Linear Algebra: A Field Guide

David Grabovsky

July 27, 2020

Contents

Part I
Basic Notions

Linear algebra is about structures called vector spaces whose elements, called vectors, may be added and scaled; and about maps between them, called linear transformations, that preserve their linear structure. In this sense, objects as different as numbers, pointed arrows in space, and even functions fall under the subject's purview.

1 Welcome to Vector Space

1.1 The Rules of the Game

Definition 1.1. A *field* F is a set of objects called *scalars* that behave in a sufficiently number-like manner. In particular, for all $x, y, z \in F$, the field must support:

(a) Associative-commutative binary operations $+$ and $\cdot$;
(b) Additive and multiplicative identity elements $0 \neq 1$;
(c) Unique additive and multiplicative inverses $-x$ and x^{-1};
(d) The distributive property: $x(y+z) = xy + xz$.

Example 1.2. $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all fields. Fields can also be finite, e.g. $\mathbb{F}_2 := \{0, 1\}$. While we will usually take $F = \mathbb{C}$, the real numbers provide the best intuition for the term "scalar."

Definition 1.3. A *vector space* V *over a field* F is a set whose elements, called *vectors*, can be added and scaled by field elements. Thus every vector space is equipped with:

(a) Associative-commutative addition: $\mathbf{u}, \mathbf{v} \in V \implies \mathbf{u} + \mathbf{v} \in V$;
(b) A unique *zero vector* $\mathbf{0} \in V$ satisfying $\mathbf{v} + \mathbf{0} = \mathbf{v}$;
(c) Unique additive inverses $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$;
(d) Associative-distributive *scalar multiplication* over F, with $1\mathbf{v} = \mathbf{v}$.

A *subspace* is a nonempty subset $W \subset V$ that inherits its structure from V, i.e. it is closed under the addition and scalar multiplication in V, and contains $\mathbf{0}$.

Example 1.4. Ordered n-tuples (lists) of numbers, directed arrows in space, $m \times n$ matrices, and functions can all be added together and scaled by real (or complex) numbers in obvious ways, so they all form vector spaces. Moreover, any field forms a vector space over itself, so we can view $\mathbb{C}$ as a complex vector space as well as a field of scalars.

The most obvious move, after making it possible to add and scale vectors arbitrarily, is to *do* it. The job of scaling and summing vectors is accomplished by linear combinations:

Definition 1.5. A *linear combination* of a set of vectors $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\} \in V$ is a new vector $\mathbf{v} \in V$, formed by scaling the original vectors by given coefficients $c_1, ..., c_n \in F$, and then adding together the scaled vectors:

$$\mathbf{v} := \sum_{i=1}^{n} c_i \mathbf{v}_i \in V. \tag{1.1}$$

Linear combination is "the only game in town," insofar as it gives the only way to generate new vectors from old ones. One might ponder two natural questions: (1) what might the set of all possible linear combinations of a given set look like? and conversely, (2) is it possible to write an arbitrary $\mathbf{v} \in V$ as a linear combination of some minimal set of vectors?

1.2 Basis and Dimension

Definition 1.6. The *linear span* of a set $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\} \subset V$ is the set of all possible linear combinations of the vectors in β. It follows (check it!) $\text{span}(\beta)$ is a subspace of V.

Definition 1.7. A set $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\} \subset V$ is *linearly independent* if no $\mathbf{v}_i \in \beta$ can be written as a (nonzero) linear combination of the others. Intuitively, removing a vector from an independent set will reduce the span of the remaining set. Conversely, a set is *dependent* if it is not independent, i.e. if it contains "redundant" vectors.

The span of β describes the subspace $W \subset V$ of every vector that can be built from β; meanwhile, if β is independent, then it has minimal size among all sets that span W. If an independent set happens to span the entire vector space, then (!) any vector $\mathbf{v} \in V$ can be written as a unique linear combination of that set.

Definition 1.8. A *basis* $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\} \subset V$ for V is a linearly independent set of vectors with $\text{span}(\beta) = V$: that is, a maximally spanning, minimally redundant set. The *dimension* of V, denoted $n = \dim V$, is the number of vectors in any basis for V.

Henceforth, we will consider only finite-dimensional spaces. A finite basis allows us to decompose every $\mathbf{v} \in V$ as a linear combination

$$\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{v}_i; \qquad \mathbf{v}_i \in \beta, \quad c_i \in F. \tag{1.2}$$

To specify an arbitrary vector, we need to (1) fix a basis β and (2) record the list of coefficients $c_i \in F$, known as the *coordinates* of $\mathbf{v}$. There is no "canonical" choice of basis for V, so a vector's coordinates are only meaningful once a basis is chosen. We will soon learn how a vector's coordinates transform when the basis is changed.

Example 1.9. Nevertheless, by regarding F^n as a vector space over F, we can always represent a vector $\mathbf{v} \in V$ by its coordinates $\mathbf{c} = (c_1, ..., c_n) \in F^n$, as follows. F^n has a *standard basis* $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$, where $\mathbf{e}_i := (0, ..., 0, 1, 0, ..., 0)$ has a 1 in the i^{th} spot and zeros everywhere else. We see that $\mathbf{c}$ decomposes along the standard basis as

$$\mathbf{c} = \sum_{i=1}^{n} c_i \mathbf{e}_i. \tag{1.3}$$

This gives a correspondence between V and F^n, whereby $\mathbf{v}$ is represented by $\mathbf{c}$ and each $\mathbf{v}_i$ by $\mathbf{e}_i$. This re-emphasizes one of the key ideas: to understand the behavior of a vector, it suffices to fix a basis and keep track of its coordinates.

2 Linear Transformations

We now introduce a special class of functions between vector spaces that "preserve linear structure," i.e. take sums to sums and scalar multiples to scalar multiples. Throughout, let V and W be vector spaces over F of dimensions n and m, respectively.

2.1 Definition and Properties

Definition 2.1. A map $T\colon V \to W$ is called a *linear transformation* (also called a *linear map*) if for all $\mathbf{u},\ \mathbf{v} \in V$ and for all $c \in F$,

$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}); \qquad T(c\mathbf{v}) = cT(\mathbf{v}). \tag{2.1}$$

Note that $T(\mathbf{0}) = \mathbf{0}$. By convention, we omit parentheses and write $T\mathbf{v}$ in place of $T(\mathbf{v})$.

Example 2.2. The zero and identity transformations 0 and id are always linear; so is the "stretching" map $\mathbf{v} \mapsto c\mathbf{v}$ for any $c \in F$. Matrices (below) are linear maps, as are differentiation and integration from calculus. On which spaces do these operators act?

Linear maps preserve sums and multiples: forming a linear combination of some $\mathbf{v}_i$ in V and then applying T is equivalent to applying T to the $\mathbf{v}_i$ and then forming the same linear combination of the $T(\mathbf{v}_i)$ in W. Since we can decompose every vector along a basis β, specifying the values of T on the vectors in β determines its action action on any $\mathbf{v} \in V$:

$$T(\mathbf{v}) = T\left(\sum_{i=1}^{n} c_i\mathbf{v}_i\right) = \sum_{i=1}^{n} c_i T(\mathbf{v}_i). \tag{2.2}$$

The notion that a linear map is determined by where it sends a basis for the space can be more precisely stated as the Construction Principle:

Theorem 2.3 (Construction)**.** *If V has a basis $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ and $\mathbf{w}_1, ..., \mathbf{w}_n \in W$ are arbitrary, then there exists a unique linear map $T\colon V \to W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$.*

Linear maps from V to W can themselves be added and scaled (pointwise) to yield linear maps. And since the zero map is linear, the set $\mathcal{L}(V, W)$ of all linear maps forms a vector space. The composition of maps, while not commutative, gives rise to an associative "multiplication" that turns $\mathcal{L}(V, V)$ into a structure called an *algebra*.[1]

So linear maps are determined by where they send the basis of V, and each $\mathbf{v} \in V$ (respectively, each $\mathbf{w} := T(\mathbf{v}) \in W$) can be represented by a list of n (resp. m) coordinates. It therefore suffices, to completely characterize T, to record the m coordinates of each of the n vectors $T(\mathbf{v}_i)$ produced by feeding the basis of V into T. These mn scalars are conventionally collected into an array of numbers called a *matrix*, where each of n columns contains the m coordinates of the $T(\mathbf{v}_i)$. Notice that while T maps n dimensions into m, the matrix describing it is of size $m \times n$.

[1] In quantum mechanics, the composition AB of linear maps A and B can be used to define the *commutator* $[A, B] := AB - BA$. This is an example of a *Lie bracket* on the *algebra of observables.*

2.2 Kernel, Image, and Isomorphism

Linear maps may be defective: they may send all of V into a proper subspace of W, or they may send two vectors in V to the same $\mathbf{w} \in W$. In the first case, not all $\mathbf{w} \in W$ come from some $\mathbf{v} \in V$, and T is not *surjective*; in the second case, some $\mathbf{w}$ may not come from a unique $\mathbf{v}$, and T is not *injective*. We can actually quantify these defects:

Definition 2.4. The *kernel* of T is the subset of V of vectors that T sends to $\mathbf{0} \in W$. The *image* of T is the subset of W "covered" by T, i.e. all vectors of the form $T(\mathbf{v}) \in W$:

$$\ker T := \{\mathbf{v} \in V \mid T\mathbf{v} = \mathbf{0}\}; \qquad \operatorname{im} T := \{\mathbf{w} \in W \mid \mathbf{w} = T\mathbf{v},\ \mathbf{v} \in V\}. \tag{2.3}$$

It follows (check it!) that $\ker T$ is a subspace of V, while $\operatorname{im} T$ is a subspace of W.

The image of T is an intuitively clear description of surjectivity: if $\operatorname{im} T \neq W$, then T does not cover W and thus fails to be surjective. Conversely, if T is not surjective, then clearly $\operatorname{im} T \neq W$. Meanwhile, the kernel of T tells us how much of V is collapsed into $\mathbf{0}$ by T. If $\ker T$ is more than just $\{\mathbf{0}\}$, then T has sent some $\mathbf{v} \neq \mathbf{0}$ to $\mathbf{0} \in W$ and fails to be injective. Conversely, if T is not injective, then $T(\mathbf{u}) = T(\mathbf{v})$ for some $\mathbf{u} \neq \mathbf{v} \in V$. Hence $T(\mathbf{u} - \mathbf{v}) = \mathbf{0} = T(\mathbf{0})$, which puts $\mathbf{u} - \mathbf{v} \neq \mathbf{0}$ inside $\ker T$, so $\ker T \neq \{\mathbf{0}\}$. Thus we have proven that $\operatorname{im} T = W$ iff T surjects, while $\ker T = \{\mathbf{0}\}$ iff T injects.

Now every $\mathbf{v} \in V$ is either killed by T or survives (either $\mathbf{v} \in \ker T$ or $T(\mathbf{v}) \in \operatorname{im} T$), so the sizes of $\ker T$ and $\operatorname{im} T$ should add up to the size of V. Let β complete a basis of $\ker T$ to a basis of V, and send this basis through T. The first $\dim(\ker T)$ vectors will be dead upon arrival in W, while the remaining ones form a basis for $\operatorname{im} T$. Thus we have proven the

Theorem 2.5 (Rank-Nullity)**.** $\dim(\operatorname{im} T) + \dim(\ker T) = \dim V$.

Linear maps with no defects, which have trivial kernel and map onto their entire target space, serve as faithful and complete dictionaries between vector spaces. They are therefore reversible, meaning that a unique map T^{-1} undoes everything that T does to a vector.

Definition 2.6. An *isomorphism* $T\colon V \to W$ is a bijective (injective and surjective) linear map. If such a map exists, we write $V \cong W$ and say that the spaces V and W are *isomorphic*.[2]

Definition 2.7. The *inverse* of a linear map $T\colon V \to W$, if it exists, is a linear map $T^{-1}\colon W \to V$ such that $T^{-1} \circ T = \mathrm{id}_V$ and $T \circ T^{-1} = \mathrm{id}_W$.

Theorem 2.8 (Inverses)**.** *A linear map $T\colon V \to W$ is an isomorphism. Equivalently:*
(1) T has a unique inverse T^{-1}, which is then also an isomorphism;
(2a) $\ker T = \{\mathbf{0}\} \iff \dim(\ker T) = 0 \iff$ *T is injective;*
(2b) $\operatorname{im} T = W \iff \dim(\operatorname{im} T) = n \iff$ *T is surjective;*
(3) $n := \dim V = \dim W =: m$.

[2]There is actually a way to "force" every linear map $T\colon V \to W$ to become an isomorphism. Since T surjects onto its image, the map $T\colon V \to \operatorname{im} T \subset W$ is surjective. To make T injective, we declare all $\mathbf{v} \in V$ which T maps to the same $\mathbf{w} \in W$ to be equivalent. This "squashes" each non-injectivity (called a *fiber*) of T into a single point. The set of these *equivalence classes* of vectors is called the *quotient space* $\widetilde{V} = V/\ker T$. The new map $\widetilde{T}\colon \widetilde{V} \to \widetilde{W}$ is an isomorphism by construction, albeit between different spaces. This construction is native to group theory, where it is known as the First Isomorphism Theorem.

In practice, an isomorphism is constructed by choosing bases $\{\mathbf{v}_i\}$ for V and $\{\mathbf{w}_i\}$ for W, and then mapping one basis to another: $T(\mathbf{v}_i) = \mathbf{w}_i$. This makes it clear that vector spaces of different dimensions cannot be isomorphic—T would either squash a basis into dependence in W, or else be unable to fill out a basis for W. In fact, we've already seen this in action: we represent a vector by its coordinates by mapping the basis $\{\mathbf{v}_i\}$ of V to the standard basis $\{\mathbf{e}_i\}$ of F^n. Moreover, the discussion of matrices above yields another isomorphism that we now have the language to give:

Theorem 2.9 (Canonical isomorphisms)**.** *Every vector space of dimension n is isomorphic to F^n, and every linear map between $V \cong F^n$ and $W \cong F^m$ is determined by the matrix $A \in \mathcal{M}_{m\times n}$ whose columns are the coordinates of $T(\mathbf{v}_i)$. Moreover, $\mathcal{L}(V,W) \cong \mathcal{M}_{m\times n}$.*

2.3 A Change of Perspective

We mentioned that the coordinates of a vector were arbitrary and dependent on a choice of basis. When the basis is changed, the vector's coordinates change; so too do the entries of a matrix, if the bases of the spaces it maps between are changed. Nevertheless, the objects themselves remain the same; hence we examine changes of basis in light of this invariance.

Let $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ and $\beta' := \{\mathbf{v}'_1, ..., \mathbf{v}'_n\}$ be two different bases for V, and similarly let $\gamma := \{\mathbf{w}_i\}$ and $\gamma' := \{\mathbf{w}'_i\}$ be bases for W. Let $\mathbf{x} \in V$ be arbitrary. We can decompose it along either basis: $\mathbf{x} = \sum_{i=1}^n c_i\mathbf{v}_i = \sum_{j=1}^n c'_j\mathbf{v}'_j$. How are the c'_j related to the c_i? The trick is to decompose each $\mathbf{v}_i \in \beta$ along β'. If $\mathbf{v}_i$ has (primed) coordinates $(b_i)_j := b_{ij} \in F$, then

$$\mathbf{v}_i = \sum_{j=1}^n b_{ij}\mathbf{v}'_j \implies \mathbf{x} = \sum_{i=1}^n c_i\mathbf{v}_i = \sum_{i=1}^n c_i\left(\sum_{j=1}^n b_{ij}\mathbf{v}'_j\right) = \sum_{i=1}^n\sum_{j=1}^n c_ib_{ij}\mathbf{v}'_j = \sum_{j=1}^n c'_j\mathbf{v}'_j. \quad (2.4)$$

Here b_{ij}, the j^{th} primed coordinate of $\mathbf{v}_i$, is the "amount" of $\mathbf{v}_i$ along $\mathbf{v}'_j$. We see that

$$c'_j = \sum_{i=1}^n c_ib_{ij} \iff \mathbf{c}' = \sum_{i=1}^n c_i\mathbf{b}_i, \qquad \mathbf{c}' = (c'_1, ..., c'_n),\ \mathbf{b}_i = \big((b_i)_1, ..., (b_i)_n\big). \quad (2.5)$$

This linear *change of basis* transformation $\phi_V \colon V \to V$ sending $\beta' \mapsto \beta$ is invertible (you can change from β to β' too!), hence an isomorphism. It acts in coordinates by $\mathbf{c}' \mapsto \mathbf{c}$, and its matrix B has elements v_{ij}, so that $B^{-1}(\mathbf{c}) = \mathbf{c}'$ performs the change of coordinates $\beta \mapsto \beta'$.

Now consider a linear map $T \colon V \to W$, and suppose that V and W suffer changes of basis $\beta \mapsto \beta'$ and $\gamma \mapsto \gamma'$, with change-of-basis maps B_V and B_W, respectively. We want to know how the matrix entries of T transform when *both* β and γ are changed. For definiteness, let A, A' be the matrices of T in bases (β, γ) and (β', γ'), respectively. One might apply T to a vector $\mathbf{v} \in V$, and use B_V and B_W to decompose $\mathbf{v}$ and $\mathbf{w} = T(\mathbf{v}_i)$ in their respective primed and unprimed bases. But this approach is tedious and uninsightful. Instead, notice that the following are equivalent: (1) first apply A to β, then change basis $\gamma \mapsto \gamma'$; (2) first change basis $\beta \mapsto \beta'$, then apply A'. In other words, the diagram below commutes:

$$\begin{array}{ccc} V(\beta) & \xrightarrow{A} & W(\gamma) \\ \downarrow{\scriptstyle B_V^{-1}} & & \downarrow{\scriptstyle B_W^{-1}} \\ V(\beta') & \xrightarrow{A'} & W(\gamma') \end{array} \tag{2.6}$$

The diagram shows that $A' = B_W^{-1} A B_V$. That is, applying $A' \colon \beta' \mapsto \gamma'$ is the same as first un-priming V through B_V, then applying $A \colon \beta \mapsto \gamma$, and finally priming W using B_W^{-1}. In the case where $V = W$, T is called a *linear operator* (also called an *endomorphism*), and we recover the identity $A' = B^{-1}AB$, where now B is the sole change-of-basis matrix.

Definition 2.10. Two linear operators $A, A' \colon V \to V$ are called *similar* or *conjugate* if they are related by $A' = B^{-1}AB$ for some invertible map $B \colon V \to V$. This means that A and A' represent the same transformation, but with respect to different bases.

Example 2.11. Alexei Petrovich wants to start a Communist revolution, but can only speak Russian. He asks his good friend Boris to translate his pamphlets into English. The words are all different, but the call to arms and the hope for a better future remain impassioned. The propaganda is distributed, and the Red Army marches on Wall Street: long live the proletariat. Alexei could have accomplished his goal by seizing the means of production, violently enforcing global equality, and drafting a Constitution for the New World Order entirely in Russian before having Boris translate it—but surely the first plan was less work. Translation to and from Russian are accomplished by Boris's changes of basis B_V and B_V^{-1}, and worldwide political overhaul is given by A (in English) or alternatively by A' (in Russian). See if you can follow the commu(nist/tative) diagram above to track Alexei's rise to power.

3 Two Extra Topics

Many respectable courses in linear algebra spend a lot of time discussing systems of linear equations as a real-world application and developing determinants as an invaluable tool. However, these topics are a bit out of the way; we will treat them only briefly.

3.1 Linear Systems

Recall that vectors and linear maps can be represented by columns and arrays of numbers, respectively. It turns out that this makes linear algebra the perfect tool for solving large systems of linear equations. The idea is that any system of m equations in n unknowns,

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2 \\ \qquad\qquad \vdots \\ c_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = c_m, \end{cases} \tag{3.1}$$

can be compactly written $A\mathbf{x} = \mathbf{c}$, where $A \in \mathcal{M}_{m\times n}$ contains the given coefficients a_{ij}, $\mathbf{x} \in F^n$ contains the unknown variables, and $\mathbf{c} \in F^m$ is a vector of given constants.

If $\mathbf{c} = \mathbf{0}$, the system is called *homogeneous*, and standard procedures called row reduction and Gaussian elimination can be applied to solve the system; that is, to find a basis for $\ker A$. Gaussian elimination involves applying "elementary row operations" to A that preserve its kernel, but put A into a form that makes solving the system much easier. If $\mathbf{c} \neq \mathbf{0}$, we consider the "augmented system" $(A \mid \mathbf{c})\mathbf{x} = \mathbf{0}$, where $(A \mid \mathbf{c}) \in \mathcal{M}_{m\times(n+1)}$ is a matrix whose whose columns are those of A, followed by $\mathbf{c}$. After solving this (homogeneous) system for $\mathbf{x}$ in the usual way, we convert back to the inhomogeneous system by choosing all solutions to the augmented system with $x_{n+1} = -1$, for this equivalently solves $A\mathbf{x} = \mathbf{c}$.

3.2 Determinants

The determinant is an alien from another planet called "multilinear algebra."

The determinant of an operator is a number that represents the volume of the n-dimensional parallelepiped formed by the vectors that make up its rows. If the rows are linearly dependent, the polytope is "squashed" and $\det T$ collapses to zero. Geometrically, $\det T$ gives the "scale factor" of the transformation. The existence and uniqueness of such a function is a hard theorem which we will state, but not prove.

Theorem 3.1 (∃! determinant)**.** *There exists a unique function* $\det\colon \mathcal{L}(V,V) \to F$*, regarded as a function of the rows* $\mathbf{a}_1, ..., \mathbf{a}_n$ *of a matrix* A *representing some linear map* $T\colon V \to V$*, satisfying the following properties:*

D1: $\det(\mathbf{a}_1, ..., c\,\mathbf{a}_i, ..., \mathbf{a}_n) = c\det(\mathbf{a}_1, ..., \mathbf{a}_n)$*;*

D2: $\det(\mathbf{a}_1, ..., \mathbf{a}_i + \mathbf{b}_i, ..., \mathbf{a}_n) = \det(\mathbf{a}_1, ..., \mathbf{a}_i, ..., \mathbf{a}_n) + \det(\mathbf{a}_1, ..., \mathbf{b}_i, ..., \mathbf{a}_n)$*;*

D3: *If* $\mathbf{a}_i = \mathbf{a}_j$ *for* $i \neq j \in \{1, ..., n\}$*, then* $\det(\mathbf{a}_1, ..., a_n) = 0$*;*

D4: $\det(I) = 1$*.*

D1 and D2 characterize the determinant as *multilinear* (that is, linear in each argument separately), D3 says that det is *alternating* (vanishing whenever it is given identical arguments), and D4 normalizes det to unity for the identity matrix I. In other words, the determinant is the unique n-linear, alternating function on operators mapping I to 1.

Proposition 3.2 (Properties of det)**.** *Let* A, B *be square matrices, with rows* $\mathbf{a}_i$ *and* $\mathbf{b}_j$*, representing linear operators* T, S *on* V*, respectively. Then:*

1. *If any row* $\mathbf{a}_i = \mathbf{0}$*, then* $\det A = 0$*. Consequently, if any two rows of* A *are linearly dependent, then* $\det A = 0$*. Exchanging any two rows of* A *incurs a sign flip in* $\det A$*.*
2. *(a)* $\det AB = (\det A)(\det B)$*; (b)* $\det cA = c^n \det A$*; and (c)* $\det A^{-1} = \frac{1}{\det A}$*.*
3. *The determinant is also well-defined when taken as a function of the columns of* A*.*

From (2c), a square matrix is invertible (i.e. the map it represents is an isomorphism) iff $\det A \neq 0$. And from (3), the determinant of the *transpose* matrix A^T, formed by switching the rows and columns of A, i.e. $(A^T)_{ij} = A_{ji}$, is equal to the determinant of A. Determinants also end up being very useful for determining the eigenvalues of a matrix: this is the topic to which we turn next, but we will try to avoid the use of determinants in our discussion.

Part II
Diagonalizable Operators

4 Finding an Eigenbasis

With the basic apparatus of linear algebra in place, we will explore the structure of linear operators, or maps from a vector space to itself. To start, note that a linear operator $T\colon V \to V$ usually rotates or shears vectors, throwing them off their span. But some vectors are resilient: those $\mathbf{v} \in V$ which T merely scales, leaving their direction unchanged, are called its eigenvectors (eigen- means "self" in German, indicating that T returns eigenvectors to their own lines); the associated scaling factor is called an eigenvalue.

Definition 4.1. A nonzero $\mathbf{v} \in V$ is called an *eigenvector* of a linear operator $T\colon V \to V$ with *eigenvalue* $\lambda \in F$ if $T(\mathbf{v}) = \lambda\mathbf{v}$.

Example 4.2. Every vector is an eigenvector of the identity operator I with eigenvalue 1. Contrariwise, the planar rotation operator $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ has *no* eigenvectors at all, because its axis of rotation points out of the plane! In 3 dimensions, however, the axis is an eigenvector, and in fact rotations in odd-dimensional spaces always have an eigenvector with $\lambda = 1$.

A natural question is whether T has enough independent eigenvectors to form a basis for V.

Definition 4.3. A linear operator $T\colon V \to V$ is called *diagonalizable* if V has a basis, called a T-*eigenbasis*, consisting entirely of T-eigenvectors.

Diagonalizable operators are interesting because they carry an "intrinsic" basis $\beta := \{\mathbf{v}_i\}$ for the space, which is handy for decomposing arbitrary vectors on which they act:

$$T(\mathbf{v}) = T\left(\sum_{i=1}^{n} c_i\mathbf{v}_i\right) = \sum_{i=1}^{n} c_i T(\mathbf{v}_i) = \sum_{i=1}^{n} c_i\lambda_i\mathbf{v}_i \implies T(\mathbf{v}) \sim (c_1\lambda_1, ..., c_n\lambda_n). \tag{4.1}$$

After changing to the the eigenbasis, T acts by stretching each (basis) eigenvector $\mathbf{v}_i$ by λ_i. You can check that the matrix of T in its eigenbasis is *diagonal*: every entry is zero except for those along the main diagonal, and moreover $a_{ii} = \lambda_i$.

Theorem 4.4. *A linear operator $T\colon V \to V$ is diagonalizable if and only if its matrix A is similar to the diagonal matrix $D := \operatorname{diag}(\lambda_1, ..., \lambda_n)$; that is, iff $A = B^{-1}DB$, where B is the matrix of a change into the eigenbasis of T.*

Example 4.5. Consider a system of coupled oscillators, i.e. point masses attached to each other by springs. The system of differential equations governing their motion may be written down as a matrix (see §3.1) whose entries give the *couplings* between the different masses. Diagonalizing this matrix *decouples* the equations of motion. The eigenvectors are called *normal modes*, and they describe different regimes of oscillations; meanwhile, the eigenvalues give the frequencies of oscillation in each normal mode.

We now take up two central questions: (1) how do diagonalizable operators behave?, and (2) what kinds of operators are diagonalizable? The second question is more difficult and will culminate in the Spectral Theorem, so we'll focus on the first for now. Let $T\colon V \to V$ be a diagonalizable operator with eigenbasis $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ and corresponding eigenvalues $\lambda_1, ..., \lambda_n$. Our first step will be to consider the λ_i. This yields a surprising result:

Proposition 4.6. *T-eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_k$ with distinct eigenvalues are independent.*

Proof. Consider the case of two vectors: if $T(\mathbf{v}_1) = \lambda_1\mathbf{v}_1$ and $T(\mathbf{v}_2) = \lambda_2\mathbf{v}_2$ for $\lambda_1 \neq \lambda_2$, then T stretches $\mathbf{v}_1$ and $\mathbf{v}_2$ by different factors. This would be impossible if $\mathbf{v}_1$ and $\mathbf{v}_2$ had been on the same line, so they must be independent. The general case follows by induction, but the argument is still geometrical.[3] Suppose that only the first $j < k$ eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_j$ are independent, so that $\mathbf{v}_{j+1} := \sum_{i=1}^{j} c_i\mathbf{v}_i$ lies in their span. Geometrically, $\mathbf{v}_{j+1}$ is the vertex opposite $\mathbf{0}$ of a parallelepiped P defined by the c_i. When we apply T to P, each side will stretch differently because the λ_i are distinct, so the long diagonal $\mathbf{v}_{j+1}$ shears and changes directions under T. But this contradicts the assumption that $\mathbf{v}_{j+1}$ is an eigenvector! Thus $\{\mathbf{v}_1, ..., \mathbf{v}_{j+1}\}$ must be an independent set. We repeat the same argument for each of the remaining eigenvectors $\mathbf{v}_{j+1}, ..., \mathbf{v}_k$ until we find that all k of them are in fact independent. ■

So if T has n distinct eigenvalues, we automatically get an eigenbasis for V, and hence T is diagonalizable. But what if two different eigenvectors share the same eigenvalue? Nothing prevents them from remaining independent, but their eigenvalues no longer guarantee this: we will have to keep track of the eigenvectors by hand.

Definition 4.7. The *λ-eigenspace* V_λ of T is the set of all the eigenvectors of T with the same eigenvalue: $V_\lambda := \{\mathbf{0}\} \cup \{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda\mathbf{v}\}$.

Each V_λ is a subspace of V, and vectors from different eigenspaces are independent. If T has n distinct eigenvalues, then each eigenspace is one-dimensional, and we can construct an eigenbasis for V using one eigenvector from each V_λ. More generally, we can choose a basis β_λ for each V_λ. Their union $\beta = \cup_{i=1}^{k}\beta_{\lambda_i}$ is guaranteed to be independent, and β forms an eigenbasis for V as long as it contains enough vectors:

Proposition 4.8. *A linear operator $T\colon V \to V$ is diagonalizable iff $\sum_{i=1}^{k} \dim V_{\lambda_i} = n$.*

Example 4.9. In quantum mechanics, the *state* of a physical system is a vector (ormally, a ray) in a complex vector space $\mathcal{H}$, and various *observables* (e.g. position, momentum, spin, etc.) are represented by diagonalizable[4] operators on $\mathcal{H}$. The *measurement* of an observable always gives one of its eigenvalues, and the system "collapses" to an eigenvector with the measured eigenvalue. The set of an operator's eigenvalues is called its *spectrum*; the spectrum is *degenerate* if multiple independent eigenvectors share an eigenvalue. In this case a measurement does not uniquely determine the state to which the system collapses. One must be content with knowing that the state lies in some (possibly large) eigenspace. Alternatively, one can add a *perturbation* that "lifts the degeneracy" by changing the eigenvalues so that they become distinct... but that is a story for a different time.

[3] Credit for this beautiful proof goes to Vedran Sego in `this` StackExchange post.

[4] "Diagonalizable" is too broad: the correct term is *hermitian*, but we haven't met hermitian operators yet, and the only substantive difference is that (as we'll prove!) hermitian operators must have real eigenvalues.

5 Direct-Sum Decomposition

We know that every $\mathbf{v} \in V$ can be decomposed along a basis for V: this is a rigid decomposition of V as the "sum" of its basis directions. We can obtain a looser, more general decomposition by introducing a diagonalizable operator T, which breaks V into independent eigenspaces V_λ. Then any $\mathbf{v} \in V$ can be written as a sum $\mathbf{v} = \sum_{i=1}^{n} \mathbf{w}_i$ of suitably chosen independent eigenvectors, one from each eigenspace. We have a handy word for this:

Definition 5.1. V is the *direct sum* of subspaces W_1 and W_2, written $V = W_1 \oplus W_2$, if $W_1 \cap W_2 = \{\mathbf{0}\}$ and every $\mathbf{v} \in V$ is uniquely written $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$, with $\mathbf{w}_1 \in W_1$, $\mathbf{w}_2 \in W_2$.

The subspaces W act like large, amorphous blobs generalizing the spans of basis vectors. They also describe our situation: under the action of T, the space decomposes into the direct sum of its eigenspaces. Even better, T acts on each eigenspace V_λ simply by scaling the vectors there. Each V_λ is like Las Vegas: vectors in V_λ stay there under the action of T.

Definition 5.2. A subspace $W \subset V$ is *invariant* under T if $T(\mathbf{w}) \in W$ for all $\mathbf{w} \in W$.

Not only are the eigenspaces all T-invariant, but within each V_λ, T acts by $T(\mathbf{w}) = \lambda\mathbf{w}$. We can describe what's going on more formally in the language of projections.

Definition 5.3. A *projection* is a linear map $P\colon V \to V$ satisfying $P^2 = P$.[5] Given a decomposition $V = W_1 \oplus W_2$, the projection along W_2 onto W_1 is defined by first writing $\mathbf{v}$ uniquely as $\mathbf{w}_1 + \mathbf{w}_2$, and then setting $P_1(\mathbf{v}) = \mathbf{w}_1$.

Example 5.4. In the Cartesian plane $V = \mathbb{R}^2$, the vector $\mathbf{v} = (a, b) = a\hat{\mathbf{x}} + b\hat{\mathbf{y}}$ may be projected along the x and y axes by $P_{\hat{\mathbf{x}}}(\mathbf{v}) = (a, 0) = a\hat{\mathbf{x}}$ and $P_{\hat{\mathbf{y}}}(\mathbf{v}) = (0, b) = b\hat{\mathbf{y}}$.

Example 5.5. In quantum mechanics, projection operators perform measurements. When a physical observable O on $\mathcal{H}$ is measured to have eigenvalue λ, the state of the system is projected onto the eigenspace $\mathcal{H}_\lambda$. Having collapsed, the "prepared" state continues to lie within $\mathcal{H}_\lambda$, and every subsequent measurement of O will yield λ.

As an exercise, prove that with $V = W_1 \oplus W_2$ as before, $\mathrm{im}(P_1) = W_1$ and $\ker(P_1) = W_2$; moreover, P_1 is the identity on W_1. Also, show that projections are diagonalizable. Anyway, we now have a beautiful story to tell. Each diagonalizable $T\colon V \to V$ decomposes V into a direct sum of k invariant eigenspaces V_λ, and within each it acts by scaling. Therefore T decomposes into a sum of scaled projection operators: indeed, for any $\mathbf{v} \in V$,

$$T(\mathbf{v}) = T\left(\sum_{i=1}^{k} \mathbf{w}_i\right) = \sum_{i=1}^{k} T(\mathbf{w}_i) = \sum_{i=1}^{k} \lambda\mathbf{w}_i. \tag{5.1}$$

Theorem 5.6 (Primary Decomposition Theorem)**.** *Every diagonalizable $T\colon V \to V$ with distinct eigenvalues $\lambda_1, ..., \lambda_k \in F$ yields the decompositions*

$$V = \bigoplus_{i=1}^{k} V_{\lambda_i}, \qquad T = \sum_{i=1}^{k} \lambda_i P_i. \tag{5.2}$$

[5] A concise, poetic explanation is due to G. Bush: "Fool me twice—you can't get fooled again!"

Part III
The Spectral Theorem

So far, we've introduced the basic apparatus of linear algebra and understood how diagonalizable operators behave. To actually characterize these operators, we'll need to describe the goemetry of vector spaces. After discussing the inner product and norm, we'll introduce adjoints by framing them in the context of linear functionals. Finally, we will conclude with the celebrated[6] Spectral Theorem, which proves that self-adjoint operators are diagonalizable.

6 Geometry and Structure

From now on, we restrict our discussion to real and complex vector spaces, i.e. those whose field of scalars is either $F = \mathbb{R}$ or $\mathbb{C}$. Throughout, the real Euclidean space $V = \mathbb{R}^n$ beautifully illustrates the definitions we introduce.

Definition 6.1. An *inner product* or *symmetric bilinear form* on a complex vector space V is a map $\langle \cdot, \cdot \rangle \colon V \times V \to F$ satisfying, for all $a, b \in F$ and all $\mathbf{u}, \mathbf{v} \in V$,

(a) Linearity: $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$;
(b) Conjugate-symmetry: $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$;
(c) Positive-definiteness: $\mathbf{v} \neq \mathbf{0} \implies \langle \mathbf{v}, \mathbf{v} \rangle > 0$.

Definition 6.2. The *norm* $\|\mathbf{v}\|$ of a vector $\mathbf{v} \in V$ is given by $\|\mathbf{v}\|^2 := \langle \mathbf{v}, \mathbf{v} \rangle$.

The inner product is a natural generalization of the dot product (indeed, the usual properties of dot products can be derived from the axioms above), and describes the projection of one vector along another: in fact, the abstract definition of the projection above can be shown to satisfy $P_{\mathbf{w}}(\mathbf{v}) = \langle \mathbf{v}, \mathbf{w} \rangle \mathbf{w}$. In other words, inner products tell us how much of one vector can be described by another: they relate *angles* between two vectors. Meanwhile, the norm is the amount of one vector along itself and therefore gives its length.

Definition 6.3. Two vectors $\mathbf{u}, \mathbf{v} \in V$ are said to be *orthogonal* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Orthogonal vectors should be thought of as perpendicular or "completely" independent. While a basis for V need not be orthogonal, we can always make it so by subtracting the projection of one basis vector along the others; this removes the components of the other vectors along the first. This process, called the *Gram-Schmidt procedure*, can be carried out inductively until the entire basis is orthogonal, at which point we can normalize it by rescaling all of the new basis vectors to unit length. We therefore say that

Theorem 6.4. *Every finite-dimensional vector space V has an* orthonormal basis*, i.e. a basis $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ where $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$.*

Example 6.5. In quantum mechanics, measuring an observable in some state $\mathbf{v} \in \mathcal{H}$ collapses that state to an eigenvector $\mathbf{v}_\lambda$ of O and yields the eigenvalue λ. The probability of finding λ is controlled by "how close" $\mathbf{v}$ is to $\mathbf{v}_\lambda$: $P(\lambda) = |\langle \mathbf{v}, \mathbf{v}_\lambda \rangle|^2$. If $\mathbf{v} = \mathbf{v}_\lambda$, then $P(\lambda) = \|\mathbf{v}\|^2$ must be unity; thus we say that "quantum states are normalized."

[6] We do not know who celebrated it, because we were not invited.

Orthogonality provides another lens through which to understand linear structure. Given any subspace $W \subset V$, every vector $\mathbf{v} \in V$ has some components that lie in W and some that lie outside it. That is, every vector can be written $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$, where $\mathbf{w}$ lies entirely in W, while $\mathbf{w}^\perp$ is orthogonal to $\mathbf{w}$.

Definition 6.6. The *orthogonal complement* of a subspace $W \subset V$ is the set of vectors orthogonal to all of W, i.e. $W^\perp := \{\mathbf{v} \in V \mid \langle \mathbf{w}, \mathbf{v} \rangle = 0 \text{ for all } \mathbf{w} \in W\} \subset V$.

Orthogonal complements are subspaces, and the decomposition above can be written as follows:

Proposition 6.7. *For any subspace $W \subset V$, $V = W \oplus W^\perp$ and $\dim V = \dim W + \dim W^\perp$.*

Example 6.8 (Euclidean space $V = \mathbb{R}^n$)**.** The inner product is the familiar dot product, and the norm is given by the Pythagorean theorem:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta); \qquad \|\mathbf{v}\|^2 = \sum_{i=1}^{n} v_i^2. \tag{6.1}$$

Any two vectors at right angles to each other are orthogonal, for instance the usual unit vectors $\hat{\mathbf{x}} := (0, 1)$ and $\hat{\mathbf{y}} := (1, 0)$ in the plane. $\mathbb{R}^n$ has a standard coordinate basis $\beta := \{\hat{\mathbf{e}}_1, ..., \hat{\mathbf{e}}_n\}$, where $\hat{\mathbf{e}}_i$ is the unit vector along the i^{th} Cartesian axis and has components $\hat{\mathbf{e}}_i = (0, ..., 0, 1, 0, ..., 0)$. This is an orthonormal basis, and $\mathbb{R}^n$ splits as the direct sum of its coordinate axes: letting $W_i := \text{span}(\hat{\mathbf{e}}_i)$, we have $\mathbb{R}^n = \bigoplus_{i=1}^n W_i$.

7 Linear Functionals and Duality

The notion of an adjoint is often introduced with no motivation and leaves many lost in a whirlwind of confusion. To remedy this, we will look at vector spaces through the mirror of *duality* and exhibit adjoints as the looking-glass versions of linear operators.

Definition 7.1. A *linear functional* on a vector space V is a linear map $\phi \colon V \to F$, i.e. one that satisfies $\phi(a\mathbf{u} + b\mathbf{v}) = a\phi(\mathbf{u}) + b\phi(\mathbf{v})$.

Linear functionals, also called *covectors* or *one-forms*, can be added and scaled just like the vectors they act on. With the observation that the zero functional is well-defined, we see that the linear functionals form a vector space:

Definition 7.2. The *dual space* V^* of a vector space V is the set of linear functionals on V.

Why do we now claim that such a construction is important enough to merit a pompous name? The key lies in the relationship between linear functionals and inner products.

Example 7.3. For $V = \mathbb{R}^3$ with the usual inner product, let $\mathbf{v} = (v_1, v_2, v_3) \in V$ be a vector. Consider the functional $\phi(\mathbf{v}) = 3v_1 + 2v_2 - 4v_3$, and note that we can write $\phi(\mathbf{v})$ as the inner product of $\mathbf{v}$ with the vector of coefficients $\mathbf{u} = (3, 2, -4)$: indeed, $\phi(\mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle$.

Notice that ϕ above is linear because the coordinates v_i of its arugment enter linearly into its definition. This requirement is fundamental, and is the reason that ϕ looks like half of an inner product, ϕ "=" $\langle \mathbf{u}, \cdot \rangle$. In fact, *every* ϕ looks like this, and there is always some vector $\mathbf{u}$ with coefficients that dot into $\mathbf{v}$ to yield the result $\phi(\mathbf{v}) \in F$.

Theorem 7.4 (Riesz Representation Theorem)**.** *For every linear functional $\phi \in V^*$, there is a unique $\mathbf{u} \in V$ such that $\phi(\mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle$ for every $\mathbf{v} \in V$. That is, every ϕ can be uniquely identified with $\mathbf{u}$, so without loss of generality we may label linear functionals by $\phi = \phi_{\mathbf{u}}$.*

The space of linear functionals is dual to the original space in the sense that every functional "is" the vector whose coefficients define its action on the original space; this is why we sometimes call linear functionals covectors. More precisely:

Theorem 7.5. *Every finite-dimensional vector space V is naturally isomorphic to V^*, with the isomorphism $\Phi \colon V \to V^*$ given by $\Phi(\mathbf{u}) = \phi_{\mathbf{u}}$. Moreover, Φ is sesquilinear and isometric:*
(a) $\Phi(a\mathbf{u} + b\mathbf{v}) = \overline{a}\Phi(\mathbf{u}) + \overline{b}\Phi(\mathbf{v})$;
(b) $\|\mathbf{u}\| = \big\|\Phi(\mathbf{u})\big\|$.
Conversely, every isomorphism $\Phi \colon V \to V^*$ gives rise to to a unique inner product on V.

We may also clarify the structure of V^* by specifying a natural *dual basis* for the space. After choosing an orthonormal basis for V, each dual basis vector is defined to be the image under Φ of the corresponding basis vector in V. In this way, each basis functional returns 1 upon eating its progenitor in V, and 0 otherwise.

Proposition 7.6. *Given an orthonormal basis $\beta := \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ for a vector space V, the dual space V^* has a natural basis $\beta^* := \{\phi_1, ..., \phi_n\}$ defined by $\phi_i(\mathbf{v}_j) = \delta_{ij}$.*

Example 7.7. For $V = \mathbb{R}^n$, consider the standard basis $\beta := \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. The dual basis is given by applying the isomorphism: $\Phi(\mathbf{e}_i) = \phi_{\hat{\mathbf{e}}_i}$. To visualize this, observe that applying an arbitrary $\mathbf{v} \in V$ gives $\phi_{\mathbf{e}_i}(\mathbf{v}) = \langle \mathbf{e}_i, \mathbf{v} \rangle = v_i$. That is, $\phi_{\mathbf{e}_i}$ picks out the i^{th} component of $\mathbf{v}$, which is consistent with the identification of $\phi_{\mathbf{e}_i}$ with $\hat{\mathbf{e}}_i$ itself, as described above.

Example 7.8. In quantum mechanics, a notational tool invented by Dirac records vectors as "kets," $\mathbf{v} \mapsto |v\rangle$ and covectors as "bras," $\phi_{\mathbf{u}} \mapsto \langle u|$. This "bra-ket" notation is justified by the one-to-one correspondence between kets and their bras, and can be used in very slick ways. For example, inner products are written $\langle u|v\rangle$, and *outer* products form linear operators by brazenly ignoring the rules: the notation $|v\rangle\langle u|$ indicates a linear operator that acts by $T\,|w\rangle = |v\rangle\langle u|w\rangle$. Upon eating the vector $|w\rangle$, the operator computes its inner product with $\langle u|$ and then redirects the resulting length in the $|v\rangle$ direction.

8 The Spectral Theorem

Having enshrined linear functionals and inner products as central to the study of vector spaces, we will use them to motivate the construction of the adjoint.

Given a linear operator $T\colon V \to V$, consider the expression $\langle \mathbf{u}, T\mathbf{v}\rangle$. As the composition of two linear maps, this expression can be exhibited as $\phi(\mathbf{v})$ for a linear functional $\phi \in V^*$. Now every linear functional can be written $\phi(\mathbf{v}) = \langle \mathbf{w}, \mathbf{v}\rangle$ for some $\mathbf{w} \in V$, so we have $\langle \mathbf{u}, T\mathbf{v}\rangle = \langle \mathbf{w}, \mathbf{v}\rangle$. The operator T seems to have jumped ship from the left-hand side and melded itself into $\mathbf{w}$, so $\mathbf{w}$ must depend on $\mathbf{u}$ and should have something to do with T. Let us therefore define the function $T^*\colon V \to V$ by $T^*(\mathbf{u}) = \mathbf{w}$ and think of $\mathbf{w}$ as the vector obtained from $\mathbf{u}$ by the action of the mysterious function T^*.

Definition 8.1. Given a linear operator $T\colon V \to V$, the *adjoint* of T is the linear operator $T^*\colon V \to V$ defined by $\langle \mathbf{u}, T\mathbf{v}\rangle = \langle T^*\mathbf{u}, \mathbf{v}\rangle$.

Verifying that T^* is in fact a linear operator on V and that it is unique is not hard, but takes a bit of computation. The basic idea is that T^* ferries us from one side of the inner product to the other, and therefore in a sense from V to V^*. Motivation notwithstanding, adjoints remain strange objects, and several basic properties, verified by quick computations using inner products, may help to make them seem less alien.

Proposition 8.2. *The adjoint T^* of an operator $T\colon V \to V$ satisfies the following properties:*
(a) $T^{**} = T$ *and* $I^* = I$*;*
(b) $(T_1 + T_2)^* = T_1^* + T_2^*$*;*
(c) $(cT)^* = \overline{c}T^*$*;*
(d) $(TS)^* = S^*T^*$*;*
(e) $\ker T^* = (\operatorname{im} T)^\perp$*;*
(f) The matrix of T^* is the conjugate transpose of the matrix of T: $T^*_{ij} = \overline{T_{ji}}$.

8.1 A Zoo of Operators

Up until now, the broad class of all linear maps has sufficed for our purposes. But with the additional structure imposed by inner products and adjoints, it makes sense to define more "rigid" operators that respect both of these notions.

Definition 8.3. A linear operator $U\colon V \to V$ is called *unitary* if it preserves the inner product, i.e. for all $\mathbf{u}, \mathbf{v} \in V$, $\langle U\mathbf{u}, U\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$. One can show that the equivalent formulation in terms of adjoints is the requirement $U^*U = UU^* = I$.

The name "unitary" is usually reserved for complex vector spaces, while for real vector spaces we call such transformations "orthogonal." In preserving the inner product, unitary operators are geometrically rigid: they are *isometries* (distance-preserving maps), and can be thought of as rotations of space. In fact, the change-of-basis maps of §2.3 were secretly unitary, and in this context we speak of the similarity transformation $A' = B^{-1}AB$ by saying that A is unitarily diagonalized by B. Unitarity leads to several cool results:

Proposition 8.4. *Every eigenvalue of a unitary operator $U: V \to V$ has unit norm (hence the name!), and eigenvectors of U corresponding to distinct eigenvaluesare orthogonal.*

Proof (Griffiths). Let $\mathbf{v} \in V$ be an eigenvector of U with eigenvalue $\lambda \in \mathbb{C}$. We then have

$$\begin{aligned} \langle U\mathbf{v}, U\mathbf{v}\rangle &= \langle \lambda\mathbf{v}, \lambda\mathbf{v}\rangle = |\lambda|^2\|\mathbf{v}\|^2; \\ \langle U\mathbf{v}, U\mathbf{v}\rangle &= \langle U^*U\mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{v}\rangle = \|\mathbf{v}\|^2. \end{aligned} \tag{8.1}$$

Therefore $|\lambda|^2\|\mathbf{v}\|^2 = \|\mathbf{v}\|^2$, so $|\lambda|^2 = 1$; hence λ lies on the unit circle in $\mathbb{C}$.

Next, suppose that $\mathbf{u}$ is another U-eigenvector with eigenvalue $\mu \neq \lambda$. We then have

$$\begin{aligned} \langle \mathbf{u}, U\mathbf{v}\rangle &= \langle \mathbf{u}, \lambda\mathbf{v}\rangle = \lambda\langle \mathbf{u}, \mathbf{v}\rangle; \\ \langle \mathbf{u}, U\mathbf{v}\rangle &= \langle U^*\mathbf{u}, \mathbf{v}\rangle = \mu\langle \mathbf{u}, \mathbf{v}\rangle, \end{aligned} \tag{8.2}$$

where in the last line we have used the fact that U^* has the same eigenvectors as U and complex-conjugate eigenvalues to those of U. Therefore $\lambda\langle \mathbf{u}, \mathbf{v}\rangle = \mu\langle \mathbf{u}, \mathbf{v}\rangle$, and since $\lambda \neq \mu$ we have $\langle \mathbf{u}, \mathbf{v}\rangle = 0$, proving that U-eigenvectors with distinct eigenvalues are orthogonal. ■

Example 8.5. In quantum mechanics, recall that the probabilities of measurements are related to the inner product between a vector and the observable's eigenvectors. Because these probabilities must sum to unity, quantum states are required to always have unit norm. Their dynamics in time must therefore be carried out by unitary operators, whose norm preservation guarantees the conservation of probability.

We now discuss the operators respecting adjoints rather than inner products: while geometrically these may seem more mysterious, they embody an austere symmetry between V and V^* and to boot are algebraically a bit nicer, as we will see.

Definition 8.6. A linear operator $H: V \to V$ is *self-adjoint* or *hermitian* if $T^* = T$.[7] In terms of inner products, it is equivalent to require that for all $\mathbf{u}, \mathbf{v} \in V$, $\langle T\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, T\mathbf{v}\rangle$.

Again there is a menagerie of names: "hermitian" is widely used for self-adjoint operators on complex vector spaces, while "symmetric" is taken up in the real case. This is because symmetric matricies satisfy $T_{ij} = T_{ji}$, i.e. they are symmetric about the diagonal, and equal their own transposes. In the hermitian case, $T_{ij} = \overline{T_{ji}}$. Much like their unitary counterparts, hermitian operators have simple eigenproperties that make them a joy to study.

Proposition 8.7. *Every eigenvalue of a hermitian operator $H: V \to V$ is real, and eigenvectors of H corresponding to distinct eigenvalues are orthogonal.*

Proof (Griffiths). Let $\mathbf{v} \in V$ be an eigenvector of H with eigenvalue $\lambda \in \mathbb{C}$. We then have

$$\begin{aligned} \langle \mathbf{v}, H\mathbf{v}\rangle &= \langle \mathbf{u}, \lambda\mathbf{v}\rangle = \lambda\|\mathbf{v}\|^2; \\ \langle \mathbf{v}, H\mathbf{v}\rangle &= \langle H\mathbf{v}, \mathbf{v}\rangle = \overline{\lambda}\|\mathbf{v}\|^2. \end{aligned} \tag{8.3}$$

Therefore $\lambda\|\mathbf{v}\|^2 = \overline{\lambda}\|\mathbf{v}\|^2$, so $\lambda = \overline{\lambda}$, and so λ is real.

[7] In infinite-dimensional spaces, there is a subtle distinction between hermitian and self-adjoint operators.

Next, suppose that $\mathbf{u}$ is another H-eigenvector with eigenvalue $\mu \neq \lambda$. We then have

$$\begin{aligned} \langle \mathbf{u}, U\mathbf{v} \rangle &= \langle \mathbf{u}, \lambda \mathbf{v} \rangle = \lambda \langle \mathbf{u}, \mathbf{v} \rangle; \\ \langle \mathbf{u}, U\mathbf{v} \rangle &= \langle U\mathbf{u}, \mathbf{v} \rangle = \mu \langle \mathbf{u}, \mathbf{v} \rangle, \end{aligned} \tag{8.4}$$

where we have used the fact that $\mu \in \mathbb{R}$. Therefore $\lambda \langle \mathbf{u}, \mathbf{v} \rangle = \mu \langle \mathbf{u}, \mathbf{v} \rangle$, and since $\lambda \neq \mu$ we have $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, proving that H-eigenvectors with distinct eigenvalues are orthogonal. ■

Example 8.8. In quantum mechanics, the outcomes of measurements must be real! This is a strong hint that observables should be represented by hermitian operators. Prove that this must be the case by recalling that measurements are performed by projection operators, and that observables must contain information about every possible measurement value.

8.2 The Stormy Finale

Let's pause to meditate on our place in the universe. We live in a vector space of additive-scalable objects roamed by linear maps that shear and distort vectors. Maps that carry all of a space's linear structure over to another space are isomorphisms that identify the spaces; other maps fail to be injective or surjective, maiming their images or swelling their kernels. Both vectors and operators can be expressed in coordinates dependent on a choice of basis, and change-of-basis maps guide us through the rugged terrain here. (Fortunately, every basis may be made orthonormal.) Enter eigenvectors, special objects that don't change direction under the action of an operator; we're on the hunt for operators with enough eigenvectors to form a basis for the space. These diagonalizable operators are complicated: they break the space into a direct sum of invariant eigenspaces, and they decompose into a sum of scaled projections. In other words, their action is given by a linear combination of their arguments' projections onto each eigenspace, weighted by their corresponding eigenvalues. There also live among us bilinear, symmetric maps called inner products and linear functionals; through the mirror of duality, they both essentially look like vectors. And from the boiling cauldron of inner products and the dark cloud of the adjoint, there bubble up and rain down upon us strange and beautiful operators, both unitary and hermitian.

To this foreboding scene we will add one fact, unproven but true, and thence, amid lightning and thunder, we will prove our main result.

Fact 8.9. *Every operator on a complex vector space has an eigenvector.*

This is emphatically false in real vector spaces: a rotation operator in the plane has no eigenvectors at all. Various proofs of this fact over $\mathbb{C}$ are available, some using determinants and some not. Fundamentally, it holds because $\mathbb{C}$ is an *algebraically closed* field.

But enough—the time has come.

Theorem 8.10 (Spectral theorem)**.** *Every hermitian operator $T\colon V \to V$ on a finite-dimensional vector space admits an orthonormal basis of T-eigenvectors for V.*

Proof. We proceed by induction on $n = \dim V$. The base case $n = 0$ is vacuous because a 0-dimensional vector space is just a single point. For $n = 1$, V is a line and is spanned by any nonzero vector. Since T has an eigenvector $\mathbf{v}$, we have $V = \text{span}(\{\mathbf{v}\})$, and $\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ is an orthonormal basis for V. Next, we assume that the theorem holds for some $n = k \in \mathbb{N}$ and set $\dim V = k + 1$. Once again, T has an eigenvector $\mathbf{v}$ with eigenvalue λ; consider the 1-dimensional subspace $W := \text{span}\,\{\mathbf{v}\} \subset V$. Since we can always write $V = W \oplus W^\perp$, we have $\dim W^\perp = k$. (We will eventually use the inductive hypothesis on $W^\perp$.)

Next we show that $W^\perp$ is T-invariant. This is due to hermiticity: for any $\mathbf{w} \in W^\perp$,

$$\langle \mathbf{v}, T\mathbf{w} \rangle = \langle T\mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle = 0, \tag{8.5}$$

where we have used that $\mathbf{v} \in W$ and $\mathbf{w} \in W^\perp$ to ensure that $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. We have just shown that applying T to any $\mathbf{w} \in W^\perp$ leaves it orthogonal to W; in other words, the image $T(W^\perp)$ is contained within $W^\perp$, so by definition $W^\perp$ is T-invariant.

We now restrict T to this invariant subspace, observing that $T\big|_{W^\perp}\colon W^\perp \to W^\perp$ is a hermitian operator on the k-dimensional vector space $W^\perp$. By the inductive hypothesis, the restricted map admits an orthonormal basis $\{\mathbf{u}_1, ..., \mathbf{u}_k\}$ of T-eigenvectors for $W^\perp$. All of these eigenvectors are orthogonal to the original eigenvector $\mathbf{v}$ whose span we called W, so define $\mathbf{u}_{k+1} := \frac{\mathbf{v}}{\|\mathbf{v}\|}$ to scale it down to unit length. By construction, the set $\beta := \{\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{u}_{k+1}\}$ furnishes an orthonormal basis of T-eigenvectors for $W \oplus W^\perp = V$. ∎

A similar but much trickier version of this argument (omitted) demonstrates a more general fact, which encompasses both unitary and hermitian operators:

Definition 8.11. A linear operator $T\colon V \to V$ is called *normal* if $TT^* = T^*T$.

Theorem 8.12. *A linear operator $T\colon V \to V$ on a finite-dimensional vector space is unitarily diagonalizable (i.e. V has an orthonormal basis of T-eigenvectors) iff T is normal.*

A Textbooks and Further Reading

Linear algebra cannot truly be done justice without a good textbook, a pad of paper, a pencil, a large eraser, and a wastebasket. Unfortunately linear algebra textbooks that belong *in* the wastebasket abound, and only a select few rise to the task of an elegant, thorough, intuitive, and motivated exposition. One text is not enough; authors differ in opinion about what should be emphasized. Below are listed a few of the texts I liked:

Paul Halmos, "Finite-Dimensional Vector Spaces."

This is *the* linear algebra textbook: it's concise, beautifully written, and develops the theory naturally and with an eye towards generalizations to the infinite-dimensional setting, which today is known as functional analysis. The book packs a punch: it treats topics deeply and once, rather than making several passes at the same material like modern texts.

Hoffman and Kunze, "Linear Algebra" (Second Edition).

This dry, exhaustive text covers everything and does it correctly. It is a very tough first introduction to linear algebra, and unapologetically presents an algebraist's take on the subject. Be wary, ye faint of heart. But read and read again; read and weep; read and love.

Sheldon Axler, "Linear Algebra Done Right" (Third Edition).

I am ambivalent on this one. Axler wrote the book out of a nostalgia for elegance and a hatred for determinants. In my opinion he prodced a decent book riddled with bad notation that nevertheless treats the theory in some generality. Axler's writing is lively and written for the modern audience, but doesn't entirely achieve his aims.

Sergei Treil, "Linear Algebra Done Wrong."

Better than "Done Right", and a deeper book. Although I am not intimately familiar with this text, it seems to accomplish what Axler tried to, minus the unfriendliness towards determinants, minus bad notation, but plus some advanced topics and plus less-than-clear writing in a few places. Overall an excellent introduction.

Barton Zweibach, MIT lecture notes on Quantum Mechanics II, especially Chapter 3.

These lecture notes contain a brief yet surprisingly clear account of many of the topics covered by the present set of notes. Zweibach follows Axler's treatment for much of his linear algebra review, but departs on key points mostly related to physics. As a supplement to a quantum mechanics course, the emphasis is of course on physical intuition rather than rigorous proof; while this may annoy some, I admire the additional clarity.

The Internet, in particular Wikipedia and Math Stack Exchange.

The Internet is always here for you. Look things up; explore, and so on. Email me at `davidgrabovsky@physics.ucsb.edu` with questions on the typos and unexplained notational quirks I've inevitably missed, or on (mostly) anything else.