonormal lists $(e_1, \ldots, e_r)$ in V and $(f_1, \ldots, f_r)$ in W such that

$$T(v) = s_1 \langle v, e_1\rangle f_1 + s_2 \langle v, e_2\rangle f_2 + \cdots + s_r \langle v, e_r\rangle f_r \qquad \forall v \in V. \tag{7.8}$$

Geometrically, the SVD decomposes the action of T into three stages: first, measure the input against the right singular vectors e_i; second, scale each measurement by the corresponding singular value s_i; third, reassemble the result in the direction of the left singular vector f_i. This is the foundation of a great deal of numerical linear algebra, dimension reduction, and low-rank approximation in applied mathematics.

8 Generalized Eigenvectors and Jordan Form

The spectral theorems of the previous chapter were beautiful but rested on hypotheses (normality, self-adjointness) that are strong. For general operators, even over $\mathbb{C}$, diagonalisation may fail: not every operator has enough eigenvectors to form a basis. In this chapter we develop tools to describe such operators. The key idea is to replace the rigid condition $T(v) = \lambda v$ with the softer condition $(T - \lambda I)^j v = 0$ for some j, which captures vectors that are sent to zero by $T - \lambda I$ after a few applications, rather than in one step. This gives the notion of a generalised eigenvector, and with it a decomposition of the space that works in every case. The culmination is the Jordan form, a canonical near-diagonal matrix representation available for any operator on a finite-dimensional complex vector space.

8.1 Generalised eigenvectors and generalised eigenspaces

Definition 8.1 (Generalised eigenvector)**.** Let $T \in \mathcal{L}(V)$ and let λ be an eigenvalue of T. A nonzero vector $u \in V$ is a generalised eigenvector of T corresponding to λ if

$$(T - \lambda I)^j u = 0 \quad \text{for some positive integer } j. \tag{8.1}$$

Equivalently, u lies in $\operatorname{null}((T - \lambda I)^j)$ for some $j \geq 1$. Every ordinary eigenvector is a generalised eigenvector with $j = 1$, so the notion is a genuine generalisation. The idea is that $T - \lambda I$ might not kill u immediately, but repeated application eventually drives it to zero; the smallest such j measures how far u is from being a true eigenvector.

Definition 8.2 (Generalised eigenspace)**.** The generalised eigenspace of T corresponding to λ is

$$G(\lambda, T) = \left\{ u \in V \,\middle|\, (T - \lambda I)^j u = 0 \text{ for some positive integer } j \right\}. \tag{8.2}$$

By inspection, $E(\lambda, T) \subseteq G(\lambda, T)$: ordinary eigenvectors are generalised ones. Both inclusions are strict in general; operators for which $E(\lambda, T) = G(\lambda, T)$ for every λ are

exactly the diagonalisable ones, as we will see.

Proposition 8.1. $G(\lambda, T)$ is a subspace of V, and it is invariant under T.

Proof. We apply the subspace test.

Zero vector. Any power of $T - \lambda I$ sends 0 to 0, so $0 \in G(\lambda, T)$.

Closure under addition. Take $u_1, u_2 \in G(\lambda, T)$ with $(T - \lambda I)^{j_1} u_1 = 0$ and $(T - \lambda I)^{j_2} u_2 = 0$. Set $j = \max(j_1, j_2)$. Applying $(T - \lambda I)^j$ to $u_1 + u_2$,

$$(T - \lambda I)^j (u_1 + u_2) = (T - \lambda I)^j u_1 + (T - \lambda I)^j u_2 = 0 + 0 = 0,$$

where the first term vanishes because $j \geq j_1$ and each additional power of $T - \lambda I$ applied to zero remains zero; similarly for the second. So $u_1 + u_2 \in G(\lambda, T)$.

Closure under scalar multiplication. If $(T - \lambda I)^j u = 0$ and $c \in \mathbb{F}$, then $(T - \lambda I)^j (cu) = c(T - \lambda I)^j u = 0$.

T-invariance. The operators T and $T - \lambda I$ commute, since $T(T - \lambda I) = T^2 - \lambda T = (T - \lambda I)T$. Hence T also commutes with every power of $T - \lambda I$. If $(T - \lambda I)^j u = 0$, then

$$(T - \lambda I)^j (Tu) = T((T - \lambda I)^j u) = T(0) = 0,$$

so $Tu \in G(\lambda, T)$. ∎

In the definition of $G(\lambda, T)$, the exponent j is allowed to depend on u. One might worry that we have to test infinitely many j's to decide membership. The next proposition shows this is unnecessary: the null spaces stabilise, and a single fixed exponent $j = \dim V$ suffices.

Proposition 8.2 (Stabilisation of null spaces)**.** Let $T \in \mathcal{L}(V)$ with $\dim V = n$. The null spaces of successive powers of T form an increasing chain,

$$\{0\} = \operatorname{null}(T^0) \subseteq \operatorname{null}(T^1) \subseteq \operatorname{null}(T^2) \subseteq \cdots, \tag{8.3}$$

and this chain stabilises: if $\operatorname{null}(T^m) = \operatorname{null}(T^{m+1})$ for some m, then equality holds for every $k \geq m$. In particular $\operatorname{null}(T^n) = \operatorname{null}(T^{n+k})$ for every $k \geq 0$.

Proof. Inclusion. If $T^k u = 0$, then $T^{k+1}u = T(T^k u) = T(0) = 0$, so $\operatorname{null}(T^k) \subseteq \operatorname{null}(T^{k+1})$.

Stabilisation. Suppose $\operatorname{null}(T^m) = \operatorname{null}(T^{m+1})$. We must show that for every $k \geq 0$, $\operatorname{null}(T^{m+k}) = \operatorname{null}(T^{m+k+1})$, whence by induction $\operatorname{null}(T^m) = \operatorname{null}(T^{m+k})$ for all k. The direction $\subseteq$ is automatic. For the reverse, suppose $T^{m+k+1}u = 0$, i.e. $T^{m+1}(T^k u) = 0$. Then $T^k u \in \operatorname{null}(T^{m+1}) = \operatorname{null}(T^m)$, so $T^m(T^k u) = 0$, i.e. $T^{m+k}u = 0$.

Bound on stabilisation index. The null spaces are subspaces of V, and the chain is increasing. As long as the chain is strictly increasing, each step adds at least one dimension, so after at most n steps we must have $\operatorname{null}(T^m) = \operatorname{null}(T^{m+1})$. From then on the chain is constant, so in particular $\operatorname{null}(T^n) = \operatorname{null}(T^{n+k})$ for every $k \geq 0$. ■

Applying this to $T - \lambda I$ in place of T gives the promised simplification.

Proposition 8.3. For every eigenvalue λ of $T \in \mathcal{L}(V)$ with $\dim V = n$,

$$G(\lambda, T) = \operatorname{null}\left((T - \lambda I)^n\right). \tag{8.4}$$

Proof. The inclusion $\operatorname{null}((T - \lambda I)^n) \subseteq G(\lambda, T)$ is immediate from the definition, taking $j = n$. For the reverse, suppose $u \in G(\lambda, T)$, so $(T - \lambda I)^j u = 0$ for some $j \geq 1$. If $j \leq n$, then by the chain inclusions $u \in \operatorname{null}((T-\lambda I)^n)$ already. If $j > n$, then by the stabilisation result applied to $T - \lambda I$, $\operatorname{null}((T - \lambda I)^n) = \operatorname{null}((T - \lambda I)^j)$, so $u \in \operatorname{null}((T - \lambda I)^n)$. ■

Analogous to the classical result that eigenvectors with distinct eigenvalues are linearly independent, generalised eigenvectors with distinct eigenvalues are also linearly independent. The proof follows the same strategy, applying a carefully chosen polynomial in T to extract vanishing conditions on the coefficients.

Theorem 8.1 (Linear independence of generalised eigenvectors)**.** Let $T \in \mathcal{L}(V)$ and let $u_1, \ldots, u_m$ be generalised eigenvectors of T corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$. Then $(u_1, \ldots, u_m)$ is linearly independent.

8.2 Nilpotent operators

The restriction of T to $G(\lambda, T)$ is going to be, essentially, a scalar λ plus a nilpotent perturbation. Nilpotent operators are thus the primitive building blocks of the Jordan form, and deserve their own treatment.

Definition 8.3 (Nilpotent)**.** An operator $N \in \mathcal{L}(V)$ is nilpotent if $N^k = 0$ for some positive integer k. The smallest such k is called the index of nilpotency of N.

The only eigenvalue of a nilpotent operator is 0, since if $N^k = 0$ and $Nv = \lambda v$ with $v \neq 0$, then $0 = N^k v = \lambda^k v$ forces $\lambda^k = 0$ and hence $\lambda = 0$. Equivalently, $G(0, N) = V$: every vector of V is eventually killed by N.

Proposition 8.4. If $N \in \mathcal{L}(V)$ is nilpotent, then $N^{\dim V} = 0$.

Proof. Since $G(0, N) = V$, by the previous section $V = \operatorname{null}(N^{\dim V})$, which means $N^{\dim V}$ vanishes on all of V, i.e. $N^{\dim V} = 0$. ∎

Nilpotent operators admit a strictly upper-triangular matrix representation — zeros on and below the diagonal — in some basis. The cleanest version of this fact identifies an especially organised basis, the so-called Jordan basis for N, in which the matrix consists of nilpotent blocks.

Theorem 8.2 (Matrix of a nilpotent operator)**.** If $N \in \mathcal{L}(V)$ is nilpotent, there is a basis of V with respect to which the matrix of N is strictly upper-triangular (zeros on and below the main diagonal).

Theorem 8.3 (Nilpotent block basis)**.** If $N \in \mathcal{L}(V)$ is nilpotent, there exist vectors $v_1, \ldots, v_r \in V$ and non-negative integers $m_1, \ldots, m_r$ such that

$$(N^{m_1}v_1,\ N^{m_1-1}v_1,\ \ldots,\ Nv_1,\ v_1,\quad N^{m_2}v_2,\ \ldots,\ v_2,\quad \ldots,\quad N^{m_r}v_r,\ \ldots,\ v_r) \tag{8.5}$$

is a basis of V. With respect to this basis the matrix of N is block-diagonal, with each

block of the form

$$\begin{pmatrix} 0 & 1 & & 0 \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}, \tag{8.6}$$

that is, zero on the diagonal with 1's on the superdiagonal.

The listing of vectors in each group is deliberate: we place the most-iterated vector $N^{m_i}v_i$ first (this is a genuine eigenvector, since N annihilates it), then the next one back $N^{m_i-1}v_i$, and so on down to v_i itself. Applying N to any of these basis vectors shifts it one slot to the right in the group or sends it to zero if it was the first — which is exactly what the 0/1 pattern of the matrix block encodes.

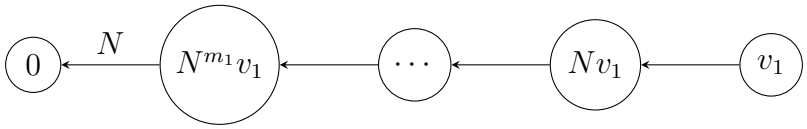

Figure 9: Action of N on one block of the Jordan basis for a nilpotent operator: the rightmost vector v_1 is shifted leftward by N, and eventually killed.

8.3 The generalised eigenspace decomposition

Over $\mathbb{C}$, the central structural result is that the whole space V splits as a direct sum of the generalised eigenspaces of T, and on each summand $T - \lambda I$ is nilpotent. This is the correct generalisation of diagonalisability to arbitrary operators.

Theorem 8.4 (Generalised eigenspace decomposition)**.** Let V be a finite-dimensional complex vector space, let $T \in \mathcal{L}(V)$, and let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then

$$V = G(\lambda_1, T) \oplus G(\lambda_2, T) \oplus \cdots \oplus G(\lambda_m, T), \tag{8.7}$$

and the restriction $(T - \lambda_i I)\big|_{G(\lambda_i, T)}$ is nilpotent.

Sketch. The proof is by induction on $\dim V$, leveraging two facts: first, every operator on a finite-dimensional nonzero complex vector space has an eigenvalue (this is where we

use $\mathbb{C}$); second, for such an eigenvalue λ_1, the space splits as

$$V = G(\lambda_1, T) \oplus \operatorname{range}\left((T - \lambda_1 I)^{\dim V}\right),$$

with both summands invariant under T. The induction hypothesis applied to T restricted to the range, which is strictly smaller than V if $G(\lambda_1, T) \neq \{0\}$, delivers the decomposition for the remaining eigenvalues. Nilpotency of $(T-\lambda_i I)$ on $G(\lambda_i, T)$ follows directly from the definition of the generalised eigenspace together with the stabilisation of null spaces. ■

Definition 8.4 (Algebraic and geometric multiplicity)**.** Let λ be an eigenvalue of $T \in \mathcal{L}(V)$.

- The algebraic multiplicity of λ is $\dim G(\lambda, T)$.
- The geometric multiplicity of λ is $\dim E(\lambda, T)$.

Since $E(\lambda, T) \subseteq G(\lambda, T)$, geometric multiplicity is at most algebraic multiplicity. Equality holds precisely when $E(\lambda, T) = G(\lambda, T)$, i.e. when there are no "defective" generalised eigenvectors for λ that fail to be true eigenvectors. We will see below that the operator T is diagonalisable if and only if every eigenvalue has equal algebraic and geometric multiplicities.

8.4 Jordan form

Assembling the pieces, we arrive at the Jordan canonical form, which provides a near-diagonal matrix representation of any operator on a finite-dimensional complex vector space. A Jordan block is the simplest kind of operator that is not diagonalisable: a scalar λ on the diagonal plus a single nilpotent perturbation pattern of 1's on the superdiagonal.

Definition 8.5 (Jordan block)**.** A Jordan block of size k with eigenvalue λ is the

$k \times k$ matrix

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix}. \tag{8.8}$$

A Jordan block is the matrix of an operator that acts on a basis $(e_1, \ldots, e_k)$ by sending each e_i to $\lambda e_i + e_{i-1}$ (with the convention $e_0 = 0$). The first basis vector is a true eigenvector; the remaining ones are generalised eigenvectors at increasing distances from being true eigenvectors, matching the structure of a nilpotent block shifted by λ.

Theorem 8.5 (Jordan form)**.** Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. There exists a basis of V with respect to which the matrix of T is block-diagonal,

$$\mathcal{M}(T) = \begin{pmatrix} J_{k_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{k_p}(\lambda_p) \end{pmatrix}, \tag{8.9}$$

where each block is a Jordan block. The list of Jordan blocks is unique up to reordering.

Sketch. Decompose $V = \bigoplus_i G(\lambda_i, T)$ using the previous theorem. On each $G(\lambda_i, T)$, the restriction of $T - \lambda_i I$ is nilpotent, and by the nilpotent block basis theorem this restriction has a block-diagonal matrix consisting of nilpotent blocks. Restoring the scalar λ_i on the diagonal of each such nilpotent block converts it into a Jordan block with eigenvalue λ_i. Assembling these bases across all generalised eigenspaces gives a basis of V in which T is block-diagonal with Jordan blocks. Uniqueness follows from the fact that the sizes and number of Jordan blocks for a given eigenvalue are determined by the dimensions of the successive null spaces $\operatorname{null}((T - \lambda I)^j)$, which are invariants of the operator. ■

Corollary 8.5.1. Every operator on a finite-dimensional complex vector space admits a basis of generalised eigenvectors with respect to which the matrix is block upper-triangular with Jordan blocks.

Remark (When is an operator diagonalisable?). The Jordan form gives a clean diagnostic: T is diagonalisable if and only if every Jordan block has size 1, equivalently $E(\lambda, T) =$

$G(\lambda, T)$ for every eigenvalue λ, equivalently the algebraic and geometric multiplicities agree for every eigenvalue. Blocks of size ≥ 2 correspond precisely to the "defective" generalised eigenvectors that cause diagonalisability to fail.

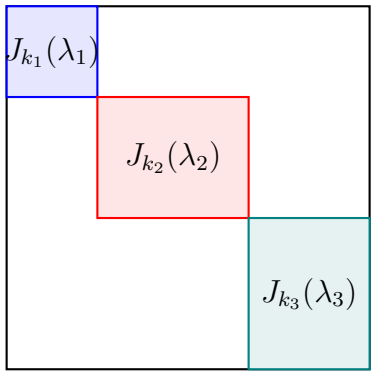

Figure 10: A Jordan form matrix is block-diagonal, with each block a Jordan block corresponding to one eigenvalue (eigenvalues may repeat across blocks).

9 Determinants

The determinant assigns a scalar to each square matrix and, via the matrix representation, to each operator on a finite-dimensional space. In $\mathbb{R}^2$ and $\mathbb{R}^3$ it has a direct geometric meaning as a signed area or volume; more generally it detects whether a matrix is invertible (nonzero determinant) or whether its columns are linearly dependent (zero determinant). Classically the determinant is defined by an explicit formula involving permutations; we instead develop it in two complementary ways, by a recursive cofactor expansion and by an axiomatic characterisation via multilinearity and alternation. Either definition uniquely determines the same function.

9.1 Definition via cofactor expansion

We define the determinant by a recursive procedure, starting from 1×1 matrices and expanding along the first row.

Definition 9.1 (Determinant)**.** The determinant is the function $\det : \mathbb{F}^{n\times n} \to \mathbb{F}$ defined recursively as follows.

- Base case: if B is the unique 0×0 empty matrix, then $\det(B) = 1$.
- $n = 1$ case: for $A = (a) \in \mathbb{F}^{1\times 1}$, $\det(A) = a$.
- Recursive case: for $A \in \mathbb{F}^{n\times n}$ with $n \geq 2$,

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} A_{1,j} \det(A_{\hat{1},\hat{j}}), \tag{9.1}$$

  where $A_{\hat{1},\hat{j}}$ denotes the $(n-1) \times (n-1)$ matrix obtained from A by deleting row 1 and column j.

The sum picks out each entry of the first row, multiplies it by the determinant of the submatrix obtained by striking through its row and column, and attaches a sign $(-1)^{1+j}$ that alternates across the row. The sign pattern $+, -, +, -, \ldots$ is dictated by the behaviour

of the determinant under column swaps, which we will derive below.

Example 18

[$n = 2$] For the 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

the cofactor expansion gives

$$\det(A) = (-1)^{1+1} a \det(d) + (-1)^{1+2} b \det(c) = ad - bc.$$

Example 19

[$n = 3$] For

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},$$

expanding along the first row,

$$\det(A) = a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}.$$

Computing the three 2×2 determinants,

$$\det(A) = a(ei - fh) - b(di - gf) + c(dh - eg).$$

Even at $n = 3$ the formula is substantial; by $n = 4$ it has 24 terms, and by $n = 10$ it has 10! terms. The definition is better understood as a specification of what the determinant is than as a practical computational rule.

Proposition 9.1 (Determinant of an upper-triangular matrix)**.** If $A \in \mathbb{F}^{n \times n}$ is upper-triangular, then

$$\det(A) = \prod_{i=1}^{n} A_{i,i}. \tag{9.2}$$

Proof. By induction on n. The $n = 1$ case is immediate from the definition. For $n \geq 2$, expand along the first row. Upper-triangularity gives $A_{1,j} = 0$ for $j > 1$ is not the right condition — what matters is that the first column of an upper-triangular matrix has $A_{i,1} = 0$ for $i > 1$. Looking at the cofactors $A_{\hat{1},\hat{j}}$ for $j \geq 2$, we strike row 1 and column j. The first column of $A_{\hat{1},\hat{j}}$ is formed from entries $A_{2,1}, A_{3,1}, \ldots$ of A, all of which are zero by upper-triangularity. An upper-triangular matrix with a zero first column has zero determinant (by the induction hypothesis, since the diagonal product includes 0), so all but the first term of the expansion vanish. The surviving term is

$$(-1)^{1+1} A_{1,1} \det(A_{\hat{1},\hat{1}}) = A_{1,1} \det(A_{\hat{1},\hat{1}}).$$

The minor $A_{\hat{1},\hat{1}}$ is itself upper-triangular of size $n-1$ with diagonal entries $A_{2,2}, A_{3,3}, \ldots, A_{n,n}$. By the induction hypothesis, $\det(A_{\hat{1},\hat{1}}) = A_{2,2} \cdots A_{n,n}$. Combining,

$$\det(A) = A_{1,1} \cdot A_{2,2} \cdots A_{n,n} = \prod_{i=1}^{n} A_{i,i}. \qquad \blacksquare$$

Transposing a matrix leaves the determinant unchanged. This is a strong statement: although the recursive definition privileges the first row, the transpose identity shows that column expansions give the same answer.

Proposition 9.2. For every $A \in \mathbb{F}^{n \times n}$, $\det(A^T) = \det(A)$.

Corollary 9.0.1 (Cofactor expansion along any row or column)**.** For any i, j and any $A \in \mathbb{F}^{n \times n}$,

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} \det(A_{\hat{i},\hat{j}}) = \sum_{i=1}^{n} (-1)^{i+j} A_{i,j} \det(A_{\hat{i},\hat{j}}). \tag{9.3}$$

The flexibility to expand along any row or column is extremely useful in practice: one chooses a row or column with many zeros to minimise the work.

9.2 Characterisation via multilinearity and alternation

There is an alternative and conceptually cleaner approach to the determinant. Rather than defining it by a formula, we can characterise it by three properties — multilinearity, alternation, and normalisation — and prove that there is a unique function satisfying them. This abstract characterisation is often easier to work with than the cofactor formula.

Theorem 9.1 (Axiomatic characterisation of the determinant)**.** The determinant is the unique function $D : \mathbb{F}^{n\times n} \to \mathbb{F}$ that, viewing a matrix as the ordered tuple of its columns $(a_1, \ldots, a_n)$, satisfies:

1. Multilinearity: D is linear in each column when the others are held fixed. Specifically, if $a_k = \lambda u + \mu v$, then

$$D(\ldots, \lambda u + \mu v, \ldots) = \lambda D(\ldots, u, \ldots) + \mu D(\ldots, v, \ldots). \tag{9.4}$$

2. Alternation: if two columns of A are equal, then $D(A) = 0$.

3. Normalisation: $D(I_n) = 1$.

The content of the theorem is that these three properties pin down a unique function, and that function agrees with the cofactor expansion. We use this characterisation to derive the practical manipulations of the determinant.

Proposition 9.3 (Consequences of alternation)**.** The determinant satisfies the following properties.

1. Swapping two columns negates the determinant.

2. Adding a scalar multiple of one column to another leaves the determinant unchanged.

3. If any column is the zero vector, the determinant is zero.

4. If the columns are linearly dependent, the determinant is zero.

Proof of (1). The key trick is to consider a matrix with specially arranged equal columns and then expand by multilinearity. Fix indices $i < j$ and consider the determinant of the matrix with its ith column replaced by $a_i + a_j$ and its jth column also replaced by $a_i + a_j$. The resulting matrix has two equal columns, so by alternation its determinant is zero.

On the other hand, expanding both replaced columns by multilinearity gives $2 \times 2 = 4$ terms:

$$D(\ldots, a_i+a_j, \ldots, a_i+a_j, \ldots) = D(\ldots, a_i, \ldots, a_i, \ldots)+D(\ldots, a_i, \ldots, a_j, \ldots)+D(\ldots, a_j, \ldots, a_i, \ldots)+D(.$$

The first and fourth terms each have two equal columns, hence by alternation both vanish. What remains is

$$D(\ldots, a_i, \ldots, a_j, \ldots) + D(\ldots, a_j, \ldots, a_i, \ldots) = 0,$$

which rearranges to $D(\ldots, a_j, \ldots, a_i, \ldots) = -D(\ldots, a_i, \ldots, a_j, \ldots)$, i.e. swapping columns negates the determinant. ■

Proof of (2). Adding λa_i to column j (with $j \neq i$) gives

$$D(\ldots, a_i, \ldots, a_j + \lambda a_i, \ldots) = D(\ldots, a_i, \ldots, a_j, \ldots) + \lambda D(\ldots, a_i, \ldots, a_i, \ldots).$$

The second term has two equal columns and vanishes by alternation. The first term is just $D(A)$. ■

Proof of (3). If the kth column is zero, use multilinearity to factor out 0, which produces 0. ■

Proof of (4). If the columns are linearly dependent, some column a_j is a linear combination of the others, $a_j = \sum_{i \neq j} \lambda_i a_i$. By multilinearity in column j,

$$D(A) = \sum_{i \neq j} \lambda_i D(\ldots, a_i, \ldots, a_i, \ldots),$$

where in each term the ith and jth columns are both a_i. Every such term has two equal columns and hence vanishes by alternation. ■

9.3 Multiplicativity and invertibility

Theorem 9.2 (Multiplicativity of the determinant)**.** For any $A, B \in \mathbb{F}^{n\times n}$,

$$\det(AB) = \det(A)\,\det(B). \tag{9.5}$$

Sketch. Fix A and consider $B \mapsto \det(AB)$ as a function of the columns of B. Writing B in terms of its columns $(b_1, \ldots, b_n)$, the kth column of AB is Ab_k, and the map $b_k \mapsto Ab_k$ is linear. Composing with det, which is multilinear and alternating in its columns, we obtain a function of $(b_1, \ldots, b_n)$ that is multilinear and alternating. By the uniqueness part of the characterisation theorem, any such function is a scalar multiple of det, and the scalar is the value at $B = I$, namely $\det(AI) = \det(A)$.

Therefore $\det(AB) = \det(A)\,\det(B)$. ■

Multiplicativity combined with the value at the identity gives the fundamental algebraic criterion for invertibility.

Theorem 9.3 (Invertibility criterion)**.** A matrix $A \in \mathbb{F}^{n\times n}$ is invertible if and only if $\det(A) \neq 0$.

Proof. Forward direction. Suppose A is invertible, so $AA^{-1} = I$. By multiplicativity,

$$\det(A)\,\det(A^{-1}) = \det(I) = 1,$$

which is possible only if $\det(A) \neq 0$ (and $\det(A^{-1})$ is its reciprocal).

Reverse direction. We prove the contrapositive: if A is not invertible then $\det(A) = 0$. A non-invertible matrix has linearly dependent columns — because the associated linear map has a nontrivial null space, and the columns spanning the range together with the dependency in the null space give a nontrivial linear combination of columns equalling

zero. By the fourth consequence of alternation above, linearly dependent columns force $\det(A) = 0$. ■

Corollary 9.3.1. If A is invertible, $\det(A^{-1}) = \det(A)^{-1}$.

9.4 Determinant of a linear operator

Having defined the determinant for matrices, we can extend it to operators on finite-dimensional spaces via any choice of basis.

Definition 9.2 (Determinant of an operator)**.** Let $T \in \mathcal{L}(V)$ with V finite-dimensional. Choose any basis of V; the determinant of T is

$$\det(T) = \det(\mathcal{M}(T)). \tag{9.6}$$

To justify this definition, we must show it does not depend on the choice of basis.

Proposition 9.4. The value of $\det(T)$ is independent of the basis used to compute it.

Proof. Matrices of the same operator T in two different bases are related by a change-of-basis transformation,

$$A = C^{-1}BC$$

for some invertible C. Applying multiplicativity,

$$\det(A) = \det(C^{-1})\det(B)\det(C) = \det(C)^{-1}\det(B)\det(C).$$

The scalars $\det(C)$ and $\det(C)^{-1}$ are reciprocals of one another in $\mathbb{F}$, so they commute and cancel (this is just a multiplication of scalars in a field, which is commutative), leaving

$$\det(A) = \det(B).$$

So both matrices of T give the same determinant. ■

Combining this basis-independence with the structural theorems of the previous chapters gives the eigenvalue formula, one of the most frequently used identities about the determinant.

Proposition 9.5. Over $\mathbb{C}$, the determinant of $T \in \mathcal{L}(V)$ equals the product of the eigenvalues of T counted with algebraic multiplicity.

Proof. By Schur's theorem, there is a basis of V with respect to which $\mathcal{M}(T)$ is upper-triangular. The diagonal entries of this upper-triangular matrix are the eigenvalues of T, with each eigenvalue appearing a number of times equal to its algebraic multiplicity. The determinant of an upper-triangular matrix is the product of its diagonal entries, so $\det(T)$ is the product of the eigenvalues of T counted with algebraic multiplicity. Since $\det(T)$ does not depend on the basis, the same value is obtained from any other basis. ∎

9.5 Cramer's rule and computation

Determinants give an explicit, though not practically efficient, formula for the solution to a linear system.

Theorem 9.4 (Cramer's rule)**.** Let $A \in \mathbb{F}^{n\times n}$ be invertible and consider the linear system $Ax = b$. Its unique solution has components

$$x_j = \frac{\det(A_j)}{\det(A)}, \qquad j = 1, \ldots, n, \tag{9.7}$$

where A_j is the matrix obtained from A by replacing its jth column with b.

Remark. Cramer's rule is theoretically clean but computationally impractical for large n. Each application requires computing $n + 1$ determinants, and a direct expansion of each determinant costs on the order of $n!$ operations. Gaussian elimination solves the same system in roughly n^3 operations, which grows vastly more slowly. Cramer's rule remains useful for small cases, for symbolic manipulations, and for the structural insight that each component of the solution is a ratio of two determinants.

10 Tensor Products

The tensor product is a construction that turns up throughout mathematics and physics: in multilinear algebra, differential geometry, representation theory, and the formulation of multi-particle quantum mechanics. It solves a specific technical problem: how to convert bilinear maps into ordinary linear maps, so that the machinery developed in the previous chapters can be brought to bear on them. We present two complementary views of the tensor product — one concrete (through bases) and one abstract (through a universal property) — and verify that they describe the same object.

10.1 Motivation: bilinear maps

Up to this point, our linear maps have taken a single vector as input. Many interesting pairings of vectors take *two* vectors as input and produce a third object in a way that is linear in each argument separately.

Definition 10.1 (Bilinear map)**.** Let V, W, Z be vector spaces over $\mathbb{F}$. A map $B : V \times W \to Z$ is called bilinear if it is linear in each argument when the other is held fixed. Explicitly, for all $v, v' \in V$, $w, w' \in W$, and $\lambda \in \mathbb{F}$,

$$B(v + v', w) = B(v, w) + B(v', w), \tag{10.1}$$

$$B(v, w + w') = B(v, w) + B(v, w'), \tag{10.2}$$

$$B(\lambda v, w) = \lambda B(v, w) = B(v, \lambda w). \tag{10.3}$$

Bilinear maps pervade linear algebra. The dot product $(u, v) \mapsto u \cdot v$ is bilinear. Matrix multiplication $(A, B) \mapsto AB$ is bilinear in the two factors. The cross product $(u, v) \mapsto u \times v$ in $\mathbb{R}^3$ is bilinear. Inner products are bilinear (over $\mathbb{R}$) or sesquilinear (over $\mathbb{C}$). Each of these is a pairing of two inputs from possibly different spaces, linear separately in each.

It is tempting to view a bilinear map $B : V \times W \to Z$ as a linear map from the product

space $V \times W$ to Z, but this does not work. A linear map from $V \times W$ would satisfy $L(\lambda(v, w)) = \lambda L(v, w)$ for a single scalar λ. But bilinearity gives

$$B(\lambda v, \lambda w) = \lambda B(v, \lambda w) = \lambda^2 B(v, w),$$

which is λ^2, not λ. So bilinear maps genuinely lie outside the linear framework.

The tensor product $V \otimes W$ is designed to solve exactly this problem: it is a new vector space, built out of V and W, together with a universal bilinear map $\otimes : V \times W \to V \otimes W$ such that every bilinear map out of $V \times W$ factors uniquely through $\otimes$ as a *linear* map on $V \otimes W$. In this sense $V \otimes W$ is the "free linear home" for bilinear data.

10.2 Construction via a basis

The most concrete way to build $V \otimes W$ is via bases.

Definition 10.2 (Tensor product via basis)**.** Let V and W be finite-dimensional vector spaces over $\mathbb{F}$ with bases $(v_1, \ldots, v_m)$ of V and $(w_1, \ldots, w_n)$ of W. Define $V \otimes W$ to be the vector space with basis $\{v_i \otimes w_j : 1 \le i \le m,\ 1 \le j \le n\}$, whose elements we call simple or decomposable tensors, together with the symbol $\otimes$ extended to arbitrary inputs by bilinearity:

$$v \otimes w = \sum_{i,j} a_i b_j \, v_i \otimes w_j, \tag{10.4}$$

whenever $v = \sum_i a_i v_i$ and $w = \sum_j b_j w_j$.

From the definition it is immediate that $\dim(V \otimes W) = mn = \dim V \cdot \dim W$: the basis of $V \otimes W$ consists of exactly mn vectors, one for each pair (v_i, w_j). The map

$\otimes : V \times W \to V \otimes W$ defined above is bilinear, essentially by construction:

$$(v + v') \otimes w = v \otimes w + v' \otimes w, \tag{10.5}$$

$$v \otimes (w + w') = v \otimes w + v \otimes w', \tag{10.6}$$

$$(\lambda v) \otimes w = \lambda(v \otimes w) = v \otimes (\lambda w). \tag{10.7}$$

Verifying these amounts to expanding both sides in the given bases and comparing.

Remark (Coordinates). Under the identification of V and W with $\mathbb{F}^m$ and $\mathbb{F}^n$ via the chosen bases, a simple tensor $v \otimes w$ corresponds to the $m \times n$ matrix vw^T with entries $(vw^T)_{i,j} = a_i b_j$, or equivalently to an mn-dimensional column vector listing the entries of this matrix. For $v = (a_1, b_1)^T$ and $w = (a_2, b_2)^T$ in $\mathbb{F}^2$,

$$v \otimes w = \begin{pmatrix} a_1 a_2 \\ a_1 b_2 \\ b_1 a_2 \\ b_1 b_2 \end{pmatrix} \in \mathbb{F}^4.$$

The pattern is that coordinates in slot (i, j) are obtained by multiplying coordinate i of v by coordinate j of w, then flattened into a single long vector.

Remark (Not every element is a simple tensor). Although simple tensors $v \otimes w$ span $V \otimes W$, most elements of the tensor product are not themselves simple. A typical element is a sum of several simple tensors that cannot be collapsed into a single product. For instance in $\mathbb{F}^2 \otimes \mathbb{F}^2$, the element

$$v_1 \otimes w_1 + v_2 \otimes w_2$$

(where v_i and w_j are the respective basis vectors) cannot be written as $v \otimes w$ for any single pair. In quantum mechanics, exactly this phenomenon gives rise to entangled states: a two-particle state that is not a product of single-particle states.

10.3 The universal property

The basis construction is explicit but depends on choices. The universal property gives an intrinsic description, which both demonstrates that the basis construction is independent of choices and illuminates why tensor products are useful.

Theorem 10.1 (Universal property of the tensor product)**.** Let V and W be finite-dimensional vector spaces. For every vector space Z and every bilinear map $B : V \times W \to Z$, there exists a unique linear map $\widetilde{B} : V \otimes W \to Z$ such that

$$B(v, w) = \widetilde{B}(v \otimes w) \qquad \forall v \in V,\ w \in W. \tag{10.8}$$

The content of the theorem is that linear maps out of $V \otimes W$ are the same thing as bilinear maps out of $V \times W$. We can translate freely between the two viewpoints.

Proof. We establish existence and uniqueness separately.

Existence. Define $\widetilde{B}$ on the basis of $V \otimes W$ by

$$\widetilde{B}(v_i \otimes w_j) = B(v_i, w_j),$$

and extend linearly to the whole space. This assignment is well-defined because $(v_i \otimes w_j)$ is a basis. To verify the factorisation property, take arbitrary $v = \sum_i a_i v_i$ and $w = \sum_j b_j w_j$. Applying $\widetilde{B}$ to $v \otimes w$ using its expression on basis elements,

$$\widetilde{B}(v \otimes w) = \widetilde{B}\left(\sum_{i,j} a_i b_j\, v_i \otimes w_j\right) = \sum_{i,j} a_i b_j\, B(v_i, w_j).$$

Using bilinearity of B to compress the right-hand side back into a single evaluation,

$$\sum_{i,j} a_i b_j\, B(v_i, w_j) = B\left(\sum_i a_i v_i,\ \sum_j b_j w_j\right) = B(v, w).$$

Hence $\widetilde{B}(v \otimes w) = B(v, w)$ for every v, w.

Uniqueness. Suppose $\widetilde{B}'$ is another linear map $V \otimes W \to Z$ satisfying the factorisation

$\widetilde{B}'(v \otimes w) = B(v, w)$. Applied to the basis elements $v_i \otimes w_j$, this gives

$$\widetilde{B}'(v_i \otimes w_j) = B(v_i, w_j) = \widetilde{B}(v_i \otimes w_j).$$

Two linear maps agreeing on a basis are equal, so $\widetilde{B}' = \widetilde{B}$. ■

Corollary 10.1.1 (Basis-independence)**.** The pair $(V \otimes W, \otimes)$ is determined uniquely up to a unique isomorphism. In particular, the tensor product does not really depend on the choice of bases used to construct it.

Proof. Suppose $(U, \star)$ is any other pair consisting of a vector space U and a bilinear map $\star : V \times W \to U$ satisfying the universal property. Apply the universal property of $(V \otimes W, \otimes)$ to the bilinear map $\star$: there is a unique linear map $\widetilde{\star} : V \otimes W \to U$ with $\widetilde{\star}(v \otimes w) = v \star w$. Apply the universal property of $(U, \star)$ to the bilinear map $\otimes$: there is a unique linear map $\widetilde{\otimes} : U \to V \otimes W$ with $\widetilde{\otimes}(v \star w) = v \otimes w$.

Composing, $\widetilde{\otimes} \circ \widetilde{\star}$ is a linear map $V \otimes W \to V \otimes W$ satisfying $\widetilde{\otimes}(\widetilde{\star}(v \otimes w)) = v \otimes w$, which is the same behaviour as the identity on $V \otimes W$. Uniqueness in the universal property applied to the identity-factoring-the-bilinear-map forces $\widetilde{\otimes} \circ \widetilde{\star} = I_{V \otimes W}$. Similarly $\widetilde{\star} \circ \widetilde{\otimes} = I_U$. So $\widetilde{\star}$ and $\widetilde{\otimes}$ are mutually inverse isomorphisms, and $(U, \star)$ is canonically identified with $(V \otimes W, \otimes)$. ■

10.4 Basic identities

The tensor product satisfies identities that look very much like those of ordinary multiplication: a commutative law, an associative law, a unit law, and distributivity over direct sums.

Proposition 10.1 (Identities for tensor products)**.** For finite-dimensional V, W, U

over $\mathbb{F}$ we have natural isomorphisms

$$V \otimes W \cong W \otimes V, \tag{10.9}$$

$$(V \otimes W) \otimes U \cong V \otimes (W \otimes U), \tag{10.10}$$

$$V \otimes \mathbb{F} \cong V, \tag{10.11}$$

$$V \otimes (W \oplus U) \cong (V \otimes W) \oplus (V \otimes U). \tag{10.12}$$

Sketch. Each isomorphism is built by the following recipe: exhibit a bilinear (or multilinear) map in one direction, use the universal property to produce its unique linear extension, exhibit an inverse in the same way, and verify the two compositions are the identity by checking on basis tensors. For the first identity, the bilinear map $(v, w) \mapsto w \otimes v$ induces a linear map $V \otimes W \to W \otimes V$ sending $v \otimes w$ to $w \otimes v$; the map in the reverse direction is constructed symmetrically, and the two are inverse to each other. ■

Linear maps between individual spaces induce linear maps between their tensor products, in a way compatible with composition.

Proposition 10.2 (Tensor product of linear maps)**.** Given linear maps $S \in \mathcal{L}(V_1, V_2)$ and $T \in \mathcal{L}(W_1, W_2)$, there is a unique linear map

$$S \otimes T : V_1 \otimes W_1 \to V_2 \otimes W_2 \tag{10.13}$$

satisfying $(S \otimes T)(v \otimes w) = S(v) \otimes T(w)$ on simple tensors.

Proof. The assignment $(v, w) \mapsto S(v) \otimes T(w)$ is a map $V_1 \times W_1 \to V_2 \otimes W_2$. It is bilinear: linearity in v follows from linearity of S and bilinearity of $\otimes$, and similarly for w. By the universal property there is a unique linear map $V_1 \otimes W_1 \to V_2 \otimes W_2$ agreeing with this bilinear map on simple tensors; this is $S \otimes T$. ■

In coordinates, if $\mathcal{M}(S) = A \in \mathbb{F}^{p\times m}$ and $\mathcal{M}(T) = B \in \mathbb{F}^{q\times n}$, then the matrix of $S \otimes T$ is the *Kronecker product* $A \otimes B \in \mathbb{F}^{pq\times mn}$, arranged in block form as

$$A \otimes B = \begin{pmatrix} A_{1,1}B & A_{1,2}B & \cdots & A_{1,m}B \\ A_{2,1}B & A_{2,2}B & \cdots & A_{2,m}B \\ \vdots & \vdots & & \vdots \\ A_{p,1}B & A_{p,2}B & \cdots & A_{p,m}B \end{pmatrix}. \tag{10.14}$$

Each scalar entry of A is replaced by that scalar times the entire matrix B, producing a larger matrix indexed by pairs.

10.5 Connection to bilinear forms and outlook

Setting $Z = \mathbb{F}$ in the universal property gives a natural identification between bilinear forms on $V \times W$ and linear functionals on $V \otimes W$. In this way $(V \otimes W)^*$ is isomorphic to the space of bilinear forms on $V \times W$, which is itself naturally isomorphic to $V^* \otimes W^*$. Bilinear algebra and tensor algebra thus become two sides of the same coin.

Concretely, when $V = W = \mathbb{F}^n$, a bilinear form $B : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ takes the form

$$B(x, y) = x^T A y$$

for some matrix $A \in \mathbb{F}^{n\times n}$, and the correspondence $B \leftrightarrow A$ is exactly the correspondence $B \leftrightarrow$ "the tensor whose (i, j) coordinate is $A_{i,j}$" in $\mathbb{F}^n \otimes \mathbb{F}^n$. The tensor product gives a coordinate-free setting in which this kind of structure can be discussed.

Tensor products generalise in several directions:

- Multilinear maps $V_1 \times \cdots \times V_k \to Z$ are classified by the iterated tensor product $V_1 \otimes \cdots \otimes V_k$, via an obvious generalisation of the universal property.
- Symmetric and alternating tensors: quotients of $V^{\otimes k}$ by the relations that permuting factors acts trivially, or by sign, produce the symmetric power $\mathrm{Sym}^k V$ and the

exterior power $\Lambda^k V$ respectively. These classify symmetric and alternating multilinear maps, and determinants live naturally in $\Lambda^n V$.

- Infinite-dimensional inner product spaces (Hilbert spaces) admit a completed tensor product whose elements underlie multi-particle quantum mechanics: the tensor product of the state spaces of two particles is the state space of the combined system, and entangled states are precisely the non-simple tensors.

The tensor product is thus not merely a formal device for handling bilinear maps; it is the algebraic machinery that organises much of multilinear and categorical algebra and is the common language of linear structures in mathematics and physics.