1 Some Basics

Sets and Lists. For us, a set will be informally thought of a collection of objects. For example, the set $\{0, 1\}$ contains the numbers 0 and 1, the set $\mathbb{R}$ is the set of real numbers, and the set $\mathbb{C}$ is the set of complex numbers. A finite set containing objects $x_1, \ldots, x_n$ will be denoted $\{x_1, \ldots, x_n\}$. The notation $(x_1, \ldots, x_n)$ will denote the set $\{x_1, \ldots, x_n\}$ plus the ordering x_1 first, then x_2, then x_3, etc. For example, $\{x_1, x_2\} = \{x_2, x_1\}$ but $(x_1, x_2) \neq (x_2, x_1)$. $(x_1, \ldots, x_n)$ will be called a "list" or "n-tuple" instead of a "set".

Given sets X and Y, a function $f : X \to Y$ is an assignment of an element of Y to every element in X. The words "map" and "transformation" are synonyms for "function". If $f : X \to Y$ is a function, then the statement $x \mapsto y$ is read "x is mapped to y" and means "$f(x) = y$".

The set $X \times Y$ is defined to be the set of lists (x, y) where $x \in X$ and $y \in Y$.

Symbols and Terminology. It is helpful to recall some common terminology about proofs. A implies B means that if A is true, then B is true. For example, in the real numbers, $x > 1$ implies that $x > 0$. The symbol $\forall$ means "for all", the symbol $\exists$ means "there exists", the symbol $\in$ means "in" of "is an element of", and the letters s.t. mean "such that". For example, the sentence "for each real number, there exists a larger real number" might be written $\forall x \in \mathbb{R},\ \exists y \in \mathbb{R}$ s.t. $y > x$.

Boolean Operations. When a mathematician says "A or B is true", they mean that either A is true, or B is true, or both A and B are true. This somewhat contradicts the common everyday usage of "or", which excludes the latter case. When a mathematician says "A and B are true", they mean that both A and B are true.

Negation. Given a statement A, its negation, denoted $\neg A$, is the statement which is true if A is false and false if A is true. For example, the negation of the statement "the sun will rise tomorrow" is the statement "the sun will not rise tomorrow". The negation of the statement "P is true $\forall x$" is "$\exists x$ such that P is false". For example, the negation of "all apples are red" is "there exists an apple that is not red." Similarly, the negation of "$\exists x$ such that P is true" is "P is false $\forall x$". Another example: the negation of the statement"there exists an even prime number" is "all prime numbers are odd". Negation flips "and" and "or". Namely, $\neg(A \text{ and } B) = (\neg A \text{ or } \neg B)$ and $\neg(A \text{ or } B) = (\neg A \text{ and } \neg B)$. For example, the negation of "p is prime and p is bigger than 2" is "p is not prime or p is less than or equal to 2".

The Contrapositive. If A implies B, then the negation of B implies the negation of A. For if A being true implies that B is true, then if B is false then A cannot be true hence must be false. If A implies B, then the (true) statement "$\neg B$ implies $\neg A$" is called "the contrapositive". Suppose you wanted to prove that "p is prime and p is bigger than 2 implies that p is odd". You could

instead prove the contrapositive: "p is even implies that p is not prime or p is at most 2." This second statement might be easier to reason with, as even numbers are defined to be those divisible by 2.

2 Vector Spaces

Definition 1 (Informal)**.** A vector space is a set V for which sums of elements are defined, as well as multiples of elements.

So if $v, w \in V$, then $v + w$ is also in V, as is $5v$, as is $3v$, as is $3.14v$, as is $5v + 3w$.

Example 2. A good example to keep in mind is the space of real-valued functions on a set X. If f and g are functions then so are $f + g$ and so is $3.14f$ and so is $5f + 3g$. In case it is not clear, $(f + g)(x) := f(x) + g(x)$ and $(cf)(x) := cf(x)$.

You can also multiply functions together, but multiplication of elements is not required as part of a vector space.

Example 3. A simpler example is the set of lists of n real numbers:

$$(a_1, \ldots, a_n),\ a_i \in \mathbb{R}.$$

For example, take $n = 2$. Then

$$(a_1, a_2) + (b_1, b_2) := (a_1 + a_2, b_1 + b_2)$$

and

$$5(a_1, a_2) := (5a_1, 5a_2).$$

Definition 4. Given V and vectors $v_1, \ldots, v_n \in V$, the element

$$a_1 v_1 + \cdots + a_n v_n$$

is called a "linear combination" of the v_is.

Remark 5. A linear combination is a sum with *finitely* many terms.

Definition 6. (Formal) A vector space V over $\mathbb{R}$ is a set together with a function $V \times V \to V$ and a function $\mathbb{R} \times V \to V$ generally denoted

$$(v, w) \mapsto v + w,\ v, w \in V$$

$$(c, v) \mapsto cv,\ c \in \mathbb{R},\ v \in V$$

that satisfy

$$v + w = w + v$$

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

$$\exists 0 \in V \text{ such that } 0 + v = v$$

$$\forall v \in V,\ \exists -v \in V \text{ such that } v + (-v) = 0$$

$$1v = v$$

$$a(bv) = (ab)v$$

$$(a + b)(v + w) = av + aw + bv + bw.$$

This axiomatizes the first informal definition of "a set where you can add things together and also multiply things by numbers." Of course there are different collections of "numbers." Here I chose the real numbers and I called V a "vector space over $\mathbb{R}$." One could also replace the real numbers by the complex numbers and get the definition for a "vector space over $\mathbb{C}$." I will use $\mathbb{F}$ to denote either $\mathbb{R}$ or $\mathbb{C}$.

Vector Spaces Over $\mathbb{Z}$?. These notes concern themselves with vector spaces over $\mathbb{R}$ and $\mathbb{C}$. One could also replace $\mathbb{R}$ by the integers $\mathbb{Z}$. For whatever reason, a "vector space over $\mathbb{Z}$" is not called a vector space but rather is called a "module over $\mathbb{Z}$." The reader may take this fastidiousness of language as an idiosyncrisy of mathematics. Modules over $\mathbb{Z}$, however, turn out to be quite different from vector spaces over $\mathbb{R}$ and $\mathbb{C}$, justifying both the differentiation in terminology and their absence from these notes.

Proposition 7. *Here are some properties of vector spaces that can be deduced from the axioms.*

- $0v = 0$
- *For each $v \in V$ there exists a unique additive inverse w such that $v+w = 0$*
- $(-1)v = -v$

Proof. $0v = (0 + 0)v = 0v + 0v$. Adding $-0v$ to each side shows that $0v = 0$.

The existence of an additive inverse is one of the axioms of a vector space. To show that the additive inverse is unique, let w be an additive inverse to v, so that $v + w = 0$. Add $-v$ to both sides so that $v + w + (-v) = -v$ and since $v + (-v) = 0$ then $w = -v$.

$0 = 0v = (1+(-1))v = 1v+(-1)v = v+(-1)v$ so that $(-1)v$ is an additive inverse to v. Because the additive inverse is unique, it must be $-v$. □

Definition 8. The set of lists of n numbers in $\mathbb{F}$, $(a_1, \ldots, a_n)$ is called $\mathbb{F}^n$. Addition and scalar multiplication are defined componentwise:

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) := (a_1 + b_1, \ldots, a_n + b_n)$$

$$c(a_1, \ldots, a_n) := (ca_1, \ldots, ca_n).$$

Definition 9. Elements of a vector space V are called "vectors." The element $0 \in V$ is often called "the zero vector" or, more simply, "zero".

Convention. For conventional reasons, vectors in $\mathbb{F}^n$ are often denoted vertically:

$$(a_1, \ldots, a_n) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Example 10. The vector $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ in $\mathbb{R}^2$ can be represented by an arrow drawn in the Cartesian plane from the point $(0, 0)$ to the point (a_1, a_2). Addition in $\mathbb{R}^2$ corresponds to appending one arrow onto the end of the other. The property $v + w = w + v$ corresponds, geometrically, to a parallelogram. The zero vector is the arrow pointing from $(0, 0)$ to itself.

Definition 11. A finite collection of vectors $v_1, \ldots v_n \in V$ is said to be linearly dependent if there exists a nonzero element $(a_1, \ldots, a_n) \in \mathbb{F}^n$ such that $a_1 v_1 + \cdots + a_n v_n = 0$.

Note that the a_is in the above definition cannot be all equal to 0.

Example 12. The vectors

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 6 \end{pmatrix}$$

are linearly dependent because

$$-2\begin{pmatrix} 1 \\ 2 \end{pmatrix} - 2\begin{pmatrix} -1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Example 13. The vectors

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

are not linearly dependent because if

$$a_1\begin{pmatrix} 1 \\ 2 \end{pmatrix} + a_2\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

then

$$\begin{pmatrix} a_1 \\ 2a_1 + a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which implies that $a_1 = 0$ and $2a_1 + a_2 = 0$ which implies $a_1 = a_2 = 0$.

Definition 14. An infinite collection of vectors $v_1, v_2, \ldots$ in V is said to be linearly dependent if it contains a finite linearly dependent subset.

Definition 15. A collection of vectors called linearly independent if they are not linearly dependent.

Example 16. Let V be the vector space of real-valued functions on $\mathbb{R}$. The functions $f(x) = x$ and $g(x) = x^2$ are linearly independent, as can be seen as follows. Suppose they were not. Then

$$a_1 f + a_2 g = 0$$

for some real numbers a_1, a_2 such that $(a_1, a_2) \neq (0, 0)$. Then

$$a_1 f(x) + a_2 g(x) = 0 \ \forall x \in \mathbb{R}$$

$$a_1 x + a_2 x^2 = 0 \ \forall x \in \mathbb{R}$$

which is only true when $(a_1, a_2) = (0, 0)$ because for a given value (a_1, a_2), the equation $a_1 x + a_2 x^2 = 0$ will have at most two solutions in x.

Example 17. Let V be the vector space of real-valued functions on $\mathbb{R}$. Then functions $f_n(x) = x^n$ form an infinite linearly independent set of vectors.

Definition 18 (Informal)**.** A subspace of a vector space is a vector space contained in another vector space.

Definition 19 (Formal)**.** A subspace W of a vector space V is a nonempty subset $W \subset V$ such that if $w_1, w_2 \in W$ then $w_1 + w_2 \in W$, and if $w \in W$ then $cw \in W$ for $c \in \mathbb{F}$.

Example 20. Let V be the vector space of all functions from $\mathbb{R}$ to itself. Let W be the space of all functions which send 0 to 0. Then W is a subspace.

Example 21. Let $V = \mathbb{F}^n$ and let $W \subset V$ be the set of lists of n numbers of the form $(a_1, \ldots, a_{n-1}, 0)$. Then W is a subspace of V.

Example 22. Let $V = \mathbb{F}^n$ and let $W \subset V$ be the set of lists of n numbers of the form $(a_1, \ldots, a_{n-1}, 1)$ then W is not a subspace of V.

Example 23. Geometrically speaking, the subspaces of $\mathbb{R}^3$ are: the whole space, the set $\{0\}$, lines passing through 0, and planes passing through 0.

Example 24. Let V be the collection of real-valued functions on $\mathbb{R}$. Polynomial functions form a subspace of V.

Proposition 25. *The intersection of two subspaces is a subspace.*

Proof. Let W_1 and W_2 be two subspaces of V and Let $v, u \in W_1 \cap W_2$. Then since $v \in W_1$ and $u \in W_1$ then $v + u \in W_1$. Simiarly, since $v \in W_2$ and $u \in W_2$, then $v + u \in W_2$. Hence $v + u \in W_1 \cap W_2$. The proof to show that $cv \in W_1 \cap W_2$ for $c \in \mathbb{F}$ is similar. □

Definition 26. Let $\{v_1, v_2, \ldots\}$ be a collection of vectors in V. The span of $\{v_1, v_2, \ldots\}$ is the set of linear combinations of elements of $\{v_1, v_2, \ldots\}$.

Remark 27. A linear combination of elements of $\{v_1, v_2, \ldots\}$ is by definition a finite sum

$$a_1 v_{i_1} + \cdots + a_n v_{i_n}$$

here $i_k \in \{1, 2, \ldots\}$. In particular, infinite sums

$$\sum_{i=1}^{\infty} a_i v_i$$

are linear combinations only if all but finitely many of the a_i are zero.

Proposition 28. *The span of $\{v_1, v_2, \ldots\} \subset V$ is a subspace of V.*

Proof. Let W be the span of $\{v_1, \ldots, v_n\}$. Two vectors in W are of the form

$$a_1 v_{i_1} + \cdots + a_n v_{i_n} \text{ and } b_1 v_{i_1} + \cdots + b_n v_{i_n}$$

Their sum

$$a_1 v_{i_1} + \cdots + a_n v_{i_n} + b_1 v_{i_1} + \cdots + b_n v_{i_n} = (a_1 + b_1) v_{i_1} + \cdots + (a_n + b_n) v_{i_n}$$

is also in W. Similarly

$$c(a_1 v_{i_1} + \cdots + a_n v_{i_n}) = c a_1 v_{i_1} + \cdots + c a_n v_{i_n}$$

is in W. □

Definition 29. An ordered list of vectors $(v_1, v_2, \ldots)$ in V is called a basis if any vector $v \in V$ can be written uniquely as a linear combination

$$v = a_1 v_{i_1} + \cdots + a_n v_{i_n}.$$

Said another way, $(v_1, v_2, \ldots)$ is a basis of V if any vector $v \in V$ can be written as

$$v = a_1 v_{i_1} + \cdots + a_n v_{i_n}$$

and if

$$v = a_1 v_{i_1} + \cdots + a_n v_{i_n} = b_1 v_{i_1} + \cdots + b_n v_{i_n}$$

then

$$a_i = b_i \ \forall i.$$

Note that the basis is not just a set of vectors but an ordered set. For example, the following are two different bases of $\mathbb{R}^2$:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Example 30. Let $e_i \in \mathbb{F}^n$ be $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$, the zero list except for a 1 in the ith position. Then $(e_1, \ldots, e_n)$ is a basis of $\mathbb{F}^n$, sometimes called the "standard basis".

Example 31. Let V be the vector space of all polynomials in the variable x with coefficients in $\mathbb{F}$. Then

$$(1, x, x^2, x^3, \ldots, x^n, \ldots)$$

is a basis of V. Another basis is given by

$$\left(1, x, \frac{x^2}{2!}, \frac{x^3}{3!}, \ldots, \frac{x^n}{n!}, \ldots\right).$$

Proposition 32. *A list of vectors $(v_1, v_2, \ldots)$ in V is a basis for V if and only if the set $\{v_1, v_2, \ldots\}$ is linearly independent and spans V.*

Proof. First I prove that the fact that $(v_1, v_2, \ldots)$ is a basis implies that $\{v_1, v_2, \ldots\}$ is linearly independent and spanning.

So suppose $(v_1, v_2, \ldots)$ is a basis. Suppose for a contradiction that the set $\{v_1, v_2, \ldots\}$ is linearly dependent. Then there would exist some equality

$$a_1 v_{i_1} + \cdots + a_n v_{i_n} = 0$$

where each v_{i_k} is different and not all of the a_is are zero. Here 0 is expressed as two different linear combinations (the nonzero coefficients are not the same on each side). This contradicts the basis assumption that each vector has a unique expression in terms of the v_i. Next observe that the fact that $\{v_1, v_2, \ldots\}$ spans is built into the definition of a basis.

Next I prove that if $\{v_1, v_2, \ldots\}$ is linearly independent and spanning then $(v_1, v_2, \ldots)$ is a basis. Suppose for a contradiction that $(v_1, v_2, \ldots)$ were not a basis. Then either of two things happen:

- There exists $v \in V$ that can be written in two different ways in terms of the basis vector

  $$v = a_1 v_{i_1} + \ldots + a_n v_{i_n} = b_1 v_{i_1} + \ldots + b_n v_{i_n}$$

  with $(a_1, \ldots, a_n) \neq (b_1, \ldots, b_n)$.
- There exists $v \in V$ that cannot be written as $v = a_1 v_{i_1} + \cdots + a_n v_{i_n}$.

The second bullet clearly contradicts the fact that $\{v_1, v_2, \ldots\}$ spans V. The first bullet point implies that

$$0 = (a_1 - b_1) v_{i_1} + \cdots + (a_n - b_n) v_{i_n}$$

where $(a_i - b_i) = 0$ for at least one i. This shows that $v_{i_1}, \ldots, v_{i_n}$ is a linearly dependent subset of V and thus contradicts the linear independence of $\{v_1, v_2, \ldots\}$. □

Definition 33. Given subspaces $W_1, W_2 \subset V$ the sum $W_1 + W_2$ is the span of vectors in W_1 and W_2.

The reader is advised that the next two definitions are two different definitions of the same term. Axler uses the second definition and so we will use that one too, though in my experience the first is more standard.

Definition 34. Let V and W be two vector spaces. The direct sum $V \oplus W$ is the vector space given by the set of ordered pairs

$$(v, w),\ v \in V,\ w \in W$$

with addition

$$(v_1, w_1) + (v_2, w_2) := (v_1 + v_1, w_1 + w_2)$$

and scalar action

$$c(v, w) := (cv, cw)$$

Definition 35. Given V, and two subspaces $W_1, W_2 \subset V$ the sum $W_1 + W_2$ is called a "direct sum" if $W_1 \cap W_2 = 0$.

Remark 36. Axler uses the notation $W_1 \oplus W_2$ to denote the set $W_1 + W_2$ *and* the assumption that it is a direct sum.

Example 37. In the vector space of all polynomials with coefficents in $\mathbb{F}$, the span of monomials of odd degree and the span of monomials of even degree form two subspaces whose sum is a direct sum.

Example 38. In $\mathbb{R}^3$, let W_1 be the span of $(1, 0, 0)$ and W_2 the span of $(0, 1, 0)$. Then $W_1 + W_2$ is a direct sum.

Proposition 39. *If $W_1 + W_2 \subset V$ is a direct sum then any vector $v \in W_1 + W_2$ can be written uniquely as $v = w_1 + w_2$ for $w_1 \in W_1$ and $w_2 \in W_2$.*

Proof. Suppose that $v = w_1 + w_2$ and $v = w_1' + w_2'$ where $w_i, w_i' \in W_i$. Then

$$0 = (w_1 - w_1') + (w_2 - w_2')$$

$$-w_1 + w_1' = w_2 - w_2'.$$

The left-hand side is in W_1 and the right-hand side is in W_2, and since the two sides are equal each side is in $W_1 \cap W_2$. Since $W_1 \cap W_2 = 0$ then $w_1 - w_1' = 0$ and $w_2 - w_2' = 0$ so $w_1 = w_1'$ and $w_2 = w_2'$. □

3 Linear Maps

Definition 40. Let V and W be two vector spaces. A linear map $T : V \to W$ is a map from V to W such that

$$T(v_1 + v_2) = T(v_1) + T(v_2),\ \forall v_1, v_2 \in V$$

$$T(cv) = cT(v),\ \forall c \in \mathbb{F},\ v \in V$$

Linear maps are also called linear transformations.

Note that T is a map that behaves nicely with respect to addition of a vector space: you can add two vectors then apply T, or you can apply T to each vector then add the two resulting vectors, and the result is the same in each case. Similarly you can multiply a vector by a scalar c then apply T, or you can apply T and then multiply the result by c, and you get same thing in each case.

Aside. A vector space is defined as a set V plus some associated operations: addition and scalar multiplication. Said another way, a vector space is not just a set but a set-with-operations. In general, to relate two sets, one uses a map between those two sets. But to relate two sets-with-operations, one has to use only those maps between the two sets that play nice with their operations. The study of sets-with-operations and the related maps between these sets-with-operations is called abstract algebra. Vector spaces are but one example of sets-with-operations. Others include groups, rings, modules, algebras, small categories, and many more.

Example 41. The map

$$\pi : \mathbb{F}^2 \to \mathbb{F}$$
$$\pi(a_1, a_2) = a_1$$

is linear.

Example 42. The map

$$f : \mathbb{F}^2 \to \mathbb{F}$$
$$f(a_1, a_2) = a_1 + 1$$

is not linear. For example

$$f(a_1, a_2) + f(a_1', a_2') = a_1 + a_1' + 2 \neq f(a_1 + a_1', a_2 + a_2') = a_1 + a_1' + 1.$$

Example 43. Differentiation gives a linear map from the vector space of polynomials over $\mathbb{F}$ to itself.

Exercise 44. *Let $T : V \to W$ be a linear map. Then $T(0) = 0$.*

Proposition 45. *Let $(v_1, v_2, \ldots)$ be a basis for V. Let $(w_1, w_2, \ldots)$ be an ordered list of vectors in W of the same length. Then there exists a unique linear map $T : V \to W$ such that $T(v_i) = w_i$.*

Proof. Set $T(v_i) = w_i$. Each $v \in V$ can be written uniquely as $a_1v_1 + \cdots + a_nv_n$ and so, in particular, uniquely determines the a_is. Therefore the definition

$$T(v) := a_1T(v_1) + \cdots + a_nT(v_n)$$

depends only v and not on the particular choice of linear combination. It is easy to check that this definition is linear.

Suppose $S(v_i) = T(v_i)$ and S is linear. Then

$$S(v) = S(a_1v_1 + \cdots + a_nv_n) = a_1S(v_1) + \cdots + a_nT(v_n) = T(v).$$

□

Remark 46. The previous proposition is extremely important. It says that in order to define a linear map you do not need to define where it sends every vector in the vector space. Rather, you only need to define where it sends a basis.

Definition 47. Let $f : X \to Y$ be a map of sets. The image of f is defined to be those points $y \in Y$ such that there exists $x \in X$ with $f(x) = y$. (Note that Axler uses the word "range" instead of "image".)

Definition 48. Let $f : X \to Y$ be a map of sets. If f does not send two different points of X to the same point in Y, f is said to be injective. If the image of f is equal to Y, then f is said to be surjective. If f is both surjective and injective, then f said to be bijective.

Proposition 49. *A map $f : X \to Y$ of sets is bijective if and only if it has an inverse. That this, if and only if there exists a map $f^{-1} : Y \to X$ such that $f^{-1} \circ f$ is the identity map on X, and $f \circ f^{-1}$ is the identity map on Y.*

Proof. Suppose that f is bijective, so that it is both surjective and injective. By surjectivity, for each $y \in Y$ there exists $x \in X$ such that $f(x) = y$. By injectivity, such an x is unique. Define $f^{-1}(y) = x$ and it follows by construction that f^{-1} has the desired properties.

Suppose that f has an inverse f^{-1}. Suppose for a contradiction that f is not surjective. Then there exists $y \in Y$ with no x mapping onto it. Then $f \circ f^{-1}$ does not include y in its image, contradicting the fact that $f \circ f^{-1}$ is the identity on Y. Suppose for a contradiction that f is not injective. Then f sends $x_1, x_2 \in X$ to the same point $y \in Y$. Then $f^{-1} \circ f$ sends x_1 and x_2 to the same point in X, contradicting the fact that $f^{-1} \circ f$ is the identity on X. □

Example 50. Let X and Y be finite sets

- There exists a injection from X into Y if and only if Y has at least as many elements as X.
- There exists a surjection from X onto Y if and only if X has at least as many elements as Y.
- There exists a bijection between X and Y if and only X and Y have the same number of elements.

Thus given sets X and Y, it is inuitive to think that X and Y are the "same size" if there exists a bijection between them. Similarly, one should think of two vector spaces as the "same size" if there exists a linear bijection between them.

Definition 51. A linear bijection is called an isomorphism. If there exists an isomorphism $T : V \to W$, then V and W are said to be isomorphic.

Proposition 52. *A linear map $T : V \to W$ is injective if and only if $T(v) = 0$ implies that $v = 0$.*

Proof. Suppose T injective. Then at most one point can map to 0. Since $T(0) = 0$ it follows that $T(v) = 0 \Rightarrow v = 0$.

Suppose that $T(v) = 0 \Rightarrow v = 0$. Let v_1 and v_2 map to the same point, so $T(v_1) = T(v_2)$. Then $T(v_1) - T(v_2) = 0 \Rightarrow T(v_1 - v_2) = 0 \Rightarrow v_1 - v_2 = 0$, so $v_1 = v_2$. Hence T is injective. □

Exercise 53. *Let $T : V \to W$ be an isomorphism of vector spaces. Let T^{-1} be an inverse. Show that T^{-1} is a linear map.*

Proposition 54. *Let $T : V \to W$ be an isomorphism. If $(v_1, v_2, \ldots)$ is a basis for V, then $(T(v_1), T(v_2), \ldots)$ is a basis for W.*

Proof. Linear independence of $\{T(v_1), T(v_2), \ldots\}$ follows from injectivity of T. That $\{T(v_1), T(v_2), \ldots\}$ spans follows from surjectivity of T. Details are left to the reader. □

A corollary of this proposition is the important

Corollary 55. *Two vector spaces are isomorphic if and only if there they have bases of the same size.*

Remark 56. If the bases are finite, the phrase "bases of the same size" means exactly what you'd think. If the bases are not finite, the phrase means "there exists a bijection between the two bases". It is a nontrivial but interesting fact that there exist different infinite sets that are not in bijection. For example, there is no bijection between the integers $\mathbb{Z}$ and the real numbers $\mathbb{R}$.

Definition 57. The dimension of a vector space V, $\dim(V)$, is the number of elements in a basis of V.

Because of the corollary, the dimension does not depend on the particular basis for V.

Corollary 58. *A vector space V of dimension n is isomorphic to $\mathbb{F}^n$.*

For $\dim(V) = n$, an explicit isomorphism $T : V \to \mathbb{F}^n$ is given by first picking a basis $(v_1, \ldots, v_n)$ for V and defining $T(v_i) = e_i$, where e_i are the standard basis vectors for $\mathbb{F}^n$.

Matrices. Suppose $T : V \to W$ is a linear map and $(v_1, \ldots, v_n)$ is a basis for V and $(w_1, \ldots, w_m)$ is a basis for W. Since T is determined by the vectors $T(v_i)$ and each $T(v_i)$ can be written uniquely in terms of the w_i, T can be encoded into nm pieces of information. Namely, if

$$T(v_i) = M_{1i}w_1 + M_{2i}w_2 + \cdots + M_{mi}w_m$$

and if

$$v = a_1v_1 + \cdots + a_nv_n$$

then

$$T(v) = a_1 T(v_1) + \cdots + a_n T(v_n)$$

$$= a_1 \left(M_{11} w_1 + M_{21} w_2 + \cdots + M_{m1} w_m\right) + \cdots + a_n \left(M_{1n} w_1 + M_{2n} w_2 + \cdots + M_{mn} w_m\right)$$

$$= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} M_{ji} a_i \right) w_j$$

so that T is determined by the nm numbers M_{ji} for $1 \leq j \leq m$ and $1 \leq i \leq n$. Note that the formula here is precisely that of matrix multiplication:

$$\begin{pmatrix} M_{11} & M_{12} & M_{13} & \cdots & M_{1n} \\ M_{21} & M_{22} & M_{23} & \cdots & M_{2n} \\ M_{31} & M_{32} & M_{33} & \cdots & M_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M_{m1} & M_{m2} & M_{m3} & \cdots & M_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix}$$

where the jth entry of the resulting $m \times 1$ matrix is the coefficient of w_j.

Said another way, given the bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of V and W, a linear map $T : V \to W$ can be encoded in the $m \times n$ matrix with entries M_{ji} where M_{ji} is the coefficient of w_j in the expansion of $T(v_i)$ in terms of the basis $(w_1, \ldots, w_m)$.

If $T : V \to W$ and $S : W \to U$ are linear maps, then their composition is a linear map $S \circ T : V \to U$. Let $(v_1, \ldots, v_n)$, $(w_1, \ldots, w_m)$, and $(u_1, \ldots, u_p)$ be bases of V, W, and U, respectively. With respect to these bases, let T be represented by the $m \times n$ matrix M with entries M_{ji} and S by the $p \times m$ matrix N with entries N_{kj}. That is,

$$T(v_i) = \sum_{j=1}^{m} M_{ji} w_j$$

$$S(w_j) = \sum_{k=1}^{p} N_{kj} u_k.$$

Then

$$(S \circ T)(v_i) = S(T(v_i)) = \sum_{j=1}^{n} M_{ji} \left(\sum_{k=1}^{m} N_{kj} \right) u_k = \sum_{k=1}^{p} \left(\sum_{j=1}^{m} N_{kj} M_{ji} \right) u_k$$

so that the matrix for $S \circ T$ with respect to the bases $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_p)$ is

$$\sum_{j=1}^{m} N_{kj} M_{ji}$$

that is, the product of the matrices NM.

Definition 59. Let $\mathcal{L}(V, W)$ denote the set of linear maps from V to W.

Proposition 60. *$\mathcal{L}(V,W)$ is a vector space when given addition and scalar multiplication as follows $(T+S)(v) := T(v) + S(v)$ and $(cT)(v) := cT(v)$.*

Fix bases $(v_1, v_2, \ldots)$ and $(w_1, w_2, \ldots)$ for V and W. A basis for $\mathcal{L}(V,W)$ is given by those transformations $T(j,i) : V \to W$ that take v_i to w_j and send all other basis vectors of V to zero. With respect to the assumed bases of V and W, $T(j,i)$ is represented by the matrix with 1 in the jith position and 0s elsewhere.

Corollary 61. $\dim \mathcal{L}(V,W) = (\dim V)(\dim W)$.

Remark 62. Given a linear map $T : V \to W$ and bases for V and W, there is a matrix representing T. Axler calls this matrix $\mathcal{M}(T)$. It depends on the bases for V and W. In the case $T : \mathbb{F}^n \to \mathbb{F}^m$, there is standard choice of basis (the standard basis). Thus to every matrix M is associated a unique linear map $T : \mathbb{F}^n \to \mathbb{F}^m$, namely the one defined by

$$T(e_i) = \sum_j M_{ji} e_j$$

where e_i denote the standard basis vectors.

Definition 63. The kernel of a linear map $T : V \to W$ is the set of $v \in V$ such that $T(v) = 0$. Axler calls the kernel the "null space".

Proposition 64. *The kernel and image of a linear map $T : V \to W$ are subspaces.*

Proof. If $T(v) = 0$ and $T(w) = 0$ then $0 = T(v) + T(w) = T(v+w)$. Similarly $T(cv) = cT(v) = 0$. Also note that $T(0) = 0$, since $T(0) = T(0v) = 0T(v) = 0$.

The image is a subspace because $T(0) = 0$, $T(v) + T(w) = T(v+w)$, and $cT(v) = T(cv)$. □

Exercise 65. *The kernel of a linear map is $\{0\}$ if and only if that map is injective.*

Linear Equations. A linear equation in m equations and n unknowns is a collection of equations of the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_n \end{cases}$$

for example

$$\begin{cases} 2x_1 + x_2 = 1 \\ x_1 + x_2 = 2 \end{cases} \quad \text{or} \quad \begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 + x_3 = 5 \\ 2x_1 - 3x_2 + x_3 = -3 \end{cases}.$$

The general form for a linear equation can be immediately expressed as

$$Ax = b$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Here the coefficients a_{ji} and b_i are known and one wants to determine the x_i that satisfy all m equations.

As an $m \times n$ matrix, A can be thought of as a linear transformation from $\mathbb{F}^n \to \mathbb{F}^m$. The equation $Ax = b$ has a solution if and only if b is in the image of A. If it does have a solution, the kernel of A determines all of the solutions:

Proposition 66. *Let $T : V \to W$ be a linear map and suppose that $Tv = w$. Then all vectors $v' \in V$ such that $Tv' = w$ form the set $v + \ker(T)$ in V.*

Proof. Suppose that $Tv' = w$. Then $T(v - v') = T(v) - T(v') = w - w = 0$. Therefore $v - v' \in \ker(T)$ so $v' \in v + \ker(T)$.

Suppose that $v' \in v + \ker(T)$. Then $v - v' \in \ker(T)$ and so $T(v - v') = 0$. In particular $T(v') = T(v) = w$. □

Unless $v = 0$, the set $v + \ker(T)$ is not a subspace. It is, however, parallel to the subspace $\ker(T)$ in the sense that it differs from that subspace by addition of a single vector. Thus the collection of all sets of the form $v + \ker(T)$ divide V up into different parallel sheets. This is best visualized for $V = \mathbb{R}^3$ and $\dim \ker(T) = 2$.

Let $w_1, \ldots, w_\ell$ be a basis for the image of T. Let $v_1, \ldots, v_\ell$ be vectors in V such that $T(v_i) = w_i$. The v_i exist because each w_i is in the image of T.

Proposition 67. *Any vector in $v \in V$ can be written as*

$$v = a_1 v_1 + \cdots + a_\ell v_\ell + u$$

where $u \in \ker(T)$ and the a_is are uniquely determined by v.

Proof. Write $T(v) = a_1 w_1 + \cdots + a_\ell w_\ell$. Because the w_i form a basis of the image of T, there is a unique choice of $a_1, \ldots, a_\ell$. Then $v - (a_1 v_1 + \cdots + a_\ell v_\ell)$ is mapped under T to $T(v - (a_1 v_1 + \cdots + a_\ell v_\ell)) = T(v) - a_1 w_1 - \cdots - a_\ell w_\ell = 0$. Hence

$$v = (a_1 v_1 + \cdots + a_\ell v_\ell) + u$$

where $u = v - (a_1 v_1 + \cdots + a_\ell v_\ell)$ is in the kernel of T. □

Proposition 68. *Let $T : V \to W$ be a linear map and let V be finite-dimensional. Let $u_1, \ldots, u_k$ be a basis for $\ker(T)$. Let $v_1, \ldots, v_\ell$ be as in the previous proposition. Then $u_1, \ldots, u_k, v_1, \ldots, v_\ell$ forms a basis of V.*

Proof. $v \in V$ can be uniquely written as

$$v = a_1 v_1 + \cdots + a_\ell v_\ell + u$$

where $u \in \ker(T)$. u can be uniquely written

$$u = b_1 u_1 + \cdots + b_k u_k$$

because $(u_1, \ldots, u_k)$ is a basis for $\ker(T)$. Hence v can be uniquely written

$$v = a_1 v_1 + \cdots + a_\ell v_\ell + b_1 u_1 + \cdots + b_k u_k.$$

□

Corollary 69 ("rank-nullity")**.** *Let $T : V \to W$ and let V be finite dimensional. Then $\dim V = \dim \ker(T) + \dim \operatorname{im}(T)$.*

The dimension of the image of T is often called the "rank" of T, hence the name of this corollary.

Corollary 70. *Let $T : V \to V$, that is, T sends a vector space to itself. Then T is invertible if and only if T is injective. T is also invertible if and only if T is surjective.*

4 Eigenvectors and Jordan Normal Form

Consider the following two matrices

$$\begin{pmatrix} 1 & 2 & -1 & 4 \\ 6 & 0 & -5 & 1 \\ 6 & 8 & -4 & 4 \\ 0 & 0 & 2 & 9 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & -1 & 2 \end{pmatrix}.$$

Anyone who has spent some time with matrices would prefer the latter to the former: it is easier to invert, to find its image, to find its kernel, et cetera. The reason it is easier to analyze the second matrix is that one needs only analze each of the three matrices (1×1, 1×1 and 2×2) on its diagonal. Three computations involving a 1×1 matrix, a 1×1 matrix, and a 2×2 matrix are almost always going to be easier than a single computation involving a 4×4 matrix.

Definition 71. A square matrix is block diagonal if it of the form

$$\begin{pmatrix} A_1 & 0 & 0 & \cdots & 0 \\ 0 & A_2 & 0 & \cdots & 0 \\ 0 & 0 & A_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_k \end{pmatrix}$$

where each A_i is a square matrix and the 0 entries in the above matrix denote matrices with all entries 0. The matrices A_i are called "blocks".

This definition is a little silly, since every $n \times n$ matrix is block diagonal with just a single $n \times n$ block. It is often to one's benefit, as suggested by two explicit 4×4 matrices at the start of the section, if one has many smaller blocks as opposed to just one $n \times n$ block.

Since $n \times n$ matrices represent linear transformations from $\mathbb{R}^n$ to itself, there should be a corresponding notion of "blocks" in the setting of abstract linear algebra:

Definition 72. Let $T : V \to V$ be linear. An invariant subspace for T is a subspace $U \subset V$ such that $T(U) \subset U$.

Oftentimes I'll just write "invariant subspace" instead of "invariant subspace for T". But whether or not a subspace is invariant depends on the particular map T.

Example 73. $\{0\}$ is an invariant subspace since $T(0) = 0$.

Example 74. V is an invariant subspace since $T(V) \subset V$.

Example 75. $\ker(T)$ is an invariant subspace since $T(\ker(T)) = \{0\} \subset \ker(T)$

Example 76. $\mathrm{im}(T)$ is an invariant subspace since $T(\mathrm{im}(T))$ consists of vectors of the form $T(T(v))$.

Exercise 77. *Let $T : V \to V$ and let U_1 and U_2 be invariant subspaces. Then $U_1 \cap U_2$ is an invariant subspace.*

Definition 78. Let $T : V \to V$. An invariant subspace U is indecomposable if $U = U_1 \oplus U_2$ implies that $U_1 = U$ or $U_2 = U$.

Proposition 79. *Let $T : V \to V$ and let V be finite dimensional. There exists a collection of indecomposable invariant subspaces $U_1, \ldots, U_k$ such that $V = U_1 \oplus \cdots \oplus U_k$.*

Proof. Note that V can be expressed as a direct sum of (possibly reducible) invariant subspaces since V itself is an invariant subspace.

Given a decomposition of V as a direct sum of invariant subspaces:

$$V = W_1 \oplus \cdots \oplus W_\ell$$

either all of the W_i are indecomposable or at least one is not. If the latter, decompose a reducible invariant subspace as a direct sum of two invariant subspaces to get a new decomposition of V as a direct sum of more subspaces. Continue in this manner until you get a decomposition where all the invariant subspaces are indecomposable. Note that this process terminates since there can be at most $\dim(V)$ subspaces in such a decomposition. □

Write $V = U_1 \oplus \cdots \oplus U_k$ as in the last proposition. Putting together bases for each U_i produces a basis for V. With respect to this basis, T can be written has a block diagonal matrix where each block cannot be written as block diagonal with two or more blocks. If all of the blocks are 1×1 matrices one says that T is diagonalizable and that the bases for U_i (necessarily each a single element) are eigenvectors.

Definition 80. An eigenvector for T is a vector $v \neq 0$ such that $Tv = \lambda v$ for some $\lambda \in \mathbb{F}$. The number λ is called the "eigenvalue" for v.

Remark 81. Note that if v is an eigenvalue for T with eigenvalue λ, then so is cv for any $c \neq 0$.

Definition 82. A linear map $T : V \to V$ is diagonalizable if there exists a basis of V consisting of eigenvectors for T.

Remark 83. With respect to a basis $(v_1, \ldots, v_n)$ of eigenvectors, the matrix for T has its nonzero entries on the diagonal:

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

This justifies the term "diagonalizable". Here λ_i is the eigenvalue for v_i.

"Most" matrices are diagonalizable. Since there are infinitely many matrices the word "most" has to be qualified, which I will not do here[1]. It might suffice to say that if you were to use a computer to randomly sample n^2 numbers from $[0, 1]$ and put them in a matrix, that matrix would be diagonalizable[2].

However, not all linear transformations are diagonalizable:

Example 84. The linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ represented by the following matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

[1]The correct way: if you topologize the space of matrices in the usual way, the set of diagonalizable matrices forms an open dense subset.

[2]Assume that the computer can work to arbitrary precision.

is not diagonalizable. To see this, it is enough to check that its eigenvectors all lie on a single line in $\mathbb{R}^2$, and so you cannot form a basis of $\mathbb{R}^2$ from them. Let $\begin{pmatrix} a \\ b \end{pmatrix}$ be an eigenvector for $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \lambda \begin{pmatrix} a \\ b \end{pmatrix}$$

for some $\lambda \in \mathbb{F}$. Then

$$\begin{pmatrix} b \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda a \\ \lambda b \end{pmatrix}$$

which means that

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c \\ 0 \end{pmatrix}$$

and $\lambda = 0$.

Definition 85. If T has an eigenvector of eigenvalue λ, the collection of eigenvectors for T with eigenvalue λ plus the 0 vector form a subspace called the λ-eigenspace of T.

Exercise 86. *If T is diagonalizable, then $V = U_1 \oplus \cdots \oplus U_k$ where the U_i are the eigenspaces of T.*

Remark 87. The 0-eigenspace of T is $\ker(T)$.

The following theorem is the first time in these notes where a distinction between $\mathbb{R}$ and $\mathbb{C}$ has been significant.

Theorem 88. *Let p be a polynomial with coefficients in $\mathbb{C}$. Then*

$$p(z) = a(z - r_1) \cdots (z - r_n)$$

for some $a, r_1, \ldots, r_n \in \mathbb{C}$.

This theorem is called the "fundamental theorem of algebra" and is usually proved in a course in complex analysis or algebraic topology. It will not be proved here. Note that it is not true if $\mathbb{C}$ is replaced by $\mathbb{R}$: $x^2 + 1$ does not factor into two polynomials of degree 1 with coefficients in $\mathbb{R}$ since there is no square root of -1 in $\mathbb{R}$.

Proposition 89. *Let $T : V \to V$ be a linear map of finite dimensional complex vector spaces. Then T has an eigenvector.*

Proof. Pick any nonzero vector $v \in V$. Consider the sequence of vectors

$$(v, T(v), T^2(v), T^3(v), \ldots,).$$

Since at most $\dim(V)$ vectors in V can be linearly independent, there has to be some m such that

$$\{v, T(v), \ldots, T^m(v)\}$$

is linearly dependent and hence there exists a relation

$$a_0 v + a_1 T(v) + \cdots + a_m T^m(v).$$

This implies that v is in the kernel of the operator

$$a_0 \operatorname{id}_V + a_1 T + \cdots + a_m T^m.$$

By the fundamental theorem of algebra, this operator is equal to

$$(T - r_1 \operatorname{id}_V) \circ \cdots \circ (T - r_m \operatorname{id}_V)$$

for some complex numbers $r_1, \ldots, r_m$. Since this operator has a kernel, it is not invertible. Therefore for some i, $(T - r_i \operatorname{id}_V)$ factors is not invertible. Rank-nullity implies that $T - r_i \operatorname{id}_V$ has a vector u in its kernel.

Then

$$(T - r_i \operatorname{id}_V)u = 0 \Rightarrow Tu = r_i u$$

so u is an eigenvector. □

Definition 90. If $T : V \to V$ is a linear map and $U \subset V$ an invariant subspace, then $T|_U$ is the linear map $U \to U$ obtained by restricting T to U.

Even though there some linear transformations that are not diagonalizable, Camille Jordan proved the following theorem in the 19th century:

Theorem 91. *Let $T : V \to V$ be linear map and V a finite dimensional complex vector space. Then there exist indecomposable invariant subspaces $U_1, \ldots, U_k$, numbers $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$, and bases $B_1, \ldots, B_k$ (where B_i is a basis of U_i) such that $V = U_1 \oplus \cdots \oplus U_k$ and, with respect to the basis B_i, $T|_{U_i}$ can be represented by a matrix that looks like*

$$\begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{pmatrix}$$

(a matrix with the same entry λ_i on the main diagonal and 1s on the diagonal above that).

With respect to the basis $(B_1, \ldots, B_k)$ in the statement of the theorem, the matrix for T might look something like

$$\begin{pmatrix} \lambda_1 & 1 & & & & & & \\ & \lambda_1 & 1 & & & & & \\ & & \lambda_1 & & & & & \\ & & & \lambda_2 & & & & \\ & & & & \lambda_3 & & & \\ & & & & & \lambda_4 & 1 & \\ & & & & & & \lambda_4 & \end{pmatrix}$$

(where the blank space means entries filled with 0). A matrix in this form is said to be in "Jordan normal form". Each block for $T|_{U_i}$ is called a "Jordan block". In the 7×7 matrix pictured above, there are four Jordan blocks of sizes 3, 1, 1, and 2. Note that it may be the case that $\lambda_i = \lambda_j$ for $i \neq j$.

Remark 92. If all the Jordan blocks are 1×1 (meaning all the subspaces U_i in the statement of the theorem are 1-dimensional) then the matrix is diagonalizable, the basis $(B_1, \ldots, B_k)$ of V is a basis of eigenvectors and the Jordan normal form is diagonal.

Note that by Proposition 79, V can be decomposed into $V = U_1 \oplus \cdots \oplus U_k$ where each U_i is an indecomposable invariant subspace. Therefore Theorem 91 follows from the following easier proposition:

Proposition 93. *Let $T : V \to V$ with V a finite dimensional complex vector space and suppose that V is an indecomposable invariant subspace. Then there exists a basis $(v_1, \ldots, v_n)$ for V and a number λ such that with respect this basis T is represented by a matrix of the form*

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}$$

Before proving this proposition, it will be necessary to discuss nilpotent transformations.

Definition 94. A linear map $N : V \to V$ is nilpotent if $N^n = 0$ for some n.

Proposition 95. *Let $N : V \to V$ be nilpotent and let k be the largest integer such that $N^k \neq 0$. Then there exist disjoint ordered finite subsets*

$$X(0) \subset V,\ X(1) \subset V,\ X(2) \subset V, \cdots, X(k) \subset V$$

such that

$$\bigcup_{j=0}^{k} \bigcup_{i=0}^{j} N^i(X(j))$$

is a basis of V.

Proof. This proof is a meditation on the sequence of linear surjections

$$V \xrightarrow{N} \mathrm{im}(N) \xrightarrow{N} \mathrm{im}(N^2) \xrightarrow{N} \mathrm{im}(N^3) \xrightarrow{N} \cdots \xrightarrow{N} \mathrm{im}(N^{k-1}) \xrightarrow{N} \mathrm{im}(N^k) \xrightarrow{N} 0.$$

To better understand what's going on, first suppose that $k = 1$, so the sequence is

$$V \xrightarrow{N} \mathrm{im}(N) \xrightarrow{N} 0$$

Pick a basis for $\mathrm{im}(N)$. This is a list of the form $(N(v_1), \ldots, N(v_n))$. Let $X(1) := (v_1, \ldots, v_n)$ so that $N(X(1))$ is the chosen basis of $\mathrm{im}(N)$. There are two things to note. First:

$$\ker(N) \cap \mathrm{Span}(X(1)) = \{0\}$$

In particular, this implies that

$$V = \ker(N) \oplus \mathrm{Span}(X(1))$$

since N maps $\mathrm{Span}(X(1))$ isomorphically to $\mathrm{im}(N)$. Second: that $N(X(1)) \subset \ker(N)$. $N(X(1))$ is a linearly independent subset and therefore can be extended to a basis of $\ker(N)$ by adding a list of vectors $X(0) \subset \ker(N)$. By rank-nullity,

$$X(1) \cup N(X(1) \cup X(0)$$

forms a basis of V. This proves the proposition in the simple case $k = 1$. A potentially helpful diagram is

$$\begin{array}{ccccc} V & \xrightarrow{N} & \mathrm{im}(N) & \xrightarrow{N} & 0 \\ X(1) & & N(X(1)) & & \\ N(X(1)) & & & & \\ X(0) & & & & \end{array}$$

The top line contains three vector spaces. Below each vector space is a collection of finite subsets that together form a basis for that space. A subset maps to the subset to the right of it. If there is no subset to the right of it, it maps to 0.

It will be additionally helpful to do the next case, where $k = 2$. Consider a basis $N^2(X(2))$ of $\mathrm{im}(N^2)$ where $X(2)$ is a linearly independent list of vectors in V. Again, $N(X(2)) \cap \ker(N) = \{0\}$ which implies that

$$\mathrm{im}(N) = N(X(2)) \oplus (\ker(N) \cap \mathrm{im}(N))$$

Since $N^2(X(2)) \subset \ker(N)$ is a linearly independent set you can add additional vectors in $\ker(N) \cap \mathrm{im}(N)$ to make it a basis. These vectors are of the form $N(X(1))$ for some list of vectors $X(1)$ in V. Therefore $N(X(2)) \cup N^2(X(2)) \cup N(X(1))$ forms a basis of $\mathrm{im}(N)$. As it stands, $X(2) \cup N(X(2)) \cup X(1)$ is a linearly independent subset of N that maps isomorphically onto $\mathrm{im}(N)$. Therefore

$$V = \mathrm{Span}(X(2) \cup N(X(2)) \cup X(1)) \oplus \ker(N)$$

As before, $N^2(X(2)) \cup N(X(1))$ forms a linearly independent subset of $\ker(N)$. Extend it to a basis by adding on vectors $X(2)$. Therefore

$$X(2) \cup N(X(2)) \cup X(1) \cup N^2(X(2)) \cup N(X(1)) \cup X(0)$$

is a basis of V. This proves the proposition in the case where $k = 2$. The corresponding picture is

$$\begin{array}{ccccc} V & \xrightarrow{N} & \mathrm{im}(N) & \xrightarrow{N} & \mathrm{im}(N^2) \xrightarrow{N} 0 \\ X(2) & & N(X(2)) & & N^2(X(2)) \\ N(X(2)) & & N^2(X(2)) & & \\ X(1) & & N(X(1)) & & \\ N^2(X(2)) & & & & \\ N(X(1)) & & & & \\ X(0) & & & & \end{array}$$

The reader who checks the case $k = 3$ in the same manner is sure to believe the proposition. For completeness, the inductive proof of the general case follows.

Suppose inductively that you have constructed sets $X(k), X(k-1), \ldots, X(j+1)$ such that the union of

$$\begin{array}{c} N^{j+1}(X(k)) \\ N^{j+2}(X(k)), N^{j+1}(X(k-1)) \\ N^{j+3}(X(k)), N^{j+2}(X(k-1)), N^{j+1}(X(k-2)) \\ \vdots \\ N^{k}(X(k)), N^{k-1}(X(k-1)), \ldots, N^{j+1}(X(j+1)) \end{array}$$

is a basis $\mathrm{im}(N^{j+1})$ satisfying the following two properties: (1) the span of the last line is $\ker(N) \cap \mathrm{im}(N^{j+1})$ and (2) the span of all but the last line maps isomorphically onto $\mathrm{im}(N^{j+2})$. Then it is routine (in a similar manner to the cases $k = 1$ and $k = 2$) to check that the union of

$$\begin{array}{c} N^{j}(X(k)) \\ N^{j+1}(X(k)), N^{j}(X(k-1)) \\ N^{j+2}(X(k)), N^{j+1}(X(k-1)), N^{j}(X(k-2)) \\ \vdots \\ N^{k-1}(X(k)), N^{k-2}(X(k-1)), \ldots, N^{j}(X(j+1)) \end{array}$$

spans a subspace of $\mathrm{im}(N^j)$ that maps isomorphically onto $\mathrm{im}(N^{j+1})$ and that

$$N^{k}(X(k)), N^{k-1}(X(k-1)), \ldots, N^{j+1}(X(j+1))$$

spans a subspace of $\ker(N) \cap \mathrm{im}(N^j)$ that can be completed to a basis by adding a linearly independent subset of the form $N^j(X(j))$ for some subset $X(j)$ in V. Hence the union of

$$N^{j}(X(k))$$

$$N^{j+1}(X(k)), N^j(X(k-1))$$
$$N^{j+2}(X(k)), N^{j+1}(X(k-1)), N^j(X(k-2))$$
$$\cdots$$
$$N^{k-1}(X(k)), N^{k-2}(X(k-1)), \ldots, N^j(X(j+1))$$
$$N^k(X(k)), N^{k-1}(X(k-1)), \ldots, N^j(X(j))$$

forms a basis of $\mathrm{im}(N^j)$ that satisfies (1) the span of the last line is $\ker(N) \cap \mathrm{im}(N^j)$ and (2) the span of all but the last line maps isomorphically onto $\mathrm{im}(N^{j+1})$.

The proposition follows from the induction when $j = 0$. □

Corollary 96. *Let $N : V \to V$ be a nilpotent linear map. Then there exists a basis of N which is the union lists of the form*

$$(v, N(v), N^2(v), \ldots, N^j(v))$$

where $N^j(v) \in \ker(N)$.

Proof. Let v be a vector in $X(j)$. □

Corollary 97. *Let $N : V \to V$ be a nilpotent linear map such that V is an indecomposable invariant subspace. Then there exists a basis of V of the form*

$$(N^{n-1}(v), N^{n-2}(v), \ldots, N(v), v)$$

where $N^{n-1}(v) \in \ker(N)$. In particular, with respect to this basis, N is represented by a matrix of the form

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Proof. In the previous corollary, each list $(v, N(v), N^2(v), \ldots, N^j(v))$ is a basis for an invariant subspace, call it U_i. Since these lists together form a basis of V, it follows that $U_i \cap U_{i'} = \{0\}$ for $i \neq i'$. Hence

$$V = U_1 \oplus \cdots \oplus U_k$$

expresses V as a direct sum of invariant subspaces. Since V is indecomposable, all but one of the U_is must be zero. □

Most of the work in the proof of Jordan normal form is contained the previous study of nilpotent linear maps. What follows are two preparatory propositions and then the proof of Proposition 93.

Proposition 98. *Let $T : V \to V$ and with V finite dimensional. If $\operatorname{im}(T^k) = \operatorname{im}(T^{k+1})$ for some k, then $\operatorname{im}(T^k) = \operatorname{im}(T^\ell)$ for all $\ell \geq k$.*

Proof. The important point to note is that

$$T^a(\operatorname{im}(T^b)) = \operatorname{im}(T^{a+b})$$

because the left side consists of vectors of the form $T^a(T^b(v))$ and the right side consists of vectors of the form $T^{a+b}(v)$. These two sets of vectors, of course, are the same.

The proof we will be by induction on ℓ. The base case $\ell = k+1$ is assumed in the statement of the proposition. Assume that $\operatorname{im}(T^\ell) = \operatorname{im}(T^k)$. You want to show that $\operatorname{im}(T^{\ell+1}) = \operatorname{im}(T^k)$ Write $\operatorname{im}(T^{\ell+1}) = T(\operatorname{im}(T^\ell)) = T(\operatorname{im}(T^k)) = \operatorname{im}(T^{k+1}) = \operatorname{im}(T^k)$. □

Proposition 99. *Let $T : V \to V$ be a linear map with $\dim(V) = n$. Then $V = \ker(T^n) \oplus \operatorname{im}(T^n)$.*

Proof. Consider the decreasing sequence of subspaces

$$V \supset \operatorname{im}(T) \supset \operatorname{im}(T^2) \supset \operatorname{im}(T^3) \supset \cdots .$$

Each subspace either has strictly smaller dimension than the last, or all subspaces after it are equal to each other (Proposition 98). Since $\dim(V) = n$, the dimension cannot decrease for more than n steps in the sequence. Hence $\operatorname{im}(T^n) = \operatorname{im}(T^{n+1}) = \operatorname{im}(T^{n+2}) = \cdots$. In particular, $\operatorname{im}(T^n) = \operatorname{im}(T^{2n})$.

Rank nullity implies that $\dim(V) = \dim(\ker(T^n)) + \dim(\operatorname{im}(T^n))$ so in order to show that $V = \ker(T^n) \oplus \operatorname{im}(T^n)$ it is enough to show that $\ker(T^n) \cap \operatorname{im}(T^n) = \{0\}$. Let $v \in \ker(T^n) \cap \operatorname{im}(T^n)$. Then $v = T^n(w)$ for some $w \in V$. Also $T^n(v) = 0$. Because $T^n : \operatorname{im}(T^n) \to \operatorname{im}(T^{2n})$ is a surjection of a vector space to itself, it is an isomorphism. Therefore if $T^n(T^n(w)) = 0$ then $T^n(w) = 0$. Hence $v = 0$. □

Proof of Proposition 93. To recap the notation: $T : V \to V$ is a linear map of a finite dimensional complex vector space to itself and V is an indecomposable invariant subspace.

Let v_0 be an eigenvalue for V with eigenvalue λ. Write $\dim(V) = n$. Then by Proposition 99

$$V = \ker((T - \lambda \operatorname{id}_V)^n) \oplus \operatorname{im}(T - \lambda \operatorname{id}_V)^n).$$

Invariant subspaces are the same for T and $T - \lambda \operatorname{id}_V$ so that V is indecomposable invariant subspace for $T - \lambda \operatorname{id}_V$. Since $\ker((T - \lambda \operatorname{id}_V)^n)$ and $\operatorname{im}((T - \lambda \operatorname{id}_V)^n)$ are invariant subspaces for $T - \lambda \operatorname{id}_V$, indecomposability of V implies that

$$V = \ker((T - \lambda \operatorname{id}_V)^n) \text{ or } V = \operatorname{im}((T - \lambda \operatorname{id}_V)^n).$$

Since v is a nonzero vector in $\ker((T - \lambda \operatorname{id}_V)^n)$ it must be that

$$V = \ker((T - \lambda \operatorname{id}_V)^n)$$

This implies that $T - \lambda \operatorname{id}_V$ is nilpotent. The proof then follows from Corollary 97. □

Definition 100. Let $T : V \to V$ be linear. A nonzero vector v for which $(T - \lambda \operatorname{id}_V)^n(v) = 0$ for some n is called a "generalized eigenvector" with eigenvalue λ.

Definition 101. The set of all generalized eigenvectors with eigenvalue λ, plus the zero vector, form a subspace of V called a "generalized eigenspace".

One consequence of the Jordan form theorem (Theorem 91) is that a linear transformation on a finite-dimensional complex vector space has a basis of generalized eigenvectors.

5 Inner Product Spaces

Definition 102. The dot product in $\mathbb{R}^n$ is a map

$$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \mapsto a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

The dot product of (v, w) is often denoted $v \cdot w$.

The benefits of the dot product are twofold: (1) it enables a better understanding of geometry in $\mathbb{R}^1$, $\mathbb{R}^2$, and $\mathbb{R}^3$ and (2) it provides a starting point to generalize our familiar notions of geometry and distance to other settings where they may be of use.

In the Cartesian plane, the distance from the origin to the point (a, b) is $\sqrt{a^2 + b^2}$. This is the content of the Pythagorean theorem. Translated into the language of the dot product, the distance from the origin to the vector $v \in \mathbb{R}^2$ is $\sqrt{v \cdot v}$.

In 3-dimensional space, a little geometric reasoning shows that the distance from the origin to the point (a, b, c) is $\sqrt{a^2 + b^2 + c^2}$. Therefore, again the distance from the origin to the vector $v \in \mathbb{R}^3$ is $\sqrt{v \cdot v}$.

It therefore makes sense to define the distance from a the origin in $\mathbb{R}^n$ to a vector $v \in \mathbb{R}^n$ as $\sqrt{v \cdot v}$.

Note that a set of points in $\mathbb{R}^2$ equidistant from the origin forms a circle. The set of points in $\mathbb{R}^3$ equidistant from the origin forms a sphere. The set of

points in $\mathbb{R}^4$ equidistant from the origin forms a ??? It might be interesting to the reader to try to visualize this set.

Because the expression $\sqrt{v \cdot v}$ is used so much, it is given a separate name: $\|v\| := \sqrt{v \cdot v}$. Also note that the dot product observes two interesting features with respect to linearity:

$$(v_1 + v_2) \cdot (u_1 + u_2) = v_1 \cdot u_1 + v_2 \cdot u_1 + v_1 \cdot u_2 + v_2 \cdot u_2$$

$$(cv) \cdot u = c(v \cdot u) = v \cdot (cu).$$

and the dot product is also "symmetric":

$$v \cdot u = u \cdot v$$

The dot product enables you to do some pretty subtle distance calculations easily.

Example 103. Consider a parallelogram $ABCD$. Given a line segment X let $\ell(X)$ denote its length. Then

$$\ell(\overline{AC})^2 + \ell(\overline{BD})^2 = \ell(\overline{AB})^2 + \ell(\overline{BC})^2 + \ell(\overline{CD})^2 + \ell(\overline{DA})^2.$$

This expresses a relation between the lengths of the edges and the lengths of the diagonals. To prove this, put A at the origin so that $\overline{AB}$ is a vector v and $\overline{AD}$ is a vector w. Then $\overline{AC}$ is $v + w$ and $\overline{BD}$ is $v - w$ (there's some abiguity about the sign because line segments don't have direction, but this doesn't matter). Therefore

$$\ell(\overline{AC})^2 + \ell(\overline{BD})^2 = \|v+w\|^2 + \|v-w\|^2 = (v+w)\cdot(v+w) + (v-w)\cdot(v-w)$$

$$= v \cdot v + 2v \cdot w + w \cdot w + v \cdot v - 2v \cdot w + w \cdot w$$

$$= 2(v \cdot v + w \cdot w) = \ell(\overline{AB})^2 + \ell(\overline{BC})^2 + \ell(\overline{CD})^2 + \ell(\overline{DA})^2.$$

Example 104. In a similar manner, it should not be hard to prove the formula on the cover of Axler.

The dot product has the curious property that it is determined by the lengths of vectors. That is, if you only knew the function $v \mapsto \|v\|$, then you could recover $v \cdot w$ for all $v, w \in \mathbb{R}^n$:

$$v \cdot w = \frac{\|v+w\|^2 - \|v-w\|^2}{4}.$$

This follows by expanding out $\|v+w\|^2 = (v+w)\cdot(v+w) = v\cdot v + 2v\cdot w + w\cdot w$ et cetera. Let $R : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map that preserves distance: $\|R(v)\| = \|v\|$ for all $v \in \mathbb{R}^n$. Then

$$(Rv) \cdot (Rw) = \frac{\|Rv + Rw\|^2 - \|Rv - Rw\|^2}{4} = \frac{\|R(v+w)\|^2 - \|R(v-w)\|^2}{4}$$

$$= \frac{\|v+w\|^2 - \|v-w\|^2}{4} = v \cdot w.$$

Therefore R preserves the dot product as well. In $\mathbb{R}^2$ and $\mathbb{R}^3$ such transformations are precisely rotations and reflections and compositions thereof.

Given two nonzero vectors $v, w \in \mathbb{R}^2$ there's always a composition of a rotation and (possibly) a reflection that takes v to a a vector Rv that is a positive multiple of e_1 and takes w to a vector Rw with positive e_2 component. Note that $Rv \cdot Rw = v \cdot w$ since rotations and reflections preserve dot product. Write

$$Rv = \begin{pmatrix} a \\ 0 \end{pmatrix}, \; Rw = \begin{pmatrix} b\cos\theta \\ b\sin\theta \end{pmatrix}.$$

Then

$$\frac{v \cdot w}{\|v\|\|w\|} = \frac{Rv \cdot Rw}{\|Rv\|\|Rw\|} = \cos\theta$$

where θ is the angle between Rv and Rw measured from Rv to Rw. Since R preserves distances and hence preserves angles, it follows that if θ is the angle between v and w then

$$\frac{v \cdot w}{\|v\|\|w\|} = \cos(\theta).$$

Note that there's some abiguity in whether the angle is measured from v to w or from w to v. Since $\cos(\theta) = \cos(-\theta)$ this doesn't matter in the above formula. The major consequence of this is that angles can expressed in terms of the dot product.

Corollary 105. *The law of cosines: given a triangle ABC then*

$$\ell(\overline{BC})^2 = \ell(\overline{AB})^2 + \ell(\overline{AC})^2 - 2\ell(\overline{AB})\ell(\overline{AC})\cos(\angle A).$$

Proof. Let v be the vector from A to B and let w be the vector from A to C. Then

$$\ell(\overline{BC})^2 = \|v-w\|^2 = \|v\|^2 + \|w\|^2 - 2v \cdot w = \|v\|^2 + \|w\|^2 - 2\|v\|\|w\|\cos\theta$$

where θ is the angle between v and w (i.e., the angle at A). □

Exercise 106. *Prove the law of sines using vectors in $\mathbb{R}^2$ and the dot product.*

Correlation. In statistics a dataset is represented by a vector $x = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$. For example, you might collect data from n people and let x_i be the height (in inches, say) of the ith person. You might define $y = (y_1, y_2, \ldots, y_n)$ be the vector where y_i is the weight (in pounds, say) of the ith person. You might have seen such data presented as a collection of n points (x_i, y_i) in the plane: the horizontal axis representing height (in inches) and the vertical axis representing weight (in pounds). But another (more correct?) way to think of x and y is as two vectors in $\mathbb{R}^n$.

Suppose that we lived a fantasy world where a person's weight (in pounds) was always exactly twice their height (in inches). Then vector y would be

exactly twice the vector x: $y = 2x$. In particular, the angle between y and x would be zero, so

$$\frac{x \cdot y}{\|x\|\|y\|} = \cos(0) = 1.$$

This is of course not the case, but something like it is the case. A person's weight (in pounds) might not be twice their weight (in inches) but it will be reasonably close to that factor of 2. For example, it won't be 1/100th of their height or 20 times their height. Therefore, while vectors x and y collected from real-world data won't be multiples of each other, they'll have a pretty small angle θ between. In other words,

$$\frac{x \cdot y}{\|x\|\|y\|} = \cos(\theta)$$

will be close to 1. The quantity

$$\frac{x \cdot y}{\|x\|\|y\|}$$

is (almost, but not quite!) what the statisticians call the "correlation" between the height and weight in the sample. It has the geometric interpretation of cosine of an angle between two vectors, but it can be computed easily from the data collected without thinking about its geometric meaning. The actual thing that statisticians call correlation is obtained by first orthogonally projecting x and y onto the subspace of $\mathbb{R}^n$ consisting of points of the form $(a_1, \ldots, a_n)$ where $a_1 + \cdots + a_n = 0$ and then computing cosine of the angle between the resulting vectors.

Special Relativity. The dot product is intimately related to the usual notion of distance in $\mathbb{R}^n$. If you modify the dot product a little, you get a new notion of distance. For example, define the "Lorentzian" product on $\mathbb{R}^2$ as a map $\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\begin{pmatrix} x_1 \\ t_1 \end{pmatrix} \cdot_L \begin{pmatrix} x_2 \\ t_2 \end{pmatrix} := x_1 x_2 - t_1 t_2.$$

This is different from the dot product by the presence of a minus sign. The Lorentzian inner product determines a norm-squared like the dot product

$$\|v\|_L^2 := v \cdot_L v.$$

One could take a square root of this norm, but this leads to sign ambiguities so let's avoid doing so. This Lorentzian norm-squared defines a new notion of distance on $\mathbb{R}^2$. For example

$$\|e_1\|_L^2 = 1$$

so if two points in $\mathbb{R}^2$ differ by e_1: $A + e_1 = B$, then the Lorentzian distance-squared between A and B is 1. Weird things can happen with this new norm. The norm-squared can be zero or negative:

$$\|e_2\|_L^2 = -1, \ \|e_1 + e_2\|_L^2 = 0.$$

This is one reason that no one talks about $\|v\|_L^2$ and $\|v\|_L$. Would $\|e_1\|_L$ be i or $-i$?

Defining a norm like this might seem absurd (imaginary distance?) but it's actually been very useful. Before discussing that, it's interesting to contrast the geometry involved in Lorentzian product to the geometry of the usual dot product in $\mathbb{R}^2$. In $\mathbb{R}^2$ with the usual dot product, vectors of the same norm form circles. On the the unit circle, one defines two functions cos and sin by declaring the the point on the unit circle a distance θ along the circle from the positive x-axis is $(\cos(\theta), \sin(\theta))$. Let $R : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map that preserves the dot product. Like any linear map, it is determined by $R(e_1)$ and $R(e_2)$. Since R preserves distance, both $R(e_1)$ and $R(e_2)$ lie on the unit circle a distance $\pi/2$ from each other. Write

$$R(e_1) = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$$

then

$$R(e_2) = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} \text{ or } R(e_2) = \begin{pmatrix} \sin(\theta) \\ -\cos(\theta) \end{pmatrix}.$$

Therefore, with respect to the standard basis, R is represented by either

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \text{ or } \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}.$$

The first represents a rotation. The second represents a reflection first, then a rotation.

In $\mathbb{R}^2$ with the Lorentzian product, the vectors of the same positive or negative norm-squared form hyperbolas. Vectors of zero-norm form the two lines spanned by $e_1 \pm e_2$. Define two functions cosh and sinh as follows. On the unit hyperbola $x^2 - t^2 = 1$, let $(\cosh(\beta), \sinh(\beta))$ be the point a (Lorentzian) distance β along the hyperbola, measured starting at the x-axis and moving upwards. Then on the negative unit hyperbola $x^2 - t^2 = -1$, the point $(\sinh(\beta), \cosh(\beta))$ is a (Lorentzian) distance β along the hyperbola, measured starting at the positive t-axis and moving right.[3] Let $R : \mathbb{R}^2 \to \mathbb{R}^2$ be a transformation that preserves the Lorentzian product. Since $\|e_1\|_L^2 = 1$, then $R(e_1)$ has to lie on either sheet of the hyperbola $x^2 - t^2 = 1$. Since $\|e_2\|_L^2 = -1$ then $R(e_2)$ has to lie on either sheet of the hyperbola $x^2 - t^2 = -1$. There are four cases to consider, depending on which sheets $R(e_1)$ and $R(e_2)$ lie on, but

[3] Because $e^{i\theta} = \cos(\theta) + i\sin(\theta)$,

$$\cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2}, \ \sin(\theta) = \frac{e^{i\theta} - e^{-i\theta}}{2i}.$$

It turns out that

$$\cosh(\beta) = \frac{e^{\beta} + e^{-\beta}}{2}, \ \sinh(\beta) = \frac{e^{\beta} - e^{-\beta}}{2}.$$

consider the simplest case where $R(e_1)$ lies on the same sheet as e_1 and $R(e_2)$ lies on the same sheet as e_2. Write

$$R(e_1) = \begin{pmatrix} \cosh\beta \\ \sinh\beta \end{pmatrix}$$

for some β. Then since $R(e_1) \cdot_L R(e_2) = e_1 \cdot_L e_2 = 0$ one sees that

$$R(e_2) = \begin{pmatrix} \sinh\beta \\ \cosh\beta \end{pmatrix}.$$

Therefore, with respect to the standard basis, R is represented by the matrix

$$R = \begin{pmatrix} \cosh\beta & \sinh\beta \\ \sinh\beta & \cosh\beta \end{pmatrix}.$$

The two lines containing vectors v such that $\|v\|_L^2 = 0$ are two eigenspaces for this transformation, with eigenvalues $\cosh\beta + \sin\beta$ and $\cosh\beta - \sinh\beta$.

Before Einstein came along, the laws of nature seemed to depend on various constants and the distances between things, distances measured using the usual dot product. Consider, for example, the attraction of two electrically charged particles or the gravitational charge between massive objects

$$F = \frac{kq_1q_2}{r^2},\ F = -\frac{km_1m_2}{r^2}.$$

Einstein pointed out that the laws of nature depend on various constants and the distances between points in space-time, not space, *and with distance measured using a Lorentzian product*:

$$\|(x, y, z, t)\|_L^2 = x^2 + y^2 + z^2 - c^2t^2$$

where c is the speed of light. This is the basis of special relativity and essentially all of special relativity's quirks can be understood in terms of the geometry of the Lorentzian product on $\mathbb{R}^4$ (and, hence, the geometry of hyperbolas).

Here's one quirk: simultaneity no longer makes sense in special relativity; that is, you cannot say that events A and B happen "at the same time". Consider the xt-plane. Two points represent points "at the same time" if they lie on the same horizontal line. The transformations R that preserve $x^2 - t^2$ do not send horizontal lines to horizonal lines. Therefore they do not preserve the notion of "at the time same time". But, according to Einstein, the laws of nature should be unchanged if you apply such a transformation R. Therefore it does not make sense to say that two events "happen at the same time".

Definition 107. Let V be a vector space over $\mathbb{F}$. A bilinear form is a map

$$V \times V \to \mathbb{F}$$

typically denoted by $(v, w) \mapsto \langle v, w\rangle$, that is "linear in each factor":

$$\langle v_1 + v_2, u_1 + u_2\rangle = \langle v_1, u_1\rangle + \langle v_1, u_2\rangle + \langle v_2, u_1\rangle + \langle v_2, u_2\rangle.$$

$$\langle cv, u\rangle = c\langle v, u\rangle = \langle v, cu\rangle.$$

Example 108. Examples of bilinear forms include the usual dot product on $\mathbb{R}^n$ and the Lorentzian product $\cdot_L$ on $\mathbb{R}^2$ described above.

Axler does not discuss bilinear forms. He instead discusses a related set of things called "inner products".

Definition 109. Let V be a vector space over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. An inner product on V is a map

$$V \times V \to \mathbb{F}$$

typically denoted $(v, w) \mapsto \langle v, w \rangle$ and satisfying[4]

$$\langle v_1 + v_2, u_1 + u_2 \rangle = \langle v_1, u_1 \rangle + \langle v_1, u_2 \rangle + \langle v_2, u_1 \rangle + \langle v_2, u_2 \rangle$$

$$\langle cv, u \rangle = c\langle v, u \rangle$$

$$\langle v, cu \rangle = \overline{c}\langle v, u \rangle$$

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

$$\langle v, v \rangle \text{ is real and nonnegative}$$

$$\langle v, v \rangle = 0 \Rightarrow v = 0$$

Note that, unlike a bilinear form, an inner product is not linear in the second factor. Instead, it's "conjugate-linear" in the second factor. It is, however, linear in the second factor if $\mathbb{F} = \mathbb{R}$.

Example 110. An example of an inner product is the dot product on $\mathbb{R}^n$. Non-examples are the dot product on $\mathbb{C}^n$ (is not conjugate-linear in the second factor), and the Lorentzian product on $\mathbb{R}^2$ (violates the last two conditions).

Example 111. A new example is the "standard inner product on $\mathbb{C}^n$":

$$\left\langle \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \right\rangle := z_1\overline{w_1} + \cdots + z_n\overline{w_n}.$$

Example 112. Related to the last example, let V be the vector space of functions $f : X \to \mathbb{C}$ where $X = \{x_1, \ldots, x_n\}$ is a set with n elements. Let δ_{x_i} be the function which is 1 on x_i are 0 on the other points. Then

$$e_i \mapsto \delta_{x_i}$$

[4]Here if $c = x + iy \in \mathbb{C}$ then $\overline{c} = x - iy \in \mathbb{C}$. If $c \in \mathbb{R}$, then $c = \overline{c}$. The complex number $\overline{c}$ is called the "complex conjugate" of c and figures into something like the Pythagorean theorem:

$$|c|^2 := c\overline{c} = (x + iy)(x - iy) = x^2 + y^2.$$

Therefore $|c| := \sqrt{|c|^2}$ is the distance of c from the origin in $\mathbb{C}$ and generalizes the usual notion of absolute value in $\mathbb{R}$.

defines an isomorphism from $\mathbb{C}^n$ to V. The standard inner product on $\mathbb{C}^n$ corresponds to the following inner product on V:

$$\sum_{i=1}^{n} f(x_i)\overline{g(x_i)}.$$

Example 113. The last example can be generalized to spaces much bigger than X by turning the finite sum into an integral. Let $L^2(\mathbb{R}^n)$ be the vector space of functions $f : \mathbb{R}^n \to \mathbb{C}$ such that

$$\int_{\mathbb{R}^n} |f(x)|^2 dx < \infty.$$

Then it turns out that for $f, g \in L^2(\mathbb{R}^n)$, the quantity

$$\int_{\mathbb{R}^n} f(x)\overline{g(x)}dx$$

is finite. Therefore define

$$\langle f, g\rangle := \int_{\mathbb{R}^n} f(x)\overline{g(x)}dx.$$

This satisfies all the axioms of an inner product.[5]

Definition 114. A vector space plus an inner product, $(V, \langle\cdot,\cdot\rangle)$, is called an "inner product space".

Definition 115. Given an inner product space $(V, \langle\cdot,\cdot\rangle)$ the norm of a vector is defined to be

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

Note that $\|v\| = 0 \Rightarrow v = 0$ and $\|cv\| = |c|\|v\|$.

Definition 116. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Two vectors v, w are called orthogonal if $\langle v, w\rangle = 0$.

Note that in the usual dot product in $\mathbb{R}^n$, two vectors are orthogonal if they are at right angles to one another:

$$\frac{v \cdot w}{\|v\|\|w\|} = \cos\theta.$$

Proposition 117. *Given an inner product space $(V, \langle\cdot,\cdot\rangle)$ and nonzero vectors $v, w \in V$, then*

$$v - \frac{\langle v, w\rangle}{\|w\|^2} w$$

is orthogonal to w.

[5]There are some fine points to be made here that require a course in measure theory to resolve, but all of this works if you restrict to continuous functions $\mathbb{R}^n \to \mathbb{C}$.

Proof.

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

□

Proposition 118 (Cauchy-Schwarz)**.** $|\langle v, w\rangle| \leq \|v\|\|w\|$.

Proof.

$$0 \leq \|v\|^2 + \left\| v - \frac{\langle v, w\rangle}{\|w\|^2} w \right\|^2$$

$$= \|v\|^2 + \left\langle v, v - \frac{\langle v, w\rangle}{\|w\|^2} w \right\rangle + \left\langle v - \frac{\langle v, w\rangle}{\|w\|^2} w, v \right\rangle + \left\| v - \frac{\langle v, w\rangle}{\|w\|^2} w \right\|^2$$

$$= \|v\|^2 + \left\| v - \frac{\langle v, w\rangle}{\|w\|^2} w \right\|^2 = \|v\|^2 - \frac{\overline{\langle v, w\rangle}}{\|w\|^2}\langle v, w\rangle - \frac{\langle v, w\rangle}{\|w\|^2}\langle w, v\rangle + \frac{|\langle v, w\rangle|^2}{\|w\|^4}\|w\|^2$$

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

which implies that

$$|\langle v, w\rangle|^2 \leq \|v\|^2\|w\|^2.$$

□

The Pythagorean theorem has an analog for arbitrary inner product spaces:

Exercise 119. *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Suppose u and v are orthogonal. Then*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Definition 120. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. A collection of vectors $v_1, \ldots, v_k$ is called "orthonormal" if

$$\langle v_i, v_j\rangle = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

An orthonormal basis is, of course, a basis which is orthonormal. Any vector in V has a simple expression in terms of an orthonormal basis:

Proposition 121. *Let $(V, \langle\cdot,\cdot\rangle)$ be a finite-dimensional inner product space with an $u_1, \ldots, u_n$ an orthonormal basis. Then for any $v \in V$:*

$$v = \langle v, u_1\rangle u_1 + \langle v, u_2\rangle u_2 + \cdots + \langle v, u_n\rangle u_n.$$

Proof. Write

$$v = \sum_i a_i u_i.$$

Then

$$\langle v, u_j \rangle = \sum_i a_i \langle u_i, u_j \rangle.$$

By orthonormality, the only term in the sum that is nonzero is when $i = j$ and therefore

$$\langle v, u_j \rangle = a_j.$$

□

Proposition 122. *Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. An orthonormal collection of vectors is linearly independent.*

Proof. Let $a_1 u_1 + \cdots + a_k u_k = 0$ be a linearly relation amongst linearly independent orthonormal vectors. Then

$$0 = \langle 0, u_j \rangle = \sum_i a_i \langle u_i, u_j \rangle = a_j.$$

□

Gram-Schmidt. Let $(V, \langle \cdot, \cdot \rangle)$ be a finite-dimensional inner product space. The Gram-Schmidt procedure produces an orthonormal basis of a inner product space from any basis. Let $(v_1, \ldots, v_n)$ be a basis of $(V, \langle \cdot, \cdot \rangle)$ an inner product space. Gram-Schmidt iteratively defines vectors $u_1, \ldots, u_n$ such that $(u_1, \ldots, u_n)$ is an orthonormal basis of V.

$$u_1 := \frac{v_1}{\|v_1\|}.$$

Note that u_1 is defined by scaling v_1 to be have norm 1.

$$u_2 := \frac{v_2 - \langle v_2, u_1 \rangle u_1}{\|v_2 - \langle v_2, u_1 \rangle u_1\|}.$$

Note that u_2 is defined as in Proposition 117 by subtracting off the right scalar multiple of u_1 so as the make it orthogonal to u_1, then rescaling to make it have norm 1.

$$u_3 := \frac{v_3 - \langle v_3, u_2 \rangle u_2 - \langle v_3, u_1 \rangle u_1}{\|v_3 - \langle v_3, u_2 \rangle u_2 - \langle v_3, u_1 \rangle u_1\|}.$$

Note that u_3 is defined by subtracting off the right scalar multiples of u_1 and u_2 to make it orthogonal to those two vectors, then rescaling to make it have norm 1.

$$u_4 := \frac{v_4 - \langle v_4, u_3 \rangle u_3 - \langle v_4, u_2 \rangle u_2 - \langle v_4, u_1 \rangle u_1}{\|v_4 - \langle v_4, u_3 \rangle u_3 - \langle v_4, u_2 \rangle u_2 - \langle v_4, u_1 \rangle u_1\|}.$$

At this point the reader should be able to understand the pattern. One stops once one produces u_n. At the end there are n orthonormal vectors, hence n linearly independent vectors of an n-dimensional vector space. Therefore $(u_1, \ldots, u_n)$ forms a basis.

Example 123. Consider $\mathbb{R}^2$ with the dot product and the basis

$$v_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \ v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Gram-Schmidt is applied as follows:

$$\|v_1\| = \sqrt{2^2 + 0^2} = 2$$

so

$$u_1 = \frac{v_1}{\|v_1\|} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Since

$$v_2 \cdot u_1 = 2$$

then

$$v_2 - (v_2 \cdot u_1)u_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and $\|v_2 - (v_2 \cdot u_1)u_1\| = 1$ so

$$u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

As a sanity check, one indeed sees that (u_1, u_2) is an orthonormal basis.

Definition 124. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Let $U \subset V$ be a subspace. Define

$$U^\perp = \{v \in V | \langle v, u \rangle = 0 \ \forall u \in U\}.$$

That is, the set of vectors in V whose inner product with everything in U is zero. $U^\perp$ is called the "orthogonal complement" of U.

In $\mathbb{R}^2$ and $\mathbb{R}^3$ with the dot proudct, $v\dot{w} = 0$ implies that v and w are at right angles, so geometrically U is perpendicular to $U^\perp$.

Example 125. Consider $\mathbb{R}^3$ with the dot product. Let $U = \text{Span}(1, 1, 1)$. Then

$$U^\perp = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} | a + b + c = 0 \right\} = \left\{ \begin{pmatrix} a \\ b \\ -a - b \end{pmatrix} | a, b \in \mathbb{R} \right\}.$$

Therefore

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

forms a basis of $U^\perp$.

Proposition 126. *Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then*

$$V = U \oplus U^\perp.$$

Proof. First note that if $u \in U \cap U^\perp$ then $\langle u, u\rangle = 0$, so that (by one of the axioms of an inner product space) $u = 0$. Therefore $U \cap U^\perp = \{0\}$.

Let $u_1, \ldots, u_k$ be an orthonormal basis of U. Extend to a basis to all of V: $(u_1, \ldots, u_k, v_{k+1}, \ldots, v_n)$. Apply Gram-Schmidt to this basis to produce an orthonormal basis $(u_1, \ldots, u_n)$. Note that Gram-Schmidt does nothing to the first k vectors (since they're already orthonormal) and that $u_{k+1}, \ldots, u_n$ are orthogonal to all the previous vectors, hence in $U^\perp$. Since $V = \mathrm{Span}(u_1, \ldots, u_n) \subset U + U^\perp$ it follows that $V = U + U^\perp$. □

Definition 127. Let V be an inner product spcae. Since $V = U \oplus U^\perp$, any vector $v \in V$ can be written uniquely as $v = u + w$ for some $u \in U$ and $w \in U^\perp$. The vector u is called the "orthogonal projection of v onto U". Let

$$P_U : V \to V$$

be the (linear) map that sends each vector to its orthogonal projection.

Remark 128. Note that $(U^\perp)^\perp = U$. Therefore $P_U + P_{U^\perp} = \mathrm{id}_V$.

Given an orthonormal basis of $U \subset V$ there's a simple formula for the orthogonal projection to U. For simplicity, assume that V is finite-dimensional. Let $(u_1, \ldots, u_k)$ be a basis for U. As in the proof of Proposition 126, extend it to an orthonormal basis $(u_1, \ldots, u_k, u_{k+1}, \ldots, u_n)$ of V. Therefore $(u_{k+1}, \ldots, u_n)$ is an orthonormal basis of $U^\perp$. Given $v \in V$ write

$$v = \langle v_1, u_1\rangle u_1 + \cdots + \langle v_k, u_k\rangle u_k + \langle v, u_{k+1}\rangle u_{k+1} + \cdots + \langle v, u_n\rangle u_n.$$

Since since

$$\langle v_1, u_1\rangle u_1 + \cdots + \langle v_k, u_k\rangle u_k \in U$$
$$\langle v, u_{k+1}\rangle u_{k+1} + \cdots + \langle v, u_n\rangle u_n \in U^\perp$$

then

$$P_U(v) = \langle v, u_1\rangle u_1 + \cdots + \langle v, u_k\rangle u_k.$$

Proposition 129. *Let V be an inner product space, v a vector in V, and U a subspace. Then $\|P_U(v) - v\| \leq \|u - v\|$ for all $u \in U$. Equality only occurs when $u = P_U(v)$.*

Proof. Write $v = u' + w$ where $u' \in U$ and $w \in U^\perp$. It is enough to show that $\|P_U(v) - v\|^2 \leq \|u - v\|^2$.

$$\|P_U(v) - v\|^2 = \|u' - (u' + w)\|^2 = \|w\|^2.$$

$$\|u - v\|^2 = \|u - (u' + w)\|^2 = \|(u - u') - w\|^2$$
$$= \|(u - u')^2\|^2 - \langle w, u - u'\rangle - \langle u - u', w\rangle + \|w\|^2$$
$$= \|u - u'\|^2 + \|w\|^2 \geq \|w\|^2 = \|P_U(v) - v\|^2.$$

There's an equality only when $\|u - u'\| = 0$, which, by one of the axioms of an inner product, only occurs when $u - u' = 0$. □

Thinking of norm as measuring a distance, then this proposition can be informally phrased as saying that $P_U(v)$ is the closest point on U to v.

Example 130. Consider $\mathbb{R}^2$ with the dot product. What is the closest point on the line $2x + 3 = 0$ to the point $(1, 3) \in \mathbb{R}^2$? $2x + 3y = 0$ is the span of $(-3, 2)$ so one wants to project $(1, 3)$ onto the span of $(-3, 2)$. Let $v = (1, 3)$ and

$$u_1 = \frac{1}{\sqrt{13}} \begin{pmatrix} -3 \\ 2 \end{pmatrix}$$

be an orthonormal basis for $U := \text{Span}(3, 2)$. Then

$$P_U(v) = \langle v, u_1 \rangle u_1 = \left(\begin{pmatrix} 1 \\ 3 \end{pmatrix} \cdot \frac{1}{\sqrt{13}} \begin{pmatrix} -3 \\ 2 \end{pmatrix} \right) \frac{1}{\sqrt{13}} \begin{pmatrix} -3 \\ 2 \end{pmatrix} = \frac{3}{13} \begin{pmatrix} -3 \\ 2 \end{pmatrix}.$$

Linear Regression. Suppose that you collect pairs of data points (x_i, y_i), $1 \leq i \leq n$. For example, you might get the heights (in inches) and weights (in pounds) of n people and x_i might be the ith person's height and y_i the ith person's weight. As with the discussion of correlation above, instead of thinking of the data as n pairs of points, one can think of it as two vectors

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \; y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

in $\mathbb{R}^n$. The following is a recipe that one learns in a first course on statistics. Suppose that you want to find a constant C such that, on average,

$$\text{weight in pounds } = C \times \text{ height in inches.}$$

C might be around 2, but is not going to be exactly 2. A reasonable thing to do would be to find the value of C that is most consistent with your observed data. One way to do this is to the find the value of C such that

$$\sum_i |y_i - Cx_i|^2$$

is minimized. Note that all of your datapoints satisfy $y_i = Cx_i$ then this quantity is 0. The lower this quantity is, the closer your data is to satisfying $y_i = Cx_i$ exactly for each i.

Given the vectors x and y in $\mathbb{R}^n$, finding a value C that minimizes $\sum_i |y_i - Cx_i|^2$ is the same thing as finding C such that minimizes $\|y - Cx\|^2$, where the norm here is from the dot product on $\mathbb{R}^n$. This is the same thing as finding the point on the span of x closest to y. Therefore, C can be calculated by

$$Cx = P_{\text{Span}(x)}(y).$$

Often you don't expect a relationship as simple as weight $= C \times$ height but instead a relation of the form weight $= C \times$ height $+ D$ for some constants C and D. To adapt the above this situation, instead project onto the span of x and $(1, 1, \ldots, 1)$. Details are left to the interested reader.

Definition 131. Recall that the transpose $M^\top$ of a matrix M is defined by $M^\top_{ij} := M_{ji}$. If M is $m \times n$ then $M^\top$ is $n \times m$.

Proposition 132. *Let V and W be inner product spaces. If $\langle Tv, w\rangle = \langle Sv, w\rangle$ for all $v \in V$ and $w \in W$ then $T = S$.*

Proof. Let $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$ be orthonormal bases for V and W. Let M be the matrix for T with respect to these bases and N be matrix for S with respect to these bases, i.e.:

$$T(v_i) = \sum_j M_{ji} w_j$$

$$S(v_i) = \sum_j N_{ji} w_j.$$

Since the basis w_j is orthonormal,

$$\langle T(v_i), w_k = \sum_j M_{ji} \langle w_j, w_k\rangle = M_{ki}$$

$$\langle S(v_i), w_k = \sum_j N_{ji} \langle w_j, w_k\rangle = N_{ki}$$

Therefore if

$$\langle Tv, w\rangle = \langle Sv, w\rangle$$

for all v, w then in particular this holds for the chosen bases of the $M = N$ and so $T = S$. □

Because of this proposition one, given $T : V \to W$ a linear map of inner product spaces, one can define $T^* : W \to V$ by

$$\langle Tv, w\rangle = \langle v, T^* w\rangle, \ \forall v, w.$$

Definition 133. T^* is called the adjoint of T.

Let $(v_1, \ldots, v_n)$ be an orthonormal basis of V and $(w_1, \ldots, w_m)$ an orthonormal basis of W. As in the proof of the preceding proposition, the matrix M for T has entries

$$M_{ji} = \langle T(v_i), w_j\rangle.$$

Let N be the matrix for T^*. Then, similarly,

$$N_{ij} = \langle T^*(w_j), v_i\rangle = \overline{\langle v_i, T^*(w_j)\rangle} = \overline{\langle T(v_i), w_j\rangle} = \overline{M_{ji}}.$$

Therefore $N = \overline{M^\top}$, the matrix for the adjoint is the complex conjugate transpose of the original matrix.

Remark 134. The map $T^* : W \to V$ is only defined when V and W are inner product spaces. If there are no inner products specified, then it does not make sense to speak of T^*.

Remark 135. The notation T^* has another usage with respect to dual spaces. These notes will not cover this usage, but it's standard enough in mathematics that it warrants a remark here. Given a vector space V, the dual of V^* is the set of linear maps $V \to \mathbb{F}$, i.e., $V^* = \mathcal{L}(V, \mathbb{F})$. Given a linear map $T : V \to W$, there's a map

$$T^* : W^* \to V^*$$

defined by

$$(T^*\phi)(v) := \phi(Tv)$$

for $\phi \in W^*$. This is not the same as the map T^* defined here. Again, these notes will not this definition of T^* and this remark is just for context in the greater mathematical world.

Proposition 136.

- $(S+T)^* = S^* + T^*$
- $(\lambda T)^* = \overline{\lambda} T^*$
- $(T^*)^* = T$
- $\mathrm{id}_V^* = \mathrm{id}_V$
- $(ST)^* = T^* S^*$.

Proof. These all have similar proofs. I'll just do the first one. For all $v \in V$ and $w \in W$:

$$\langle v, (S+T)^* w \rangle = \langle (S+T)v, w \rangle = \langle Sv, w \rangle + \langle Tv, w \rangle$$

$$= \langle v, S^* w \rangle + \langle v, T^* w \rangle = \langle v, (S^* + T^*)w \rangle.$$

Since

$$\langle v, (S+T)^* w \rangle = \langle v, (S^* + T^*)w \rangle$$

for all $v \in V$ and $w \in W$ then $(S+T)^* = S^* + T^*$. □

Proposition 137. $\ker(T^*) = (\operatorname{Im} T)^\perp$ *and* $\operatorname{im}(T^*) = (\ker(T))^\perp$.

Proof. These two equalities are similar so I'll just do the first one. I'll show that $\ker(T^*) \subset (\operatorname{im} T)^\perp$ and then $(\operatorname{im} T)^\perp \subset \ker(T^*)$.

Let $w \in \ker(T^*)$. Then $\langle T(v), w \rangle = \langle v, T^*(w) \rangle = 0$ so w is orthogonal to everything in the image of T.

Let $w \in (\operatorname{im} T)^\perp$. Then $0 = \langle w, T(v) \rangle = \langle T^* w, v \rangle$ for all $v \in V$. In particular, this holds for $v = T^* w$. Then $\langle T^* w, T^* w \rangle = 0$ so $T^* w = 0$. □

Definition 138. Let V be an inner product space. A map $T : V \to V$ is called self-adjoint if $T^* = T$.

Equivalently, self-adjoint maps are those maps such that $\langle Tv, u\rangle = \langle v, Tu\rangle$ for all $v, u \in V$.

An Analogy. adjoint:linear maps::complex conjguation:complex numbers

Proposition 139. *Let V be an inner product space and $T : V \to V$ self-adjoint. The eigenvalues of T are real.*

Proof. Suppose that $Tv = \lambda v$, then

$$\langle Tv, v\rangle = \lambda\langle v, v\rangle$$

but also

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

If v is an eigenvector then, by definition v is nonzero, so $\langle v, v\rangle \neq 0$. Therefore $\lambda = \overline{\lambda}$ so λ must be real. □

To Continue The Analogy. self-adjoint operators:linear maps::real numbers:complex numbers

Quantum Mechanics. Historically, the most significant application of self-adjoint operators comes from quantum mechanics. Quantum mechanics is the physical theory of very small particles. Whereas macroscopic objects are governed by Newton's laws, submicroscopic (think like a proton or an electron) objects are governed by a different set of laws. It turns out that linear algebra is the right mathematical language for these different sets of laws. In fact, when Werner Heisenberg first formulated quantum mechanics, he re-discovered the concept of a matrix.

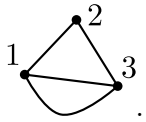
The strangest thing about quantum mechanics is the notion that you cannot always say "the particle is at the point x". Rather you might only be able to say something like "there is a a 30% chance the particle is at x and a 70% change the particle is at y."

Here is what is happening linear-algebraically. For simplicity assume the particles are constrained to lie on a line (instead of three-dimensional space) so their positions and momentums are real numbers. In classical mechanics a particle is described by its position x and its momentum p. The momentum is the particle's mass times its velocity, so the position-momentum pair is essentially describes where the particle is and where it is going. In quantum mechanics, a particle is instead described by a vector ψ in some complex inner product space V. For simplicity, one assumes that $\|\psi\|^2 = 1$. The vector space V depends on what sort of particle you're studying. There is a "position operator" $X : V \to V$ and a "momentum operator" $P : V \to V$. X and P are both self-adjoint and the (necessarily real) eigenvalues of X and P are the possible positions and momentums you might observe. When you observe a

particle, you're not guaranteed to get a predictable position or momentum. Rather you'll get one randomly selected from the eigenvalues of X and P. The chance a particular eigenvalue is selected depends on what the vector ψ is. Exactly how this randomness works is still mysterious today.

What can be said is that $\langle \psi, X\psi \rangle$ is the average position. That is, if you you set-up many identical particles in the state ψ, measure the positions in each of them, and find the average then you end up with $\langle \psi, X\psi \rangle$. Similarly $\langle \psi, P\psi \rangle$ is the average momentum. If ψ happens to be an eigenvector of X with eigenvalue x then there is a 100% chance you measure the position as x. Similarly for eigenvectors of P. The Heisenberg uncertainty principle, which states that you cannot know the position and momentum of a particle at the same time, can be restated linear-algebraically as the fact that X and P do not have any common eigenvectors.

Self-Adjoint Operators and Graphs. A more recent and simpler application of self-adjoint operators comes from the theory of graphs. A graph[6] is a collection of dots (called vertices) and edges connecting those dots, e.g.,



Here I have numbered the vertices. Graphs model a set (the set of vertices) plus relationships between the elements (the edges). For example the vertices might model people and the edges friendships. Or the vertices might model webpages and edges links. The adjaceny matrix of a graph Γ, $A(\Gamma)$ is a matrix where $A(\Gamma)_{ij}$ is the number of edges between i and j. Note that the adjacency matrix needs a numbering of the vertices. For example:

$$A\left(\begin{array}{c}\text{2}\\ \text{1}\quad\text{3}\end{array}\right) = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

$A(\Gamma)$ is always symmetric and therefore self-adjoint for the inner product on $\mathbb{R}^n$. Here n is the number of vertices of Γ. Note that

$$A(\Gamma)_{ij}A(\Gamma)_{jk}$$

is the number of paths in Γ from i to k that pass through j. Hence

$$(A(\Gamma)^2)_{ik} = \sum_j A(\Gamma)_{ij}A(\Gamma)_{jk}$$

is the total number of two-step paths from i to k. Similarly, the ijth entry of $A(\Gamma)^n$ is the number of n step paths from i to j. Recall that computing

[6]Terminology note: sometimes people call this a "mulitgraph" and the reserve the term "graph" for the situation where each edge is determined by its endpoint vertices.

the powers of matrices is made easy by finding eigenvectors. Therefore there is a solution to the combinatorial problem of finding n step paths from i to j solved in terms of the eigenvalues of $A(\Gamma)$. This is a recurring theme in graph theory: to Γ you associate some matrix and determine properties of that graph from the eigenvalues of that matrix.

It turns out that self-adjoint operators are diagonalizable. In fact, self-adjoint operators are part of a larger class of operators called "normal operators". Normal operators are diagonalizable. The proof will involve several propositions.

Definition 140. Let V be an inner product space. $T : V \to V$ is called normal if $T^*T = TT^*$.

Example 141. Suppose $T^* = T$ and $T^* = -T$ or $T^* = T^{-1}$. Then T is normal.

The next few propositions will all be preparations for proving that normal operators are diagonalizable.

Proposition 142. *Let V be a complex inner product space and $T : V \to V$ any linear map. Then $\langle Tv, v\rangle = 0$ if and only if $T = 0$.*

Proof. Assume $\langle Tw, w\rangle = 0$ for all $w \in V$.

$$0 = \langle T(Tv + v), Tv + v\rangle = \langle T^2v, Tv\rangle + \langle Tv, Tv\rangle + \langle Tv, v\rangle + \langle T^2v, v\rangle.$$

By assumption, $\langle T^2v, Tv\rangle = 0$ and $\langle Tv, v\rangle = 0$. Therefore

$$\langle Tv, Tv\rangle + \langle T^2v, v\rangle = 0.$$

A similar computation for

$$0 = \langle T(Tv + iv), Tv + iv\rangle$$

shows that

$$\langle Tv, Tv\rangle - \langle T^2v, v\rangle = 0$$

and hence $\langle Tv, Tv\rangle = 0$, so that $Tv = 0$. Since this holds for any v, it follows that $T = 0$. □

Proposition 143. *T is normal if and only if $\|Tv\| = \|T^*v\|$.*

Proof.

$$T \text{ normal} \Leftrightarrow T^*T - TT^* = 0 \Leftrightarrow \langle (T^*T - TT^*)v, v\rangle = 0$$

$$\Leftrightarrow \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \Leftrightarrow \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle.$$

□

Exercise 144. *If $T : V \to V$ is normal then $(T - \lambda \operatorname{id}_V)^* = T^* - \overline{\lambda} \operatorname{id}_V$.*

Proposition 145. *Let $T : V \to V$ be normal. Then T and T^* have the same eigenvectors. In fact, if $Tv = \lambda v$ then $T^*v = \overline{\lambda}v$.*

Proof. Suppose v is an eigenvector of T with eigenvalue λ. Then

$$\|(T - \lambda \operatorname{id}_V)v\| = 0 \Rightarrow \|(T - \lambda \operatorname{id}_V)^* v\| = 0 \Rightarrow \|(T^* - \overline{\lambda} \operatorname{id}_V)v\| = 0.$$

Remember that a vector of norm zero must be the zero vector. □

Proposition 146. *Let T be normal and let v_1, v_2 be eigenvectors with eigenvalues λ_1, λ_2. Suppose $\lambda_1 \neq \lambda_2$. Then $\langle v_1, v_2 \rangle = 0$.*

Proof. $\langle Tv_1, v_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \lambda_1 \langle v_1, v_2 \rangle$. On the other hand, $\langle Tv_1, v_2 \rangle = \langle v_1, T^* v_2 \rangle = \langle v_1, \overline{\lambda_2} v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle$. Since $\lambda_1 \neq \lambda_2$, it must be the case that $\langle v_1, v_2 \rangle = 0$. □

Exercise 147. *Let $T : V \to V$. Show that U is an invariant subspace for T^* if and only if $U^\perp$ is an invariant subspace for T.*

Theorem 148 (Spectral Theorem for Normal Operators)**.** *Let V be a complex inner product space. Then $T : V \to V$ is normal if and only if T has an orthonormal eigenbasis.*

Proof. Suppose T has an orthonormal eigenbasis. By the last proposition these are also orthonormal eigenvectors for T^* and hence with respect to this basis the matrices for T and T^* are

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \lambda_n \end{pmatrix}, \quad \begin{pmatrix} \overline{\lambda_1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \overline{\lambda_n} \end{pmatrix}.$$

Hence both T^*T and TT^* are repsented by the same matrix

$$\begin{pmatrix} |\lambda_1|^2 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & |\lambda_n|^2 \end{pmatrix}$$

so $T^*T = TT^*$.

The proof of the other direction is by induction on $\dim(V)$. The case $\dim(V) = 1$ is easy. Assume that if a normal operator on a vector space of $\dim(V) - 1$ has an orthonormal eigenbasis. Let $v \in V$ be an eigenvector of T. This is the place where the fact that V is a complex vector space is used. Normalize v:

$$u_1 := \frac{v}{\|v\|}$$

and let $U = \operatorname{Span}(u_1)$. Since u_1 is an eigenvector for T and since T is normal, u_1 is an eigenvector for T^*. Therefore U is invariant for T^*. Therefore $U^\perp$ is invariant for T. Therefore T, restricted to $U^\perp$, (denoted $T|_{U^\perp}$) is a normal operator on a vector space of dimension $\dim(V) - 1$. Therefore $T|_{U^\perp}$ has an orthonormal eigenbasis $(u_2, \ldots, u_n)$. Since this is an orthonormal basis of $U^\perp$, the basis $(u_1, u_2, \ldots, u_n)$ is an orthonormal eigenbasis for T. □

This proof doesn't work for the case of real numbers becuase T might not have an eigenvector. However, if T is self-adjoint you can get an eigenvector for T:

Theorem 149 (Spectral Theorem for Self-Adjoint Operators)**.** *Let V be an inner product space. $T : V \to V$ is self-adjoint if and only if it has an orthonormal eigenbasis with real eigenvalues.*

Proof. If T has an orthonormal basis with real-eigenvalues then the matrix for T with respect to this basis is diagonal with real entries. The matrix for the adjoint will be complex conjugate transpose of the matrix for T. Since the eigenvalues of T are real, these two matrices are the same, so $T^* = T$.

Conversely, suppose T is self-adjoint and pick a basis $(v_1, \ldots, v_n)$ of V. Let M be the matrix for T with respect to $(v_1, \ldots, v_n)$. M can be thought of as a linear map $M : \mathbb{R}^n \to \mathbb{R}^n$. If $(a_1, \ldots, a_n) \in \mathbb{R}^n$ is an eigenvector for M then $a_1 v_1 + \cdots + a_n v_n$ is an eigenvector for T. Therefore, in order to find an eigenvector for T, it is enough to find an eigenvector for M.

The key point is that M can also be thought of as a linear map $M : \mathbb{C}^n \to \mathbb{C}^n$. Let v be an eigenvector for M in $\mathbb{R}^n$ with eigenvalue λ. Note that

$$\overline{Mv} = \overline{\lambda v} \Rightarrow M\overline{v} = \lambda\overline{v}$$

since M has real entries and λ is real (remember T is self-adjoint, so its eigenvalues, and hences the eigenvalues of M, are real). Therefore $\overline{v}$ is also an eigenvector of M of eigenvalue λ and therefore so are

$$\mathrm{Re}(v) = \frac{v + \overline{v}}{2} \text{ and } \mathrm{Im}(v) = \frac{v - \overline{v}}{2i}.$$

Since v is nonzero, at least one of these is an eigenvector for M that lies in $\mathbb{R}^n$.

Therefore T has an eigenvector. The rest of the proof of the theorem is the same as in the complex case. □

Definition 150. Let V be an inner product space. A linear map $T : V \to V$ is called positive semidefinite if it is self-adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$. It is called positive definite if it is self-adjoint and $\langle Tv, v\rangle > 0$ for all $v \neq 0$.

Example 151. Let M be an $n \times n$ diagonal matrix with nonnegative entries. Equip $\mathbb{R}^n$ with the dot product. Then M, viewed as a map $\mathbb{R}^n \to \mathbb{R}^n$, is positive semidefinite. If the diagonal entries are all positive, the M is positive definite.

The previous example in fact is typical of positive operators.

Proposition 152. *Let V be a finite-dimensional inner product space. $T : V \to V$ is positive semidefinite if and only if it is self-adjoint and all its eigenvalues are real and nonnegative.*

Proof. Suppose T is positive semidefinite. Let v be an eigenvector. Then $\langle Tv, v\rangle = \lambda\langle v, v\rangle$ is nonnegative, so λ is nonnegative.

Suppose that T is self-adjoint and all its eigenvalues are nonnegative. Let $(v_1, \ldots, v_n)$ be an orthonormal eigenbasis. Write

$$v = \sum_i a_i v_i.$$

Then

$$\langle Tv, v\rangle = \sum_i \lambda_i |a_i|^2 \geq 0.$$

□

Square Roots. Let $T : V \to V$ be positive semidefinite. Since T is self-adjoint (as part of the definition of positive semidefinite), then T has an orthonormal eigenbasis $(v_1, \ldots, v_n)$. Let λ_i be the eigenvalue for v_i. Since $\lambda_i \geq 0$, $\sqrt{\lambda_i}$ makes sense. Define $T^{1/2} : V \to V$ by $T^{1/2}(v_i) = \sqrt{\lambda_i} v_i$. $(v_1, \ldots, v_n)$ is an orthonormal eigenbasis for $T^{1/2}$ and, since the eigenvalues are nonnegative, $T^{1/2}$ is positive semidefinite.

Here is a useful example of the square root. Equip $\mathbb{R}^2$ with the dot product and let $A : \mathbb{R}^2 \to \mathbb{R}^2$ be a positive definite map. In particular, 0 is not an eigenvalue for A, so A is invertible. A homework assignment is to show that $\langle x, y\rangle_A := x \cdot Ay$ defines a *different* inner product on $\mathbb{R}^2$. It therefore gives a different notion of distance on $\mathbb{R}^2$. You can ask, what are the sets of points equidistant from the origin? That is, what are the points $x \in \mathbb{R}^2$ such that $\|x\|_A^2 = c$ (c is some constant)? You can do this by hand by writing out a matrix for A and, as on the question from the worksheet from discussion, you'll find that the constant norm sets are conic sections that turn out to be ellipses. Exactly why ellipses appear is explained by the square root. Since A is positive with respect to the dot product, it has a positive square root $A^{1/2}$. Then $\langle x, y\rangle = (A^{1/2}x) \cdot (A^{1/2}y)$. Let

$$X = \{x \in \mathbb{R}^2 | \|x\|_A^2 = c\}.$$

Then

$$X = \{x \in \mathbb{R}^2 | (A^{1/2}x) \cdot (A^{1/2}x) = c\} = \{A^{-1/2}y | y \cdot y = c\}.$$

Here $A^{-1/2}$ is the inverse of $A^{1/2}$. The set of points $y \in \mathbb{R}^2$ such that $y \cdot y = c$ is a circle, so X is $A^{-1/2}$ applied to a circle. Since $A^{-1/2}$ is positive with respect to the dot product, it has an orthonormal eigenbasis. Each of these eigendirections is stretched by some positive amount, and the two eigendirections are at right angles. Therefore $A^{-1/2}$ applied to a circle gives an ellipse, so X is an ellipse.

Proposition 153. *Let V be a finite-dimensional inner product space and $T : V \to V$ positive semidefinite. Then T has a unique positive semidefinite square root.*

Proof. The previous discussion already shows that T has a positive semidefinite square root. It remains to show that it is unique.

Let R be positive semidefinite and such that $R^2 = T$. Since R is positive semidefinite, it has an orthonormal eigenbasis $(v_1, \ldots, v_n)$ such that $Rv_i = \mu_i v_i$ where $\mu_i \geq 0$. Note that $Tv_i = R^2 v_i = \mu_i^2 v_i$ so $(v_1, \ldots, v_n)$ forms an orthonormal eigenbasis of T. Therefore the eigenvalues of T are $\lambda_i = \mu_i^2$. And hence R can is the map defined by $Rv_i = \sqrt{\lambda_i} v_i$. □

Definition 154. Let V be an inner product space. An isometry $S : V \to V$ is a linear map such that $\|Sv\| = \|v\|$ for all $v \in V$.

Said informally: S is an isometry if it preserves distances.

Proposition 155. *S is an isometry of V if and only if $\langle Sv, Sw\rangle = \langle v, w\rangle$ for all $v, w \in V$.*

Proof. One direction follows from the definition of isometry. For the other, suppose that S is an isometry. If $\mathbb{F} = \mathbb{R}$, then

$$\langle v, w\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}$$

so if $\|Su\| = \|u\|$ for all u, then $\langle Sv, Sw\rangle = \langle v, w\rangle$ for all $v, w \in V$. There's a similar formula for a complex inner product spaces that shows the same result for the complex case. □

Proposition 156. *Let V be a finite-dimensional inner product space. S is an isometry if and only if it is invertible and $S^{-1} = S^*$.*

Proof. Suppose that S is an isometry. If $\{u_1, \ldots, u_k\}$ is an orthonormal set then so is $\{Su_1, \ldots, Su_k\}$ since $\langle Su_i, Su_j\rangle = \langle u_i, u_j\rangle$. Since orthornomal sets are linearly independent it follows that $\{Su_1, \ldots, u_k\}$ is linearly independent. In particular, if $(u_1, \ldots, u_n)$ is an orthonormal basis of V then so is $(Su_1, \ldots, Su_n)$. Therefore S is invertible (its inverse is defined by sending Su_i to u_i).

Remember that if

$$\langle T_1 v, w\rangle = \langle T_2 v, w\rangle$$

for all $v, w \in V$, then $T_1 = T_2$. Since

$$\langle S^* Sv, w\rangle = \langle Sv, Sw\rangle = \langle v, w\rangle = \langle \mathrm{id}_V\, v, w\rangle$$

then $S^* S = \mathrm{id}_V$ hence $S^{-1} = S^*$.

Conversely, suppose that $S^* S = \mathrm{id}_V$. Then $\langle S^* Sv, w\rangle = \langle v, w\rangle$ for all $v, w \in V$ and hence $\langle Sv, Sw\rangle = \langle v, w\rangle$ for all $v, w \in V$. □

Example 157. Consider $\mathbb{R}^n$ with the dot product and let $S : V \to V$ be an isometry. Since S takes an orthonormal basis to an orthonormal basis, since the standard basis is an orthonormal basis for the dot product, and since the

ith column of the matrix for S with respect to the standard basis vector is Se_i, then the columns of this matrix form an orthonormal basis. Such a matrix is typically called an "orthogonal matrix" (though of course it really should be called an "orthonormal matrix").

Example 158. Consider $\mathbb{C}^n$ with its standard inner product and let $S : V \to V$ be an isometry. Similarly to the last example, the columns of the matrix for S with respect to the standard basis are orthonormal. Such a matrix is called a "unitary matrix".

Another way of stating the fact that self-adjoint operators have orthonormal eigenbases is the following: if $T : V \to V$ is self-adjoint there exists an orthonormal basis of V, call it $(v_1, \ldots, v_n)$, such that T takes each v_i to a multiple of itself. Of course, self-adjoint operators on an inner product space are a special kind of operator and many transformations you run into in nature aren't self-adjoint. However, the following remarkable fact is true for any transformation between inner product spaces: if $T : V \to W$ is a linear transformation between inner product spaces, there exist orthonormal bases $(v_1, \ldots, v_n)$ of V and $(w_1, \ldots, w_m)$ of W such that T takes each v_i to a multiple of w_i. If $n > m$ then $v_{m+1}, \ldots, v_n$ are sent to zero. This is formally stated in the following theorem:

Theorem 159 (Singular Value Decomposition)**.** *Let V and W be finite-dimensional inner product spaces and let $T : V \to W$ be any linear map. Then there exist orthonormal bases $(v_1, \ldots, v_n)$ of V and $(w_1, \ldots, w_m)$ of W and nonnegative numbers s_i such that*

$$T(v) = \sum_i s_i \langle v, v_i \rangle w_i.$$

Here the sum is over $1 \leq i \leq n$ if $m \geq n$ and is over $1 \leq i \leq m$ if $m \leq n$.

Remark 160. Note that $T(v_i) = s_i w_i$.

I delay the proof of this theorem for a moment. The values s_i are called the "singular values" of the transformation T. There is a unique set of them, in fact:

Proposition 161. *The values s_i are the nonnegative square roots of the eigenvalues of T^*T.*

The proposition will follow immediately from the proof of Theorem 159.

Remark 162. The singular value decomposition theorem can be stated in this way: there exist numbers s_i and orthonormal bases of V and W such that

the matrix with respect to these bases is of the form

$$\begin{pmatrix} s_1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & s_m & 0 & \cdots & 0 \end{pmatrix} \text{ or } \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & s_n \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

depending on whether $m \leq n$ or $m \geq n$.

Example 163. Consider $\mathbb{R}^2$ with the dot product. The matrix

$$\begin{pmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{pmatrix}$$

is not diagonalizable. However, it does take a pair of orthonormal vectors in $\mathbb{R}^2$ to a different pair of orthogonal vectors. Here is the pair:

$$\begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix}, \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}.$$

These are sent to

$$\begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix}, \begin{pmatrix} 3\sqrt{3}/2 \\ 3/2 \end{pmatrix}$$

respectively. Therefore, if we let

$$v_1 = \begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix}, \ v_2 = \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}$$

$$w_1 = \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix}, \ w_2 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}$$

and call the transformation T then (v_1, v_2) and (w_1, w_2) are each orthonormal bases of $\mathbb{R}^2$ and

$$Tv_1 = w_1, \ Tv_2 = 3w_2.$$

Writing an arbitrary vector in terms of v_1 and v_2:

$$v = \langle v, v_1 \rangle v_1 + \langle v, v_2 \rangle v_2$$

then shows that

$$T(v) = \langle v, v_1 \rangle w_1 + 3\langle v, v_2 \rangle w_2.$$

The singular values of T are 1 and 3.

One might wonder how the singular vectors and singular values in the last example were computed. Here is the trick:

- Find an orthonormal eigenbasis of T^*T. This is $(v_1, \ldots, v_n)$.
- Reorder the basis so that $(v_1, \ldots, v_k)$ are the eigenvectors corresponding to nonzero eigenvalues.
- For $1 \leq i \leq k$ set $w_i = \frac{Tv_i}{\|Tv_i\|} = \frac{Tv_i}{\sqrt{\lambda_i}}$.
- Extend $(w_1, \ldots, w_k)$ to an orthonormal basis $(w_1, \ldots, w_m)$ of W.

That this trick works follows from the proof of the Theorem 159. Here's a fully worked example:

Example 164. Consider $\mathbb{R}^2$ with the dot product and the matrix

$$T = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix}$$

which is a projection onto the subspace spanned by $(1, -1)$, but is not an orthogonal projection. Since

$$\begin{pmatrix} 2 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} 8 & 4 \\ 4 & 2 \end{pmatrix}$$

and this matrix has eigenvalues 10 and 0, the singular values of T are $\sqrt{10}$ and 0. An orthonormal eigenbasis for T^*T is

$$v_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \; v_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

where $\lambda_1 = 10$ and $\lambda_2 = 0$. Therefore set

$$w_1 = \frac{1}{\sqrt{10}} T v_1 = \frac{1}{\sqrt{50}} \begin{pmatrix} 5 \\ -5 \end{pmatrix}.$$

And let w_2 be any unit length vector orthogonal to w_1, for example

$$w_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then

$$Tv_1 = \sqrt{10} w_1$$

$$Tv_2 = 0.$$

In decimal approximations:

$$s_1 = 3.1623$$

$$s_2 = 0$$

$$v_1 = \begin{pmatrix} .8944 \\ .4472 \end{pmatrix}$$

$$v_2 = \begin{pmatrix} .4472 \\ -.8944 \end{pmatrix}$$
$$w_1 = \begin{pmatrix} .7071 \\ -.7071 \end{pmatrix}$$
$$w_2 = \begin{pmatrix} .7071 \\ .7071 \end{pmatrix}.$$

Example 165. Consider changing the matrix for T just a little bit

$$S := \begin{pmatrix} 2.1 & 1.3 \\ -2.2 & -1.1 \end{pmatrix}$$

Then the singular values and singular vectors of S can be computed:

$$s_1 = 3.4821$$
$$s_2 = .1580$$
$$v_1 = \begin{pmatrix} .8732 \\ .4874 \end{pmatrix}$$
$$v_2 = \begin{pmatrix} .4874 \\ -.8732 \end{pmatrix}$$
$$w_1 = \begin{pmatrix} .7056 \\ .7086 \end{pmatrix}$$
$$w_2 = \begin{pmatrix} .7086 \\ -.7056 \end{pmatrix}.$$

As might be expected, the singular values and singular vectors didn't change by much. What's interesting here is that, since s_2 is so small, S can be approximated by the map $\widetilde{S}$ defined by

$$\widetilde{S}(v_1) = 3.4821 w_1$$
$$\widetilde{S}(v_2) = 0.$$

This is a simpler map, algebraically, since its image is one-dimensional. If S represents some observed data or relationships, then $\widetilde{S}$ represents a lower-dimensional approximation of the same data or relationships. See, for example, the next example.

Example 166. The singular values and singular vectors are used in image processing. Consider, for example, an 8×6 pixel image of the number 0. It might be represented in a matrix in the following form:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Each entry, viewed as a number from 0 to 1, represents how gray the pixel is, where 0 is completely white and 1 is completely black. Viewed as a transformation $\mathbb{R}^6 \to \mathbb{R}^8$, the image of the matrix is dimension 2, and therefore you'd expect two nonzero singular values.

If the image of the number 0 had some noise, the matrix might look like

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .98 & 1 & .98 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ .06 & 1 & 0 & .02 & 1 & 0 \\ 0 & 1 & .01 & 0 & 1 & 0 \\ 0 & 1 & .01 & 0 & .99 & 0 \\ 0 & 1 & .96 & .98 & 1 & 0 \\ 0 & 0 & 0 & 0 & .05 & 0 \end{pmatrix}$$

The dimension of the image of this matrix is not 2. You'd expect many nonzero singular values. In fact

$$s_1 = 3.68$$

$$s_2 = 1.4974$$

$$s_3 = .0549$$

$$s_4 = .0376$$

$$s_5 = .0191$$

$$s_6 = 0.$$

(The vectors $(v_1, \ldots, v_6)$ and $(w_1, \ldots, w_8)$ can also be computed, though I won't list them here.) While this new matrix has a 5-dimensional image, its image is still "approximately" 2-dimensional: two of singular values are much larger than the others. What this means is that this matrix can be approximated by the transformation

$$v_1 \mapsto 3.68 w_1$$

$$v_2 \mapsto 1.4974 w_2$$

$$v_3 \mapsto 0$$

$$v_4 \mapsto 0$$

$$v_5 \mapsto 0$$

$$v_6 \mapsto 0$$

One would naively expect this approximation to still have a matrix, with respect to the standard basis, resembling "0". Indeed, here it is:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .98 & .99 & .99 & 1 & 0 \\ .02 & 1 & 0 & 0 & 1 & 0 \\ .01 & 1 & .01 & .01 & 1 & 0 \\ .01 & 1 & .01 & .01 & 1 & 0 \\ .01 & 1 & .01 & .01 & .99 & \\ 0 & 1 & .97 & .97 & 1 & 0 \\ 0 & .02 & 0 & 0 & .02 & 0 \end{pmatrix}.$$

The 0 is still easy to see, but this matrix can be encoded in very few pieces of information. One need only specify the vectors v_1, v_2, w_1, and w_2 and the two singular values s_1 and s_2. This is a total of $6+6+8+8+1+1 = 30$ pieces of information, less than the 48 required to represent the original matrix. As the matrices become much larger, this sort of approximation becomes more and more efficient.

Proof of Theorem 159. The trick is to let $(v_1, \ldots, v_n)$ be an orthonormal eigenbasis of T^*T. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues. Since T^*T is self-adjoint, such an eigenbasis always exists. Moreover, T^*T is positive semidefinite:

$$\langle T^*Tv, v\rangle = \|Tv\|^2 \geq 0$$

so that λ_i is real and nonnegative for all i.

Reorder the eigenbasis so that $v_1, \ldots, v_k$ are the eigenvectors with nonzero eigenvalues. For $1 \leq i \leq k$, define

$$w_i := \frac{Tv_i}{\sqrt{\lambda_i}}.$$

Note that $\{Tv_1, \ldots, Tv_k\}$ are linearly independent: if

$$\sum_{i=1}^{k} a_i Tv_i = 0$$

then

$$\sum_{i=1}^{k} a_i T^*Tv_i = 0 \Rightarrow a_i\lambda_i v_i = 0$$
$$\Rightarrow a_i\lambda_i = 0\ \forall i \Rightarrow a_i = 0\ \forall i.$$

Since for $k+1 \leq i \leq n$, $T^*Tv_i = 0 \Rightarrow Tv_i = 0$, the set $\{v_1, \ldots, v_k\}$ spans $\mathrm{im}(T)$. Hence $(T(v_1), \ldots, T(v_k))$ forms a basis for $\mathrm{im}(T)$. Since

$$\langle Tv_i, Tv_j\rangle = \langle T^*Tv_i, v_j\rangle = \lambda_i\langle v_i, v_j\rangle$$

and $(v_1, \ldots, v_n)$ is orthonormal, then $(Tv_1, \ldots, Tv_k)$ is an orthogonal collection. For $1 \leq i \leq k$, let

$$w_i := \frac{Tv_i}{\|Tv_i\|} = \frac{Tv_i}{\sqrt{\lambda_i}}.$$

For $k+1 \leq i \leq m$, let $w_{k+1}, \ldots, w_m$ be an orthonormal extension to a basis of W. Then

$$T(v) = T\left(\sum_{i=1}^{n} \langle v, v_i \rangle v_i\right) = \sum_{i=1}^{n} \langle v, v_i \rangle T(v_i) = \sum_{i=1}^{k} \sqrt{\lambda_i} \langle v, v_i \rangle w_i$$

$$= \sum_{i=1}^{\min(m,n)} \sqrt{\lambda_i} \langle v, v_i \rangle w_i$$

where the last equality uses the fact that $\sqrt{\lambda_i} = 0$ for $k+1 \leq i \leq m$. □

6 Changing Bases

Most linear transformations $T : \mathbb{F}^n \to \mathbb{F}^m$ are described as matrices written in terms of the standard basis vector. That is, T is presented to the reader by the matrix M such that

$$T(e_i) = \sum_{j=1}^{m} M_{ji} e_j.$$

For example, suppose that the matrix for T with respect to the standard basis is

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

This describes the map

$$e_1 \mapsto 2e_1 + e_2$$

$$e_2 \mapsto e_1 + 2e_2.$$

Often, though, the most useful basis is not the standard basis. For example,

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \; v_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

is (using the dot product on $\mathbb{R}^2$) an orthonormal eigenbasis for T. Then the matrix for T with respect to (v_1, v_2) is

$$\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

You can relate the two matrix descriptions of T. Define a transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$S(e_1) = v_1, \; S(e_2) = v_2.$$

Then

$$S^{-1}(v_1) = e_1,\ S^{-1}(v_2) = e_2.$$

Then

$$S^{-1}TS(e_1) = S^{-1}Tv_1 = S^{-1}3v_1 = 3e_i.$$

Similarly $S^{-1}TS(e_2) = e_2$. Therefore the matrix for $S^{-1}TS$ with respect to the standard basis is a diagonal matrix. Let P be the matrix for S (with respect to the standard basis) and M the matrix for T (with respect to the standard basis). Then the matrix for $S^{-1}TS$ is

$$P^{-1}MP.$$

Since

$$S(e_1) = v_1 = 2e_1 + e_2$$
$$S(e_2) = v_2 = e_1 + 2e_2$$

the columns of P are the eigenvectors of S. Explicitly:

$$P = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}.$$

$$P^{-1} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}.$$

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$P^{-1}MP = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

This generalizes to an arbitrary diagonalizable map $T : \mathbb{F}^n \to \mathbb{F}^n$:

Proposition 167. *Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be diagonalizable with eigenbasis $(v_1, \ldots, v_n)$. Define S by*

$$S(e_i) = v_i$$

. Then the matrix P for S with respect to the standard basis vectors is the matrix whose ith column is v_i. Also if M is the matrix for T with respect to the standard basis vectors then

$$P^{-1}MP$$

is a diagonal matrix whose ith diagonal entry is the eigenvalue for v_i.

Proof. Write $v_i = \sum_j a_j e_j$. Then $S(e_i) = \sum_j a_j e_j$ and, by definition the entries $(a_1, \ldots, a_n)$ form the ith column of the matrix for S with respect to the standard basis.

Since $S(e_i) = v_i$ and $T(v_i) = \lambda_i v_i$, then

$$S^{-1}TS(e_i) = \lambda_i e_i$$

so that the matrix for $S^{-1}TS$ (which is $P^{-1}MP$) is diagonal with $\lambda_1, \ldots, \lambda_n$ down the diagonal. □

Remark 168. Computing inverses of matrices can be annoyingly difficult to do by hand. However, if you're in $\mathbb{R}^n$ with the dot product or $\mathbb{C}^n$ with the standard inner product, and the matrix P has orthonormal columns, then P is an isometry of $\mathbb{F}^n$ and hence its inverse is its complex-conjugate transpose. For example

$$\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^{-1} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

This being said, it might be helpful to know the formula for inverting an arbitrary 2×2 matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

If $ad - bc = 0$ then the matrix is not invertible.

Proposition 169. *Equip $\mathbb{R}^n$ with the dot product. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map with singular vectors $(v_1, \ldots, v_n)$, $(w_1, \ldots, w_m)$ and singular values $(s_1, \ldots, s_{\min(m,n)})$. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be defined by*

$$F(e_i) = v_i$$

and let $G : \mathbb{R}^m \to \mathbb{R}^m$ be defined by

$$G(e_i) = w_i.$$

Let $\mathbf{V}$ and $\mathbf{U}$ be the matrices (with respect to the standard basis) for F and G, respectively. Let Σ be the matrix

$$\begin{pmatrix} s_1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & s_m & 0 & \cdots & 0 \end{pmatrix} \text{ or } \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & s_n \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

depending on whether $m \leq n$ or $n \leq m$. Then

$$\mathbf{U}\Sigma\mathbf{V}^\top$$

is the matrix for T with respect to the standard basis.

Proof. Since $T(v_i) = s_i w_i$ for $1 \leq i \leq \min(m, n)$ and 0 otherwise, it follows that

$$G^{-1}TF(e_i) = \begin{cases} s_i e_i & 1 \leq i \leq \min(m,n) \\ 0 & \text{otherwise} \end{cases}$$

Hence the matrix for $G^{-1}TF$, with respect to the standard basis, is Σ. Let M be the matrix for T with respect to the standard basis. Then

$$\mathbf{U}^{-1}M\mathbf{V} = \Sigma.$$

Since $\mathbf{V}$ and $\mathbf{U}$ have orthonormal columns, they are isometries and hence $\mathbf{V}^{-1} = \mathbf{V}^\top$, $\mathbf{U}^{-1} = \mathbf{U}^\top$, so

$$M = \mathbf{U}\Sigma\mathbf{V}^\top.$$

□

The decomposition $M = \mathbf{U}\Sigma\mathbf{V}^\top$ is called the "singular value decomposition" of the matrix M and explains the usage of the term "decomposition".

As mentioned before, one can approximate T by only using the largest few singular values. For example, suppose you want to approximate T by a linear map of rank 5. If the singular values are listed in decreasing order, then this amounts to setting all but the first five diagonal entries of Σ equal to 0, all but the first five columns of $\mathbf{V}$ equal to 0, and all but the first five columns of $\mathbf{U}$ equal to 0.

7 The Determinant

A permutation is a bijection from the set $\{1, \ldots, n\}$ to itself. For example

$$1 \mapsto 2$$

$$2 \mapsto 1$$

is a permutation of $\{1, 2\}$. There are $n!$ permutations of $\{1, \ldots, n\}$. Notation for permutations can be cumbersome. I like "cycle notation", which is best explained by an example

$$(143)(67)$$

this is a permutation of $\{1, \ldots, 7\}$ that sends 1 to 4, 4 to 3, 3 to 1 as well 6 to 7 and 7 to 6. It sends 2 and 5 to themselves. Another example:

$$(1234)$$

is the permutation of $\{1, 2, 3, 4\}$ that sends 1 to 2, 2 to 3, 3 to 4, and 4 to 1. I'll call the identity permutation simply e, so there are two permutations of $\{1, 2\}$

$$e,\ (12)$$

and six permutations of $\{1, 2, 3\}$

$$e,\ (12),\ (13),\ (23),\ (123),\ (132).$$

The set of permutations of $\{1, \ldots, n\}$ is denoted by S_n (and is referred to as "the symmetric group").

Definition 170. Let $x_1, x_2, \ldots, x_n$ be n variables. Given $\sigma \in S_n$, define

$$\operatorname{sign}(\sigma) = \frac{\prod_{i<j}(x_{\sigma(i)} - x_{\sigma(j)})}{\prod_{i<j}(x_i - x_j)}.$$

This is always either 1 or -1.

To see that $\operatorname{sign}(\sigma) = \pm 1$, note that the numerator and denominator have the same $\frac{n(n-1}{)}2$ factors, just that some might have been multiplied by -1. For example, for $n = 3$:

$$\operatorname{sign}((123)) = \frac{(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)}$$

$$= \frac{(x_2 - x_3)(x_2 - x_1)(x_3 - x_1)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = 1$$

and

$$\operatorname{sign}((12)) = \frac{(x_2 - x_1)(x_2 - x_3)(x_1 - x_3)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = -1.$$

Since $\operatorname{sign}(\sigma)$ is either the constant function 1 or -1, then if you apply a permutation τ to $\operatorname{sign}(\sigma)$ (i.e., to *both* the numerator and denominator) then it does not change. Explicitly:

$$\operatorname{sign}(\sigma) = \frac{\prod_{i<j}(x_{\tau(\sigma(i))} - x_{\tau(\sigma(j))})}{\prod_{i<j}(x_{\tau(i)} - x_{\tau(j)})}.$$

Therefore

$$\operatorname{sign}(\sigma)\operatorname{sign}(\tau) = \frac{\prod_{i<j}(x_{\tau(\sigma(i))} - x_{\tau(\sigma(j))})}{\prod_{i<j}(x_{\tau(i)} - x_{\tau(j)})} \frac{\prod_{i<j}(x_{\tau(i)} - x_{\tau(j)})}{\prod_{i<j}(x_i - x_j)} = \operatorname{sign}(\sigma\tau).$$

A transposition is a permutation which switches precisely two elements of $\{1, \ldots, n\}$. For example, (12), (23), and (13) are the transpositions of S_n.

Exercise 171. *The sign of a transposition is* -1.

So if $\sigma = \tau_1\tau_2\cdots\tau_k$ where each τ_i is a transposition, then $\operatorname{sign}(\sigma) = (-1)^k$.

Example 172. Since the permutations (123) and (132) in S_3 are each the product of two transpositions, their signs are $+1$.

Definition 173. Let V be a vector space over $\mathbb{F}$. A map $\phi : V \times V \times \cdots \times V \to \mathbb{F}$ is multilinear if it is linear in each factor. That is,

$$\phi(v_1, \ldots, v_{i_1}, aw + bu, v_{i+1}, \ldots, v_k)$$

$$= a\phi(v_1, \ldots, v_{i_1}, w, v_{i+1}, \ldots, v_k) + b\phi(v_1, \ldots, v_{i_1}, u, v_{i+1}, \ldots, v_k)$$

for each i.

Definition 174. Let V be a vector space over $\mathbb{F}$. A map $\phi : V \times V \times \cdots \times V \to \mathbb{F}$ is called alternating if switching two arguments multiplies it by -1:

$$\phi(v_1, v_2, \ldots, w, \ldots, u, \ldots v_k) = -\phi(v_1, v_2, \ldots, u, \ldots, w, \ldots, v_k).$$

Example 175. Define $\phi : \mathbb{F}^2 \times \mathbb{F}^2 \to \mathbb{F}$ by

$$\phi\left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) = ad - bc.$$

The reader can check that this is alternating and multilinear.

Remark 176. If $\phi : \underbrace{V \times V \times \cdots \times V}_{k \text{ times}} \to \mathbb{F}$ is alternating and $\sigma \in S_k$, then

$$\phi(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(k)}) = \text{sign}(\sigma)\phi(v_1, \ldots, v_k)$$

because σ can be written as a product of transpositions and applying each transposition separately multiplies ϕ by -1.

Remark 177. If $\phi : V \times V \times \cdots \times V \to \mathbb{F}$ is alternating, then

$$\phi(v_1, \ldots, w, \ldots, w, \ldots v_k) = -\phi(v_1, \ldots, w, \ldots, w, \ldots, v_k)$$

$$\Rightarrow \phi(v_1, \ldots, w, \ldots, w, \ldots v_k) = 0.$$

So ϕ is nonzero only when all of its arguments are different.

Theorem 178. *There exists a unique multilinear alternating map*

$$\det : \underbrace{\mathbb{F}^n \times \mathbb{F}^n \times \cdots \mathbb{F}^n}_{n \text{ times}} \to \mathbb{F}$$

such that $\det(e_1, e_2, \ldots, e_n) = 1$.

The map det in the theorem is called the determinant. Note that an element of $\underbrace{\mathbb{F}^n \times \mathbb{F}^n \times \cdots \mathbb{F}^n}_{n \text{ times}}$ can be represented by a matrix M: the ith factor is the ith column of M.

Proof of Theorem 178. This is a proof best done backwards. That is, it is best to suppose that det exists and simplify it a bit. This will provide some intuition and then, with enough intuition, the proof will be easy.

Think of the list of n vectors $(v_1, \ldots, v_n)$ as a matrix M. That is

$$v_i = \sum_j M_{ji} e_j.$$

Then

$$\det(M) := \det(v_1, \ldots, v_n) = \det\left(\sum_{i_1} M_{i_1 1} e_{i_1}, \sum_{i_2} M_{i_2 2} e_{i_2}, \ldots, \sum_{i_n} M_{i_n n} e_{i_n}\right).$$

By the multilinearity property of det, this is equal to

$$= \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} M_{i_1 1} M_{i_2 2} \cdots M_{i_n n} \phi(e_{i_1}, e_{i_2}, \ldots, e_{i_n}).$$

This is a sum with n^n terms. However, because det is alternating, whenever two of the indices are the same, i.e., $i_k = i_j$ for some k and j, then $\phi(e_{i_1}, e_{i_2}, \ldots, e_{i_n}) = 0$. Therefore the only terms that contibute nonzero amounts are where $i_1, i_2, \ldots, i_n$ contains all the elements in $\{1, \ldots, n\}$. That is to say, the only terms that contribute nonzero amounts are where $i_1, i_2, \ldots, i_n$ is a permutation of $1, 2, \ldots, n$. That is to say, the only terms that contribute nonzero amounts are where $i_1, i_2, \ldots, i_n = \sigma(1), \sigma(2), \ldots, \sigma(n)$ for some $\sigma \in S_n$. Therefore

$$\det(M) = \sum_{\sigma} M_{\sigma(1)1} M_{\sigma(2)2)} \cdots M_{\sigma(n)n} \det(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}).$$

Also by the alternating property,

$$\det(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}) = \operatorname{sign}(\sigma).$$

So

$$\det(M) = \sum_{\sigma} \operatorname{sign}(\sigma) M_{\sigma(1)1} M_{\sigma(2)2)} \cdots M_{\sigma(n)n} \det(e_1, e_2, \ldots, e_n)$$

and since $\det(e_1, e_2, \ldots, e_n) = 1$, then

$$\det(M) = \sum_{\sigma} \operatorname{sign}(\sigma) M_{\sigma(1)1} M_{\sigma(2)2} \cdots M_{\sigma(n)n}.$$

Remember that we started this proof supposing that det exists but really we want to prove that it exists. At this point, the choice is obvious. Simply *define*

$$\det(M) := \sum_{\sigma} \operatorname{sign}(\sigma) M_{\sigma(1)1} M_{\sigma(2)2} \cdots M_{\sigma(n)n}.$$

Then you can check by hand that det is alternating, multilinear, and satisfies $\det(e_1, e_2, \ldots, e_n) = \det(I) = 1$. I'll just check that it's multilinear. Let $\tau \in S_n$ be a transposition. I'd like to show that

$$\sum_{\sigma} \operatorname{sign}(\sigma) M_{\sigma(1)1} M_{\sigma(2)2} \cdots M_{\sigma(n)n} = \sum_{\sigma} \operatorname{sign}(\sigma) M_{\sigma(1)\tau(1)} M_{\sigma(2)\tau(2)} \cdots M_{\sigma(n)\tau(n)}$$

The right hand side is $\det(M')$ where M' is M but with τ applied to the columns. Since permutations are invertible, $\sigma = (\sigma\tau^{-1})\tau$.

$$\sum_{\sigma} \operatorname{sign}(\sigma) M_{\sigma(1)\tau(1)} M_{\sigma(2)\tau(2)} \cdots M_{\sigma(n)\tau(n)}$$

$$= \sum_{\sigma} \operatorname{sign}(\sigma) M_{(\sigma\tau^{-1})\tau(1)\tau(1)} M_{(\sigma\tau^{-1})\tau(2)\tau(2)} \cdots M_{(\sigma\tau^{-1})\tau(n)\tau(n)}$$

Applying τ to the second argument has simply scrambled up the factors in each summand, so this is the same thing as

$$\sum_{\sigma} \operatorname{sign}(\sigma) M_{(\sigma\tau^{-1})(1)1} M_{(\sigma\tau^{-1})(2)2} \cdots M_{(\sigma\tau^{-1})(n)n}.$$

Since τ is fixed, as σ varies over all permutations, so does $\sigma\tau^{-1}$. Therefore setting $\sigma' = \sigma\tau^{-1}$ shows this is equal to

$$\sum_{\sigma'} \operatorname{sign}(\sigma'\tau) M_{\sigma'(1)1} M_{\sigma'(2)2} \cdots M_{\sigma'(n)n}$$

$$= \operatorname{sign}(\tau) \sum_{\sigma'} M_{\sigma'(1)1} M_{\sigma'(2)2} \cdots M_{\sigma'(n)n} = -\det(M).$$

□

Exercise 179. *Prove that* $\det(M) = \det(M^\top)$.

Corollary 180. $\det(M)$ *is alternating the in the rows too: if* M' *is obtained from* M *by switching two rows, then* $\det(M') = -\det(M)$.

There's an inductive way to compute the determinant called "expansion by minors". For simplicity, here it is described by expanding the down the first column:

Proposition 181.

$$\det(M) = \sum_{i=1}^{n} (-1)^{i-1} M_{i1} \det(M(i))$$

where here $M(i)$ *is the* $(n-1) \times (n-1)$ *matrix obtained from* M *by removing the 1st column and the* i*th row.*

Proof. Let $M_{(i)}$ be obtained from M by replacing the first column of M by e_i. Then by multilinearity of det,

$$\det(M) = \sum_{i=1}^{n} M_{i1} M_{(i)}.$$

Note that

$$M_{(i)} = \begin{pmatrix} 0 & M_{12} & M_{13} & \cdots & M_{1n} \\ 0 & M_{22} & M_{23} & \cdots & M_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & M_{(i-1)2} & M_{(i-1)3} & \cdots & M_{(i-1)n} \\ 1 & M_{i2} & M_{i3} & \cdots & M_{in} \\ 0 & M_{(i+1)2} & M_{(i+1)3} & \cdots & M_{(i+1)n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & M_{n2} & M_{n3} & \cdots & M_{nn} \end{pmatrix}.$$

After switching the ith row $(i-1)$ times with the row above it you get

$$\begin{pmatrix} 1 & M_{i2} & M_{i3} & \cdots & M_{in} \\ 0 & M_{12} & M_{13} & \cdots & M_{1n} \\ 0 & M_{22} & M_{23} & \cdots & M_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & M_{(i-1)2} & M_{(i-1)3} & \cdots & M_{(i-1)n} \\ 0 & M_{(i+1)2} & M_{(i+1)3} & \cdots & M_{(i+1)n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & M_{n2} & M_{n3} & \cdots & M_{nn} \end{pmatrix}.$$

and it's not hard to see that the determinant of this matrix is $\det(M(i))$. By the last corollary, det is alternating in the rows, so $\det(M_{(i)}) = (-1)^{i-1}\det(M(i))$. □

A similar result works, but with an arbitrary column instead of the first column. Details are left to the interested reader.

Note that alternating multilinear maps

$$\underbrace{\mathbb{F}^n \times \cdots \times \mathbb{F}^n}_{n \text{ times}} \to \mathbb{F}$$

form a vector space with addition

$$(\phi + \phi')(v_1, \ldots, v_n) := \phi(v_1, \ldots, v_n) + \phi'(v_1, \ldots, v_n)$$

and scalar multiplication

$$(c\phi)(v_1, \ldots, v_n) := c\phi(v_1, \ldots, v_n).$$

Call this vector space $\mathcal{A}$. The proof of Theorem 178 shows that there exists a nonzero element in $\mathcal{A}$ (the determinant) and that any element in $\mathcal{A}$ is a multiple of the determinant: if $\phi \in A$, then the theorem says that $\frac{1}{\phi(I)}\phi = \det$. Therefore $\dim(\mathcal{A}) = 1$ and so any linear map $\mathcal{A} \to \mathcal{A}$ is multiplication by a scalar.

Proposition 182. *Given an $n \times n$ matrix M, define a map*

$$T_M : \mathcal{A} \to \mathcal{A}$$

by

$$(T_M\phi)(v_1, \ldots, v_n) = \phi(Mv_1, \ldots, Mv_n).$$

Then T_M is multiplication by $\det(M)$.

Proof. It is not hard to check that T_M is linear. Since det forms a basis of $\mathcal{A}$, $T_M \det = c \det$ for some constant c. To determine c, note that

$$(T_M \det)(e_1, \ldots, e_n) = \det(Me_1, \ldots, Me_n) = \det(M) = \det(M)\det(e_1, \ldots, e_n).$$

Therefore $c = \det(M)$. □

Proposition 183. $\det(MN) = \det(M)\det(N)$.

Proof. $T_N(T_M(\phi)) = \det(N)\det(M)\phi$. Explicitly, however,

$$(T_N(T_M(\phi)))(v_1, \ldots, v_n) = \phi(MNv_1, \ldots, MNv_n) = T_{MN}\phi = \det(MN)\phi$$

$$\Rightarrow \det(MN)\phi = \det(M)\det(N)\phi.$$

□

Proposition 184. $\det(M) \neq 0$ *if and only if* M *is invertible. If* M *is invertible, then* $\det(M^{-1}) = \frac{1}{\det(M)}$

Proof. If M is invertible, then $\det(M)\det(M^{-1}) = \det(MM^{-1}) = \det(I) = 1$.

If M is not invertible, then the columns are linearly dependent, so one column is a linear combination of the others. Then $\det(M) = 0$ by multilinearity. □

Corollary 185. λ *is an eigenvalue of* M *if and only if* $\det(M - \lambda I) = 0$. *Here* I *is the identity matrix.*

Definition 186. The characteristic polynomial of M is defined by $\det(tI - M)$. It is a polynomial of degree n in t.

By the last corollary, λ is a root of $\det(tI - M)$ if and only if λ is an eigenvalue of M.

Example 187. Suppose you're interested in finding the eigenvectors of

$$\begin{pmatrix} -7 & -8 & -9 \\ 0 & 1 & 0 \\ 6 & 6 & 8 \end{pmatrix}.$$

The eigenvalues are the roots of

$$\det \begin{pmatrix} -7-\lambda & -8 & -9 \\ 0 & 1-\lambda & 0 \\ 6 & 6 & 8-\lambda \end{pmatrix} = -(\lambda^3 - 2\lambda^2 - \lambda + 2) = -(\lambda^2 - 1)(\lambda - 2)$$

$$= -(\lambda - 1)(\lambda + 1)(\lambda - 2).$$

The eigenvectors can be calculated by solving

$$\begin{pmatrix} -7 & -8 & -9 \\ 0 & 1 & 0 \\ 6 & 6 & 8 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \lambda \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

for each of $\lambda = 1$, $\lambda = -1$, and $\lambda = 2$. For example, for $\lambda = 2$:

$$-7a - 8b - 9c = 2a$$

$$b = 2b$$

$$6a + 6b + 8c = 2c$$

implies that $b = 0$ and $a = -c$ so the $\lambda = 2$ eigenspace consists of vectors of the form

$$\begin{pmatrix} a \\ 0 \\ -a \end{pmatrix}.$$

Proposition 188. *Let M be an upper triangular matrix (that is, it's a matrix with only zeroes below the main diagonal). Then* $\det(M)$ *is the product of the diagonal entries of M.*

Proof. In the definition of the determinant,

$$\det(M) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) M_{\sigma(1)1} \cdots M_{\sigma(n)n}$$

the only term that doesn't include a factor from below the diagonal is the one corresponding to the identity permutation.

An alternate proof uses expansion by minors along the first column and induction on the size of the matrix. □

Given a linear map $T : \mathbb{C}^n \to \mathbb{C}^n$ you can find a basis $(v_1, \ldots, v_n)$ of $\mathbb{C}^n$ so that the matrix for T with respect to $(v_1, \ldots, v_n)$ is in Jordan normal form. Call this matrix J. Let P be the change of basis matrix

$$P(e_i) = v_i$$

and let M be the matrix for T with respect to the standard basis. Then

$$M = P^{-1}JP.$$

Therefore

$$\det(M) = \det(J) = \prod_i \lambda_i$$

where the here the product is over all the diagonal entries of J. Thus the determinant of M is the product of the eigenvalues, counted with multiplicity. Here the phrase "counted with multiplicity" just means that λ_i appears in the product possibly several times since it may appear on the diagonal of J multiple times.

Theorem 189 (Cayley-Hamilton)**.** *Let $p_M(t) = \det(tI - M)$ be the characteristic polynomial. Then $p_M(M) = 0$.*

Remark 190. This theorem famously has a false proof: simply plug M into $\det(tI - M)$ to get $p_M(M) = \det(MI - M) = \det(0) = 0$. This doesn't work because you want to show that $p_A(A)$ is the zero matrix and $\det(MI - M) = 0$ is the number zero. The lesson is that you need to be a little careful substituting matrices into variables.

Proof. Whether M is a real or complex matrix, you can always think of it as a complex matrix. Let J be the Jordan form for M. Let $\{\lambda_1, \ldots, \lambda_k\}$ be the numbers that appear on the diagonal of J.

Recall that the generalized λ_i-eigenspace, call it E_i, is the invariant subspace corresponding to all the Jordan blocks with λ_i on the diagonal. Let $\dim(E_i) = n_i$. Looking at the Jordan blocks it is not hard to see that $\lambda_i I - J$ is nilpotent and $(\lambda_i I - J)^{n_i}$ sends any vector in E_i to 0.

Note that $p_M(t) = \det(tI - M) = \det(P(tI - M)P^{-1}) = \det(tI - J) = \prod_{i=1}^{k}(t - \lambda_i)^{n_i}$. If I plug J into this polynomial I get

$$\prod_{i=1}^{k}(J - \lambda_i I)^{n_i}.$$

This matrix acts as zero on each E_i. Since

$$\mathbb{C}^n = \bigoplus_i E_i$$

that means that this matrix is the zero matrix. Therefore

$$\prod_{i=1}^{k}(J - \lambda_i I)^{n_i} = 0$$

so

$$\prod_{i=1}^{k}(P^{-1}(J - \lambda_i I)P)^{n_i} = 0$$

so

$$\prod_{i=1}^{k}(M - \lambda_i I)^{n_i} = 0$$

so

$$p_M(M) = 0.$$

□

Volume. It turns out determinants of $n \times n$ matrices with real coefficients are related to volume. The volume of a box in $\mathbb{R}^n$ of side lengths $(a_1, \ldots, a_n)$ (where each $a_i \geq 0$) is $a_1 a_2 \cdots a_n$. The volume of any region in $\mathbb{R}^n$ is defined by cutting it up into (potentially countably many) boxes and adding up their volumes.[7] If you apply a linear transformation to this box you get a region with n pairs of parallel opposite faces. Such a region is called a parallelepiped. (When $n = 2$ then you get a parallelogram.) If the original box is determined by sides $a_1 e_1, \ldots, a_n e_n$, then the parallelepiped is determined by sides $T(a_1 e_1), \ldots, T(a_n e_n)$. In general, a list $v_1, \ldots, v_n$, determines a parallelepiped. (If $(v_1, \ldots, v_n)$ is not a basis, then the parallelepiped will have some pairs of opposite faces smushed together.)

[7]There are some technical details here that are resolved by a course in measure theory. This process of measuring volume of subsets of $\mathbb{R}^n$ works for all reasonable subsets, but not for every subset.

Proposition 191. $|\det(v_1, \ldots, v_n)|$ *is the volume of the parallelpiped spanned by* $v_1, \ldots, v_n$.

The proof is delayed for a bit because the proof has an unlikely input: row reduction. There are three row operations:

(i) Add a multiple of one row to another row

(ii) Switch two rows

(iii) Multiply a row by a nonzero scalar

After performing these operations (perhaps many times over) on a matrix M, the matrix can be transformed into one of the form:

$$M' = \begin{pmatrix} 0 & I & A \\ 0 & 0 & 0 \end{pmatrix}$$

where I is the identity matrix, A is some matrix, and 0 stands for the 0 matrix. Note that, except for the matrix I, none of the blocks M' have to be square. The proof that row operations can always transform the matrix M into the form M' is not hard. It uses double induction on the dimensions of the matrix. I leave it to the interested reader.

The most important part about row operations for the present purposes is that each row operation in realized by left multiplication by a matrix.

(i) A row operation of type (i) corresponds to left multiplication by

$$I + aF_{ij}$$

where $a \in \mathbb{R}$ and F_{ij} is the matrix which is everywhere 0 except for a 1 in the ith entry. For example, in the 2×2 case:

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} M_{11} + aM_{21} & M_{12} + aM_{22} \\ M_{21} & M_{22} \end{pmatrix}$$

so multiplying by

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

on the left adds a times the second row to the first row.

A matrix E of this form is called an elementary matrix of type (i). Note that $\det(E) = 1$.

(ii) A row operation of type (ii) corresponds to left multiplication by a matrix which is the identity matrix except that the ith column is e_j and the jth column is e_i. The is the matrix which sends all the standard basis vectors to themselves except it switches e_i an e_j. For example

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} M_{21} & M_{22} \\ M_{11} & M_{12} \end{pmatrix}$$

so multiplying by

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

on the left switches the first and the second rows.

A matrix E of this form is called an elementary matrix of type (ii). Note that $\det(E) = -1$.

(iii) A row operation of type (iii) corresponds to left multiplication by a diagonal matrix with 1s down the diagonal except the ith entry is allowed to be any nonzero number a. For example

$$\begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ aM_{21} & aM_{22} \end{pmatrix}$$

so multiplying by

$$\begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}$$

on the left multiplies the second row by a.

A matrix E of this form is called an elementary matrix of type (iii). Note that $\det(E) = a$.

Proposition 192. *Suppose that M is a square matrix whose columns form a basis. Then M can be written as a product of elementary matrices of type (i), (ii), and (iii).*

Proof.

$$M' = E_1 E_2 \cdots E_k M$$

where M' is in the form as above. Since the right side is invertible, $M' = I$, so

$$M = E_k^{-1} E_{k-1}^{-1} \cdots E_1^{-1}.$$

□

Proof of Proposition 191. Note that an elementary matrix of type (i) or (ii) doesn't change volume, and the absolute value of its determinant is 1. Note that an elementary matrix of type (iii) with nontrivial entry a multiplies volume by $|a|$, and its determinant is a. Since

$$\det(M) = \det(E_1) \det(E_2) \cdots \det(E_k)$$

for some elementary matrices $E_1, \ldots, E_k$, it follows M changes volume by a factor of $|\det(M)|$. If $M = (v_1, \ldots, v_n)$, then M applied to the standard cube spanned by $(e_1, \ldots, e_n)$ is the parallelepiped spanned by $(v_1, \ldots, v_n)$. Therefore the volume of this parallelepiped is $|\det(M)|$. □