Math 340: Abstract Linear Algebra

Alex Wang

Summer 2023

Contents

1 Fields

Planned Lecture Date(s): June 21, 2023.

This course is a second course in linear algebra, and will approach things from a more theoretical perspective than a typical first course, which focuses primarily on matrices and matrix operations. In particular, I will try my best to use as few matrices as possible, to emphasize the abstraction of linear algebra away from the computational aspects.

Definition. A **field** is a set F with two binary operations, usually denoted $+$ and $\cdot$, such that

- Commutativity: $a + b = b + a$ and $a \cdot b = b \cdot a$.
- Associativity: $a + (b + c) = (a + b) + c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- Identity: There exists an element $0 \in F$ such that $0 + a = a$ for all $a \in F$, and similarly there exists an element $1 \in F$ such that $1 \cdot b = b$ for all $b \in F$.
- Inverses: For every element $a \in F$, there exists an element, denoted $-a$, such that $a + (-a) = 0$. Similarly, for every *non-zero* element $b \in F$, there exists an element, denoted b^{-1}, such that $b \cdot b^{-1} = 1$.
- Distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$.
- $0 \neq 1$.

These rules are often referred to as the *field axioms*. Since this resembles many number systems that we are familiar with, we often call these operations "addition" and "multiplication", and abbreviate them in the usual ways (for example, we will often abbreviate $a \cdot b$ to ab). Note that subtraction is addition of the inverse ($a - b = a + (-b)$), and division is multiplication by the inverse ($a/b = a \cdot b^{-1}$).

Example. The real numbers, denoted $\mathbb{R}$, form a field. Other fields you may be familiar with include $\mathbb{C}$, the complex numbers, and $\mathbb{Q}$, the rational numbers. However, the integers $\mathbb{Z}$ do not form a field, since 2 (for example) does not have a multiplicative inverse.

The idea of a field is that it provides us a setting in which to do "numerical operations": if we would like to be able to add, subtract, multiply, and divide, the field axioms are the "minimum" rules we need to do so. In particular, in the context of linear algebra, this is frequently all we need to make our theory work. However, fields can often look very different than the examples above.

Example. Let $\mathbb{F}_5$ denote the set $\{0, 1, 2, 3, 4\}$ together with the usual operations of addition and subtraction, with the caveat that all operations are taken *modulo 5* (with 0 and 1 being the additive and multiplicative identities). This means that if an operation would result in a number outside of the range 0 through 4, we can add or subtract multiples of 5 to bring it into the range. For example, here are some computations in $\mathbb{F}_5$:

- $2 + 2 = 4$, as usual.
- $2 + 3 = 0$, since $2 + 3 = 5$, but 5 does not exist in $\mathbb{F}_5$, so we can instead subtract 5 to obtain 0, which is in $\mathbb{F}_5$. This shows that an additive inverse of 2, denoted -2, is 3 (in this field).
- $2 \cdot 4 = 3$, since $2 \cdot 4 = 8$ and $8 - 5 = 3$, which is an element of $\mathbb{F}_5$.
- $3 \cdot 2 = 1$, since $3 \cdot 2 = 6$ and $6 - 5 = 1$, which is an element of $\mathbb{F}_5$. This shows that 2 is a multiplicative inverse for 3.
- By commutativity, we also know that $3 + 2 = 2 + 3 = 0$, so $-3 = 2$, and $2 \cdot 3 = 3 \cdot 2 = 1$, so $3^{-1} = 2$.

Note that taking numbers modulo 5 can also be thought of as taking the remainder when dividing by 5 (e.g., 29 modulo 5 is 4, since 29 divided by 5 is 5 with remainder 4).

It turns out that field identities and inverses are actually unique, but I'll leave this to homework (possibly) for you to prove it.

At this point, we haven't actually checked all of the axioms: while commutativity, associativity, distributivity, and identity are fairly straightforward to show, the existence of inverses is perhaps harder. For this case, we can simply check that $1 \cdot 1 = 2 \cdot 3 = 4 \cdot 4 = 1$, but in general, this may be more difficult.

This gives us our first example of a field that is not quite what we're used to - in fact, this field is *finite*! In contrast, the fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all infinite. It turns out that for any prime number p, the above construction will create a field $\mathbb{F}_p$ with the p elements $\{0, 1, \ldots, p-1\}$. These finite fields have numerous applications in number theory, cryptography, etc.

For any field F, if we start with our multiplicative identity 1_F (which I will abbreviate as 1), we can add it to itself repeatedly.

$$1 + 1 + \cdots + 1$$

In this sense, we can define the integer n to be the sum of 1 with itself n times. This allows us to talk about any integer in any field! For example, in $\mathbb{F}_5$, the integer 8 exists, and represents whatever element of the field that we get when we add 1 to itself 8 times.

Doing this in $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$ is somewhat boring, but something weird can happen in these finite fields we've constructed! In $\mathbb{F}_p$, what happens if we add 1 to itself repeatedly?

$$\underbrace{1 + \cdots + 1}_{p \text{ times}} = p = 0$$

This is very unusual! In the field $\mathbb{F}_p$, the integer p (and all of its multiples) is equal to 0.

Definition. We say that for a field F, the smallest number of times we can add 1 to itself to get 0 is called the **characteristic** of F. If this does not occur, then we say that the characteristic of F is 0.

For example, the characteristic of $\mathbb{F}_p$ is p, and the characteristic of $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all 0. For the most part in this class, we will focus on fields of characteristic 0, but much of the theory still works in general! Something we have to be very careful about is that when we divide, we need to make sure to not divide by p (since $p = 0$!).

Fields are also a general setting to solve equations: for example, I can ask about solutions to

$$x^2 - 4x + 3 = 0$$

and from either factoring or the quadratic formula, we know that solutions are given by $x = 1$ and $x = 3$. However, if we are handed an equation like

$$x^2 + 1 = 0$$

we know that over $\mathbb{R}$, this equation has no solutions, since the square of any real number must be non-negative. However, over $\mathbb{C}$, this equation does have a solution, which is i (or $-i$). What about over $\mathbb{F}_5$? It turns out that we have

$$2^2 + 1 = 4 + 1 = 5 \equiv 0 \pmod 5$$

and so this equation does indeed have a solution in $\mathbb{F}_5$. What about $\mathbb{F}_7$? Or $\mathbb{F}_{11}$? Answering the question of when $\mathbb{F}_p$ has a solution to $x^2 + 1 = 0$ is beyond the scope of this course, but is an introduction to the branch of math known as **number theory**.

One thing we can try is the quadratic formula, which says that for a polynomial $ax^2 + bx + c = 0$ (with $a \neq 0$), the roots are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

However, what does this equation actually mean? We know how to add, subtract, multiply, and divide, but the square root symbol is not something that must exist in our field axioms. We interpret this symbol to mean "what element of my field do I have to square to get the value inside?". However, sometimes the answer to this question is that there is *no element* which squares to what we want. For example, in $\mathbb{R}$, if we write $\sqrt{-1}$, no real number squares to -1, so the quadratic formula actually fails! We usually resolve this implicitly by passing to the complex numbers, where we have $\pm i$ to solve our equation. This is an example of **extending** our field to solve an equation, but I will not go into too much detail about this now.

Something else we should worry about is that we're dividing by $2a$. Although we require that $a \neq 0$, it's possible that $2 = 0$. What happens if $\text{char}(F) = 2$?

Let's denote polynomials with coefficients in the field F by $F[x]$. The fundamental theorem of algebra tells us that if we pick any polynomial $f(x) \in \mathbb{C}[x]$, it has a root in $\mathbb{C}$. However, this is very not true in $\mathbb{R}$ or $\mathbb{Q}$, as the polynomial

$$f(x) = x^2 + 1$$

has all coefficients in $\mathbb{Q}$ (which is a subset of $\mathbb{R}$), but as previously shown does not have any roots in $\mathbb{Q}$ (nor $\mathbb{R}$). A field which satisfies this property is very special.

Definition. A field F is said to be **algebraically closed** if any polynomial $f(x) \in F[x]$ has a root in F.

Solving polynomials is generally much easier in an algebraically closed field, since solutions are guaranteed to exist. A field F that is not algebraically closed can generally by extended to an **algebraic closure**, usually denoted $\overline{F}$, but the construction is beyond the scope of this course. For example, $\overline{\mathbb{R}} = \mathbb{C}$, but $\overline{\mathbb{Q}}$ is a very interesting field! Can you think of an example of an element of $\overline{\mathbb{Q}}$ which is not in $\mathbb{R}$?

2 Vector Spaces

Planned Lecture Date(s): June 23, 2023.

We now proceed to define the fundamental object of linear algebra.

Definition. A **vector space** over a field F is a set V together with two binary operations:

1. Addition: a map $V \times V \to V$ mapping (v, w) to $v + w$.
2. Scalar multiplication: a map $F \times V \to V$ mapping (a, v) to av.

These operations satisfy the following properties:

1. Associativity: $u + (v + w) = (u + v) + w$.
2. Commutativity: $u + v = v + u$.
3. Identity: There exists an element, denoted $\vec{0} \in V$, such that $v + 0 = v$ for all $v \in V$. Note that this is an element of V, different than $0 \in F$.
4. Inverse: For every $v \in V$, there is an element, denoted $-v$, such that $v + (-v) = \vec{0}$.
5. Compatibility: $a(bv) = (ab)v$, for $a, b \in F$ and $v \in V$.
6. Distributivity: $a(u + v) = au + av$ and $(a + b)v = av + bv$.
7. $1v = v$.

Elements of the vector space V are called **vectors**.

These axioms give us rules for what operations we are permitted to perform in a vector space. Note that a vector is *defined* to be an element of a vector space - the definitions that you may be familiar with are simply a consequence of this, as we'll see in a moment.

In practice, many of these axioms will come for free - the important axioms to check are that addition and scalar multiplication both land in the right place (we'll see examples of this).

Example. From your previous linear algebra class, $\mathbb{R}^n$ is a vector space over the field $\mathbb{R}$. In general, for any field F, the set F^n consisting of ordered n-tuples of elements of F is a vector space, under elementwise addition and multiplication by F.

In all of these cases, elements of the vector space are ordered lists of numbers, as you may have previously defined them. However, we can abstract away from this idea, and create some more interesting vector spaces.

Example. Let $F[x]$ be the set of polynomials with coefficients in F. We can check that

1. The polynomial $f(x) = 0$ is the zero vector in $F[x]$.
2. The sum of two polynomials with coefficients in F is again a polynomial with coefficients in F.
3. We can scale a polynomial by an element $a \in F$ by multiplying the polynomial by a, and this again gives us a polynomial with coefficients in F. In particular, choosing $a = -1$ gives an inverse.

We should also check the remaining axioms, but this can be done without much work, so I will omit the details here. This allows us to conclude that $F[x]$ is a vector space over F.

This example is not too fancy, since if we simply collect the coefficients of a polynomial $f(x) \in F[x]$, we can think of polynomials as "infinite lists" of elements of F. Some other examples in this vein include $M_{n\times m}(F)$, the set of $n \times m$ matrices over a field F, as well as the set of $\mathbb{R}$-valued sequences (x_n). However, we can do things that are a bit weirder.

Example. Let $C(\mathbb{R})$ be the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$. We can check that

1. The function $f(x) = 0$ is the zero vector in $C(\mathbb{R})$.
2. The sum of two continuous functions f and g is continuous, thus $f + g \in C(\mathbb{R})$.
3. For a real number $a \in \mathbb{R}$ and continuous function f, the function $(af)(x)$ is continuous, and in particular we have $-f \in C(\mathbb{R})$, which is the additive inverse of f.

Again, to be thorough, we should check the remaining axioms, but this can be verified in a fairly straightforward manner. We can then conclude that $C(\mathbb{R})$ is a vector space (over $\mathbb{R}$).

We can also impose conditions to "limit" our vector spaces: for example, instead of taking all continuous functions from $\mathbb{R}$ to $\mathbb{R}$, we could add some constraints.

Example. Which of the following are vector spaces (under the usual addition/scalar multiplication)?

- The set of continuous functions $f : \mathbb{R} \to \mathbb{R}$, with $f(0) = 0$.
- The set of continuous functions $f : \mathbb{R} \to \mathbb{R}$, with $f(0) = 1$.
- The set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ which are differentiable.
- The set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ which are infinitely differentiable, and furthermore for some n, the n-th derivative of f is 0.

The answer is that all of them except for the 2nd are vector spaces (why?). In fact, the last example is precisely $\mathbb{R}[x]$.

Note that all of these examples are vector spaces inside the bigger vector space $C(\mathbb{R})$. These are examples of **subspaces**, which we will explore next class.

We also have another interesting class of examples.

Example. The field of complex numbers, $\mathbb{C}$, is a vector space over itself, as we can think of it as $\mathbb{C}^n$ for $n = 1$. However, $\mathbb{C}$ can also be considered a vector space over $\mathbb{R}$. We check that

1. The complex number 0 is the zero vector in $\mathbb{C}$.
2. The sum of two complex numbers is again a complex number.
3. We can multiply any complex number by a real number to again obtain a complex number.

We again omit the checking of the remaining axioms. This shows that $\mathbb{C}$ is a vector space over $\mathbb{R}$!

This process generalizes to any field K which contains another field F: we can add elements of K, and we can scale elements of K by elements of F, since F is a subset of K, so we can multiply as if everything is in K. Note however that the field axioms of K allow us to multiply elements of K, but if we think of K as an F-vector space, we are only allowed (under the vector space axioms) to multiply elements of K by elements of F.

Let's prove some things about vector spaces.

Theorem. Let V be a vector space over a field F. Then,

- The additive identity is unique.
- The additive inverse is unique.
- For any $v \in V$, we have that $0v = 0$.
- For any $v \in V$, we have that $-v = (-1)v$.

Proof. We prove each of the above claims.

- Suppose that a and b are both additive identities. Then, we have that

$$a = a + b = b + a = b$$

and thus $a = b$.

- Fix $v \in V$. Suppose w and w' are both additive inverses. Then, we have that

$$w = w + 0 = w + (v + w') = (w + v) + w' = (v + w) + w' = 0 + w' = w'$$

and so $w = w'$.

- We have that

$$0v = (0 + 0)v = 0v + 0v$$

and adding $-(0v)$ (the additive inverse of $0v$) on both sides gives that $0 = 0v$.

- We have that

$$0 = 0v = (1 + (-1))v = 1v + (-1)v = v + (-1)v$$

and so we have found an inverse for v. Since additive inverses are unique, we have that $-v = (-1)v$.

□

We will often be sloppy and use 0 to denote both $0 \in F$ as well as $0 \in V$, but remember that these are different things.

The intuition to take away from vector spaces is that they are an object in which we can add and scale. We'll see that despite all these examples looking very different, the operations of addition and scalar multiplication force structure upon our vector spaces, which we will explore in this class.

3 Building New Vector Spaces out of Old Ones

Planned Lecture Date(s): June 26, 2023.

If we start with a vector space V over a field F, there are lots of ways we can create new ones! One such way is by taking a smaller vector space inside V.

Definition. A **vector subspace** of V is a subset $W \subset V$ which is itself a vector space over F, under the same operations of addition and scalar multiplication in V.

Example. Let's look at some easy examples of subspaces.

- Let V be a vector space. Then, the set $\{0\}$ is a subspace of V, known as the **zero subspace**.
- Let V be a vector space. Then, the entire space V is a subspace of V.
- Let $V = F^n$, and let W_i be the set
$$W_i = \{\vec{x} \in F^n \,:\, x_i = 0\}$$
for all $1 \leq i \leq n$. Then, W_i is a subspace of V.

Intuitively, we think of W_i as "being" F^{n-1}, since it's as if we just throw out the i-th coordinate. Geometrically, we can picture it as the plane $z = 0$ in $\mathbb{R}^3$, which we think of as essentially being $\mathbb{R}^2$. It turns out that these vector spaces are isomorphic, but we won't define what that means until a little bit later.

Example. Let $V = \mathbb{R}[x]$, and let W be the set of polynomials $f \in \mathbb{R}[x]$ with $f(0) = 0$. Then, W is a subspace of V.

Since the operations are inherited from V, if we want to check if $W \subset V$ is a vector space, most of the axioms are also inherited from V. The most important thing to check is that we can't "leave" W using our operations (in other words, W is closed under addition and scalar multiplication).

Theorem. Let V be a vector space, and let $W \subset V$ be a nonempty subset. Then, W is a subspace of V if and only if for every $v, w \in W$ and $c \in F$, we have that $cv + w \in W$.

Proof. Suppose W is a subspace. Then, $cv + w \in W$ since W is closed under addition and scalar multiplication.

Suppose that $cv + w \in W$ for all $v, w \in W$ and $c \in F$. Since W is nonempty, choose $v \in W$, and so $(-1)v + v = 0 \in W$. Then, choosing $c = 1$ shows that W is closed under vector addition, and choosing $w = 0$ shows that W is closed under scalar multiplication. Note additionally that $-v = (-1)v$, so choosing $c = -1$ shows that W has additive inverses. The remaining axioms are inherited from the analogous axioms in V. □

This gives us a convenient way to construct subspaces: given a larger vector space V, we can construct a subspace W simply by imposing a condition, and checking that it is preserved in the above manner.

Example. Let V be the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. Let $W = C(\mathbb{R})$ be the set of continuous functions. Then, W is a subspace of V.

Proof. Let $c \in R$, and $f, g \in C(\mathbb{R})$. Then, the function $cf + g$ is continuous (from your favorite real analysis textbook/class). Thus, $C(\mathbb{R})$ is a vector subspace of V. □

There is a bit of a subtlety here: if W is a subspace of V, we know that both V and W have additive identities, denoted 0_V and 0_W. However, a priori, these elements need not be the same.

Theorem. Let W be a subspace of V. Then, $0_W = 0_V$.

Proof. Fix $w \in W$. Then, $w + 0_V = w$, since w and 0_V are both elements of V. Since $w \in W$ was arbitrary, 0_V is an identity element for W. However, since we have proven that the additive identity is unique, we conclude that $0_W = 0_V$. □

Theorem. Let $W_1, W_2 \subset V$ be subspaces of V. Then, $W_1 \cap W_2$ is a subspace of V.

Proof. Since $0 \in W_1$ and $0 \in W_2$, $0 \in W_1 \cap W_2$, so $W_1 \cap W_2$ is nonempty. Fix $c \in F$, and $v, w \in W_1 \cap W_2$. Then, v and w are in both W_1 and W_2. Thus, we have that $cv + w \in W_1$ since $v, w \in W_1$, and similarly $cv + w \in W_2$ since $v, w \in W_2$. Thus, $cv + w \in W_1 \cap W_2$. We conclude that $W_1 \cap W_2$ is a subspace of V. □

We can extend this proof to arbitrary intersections: that is, if $\{W_i\}_{i \in I}$ are a collection of subspaces of V, then the intersection $\bigcap_{i \in I} W_i$ is also a subspace of V. If we start with two subspaces $W_1, W_2 \subset V$, we can also try to put these vector spaces together.

Definition. Let $W_1, W_2 \subset V$ be two subspaces. Then, the **sum** of W_1 and W_2, denoted $W_1 + W_2$, is given by

$$W_1 + W_2 = \{w_1 + w_2 \,|\, w_i \in W_i\}$$

where the elements are sums of elements in W_1 and W_2.

Theorem. We prove some facts about sums of vector spaces. Let $W_1, W_2 \subset V$ be subspaces. Then,

- $W_1 + W_2$ is a subspace of V.
- $W_1, W_2 \subset W_1 + W_2$.
- $W_1 + W_2$ is the smallest (under containment) subspace of V which contains both W_1 and W_2.

Proof. We prove these results.

- Since V is closed under addition, $W_1 + W_2 \subset V$, and furthermore $0 = 0 + 0$, so $W_1 + W_2$ is nonempty. Let $c \in F$, and let $v, v' \in W_1 + W_2$. Write $v = w_1 + w_2$ and $v' = w_1' + w_2'$. Then, we have that

  $$cv + v' = c(w_1 + w_2) + (w_1' + w_2') = (cw_1 + w_1') + (cw_2 + w_2')$$

  and since W_1 and W_2 are subspaces, this is a sum of elements in W_1 and W_2 respectively, so $cv + v' \in W_1 + W_2$.
- For any $w \in W_1$, we can write $w = w + 0$, so $w \in W_1 + W_2$ (and analogously for W_2).
- Let W be the smallest subspace of V which contains both W_1 and W_2, defined to be

  $$W = \bigcap_{i \in I} V_i$$

  where I is some indexing set, and the set $\{V_i\}_{i \in I}$ contains all subspaces of V which contain W_1 and W_2. Then, for any $w = w_1 + w_2 \in W_1 + W_2$, we have that $w_1, w_2 \in V_i$, so $w \in V_i$, and since V_i was arbitrary, we have that $w \in \bigcap_{i \in I} V_i = W$. This shows that $W_1 + W_2 \subset W$. Since $W_1 + W_2$ is a subspace which contains W_1 and W_2, we have that for some $j \in I$, $W_1 + W_2 = V_j$, and so $W \subset V_j = W_1 + W_2$. This shows that $W = W_1 + W_2$, as desired.

□

We can extend this process to the sum of any *finite* number of vector spaces, but passing to infinitely many may result in some issues, so we ignore this case for now. Summing vector spaces is not always the nicest operation, because although elements are defined as sums of vectors of the summands, this decomposition is not always unique.

Example. Fix a field F, and let $V = F^3$. As before, write

$$V_i = \{\vec{x} \in V \,:\, x_i = 0\}$$

Then, let $W = V_1 + V_3$. Note then that

$$v = \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0\\1\\1\end{pmatrix} + \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\0\\1\end{pmatrix} + \begin{pmatrix}1\\1\\0\end{pmatrix}$$

and so we have written v in two ways as the sum of a vector from V_1 and a vector from V_3. This is due to the fact that we could move the 1 in the middle entry into either component, which we can mathematically indicate as

$$\begin{pmatrix}0\\1\\0\end{pmatrix} \in V_1 \cap V_3$$

to say that this "choice" can be absorbed into either V_1 or V_3.

This is a somewhat undesirable property of sums of vector spaces, as we'd prefer if our decompositions are unique.

Theorem. Let V be a vector space over a field F, and let $W_1, W_2 \subset V$ be subspaces. The following are equivalent.

1. If $0_V = w_1 + w_2$ with $w_1 \in W_1$ and $w_2 \in W_2$, then $w_1 = w_2 = 0$.
2. Every vector $w \in W_1 + W_2$ has a **unique** representation as $w = w_1 + w_2$, where $w_1 \in W_1$ and $w_2 \in W_2$.
3. $W_1 \cap W_2 = \{0\}$.

When this is the case, we say that the sum $W_1 + W_2$ is a **direct sum**, and denote it $W_1 \oplus W_2$.

Proof. We prove these equivalences.

- (1 $\Rightarrow$ 2) Fix $w \in W_1 + W_2$, and suppose we can write $w = v_1 + v_2 = w_1 + w_2$, where $v_i, w_i \in W_i$. Then, we have that
$$0 = w - w = (v_1 + v_2) - (w_1 + w_2) = (v_1 - w_1) + (v_2 - w_2)$$
Since $v_i - w_i \in W_i$, we have written 0 as a sum of a vector in W_1 together with a vector in W_2. Thus, by assumption, $v_i - w_i = 0$, so $v_i = w_i$, and this decomposition is unique, as desired.
- (2 $\Rightarrow$ 3) Fix $v \in W_1 \cap W_2$. Then, we can write
$$v = v + 0 = 0 + v$$
and so if $v \neq 0$, this provides two different decompositions of v into a sum of elements of W_1 and W_2. Thus, $v = 0$, as desired.
- (3 $\Rightarrow$ 1) Let $0_V = w_1 + w_2$ with $w_i \in W_i$. Since $0_V \in W_1$ and $0_V \in W_2$, and $0_V = w_1 + w_2$, we can write
$$0_V - w_2 = w_1$$
Then, both terms on the left are in W_2, and thus their difference must also. We therefore have that $w_1 \in W_2$, and thus $w_1 \in W_1 \cap W_2 = \{0\}$, so $w_1 = 0$. The argument for w_2 is analogous.

□

Example. Here are some examples of direct sums.

- F^n is the direct sum of the subspaces
$$F^n = \bigoplus_{i=1}^{n} \{\vec{x} \in F^n \,:\, x_j = 0 \;\; \forall \; j \neq i\}$$
where each component can be thought of as the i-th entry of the vector.

- $F[x]$ can be written as

$$F[x] = \{a_0 + a_2x^2 + \cdots + a_{2n}x^{2n} \,:\, a_i \in F\} \oplus \{a_1 + a_3x^3 + \cdots + a_{2n+1}x^{2n+1} \,:\, a_i \in F\}$$

  where the polynomial is split into odd and even degrees. Note that the odd degree component does indeed contain 0, since we can choose each $a_i = 0$.

- $C(\mathbb{R})$ can be written as a direct sum

$$C(\mathbb{R}) = \{f \in C(\mathbb{R}) \,:\, f(-x) = f(x)\} \oplus \{f \in C(\mathbb{R}) \,:\, f(-x) = -f(x)\}$$

  and the construction is left as an exercise.

Direct sums are the "nicest" type of sums that we can have, and they provide us a construction to "glue together" two subspaces into a larger subspace, provided that the subspaces are disjoint (except for 0). We can emulate this construction for two arbitrary vector spaces, not necessarily contained within some common vector space (which are "automatically" disjoint).

Definition. Let V, W be F-vector spaces. Then, we can define the **direct sum** of V and W to be the vector space

$$V \oplus W = \{(v, w) \,:\, v \in V, w \in W\}$$

under the addition and scalar multiplication rules

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2) \qquad c(v, w) = (cv, cv)$$

and one can check that this indeed forms a vector space. Note that instead of actually adding the vectors v and w, which we cannot do since they live in different vector spaces, we just keep track of the ordered pairs (and intuitively think of them as a sum).

One other way we can create new vector spaces from old ones is through taking **quotients**. This process is a little bit abstract, so I will defer much of the details to homework.

Definition. An **equivalence relation** on a set S is a set $T \subset S \times S$ which satisfies the following properties.

- Reflexivity: $(x, x) \in T$ for all $x \in S$.
- Symmetry: $(x, y) \in T$ if and only if $(y, x) \in T$, for all $x, y \in S$.
- Transitivity: If $(x, y) \in T$ and $(y, z) \in T$, then $(x, z) \in T$ for all $x, y, z \in S$.

If $(x, y) \in T$, we write $x \sim y$. We denote by $[x]$ the **equivalence class** of x, meaning

$$[x] = \{y \in S \,:\, x \sim y\}$$

The idea is that an equivalence relation tells us when two things are "similar", and so we treat them as the same object "up to equivalence" In the context of vector spaces, we'll see that if we choose our equivalence relation appropriately, the resulting set of classes still forms a vector space.

Definition. Let V be a vector space over F, and let $W \subset V$ be a subspace. We define an equivalence relation on V via the rule $x \sim y$ if $x - y \in W$. Then, the set of equivalence classes of vectors under this equivalence relation form an F-vector space, under the addition and scalar multiplication given by

$$[v] + [w] = [v + w] \qquad c[v] = [cv]$$

We denote the quotient space V/W. Note that we can naturally think of any element $v \in V$ as an element of V/W by considering $[v]$.

There are many details to check here to ensure that this does indeed form a vector space, which you will do on homework.

Example. Let $V = F^n$, and let W be the subspace where all entries are 0 except for the n-th coordinate. Then, two vectors $v_1, v_2 \in V$ are the same in V/W ($[v_1] = [v_2]$) if $v_1 - v_2 \in W$. Thus, v_1 and v_2 must agree in every coordinate except for the n-th. Therefore, the "distinct" vectors that can arise in V/W are in correspondence with the distinct entries in the first $n - 1$ coordinates (and thus "looks like" F^{n-1}).

4 Linear Independence and Span

Planned Lecture Date(s): June 28, 2023.

We'll use the structure that we've developed about vector spaces and subspaces to define some important notions.

Definition. Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ be a collection of vectors in V. A **linear combination** of the vectors of S is a sum of the form

$$\sum_{i=1}^{n} c_i v_i = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

where each $c_i \in F$. Note that by convention, the empty sum is $\vec{0}$.

Given a set of vectors $S = \{v_1, \ldots, v_n\}$, we can examine what vectors we can "create" from using only the vectors in S, together with our vector space operations.

Definition. Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ be a collection of vectors in V. The **span** of S, or the span of $\{v_1, \ldots, v_n\}$, is the set

$$\text{span}(S) = \text{span}(v_1, \ldots, v_n) = \left\{ \sum_{i=1}^{n} c_i v_i \ : \ c_i \in F \right\}$$

consisting of all linear combinations of vectors in S.

Theorem. For any set of vectors $S \subset V$, the span of S is a vector subspace of V.

Proof. Since the empty sum is a linear combination of elements of S, we have that $0 \in \text{span}(S)$, so $\text{span}(S)$ is not empty. We can then check that for $c \in F$ and $v, w \in \text{span}(S)$, the vector $cv + w$ is also a linear combination of elements of S. □

Example. Consider the space

$$W = \text{span}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right) \subset F^3$$

This consists of the space of all vectors with 0 in the third coordinate, and is a subspace of F^3.

Example. If S consists of only one vector, call it v, then

$$\text{span}(S) = \text{span}(v) = \{cv \ : \ c \in F\}$$

is simply the set of all vectors we can obtain by scaling v. Then, we have that

$$\text{span}(v_1, \ldots, v_n) = \text{span}(v_1) + \text{span}(v_2) + \cdots + \text{span}(v_n)$$

where the sum is taken as vector subspaces of V. The span of S is the smallest subspace of V which contains all the vectors in S.

Intuitively, we think of the vectors in S as our "allowed directions": we can move however we want, but only in the given directions. If S "has enough directions", then we can get to every vector in V, and this is a very special condition.

Definition. Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ be a collection of vectors in V. We say that S **spans** V if

$$\text{span}(S) = \text{span}(v_1, \ldots, v_n) = V$$

Equivalently, S spans V if every vector $v \in V$ can be written (not necessarily uniquely) as

$$v = \sum_{i=1}^{n} c_i v_i = c_1 v_1 + \cdots + c_n v_n$$

where each $c_i \in F$. When this is the case, we say that S is a **spanning set** for V.

Note that even if a set S does not span V, it still spans **some** vector subspace of V, and one such space that it spans is $\text{span}(S)$ (by definition).

Another notion that we can associate to a set of vectors is the following.

Definition. Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ be a collection of vectors in V. We say that S is **linearly dependent** if there exists $c_i \in F$, not all of which are 0 (such a linear combination is called **non-trivial**), such that

$$\sum_{i=1}^{n} c_i v_i = 0$$

If no such collection of c_i exist, we say that S is **linearly independent**. Such a linear combination is called a **linear dependence**.

The first thing to emphasize is that being linearly independent (or dependent) is a property of a **collection** of vectors, rather than any individual vector within the set. We can prove some properties of linearly independent sets.

Theorem. Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ be a collection of vectors in V.

- If S is linearly dependent, and $S' \supset S$, then S' is linearly dependent.
- If S is linearly independent, and $S' \subset S$, then S is linearly independent.
- If $0 \in S$, then S is linearly dependent.
- Every vector in $\text{span}(S)$ has a unique representation as a linear combination of vectors in S.

Proof. We prove the above claims.

- Suppose S is linearly dependent. Then, we can write

  $$\sum_{i=1}^{n} c_i v_i = 0$$

  with some $c_j \neq 0$. Then, this is also a linear combination of elements of S', with the coefficients on the elements of S' being all 0. Thus, we have found a linear dependence in S', so S' is linearly dependent.

- Suppose S is linearly independent, but $S' \subset S$ is not. Suppose S' has $m < n$ elements, and without loss of generality, reorder S so that $S' = \{v_1, \ldots, v_m\}$. Then, we can write

  $$\sum_{i=1}^{m} c_i v_i = 0$$

  with some $c_j \neq 0$. Then, this is also a linear combination of elements of S, with the coefficnets on the elements of S being all 0. Thus, we have found a linear dependence in S, which is a contradiction. We conclude that S' must be linearly independent.

- Without loss of generality, suppose $v_1 = 0$. Then, the linear combination with $c_1 = 1$ and $c_i = 0$ for all $i \neq 1$ is a non-trivial linear combination which sums to 0, so S is linearly dependent.

- Fix $v \in \text{span}(S)$, and suppose we can write v in two ways, as follows:

  $$v = \sum_{i=1}^{n} a_i v_i = \sum_{i=1}^{n} b_i v_i$$

  Then, we have that

  $$0 = v - v = \sum_{i=1}^{n} (a_i - b_i) v_i$$

  and thus $a_i - b_i = 0$ for all i.

□

Our intuition should be that a linear dependence occurs when our set of vectors becomes "redundant": we have too many vectors for each one to provide "unique" information. We can validate this intuition using the following.

Theorem. Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ be a collection of vectors in V. Suppose S is linearly dependent. Then, there exists some $1 \leq j \leq n$ such that v_j can be written as a linear combination of $S \backslash \{v_j\}$.

Proof. Since S is linearly independent, we can write

$$0 = \sum_{i=1}^{n} c_i v_i$$

for some $c_j \neq 0$. Then, subtracting, we have that

$$c_j v_j = -\sum_{\substack{i=1 \\ i \neq j}}^{n} c_i v_i$$

$$v_j = -\frac{1}{c_i} \sum_{\substack{i=1 \\ i \neq j}}^{n} c_i v_i$$

and so we have produced such a linear combination. Note that we crucially here divide by c_j, which is only possible because $c_j \neq 0$, so it has a multiplicative inverse in the field F. □

Note that this theorem does not guarantee that *every* vector can be written as a linear combination of the others, only that one such vector exists.

Example. In F^2, we have the vectors

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

and here any of the three vectors can be written as a linear combination of the other two. However, if we instead have the vectors

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

then the first cannot be written as a linear combination of the second and third (this statement is still true even if $\text{char}(F) = 2$).

Intuitively, we should think of span and linear independence as "opposite" notions: if S is linearly independent, we cannot have too many vectors, or they will become redundant, whereas if S is a spanning set, we cannot have too few vectors, or we do not have enough directions to go.

If V is a vector space over a field F, and $S = \{v_1, \ldots, v_n\}$ is a set of vectors in V, one can ask the question: if I choose a vector $v \in V$, how can I write it as a linear combination of vectors in S?

- If S is linearly independent, then we can write v in **at most one** way: if $v \in \text{span}(S)$, we can write it as a linear combination of elements of S, and by our result above, this is unique. However, if $v \notin \text{span}(S)$, we cannot write v as a linear combination of elements of S.
- If S is a spanning set for V, then we can write v in **at least one** way: since S spans V, every element of V is a linear combination of elements of S, but we may be able to write down several different linear combinations which sum to v.

However, if S is both linearly independent and a spanning set for V, this is a very special condition! We'll explore this in the next lecture.

5 Basis and Dimension

Planned Lecture Date(s): June 30, 2023.

From last lecture, we saw that if a set of vectors S is both linearly independent and a spanning set for V, then every vector in V can be written in exactly one way (uniquely) as a linear combination of vectors in S. This condition is very special, so we'll give it a name.

Definition. Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ be a collection of vectors in V. If S is both linearly independent and a spanning set for V, then S is said to be a **basis** for V.

We should think of our basis as our "coordinates" on our vector space: every vector has a coordinate representation, and furthermore this representation is unique.

Example. In F^n, we have a standard basis, denoted by $\{e_i\}_{1 \leq i \leq n}$, where e_i is the vector with a 1 in the i-th coordinate, and a 0 in every other coordinate. Any vector can then be uniquely written as

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + a_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

Example. Let $F[x]_{\leq n}$ denote the polynomials of degree at most n. Then, the set $\{1, x, x^2, \ldots, x^n\}$ forms a basis for $F[x]_{\leq n}$. However, we can also choose an alternative basis, perhaps $\{1, x-1, (x-1)^2, \ldots, (x-1)^n\}$. For example, we can write

$$1 + x^2 = 2 + 2(x-1) + (x-1)^2$$

and this allows us to express the same element of $F[x]_{\leq n}$ in two different bases. Note that the coefficients are different, but they sum to the same element.

Finding bases is incredibly useful, as we can use them to do computations (as we'll see later on in the course). However, there is a certain elegance to proofs which do not require us to choose a basis, and mathematicians generally prefer these proofs as they require less "arbitrary choice". We'll often say that a proof or argument is **canonical** if choosing a basis is not necessary.

We'll now work towards proving some very important properties of bases.

Lemma. (Steinitz Exchange Lemma) Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ and $T = \{w_1, \ldots, w_m\}$, with S linearly independent and T spanning V. Then, the following hold:

- $n \leq m$.
- After possibly reordering T, the set $\{v_1, \ldots, v_n, w_{n+1}, \ldots, w_m\}$ spans V.

Proof. We induct on n. When $n = 0$, we have that both $0 \leq m$, and the set T spans V by assumption.

Suppose the claim holds for $n-1$. By the inductive hypothesis, after possible reordering of T, the set $\{v_1, \ldots, v_{n-1}, w_n, \ldots, w_m\}$ spans V. If $n > m$, the given set is $\{v_1, \ldots, v_m\}$, and $v_n \in V$ is in the span of this set, violating the linear independence of S. This shows that $n \leq m$. We can then write

$$v_n = \sum_{i=1}^{n-1} c_i v_i + \sum_{i=n}^{m} c_i w_i$$

We cannot have $c_i = 0$ for all $i \geq n$, or this would write v_n as a linear combination of $\{v_1, \ldots, v_{n-1}\}$, but S is assumed to be linearly independent. Thus, some $c_j \neq 0$ for $n \leq j \leq m$, so without loss of generality, let $j = n$ (up to reordering). Then, we have that

$$w_n = \frac{1}{c_n} \left(v_n - \sum_{i=1}^{n-1} c_i v_i - \sum_{i=n+1}^{m} c_i w_i \right)$$

and so $w_n \in \text{span}(v_1, \ldots, v_{n-1}, v_n, w_{n+1}, \ldots, w_m)$, and so we have that

$$V = \text{span}(v_1, \ldots, v_{n-1}, w_n, w_{n+1}, \ldots, w_m) \subset \text{span}(v_1, \ldots, v_{n-1}, v_n, w_{n+1}, \ldots, w_m) \subset V$$

and thus we have that $\text{span}(v_1, \ldots, v_{n-1}, v_n, w_{n+1}, \ldots, w_m) = V$ as desired. □

This gives us a very important result.

Theorem. Let V be a vector space over F, and let $S = \{v_1, \ldots, v_n\}$ and $T = \{w_1, \ldots, w_m\}$. Suppose S and T are both bases for V. Then, $n = m$.

Proof. Apply the Steinitz Exchange Lemma to S and T to obtain $n \leq m$, and again to T and S to obtain $m \leq n$. □

This shows that if our bases are of finite length, they must be the same length! We therefore make the following definition.

Definition. Let V be a vector space with basis $S = \{v_1, \ldots, v_n\}$. The **dimension** of V, denoted $\dim(V)$, is equal to n, the length of one (and by the above theorem, any) basis. If V does not possess a finite basis, we say that V is infinite-dimensional.

We now have a way to quantify how large our vector spaces are.

Example. Let's think about some vector spaces we're familiar with.

- The vector space F^n is n-dimensional, with standard basis $\{e_1, \ldots, e_n\}$.
- The vector space $F[x]_{\leq n}$ is $(n+1)$-dimensional, with a basis given by $\{1, x, \ldots, x^n\}$.
- The $\mathbb{R}$-vector space $\mathbb{C}$ is two dimensional, given by basis $\{1, i\}$.
- The vector space $F[x]$ has a basis, given by $\{1, x, x^2, \ldots\}$. However, this basis is infinite - this does not prove that $F[x]$ does not possess a finite basis (and by consequence be finite dimensional), but $F[x]$ is indeed infinite-dimensional.

Another consequence of the Steinitz Exchange Lemma is the following.

Theorem. Let V be a finite-dimensional vector space of dimension n. Then, we have that

- Any subset $P \subset V$ with more than n vectors is linearly dependent.
- Any subset $Q \subset V$ with fewer than n vectors cannot span V.

Proof. We prove both statements using Steinitz Exchange Lemma. Fix a basis B for V.

- If P is linearly independent, choose $S = P$ and $T = B$. This is a contradiction, since we then must have $|P| \leq |B| = n$.
- If Q is a spanning set, choose $S = B$ and $T = Q$. This is a contradiction, since we then must have that $|Q| \geq |B| = n$.

□

Another important property of bases is the following.

Theorem. Let V be a finite dimensional vector space, and let $S = \{v_1, \ldots, v_n\}$ be a linearly independent set. Then, we can find a basis T of V, with $S \subset T$ (we say that we can "complete" S to a basis of V).

Proof. If S spans V, then S is a basis for V, and we are done. Otherwise, pick $w \in V \setminus \text{span}(S)$. We claim that $S \cup \{w\}$ is linearly independent. Suppose that $S \cup \{w\}$ is linearly dependent. Then, we have that

$$0 = c_0 w + \sum_{i=1}^{n} c_i v_i$$

with not all $c_i = 0$. Notice that if $c_0 = 0$, we have produced a non-trivial linear combination of the v_i which give 0, which cannot occur as the v_i are linearly independent. Thus, we have that $c_0 \neq 0$, so we can write

$$w = -\frac{1}{c_0}\left(\sum_{i=1}^{n} c_i v_i\right)$$

and so $w \in \text{span}(S)$, which is a contradiction. Thus, the set $S \cup \{w\}$ must be linearly independent. We can repeat this process finitely many times unless $|S| > \dim(V)$, at which point our previous theorem tells us that S is linearly dependent, which is a contradiction. Thus, we must be in the case where S spans V, and thus S is a basis. □

This theorem allows us to prove several results.

Theorem. Let V be a finite dimensional vector space, and let $W \subset V$ be a subspace.

- $\dim(W) \leq \dim(V)$
- Every basis for W can be completed to a basis for V.

Proof. Fix a basis B for W. Then, B is linearly independent (both as a subset of W and as a subset of V), so we can extend it to a basis for V. Extending our basis adds a non-negative number of elements to our basis, so $\dim(W) \leq \dim(V)$. □

Example. Let V be a vector space over F, and let $W_1, W_2 \subset V$ be subspaces.

- If the sum is direct, $\dim(W_1 \oplus W_2) = \dim(W_1) + \dim(W_2)$.
- $\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$.
- $\dim(V/W_i) = \dim(V) - \dim(W_i)$.

The second and third require some work to prove, but for the first, we can simply concatenate bases for W_1 and W_2.

Finally, a few words about infinite dimensional vector spaces: many of the results that we'll prove are easiest in the finite dimensional case, so we'll generally stick to these situations. However, we do have some examples of infinite-dimensional vector spaces, such as $F[x]$, $C(\mathbb{R})$, or $\mathbb{R}$ as a $\mathbb{Q}$-vector space. Showing that every finite dimensional vector space has a basis is not hard, as we can simply apply the above theorem to an empty set of vectors. However, showing that every (not necessarily finite dimensional) vector space has a basis is much more difficult (and in fact equivalent to the axiom of choice!).

6 Linear Transformations

Planned Lecture Date(s): July 3, 2023.

One of the perspectives that we take in mathematics is that once we understand a certain type of object, we should understand functions between these objects. In our case, our objects are vector spaces, and our functions are linear transformations.

Definition. Let V, W be F-vector spaces. A **linear transformation** T is a function $T : V \to W$ satisfying the following properties:

- Superposition: $T(v_1 + v_2) = T(v_1) + T(v_w)$ for all $v_1, v_2 \in V$.
- Proportionality: $T(cv) = cT(v)$ for all $c \in F$ and $v \in V$.

We will sometimes refer to linear transformations as **linear maps**, and a function which satisfies these properties is said to be **linear**. The space V is referred to as the **domain**, and the space W is referred to as the **codomain**.

The idea is that a vector space is a set together with certain additional structure, namely addition and scalar multiplication, so our functions between vector spaces should "preserve" this structure. One way you can think of this property is that for either addition or scalar multiplication, we can perform the operation either before or after applying our linear transformation, and we will get the same result.

One way to check if a transformation is linear is the following (hopefully familiar) result.

Theorem. Let V, W be F-vector spaces, and let $T : V \to W$ be any function. Then, T is a linear transformation if and only if $T(cv + w) = cT(v) + T(w)$ for all $c \in F$ and $v, w \in V$.

Proof. Choosing $c = 1$ gives superposition, and choosing $w = 0$ gives proportionality. □

Example. We explore some examples of linear transformations.

- Let V be an F-vector space. Then, the identity map $\mathrm{Id} : V \to V$ which takes any $v \in V$ to itself is a linear map.
- Let V, W be F-vector spaces. Then, the zero map $0 : V \to W$ which takes any $v \in V$ to $0 \in W$ is a linear map.
- Let $V = F^n$ and $W = F^m$. Then, any $m \times n$ matrix is a linear map.

Let's prove a property of linear maps.

Theorem. Let V, W be F-vector spaces, and let $T : V \to W$ be a linear map. Then, $T(0) = 0$.

Proof. We have that

$$T(0) = T(0 + 0) = T(0) + T(0)$$

and subtracting $T(0)$, we have that $T(0) = 0$ as desired. □

It turns out that if our vector spaces are finite-dimensional, then matrices are, in some sense, the *only* linear maps that we can have, and we will build towards this result.

Suppose V and W are finite-dimensional F-vector spaces. If we pick a basis $B = \{v_1, \ldots, v_n\}$ for V, then we can make the following observation: for any $v \in V$, we can write

$$v = \sum_{i=1}^{n} c_i v_i$$

and furthermore the choice of c_i are unique. Then, if $T : V \to W$ is a linear map, we have that

$$\begin{aligned} T(v) &= T\left(\sum_{i=1}^{n} c_i v_i\right) \\ &= \sum_{i=1}^{n} T\left(c_i v_i\right) \\ &= \sum_{i=1}^{n} c_i T\left(v_i\right) \end{aligned}$$

This shows the following important result: the value of T on **any** vector $v \in V$ is fully determined by $T(v_i)$ for each v_i in B! In other words, in order to define a linear map $T : V \to W$, it suffices to decide where we would like to map each v_i, and so we only need to make n choices. Once we have done so, the value of T on any vector in V is determined. This is the power that comes from the structure of linear maps between vector spaces!

Example. Let $D : \mathbb{R}[x]_{\leq n} \to \mathbb{R}[x]_{\leq n}$ be the map that sends $f(x)$ to $f'(x)$ (i.e. the derivative map). One can check that this map is indeed linear, and furthermore $\mathbb{R}[x]_{\leq n}$ has the basis $\{1, x, x^2, \ldots, x^n\}$. We see that

$$D(x^n) = nx^{n-1}$$

and this definition is enough to compute $D(f(x))$ for any $f(x) \in \mathbb{R}[x]_{\leq n}$.

Now, let's suppose that we've fixed a basis $B_V = \{v_1, \ldots, v_n\}$ on V, and a basis $B_W = \{w_1, \ldots . w_m\}$ on W. Then, for each v_i, we can write

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

and we can implicitly think of this as represented by the vector

$$\begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \in F^m$$

If we package this into a matrix, it looks like the following.

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

and we see that this matrix acting on the vector e_i (which has a 1 in the i-th component, and 0 elsewhere) gives precisely the vector with c_{ij} in the j-th component, and so this matrix does indeed send each v_i to the prescribed position of $T(v_i)$, as desired. By our above work, we see that this completely determines the linear transformation. For this reason, we think of choosing bases as picking **coordinates**, and we will sometimes refer to this process in this way.

We therefore make the following observation: a linear transformation T is a linear function between two F-vector spaces V and W, and if we choose bases for both V and W, then in these chosen coordinates, T takes the form of a matrix, and acts on vectors under standard matrix multiplication. Furthermore, the columns of the matrix tell us exactly where each basis vector is sent under T. We'll go back and forth between the matrix perspective and the more abstract perspective, to see things both ways.

When we choose coordinates, we are implicitly choosing an isomorphism of V to F^n and W to F^m, and we will make this rigorous later on once we've defined isomorphisms.

We will often denote the set of linear maps from V to W (as vector spaces over F) using the notation $\mathrm{Hom}_F(V, W)$, often suppressing the F when the context is clear. If $V = W$, we will use $\mathrm{End}(V)$ to denote

$\mathrm{Hom}(V, V)$. The symbol "Hom" is short for "homomorphism", which is a general mathematical term used to describe structure-preserving functions.

We conclude this section by defining some properties that linear maps can have.

Definition. Let V, W be F-vector spaces, and let $T : V \to W$ be a linear transformation. We say that T is...

- ...**injective** if for all $v_1, v_2 \in V$, $T(v_1) = T(v_2)$ implies $v_1 = v_2$.
- ...**surjective** if for all $w \in W$, there exists $v \in V$ with $T(v) = w$.
- ...**bijective** if T is both injective and surjective.

Intuitively, we want to think of an injective map which doesn't take two different things and send them to the same thing, and a surjective map as one which hits everything in the codomain. In your first linear algebra class, you've probably seen "injective" as **one-to-one**, and "surjective" as **onto**. These are also acceptable names, but the nomenclature we use here generalizes to other algebraic objects.

7 Kernel, Image, and the First Isomorphism Theorem

Planned Lecture Date(s): July 5, 2023 - July 7th, 2023.

Given a linear map $T : V \to W$, we can ask several different questions about it: is T injective or surjective? How can we tell? Let's define some terms.

Definition. Let V, W be F-vector spaces, and let $T : V \to W$ be a linear map. The **kernel** of T, denoted $\ker(T)$, is given by

$$\ker(T) = \{v \in V \mid T(v) = 0\} \subset V$$

Furthermore, the **image** of T, denoted $\mathrm{im}(T)$, is given by

$$\mathrm{im}(T) = \{w \in W \mid \exists\ v \in V \text{ such that } T(v) = w\} \subset W$$

We refer to $\dim(\mathrm{im}(T))$ as the **rank** of T, and $\dim(\ker(T))$ as the **nullity** of T.

The first thing to notice is that if we choose coordinates and represent T as a matrix M, then $\ker(T)$ corresponds precisely to the **null space** of M, and $\mathrm{im}(T)$ corresponds to the **range**, or **column space** of M, and furthermore the rank and nullity correspond to the dimensions of these spaces. However, we can refer to these spaces entirely independently of the matrix representation, as we have done above.

Theorem. Let V, W be F-vector spaces, and let $T : V \to W$ be a linear map. Then, we have that

- $\ker(T)$ is a subspace of V.
- $\mathrm{im}(T)$ is a subspace of W.

Proof. We prove these statements.

- Note that $T(0) = 0$, so $0 \in \ker(T)$, and so $\ker(T)$ is nonempty. Suppose $c \in F$, and $v, w \in \ker(T) \subset V$. Then, we have that
$$T(cv + w) = cT(v) + T(w) = c0 + 0 = 0$$
and so $cv + w \in \ker(T)$, so $\ker(T)$ is a subspace of V.
- Note that $T(0) = 0$, so $0 \in \mathrm{im}(T)$, and so $\mathrm{im}(T)$ is nonempty. Suppose $c \in F$, and $v, w \in \mathrm{im}(T)$. Then, there exists $v', w' \in V$ with $T(v') = v$ and $T(w') = w$. We therefore have that
$$T(cv' + w') = cT(v') + T(w') = cv + w$$
and so $cv + w \in \mathrm{im}(T)$. Thus, $\mathrm{im}(T)$ is a subspace of V.

□

This shows that these objects we've defined are indeed subspaces, and are natural objects to consider.

Theorem. Let V, W be F-vector spaces, and let $T : V \to W$ be a linear map. Then, we have the following.

1. T is injective if and only if $\ker(T) = \{0\}$.
2. T is surjective if and only if $\mathrm{im}(T) = W$.
3. T is bijective if and only if both $\ker(T) = \{0\}$ and $\mathrm{im}(T) = W$.

Proof. We prove each of these statements.

1. Suppose T is injective. Then, we have proved that $T(0) = 0$, so if $v \in \ker(T)$, then $T(v) = 0 = T(0)$, so $v = 0$.

   Suppose $\ker(T) = 0$, and suppose $T(v_1) = T(v_2)$ for some $v_1, v_2 \in V$. Then, we have
   $$0 = T(v_1) - T(v_2) = T(v_1 - v_2)$$
   and so $v_1 - v_2 \in \ker(T) = \{0\}$. Thus, $v_1 = v_2$, as desired.

2. By definition, T being surjective is precisely the statement that $\text{im}(T) = W$.

3. Combining the above two statements gives the desired result.

□

We therefore conclude that checking if a linear map is injective, surjective, or bijective is equivalent to checking the appropriate condition on the kernel or image. Thus, we can choose coordinates to express our map in its matrix representation, and from our previous linear algebra class, we have methods of determining the null space and column space of such a matrix.

It is a very special condition when a linear map is bijective, as we can attempt to construct an inverse. Suppose $T : V \to W$ is a bijective linear map, and we would like to invert T with some function $S : W \to V$. This means that for any $v \in V$, $S(T(v)) = v$, and similarly for any $w \in W$, $T(S(w)) = w$. How would we go about doing this? Let's try to define S first.

Fix $w \in W$. We'd like to decide what $S(w)$ is: since T is surjective, there exists some $v \in V$ with $T(v) = w$, and furthermore since T is injective, this v is unique. Thus, we assign $S(w)$ to be this v. It remains to check that the map S does indeed provide an inverse for T, and furthermore, the inverse function S is linear! A priori, we have no reason to expect this to be true, but this is indeed the case, and you will prove this on homework. Since we can construct an inverse map when T is bijective, we will often times refer to bijective linear maps as **invertible**.

If we can find a bijective linear map T between two F-vector spaces V and W, then any calculations we would like to carry out in V, we could instead map over to W via T, do our calculations in W, and map back to V using T^{-1}. In this sense, V and W are "the same" vector space, since anything we'd like to do in V we can do in W, and vice versa. This leads us to a definition.

Definition. Let V and W be two F-vector spaces. Then, V and W are said to be **isomorphic** if there exists a bijective linear map $T : V \to W$. We say that T (alternatively, T^{-1}) is an **isomorphism** between V and W.

In essence, two vector spaces are isomorphic if the tools of linear algebra cannot tell the difference between the two: anything we do in either V or W can be moved to the other via either T or T^{-1}. We prove an important property of bijective maps.

Theorem. Let V, W be finite-dimensional F-vector spaces, and let $T : V \to W$ be a linear map. Then, T is invertible if and only if T sends bases of V to bases of W.

Proof. Suppose T is invertible. Fix a basis $\{v_1, \ldots, v_n\}$ for V, and examine $\{T(v_1), \ldots, T(v_n)\}$. Suppose we have that

$$0 = c_1T(v_1) + \cdots + c_nT(v_n) = T(c_1v_1 + \cdots + c_nv_n)$$

Since T is injective, each $c_i = 0$, showing linear independence. Similarly, fix $w \in W$. Since T is surjective, we can find $v \in V$ with $T(v) = w$, so

$$v = c_1v_1 + \cdots + c_nv_n$$

and mapping everything forward by T and applying linearity shows that w is in fact in the span, so this set is a spanning set for W, and therefore a basis.

Suppose T sends bases to bases. Fix a basis $\{v_1, \ldots, v_n\}$ for V, and by assumption we have that $\{T(v_1), \ldots, T(v_n)\}$ is a basis for W. Then, suppose some $v = c_1v_1 + \cdots + c_nv_n$ is in $\ker(T)$. We have

$$0 = T(v) = T(c_1v_1 + \cdots + c_nv_n) = c_1T(v_1) + \cdots + c_nT(v_n)$$

and since the $T(v_i)$ form a basis, each $c_i = 0$, so $v = 0$, and $\ker(T) = 0$, so T is injective. Similarly, fix $w \in W$. Then, we have that

$$w = c_1T(v_1) + \cdots + c_nT(v_n) = T(c_1v_1 + \cdots + c_nv_n)$$

and so $w \in \text{im}(T)$. Thus, we have that $\text{im}(T) = W$ and therefore T is surjective. We conclude that T is invertible. □

We can additionally conclude that in order to construct an isomorphism between two vector spaces, we can simply choose our linear map to map bases to bases, and this is equivalent to producing a linear map.

Corollary. Let V, W be finite-dimensional F-vector spaces. If V and W are isomorphic, then $\dim(V) = \dim(W)$.

Proof. Suppose $\dim(V) = n$. Choose an isomorphism $T : V \to W$, and fix a basis $\{v_1, \dots, v_n\}$ for V. Then, by the above theorem, the set $\{T(v_1), \dots, T(v_n)\}$ is a basis for W, so $\dim(W) = n$. □

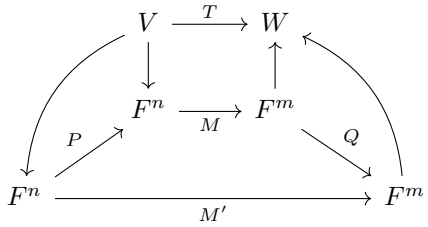
The above statements also hold when V and W are not finite-dimensional, but the proof is more involved, so I'll omit them here. This in fact shows that any finite-dimensional vector space is isomorphic to F^n, and on homework you will show that isomorphisms are transitive, so we can conclude that every finite-dimensional F-vector space of the same dimension is isomorphic!

Furthermore, choosing bases for V and W (with an ordering) implicitly gives a choice of isomorphism to F^n, so for $T : V \to W$ with $\dim(V) = n$ and $\dim(W) = m$, we can draw the diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \downarrow & & \uparrow \\ F^n & \xrightarrow[M]{} & F^m \end{array}$$

where the arrows on the sides are isomorphisms, and go both ways. Thus, following a linear map from V to W along T is the same as following the isomorphism on the left, following the matrix M representing T from F^n to F^m, and then following the isomorphism on the right.

If we instead choose a different basis, then we have the following diagram.



Note that P and Q represent isomorphisms of F^n and F^m with each other, and we can write the matrix M' as $M' = Q^{-1}MP$. Thus, the matrices M and M' differ by a similarity transformation, where the matrices P and Q represent change of basis matrices for both V and W in F^n and F^m, respectively. Furthermore, note that if $V = W$ and the same basis is chosen for both copies of V, then our diagram says that $M' = P^{-1}MP$, which is precisely the change of basis for square matrices from a first course in linear algebra.

Now that we have the language of isomorphisms, we can prove one of the major results in linear algebra.

Theorem. (First Isomorphism Theorem) Let V, W be F-vector spaces, and let $T : V \to W$ be a linear map. Then, there is a canonical isomorphism

$$V/\ker(T) \cong \operatorname{im}(T)$$

as F-vector spaces.

Proof. We define a map $\widehat{T} : V/\ker(T) \to \operatorname{im}(T)$ by sending $[v] \mapsto T(v)$. Note first that by definition, $T(v) \in \operatorname{im}(T)$ for all $v \in V$. We first show that this map is well-defined and independent of choice of representative in $[v]$. Suppose $[v] = [w]$. Then, $v \sim w$, so $v - w \in \ker(T)$. Then, we have that

$$T(v) - T(w) = T(v - w) = 0$$

and so $T(v) = T(w)$, so $\widehat{T}$ is indeed well-defined.

We show that $\widehat{T}$ is injective. Fix $[v] \in V/\ker(T)$, and suppose $\widehat{T}([v]) = 0$. Then, we have that

$$0 = \widehat{T}([v]) = T(v)$$

and so $v \in \ker(T)$, so $[v] = 0$. We conclude that $\ker(\widehat{T}) = 0$, and so $\widehat{T}$ is injective.

We show that $\widehat{T}$ is surjective. Fix $w \in \text{im}(T)$. Then, since $w \in \text{im}(T)$, we can find $v \in V$ such that $T(v) = w$. Then, we have that

$$\widehat{T}([v]) = T(v) = w$$

and so $w \in \text{im}(\widehat{T})$. Since W was arbitrary, $\widehat{T}$ is surjective.

We conclude that $\widehat{T}$ is indeed an isomorphism, as desired. □

From this, we have an immediate corollary, which might be familiar from a first course in linear algebra.

Corollary. (Rank-Nullity Theorem) Let V, W be F-vector spaces, and let $T : V \to W$ be a linear map. Then, we have that

$$\dim(\ker(T)) + \dim(\text{im}(T)) = \dim(V)$$

Proof. We have that $V/\ker(T) \cong \text{im}(T)$, so applying the homework result that $\dim(V/W) = \dim(V) - \dim(W)$ and that isomorphic vector spaces have the same dimension, we have that

$$\dim(V) - \dim(\ker(T)) = \dim(\text{im}(T))$$

and the result follows. □

The First Isomorphism Theorem motivates our intuition of quotient spaces: for any $W \subset V$, we can choose a linear transformation T which "collapses" W to 0, therefore making $\ker(T) = W$, and the image of T, which maps everything outside of W, is isomorphic to the quotient space V/W.

8 The Vector Space of Linear Transformations

Planned Lecture Date(s): July 10th, 2023.

In this section, we take a step back from linear transformations, and instead view the set of linear transformations between two vector spaces as its own object. Recall that for F-vector spaces V and W, we have

$$\mathrm{Hom}_F(V,W) = \{T \mid T : V \to W \text{ is linear}\}$$

However, since these are just functions, we can perform some operations on them. For $S, T \in \mathrm{Hom}_F(V,W)$, we can define $S+T \in \mathrm{Hom}_F(V,W)$ to be

$$(S+T)(v) = S(v) + T(v)$$

and similarly for $c \in F$ and $T \in \mathrm{Hom}_F(V,W)$, we can define $cT \in \mathrm{Hom}_F(V,W)$ to be

$$(cT)(v) = cT(v)$$

This defines addition and scalar multiplication on the set $\mathrm{Hom}_F(V,W)$, and furthermore the remaining axioms can be checked to verify that the set of linear transformations from V to W forms a vector space! Note that the additive identity in this vector space is the zero map, which sends every vector in V to 0 in W. What can we say about this vector space?

Theorem. Let V, W be finite-dimensional F-vector spaces. Then, $\dim(\mathrm{Hom}_F(V,W)) = \dim(V)\dim(W)$.

Proof. Fix bases for V and W, and let $n = \dim(V)$ and $m = \dim(W)$. Then, any $T \in \mathrm{Hom}_F(V,W)$ has a unique matrix representation, as an $m \times n$ matrix from F^n to F^m. Since a matrix of this form has mn (independent) entries, the vector space of matrices is isomorphic to F^{mn}. Thus, $\dim(\mathrm{Hom}_F(V,W)) = mn$, as desired □

This gives us another way to construct new vector spaces! Given any two vector spaces over F, we can construct a larger vector space. When $V = W = F^n$, we have that $\mathrm{Hom}_F(V,W)$ is precisely the $n \times n$ square matrices with entries in F. One observation you might make is that we can do more than add and scalar multiply square matrices, but we can actually multiply them as well - it turns out that this makes $\mathrm{Hom}_F(V,V)$ more than just a vector space over F, but in fact an **algebra** over F. We won't go into too much detail about algebras, but one can think of them as a vector space together with additional multiplicative structure.

Example. We examine some examples of $\mathrm{Hom}_F(V,W)$. Note that in the finite dimensional case, any claims of isomorphism can be accomplished simply by counting dimension, but I attempt to give a basis-independent isomorphism.

- $\mathrm{Hom}_F(F^n, F^m)$ is the set of $m \times n$ matrices.
- $\mathrm{Hom}_F(0,W) \cong \{0\}$, since the only linear transformation out of the zero vector space is the map sending 0 to 0.
- $\mathrm{Hom}_F(F,W) \cong W$, since for $T \in \mathrm{Hom}_F(F,W)$, if $T(1) = w \in W$, then $T(c) = cT(1) = cw$.
- $\mathrm{Hom}_F(V,F) \cong V$, but this isomorphism is a bit more involved, so we will investigate it further.

The last example above, the set $\mathrm{Hom}_F(V,F)$, has a special name.

Definition. Let V be a vector space over F. Then, we define the **dual space** of V, denoted V^*, to be $\mathrm{Hom}_F(V,F)$.

The dual space contains **linear functionals**: maps from V into the base field F which are linear. If V is finite dimensional and $\dim(V) = n$, then we can perform the following construction.

Definition. Let V be a finite-dimensional vector space over F, with $\dim(V) = n$. Let $B = \{v_1, \ldots, v_n\}$ be a basis for V. Then, we define the **dual basis** for V to be $B^* = \{\varepsilon_1, \ldots, \varepsilon_n\}$, where

$$\varepsilon_i(v_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Recall that defining a linear map on a basis determines the entire linear map, so I have defined each ε_i fully. Since this is called the dual basis, it had better be a basis for the dual space.

Theorem. Let V be a finite-dimensional vector space over F, with $\dim(V) = n$. Let $B = \{v_1, \ldots, v_n\}$ be a basis for V. Then, the dual basis $B^* = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is a basis for V^*, the dual space of V.

Proof. We show that B^* is linearly independent. Suppose that

$$Z = c_1\varepsilon_1 + \cdots + c_n\varepsilon_n$$

where Z represents the zero function in $V^* = \mathrm{Hom}_F(V, F)$. Then, we have that $Z(v) = 0$ for all $v \in V$, so we compute

$$0 = Z(v_i) = (c_1\varepsilon_1 + \cdots + c_n\varepsilon_n)(v_i) = a_i$$

and thus each $c_i = 0$.

We show that B^* is a spanning set. Fix $T \in \mathrm{Hom}_F(V, F)$. Then, set

$$c_i = T(v_i)$$

We claim that

$$T = c_1\varepsilon_1 + \cdots + c_n\varepsilon_n$$

Let S denote the right hand side. Then, $S(v_i) = T(v_i)$ for each i by construction, so we have that $(S - T)(v_i) = 0$ for all i. Since the v_i form a basis for V, $(S - T)$ must be the zero function, and so $S = T$ as desired.

We conclude that B^* is a basis for V^*. □

Not only does this theorem show us that $\dim(V^*) = \dim(V)$, which we already knew, but it also allows us to construct an explicit isomorphism, only by choosing a basis for V (and not additionally V^*), by sending v_i to the corresponding ε_i. It turns out that in general, there is no way to construct a canonical isomorphism between V and V^*, but this discussion is beyond the scope of this course.

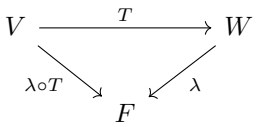
When V is infinite dimensional, it is in fact true that the cardinality of $\dim(V^*) > \dim(V)$, and this inequality is strict! However, this proof is beyond the scope of this course.

In mathematics, a general viewpoint is to consider mathematical objects and their associated functions together, and this is no exception. We can take the dual of a vector space, but we'd like to define the notion of "dualization" to not just vector spaces, but the linear transformations between them.

Definition. Let V, W be F-vector spaces, and let $T : V \to W$ be a linear map. Then, we define the **dual map** associated to T, denoted T^*, via

$$\begin{aligned} T^* : W^* &\to V^* \\ \lambda &\mapsto \lambda \circ T \end{aligned}$$

This definition is confusing, and quite unintuitive, so we illustrate it with the following diagram.



In order to construct a map from W^* to V^*, we must take in an element of $W^* = \mathrm{Hom}_F(W, F)$, and return an element of $V^* = \mathrm{Hom}_F(V, F)$. Given such a $\lambda \in W^*$, the natural way to construct a map from V to F is to first map from V to W along T, and then to map from W to F along λ. Note that this assignment does not require us to choose bases on either V or W, and additionally this reverses the direction of the original map T.

Suppose V is a finite-dimensional F-vector space, and we pick a basis (and therefore isomorphism to F^n). Then, an element $f \in V^*$ would map F^n to F, and is therefore represented by a $1 \times n$ matrix (or a row vector). Then, let's consider the dual basis to the standard basis $\{e_1, \ldots, e_n\}$ for F^n. We have that

$$\varepsilon_j(e_i) = \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = c_i$$

where the 1 occurs in the i-th position. Thus, for a fixed j, we must have that $c_j = 1$ and $c_i = 0$ for all $i \neq j$, and so in fact, in our matrix representation, the vector ε_i is simply the transpose of the vector e_i! This in fact holds in general, as when we've fixed bases, $\dim(V) \cong \dim(V^*)$, and furthermore the image of any vector under the isomorphism is precisely the transpose. This in fact extends even further, as you will show on homework that for $T : V \to W$, if a basis for V and W are chosen, then the matrix associated to $T^* : W^* \to V^*$ (in the dual basis in both V and W) is precisely the transpose of the matrix associated to T.

We end this section with an observation: the dualization construction we've defined allows us to take the dual of any vector space over F, so why not take the dual of V^* itself? We can define the **double dual** of V, and from our previous work, we know that $\dim(V^{**}) = \dim(V^*) = \dim(V)$, and so $V \cong V^{**}$. Elements of V^{**} take elements of V^*, and send them to F. One such example is the evaluation map: for $v \in V$, we can define

$$\begin{aligned} \mathrm{ev}_v : V^* &\to F \\ \lambda &\mapsto \lambda(v) \end{aligned}$$

and we can check that this is indeed an element of V^{**}. It turns out that the map

$$\begin{aligned} T : V &\to V^{**} \\ v &\mapsto \mathrm{ev}_v \end{aligned}$$

is in fact an isomorphism, and furthermore does not require a choice of basis for V! This shows that not only is V^{**} isomorphic to V, but it is *canonically* isomorphic to V, as we can produce a basis-independent isomorphism.

Example. We work through an example of showing the matrix representation of the dual map is the transpose of the original map. Let $V = \mathbb{R}^2$, let $W = \mathbb{R}^3$, and let

$$T = \begin{pmatrix} 2 & -1 \\ 4 & 0 \\ -3 & 2 \end{pmatrix}$$

Choose the standard bases $\{e_1, e_2\}$ on V and $\{e_1', e_2', e_3'\}$ on W, and let $\{\varepsilon_1, \varepsilon_2\}$ and $\{\varepsilon_1', \varepsilon_2', \varepsilon_3'\}$ be the corresponding dual bases, respectively. We'd like to determine the matrix representation of T^* in the dual bases. We have that

$$T^*(\varepsilon_1') = \varepsilon_1' \circ T = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 4 & 0 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 \end{pmatrix} = 2\varepsilon_1 - \varepsilon_2$$

$$T^*(\varepsilon_2') = \varepsilon_2' \circ T = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 4 & 0 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 0 \end{pmatrix} = 4\varepsilon_1$$

$$T^*(\varepsilon_3') = \varepsilon_3' \circ T = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 4 & 0 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} -3 & 2 \end{pmatrix} = -3\varepsilon_1 + 2\varepsilon_2$$

Since we have determined where T sends each basis vector ε_i' in terms of the ε_j, we can write down the matrix for T^*, since the columns correspond to the images of the ε_i', in the ε_j basis. We therefore have

$$T^* = \begin{pmatrix} 2 & 4 & -3 \\ -1 & 0 & 2 \end{pmatrix}$$

which is the transpose, as desired.

9 Multilinear Maps

Planned Lecture Date(s): July 12th, 2023.

Linear maps are a central object in the study of linear algebra, and a natural question to ask may be how we can extend linear maps to a broader class of objects. For example, if we are given a collection of vector spaces $V_1, V_2, \ldots, V_n$, and a target vector space W, we can define functions

$$T : V_1 \times \cdots \times V_n \to W$$

Note here that the domain is the *Cartesian* product of the V_i, and so we think of it setwise as simply ordered n-tuples $(v_1, \ldots, v_n)$, where each $v_i \in V_i$. This differs from the direct sum, as the direct sum of these vector spaces would have a vector space structure itself, whereas here we are not considering the product as having any vector space structure.

Definition. Let $V_1, \ldots, V_n, W$ be F-vector spaces, and let

$$T : V_1 \times \cdots \times V_n \to W$$

be a function. Then, T is said to be **multilinear** if for each i, the map

$$\begin{aligned} T_i : V_i &\to W \\ v &\mapsto T(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_n) \end{aligned}$$

is an F-linear map.

We think of T as a map which is linear *in each component*, so if we hold all inputs constant except for one, T would be a linear map in that component.

Warning: This is **not** the same thing as being a linear map from the direct sum of the V_i. For example, if T were a linear map on the direct sum, then for $c \in F$ we would have

$$T(cv_1, cv_2, \ldots, cv_n) = cT(v_1, v_2, \ldots, v_n)$$

but since T is multilinear, it is linear in each component, so we would have

$$T(cv_1, cv_2, \ldots, cv_n) = c^n T(v_1, v_2, \ldots, v_n)$$

where we take the scalar c out of each component using multilinearity. We conclude that the direct sum of the V_i is not quite the correct place to capture the multilinearity of the map T. Instead, we can define what is known as the **tensor product** of the V_i, but this is a very involved topic which I will not cover (for now).

We'll denote the set of multilinear transformations using the notation

$$\mathrm{Hom}_F(V_1, \ldots, V_n; W) = \{T : V_1 \times \cdots \times V_n \to W \mid T \text{ is multilinear}\}$$

Note that the semicolon emphasizes that the last entry, W, is the codomain, whereas the entries before are factors in the domain. We'll often say that such a map is n-linear if the input is an n-fold product, or in other words, takes in n inputs. When $n = 1$, this notation precisely agrees with the notation we've previously had for the vector space of linear transformations, perhaps without a semicolon. If the V_i are all the same V, we will often use the shorthand V^n to denote V_1 through V^n. A straightforward check will show that $\mathrm{Hom}_F(V_1, \ldots, V_n; W)$ form a vector space itself, and as usual, the first question we can ask about this vector space is what the dimension is.

Theorem. Let $V_1, \ldots, V_n, W$ be F-vector spaces. Then, we have that

$$\dim(V_1, \ldots, V_n, W) = \dim(V_1)\dim(V_2)\cdots\dim(V_n)\dim(W)$$

Proof. We begin the proof by making a very important observation. Suppose we fix a vector $v \in V_n$. Then, the map

$$\begin{aligned} T' : V_1 \times \cdots \times V_{n-1} &\to W \\ (v_1, \ldots, v_{n-1}) &\mapsto T(v_1, \ldots, v_{n-1}, v) \end{aligned}$$

is still multilinear (but now only $(n-1)$-linear rather than n-linear), since if we take all but one entry fixed, linearity of that entry is inherited from the multilinearity of T. This means that for a fixed T and any $v \in V_n$, we have a way of producing an element of $\mathrm{Hom}_F(V_1, \ldots, V_{n-1}; W)$ by simply fixing the last element of T.

Lemma. Let $V_1, \ldots, V_n, W$ be F-vector spaces. Then, we have that

$$\mathrm{Hom}_F(V_1, \ldots, V_n; W) \cong \mathrm{Hom}_F(V_n, \mathrm{Hom}_F(V_1, \ldots, V_{n-1}; W))$$

where the Hom-set on the right is the set of linear transformations between two vector spaces.

Proof. The description above outlines a map which takes an element of $\mathrm{Hom}_F(V_1, \ldots, V_n; W)$, and returns a method of taking an element of V_n to an element of $\mathrm{Hom}_F(V_1, \ldots, V_{n-1}; W)$. This is precisely an element of the right hand side, and formally looks like

$$\begin{aligned} \varphi : \mathrm{Hom}_F(V_1, \ldots, V_n; W) &\to \mathrm{Hom}_F(V_n, \mathrm{Hom}_F(V_1, \ldots, V_{n-1}; W)) \\ T &\mapsto (v \mapsto ((v_1, \ldots, v_{n-1}) \mapsto T(v_1, \ldots, v_{n-1}, v))) \end{aligned}$$

A slightly involved check shows that φ is linear, and we can produce an inverse in a similar manner, by starting with a map S from V_n to $\mathrm{Hom}_F(V_1, \ldots, V_{n-1}; W)$, and constructing an element $S' \in \mathrm{Hom}_F(V_1, \ldots, V_n; W)$ by sending $(v_1, \ldots, v_n)$ to $S(v_n)(v_1, \ldots, v_{n-1})$. Mathematically, we write this as

$$\begin{aligned} \varphi^{-1} : \mathrm{Hom}_F(V_n, \mathrm{Hom}_F(V_1, \ldots, V_{n-1}; W)) &\to \mathrm{Hom}_F(V_1, \ldots, V_n; W) \\ S &\mapsto ((v_1, \ldots, v_n) \mapsto S(v_n)(v_1, \ldots, v_{n-1})) \end{aligned}$$

A tedious computation shows that this is indeed an inverse, and so we conclude these spaces are isomorphic. □

We now prove our theorem by induction.

Base Case: When $n = 1$, we have that $\dim(\mathrm{Hom}_F(V_1; W)) = \dim(V_1)\dim(W)$, by previous result.

Inductive Step: Suppose this holds for some $n - 1$. Then, we have that

$$\begin{aligned} \dim(\mathrm{Hom}_F(V_1, \ldots, V_n; W)) &= \dim(\mathrm{Hom}_F(V_n, \mathrm{Hom}_F(V_1, \ldots, V_{n-1}; W)) \\ &= \dim(V_n)\dim(\mathrm{Hom}_F(V_1, \ldots, V_{n-1}; W)) \\ &= \dim(V_n)\dim(V_1)\dim(V_2)\cdots\dim(V_{n-1})\dim(W) \end{aligned}$$

which gives the desired result. □

This proof actually illustrates an incredibly useful fact, which is that taking one entry constant in an n-linear map is the same thing as defining a map from that entry to an $(n-1)$-linear map in the other entries. Students of computer science may recognize this as **currying**.

We'll frequently restrict ourselves to the case where each V_i is the same, and so we'll consider $\mathrm{Hom}_F(V^n; W)$. When $W = F$, we say that elements of $\mathrm{Hom}_F(V^n; F)$ are n-linear **forms**, and we can investigate these objects.

Definition. Let V be an F-vector space, and consider $\mathrm{Hom}_F(V^k; F)$, the set of k-linear forms. We say that an element $f \in \mathrm{Hom}_F(V^k; F)$ is...

- ...**symmetric** if for any permutation σ of $\{1, 2, \ldots, k\}$, we have that

  $$f(v_1, \ldots, v_k) = f(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(k)})$$

  In other words, f does not depend on the order of the inputs.

- ...**anti-symmetric** (or **skew-symmetric**) if for all $1 \leq i, j \leq k$, we have that

$$f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = -f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k)$$

  In other words, swapping any two entries of f gives a minus sign.

- ...**alternating** if whenever $v_i = v_j$ for $i \neq j$, we have that

$$f(v_1, \ldots, v_k) = 0$$

  In other words, if f has a repeated input, then f will be 0.

A clever observer may notice that the last two definitions are similar, and we validate this intuition with the following theorem.

Theorem. Let V be an F-vector space. Then, if $f \in \mathrm{Hom}_F(V^k; F)$ is alternating, then f is anti-symmetric. Furthermore, if $\mathrm{char}(F) \neq 2$, then the reverse implication holds.

Proof. Suppose f is alternating. Then, we have that

$$\begin{aligned} 0 &= f(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j \ldots, v_k) \\ &= f(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_k) + f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) \\ &\quad + f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k) + f(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_k) \\ &= 0 + f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) + f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k) + 0 \\ f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) &= -f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k) \end{aligned}$$

Suppose f is anti-symmetric. Then, we have that

$$\begin{aligned} f(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_k) &= -f(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_k) \\ 2f(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_k) &= 0 \end{aligned}$$

and since $\mathrm{char}(F) \neq 2$, we conclude that $f(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_k) = 0$. □

When $\mathrm{char}(F) = 2$, we know that $1 = -1$, and so the symmetric and anti-symmetric forms are actually the same! There also exist alternating forms which are not (anti-)symmetric - can you find one?

Theorem. Let V be an F-vector space of dimension n. Then, the following sets are subspaces of $\mathrm{Hom}_F(V^k; F)$.

- The symmetric forms, denoted

$$\mathrm{Sym}^k(V) = \{f \in \mathrm{Hom}_F(V^k, F) \mid f \text{ is symmetric}\}$$

  We have that $\dim(\mathrm{Sym}^k(V)) = \binom{n+k-1}{k}$.

- The alternating forms, denoted

$$\mathrm{Alt}^k(V) = \{f \in \mathrm{Hom}_F(V^k, F) \mid f \text{ is alternating}\}$$

  We have that $\dim(\mathrm{Alt}^k(V)) = \binom{n}{k}$.

Proof. Homework. □

One of the benefits of alternating forms is that we can use them to detect linear dependence. Suppose V is an F-vector space, and $S = \{v_1, \ldots, v_k\}$ are a set of linearly dependent vectors in V. Then, without loss of generality, we can write v_1 as a linear combination of the remaining vectors.

$$v_1 = c_2 v_2 + \cdots + c_n v_n$$

Then, if $f \in \mathrm{Alt}^k(V)$, we have that

$$\begin{aligned} f(v_1, v_2, \ldots, v_k) &= f(c_2 v_2 + \cdots + c_n v_n, v_2, \ldots, v_n) \\ &= c_2 f(v_2, v_2, \ldots, v_n) + \cdots + c_n f(v_n, v_2, \ldots, v_n) \\ &= 0 \end{aligned}$$

since each entry in the sum will have a repeated input. Note that this has no guarantees if S is linearly independent - f could also evaluate to 0. We can use these alternating forms to give us information about the linear dependence of the entries. In the special case where $k = \dim(V) = n$, we have that $\dim(\mathrm{Alt}^n(V)) = \binom{n}{n} = 1$, so the space of alternating n-linear forms on V are all simply scalar multiples of each other. Furthermore, if $f \in \mathrm{Alt}^n(V)$ and $S = \{v_1, \ldots, v_n\}$ is linearly independent, then S is a basis for V, and so if $f(v_1, \ldots, v_n) = 0$, then one can show (by expanding each w_i in the given basis and using multilinearity of f) that f on any set of vectors $(w_1, \ldots, w_n) \in V^n$ will be 0, and so f will be the zero map.

However, since $\dim(\mathrm{Alt}^n(V)) = 1$, any nonzero map in $\mathrm{Alt}^n(V)$ will take linearly independent sets to a nonzero value, and linearly dependent sets to 0. If we fix a basis $\{v_1, \ldots, v_n\}$ for V, and require that $f(v_1, \ldots, v_n)$ to be 1, then we have uniquely determined an alternating n-linear map - and this will turn out to be the determinant.

10 Determinants

Planned Lecture Date(s): July 14th, 2023.

From the last lecture, we saw that if V is an F-vector space of dimension n, there is (up to scaling) only one n-linear alternating map from V^n to F, and furthermore, this map detects the linear independence of n vectors (and therefore, whether or not they form a basis for V) in that it evaluates to a nonzero value if and only if the inputs form a basis. In this section, we give a more precise definition of the determinant. We begin with some preliminary definitions.

Definition. Let $[n] = \{1, 2, \ldots, n\}$. A **permutation** of $[n]$ is a bijection $\sigma : [n] \to [n]$, and we let S_n denote the set of all permutations of $[n]$.

Students with experience in group theory will recognize this as the symmetric group on n elements. Many of the results we use today will follow from results of group theory, but I will do my best to avoid them when possible. Note that $|S_n| = n!$, since there are $n!$ ways to permute n elements. For any permutation σ, we can think of it as the (ordered) image of $\sigma(1), \sigma(2), \ldots, \sigma(n)$.

Example. One such $\sigma \in S_3$ would be defined by $\sigma(1) = 2$, $\sigma(2) = 1$, and $\sigma(3) = 3$. We would denote this by $(2, 1, 3)$.

A **transposition** is the act of swapping any two entries. For example, $(2, 1, 3)$ and $(2, 3, 1)$ differ by one transposition.

Definition. For a permutation $\sigma \in S_n$, let $T(\sigma)$ denote the number of transpositions from σ to the identity permutation. We can define the **sign** of σ to be

$$\operatorname{sgn}(\sigma) = (-1)^{T(\sigma)}$$

The fact that this definition is well-defined is very nontrivial - two different sequences of transpositions from the identity permutation to σ will very possibly have different lengths, but the fact that they must differ by an even number (and thus making $\operatorname{sgn}(\sigma)$ well-defined) is a fact from group theory which I will not prove (you can find this in most group theory textbooks). Perhaps an indication that the sign function is indeed well-defined is the following lemma.

Lemma. Let V be an F-vector space of dimension n, and suppose $f \in \operatorname{Alt}^k(V)$. Then, we have that

$$f(v_1, \ldots, v_k) = \operatorname{sgn}(\sigma) f(v_{\sigma(1)}, \ldots, v_{\sigma(k)})$$

Proof. Note that each transposition in σ contributes exactly one minus sign, as f is alternating, so the total number of minus signs contributed will multiply to $\operatorname{sgn}(\sigma)$. □

Note that if $\operatorname{sgn}(\sigma)$ were not well-defined, we could show that

$$f(v_1, \ldots, v_k) = -f(v_1, \ldots, v_k)$$

and if $\operatorname{char}(F) \neq 2$ this would imply that $f = 0$, so the existence of nonzero alternating maps verifies the well-definedness of the sign function.

We now attempt to construct the determinant. Let V be an F-vector space of dimension n, and fix $f \in \operatorname{Alt}^n(V)$. Fix a basis $B = \{v_1, \ldots, v_n\}$ for V. Since $f(v_1, \ldots, v_n) = c \neq 0$, without loss of generality, we can replace f with $\frac{1}{c} f$ to ensure that $f(v_1, \ldots, v_n) = 1$. Then, suppose that $(w_1, \ldots, w_n) \in V^n$ is an ordered set of n vectors in V. We can write

$$w_i = \sum_{j=1}^{n} c_{ij} v_j$$

We compute

$$f(w_1, \ldots, w_n) = f\left(\sum_{j_1=1}^{n} c_{1j_1} v_{j_1}, \ldots, \sum_{j_n=1}^{n} c_{nj_n} v_{j_n} \right)$$

$$= \sum_{j_1=1}^{n} \cdots \sum_{j_n=1}^{n} c_{1j_1} \cdots c_{nj_n} f(v_{j_1}, \ldots, v_{j_n})$$

At this point, we exploit the fact that f is alternating. Whenever any of the v_{j_k} agree, f will evaluate to 0, so the only terms remaining are the terms in which each of $\{1, 2, \ldots, n\}$ appears exactly once. These are precisely permutations of $[n]$, so we rewrite as

$$= \sum_{\sigma \in S_n} c_{1,\sigma(1)} \cdots c_{n,\sigma(n)} f(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$$

However, by our lemma above, the value of f is simply $\operatorname{sgn}(\sigma)$, since the value of f on $\{v_1, \ldots, v_n\}$ was chosen to be 1.

$$= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) c_{1,\sigma(1)} \cdots c_{n,\sigma(n)}$$

$$= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} c_{i,\sigma(i)}$$

and so we have produced a formula for the determinant. We take a moment to emphasize that since the determinant is a map into F, the value of the determinant is always an element of the field F over which V is a vector space.

Let's play with this formula concretely: suppose $V = F^n$, and our chosen basis is $B = \{e_1, \ldots, e_n\}$. Then, c_{ij} is the j-th component of w_i, and so if we put $(w_1, \ldots, w_n)$ into the columns of a matrix, we have

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

and our formula asks us to choose one entry from each row and column, such that no row or column is repeated, take the product of those entries together with a sign of the permutation, and sum over all permutations. In the 2×2 case, this is done by computing

$$\det\left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}$$

and in the 3×3 case, this is done by computing

$$\begin{aligned} &\det\left(\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\right) \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{aligned}$$

Note that many of the techniques you've previously seen in computing determinants (such as basket-weaving, expansion by minors, etc.) are simply mnemonic techniques for computing this sum over all permutations.

The main takeaways of our work so far is the following: for V an F-vector space of dimension n, and $B = \{v_1, \ldots, v_n\}$ a basis for V, the determinant is the unique alternating n-linear map from $V^n \to F$ satisfying $f(v_1, \ldots, v_n) = 1$. Furthermore, we have a formula to compute f on any set of n vectors in V, and we know that $f(w_1, \ldots, w_n) \neq 0$ if and only if $\{w_1, \ldots, w_n\}$ form a basis for V.

Thus far, we've only taken determinants of sets of n vectors, but we commonly think of determinants as a property of matrices. This is a fairly easy identification to make, since given a matrix, we can simply take the determinant of the n columns of the matrix. However, this identification is more insightful than it may first appear.

Theorem. Let M be an $n \times n$ matrix. Then, $\det(M) \neq 0$ if and only if M is invertible.

Proof. Since M is $n \times n$, we think of M as a map from F^n to itself with some fixed basis $\{v_1, \ldots, v_n\}$. We recall that M is invertible if and only if it takes bases to bases. Furthermore, the columns of M are precisely the images of $\{v_1, \ldots, v_n\}$, and so this set is a basis if and only if the determinant of the columns of M is nonzero. □

We'll recall one of the most important properties of determinants.

Theorem. Let A, B be $n \times n$ matrices. Then, we have that

$$\det(AB) = \det(A)\det(B)$$

Proof. This follows from a lot of tedious computation, which I will omit. □

Corollary. If A is an invertible matrix, then $\det(A^{-1}) = \det(A)^{-1}$.

Proof. We have that

$$1 = \det(\mathrm{Id}) = \det(AA^{-1}) = \det(A)\det(A^{-1})$$

which gives our desired result. □

This theorem provides many consequences: for example, we can frequently compute determinants by performing row reduction, since we can represent row operations as matrices of determinant 1. However, the most important of these is the following.

Theorem. Suppose V is an F-vector space, and $T \in \mathrm{End}(V)$. Then, the determinant of any matrix representation of T is the same.

Proof. From previous work, we've seen that if M and M' are two matrix representations of T in different bases, then they are related by the relation

$$M' = P^{-1}MP$$

We therefore have

$$\det(M') = \det(P^{-1}MP) = \det(P^{-1})\det(M)\det(P) = \det(P)^{-1}\det(P)\det(M) = \det(M)$$

□

This allows us to make a very powerful statement: if the determinant of any matrix representation of T is the same, we can in fact define the determinant of the *linear transformation* T!

Definition. Let V be a finite-dimensional vector space, and let $T \in \mathrm{End}(V)$. Then, the **determinant** of T is the determinant of any matrix representation of T.

Many nice properties follow from this.

Corollary. Let V be an F-vector space, and $S, T \in \mathrm{End}(V)$.

- T is invertible if and only if $\det(T) \neq 0$.
- $\det(S \circ T) = \det(S)\det(T)$.
- $\det(T^{-1}) = \det(T)^{-1}$.
- $\det(T^*) = \det(T)$.

The last property follows from the dual map of T being represented as a matrix by the transpose, which has the same determinant.

11 Eigenvalues and Eigenvectors

Planned Lecture Date(s): July 17th, 2023.

Recall that for a finite-dimensional F-vector space V, and a linear map $T : V \to V$, after choosing a basis for V, the map T has a matrix representation. However, this matrix representation depends heavily on the basis, so a natural question to ask may be: what choice of basis for V makes the most sense for the map T? Let's try to motivate this with some intuition. Consider the following matrix in the standard basis e for $\mathbb{R}^2$.

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

We see that M in the e-basis takes e_1 to $4e_1 + e_2$, and takes e_2 to $-2e_1 + e_2$. This is, in some sense, not desirable, as our bases "mix": M maps one basis to some combination of the two bases. However, if we instead choose the basis

$$B = \left\{ v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right\}$$

then we can write

$$M_B = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

and M in the B-basis takes v_1 to $3v_1$ and v_2 to $2v_2$. In some sense, we have separated the action of M into two components: the action of M on $\operatorname{span}(v_1)$ is completely independent from the action of M on $\operatorname{span}(v_2)$. Thus, we could write

$$\mathbb{R}^2 = \operatorname{span}(v_1) \oplus \operatorname{span}(v_2)$$

and furthermore, M restricts to a linear transformation M_i on $\operatorname{span}(v_i)$. Note that this would not be true if M were to "mix" the components! This allows us to treat these two subspaces completely separate from each other, and by decomposing our space into separate components, this gives us a more "natural" basis to express M in.

The key observation here is that for any $v \in \operatorname{span}(v_i)$, $M(v) \in \operatorname{span}(v_i)$, so this subspace is closed under the operation of M. This leads us to a definition.

Definition. Let V be an F-vector space, and $T : V \to V$. Then, a subspace $W \subset V$ is said to be an **invariant subspace** if $T(W) \subset W$. In other words, for all $w \in W$, we have that $T(w) \in W$.

If we can write V as a direct sum of invariant subspaces, such as

$$V = V_1 \oplus \cdots \oplus V_k$$

where each V_i is an invariant subspace of T, then choosing bases for each V_i and concatenating to obtain a basis for V, we can write a matrix representation for T as

$$M_T = \begin{pmatrix} M_{1,1} & & & \\ & M_{2,2} & & \\ & & \ddots & \\ & & & M_{k,k} \end{pmatrix}$$

where each $M_{i,i}$ is a square block matrix of size $\dim(V_i)$, representing the restriction of T to V_i. Furthermore, all entries outside of these blocks are zero, since none of the V_i "interact" when acted upon by T. Thus, if we'd like to put our matrix in block diagonal form, we need to find invariant subspaces which direct sum to V.

Example. We have several examples of invariant subspaces.

- For any vector space V and map $T : V \to V$, both $\{0\}$ and V are invariant subspaces.

- When $V = \mathbb{R}^3$, and T is a rotation matrix about the z-axis, given by

$$M_T = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

  we note that the 2-dimensional subspace given by the xy-plane is an invariant subspace of $\mathbb{R}^3$.

- If v is a vector satisfying $T(v) = \lambda v$ for some $\lambda \in F$, then $\operatorname{span}(v)$ is an invariant subspace of V.

This last example should look familiar to students having taken a first course in linear algebra! We'll focus on this case for now.

Definition. Let V be an F-vector space, and let $T : V \to V$ be a linear map. An **eigenvector** of T is a **non-zero** vector $v \in V$ such that for some $\lambda \in F$, we have that $T(v) = \lambda v$. The scalar λ is said to be the **eigenvalue** associated to the eigenvector v.

Note that while the eigenvector v is not allowed to be 0 (if it were, what would its eigenvalue be?), the eigenvalue λ is certainly permitted to be 0. Let's prove some statements about eigenvectors and eigenvalues.

Theorem. Let V be an F-vector space, and $T : V \to V$ a linear map.

- If v is an eigenvector of T, then the eigenvalue associated to v is unique.
- For a given eigenvalue λ, the space of eigenvectors of T with eigenvalue λ (together with 0) form a subspace of V. This space is called the **eigenspace** associated to λ.
- If $\{v_1, \ldots, v_n\} \subset S$ are a collection of eigenvectors with **distinct** eigenvalues λ_i corresponding to v_i, then $\{v_1, \ldots, v_n\}$ is linearly independent.

Proof. We prove these claims.

- Suppose $T(v) = \lambda_1 v$ and also $T(v) = \lambda_2 v$. Then, $\lambda_1 v = \lambda_2 v$, and since $v \neq 0$ by assumption, by previous homework we conclude $\lambda_1 = \lambda_2$.
- Fix an eigenvalue λ, and let V_λ denote the set of eigenvectors of V with eigenvalue λ, together with 0. Fix $c \in F$ and $v, w \in V_\lambda$. Note that $0 \in V_\lambda$, so $V_\lambda \neq \emptyset$. We then have that

$$T(cv + w) = cT(v) + T(w) = c\lambda v + \lambda w = \lambda(cv + w)$$

  so $cv + w \in V_\lambda$, as desired.

- We induct on n. Suppose $n = 1$. Since v_1 is an eigenvector, it is not 0, so the set $\{v_1\}$ is linearly independent.

  Suppose this holds for $n - 1$. Write a linear combination

$$c_1 v_1 + \cdots + c_n v_n = 0$$

  Then, we have that

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 \lambda_1 v_1 + \cdots + c_n \lambda_n v_n$$

  However, multiplying our original expression by λ_1, we have that

$$\lambda_1(c_1 v_1 + \cdots + c_n v_n) = c_1 \lambda_1 v_1 + \cdots + c_n \lambda_1 v_n$$

  and subtracting, we have that

$$0 = c_1(\lambda_1 - \lambda_1)v_1 + c_2(\lambda_2 - \lambda_1)v_2 + \cdots + c_n(\lambda_n - \lambda_1)v_n$$

Since the first term is 0, this is a linear combination of $\{v_2, \dots, v_n\}$, so by the inductive hypothesis these vectors are linearly independent. Thus, for $2 \leq i \leq n$, each $c_i(\lambda_i - \lambda_1) = 0$, but since the eigenvalues are distinct, we must have that $c_i = 0$. Thus, our original equation reads

$$0 = c_1 v_1$$

and since $v_1 \neq 0$, we must have $c_1 = 0$ as well.

□

One conclusion we can draw from this theorem is that if we find enough eigenvectors, we can use our eigenspaces as invariant subspaces, and furthermore since they are linearly independent, we can show that the sum will be direct, and this will give us the decomposition we want. The next question to ask is: how do we find eigenvectors?

We return to our original equation, where we try to find non-zero vectors v satisfying $T(v) = \lambda v$. We then have that

$$\begin{aligned} T(v) &= \lambda v \\ T(v) - \lambda v &= 0 \\ (T - \lambda \operatorname{Id})v &= 0 \end{aligned}$$

and so v must be in $\ker(T - \lambda \operatorname{Id})$. Since v is non-zero, the kernel of $T - \lambda \operatorname{Id}$ has a non-trivial element, so $T - \lambda \operatorname{Id}$ cannot be invertible. How do we detect invertibility of a linear transformation? We must have that $\det(T - \lambda \operatorname{Id}) = 0$. In any basis, the computation of the determinant is simply sums and products of the entries, so expanding the determinant as a function of λ, we obtain some polynomial in λ.

Definition. Let V be an F-vector space, and $T : V \to V$ a linear map. Then, the expression $\det(T - \lambda \operatorname{Id})$ is a polynomial in λ, and is referred to as the **characteristic polynomial** of T. Note that when we express T in a matrix, we see that the term λ occurs in n different entries, and thus the degree of the characterstic polynomial is n.

Note first that since the determinant does not depend on the choice of basis, neither does the characteristic polynomial, so this is indeed well-defined for linear transformations. Furthermore, solutions (roots) to the characteristic polynomial correspond precisely to the eigenvalues which have (non-zero) eigenvectors, and so if we are able to find an eigenvalue λ, then the eigenspace V_λ is given by

$$V_\lambda = \ker(T - \lambda \operatorname{Id})$$

Since a polynomial over a field of degree n can only have n roots, T has at most n distinct eigenvalues. Note that by construction, if λ is an eigenvalue, then $\dim(\ker(T - \lambda \operatorname{Id})) \geq 1$. Let $p(t)$ be the characteristic polynomial of T. If $p(t)$ has $n = \dim(V)$ distinct roots $\lambda_1, \dots, \lambda_n$ in F, then each corresponds to an eigenspace V_{λ_i}, and furthermore these eigenspaces must be disjoint (aside from 0) since the eigenvectors associated with distinct eigenvalues are linearly independent. Thus, we have

$$V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_n} \subset V$$

but the dimension of the left is at least n since each summand has dimension at least 1, so we conclude the left side must have exactly dimension n, and thus be equal to V. This gives us our desired decomposition! We can then write

$$M_T = \begin{pmatrix} M_{1,1} & & & \\ & M_{2,2} & & \\ & & \ddots & \\ & & & M_{n,n} \end{pmatrix}$$

where each $M_{i,i}$ is a 1×1 square matrix (and thus a single element of F). How do we know what $M_{i,i}$ is? Restricting to V_{λ_i}, we see that $T|_{V_{\lambda_i}} = \lambda_i \operatorname{Id}$, so $M_{i,i} = \lambda_i$ in some (and actually any) choice of basis. Thus,

if we simply pick $v_i \in V_{\lambda_i}$, then the set $\{v_i\}$ form a basis of eigenvectors (or **eigenbasis**) for V. When an eigenbasis exists, we say that T is **diagonalizable** (since the only nonzero entries are on the diagonal). What we have shown here is that when the characteristic polynomial has $n = \dim(V)$ distinct roots, T will always be diagonalizable, using this construction. However, we have required some fairly strong assumptions, and in the next lecture, we'll explore what happens if those assumptions are relaxed.

Since the determinant is invariant under change of basis, if we change our basis so that M is diagonal, the determinant is simply the product of the diagonal entries. Thus, we conclude that when T is diagonalizable, the determinant is given by the product of the roots. It turns out that this will be true even when T is not diagonalizable, but we'll need a bit more machinery to see this. However, the determinant is not the only quantity associated to a matrix that has an interesting interpretation in terms of eigenvalues.

Definition. Let M be an $n \times n$ matrix. Then, the **trace** of M, denoted $\mathrm{tr}(M)$ is defined to be the sum of the diagonal elements of M.

Theorem. Let A and B be $n \times n$ matrices. Then, we have that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

Proof. This can be checked using a tedious computation. □

Theorem. Let V be a F-vector space, and let $T : V \to V$. Then, the trace of T, denoted $\mathrm{tr}(T)$, is the trace of any matrix representation of T, and is well-defined.

Proof. We know that if M and M' are two matrix representations of T, then $M' = P^{-1}MP$. We therefore have

$$\mathrm{tr}(M') = \mathrm{tr}(P^{-1}MP) = \mathrm{tr}(PP^{-1}M) = \mathrm{tr}(M)$$

and thus the trace is independent of basis. □

Corollary. Let V be an F-vector space, and let $T : V \to V$ be a diagonalizable linear map. Then, $\mathrm{tr}(T)$ is the sum of the eigenvalues, and $\det(T)$ is the product of the eigenvalues.

Proof. Since T is diagonalizable, choose a basis for V in which T is diagonal, with the eigenvalues on the diagonal. Then, both trace and determinant are independent of basis, so they can be computed in this basis. □

One observation to make is that if 0 is an eigenvalue of T, then $\det(T) = 0$, so T is not invertible. This can also be seen by the definition of an eigenvalue, as $T(v) = 0$ has a nonzero solution.

It turns out that in general, even if our matrix is not diagonalizable, we can still make a similar statement, but we will return to this in a little bit.

12 Nilpotent Maps and Cyclic Subspaces

Planned Lecture Date(s): July 19th, 2023.

We take a brief interlude from eigenvalues and eigenvectors to discuss a topic which will be very useful to us soon.

Definition. Let $T : V \to V$ be a linear map. T is said to be **nilpotent** if $T^k = 0$ for some natural number k. The smallest k such that $T^k = 0$ is the **index of nilpotency**.

Our goal is to show that if a linear map $T : V \to V$ is nilpotent, we can find a basis for V in which the matrix representation of T is particularly nice.

Definition. Let $T : V \to V$ be a linear map. Then, for $v \in V$, the **cyclic subspace** associated to v is given by

$$C(v) = \text{span}\{v, T(v), T^2(v), \dots\}$$

Note that this definition depends on T.

Note that since T is nilpotent, T^k will eventually be 0, and so this is the span of a finite set.

Lemma. The set $C(v)$ is a subspace with basis $\{v, T(v), T^2(v), \dots\}\backslash\{0\}$. A **cyclic basis** is a (union of) bases of this form.

Proof. If $v = 0$, $C(v)$ is the zero space, so this statement is true. Otherwise, assume $v \neq 0$. Since $C(v)$ is the span of a set of vectors, it is automatically a subspace, and the given basis spans $C(v)$, so it is enough to show that this set is linearly independent. Suppose k is the smallest integer such that $T^{k+1}(v) = 0$, and therefore our proposed basis is $\{v, T(v), T^2(v), \dots, T^k(v)\}$. We have that

$$c_0 v + c_1 T(v) + c_2 T^2(v) + \cdots + c_k T^k(v) = 0$$

for some coefficients c_i. Suppose not all c_i are 0. Then, let j be the minimum integer such that $c_j \neq 0$. Then, applying T^{k-j} to the entire expression, we have

$$c_j T^k(v) + c_{j+1} T^{k+1}(v) + \cdots + c_k T^{2k-j}(v) = c_j T^k(v) = T^{k-j}(0) = 0$$

and since $T^k(v) \neq 0$ by assumption, we have that $c_j = 0$, which is a contradiction. We conclude each $c_i = 0$, as desired. □

Lemma. The index of nilpotency of T is equal to the dimension of the largest cyclic subspace of V (associated to T).

Proof. Let k be the index of nilpotency of T. Since $T^k = 0$, by the above lemma, any cyclic subspace $C(v)$ must have basis

$$\{v, T(v), T^2(v), \dots, T^{k'}(v)\}$$

where $k' < k$, where $T^{k'+1}(v) = 0$. Thus, we have that $k' + 1 = \dim(C(v)) \leq k$ for all $v \in V$.

Similarly, suppose every cyclic subspace $C(v)$ has dimension at most k'. Suppose $T^{k'} \neq 0$. Then, for some $v \in V$, $T^{k'}(v) \neq 0$, so the set

$$\{v, T(v), \dots, T^{k'}(v), \dots\}$$

would be a basis for for a cyclic subspace of dimension at least $k' + 1$. Thus, we must have that $T^{k'} = 0$, so $k \leq k'$. We conclude that k is precisely the dimension of the largest cyclic subspace of V. □

Theorem. Let V be an F-vector space of dimension n, and let $T : V \to V$ be a nilpotent linear map. Then, we can decompose

$$V = C(v_1) \oplus \cdots \oplus C(v_\ell)$$

as a direct sum of cyclic subspaces. Furthermore, this decomposition (associated to T) is unique up to the number of subspaces (ℓ) and the dimension of each subspace.

Proof. We proceed by induction on the index of nilpotency of T. Suppose the index of nilpotency of T is 1. Then, $T = 0$, and any basis for V given by $\{v_1, \dots, v_n\}$ is a cyclic basis, where

$$V = C(v_1) \oplus \cdots \oplus C(v_n)$$

Suppose the claim holds for when T has index of nilpotency $k-1$. Restricting T to a map $T : \mathrm{im}(T) \to \mathrm{im}(T)$, since we have already applied T once, the restriction of T has index of nilpotency $k-1$. Thus, by the inductive hypothesis, we have that

$$\mathrm{im}(T) = C(v_1) \oplus \cdots \oplus C(v_\ell)$$

for some ℓ', and this is unique up to dimension of the cyclic subspaces and number of summands. Let $k_i = \dim(C(v_i))$. Then, the set

$$\bigcup_{i=1}^{\ell} \{v_i, T(v_i), \dots, T^{k_i-1}(v_i)\}$$

is a basis for $\mathrm{im}(T)$. By the Rank-Nullity Theorem, we have that $\dim(V) = \dim(\mathrm{im}(T)) + \dim(\ker(T))$, so in order to extend this to a basis of V, we must produce $\dim(\ker(T))$ new basis vectors. We notice that

$$\{T^{k_1-1}(v_1), \dots, T^{k_\ell-1}(v_\ell)\}$$

are ℓ basis vectors which are already in $\ker(T)$, so choose $\{u_1, \dots, u_j\}$ (with $j = \dim(\ker(T)) - \ell$) to extend this to a basis of $\ker(T)$, which is linearly independent from the basis for $\mathrm{im}(T)$. It remains to find ℓ more basis vectors for V. Since $v_i \in \mathrm{im}(T)$ for all i, choose w_i such that $T(w_i) = v_i$, and we note there are precisely ℓ such w_i. We claim that the set

$$\{w_1, \dots, w_\ell\} \cup \{u_1, \dots, u_j\} \cup \left(\bigcup_{i=1}^{\ell} \{v_i, T(v_i), \dots, T^{k_i-1}(v_i)\} \right)$$

form a basis for V. Since the number of vectors is equal to $\dim(V)$, it suffices to check that these are linearly independent. Suppose we have a linear combination

$$a_1 w_1 + \cdots + a_\ell w_\ell + b_1 u_1 + \cdots + b_j u_j + \sum_{p=1}^{\ell} \sum_{q=0}^{k_i-1} T^q(v_p) = 0$$

Then, applying T to the entire expression, we have that

$$a_1 v_1 + \cdots + a_\ell v_\ell + \sum_{p=1}^{\ell} \sum_{q=0}^{k_i-1} T^{q+1}(v_p) = 0$$

which is a linear combination of the basis for $\mathrm{im}(T)$, and thus each a_i must be 0. What remains is a basis of $\mathrm{im}(T) + \ker(T)$ by construction, and therefore must also satisfy linear independence. Note that adding w_i to the basis replaces $C(v_i)$ with $C(w_i)$, so we write

$$V = C(w_1) \oplus \cdots \oplus C_{w_\ell} \oplus C(u_1) \oplus \cdots C(u_j)$$

which gives the desired decomposition. Furthermore, note that ℓ is determined by the decomposition of $\mathrm{im}(T)$, which is assumed to be unique, and $j = \dim(\ker(T)) - \ell$ is determined by $\ker(T)$ and ℓ, so this decomposition is indeed unique up to reordering and dimension of components. □

Using this decomposition, we can make a very important statement about matrices.

Definition. An $n \times n$ square matrix is said to be a **nilpotent Jordan block** if it is of the form

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

where all entries are 0 aside from the superdiagonal, on which all entries are 1.

Corollary. Let $T : V \to V$ be a nilpotent map. Then, we can choose a basis for V in which T has matrix representation

$$M_T = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix}$$

where each J_i is a nilpotent Jordan block, and this representation is unique up to reordering of the blocks.

Proof. Suppose that T is nilpotent. Then, decompose V as a direct sum of cyclic subspaces as

$$V = C(v_1) \oplus \cdots \oplus C(v_k)$$

Then, restricting T to $C(v_i)$ and setting $k_i = \dim(C(v_i))$, we see that on the basis $\{w_0, \ldots, w_{k_i - 1}\} = \{v, T(v), T^2(v), \ldots, T^{k_i - 1}(v)\}$, the map T acts as

$$T(w_j) = \begin{cases} w_{j+1} & j < k_i - 1 \\ 0 & j = k_i - 1 \end{cases}$$

and so restricted to $C(v_i)$, the map T has the matrix representation

$$T|_{C(v_i)} = \begin{pmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & \ddots & \ddots & \\ & & & & 0 & 1 \\ & & & & & 0 \end{pmatrix}$$

of size $k_i \times k_i$. Note that this takes our above basis in the reverse order. A similar argument can be used to instead place all ones on the subdiagonal, using the above basis ordering, so the choice of diagonal is generally taken as a convention to be above the main diagonal (although many authors will disagree). Thus, taking such a choice of basis on each $C(v_i)$ gives precisely a nilpotent Jordan block of size k_i, which gives our desired decomposition. □

An interesting consequence of this fact is the following.

Corollary. Let $T : V \to V$ be a nilpotent linear map. Then, $\operatorname{tr}(T) = 0$.

13 Generalized Eigenspaces and Jordan Normal Form

Planned Lecture Date(s): July 21st, 2023.

We saw previously that under certain very nice conditions, we could diagonalize a linear map $T : V \to V$, and decompose it into a collection of linear maps on invariant subspaces. However, this required some restrictive assumptions, and we attempt to relax these assumptions here. Let's illustrate some subtleties of what could go wrong.

Let $V = \mathbb{R}^3$. Suppose T is represented by the matrix

$$M_T = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then, we compute the characteristic polynomial to be

$$\det(T - \lambda \operatorname{Id}) = \det\left(\begin{pmatrix} 3-\lambda & 0 & 0 \\ 0 & 3-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix}\right) = -(\lambda-3)^2(\lambda-1)$$

and we see that $\lambda = 3$ is a root of this polynomial, with multiplicity 2. Since the characteristic polynomial has a repeated root, we cannot simply apply our previous construction for an invariant subspace decomposition. The only possible eigenvalues for M_T is $\lambda = 3$ and $\lambda = 1$. We then consider

$$\ker(T - 3\operatorname{Id}) = \ker\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}\right) = \left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \;\middle|\; a, b \in F \right\}$$

and we see that $V_3 = \ker(T - 3\operatorname{Id})$ is a 2-dimensional subspace of $\mathbb{R}^3$. Since $\lambda = 1$ is also an eigenvalue, V_1 has dimension at least 1, and is linearly independent with any two basis vectors in V_3, so V_1 has dimension exactly 1. We can then write

$$V = V_3 \oplus V_1$$

and we have expressed V as a direct sum of invariant subspaces, and we can write M_T in block form as

$$M_T = \begin{pmatrix} M_{1,1} & \\ & M_{2,2} \end{pmatrix}$$

where $M_{1,1}$ is a 2×2 square matrix corresponding to V_3, and $M_{2,2}$ is a 1×1 square matrix corresponding to V_1. However, since M maps any $v \in V_3$ to $3v$, we know that $M|_{V_3} = 3\operatorname{Id}$, so any basis for V_3 will consist of two eigenvectors of M. Using a choice of basis for V_3, we have that

$$M_{1,1} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

and combining that with a choice of basis for V_1, we can write M_T in diagonal form (and it turns out that it already is in diagonal form). In this case, we were still able to diagonalize M_T, even though there was a repeated root in the characteristic polynomial. However, this is not always true: consider the example of

$$M_T = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

We compute the characteristic polynomial to be the same, as we have

$$p(\lambda) = -(\lambda-3)^2(\lambda-1)$$

Solving for V_3, we see that

$$\ker(T - 3\,\mathrm{Id}) = \ker\left(\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}\right) = \left\{ \begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix} \middle| \, a \in F \right\}$$

and unfortunately, this space is only dimension 1. It turns out that V_1 is also dimension 1, and so we have that $V_3 \oplus V_1$ is only dimension 2, inside of V, which is dimension 3. The issue appears to be the following: the eigenvalue 3 occupied 2 roots worth of space, but only contributed 1 dimension of eigenspace. This leads us to make the following definitions.

Definition. Let V be an F-vector space, and let $T : V \to V$ be a linear map. Let λ be an eigenvalue of T. Then, we define...

- ...the **algebraic multiplicity** of λ to be multiplicity of the root $t = \lambda$ in the characteristic polynomial $p(t)$ for T. In other words, it is the number of times the factor $(x - \lambda)$ occurs in the factorization of $p(t)$.
- ...the **geometric multiplicity** of λ to be $\dim(\ker(T - \lambda\,\mathrm{Id}))$. In other words, it is the dimension of the eigenspace associated to λ.

Theorem. Let V be an F-vector space, and let $T : V \to V$ be a linear map. Let λ be an eigenvalue of T. Then, the algebraic multiplicity of λ is at least the geometric multiplicity of λ.

Proof. Fix an eigenvalue λ of T, and suppose it has geometric multiplicity k. Then, we have that $\dim(\ker(T - \lambda\,\mathrm{Id})) = k$, so choose a basis $\{v_1, \ldots, v_k\}$ for $\ker(T - \lambda\,\mathrm{Id})$. Extend this to a basis $\{v_1, \ldots, v_n\}$ of V. Let M be the matrix with columns

$$M = \begin{pmatrix} \vdots & \vdots & & \vdots & \vdots \\ v_1 & v_2 & \cdots & v_{n-1} & v_n \\ \vdots & \vdots & & \vdots & \vdots \end{pmatrix}$$

Then, we have that

$$(A - t\,\mathrm{Id})M = \begin{pmatrix} \vdots & & \vdots & \vdots & & \vdots \\ (\lambda - t)v_1 & \cdots & (\lambda - t)v_k & (A - t\,\mathrm{Id})v_{k+1} & \cdots & (A - t\,\mathrm{Id})v_n \\ \vdots & & \vdots & \vdots & & \vdots \end{pmatrix}$$

Taking a determinant on both sides, we have that

$$\det(A - t\,\mathrm{Id})\det(M) = (\lambda - t)^k \det\left(\begin{pmatrix} \vdots & & \vdots & \vdots & & \vdots \\ v_1 & \cdots & v_k & (A - t\,\mathrm{Id})v_{k+1} & \cdots & (A - t\,\mathrm{Id})v_n \\ \vdots & & \vdots & \vdots & & \vdots \end{pmatrix}\right)$$

and since $\det(M)$ is a constant which does not depend on t, we have that the characteristic polynomial $\det(A - t\,\mathrm{Id})$ has a factor of $(\lambda - t)^k$, so the algebraic multiplicity of λ is at least k, as desired. □

We've therefore created a condition for diagonalizability: for each eigenvalue of T, the algebraic multiplicity is always greater than or equal to the geometric multiplicity. If these quantities are equal *for every eigenvalue*, then using our previous construction, each V_λ has the right dimension corresponding to the number of times λ is a root of the characteristic polynomial. If furthermore, the characteristic polynomial $p(t)$ has all n roots (counting multiplicity) in F, then by our previous argument, we can direct sum these to create all of V. We summarize this in the following manner.

Theorem. Let V be an F-vector space of dimension n, and let $T : V \to V$ be a linear map. Let $p(t)$ denote the characteristic polynomial of T. Suppose $p(t)$ has n roots in F (counting multiplicity), and each root λ_i has algebraic multiplicity equal to its geometric multiplicity. Then, we can write

$$V = \bigoplus_{i=1}^{k} V_{\lambda_i}$$

where k is the number of distinct roots. Furthermore, $T|_{V_{\lambda_i}} = \lambda_i \operatorname{Id}$ on V_{λ_i}, and so we can choose a basis of eigenvectors in which T is of the form

$$T = \begin{pmatrix} M(\lambda_1) & & & \\ & M(\lambda_2) & & \\ & & \ddots & \\ & & & M(\lambda_k) \end{pmatrix}$$

where $M(\lambda_i)$ denotes the block $\lambda_i \operatorname{Id}$ of size equal to the multiplicity of λ_i. Therefore, T is diagonalizable.

If the geometric multiplicity does not agree with the algebraic multiplicity, is there anything that we can do? It turns out the answer is yes. For now, we will assume that F is algebraically closed (often abbreviated as $F = \overline{F}$), to ensure that the characteristic polynomial does indeed have all of its roots (but possibly not distinct).

Definition. Let V be an F-vector space of dimension n, and let $T : V \to V$ be a linear map. Let λ be an eigenvalue of T. Then, we can define the **generalized eigenspace** associated to λ as

$$V_{[\lambda]} = \{v \in V \mid \exists\ n \geq 1 \text{ such that } (T - \lambda \operatorname{Id})^n(v) = 0\}$$

One can check that this indeed forms a subspace, and note that this would agree with the definition of an eigenspace if we only allowed $n = 1$. Our goal is to use these generalized eigenspaces in place of our regular eigenspaces. Note that

$$\ker((T - \lambda \operatorname{Id})) \subset \ker((T - \lambda \operatorname{Id})^2) \subset \ker((T - \lambda \operatorname{Id})^3) \subset \cdots \subset \bigcup_{i=1}^{\infty} \ker((T - \lambda \operatorname{Id})^i) = V_{[\lambda]}$$

and furthermore, each $\ker(T - \lambda \operatorname{Id})^k$ is a subspace of V. Since this is an increasing sequence of subspaces of V, and V is finite-dimensional, this chain must eventually stabilize for some N, and so we have that

$$V_{[\lambda]} = \ker((T - \lambda \operatorname{Id})^N) = \ker((T - \lambda \operatorname{Id})^M)$$

for all $M > N$.

Lemma. We have that $V = \ker((T - \lambda \operatorname{Id})^N) \oplus \operatorname{im}((T - \lambda \operatorname{Id})^N)$.

Proof. We first show $\ker((T - \lambda \operatorname{Id})^N) \cap \operatorname{im}((T - \lambda \operatorname{Id})^N) = \{0\}$. Let $v \in \ker((T - \lambda \operatorname{Id})^N) \cap \operatorname{im}((T - \lambda \operatorname{Id})^N)$. Then, $(A - \lambda \operatorname{Id})^N v = 0$, and furthermore we can find $w \in V$ with $(A - \lambda \operatorname{Id})^N w = v$. Then, we have that

$$(A - \lambda \operatorname{Id})^{2N}(w) = (A - \lambda \operatorname{Id})^N (A - \lambda \operatorname{Id})^N (w) = (A - \lambda \operatorname{Id})^N (v) = 0$$

Thus, $w \in \ker((A - \lambda \operatorname{Id})^{2N}) = \ker((A - \lambda \operatorname{Id})^N)$, so $v = 0$ as desired.

Then, we have that $\ker((T - \lambda \operatorname{Id})^N) \oplus \operatorname{im}((T - \lambda \operatorname{Id})^N)$ is a subspace of V of dimension $\dim(\ker((T - \lambda \operatorname{Id})^N)) + \dim(\operatorname{im}((T - \lambda \operatorname{Id})^N)) = \dim(V)$ by Rank-Nullity, so these are equal, as desired. □

Theorem. Let V be an F-vector space of dimension n, and let $T : V \to V$ be a linear map. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then, we can decompose

$$V = \bigoplus_{i=1}^{k} V_{[\lambda]}$$

Proof. We first check that $\ker((T-\lambda\,\mathrm{Id})^N)$ and $\mathrm{im}((T-\lambda\,\mathrm{Id})^N)$ are invariant subspaces. Note that T commutes with $T-\lambda\,\mathrm{Id}$, so we have that

- For $v \in \ker((T-\lambda\,\mathrm{Id})^N)$, we have that $(T-\lambda\,\mathrm{Id})^N(T(v)) = T((T-\lambda\,\mathrm{Id})^N(v)) = 0$, so $T(v) \in \ker((T-\lambda\,\mathrm{Id})^N)$.
- For $v \in \mathrm{im}((T-\lambda\,\mathrm{Id})^N)$, there exists $w \in V$ with $(T-\lambda\,\mathrm{Id})^N(w) = v$. Then, $(T-\lambda\,\mathrm{Id})^N(T(w)) = T((T-\lambda\,\mathrm{Id})^N(w)) = T(v)$, so $T(v) \in \mathrm{im}((T-\lambda\,\mathrm{Id})^N)$.

We now show the following: the only eigenvalue of T on $\ker((T-\lambda\,\mathrm{Id})^N)$ is λ, and furthermore λ is not an eigenvalue of T on $\mathrm{im}((T-\lambda\,\mathrm{Id})^N)$.

- Note that $(T-\lambda\,\mathrm{Id})^{N-1}(w)$ for any $w \in V$ will be an eigenvector of T in $\ker((T-\lambda\,\mathrm{Id})^N)$ with eigenvalue λ, so λ is certainly an eigenvalue of T on $\ker((T-\lambda\,\mathrm{Id})^N)$. If λ' is another eigenvalue with eigenvector $0 \neq w \in \ker((T-\lambda\,\mathrm{Id})^N)$, then $(T-\lambda\,\mathrm{Id})w = (\lambda'-\lambda)w$, but since w is in the kernel, we have that
$$0 = (T-\lambda\,\mathrm{Id})^N(w) = (\lambda'-\lambda)^N(w)$$
and so $(\lambda'-\lambda)^k = 0$, which implies $\lambda = \lambda'$.
- Suppose that λ was an eigenvalue of T in $\mathrm{im}((T-\lambda\,\mathrm{Id})^N)$. Then, for some $v \in \mathrm{im}((T-\lambda\,\mathrm{Id})^N)$, we would have $T(v) = \lambda v$. But then $(T-\lambda\,\mathrm{Id})v = 0$, so $v \in \ker(T-\lambda\,\mathrm{Id}) \subset \ker((T-\lambda\,\mathrm{Id})^N)$. However, we proved that $\ker((T-\lambda\,\mathrm{Id})^N) \cap \mathrm{im}((T-\lambda\,\mathrm{Id})^N) = \{0\}$, which is a contradiction. Thus, λ is not an eigenvalue for T in $\mathrm{im}((T-\lambda\,\mathrm{Id})^N)$.

We now proceed by strong induction on $\dim(V)$. When $\dim(V) = 0$, this statement holds vacuously. Suppose $\dim(V) = n$, and let T have eigenvalues $\lambda_1, \ldots, \lambda_k$. Then, we can write
$$V = \ker((T-\lambda_1\,\mathrm{Id})^{N_1}) \oplus \mathrm{im}((T-\lambda_1\,\mathrm{Id})^{N_1})$$
Since λ_1 is an eigenvalue, we have that $\dim(\ker((T-\lambda\,\mathrm{Id})^{N_1})) \geq 1$ since it contains at least the eigenvectors associated to λ_1, so $\dim(\mathrm{im}((T-\lambda_1\,\mathrm{Id})^{N_1})) < \dim(V)$. Furthermore, by our above work, the eigenvalues associated with $\mathrm{im}((T-\lambda_1\,\mathrm{Id})^{N_1})$ are precisely $\lambda_2, \ldots, \lambda_k$. We then obtain our result by induction. □

Corollary. We have that $\dim(V_{[\lambda]})$ is the algebraic multiplicity of λ.

Proof. Let $n(\lambda)$ denote the algebraic multiplicity of λ. Note that since V is the direct sum of the $V_{[\lambda]}$, the characteristic polynomial is the product of the characteristic polynomials restricted to each component (this follows from picking a block diagonal matrix). Since the only component which contributes factors of $(t-\lambda)$ to the characteristic polynomial of T is $\ker((T-\lambda\,\mathrm{Id})^N)$, and T restricted to this space has only eigenvalue λ, and thus characteristic polynomial $(t-\lambda)^{\dim(V_{[\lambda]})}$, we have that $n(\lambda) = \dim(V_{[\lambda]})$ as desired. □

We've therefore concluded that we can write T in block form as
$$T = \begin{pmatrix} M(\lambda_1) & & & \\ & M(\lambda_2) & & \\ & & \ddots & \\ & & & M(\lambda_k) \end{pmatrix}$$
where each matrix $M(\lambda_i)$ is a square matrix of size $\dim(V_{[\lambda]})$ on the $V_{[\lambda]}$ component. The question remains to figure out what each $M(\lambda_i)$ should be.

We notice that $T-\lambda_i\,\mathrm{Id}$ restricted to $V_{[\lambda_i]}$ is nilpotent on $V_{[\lambda_i]}$. Thus, we can make a definition to finish off our work.

Definition. An $n \times n$ square matrix is said to be a **Jordan block** (associated to the eigenvalue λ) if it is of the form

$$\begin{pmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \lambda & 1 & & \\ & & & \ddots & \ddots & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{pmatrix}$$

where all entries are 0 aside from the diagonal, on which all entries are λ, and the superdiagonal, on which all entries are 1.

Using this theorem, we can conclude the following result.

Theorem. (Jordan Normal/Canonical Form) Let V be an F-vector space of dimension n, let $T : V \to V$ be a linear map, and suppose that the characteristic polynomial $p(t)$ has n roots in F, counted with multiplicity, denoted $\lambda_1, \ldots, \lambda_k$. Then, we can find a basis of V such that the matrix representation of T is given by

$$M_T = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_\ell \end{pmatrix}$$

where each J_i is a Jordan block associated to some eigenvalue λ_j. Furthermore, this matrix representation is unique up to reordering the Jordan blocks. Additionally, we have that for any eigenvalue λ_j, the geometric multiplicity is given by the number of Jordan blocks corresponding to λ_j, and the algebraic multiplicity is given by the sum of the sizes of all Jordan blocks corresponding to λ_j.

Proof. We've shown previously that T decomposes into $M(\lambda_i)$ on $V_{[\lambda_i]}$, and $T - \lambda_i \operatorname{Id}$ is nilpotent on $V_{[\lambda_i]}$. Thus, we can choose a basis in which $T - \lambda_i \operatorname{Id}$ is in nilpotent Jordan block form, and thus T restricted to $V_{[\lambda_i]}$ is in Jordan block form (associated to the same eigenvalue). Plugging this in for each $M(\lambda_i)$ gives the desired result, and the statements about geometric and algebraic multiplicity are true by construction. □

This gives us a canonical form to represent linear maps! Even if our matrix is not diagonalizable, we can still put it into Jordan Normal Form, as long as F is algebraically closed. The advantage of canonical forms is that we have a standard way to make arguments about matrices. For example, here is a very important theorem.

Theorem. (Cayley-Hamilton) Let $T : V \to V$ be a linear map with characteristic polynomial $p(t)$. Then, $p(T) = 0$.

Proof. I'll prove this for a matrix in Jordan Normal Form, and leave the rest as homework. Note that $p(t)$ factors as

$$p(t) = (t - \lambda_1)^{N_1} \cdots (t - \lambda_k)^{N_k}$$

and note that for any $v \in V$, we can decompose v into a linear combination of elements of the generalized eigenspaces $V_{[\lambda_i]}$. Then, applying $p(T)$ to v with T, we see that each factor $(T - \lambda_i)^{N_i}$ either leaves each component untouched, or sends it to 0. Thus, all vectors $v \in V$ are sent to 0, so $p(T) = 0$ as desired. □

We can also return to some previous statements, which we could only prove when T was diagonalizable.

Theorem. Let V be an F-vector space ($F = \overline{F}$), and $T : V \to V$ a linear map. Then, we have that $\det(T)$ is the product of the eigenvalues (counting multiplicity), and $\operatorname{tr}(T)$ is the sum of the eigenvalues (counting multiplicity).

Proof. We put T in Jordan Normal Form, and compute trace and determinant. □

There is one last issue to address, however: if F is not algebraically closed, we may be in trouble. There are a few ways to fix this situation: if we are computing matrices over any field F, we could simply pass to $\overline{F}$, and pretend that our matrix T is really defined over the larger field $\overline{F}$. We frequently see this done implicitly when computing eigenvalues of $\mathbb{R}$-matrices by passing to $\mathbb{C}$. However, we can make the following observation: if for some fixed transformation T, if F contains all eigenvalues of T, then we can put T in Jordan Normal Form even if F is not algebraically closed! We only need F to contain the eigenvalues of T in order to do so.

Another final observation is the following. If F does not contain the eigenvalues of T, we know that $\overline{F}$ certainly does. Then, both the sum and product of the eigenvalues of T can be computed in $\overline{F}$, but they can also be computed with a matrix representation for T over F, which we then implicitly think of as a matrix with $\overline{F}$ entries. Since these must agree, as determinant and trace are invariant under change of basis, we conclude that the sum and product of the eigenvalues in $\overline{F}$ must actually be an element of F! You'll see on homework that this can be seen through another viewpoint as coefficients of the characteristic polynomial.

Example. Consider the matrix

$$M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

as a map $M : \mathbb{R}^2 \to \mathbb{R}^2$. One may notice that this corresponds to the matrix which rotates $\mathbb{R}^2$ by an angle of $\frac{\pi}{2}$. We compute the characteristic polynomial to be

$$p(\lambda) = \lambda^2 + 1$$

which has no roots over $\mathbb{R}$, but over $\mathbb{C}$ has roots $\pm i$. We have that $\operatorname{tr}(M) = i + (-i) = 0$ and $\det(M) = (i)(-i) = 1$, and so the rules regarding sums and products of eigenvalues still holds, even though the trace and determinant are purely real numbers.

One interesting application of Jordan Normal Form occurs in the following manner. We can define the exponential of a matrix as follows.

Definition. Let M be an $n \times n$ matrix over a field F, with $\operatorname{char}(F) = 0$. Then, we define

$$\exp(M) = \mathrm{Id} + M + \frac{1}{2}M^2 + \cdots = \sum_{i=1}^{\infty} \frac{1}{i!} M^i$$

We will sometimes write e^M to mean $\exp(M)$.

Taking the exponential of a matrix often shows up in various applications, such as solving systems of ordinary differential equations, or relating Lie Algebras to Lie Groups. Showing that this infinite sum converges (or even makes sense) is a question for analysis, but if we suspend our disbelief momentarily, we can consider the following.

Theorem. Let M be an $n \times n$ diagonalizable matrix, and let $M = P^{-1}DP$, where D is a diagonal $n \times n$ matrix, and P is an invertible $n \times n$ matrix. Then, we have that

$$\exp(M) = P^{-1} \exp(D) P$$

Proof. We have that

$$\begin{aligned} \exp(M) &= \exp(P^{-1}DP) \\ &= \sum_{i=1}^{\infty} \frac{1}{i!} (P^{-1}DP)^i \\ &= \sum_{i=1}^{\infty} \frac{1}{i!} P^{-1}(D)^i P \end{aligned}$$

$$= P^{-1}\left(\sum_{i=1}^{\infty}\frac{1}{i!}(D)^i\right)P$$
$$= P^{-1}\exp(D)P$$

which is our desired result. □

The interesting consequence here is that $\exp(D)$ is incredibly easy to compute, since D is diagonal. We see that sums and products of the diagonal entries do not interact, and thus we have that if

$$D = \begin{pmatrix} a_{1,1} & & & \\ & a_{2,2} & & \\ & & \ddots & \\ & & & a_{n,n} \end{pmatrix}$$

$$\exp(D) = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} + \begin{pmatrix} a_{1,1} & & & \\ & a_{2,2} & & \\ & & \ddots & \\ & & & a_{n,n} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} a_{1,1}^2 & & & \\ & a_{2,2}^2 & & \\ & & \ddots & \\ & & & a_{n,n}^2 \end{pmatrix}$$
$$= \begin{pmatrix} 1+a_{1,1}+\frac{1}{2}a_{1,1}^2 & & & \\ & 1+a_{2,2}+\frac{1}{2}a_{2,2}^2 & & \\ & & \ddots & \\ & & & 1+a_{n,n}+\frac{1}{2}a_{n,n}^2 \end{pmatrix}$$
$$= \begin{pmatrix} \exp(a_{1,1}) & & & \\ & \exp(a_{2,2}) & & \\ & & \ddots & \\ & & & \exp(a_{n,n}) \end{pmatrix}$$

and we can simply perform two matrix multiplications with P and P^{-1} to compute $\exp(M)$. However, if D is not diagonalizable, this trick gets slightly more complicated (but we can still salvage something!). If D is instead taken to be in Jordan Normal Form, then on each block, we have that

$$D|_{V_\lambda} = \lambda\,\mathrm{Id} + N$$

where N is a nilpotent operator. Thus, we have that

$$\exp(M) = P^{-1}\exp(D)P = P^{-1}\exp(\lambda\,\mathrm{Id}+N)P = P^{-1}\exp(\lambda\,\mathrm{Id})\exp(N)P = e^{\lambda}P^{-1}\exp(N)P$$

where one can check that the exponential is still multiplicative on matrices. Note that taking the exponential of $\lambda\,\mathrm{Id}$ is easy, since this is diagonal, but furthermore, taking $\exp(N)$ is also not too difficult, since N is nilpotent! Thus, some finite power of N will be 0, so $\exp(N)$ is in fact a finite sum of powers of N, together with coefficients.

14 Bilinear Forms

Planned Lecture Date(s): July 31st, 2023.

We recall from a few weeks ago that if V is an F-vector space of dimension n, then we can define the set $\mathrm{Hom}(V^2; F)$ to be the set of **bilinear forms**, or in other words, multilinear maps from $V \times V$ to F. How can we represent bilinear forms in coordinates? Let's fix a basis $\{v_1, \ldots, v_n\}$ for V. Then, fix $w, w' \in V$ with

$$w = a_1 v_1 + \cdots + a_n v_n \qquad w' = b_1 v_1 + \cdots + b_n v_n$$

Then, we compute that

$$\begin{aligned}
B(w, w') &= B\left(\sum_{i=1}^{n} a_i v_i, \sum_{j=1}^{n} b_i v_i\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j B(v_i, v_j) \\
&= \sum_{i=1}^{n} a_i \sum_{j=1}^{n} b_j B(v_i, v_j) \\
&= \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n} b_j B(v_1, v_j) \\ \sum_{j=1}^{n} b_j B(v_2, v_j) \\ \vdots \\ \sum_{j=1}^{n} b_j B(v_3, v_j) \end{pmatrix} \\
&= \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \begin{pmatrix} B(v_1, v_1) & B(v_1, v_2) & \cdots & B(v_1, v_n) \\ B(v_2, v_1) & B(v_2, v_2) & & \\ \vdots & & \ddots & \\ B(v_n, v_1) & & & B(v_n, v_n) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}
\end{aligned}$$

and we notice that the matrix in the middle is constant, once we have fixed a basis for V. This matrix therefore represents a bilinear form! We want to think about this matrix in a slightly different manner than usual, however, since we are used to matrices representing linear transformations, but here our matrix acts on a vector on both sides.

More abstractly, we can think of this in the following way: if we fix a basis for V, and construct the corresponding dual basis for V^*, then for $B \in \mathrm{Hom}(V^2; F)$ and $v, w \in V$, then B is of the form

$$B(v, w) = \lambda_v(T(w))$$

for some uniquely determined linear map $T : V \to V$.

Definition. Let $B \in \mathrm{Hom}(V^2; F)$. Then, B is said to be...

- ...**symmetric** if $B(v, w) = B(w, v)$ for all $v, w \in V$.
- ...**anti-symmetric** if $B(v, w) = -B(w, v)$ for all $v, w \in V$.
- ...**alternating** if $B(v, v) = 0$ for all $v \in V$.

These definitions should be familiar from when they were defined for multilinear maps, but since we can relate these to matrices, we have a useful way of checking these properties.

Theorem. Let $B \in \mathrm{Hom}(V^2; F)$. Then, B is symmetric (resp. anti-symmetric) if and only if the matrix representation of B is symmetric (resp. anti-symmetric).

Proof. This follows from the entries of the matrix representation being the evaluation of B on the basis vectors. □

We have an additional definition, which is something we might not have seen before.

Definition. Let $B \in \mathrm{Hom}(V^2; F)$. Then, B is said to be **nondegenerate** if $B(v, w) = 0$ for all $w \in V$ implies $v = 0$.

Let's first relate this to a matrix condition.

Theorem. Let $B \in \mathrm{Hom}(V^2; F)$. Then, B is nondegenerate if and only if the matrix associated to B is invertible.

Proof. Let A be the matrix associated to B.

Suppose B is nondegenerate. We have that

$$B(v, w) = v^T A w$$

and so fixing $v \in V$, we suppose that $v^T A$ as a map from V to F sends every w to 0. Then, by Rank-Nullity, we have that

$$\dim(V) = \dim(\ker(v^T A)) + \dim(\mathrm{im}(v^T A)) = \dim(V) + \dim(\mathrm{im}(v^T A))$$

and thus $\mathrm{im}(v^T A) = \{0\}$. Thus, $v^T A$ sends every vector to 0, and therefore is the 0 map. Since $0 = v^T A = (A^T v)^T$, we must have that $v \in \ker(A^T)$. Since B is nondegenerate, $v = 0$, so we have that $\ker(A^T) = \{0\}$. Thus, $0 \neq \det(A^T) = \det(A)$, as desired.

Suppose $\det(A) \neq 0$. Fix $v \in V$, and suppose that $B(v, w) = 0$ for all $w \in W$. Then, we have that $v^T A w = 0$ for all w, so $v^T A$ is the zero map from V to F. Since $0 \neq \det(A) = \det(A^T)$, we must have that A^T is invertible, so $0 = v^T A = (A^T v)^T$ implies that $v = 0$. □

The idea of the theorem is that we treat $v^T A$ as a linear map from V to F. We have a name for such maps! This is an element of the dual space, and we have a more formal way of thinking about this which we have seen before.

Theorem. Let $B \in \mathrm{Hom}(V^2; F)$. Then, for all $v \in V$, the map

$$\begin{aligned} B(v, -) : V &\to F \\ w &\mapsto B(v, w) \end{aligned}$$

is an element of V^*.

Proof. This follows from the fact that B is multilinear. □

We previously saw this in the context of currying, where we input one element into a k-linear form, and treat the rest as a $(k-1)$-linear form. However, when B is nondegenerate, we can actually say something more!

Theorem. Let $B \in \mathrm{Hom}(V^2; F)$. Then, B is nondegenerate if and only if the map

$$\begin{aligned} \varphi : V &\to V^* \\ v &\mapsto B(v, -) \end{aligned}$$

is an isomorphism of vector spaces.

Proof. Suppose B is nondegenerate. First note that the assignment $v \mapsto B(v, -)$ is linear since B is multilinear, and thus linear in the first component. Then, if $\varphi(v) = B(v, -)$ is the zero map, then by nondegeneracy of B, we must have that $v = 0$. Thus, φ is injective. Furthermore, since $\dim(V) = \dim(V^*)$, we must have that φ is an isomorphism, as desired.

Suppose that φ is an isomorphism. Then, φ is injective, so if $\varphi(v) = B(v, -) = 0$, then we must have $v = 0$, which implies that B is nondegenerate. □

The idea here is that nondegenerate bilinear forms give us an isomorphism between V and V^*. This doesn't necessarily require us to pick a basis to define, but it does require the additional information of a nondegenerate bilinear form, and different choices of this form will give different isomorphisms.

Our nondegenerate form is actually slightly stronger, since not only does it give us an isomorphism, but it can also provide us a basis for the dual! Recall that given a basis $\{v_1, \dots, v_n\}$ for V, we can construct a dual basis $\{\varepsilon_1, \dots, \varepsilon_n\}$ for V, with the property that

$$\varepsilon_i(v_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Given a basis $\{v_1, \dots, v_n\}$, we can construct a dual basis in a similar manner, by considering

$$B(v, w) = v^T A w$$

Then, from our previous dual basis construction, we'd like $v^T A = w^T$, so we can simply choose $v^T = (v_i')^T = w^T A^{-1}$, and the collection $\{v_1', \dots, v_n'\}$ forms a basis for V which is *dual* to $\{v_1, \dots, v_n\}$, in the sense that

$$B(v_i, v_j') = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and this construction creates a basis "dual" to our original, with respect to B.

We finish off this section with an interesting topic: suppose we have a nondegenerate bilinear form $B \in \mathrm{Hom}(V^2; F)$, and we examine

$$B(v, T(w))$$

for some linear map T. In terms of matrices, if A is the matrix which represents B, and M_T is the matrix which represents T, we have

$$B(v, M_T(w)) = v^T A(M_T w) = (v^T A M_T)w = v^T (A M_T A^{-1}) A w = ((M_T')^T v)^T = B(M_T'^T(v), w)$$

where A^{-1} exists as B is nondegenerate, so A is invertible. Note that M_T' represents T after a change of basis via A from M_T. In some sense, we've related performing T in the second component to performing a related map for T in the first component. This motivates us to make a definition.

Definition. Suppose $T : V \to V$ is a linear map, and $B \in \mathrm{Hom}(V^2; F)$ is a bilinear form. Then, if it exists, the map $T^\dagger : V \to V$ is said to be the **adjoint** of T if for all $v, w \in V$, we have that

$$B(v, T(w)) = B(T^\dagger(v), w)$$

Theorem. If B is nondegenerate, then adjoint maps always exist.

Proof. Fix a basis $\{v_1, \dots, v_n\}$ for V. Then, we want to define a linear map $T^\dagger : V \to V$, so it suffices to define it on the basis. We therefore need to find images for $T^\dagger(v_i)$ in V such that

$$B(T^\dagger(v_i), w) = B(v_i, T(w))$$

for all $w \in W$. However, since B is nondegenerate, the map $v \mapsto B(v, -)$ is an isomorphism, and so defining the map $B(v, -)$ uniquely determines v. Since we have defined

$$B(T^\dagger(v_i), -) : w \mapsto B(v_i, T(w))$$

we have completely defined the map $B(T^\dagger(v_i), -)$, so this uniquely determines a vector $T^\dagger(v_i)$ via our isomorphism. Thus, we have defined $T^\dagger$, which has the adjoint property by construction. □

However, we can also think about the adjoint map in a different way! Instead of thinking about the adjoint as the map which satisfies

$$B(v, T(w)) = B(T^\dagger(v), w)$$

we can instead think of the operation on the left as taking an element $B(v, -) \in V^*$, and precomposing with the map T. The map

$$\begin{aligned} T^* : V^* &\to V^* \\ \lambda &\mapsto \lambda \circ T \end{aligned}$$

is one we've encountered before: this is precisely the dual map! Thus, the composition of $B(v, -)$ with T is precisely the image of $B(v, -)$ under the dual map, which outputs an element of the dual. Composing this with the isomorphism (in reverse) of $v \mapsto B(v, -)$ precisely identifies each $B(w, -)$ with an element of V under a composition of linear maps, and so this map is precisely the adjoint! We can attempt to understand this through the following diagram.

$$\begin{array}{ccc} V & \xrightarrow{T} & V \\ {\scriptstyle v \mapsto B(v,-)}\downarrow & & \downarrow{\scriptstyle v \mapsto B(v,-)} \\ V^* & \xleftarrow[T^*]{} & V^* \end{array}$$

Starting with $w \in V$ on the top right, we follow the isomorphism downward to get $B(w, -)$, and then the dual map identifies with $B(w, -) \circ T$, which is precisely $B(w, T(-))$. However, this element of V^* must correspond to $B(w', -)$ for some $w' \in V$, and the choice of w' comes from the isomorphism on the left (in reverse). Thus, this entire map is linear, so we've defined $T^\dagger$ to satisfy precisely the properties we want. You can think of the adjoint as corresponding to the dual map, once we've used B to create an isomorphism from V to V^*.

15 Inner Product Spaces

Planned Lecture Date(s): August 2nd, 2023.

So far, our vector spaces have not had the structure to define a notion of length. Using the tools of bilinear forms, we will work towards a setting in which we can talk about "distance". The first thing to note is that distance is intrinsically a notion that requires order (we need a notion of greater than or less than), and so the most natural setting to work in is a field with an ordering. There are many ways we can define this, but the main takeaway is that we should use the field of real numbers, $\mathbb{R}$, as the field over which our distances are defined.

For now, we'll restrict to the case where $F = \mathbb{R}$. It turns out that much of the theory still goes through when $F = \mathbb{C}$, but we have to make some slight changes to keep the desired properties, so we'll address this later.

Definition. Let V be an $\mathbb{R}$-vector space. A **inner product** on V is a symmetric bilinear form on V, often denoted $\langle \cdot, \cdot \rangle$, satisfying the property that for all $v \in V$, $\langle v, v \rangle \geq 0$, and furthermore $\langle v, v \rangle = 0$ if and only if $v = 0$. The $\mathbb{R}$-vector space V, together with the inner product $\langle \cdot, \cdot \rangle$, is said to be a **inner product space**.

The property that an inner product satisfies is often called **positive definiteness**, and we think of the quantity $\langle v, v \rangle$ as the "length" of v (possibly squared), so it should be a non-negative quantity. Thus, having an inner product allows us to define a notion of length in V. Note that positive-definiteness requires a notion of inequality, which only makes sense in an ordered field.

The most common example of an inner product that we are likely familiar with is the Euclidean dot product, as given $v, w \in \mathbb{R}^n$, we can define

$$\vec{v} \cdot \vec{w} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^{n} a_i b_i$$

Recall from vector calculus that $\vec{v} \cdot \vec{w} = ||\vec{v}|| \cdot ||\vec{w}|| \cos(\theta)$, where θ is the angle between $\vec{v}$ and $\vec{w}$. We can think of this as a measure of alignment between v and w: if v and w are generally in the same direction, then θ should be small, so the dot product comes close to attaining it's maximum possible value, $||\vec{v}|| \cdot ||\vec{w}||$. Similarly, if v and w point in very different directions, the dot product will be close to 0. This is the intuition we want to keep when dealing with general inner products: the inner product $\langle v, w \rangle$ is a measure of how closely v and w "align", as determined by this inner product.

Definition. Let V be an inner product space. Two vectors $v, w \in V$ are said to be **orthogonal** if $\langle v, w \rangle = 0$. A set of vectors is said to be orthogonal if they are pairwise orthogonal.

Definition. Let V be an inner product space, and let $W \subset V$. Then, we can define the **orthogonal complement** of W to be the space

$$W^{\perp} = \{v \in V \mid \langle v, w \rangle = 0 \ \forall \ w \in W\}$$

It turns out that $W^{\perp}$ is indeed a subspace (with some interesting properties!), which you will verify on homework.

Intuitively, orthogonal vectors are vectors which are maximally distinct, and align with each other as little as possible, so there is minimal redundant information carried by the vectors. As such, it may be a good idea to find a basis which is orthogonal.

Definition. Let V be an inner product space. Then, an **orthogonal basis** is a basis for V which is (pairwise) orthogonal. Furthermore, if $\{v_1, \ldots, v_n\}$ is a basis for V, then it is said to be an **orthonormal basis** if

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

Note that an orthonormal basis is an orthogonal basis which additional requires each basis vector to have "length" equal to 1. Given an orthogonal basis, we can simply scale each v_i by $\frac{1}{\sqrt{\langle v_i, v_i \rangle}}$ to normalize v_i to have length 1.

A very convenient trick is the following: if $\{v_1, \ldots, v_n\}$ is an orthonormal basis for V, then we can write any $v \in V$ as

$$v = c_1 v_1 + \cdots + c_n v_n$$

If V were simply a vector space without an inner product, it would be fairly difficult to find the coefficients c_i, but using our inner product, we see that

$$\begin{aligned} \langle v_i, v \rangle &= \langle v_i, c_1 v_1 + \cdots + c_n v_n \rangle \\ &= \sum_{j=1}^{n} c_j \langle v_i, v_j \rangle \\ &= c_i \end{aligned}$$

and therefore in order to recover the coefficients, we can simply use the inner product to "measure" them out.

The question to ask is then: how can we construct an orthonormal basis? Let's develop a tool to create them.

Definition. Let V be an inner product space, and let $v, w \in V$. Then, we define the projection of v onto w as

$$\text{proj}_w(v) = \frac{\langle v, w \rangle}{\langle w, w \rangle} w$$

Note that this is an element of $\text{span}(w)$.

We want to think of this as the shadow of v onto $\text{span}(w)$, as the inner product $\langle v, w \rangle$ tells us how much v aligns with w, and the vector $\frac{1}{\langle w, w \rangle} w$ is a vector of length 1 in $\text{span}(w)$. Using these projections, we will create a process that constructs orthonormal bases.

Theorem. Let V be an inner product space, and let $\{v_1, \ldots, v_n\}$ be any basis for V. Perform the following process:

- Set $w_1 = v_1$.
- For $1 \leq k \leq n$ (sequentially), let $w_k = v_k - \sum_{i=1}^{k-1} \text{proj}_{w_i}(v_k)$.
- Set $u_k = \frac{1}{\sqrt{\langle w_k, w_k \rangle}} w_k$.

Then, the set $\{u_1, \ldots, u_n\}$ is an orthonormal basis for V. This process is known as the **Gram-Schmidt Orthonormalization Process**.

Proof. We first show that the w_k, as constructed, form an orthogonal basis. We proceed by induction on k, proving the statement that the set $\{w_1, \ldots, w_k\}$ are (pairwise) orthogonal. When $k = 1$, this is vacuously true. Suppose this holds for $k - 1$. We show that w_k is orthogonal with w_j for $j < k$. We compute

$$\begin{aligned} \langle w_k, w_j \rangle &= \left\langle v_k - \sum_{i=1}^{k-1} \text{proj}_{w_i}(v_k), w_j \right\rangle \\ &= \langle v_k, w_j \rangle - \sum_{i=1}^{k-1} \left\langle \text{proj}_{w_i}(v_k), w_j \right\rangle \end{aligned}$$

$$= \langle v_k, w_j \rangle - \sum_{i=1}^{k-1} \left\langle \frac{\langle v_k, w_i \rangle}{\langle w_i, w_i \rangle} w_i, w_j \right\rangle$$
$$= \langle v_k, w_j \rangle - \sum_{i=1}^{k-1} \frac{\langle v_k, w_i \rangle}{\langle w_i, w_i \rangle} \langle w_i, w_j \rangle$$

By the inductive hypothesis, we have $1 \leq i, j \leq k-1$, so w_i and w_j are orthogonal when $i \neq j$, so we have

$$= \langle v_k, w_j \rangle - \frac{\langle w_i, w_i \rangle}{\langle w_i, w_i \rangle} \langle v_k, w_j \rangle$$
$$= \langle v_k, w_j \rangle - \langle v_k, w_j \rangle$$
$$= 0$$

We therefore conclude that w_k is orthogonal with each w_j for $1 \leq j < k$. Thus, the set $\{w_1, \ldots, w_n\}$ is orthogonal. By construction, the set $\{u_1, \ldots, u_n\}$ is then orthonormal, as desired. □

The strength of Gram-Schmidt is that given any basis, we can produce an orthonormal basis which "agrees" with our original basis as much as possible. We can think of the process as iteratively "throwing out" the parts of each basis vector which are not orthogonal, and thus we are left with only the orthogonal parts.

Although we have our standard Euclidean inner product, we can actually think about alternative inner products. Here are some examples.

Example. Let $V = \mathbb{R}^2$, and consider...

- the candidate inner product
$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = 2x_1y_1 + 2x_2y_2$$
- the candidate inner product
$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = 2x_1y_1 - x_1y_2 - x_2y_1 + 3x_2y_2$$
- the candidate inner product
$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = 4x_1y_1 - 3x_1y_2 - 3x_2y_1 + 2x_2y_2$$

Which of these are actually inner products?

The answer is a bit tricky: we can easily check that they're all symmetric and bilinear, but checking positive definiteness is harder. If we write these symmetric bilinear forms as matrices, we see that they are represented by the matrices

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}, \begin{pmatrix} 4 & -3 \\ -3 & 2 \end{pmatrix}$$

and we can rephrase the question using the following definition.

Definition. A matrix M is **positive definite** if $v^T M v > 0$ for all $v \neq 0$.

We can then ask which of these matrices are positive definite. This can be checked using an equivalent condition.

Theorem. Let M be a symmetric $n \times n$ matrix. Then, M is positive definite if and only if all eigenvalues of M are real and positive.

Proof. Homework. □

We see that the first matrix has eigenvalue 2 (with multiplicity 2), which is positive, the second matrix has eigenvalues $\frac{1}{2}\left(5 \pm \sqrt{5}\right)$ which are both positive, and the third matrix has eigenvalues $3 \pm \sqrt{10}$, which are not all positive. Thus, the first two define inner products, whereas the third does not.

When the underlying vector space is more complicated, we can also have other inner products.

Example. Let $V = C([0,1])$, the space of continuous functions from $[0,1]$ to $\mathbb{R}$. Then, we can define an inner product

$$\langle f, g \rangle = \int_0^1 f(t)g(t)\,dt$$

One can check that this does indeed satisfy the inner product axioms. For example, we can construct an orthogonal basis for this space relative to this inner product!

We also make the following observation.

Theorem. Any inner product is a nondegenerate form.

Proof. Suppose $\langle v, w \rangle = 0$ for all $w \in W$. Then, when $v = w$, we have that $\langle w, w \rangle = 0$, so $v = w = 0$, as desired. □

This tells us that given an inner product space V, we have a canonical isomorphism between V and V^*. This does require the additional structure of an inner product, but does not require a basis!

We have a theorem which we may have previously seen in a vector calculus class, but generalizes to the setting of inner products.

Theorem. (Cauchy-Schwarz) Let V be an inner product space, and let $v, w \in V$. Then, we have that

$$\langle v, w \rangle^2 \leq \langle v, v \rangle \cdot \langle w, w \rangle$$

Furthermore, we have equality if and only if v and w are linearly dependent.

Proof. If $v = 0$ or $w = 0$, both sides of the inequality are 0, so this statement holds. Otherwise, we show the equivalent statement that

$$\frac{\langle v, w \rangle^2}{\langle v, v \rangle \cdot \langle w, w \rangle} \leq 1$$

We have that

$$\frac{\langle v, w \rangle^2}{\langle v, v \rangle \cdot \langle w, w \rangle} = \left\langle \frac{v}{\sqrt{\langle v, v \rangle}}, \frac{w}{\sqrt{\langle w, w \rangle}} \right\rangle$$

and so it suffices to show that for unit vectors $v, w \in V$, we have that

$$\langle v, w \rangle \leq 1$$

We have that

$$\begin{aligned} 0 &\leq \langle v - w, v - w \rangle \\ &= \langle v, v \rangle - 2\langle v, w \rangle + \langle w, w \rangle \\ &= 2 - 2\langle v, w \rangle \\ &= 2(1 - \langle v, w \rangle) \end{aligned}$$

and so we must have $\langle v, w \rangle \leq 1$, as desired.

If v and w are dependent, and either $v = 0$ or $w = 0$, then equality holds. Otherwise, without loss of generality, we have that $v = cw$ for some $c \in F$, and so we have

$$c^2 \langle v, w \rangle^2 = \langle v, cw \rangle^2 \leq \langle v, v \rangle \langle cv, cv \rangle = c^2 \langle v, v \rangle^2$$

Similarly, if we have equality, then we can reduce to the unit vector case, where $\langle v, w\rangle^2 = 1$. Then, $\langle v, w\rangle = \pm 1$, so we have that

$$\langle v \mp w, v \mp w\rangle = 2(1 \mp \langle v, w\rangle)$$

and so by positive definiteness, $v \mp w = 0$, which provides a linear dependence. □

Finally, as promised, we discuss what happens if we replace $\mathbb{R}$ with $\mathbb{C}$ instead. Since we want to associate $\langle v, v\rangle$ to the length of the vector v, we need to ensure that however we define it, $\langle v, v\rangle$ will still be a real number (note that it doesn't make sense to talk about $\langle v, v\rangle > 0$ if it is not real). It turns out that the correct modifications are the following:

Definition. If V is a $\mathbb{C}$-vector space, an **inner product** on $\mathbb{C}$ is a map $\langle \cdot, \cdot\rangle : V \times V \to \mathbb{C}$ satisfying

1. Conjugate Symmetry: $\langle v, w\rangle = \overline{\langle w, v\rangle}$.
2. Linearity in the first argument: $\langle cv_1 + v_2, w\rangle = c\langle v_1, w\rangle + \langle v_2, w\rangle$.
3. Positive-definiteness: $\langle v, v\rangle \geq 0$, and $\langle v, v\rangle = 0$ if and only if $v = 0$.

A few things to note: firstly, the convention of linearity in the first argument rather than the second is a choice. Mathematicians frequently make this choice, whereas physicists tend to make the choice of linearity in the second component. Note however that the inner product is *almost* linear in the second component, as we can use conjugate symmetry to swap to the first component, take out a $\mathbb{C}$-scalar, and conjugate back (also conjugating the scalar). People will often call this property **sesquilinearity**, as the prefix "sesqui-" means "one and a half", as the inner product is linear in one component and "half"-linear in the other. Finally, note that conjugate symmetry implies that $\langle v, v\rangle = \overline{\langle v, v\rangle}$, so $\langle v, v\rangle$ is real, so it makes sense to require it to be positive.

One reason we need to deal with all this complex conjugation is to ensure that our lengths stay real: without this, if we choose $w = iv$, then we have that $\langle w, w\rangle = \langle iv, iv\rangle = -\langle v, v\rangle$, which is problematic, since we'd like both of these to be positive.

16 The Spectral Theorem for Self-Adjoint Operators

Planned Lecture Date(s): August 4th, 2023.

Let's return to our discussion of adjoint maps for a moment, specifically in the context of the inner product. If we consider the standard inner product on $\mathbb{R}^n$ as

$$\langle v, w \rangle = v^T w$$

we can consider some matrix $A : \mathbb{R}^n \to \mathbb{R}^n$, and examine

$$\langle v, A(w) \rangle = v^T(Aw) = (v^T A)w = (A^T v)^T w = \left\langle A^T(v), w \right\rangle$$

and we see that A^T indeed represents the adjoint of A. However, when we pass to the case where our base field is $\mathbb{C}$, we need to make some adjustments to the standard inner product, specifically

$$\langle v, w \rangle = v^* w$$

where v^* denotes the conjugate transpose of v, in that we take the transpose, and also complex conjugate each entry. One can check that this construction does indeed satisfy sesquilinearity and conjugate symmetry, as well as positive definiteness, so this does indeed work as an inner product. However, in this new setting, our adjoint map also changes. Given some matrix A, we can compute

$$\langle v, A(w) \rangle = v^*(Aw) = (v^* A)w = (A^* v)^* w = \langle A^*(v), w \rangle$$

and we see that the adjoint is instead A^*, the conjugate transpose of A. This is sometimes called the **Hermitian** adjoint of A.

Definition. Let V be an inner product space, and let $T : V \to V$ be a linear map. T is said to be **self-adjoint** if $T^\dagger = T$.

Note that since the definition of the adjoint map relies upon the choice of inner product, this definition also depends on which inner product is chosen. The reason that self-adjoint maps are nice is the following major theorem.

Theorem. (Spectral Theorem) Let V be a finite-dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$, and let $T : V \to V$ be self-adjoint. Then, we have that:

- Every eigenvalue of T is real.
- T is diagonalizable.

Proof. We first prove that every eigenvalue of T is real. If V is over $\mathbb{R}$, this statement is automatically true, so assume V is over $\mathbb{C}$. Suppose λ is an eigenvalue of T, and let v be an associated eigenvector. Then, we have that

$$\lambda \langle v, v \rangle = \langle v, \lambda v \rangle = \langle v, T(v) \rangle = \langle T(v), v \rangle = \langle \lambda v, v \rangle = \overline{\lambda} \langle v, v \rangle$$

and since v is an eigenvector, $\langle v, v \rangle \neq 0$, so $\lambda = \overline{\lambda}$ and thus $\lambda \in \mathbb{R}$, as desired.

We then proceed by induction (on the variable n) on the following statement: when $\dim(V) = n$, we can decompose V into

$$V = \bigoplus_{i=1}^{n} V_i$$

where each V_i is a dimension 1 invariant subspace of V under T. Note that this statement implies that T is diagonalizable, since choosing one nonzero vector from each V_i forms an eigenbasis for T.

When $\dim(V) = 0$, V is the empty direct sum of eigenspaces, so we have established the base case.

Suppose this holds for vector spaces of dimension $n-1$. We would like to show that this holds when $\dim(V) = n$. Choose an eigenvalue λ of T, and associated eigenvector v. Then, let $W = \operatorname{span}(v)$, and consider

$$V = W \oplus W^{\perp}$$

where the sum is direct using a result from homework. We show that $W^{\perp}$ is an invariant subspace under T. Fix $w \in W^{\perp}$. We have that

$$\langle v, T(w)\rangle = \langle T(v), w\rangle = \langle \lambda v, w\rangle = \lambda \langle v, w\rangle = 0$$

since $\langle v, w\rangle = 0$, and v is an eigenvector of T. Since T is self-adjoint, it restricts to a self-adjoint operator on $W^{\perp}$ (think about why!), and so we can apply the inductive hypothesis to $W^{\perp}$, since $\dim(W^{\perp}) = n-1$. Then, relabelling W as V_n, we have that

$$V = W \oplus W^{\perp} = V_n \oplus W^{\perp} = V_n \oplus \left(\bigoplus_{i=1}^{n-1} V_i\right)$$

which gives our desired result. □

Corollary. Let V be a finite-dimensional inner product space, and $T : V \to V$ a self-adjoint map. Then, eigenvectors associated with distinct eigenvalues are orthogonal.

Proof. From the construction, we see that each V_i is orthogonal, and so the inner product of vectors contained purely in the direct sum of different V_i must be 0. We can also see this directly: Let v_1 and v_2 be eigenvectors associated to eigenvalues λ_1 and λ_2, respectively. Then, we have that

$$\begin{aligned}
\lambda_1 \langle v_1, v_2\rangle &= \langle \lambda_1 v_1, v_2\rangle \\
&= \langle T(v_1), v_2\rangle \\
&= \langle v_1, T(v_2)\rangle \\
&= \langle v_1, \lambda_2 v_2\rangle \\
&= \lambda_2 \langle v_1, v_2\rangle \\
(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle &= 0
\end{aligned}$$

and so if the eigenvalues are distinct, we must have $\langle v_1, v_2\rangle = 0$, as desired. Note that no complex conjugation is needed, since all eigenvalues are real. □

The Spectral Theorem generalizes to the case when V is not finite-dimensional, and is an incredibly important result in the field of functional analysis. However, the proofs get much more difficult in this case, so we will not explore this further.

Example. One application of the Spectral Theorem is in multivariable calculus, when one defines the Hessian matrix of second derivatives of a function $f(x_1, \ldots, x_n)$ as

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & & \\ \vdots & & \ddots & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & & & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

Clairaut's Theorem tells us that mixed partial derivatives commute, which means that this matrix is symmetric! Thus, this matrix is diagonalizable, and (at any point) the eigenvectors correspond to the principal axes of the surface which best approximates f near that point, and the eigenvalues represent the second derivative of the curves in those directions.

Example. For a vector space V, a quadratic form $q : V \to F$ is a (not necessarily linear!) map such that there exists a symmetric bilinear form B with

$$q(v) = B(v, v) \ \ \forall \ v \in V$$

Thus, we can write $q(v) = v^T M v$ for some matrix M, and since B is symmetric, M is diagonalizable. Thus, we can separate q into separate components. If $V = \mathbb{R}^n$, then for some choice of coordinates, we can write

$$q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i x_i^2$$

where the λ_i are the eigenvalues of M.

Example. In probability theory, if one has a set of random variables $\{x_1, \ldots, x_n\}$, then we can define a covariance matrix

$$M = \begin{pmatrix} \text{Var}(x_1) & \text{Covar}(x_1, x_2) & \cdots & \text{Covar}(x_1, x_n) \\ \text{Covar}(x_2, x_1) & \text{Var}(x_2) & & \\ \vdots & & \ddots & \\ \text{Covar}(x_n, x_1) & & & \text{Var}(x_n) \end{pmatrix}$$

and since covariance is symmetric, this matrix is symmetric. Thus, we can find a change of coordinates in which this matrix is diagonalizable, and thus making our new variables independent since their covariance is 0.

Example. In physics, one frequently sees the equation

$$E = \frac{1}{2} I \omega^2$$

where E denotes energy, I denotes the moment of inertia, and ω denotes angular velocity. However, you may have also seen that in the rotational setting, many of our quantities are vectors, such as angular position or velocity, whose direction points in the axis of rotation (taken with right hand rule) and magnitude represents the desired scalar quantity. This doesn't quite make sense! How can I square the vector ω? It turns out that the moment of inertia is actually properly interpreted as a matrix, and the equation should really read

$$E = \frac{1}{2} \omega^T I \omega$$

Note that physicists often call this the **inertia tensor**, since it is essentially the bilinear form associated to I evaluated on (ω, ω). The moment of inertia matrix in three dimensions is often written

$$I = \begin{pmatrix} \int y^2 + z^2 \, dm & -\int xy \, dm & -\int xz \, dm \\ -\int xy \, dm & \int x^2 + z^2 \, dm & -\int yz \, dm \\ -\int xz \, dm & -\int yz \, dm & \int x^2 + y^2 \, dm \end{pmatrix}$$

where the diagonal entries are the standard moments of inertia about the usual coordinate axes, and the off-diagonal terms are the products of inertia. One can think of the off-diagonal entries as mixing between the coordinate axes, so an object spinning around one axis will also try to spin around the others. However, since this matrix is symmetric, it is diagonalizable, and the eigenvectors which diagonalize it correspond to the principal axes of rotation, and the eigenvalues correspond to the moments of inertia around these axes. These axes are called principal, since the off-diagonal products of inertia are 0, so the object rotating around these axes will not try to spin around the others.