abbreviated course schedule

1. **Vectors.** Vector spaces, subspaces, linear dependence, dimension, reduction, finding bases.

2. **Maps.** Systems of linear equations, linear mappings, action on bases, action on dimension, complimentary subspace mapped surjectively (bijectively), range/nullspace+rank/nullity, range versus columnspace of matrices, more row reduction.

3. **Maps with equal dimension; invertibility.** Invertibility, preservation of linear (in)dependence, change of basis, similarity, elementary row ops, determinants, equivalent conditions, methods of computation, cramer’s rule.

4. **Solidification and applications.** Geometry, computer graphics, and some combinatorics (graphs as matrices, Markov chains).

5. **Eigenstuff.** Stable subspaces, picture of eigenvecs, picture of diag’ability, finding with det, characteristic poly (det/tr), non-diag’ablity, Jordan form, positive/negative definiteness.

6. **Solidification and applications.** Linear systems of differential equations and other real-world examples (e.g. circuits).

7. **Inner products.** Dot product, transpose, orthogonality/normality, ortho projections, Gram-Schmidt, orthogonal transforms, symmetric maps, spectral theorem.

8. **Least squares, singular value decomposition** Several concrete uses.
1 Vector Spaces

1.1 Spaces, Subspaces

In highschool mathematics, and even in introductory college calculus courses,

You should already be familiar with the notions of “scalars” - that is, real or complex numbers, typically representing an amount - and “vectors”, which are said to capture both a size and a direction, typically in two or three dimensional space. One of the major concerns in the study of linear algebra is the generalization of “direction” to larger numbers of dimensions.

People typically perceive the space around them to be Euclidean 3-space, and base their intuition of vector spaces on this model.

Let’s state more precisely what we mean by $\mathbb{R}^n$. We define $\mathbb{R}^n$ to be the set of ordered lists of real numbers with length $n$. We will call each such ordered list an ordered $n$-tuple, or a Euclidean vector (of length $n$). We will use several simple notations to refer to particular vectors, some which explicitly refer to the entries in the list (which we will call the components of the vector):

$$x = \bar{x} = \mathbf{x} = (x_i) = (x_1, x_2, \ldots, x_n) = [x_1 \ x_2 \ \cdots \ x_n] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

As you know from your experience with $\mathbb{R}^2$ and $\mathbb{R}^3$, there are many features shared by each $\mathbb{R}^n$:

- Each contains a zero vector: $0 = (0, \ldots, 0)$.
- Any two vectors may be “added” together.
- Any vector may be “multiplied” by a scalar.
- The standard properties of arithmetic manipulation of scalars (associativity of addition and multiplication, commutativity of addition and multiplication and distribution of multiplication over addition) hold for these operations.

We also think of each (nonzero) vector in $\mathbb{R}^n$ as pointing in a direction out of the origin, and all stretches of the vector also pointing in that direction. However, we have not written down anything precise which captures this idea.

Linear algebra is concerned with the study of generalizations of $\mathbb{R}^n$—that is, abstract objects which share the above properties of some $\mathbb{R}^n$. In linear algebra such an object is called a vector space.

When we try to write down a formal definition of what we see above, we get stuck on the following problem: What would a general notion of “direction” be? We see that in $\mathbb{R}^n$, every vector determines a direction, namely that direction that it points in. But it is actually fairly meaningless to say anything more. So let’s define a vector space to be something which behaves a lot like $\mathbb{R}^n$, and then say that two vectors in this space have the same direction if
they are scalar multiples of each other. We will return to the question of comparing different
directions later in the course.

The properties above lead to the following definition. the axioms we lay down here are
merely more precise versions of these properties. you do not need to memorize this defini-
tion, so long as you understand its important features well enough the say when something
“should” be a vector space.

A vector space is a set \( V \), whose elements we call vectors, along with two operations called
addition and scalar multiplication, which satisfy the following axioms.

- the addition operation, written \(+\), takes two vectors \( \vec{x} \) and \( \vec{y} \), and gives you another
  one called \( \vec{x} + \vec{y} \). this implicitly means that \( V \) is closed under \(+\)!

- addition satisfies the normal rules of associativity and commutativity. namely, \((\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})\) and \(\vec{x} + \vec{y} = \vec{y} + \vec{x}\).

- there is a unique element of \( V \), called the zero element and written \( 0 \) or \( \vec{0} \) or \( 0 \), such
  that \( 0 + \vec{x} = \vec{x} + 0 = \vec{x} \) for all vectors \( \vec{x} \). also, for every vector \( \vec{x} \) there is a unique
  vector, called the negative of \( \vec{x} \) and written \( -\vec{x} \), such that \( \vec{x} + (-\vec{x}) = \vec{0} \).

- the scalar multiplication operation, written \( \cdot \) or just by juxtaposition, takes a scalar \( a \)
  and a vector \( \vec{x} \) and gives you a vector called \( a \cdot \vec{x} \) or \( a\vec{x} \). again, this implicitly means
  that \( V \) is closed under multiplication by scalars!

- scalar mult satisfies usual rules of manipulation as well: \( a(b\vec{x}) = (ab)\vec{x} \), \( (a + b)\vec{x} = a\vec{x} + b\vec{x} \), \( a(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y} \), \( 0\vec{x} = \vec{0} \), \( 1\vec{x} = \vec{x} \), \( (-1)\vec{x} = -\vec{x} \).

Clearly, each \( \mathbb{R}^n \) is a vector space. what about other examples? The easiest way to get
new vector spaces is to find them inside already known vector spaces. This leads us to the
topic of subspaces, which we will explore next.

Suppose \( V \) is a vector space, and suppose that \( W \) is a collection of vectors in \( V \). We say
that \( W \) is a subspace of \( V \) if, roughly speaking, it is a vector space itself under the same
operations as \( V \). More precisely, we assume that

- if \( \vec{u}, \vec{v} \) are any two vectors in \( W \), then \( \vec{u} + \vec{v} \) also belongs to \( W \), and

- if \( \vec{u} \) is any vector in \( W \) and \( a \) is any scalar, then \( a\vec{u} \) is in \( W \).

It can be shown that, granted these two assumptions, \( W \) inherits all the other vector space
axioms from \( V \) automatically. For example, we get that \( 0 \) is in \( W \).

Let’s do some examples/exercises right now, in order to get a feel for what’s going on.
Take our first \( V = \mathbb{R}^2 \). Here are some vector spaces inside of it:

- The zero-space \( W = \{0\} \). We will often write \( 0 \) for this set.

- The \( x \)-axis, \( W = \{(x,0) \mid \text{all scalars } x \} \), and likewise the \( y \)-axis.

- For any fixed vector \( \vec{v} = (x,y) \), the set \( W = V(\vec{v}) = \{a(x,y) \mid \text{all scalars } a \} = \{(ax,ay) \mid \text{all scalars } a \} \). This includes the two axis examples, and the zero-example
  (take \( (x,y) = (0,0) \)). This is the also the easiest example of a construction we will
be seeing often: we call \( V(\vec{v}) \) the span of \( \vec{v} \) in \( \mathbb{R}^2 \). It is the smallest subspace of \( \mathbb{R}^2 \) containing \( \vec{v} \). More on this, in full detail, to follow. When \( \vec{v} \neq \vec{0} \), notice that this set takes the form of a line through \( \vec{v} \) and the origin.

- For any two fixed vectors \( \vec{u} \) and \( \vec{v} \), the set \( W = V(\vec{u}, \vec{v}) = \{a\vec{u} + b\vec{v} \mid \text{all scalars } a, b\} \). Notice that by taking \( \vec{u} = \vec{0} \) and possibly \( \vec{v} = \vec{0} \), we recover our other examples. Another case in which this degenerates this into a prior example is when \( \vec{u} \) and \( \vec{v} \) point in the same direction, whence \( V(\vec{u}, \vec{v}) = V(\vec{v}) \) as above. However, when this is not the case, \( V(\vec{u}, \vec{v}) \) is much bigger than \( V(\vec{v}) \); in fact, it is all of \( \mathbb{R}^2 \)!

- We will come to understand later in this chapter that these are all the vector spaces inside of \( \mathbb{R}^2 \): the zero space (the single point set consisting of the origin), a line through the origin, and the whole plane.

Our second big example is \( V = \mathbb{R}^3 \). The results are completely analogous. Each vector spaces inside of it is one of the following items:

- The zero-space \( \{0\} \), which is the set consisting of the single point, the origin.

- The set of multiples of a vector \( \vec{v} \neq \vec{0} \), which takes the form of a line through the origin and \( \vec{v} \).

- The set of sums-of-scalar-multiples \( a\vec{u} + b\vec{v} \), for all \( a, b \in \mathbb{R} \), where \( \vec{u} \) and \( \vec{v} \) are not proportional. This takes the form of the plane through the origin, containing the points \( \vec{u} \) and \( \vec{v} \).

- The whole space \( \mathbb{R}^3 \).

Let us turn now to give examples of vector spaces and subspaces in general. The collection \( M_{n \times n} \) of all \( n \times n \) matrices is a vector space.

Let \( p(x) \) denote a polynomial in the variable \( x \). Then the set

\[
P_n = \{\text{polynomials } p(x) \mid \deg p(x) < n\}
\]

of polynomials of degree less than \( n \) forms a vector space under addition of polynomials, and scalar multiplication via multiplication-by-constants. One can take the union of the \( P_n \), and get

\[
P = \{\text{all polynomials } p(x)\},
\]

which is a rather large vector space.

We can adjust the last space to get something even larger. Recall that a power series is an expression of the form \( p(x) = \sum_{n=0}^{\infty} a_n x^n \), with possibly infinitely many nonzero coefficients. Then taking

\[
S = \{\text{all power series } p(x)\},
\]

one gets an even bigger vector space.

You shouldn’t have thought of this, but here is another example. Let \( S \) be a set with \( n \) elements, say \( S = \{a_1, a_2, \ldots, a_n\} \), and define \( V \) to be the collection of functions \( f: S \to \mathbb{R} \). For any \( f, g \in V \) define \( f + g \) to be the function satisfying \( [f + g](a_i) = f(a_i) + g(a_i) \) for all
Let $\vec{v}$ be a vector in $V$. Then span($\vec{v}$), written $\text{span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k)$ or $\text{span}\{\vec{v}_i\}_{i=1}^k$, be the set of all linear combinations of these vectors. Thus a vector $\vec{u}$ in $V$ is a member of $\text{span}\{\vec{v}_i\}$ if and only if $\vec{u}$ can be written in the form $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_k\vec{v}_k = \sum_{i=1}^{k} a_i\vec{v}_i$ \hfill ($\ast$)

where the $a_i$ are scalars.

Next we define the span of the $\vec{v}_i$, written $\text{span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k)$ or $\text{span}\{\vec{v}_i\}_{i=1}^k$, be the set of all linear combinations of these vectors. Thus a vector $\vec{u}$ in $V$ is a member of $\text{span}\{\vec{v}_i\}$ if and only if $\vec{u}$ can be written in the form $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_k\vec{v}_k = \sum_{i=1}^{k} a_i\vec{v}_i$ \hfill ($\ast$).

The collection span$(S)$ is always a subspace of $V$, since it is defined to be closed under the operations. One can show that it is the minimal subspace of $V$ containing $S$. One way of interpreting this is that it is a subspace of $V$, and every subspace $W'$ of $V$ which contains $S$ must contain span$(S)$ as well. Another interpretation of this is that span$(S)$ is the intersection of all subspaces of $V$ which contain $S$.

If $W$ is a subspace of $V$ and $S$ is a subset of $V$ for which span$(S) = W$, then we call $S$ a spanning set for $W$.

Now let $W$ be any subspace of $V$. Since $W$ is closed under $+$ and $\cdot$, any linear combination in $W$ has to belong to $W$. Moreover, any $\vec{w}$ is a linear combination in $\vec{w}$, and so one has that span$(W) = W$. Therefore, every subspace can be obtained as the span of something. Of course, this choice of spanning set is far from optimal, but it does prove the point that spans are useful at constructing subspaces.

\[ \text{The fact that we had a finite list was not in any way important. We could have let } S \text{ be any arbitrary subset of } V, \text{ and define } \vec{u} \text{ to be a “linear combination” in } S \text{ if there exist } \vec{v}_i \in S \text{ and scalars } a_i \text{ for which } \vec{u} = \sum_{i=1}^{k} a_i\vec{v}_i. \text{ Then span}(S) \text{ is the set of all linear combinations in } S. \]
Moreover, spans can be useful tools for explicit calculation. Once you find a spanning set for $W$, you can write down what all its vectors are. Let’s look at some examples to make this more clear.

Take $V = \mathbb{R}^3$, and $\vec{v}_1 = (1, 0, 0), \vec{v}_2 = (0, 1, 0), \vec{v}_3 = (1, 1, 0)$. What is $W = \text{span}\{\vec{v}_i\}$? The thing to do here is to literally look at the definition we were given. An element of the span is just a linear combination. So an element of $W$ is something of the form $a_1 \vec{v}_1 + a_2 \vec{v}_2 + a_3 \vec{v}_3$. Let’s simplify this some:

$$a_1 \vec{v}_1 + a_2 \vec{v}_2 + a_3 \vec{v}_3 = a_1 (1, 0, 0) + a_2 (0, 1, 0) + a_3 (1, 1, 0) = (a_1, 0, 0) + (0, a_2, 0) + (a_3, a_3, 0) = (a_1 + a_3, a_2 + a_3, 0).$$

This is a general form for an element of $W$. However, there is some redundancy, because we can always absorb $a_3$ into $a_1$ and $a_2 \ldots$ so the simplest general form of an element of $W$ is

$$(b_1, b_2, 0).$$

Writing $b_1 = x$ and $b_2 = y$, we have found that $W$ is the set $\{(x, y, 0) \mid x, y \in \mathbb{R}\}$. Thus $W$ is merely the $xy$-plane in 3-space.

This is not the only way to write $W$ in terms of two parameters. Notice also that all the following general elements in two variables describe $W$ (to derive each, use an analogous “absorbtion” argument):

$$(a_1 + a_3, a_3, 0), \quad (a_3, a_2 + a_3, 0), \quad (5a_1, 7a_2, 0), \quad (3a_1 + 8a_3, 8a_3, 0).$$

And there are infinitely more we can come up with. Basically, we have some freedom which comes from the existence of the relation/constraint

$$\vec{v}_1 + \vec{v}_2 - \vec{v}_3 = 0.$$

the existence of a nontrivial relation is called a linear dependence, and will be our next topic of discussion.

1.2 Linear Dependence, Bases

Let $V$ be a vector space and let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ be vectors in $V$. We say that the $\vec{v}_i$ are linearly dependent if we can find nontrivial a linear combination of them which vanishes, that is, if we can obtain

$$a_1 \vec{v}_1 + a_2 \vec{v}_2 + \cdots + a_k \vec{v}_k = \vec{0}$$

with not all $a_i$ equal to zero. (If we were to allow all $a_i = 0$, then we would have a silly definition, because ALL vectors would be linearly dependent!)

Here is another way of defining it. By a representation of zero, or an equation of linear dependence, we mean an equation of the form immediately above,

$$a_1 \vec{v}_1 + \ldots + a_k \vec{v}_k = \vec{0}.$$
such an equation is called \textit{trivial} if all the \( a_i = 0 \), and \textit{nontrivial} otherwise. There is always at least one representation of zero, namely

\[ 0 = 0. \]

Any \textit{trivial} equation of linear dependence expresses merely this. A \textit{nontrivial} equation of linear dependence expresses something more: it tells you that some collection of nonzero vectors are interrelated. It also gives you an alternate way of writing \( 0 \), aside from the trivial representation above. Therefore we may also think of a nontrivial equation of linear dependence as giving you a \textit{redundant} way of writing the vector \( \vec{0} \).

This second viewpoint justifies some conventions that are common when dealing with linear dependence in the study of linear algebra. We really want linear dependence to express something nontrivial. We want it to give us a truly redundant representation of zero. Otherwise, there is no content. If we are given a list of vectors \( \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \), and one of them is zero, say \( \vec{v}_j = 0 \), then this list is \textit{always} linearly dependent, since

\[ 1 \cdot \vec{0} + \sum_{i \neq j} 0 \vec{v}_i = 0. \]

So saying that some collection of vectors, including the zero vector, is linearly dependent, gives trivial representations of zero always. To ask if there are \textit{nontrivial} representations of zero, one must remove \( \vec{v}_j \) from the list and start over. Therefore, we adopt the following.\textbf{Convention:} when we are told that a list of vectors is linearly dependent, we will generally assume that \( 0 \) is not in this list, or otherwise we throw that vector away.

The second convention is just an expansion of the first one. Suppose our list of vectors has a certain element listed (at least) twice: \( \vec{v}_j = \vec{v}_l \). Then again, the vectors are automatically linearly dependent:

\[ 1 \vec{v}_j + (-1)\vec{v}_l + \sum_{i \neq j,l} 0 \vec{v}_i = 0. \]

This leads us to impose the following.\textbf{Convention:} when we are told that a list of vectors is linearly dependent, we assume that this list contains no element repeatedly, or otherwise we will throw away all but one of them and start over.

The third convention is along a similar vein. Say we have an equation of linear dependence, \( \sum a_i \vec{v}_i = 0 \). Suppose that some coefficient is zero, say \( a_j = 0 \). Then the \( j \)-th term in the sum is \( a_j \vec{v}_j = 0 \vec{v}_j = \vec{0} = 1 \cdot \vec{0} \). This means that having a zero scalar coefficient contributes the same as having a zero vector in our list. The first convention above says that we should not consider this as part of the equation of linear dependence.\textbf{Convention:} given an equation of linear dependence, we assume that all the scalar coefficients are nonzero, or otherwise we throw away those terms from the list.

[Picture of three vectors in space, linearly independent; then three vectors in space, lying on same plane.]

Look back to the last example in the span section... \( \text{span}\{\vec{v}_1, \vec{v}_2\} = \text{span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \), and every \( \vec{u} \) here is of form \( \vec{u} = a_1 \vec{v}_1 + a_2 \vec{v}_2 \) \textit{uniquely}. Proof: If also \( \vec{u} = b_1 \vec{v}_1 + b_2 \vec{v}_2 \) then \( 0 = \vec{u} - \vec{u} = \ldots = (a_1 - b_1)\vec{v}_1 + (a_2 - b_2)\vec{v}_2 \). By linear independence, \( a_1 - b_1 = 0 \) so \( a_1 = b_1 \), and similarly \( a_2 = b_2 \).
Now let us prove a general result. If $V$ is a vector space and $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ are vectors in $V$, then $\{\vec{v}_i\}$ is linearly independent if and only if each $\vec{u} \in \text{span}\{\vec{v}_i\}$ may be written uniquely as a linear combination of the $\vec{v}_i$. The proof goes like this. Suppose $\{\vec{v}_i\}$ is linearly dependent, and we have a nontrivial relation,

$$a_1 \vec{v}_1 + \cdots + a_k \vec{v}_k = \vec{0}.$$  

Every $\vec{u}$ in $\text{span}\{\vec{v}_i\}$ is by definition a linear combination of the $\vec{v}_i$, say

$$b_1 \vec{v}_1 + \cdots + b_k \vec{v}_k = \vec{u}.$$  

Then by adding the last two equations together, and simplifying ($\vec{u} + \vec{0} = \vec{u}$, etc.), we get

$$(a_1 + b_1) \vec{v}_1 + \cdots + (a_k + b_k) \vec{v}_k = \vec{u},$$

which is a different representation, and so the original representation was not unique. Conversely, given two different representations

$$b_1 \vec{v}_1 + \cdots + b_k \vec{v}_k = \vec{u} = c_1 \vec{v}_1 + \cdots + c_k \vec{v}_k$$

of $\vec{u}$, we can subtract them from each other, and get

$$\vec{0} = \vec{u} - \vec{u} = (b_1 - c_1) \vec{v}_1 + \cdots + (b_k - c_k) \vec{v}_k,$$

and the fact that the two representations were different gives us the nontriviality of the representation of $\vec{0}$.

Behold, the proof in general is exactly the same as the proof of the specific example preceding it! This is not all that abstract! Moreover, the process carried out in the proof really drills home the viewpoint that the span produces you “all vectors which can written in terms of the given vectors.” It also shows that linear dependence says something about how redundant these representations can get: the more the nontrivial representations of $\vec{0}$ one has, then by “adding zero to both sides,” one can get more nontrivial representations of every vector.

Recapping slightly differently, the proof went like this. Step 1: show that in order to check the uniqueness of representation of all vectors in $\text{span}\{\vec{v}_i\}$, it suffices to check the uniqueness of representation of the single vector $\vec{0}$. Step 2: the lack of redundant representations of $\vec{0}$ is precisely the statement of linear independence.

Say we are given any vector space $V$, and we want a way to manage it and its information cleanly. One approach is to try to manage all of its vectors efficiently. As we have seen, a spanning set gives a way to represent all vectors. By the general fact we just proved, if a spanning set is linearly independent as well then it gives unique representations of all vectors. Since this seems like an efficient mechanism for managing all the vectors, we put a name on this notion: we define a basis is a linearly independent spanning set.

Let us again quickly reinforce this concept with some examples. $\mathbb{R}^2$ has, for starters, the standard basis $e_1 = (1, 0), e_2 = (0, 1)$. One can also rotate these by $45^\circ$, and get the basis $f_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), f_2 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. (One can rotate it by any angle and still get a basis.) In fact, we will always see that there is a wide variety of bases for any vector space. For instance, any two nonproportional vectors in $\mathbb{R}^2$ form a basis for it.
Any Euclidean space \( \mathbb{R}^n \) also has a **standard basis**:

\[
e_1 = (1, 0, \ldots, 0), \ e_2 = (0, 1, \ldots, 0), \ldots, e_n = (0, 0, \ldots, n).
\]

We will see shortly that one has a lot of freedom, however.

The subspace \( V \) of \( \mathbb{R}^3 \) spanned by \( \vec{v}_1, \vec{v}_2, \vec{v}_3 \) subject to \( \vec{v}_1 + \vec{v}_2 - \vec{v}_3 = 0 \), as defined at the end of the preceding section, has three easy choices of basis: \( \{\vec{v}_1, \vec{v}_2\}, \{\vec{v}_2, \vec{v}_3\}, \{\vec{v}_3, \vec{v}_1\} \). There are many more, and the reader is encouraged to find some of them, but notice that all bases here have exactly two vectors. The three-vector spanning set \( \{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \) is not a basis, since it is linearly dependent.

In general, if we take \( \{\vec{v}_i\} \) to be any linearly independent collection of vectors in a space \( V \), and set \( W = \text{span}\{\vec{v}_i\} \), then \( \{\vec{v}_i\} \) is a basis for \( W \). This claim is just a restatement of the definition: by construction, \( \{\vec{v}_i\} \) is a linearly independent spanning set for \( W \).

### 1.3 dimension

In the examples at the end of the preceding section, we noticed a curious thing. We had a vector space \( V \), and a basis \( \vec{v}_1, \ldots, \vec{v}_n \) for \( V \), and a second basis \( \vec{u}_1, \ldots, \vec{u}_m \) for \( V \). We noticed that both lists had the same size, that \( n = m \). It turns out that this is true for all vector spaces and all bases.

**Theorem.** If \( V \) is a vector space, and \( \{\vec{v}_i\} \) and \( \{\vec{u}_j\} \) are any two bases for \( V \), then \( \{\vec{v}_i\} \) and \( \{\vec{u}_j\} \) have the same number of elements.

For a proof of this fact, see for example G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*, Section 7.4, Theorem 5, Corollary 1.

We define the **dimension** of a vector space \( V \) is the number of elements in any basis for it. We denote this quantity by \( \dim V \). **Convention.** we will always assume our vector spaces are finite-dimensional. Examples of infinite-dimensional vector spaces will occur only in the extended exploration problems. (However, they will turn up in any course which uses Fourier theory.)

Note: only the zero vector space \( \{0\} \) has dimension zero.

We now go on to give several other ways to see the concept of dimension. We should probably move these up to the end of the preceding section, since they are really theorems about bases, and then by “reducing to the cardinality information” deduce the corresponding results about dimension.

Let \( V \) be a vector space, and let \( n \geq 0 \) be an integer. Then the following statements hold:

- \( \dim V \leq n \) if and only if \( V \) has a (at least one) spanning set with \( n \) elements.
- \( \dim V \geq n \) if and only if all spanning sets for \( V \) have at least \( n \) elements.
- Therefore, \( \dim V \) is the size of the smallest possible spanning set for \( V \).
- Call a spanning set \( S \) **minimal** if, for each element \( \vec{v}_i \) in \( S \), removing \( \vec{v}_i \) from \( S \) would cause \( S \) to no longer span \( V \).
• **Theorem.** Any minimal spanning set is a smallest possible spanning set, and therefore the dimension is the size of any minimal spanning set.

• \( \dim V \geq n \) if and only if there exists a set of \( n \) linearly independent vectors.

• \( \dim V \leq n \) if and only if all linearly independent sets in \( V \) have not more than \( n \) elements.

• Therefore, the dimension is equal to the size of a largest possible linearly independent set.

• A linearly independent set \( S \) is called *maximal* in \( V \) when adding any element of \( V \) to \( S \) forces \( S \) to no longer be linearly independent.

• **Theorem.** Every maximal linearly independent set is a largest possible linearly independent set, and therefore the dimension is the size of any maximal linearly independent set.

At the beginning of the next lecture we use these facts to show how to compute spanning sets and bases.
2 maps

2.1 transformations

• explain functions somehow. theology, blah.

• examples of functions:
  – examples here

• definition. let $U$ and $V$ be vector spaces, and $L : U \rightarrow V$ be a function (mapping, etc). we say that $L$ is a linear mapping (transformation, etc...) when $L$ distributes through addition and scalar mult. specifically, if $\vec{u}, \vec{v}$ are vectors $U$ and $a, b$ are scalars, then we require that $L(a\vec{u} + b\vec{v}) = aL(\vec{u}) + bL(\vec{v})$.

• note: “superposition”

• note: can calculate values of $L$ “term-by-term”

• examples:
  – two trivial examples: the zero map, and the identity map.
  – scalar mult. includes above.
  – $(x_1, \ldots, x_n) \mapsto (\lambda_1 x_1, \ldots, \lambda_n x_n)$. DRAW PICTURE.

• next recall the key property of a basis: it is a set of vectors $\vec{v}_1, \ldots, \vec{v}_n$ in $V$ such that each vector $\vec{u}$ in $V$ has a unique expression as a linear combination, $\vec{u} = \sum a_i \vec{v}_i$.

• when $\vec{u}$ is written in this form, and $L$ is a linear function, we may calculate $L(\vec{u}) = L(\sum a_i \vec{v}_i) = \sum a_i L(\vec{v}_i)$. in other words, the value $L(\vec{u})$ depends only on the fixed values $L(\vec{v}_i)$ and the coefficients $a_i$ in the representation of $\vec{u}$.

• therefore, if we know how a linear transformation $L$ acts on a basis, on just this finite set of vectors, then we know how it acts on everything in $V$. this is what we mean, when we say that “a linear map is determined by its action on a basis”.

• figuratively, we may view a basis as the “skeleton” of a vector space. we “fill in the flesh”, obtaining all other vectors, by taking all linear combinations. we picture a linear map as moving or twisting a basis in some manner, and the basis dragging the rest of the vector space along with it.

• let’s draw some pictures of some linear maps which make this clear. (PICTURES HERE.)

• in the examples above, we chose values to assign the $L(\vec{v}_i)$ (PHRASE BETTER), and then filled in this definition to make a linear map $L(\vec{u})$ on all of the vector space $V$. although we made particular choices for the $L(\vec{v}_i)$, the choices are entirely arbitrary. more precisely, given a vector space $V$, basis $\vec{v}_i$, and another vector space $U$, for each possible choice of $\vec{u}_1, \ldots, \vec{u}_n$ in $U$, there is a unique linear mapping determined by $L(\vec{v}_i) = \vec{u}_i$ for all $i$. furthermore, all linear mappings from $V$ to $U$ are of this form.
• in other words, a linear mapping is determined by the finite set of data, $\vec{u}_i = L(\vec{v}_i)$, the values $L$ takes on a basis; and conversely, every finite set of data, telling one what to do with a basis, can be uniquely extrapolated to a linear mapping on the whole vector space.

• segue: what does the map look like, given the $\vec{u}_i$ and $\vec{v}_i$? what if the $\vec{u}_i$ are linearly dependent? and what if not?

• give ($\vec{u}_i$ lin dep) iff (nullspace nonzero) iff (dimension decreases), and relate to chapter 1. engineer showing this, in such a way as to make the conservation of dimension law easy.

  effect on dimension (it decreases or stays the same), range/nullspace+rank/nullity, range equals columnspace, complimentary subspace of the kernel is mapped surjectively (bijectively), more row reduction

note: when talking about action on bases, use skeleton/body analogy.

examples: complimentary subspaces, projections, diagonals, rotations, matrix mult vs. composition of maps.

goal: pick up a matrix, and view its linear map geometrically.

3 maps with equal dimension; invertibility

3.1 invertibility

- asdfasdf

preservation of linear (in)dependence, change of basis, similarity, elementary row ops, determinants, equivalent conditions, methods of computation, cramer’s rule.

examples: \( \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \), (after dets) 2x2 rule.

other topics: expansion by minors.

further exploration: det and multilinear algebra.

application: cramer’s rule = way2 of solving linear systems.
4 solidification/applications

• similar matrices / conjugation
• geometry
• least squares approximations
• “linear systems” (ask vkm for definition), and particularly state-space methods
• stress/strain and bases
• isotropic material
• rigid body dynamics
5 eigenstuff

stable subspaces, picture of eigenvcs, picture of diag’ability, finding with det, characteristic poly (det/tr), non-diag’ability, Jordan form, positive/negative-defness.

In this section we take up the study of eigenvectors. Eigenvectors are an extremely important tool in linear algebra, and they are used in almost every application of the field. When correctly understood, they provide a way to see the entire effect of a linear mapping at a glance. In applications, they often isolate some of the most important qualitative features of the system in question.

An eigenvector is a special vector which is attached to a square matrix, or to a linear mapping of a single vector space to itself. The collection of eigenvectors of a matrix tell you everything there is to know about it, in many cases. So which vectors are these? They are the ones on which the linear mapping acts in a particularly simple way. Let us begin with the prototypical example.

5.1 Motivation

We take the standard vector space $V = \mathbb{R}^n$ with the standard basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. The simplest linear mappings $L: V \rightarrow V$ are those of the form $L(v) = \lambda v$ for some scalar $\lambda$. These mappings have a matrix which looks like

$$
\begin{bmatrix}
\lambda \\
\lambda \\
\vdots \\
\lambda
\end{bmatrix}.
$$

A matrix of this form is called a scalar matrix because, geometrically, this merely stretches the whole space by a factor of $\lambda$. (FIGURE!)

Making the situation only a little bit messier, suppose we are given scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$, and we form the matrix

$$
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_n
\end{bmatrix}.
$$

Such a matrix is called a diagonal matrix. This matrix sends each $\vec{v}_i$ to $\lambda_i \vec{v}_i$. And we know that for an arbitrary vector $\vec{v} \in V$, if we write $v = \sum a_i \vec{v}_i$, then the matrix sends $\vec{v}$ to the vector

$$
\sum_{i=1}^{n} \lambda_i a_i \vec{v}_i,
$$

since knowing how a linear mapping acts on a basis tells you how it acts on every vector. The formula for $\vec{v}$ immediately preceding should also be clear from the matrix we have written down.

Considering this matrix as a linear mapping, what does its action look like geometrically? The answer to this is simple as well. One first draws the axes corresponding to the basis
The action of this linear mapping on the basis is merely to stretch each axis $\vec{v}_i$ to the size $\lambda_i$. Then, for a general point $\vec{v} \in V$, to find where the linear mapping sends $\vec{v}$, just apply a stretch by $\lambda_i$ in each $\vec{v}_i$-direction, each time taking $\vec{v}$ along with it parallel to that axis. (FIGURE!)

Here is another geometrical insight, which will be important. For each $i$, the matrix sends the subspace span{$v_i$} to itself, and in fact just acts as a scalar matrix on this subspace. Further, all of these subspaces span $V$.

This example, with its simple geometrical picture, and its ease of computation, is the prototype. Through the theory of eigenvectors, we will show that for many matrices, the picture is exactly the same as this one, after possibly a change of basis. We will also show that for all matrices, the picture can be made very similar to this one, after we enlarge our experience to include the phenomenon of nilpotence.

5.2 Definitions

We now turn to the general case.

Definition.

• Let $A$ be a square matrix. A nonzero column vector $\vec{v}$ is called an eigenvector of $A$ if $A$ sends $\vec{v}$ to a scalar multiple of itself. In other words, $\vec{v}$ is an eigenvector of $A$ if $A\vec{v} = \lambda \vec{v}$ for some scalar $\lambda$. We call $\lambda$ the eigenvalue of $\vec{v}$ under $A$. We say that $\lambda$ is an eigenvalue of $A$ if it is the eigenvalue of $\vec{v}$ under $A$ for some nonzero $\vec{v}$.

• If $V$ is a vector space and $L: V \rightarrow V$ is a linear mapping, one says that $\vec{v} \in V$ is an eigenvector for $L$ if $\vec{v}$ is nonzero and $L(\vec{v}) = \lambda \vec{v}$ for some scalar $\lambda$. In this situation, we call $\lambda$ the eigenvalue of $\vec{v}$ under $L$. We call $\lambda$ an eigenvalue of $L$ if it is the eigenvalue of some nonzero $\vec{v}$ under $L$.

We require that $\vec{v} \neq \mathbf{0}$, because $A \mathbf{0} = \lambda \mathbf{0}$ for all $\lambda$. If we allowed $\vec{v} = \mathbf{0}$, then every $\lambda$ would be an eigenvalue, specifically for the eigenvector $\mathbf{0}$, and therefore the definition would not be very meaningful. Also, by requiring that $\vec{v} \neq \mathbf{0}$, we see that if $\lambda \vec{v} = A\vec{v} = \lambda' \vec{v}$ for some other scalar $\lambda'$, then we must actually have $\lambda = \lambda'$. Therefore “the” eigenvalue of $\vec{v}$ is a well-defined notion.

Example. Let $I$ denote the identity matrix, as usual. Then $I\vec{v} = \vec{v} = 1 \cdot \vec{v}$ for all $\vec{v}$, whence every nonzero vector is an eigenvector for $I$, with eigenvalue 1. But since there is only one eigenvalue for any eigenvector, we find that 1 is the only eigenvalue of $I$.

Example. Let $\lambda$ be a scalar, and consider the matrix $A = \lambda I$. Then in the exact same manner as in the preceding example, one has $A\vec{v} = \lambda I \vec{v} = \lambda \vec{v}$ for all $\vec{v}$, and therefore every nonzero vector is an eigenvector with eigenvalue $\lambda$, and furthermore $\lambda$ is the only eigenvalue of $A$.

Example. In the prototypical example of the preceding subsection, where we are dealing with the matrix

$$A = \begin{bmatrix} \lambda_1 \\ & \lambda_2 \\ & & \ddots \\ & & & \lambda_n \end{bmatrix},$$

16
one sees that $A\vec{v}_i = \lambda_i \vec{v}_i$. Therefore, each $\vec{v}_i$ is an eigenvector of $A$ with eigenvalue $\lambda_i$. We will soon show how to prove that these are the only eigenvalues of $A$.

5.3 Eigenspaces

The examples at the end of the preceding subsection are too simplistic to show a general method for computing the eigenvectors of a matrix. The primary tool for this job is the characteristic polynomial, which we will discuss in the next section. In this section, we introduce a eigenspaces, which are used in conjunction with the characteristic polynomial. We use it here to complete our discussion of diagonal linear mappings, and derive some important facts. However, since this section is somewhat more technical than the others nearby, it may be preferable for the reader to start with the basic definition and comments, and then skip ahead to the next section with little harm in understanding.

Definition. Let $L: V \to V$ be a linear mapping, and let $\lambda$ be a scalar. The $\lambda$-eigenspace of $L$ is the set

$$V(L, \lambda) = \{ \vec{v} \in V \mid L(\vec{v}) = \lambda \vec{v} \},$$

that is, the set of eigenvectors of $L$ with the fixed eigenvalue $\lambda$, together with the zero vector.

A word on notation. There is no standard notation for an eigenspace. In these notes we will write $V(L, \lambda)$, but other authors will often use their own notation. Also, it will often be clear from context which linear mapping $L$ we are talking about. In such a case, it will often be easier to relax the notation and simply write $V(\lambda)$.

We make note here of two facts concerning an eigenspace. First, for every scalar $\lambda$, the set $V(\lambda)$ is a bona fide subspace $V$. Here is the proof: if $\vec{u}, \vec{v} \in V(\lambda)$, and $a, b$ are scalars, then $L(a\vec{u} + b\vec{v}) = aL(\vec{u}) + bL(\vec{v}) = a\lambda \vec{u} + b\lambda \vec{v} = \lambda(a\vec{u} + b\vec{v})$, whence $a\vec{u} + b\vec{v} \in V(\lambda)$.

Second, the subspace $V(\lambda)$ is a nonzero subspace if and only if $\lambda$ is an eigenvalue of $L$ on $V$. This is a result of merely putting the definitions of $V(\lambda)$ and of an eigenvalue together.

These two facts together explain the name “eigenspace”.

Next we state a result which tells us roughly that the various eigenspaces are independent of each other. This result, as we will see in the example that follows it, will sometimes be useful in determining when a subset of some $V(\lambda)$ spans $V(\lambda)$. We will state the result in two ways, which will turn out to be equivalent. Here is the first one:

Eigenspace Independence Theorem. Let $L: V \to V$ be a linear mapping, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct scalars. For each $\lambda_i$, let $\vec{v}_i$ be an eigenvector with eigenvalue $\lambda_i$. Then the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is linearly independent.

Before stating the theorem in the second form, we will need to introduce a little bit of terminology.

Let $V$ be a vector space, and $\{V_i \mid i \in J\}$ a collection of subspaces of $V$. We say that the $V_i$ are linearly independent of each other if the following condition is satisfied: whenever we choose a finite list of vectors $\vec{v}_i$ from various distinct $V_i$, in such a manner that $\sum \vec{v}_i = 0$, the vectors are linearly independent.
it must be the case that every $\vec{v}_i = 0$. (A set of nonzero vectors $\{\vec{v}_i \mid i \in J\}$ is linearly independent in the old sense if and only if the collection $\{\text{span}\{\vec{v}_i\} \mid i \in J\}$ of subspaces is linearly independent in this new sense.)

One can show that the theorem above is equivalent to the following statement:

**Eigenspace Independence Theorem.** Let $L: V \to V$ be a linear mapping, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct scalars. Then the eigenspaces $V(\lambda_1), V(\lambda_2), \ldots, V(\lambda_k)$ are all linearly independent of each other.

**Idea of Proof.** In order to turn the underlying geometrical idea into a formal proof, one must perform some alterations on it which obscure the real idea. Therefore, we present this idea as a sketch, so that it is not lost. The reader who is satisfied with this loose geometrical reasoning is invited to skip the proof which follows.

We suppose that we have vectors $\vec{v}_i \in V(\lambda_i)$ for each $i$, and that they satisfy $\sum \vec{v}_i = 0$. Pick the first one and move it over to the other side, so that $\vec{v}_1 = -\sum_{i>1} \vec{v}_i$. The vector $\vec{v}_1$ is an eigenvector with value $\lambda_1$, where each vector on the right hand side is an eigenvector with some different eigenvalue. So when we apply $L$ to both sides, the left hand side is stretched by $\lambda_1$, and the right hand side is stretched, piece by piece, by a host of different scalars. These distinct stretches are incompatible in such a way that the only possible vector which transforms the same way under both of them is $0$. This proves that $\vec{v}_1 = 0$. Repeating this argument over again with $\vec{v}_2$ in place of $\vec{v}_1$, and then with $\vec{v}_3$, and so on, shows that all the $\vec{v}_i$ must be $0$. This would prove the theorem.

**Proof.** Let the $\vec{v}_i \in V(\lambda_i)$ be given, and suppose that $\sum \vec{v}_i = 0$. We will show by contradiction that we must have all $\vec{v}_i = 0$.

We may assume, when we are given an expression $\sum \vec{v}_i = 0$, that each $\vec{v}_i$ we are given is nonzero; otherwise, we remove $\lambda_i$ and $\vec{v}_i$ from the list, and apply the argument again to the shortened list.

Next, consider all expressions $\sum \vec{v}_i = 0$ for various collections of $\lambda_i$ and $\vec{v}_i$. Some of these will have different lengths $k$. If there are any expressions at all, there will have to be one with the smallest length $k$. Let us assume we are working with an expression which has $k$ minimal. To find a contradiction, we will produce an expression of the smaller length.

We start by taking the expression $\sum \vec{v}_i = 0$, and moving the $\vec{v}_1$ over to its own side:

$$-(\vec{v}_2 + \vec{v}_3 + \cdots + \vec{v}_k) = \vec{v}_1.$$  

Applying $L$ to both sides of this expression, we get

$$-(\lambda_2 \vec{v}_2 + \lambda_3 \vec{v}_3 + \cdots + \lambda_k \vec{v}_k) = \lambda_1 \vec{v}_1$$

Alternatively, multiplying both sides of the first expression by $\lambda_1$ we get

$$-(\lambda_1 \vec{v}_2 + \lambda_1 \vec{v}_3 + \cdots + \lambda_1 \vec{v}_k) = \lambda_1 \vec{v}_1.$$  

Subtracting these last two lines gives

$$(\lambda_1 - \lambda_2)\vec{v}_2 + (\lambda_1 - \lambda_3)\vec{v}_3 + \cdots + (\lambda_1 - \lambda_k)\vec{v}_k = 0.$$
Define $\mu_i = \lambda_{i+1}$ and $\vec{u}_i = (\lambda_1 - \lambda_{i+1})\vec{v}_{i+1}$ for $i = 1, 2, \ldots, k - 1$. Since all the $\lambda_i$ were distinct so are all the $\mu_i$, and furthermore we have $\lambda_1 - \lambda_i \neq 0$ for $i > 1$, so that all the $\vec{u}_i$ are nonzero. Thus the expression $\sum \vec{u}_i = 0$ has is of the type we are considering, and it has length $k - 1 < k$, a contradiction.

Since there is no shortest expression, there is no expression at all, and the spaces $V(\lambda_i)$ are linearly independent.

**Corollary.** If we let $U = \text{span}(V(\lambda_1), V(\lambda_2), \ldots, V(\lambda_k))$, with the $\lambda_i$ distinct, then $\dim(U) = \sum \dim(V(\lambda_i))$.

**Proof.** For each $i$, choose a basis $\{\vec{v}_{i1}, \vec{v}_{i2}, \ldots, \vec{v}_{in_i}\}$ for $V(\lambda_i)$, where $n_i = \dim(V(\lambda_i))$. Then the set $\{v_{ij} \mid 1 \leq i \leq k, 1 \leq j \leq n_i\}$ spans $U$, and is linearly independent as a consequence of the above theorem.

**Corollary.** If $V$ is a vector space with $n = \dim(V)$, and $L: V \to V$ is a linear mapping, then $L$ has at most $n$ distinct eigenvalues.

**Proof.** In the corollary above, each distinct eigenvalue $\lambda_i$ has a nonzero eigenspace $V(\lambda_i)$, which contributes at least 1 to the sum on the right hand side of $\dim(U) = \sum \dim(V(\lambda_i))$. Since $\dim(U) \leq \dim(V) = n$, we get the desired result.

**Corollary.** Suppose $L: V \to V$ is a linear mapping, and $\lambda_1, \lambda_2, \ldots, \lambda_k$ are scalars such that $\text{span}\{V(\lambda_i) \mid i = 1, 2, \ldots, k\}$ is all of $V$. Then every eigenvalue $\lambda$ of $L$ must be in the list $\lambda_1, \lambda_2, \ldots, \lambda_k$.

**Proof.** Let $\lambda$ be a scalar which is different from each $\lambda_i$. By the theorem, the subspaces $V(\lambda)$, $V(\lambda_1)$, $V(\lambda_2)$, $\ldots$, and $V(\lambda_k)$ are linearly independent of each other. But $V(\lambda_1), V(\lambda_2), \ldots, V(\lambda_k)$ taken together span all of $V$. Therefore $V(\lambda)$ is linearly independent of all of $V$, and must be the zero space. But $V(\lambda)$ is the zero space if and only if $\lambda$ is not an eigenvalue. Therefore, in order for $\lambda$ to be an eigenvalue, one must have $\lambda$ in the list.

Now we use the above theorem to conclude our study of diagonal matrices, starting where we left them at the end of the preceding subsection. Recall that we have fixed a basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ of $V$, and that $L$ is a linear map determined by the rule $L(\vec{v}_i) = \lambda_i \vec{v}_i$ for all $i$. We know that each $\vec{v}_i$ is an eigenvector of $L$ with eigenvalue $\lambda_i$. According to the second corollary to the above theorem we now know that if the $\lambda_i$ are distinct, then there are $n$ of them, and there can be no others, and therefore the $\lambda_i$ are all of the eigenvalues of $L$. In fact, we have actually gotten a stronger result:

**Distinct Eigenvalues Theorem.** If $\dim(V) = n$ and $L$ has $n$ distinct eigenvalues, then there exists a basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ with respect to which the matrix for $L$ is diagonal (and hence we are in the simple geometrical picture described in the first section).

**Proof.** For each $i$, we know that $\lambda_i$ is an eigenvalue for $L$, and therefore that there exists a $\vec{v}_i \in V$ which is an eigenvector for $L$ with eigenvalue $\lambda_i$. Apply the theorem to the collection of subspaces $\{V(\lambda_i) \mid i = 1, \ldots, n\}$ to conclude that the various $\vec{v}_i$, each belonging to different $V(\lambda_i)$, are linearly independent. But there are $n = \dim(V)$ of them, and therefore they must be a basis. Since $L$ has the form $L(\vec{v}_i) = \lambda_i \vec{v}_i$ on this basis, we see that the matrix for $L$ is
diagonal with respect to this basis.

This is a powerful fact, which tells us about a wide class of linear mappings or matrices. We still need to treat the case of a diagonal matrix in which not all of the $\lambda_i$ are distinct. Since we have counted less than $n$ distinct eigenvalues, one might ask if it is possible for other eigenvalues to occur. We will show now that this is impossible—that we still have all possible eigenvalues—and along the way we will show that the eigenspaces for these eigenvalues are exactly what we would expect them to be.

As in the proof of the eigenspace independence theorem, this discussion gets technical, and so we will state simply what the picture is. The reader which is contented by this reasoning is invited to skip ahead to the statement of the next theorem. Essentially the same ideas are at work as those in the eigenspace independence theorem. In fact, the proof calls upon this theorem for support at an essential venture. One begins by writing out all the spaces $V(\lambda_i)$, where the $\lambda_i$ are the various diagonal entries. Each one contains a basis eigenvector for each $\vec{v}_j$ which has $\lambda_j = \lambda_i$. This would lead us to expect that $V(\lambda_i)$ has dimension equal to the number of diagonal entries equal to $\lambda_i$. Since $V(\lambda_i)$ has at least this dimension (for it contains all these $\vec{v}_j$ with $\lambda_j = \lambda_i$), it suffices to show that it cannot have any larger dimension. To do this, we have to consider all the $V(\lambda_i)$ at once. The eigenspace independence theorem tells us that they are all linearly independent of each other, and therefore their common span has dimension $\geq \sum \dim(V(\lambda_i))$. But since each $\vec{v}_i$ is contained in some $V(\lambda_i)$, and all these vectors are linearly independent, the right hand side is at least $n = \dim(V)$. Now, if any $V(\lambda_i)$ were strictly bigger than the span of the $\vec{v}_j$ with $\lambda_j = \lambda_i$, then one would have that their common span has dimension at least $n + 1$, which is impossible. Any eigenvalue $\mu$ that is distinct from all the $\lambda_i$ will contradict the third corollary to the theorem, and therefore cannot exist. Therefore, the $\lambda_i$ are the only eigenvalues, and for any basis with respect to which $L$ is diagonal, each $\lambda_i$ appears precisely $\dim(V(\lambda_i))$ times along the diagonal.

So now suppose that the $\lambda_i$ are arbitrary, and might not be distinct. We see immediately that $\lambda_i \in V(\lambda_i)$ and therefore $\operatorname{span}\{\lambda_i\} \subseteq V(\lambda_i)$. However, if for some value of $j$ we have $\lambda_j = \lambda_i$, then also $\operatorname{span}\{\lambda_i\} \subseteq V(\lambda_j) = V(\lambda_i)$. Therefore taking all such $j$ into account, we find that $\operatorname{span}\{v_j \mid \lambda_j = \lambda_i\} \subseteq V(\lambda_i)$. Using the theorem just proved, we will show that in fact this last $\subseteq$ is an equality:

For every $i$, $V(\lambda_i) = \operatorname{span}\{v_j \mid \lambda_j = \lambda_i\}$.

**Proof.** According to the first corollary to the theorem,

$$n = \dim(V) \geq \dim(U) = \sum_{\lambda_i} V(\lambda_i),$$

the sum being taken over all distinct values of $\lambda_i$. But we know that $\operatorname{span}\{\vec{v}_j \mid \lambda_j = \lambda_i\} \subseteq V(\lambda_i)$, and so by taking the dimensions of both sides of this inclusion and summing over all distinct $\lambda_i$, we find that

$$\sum_{\lambda_i} V(\lambda_i) \geq \sum_{\lambda_i} \dim(\operatorname{span}\{v_j \mid \lambda_j = \lambda_i\}).$$
However, since \( \{ \vec{v}_1, \ldots, \vec{v}_n \} \) is assumed to be a basis, we know that the \( v_i \) are all linearly independent. Therefore, \( \dim(\text{span}\{v_j \mid \lambda_j = \lambda_i\}) \) is just the number of \( v_j \) we are taking the span of, and this is just \#\( \{ j \mid \lambda_j = \lambda_i \} \). Requiring the last inequality with this in mind, we have

\[
\sum_{\lambda_i} V(\lambda_i) \geq \sum_{\lambda_i} \#\{ j \mid \lambda_j = \lambda_i \}.
\]

Finally, consider the sum on the right hand side of the last inequality, \( \sum_{\lambda_i} \#\{ j \mid \lambda_j = \lambda_i \} \). If we take the set \( \{ 1, 2, \ldots, n \} \), and partition it into subsets such that \( i \) and \( j \) are grouped together if and only if \( \lambda_i = \lambda_j \), then the size of the group containing \( i \) is just \( \#\{ j \mid \lambda_j = \lambda_i \} \). By summing over the distinct values of \( \lambda_i \), we are summing over all groups. Therefore, the sum \( \sum_{\lambda_i} \#\{ j \mid \lambda_j = \lambda_i \} \) counts all elements in every group, and hence just counts the elements in the original set, \( \{ 1, 2, \ldots, n \} \). Therefore that sum is equal to \( n \). Putting everything together, we have the chain of inequalities,

\[
n \geq \sum_{\lambda_i} \dim(V(\lambda_i)) \geq \sum_{\lambda_i} \#\{ j \mid \lambda_j = \lambda_i \} = n,
\]

and the fact that the two ends are equal force that the \( \geq \) in the middle is actually an \( = \). Since \( \dim(V(\lambda_i)) \geq \#\{ j \mid \lambda_j = \lambda_i \} \) for every \( i \), if we were to have a strict inequality (\( > \)) for any \( i \), we would get a strict inequality (\( > \)) in the big above inequality, which would be a contradiction. Therefore, for each \( i \), we get the equality \( \dim(V(\lambda_i)) = \#\{ j \mid \lambda_j = \lambda_i \} = \dim(\text{span}\{ \vec{v}_j \mid \lambda_j = \lambda_i \}) \), and finally we conclude that

\[
V(\lambda_i) = \text{span}\{ \vec{v}_j \mid \lambda_j = \lambda_i \}, \quad \text{for every } i.
\]

We may continue with no extra effort to find the other fact that we sought to prove. Since now we know that \( n = \sum \dim(V(\lambda_i)) \), we see that there can be no other eigenvalues. For if \( \mu \) is any scalar distinct from all the \( \lambda_i \), then the theorem implies that \( V(\mu) \) is linearly independent of all the \( V(\lambda_i) \). But the \( V(\lambda_i) \) span \( V \) (since their span contains a basis for \( V \)), and hence \( V(\mu) \) is linearly independent of \( V \), which forces \( V(\mu) \) to be the zero subspace of \( V \). Therefore, \( \mu \) is not an eigenvalue, and there are no eigenvalues other than the \( \lambda_i \).

All of this amounts to the proof of the following theorem.

**Theorem.** Let \( L: V \to V \) be a linear mapping which is diagonal with respect to at least one basis. Then for every basis of \( V \) under which \( V \) is diagonal, the diagonal entries \( \lambda_i \) are precisely the eigenvalues of \( L \), and each \( \lambda_i \) appears \( \dim(V(L, \lambda_i)) \) times along the diagonal.

To recap, we have shown the following. If \( L \) is a linear mapping which is diagonal with respect to some basis, then the information of the diagonal matrix is completely determined by the theory of eigenvectors. (In the exercises, you will be asked to show how the theory of eigenvectors determines all bases are that diagonalize such a linear mapping.) Therefore, the theory of eigenvectors achieves its goal of algebraically encapsulating the geometrical invariants of a diagonal transformation.
5.4 The Characteristic Polynomial

Now that we have finished working out our formalism in the case of diagonal matrices, we return to a couple of examples, whose simplicity makes their deep usefulness a surprise. In fact, we will use them to produce an algorithm for finding all the eigenvectors of any matrix.

Example. For \( \lambda = 0 \), the 0-eigenspace of \( L \) is the set \( \{ \vec{v} \in V \mid L(\vec{v}) = \vec{0} \} \), by definition. But this set is precisely the nullspace (kernel) of \( L \)! Therefore, 0 is an eigenvalue of \( L \) if and only if \( L \) has nonzero nullspace, if and only if \( L \) is singular. Therefore, we can test whether 0 is an eigenvalue simply by checking whether or not \( \det L = 0 \).

Example. When \( \mu \) is a scalar, the linear map \( L(\vec{v}) = \mu \vec{v} \) has \( V(\mu) = V \), because every nonzero vector in \( V \) literally satisfies the definition of an eigenvector for \( L \) with eigenvalue \( \mu \). If \( \lambda \) is any scalar not equal to \( \mu \), then any eigenvector for \( L \) with eigenvalue \( \lambda \) would have \( \lambda \vec{v} = L(\vec{v}) = \mu \vec{v} \), and therefore \((\lambda - \mu)\vec{v} = \vec{0} \). But \( \lambda \neq \mu \) and therefore we can multiply both sides by \( 1/(\lambda - \mu) \) to conclude that \( \vec{v} = 0 \). Therefore, \( V(\lambda) \) is the zero vector space for \( \lambda \neq \mu \), while \( V(\mu) = V \).

We want to exploit the last two examples to produce a method of computing eigenvectors. The first step is to use the following trick.

Let \( L_1, L_2 : V \to V \) be two linear mappings. Suppose that \( \vec{v} \) is a common eigenvector for both of them, say with eigenvalue \( \lambda_i \) under \( L_i \). Then \( \vec{v} \) is an eigenvector for \( L_1 + L_2 \) with eigenvalue \( \lambda_1 + \lambda_2 \). Proof: \((L_1 + L_2)\vec{v} = L_1\vec{v} + L_2\vec{v} = \lambda_1\vec{v} + \lambda_2\vec{v} = (\lambda_1 + \lambda_2)\vec{v} \).

We use this trick in the following manner. Let \( L \) be our given linear function, and let \( \lambda, \mu \) be scalars. Suppose that \( \vec{v} \in V(\lambda) \). Since every vector is an eigenvector for \( \mu I \) with eigenvalue \( \mu \), then by the preceding remark, applied to \( L_1 = L \), \( \lambda_1 = \lambda \), \( L_2 = \mu I \), \( \lambda_2 = \mu \), we have that \( \vec{v} \) is an eigenvalue for \( L + \mu I \) with eigenvalue \( \lambda + \mu \). Thus one has

\[
V(L, \lambda) = V(L + \mu I, \lambda + \mu).
\]

In particular, when we specialize to \( \mu = -\lambda \), we get the following fundamental fact:

**Theorem.** \( V(L, \lambda) = V(L - \lambda I, 0) = \text{nullspace}(L - \lambda I) \).

**Proof.** We have just shown the left equality, and the right equality is merely the result of the first example of this section.

Therefore, to test whether \( \lambda \) is an eigenvalue of \( L \), i.e. whether \( V(L, \lambda) \) is nonzero, it suffices to check whether or not \( L - \lambda I \) is singular. And for this, it just suffices to check whether \( \det(L - \lambda I) = 0 \). Thus, we have proved another fundamental fact:

**Theorem.** \( \lambda \) is an eigenvalue of \( L \) if and only if \( \det(L - \lambda I) = 0 \).

We are thus motivated to take a closer look at the function \( \phi(t) = \det(L - tI) \), since the roots of \( \phi(t) \) are precisely the eigenvalues of \( L \). To do this, let us choose a basis \( \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \)
for $V$, and write $A = (a_{ij})$ for the matrix of $L$ with respect to this basis. Then we have

$$\det(L - tI) = \det(A - tI) = \begin{vmatrix} a_{11} - t & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - t & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - t \end{vmatrix}.$$ 

Recall from our study of determinants that this quantity will be a sum and difference of monomials in the various entries listed above. In particular, each entry is a polynomial in $t$ (more precisely, either a constant $a_{ij}$ or a linear term $a_{ii} - t$), and therefore any monomial in the entries is a polynomial in $t$, and hence any sum and difference of them is also a polynomial in $t$. Therefore, $\det(A - tI)$ is a polynomial in $t$. But we can say more. In the expression for the determinant, there is a term which is the product of the entries along the main diagonal: $(a_{11} - t)(a_{22} - t) \cdots (a_{nn} - t)$, which is a polynomial of degree $n$, and whose coefficient of $t^n$ is $\pm 1$. All other monomials in the expression for the determinant have at most $n - 1$ entries along the main diagonal, and therefore contribute to the determinant a polynomial whose degree is at most $n - 1$. So no other terms can contribute to the coefficient of $t^n$, and in particular cannot cancel it out. Therefore, $\det(A - tI)$ is a polynomial in $t$ of degree $n$ and with highest degree term $\pm t^n$.

In order to make this calculation, we had to choose a basis $\{v_1, v_2, \ldots, v_n\}$. However, we do know that the determinant is an intrinsic quantity, and is invariant of what basis we are working in. Therefore, our calculation of $\phi(t)$ did not depend on the choice of basis either, and $\phi(t)$ is intrinsically attached to $L$, however we may compute it using any basis for $V$ that we please.

We now summarize all of the preceding discussion.

**Definition-Theorem.** Let $V$ be a vector space, and $L: V \to V$ a linear mapping. We call the function $\phi(t) = \det(L - tI)$ the *characteristic polynomial* of $L$. It is a polynomial of degree exactly $n = \dim(V)$, and whose coefficient of $t^n$ is $\pm 1$. A scalar $\lambda$ is an eigenvalue for $L$ if and only if $\phi(\lambda) = \det(L - \lambda I) = 0$. Furthermore, the eigenvectors of $L$ with eigenvalue $\lambda$ are precisely the nonzero vectors of nullspace($L - \lambda I$).

**Corollary.** A linear mapping $L$ on a vector space $V$ can have at most $n = \dim(V)$ eigenvalues.

**Proof.** This is true because the polynomial $\phi(t)$ of degree $n$ can have at most $n$ roots. (Yes, this was proved in the preceding section as well. But the extreme difference between the two methods of proof merited seeing this proof as well.)

*A few words on notation.* Many authors define the characteristic polynomial of a matrix by $\phi_1(t) = \det(\lambda I - L)$. It is easy to see that $\phi_1(t) = (-1)^n \phi(t)$, and therefore $\phi_1(\lambda) = 0$ if and only if $\phi(\lambda) = 0$, so that one gets an almost identical theory. One might argue that $\phi_1(t)$ is “simpler” because its coefficient of $t^n$ is always $+1$. Perhaps they are correct.

Let us take a step back and see what this theorem does for us. We have a linear mapping $L$, whose eigenvectors we wish to find. Then the theorem tells us to do the following:
1. Take a determinant and compute the polynomial $\phi(t)$.

2. Find the roots of $\phi(t)$; these roots are the eigenvalues of $L$.

3. For each root $\lambda$ of $\phi(t)$, using the methods of Day 2, compute the nullspace of $L - \lambda I$; the nonzero vectors of this subspace are the eigenvectors of $L$ with eigenvalue $\lambda$.

Therefore, we have a complete procedure for finding eigenvectors, thanks to the content of this theorem!

One last thing on this topic. It might be useful, when carrying out the above program, to know how large to expect $\text{nullspace}(L - \lambda I)$ to be. It turns out that one can find a very easy bound, just from looking at the form of $\phi(t)$. In order to describe this bound, we will need to define a few terms first.

We begin by recalling the following theorem, which is common to most highschool programs:

**The Root-Factor Theorem.** Let $f(x)$ be a polynomial. Then $a$ is a root of $f(x)$ if and only if the polynomial $(x - a)$ is a factor of $f(x)$.

**Definition.** Let $f(x)$ be a polynomial. We say that $a$ is a root of $f(x)$ of multiplicity $n$ if $(x - a)^n$ is a factor of $f(x)$, but $(x - a)^{n+1}$ is not. In other words, $n$ is the largest exponent for which $(x - a)^n$ divides $f(x)$. We write $n = \text{mult}(f, a)$.

Thus, by the Root-Factor Theorem, $a$ is a root of $f(x)$ if and only if it is a root of $f(x)$ of multiplicity $\geq 1$, and a nonroot if and only if it is a root of multiplicity zero.

It is a fact that the sum of the multiplicities of the roots of $f(x)$ does not exceed the degree of $f(x)$, and the Fundamental Theorem of Algebra says that the sum of the multiplicities of the complex roots is exactly equal to the degree of $f(x)$.

Our bound on the number of linearly independent eigenvectors for a given eigenvalue now goes like this:

**Theorem.** Let $L: V \to V$ be a linear mapping with characteristic polynomial $\phi(t)$, and let $\lambda$ be a scalar. Then $\dim V(L, \lambda) \leq \text{mult}(\phi(t), \lambda)$.

First notice the usefulness of the theorem. Suppose we have an eigenvalue $\lambda$, found by solving $\phi(t) = 0$, and we want to search for eigenvectors for with eigenvalue $\lambda$. Without any other information, we may begin looking for up to $n$ of them! But instead, we consider the quotients $\phi(t)/(t - \lambda)^k$, for $k = 1, 2, \ldots$, until $\lambda$ is no longer a root of the result. This computes $\text{mult}(\phi(t), \lambda)$ for us, and gives us an upper bound on how many eigenvalues to look for.

Of particular interest is the special case when $\text{mult}(\phi(t), \lambda) = 1$. When this occurs, we know that there is at least one eigenvector $\vec{v}$, since $\phi(\lambda) = 0$, but also that there is at most one eigenvector, by the theorem above. Therefore, in this case, there is a unique eigenvector with eigenvalue $\lambda$, up to scalar multiple. Unfortunately, we will soon produce counterexamples to this for any larger value of $\text{mult}(\phi(t), \lambda)$.
Proof of the theorem. Let \( \{ \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \} \) be a basis for \( V(L, \lambda) \). Then \( L(\vec{v}_i) = \lambda \vec{v}_i \) for each \( i \). Choose vectors \( \vec{v}_{k+1}, \vec{v}_{k+2}, \ldots, \vec{v}_n \) so that \( \{ \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \} \) is a basis for \( V \). Let \( A = (a_{ij}) \) be the basis for \( L \) with respect to this matrix. Then \( A \) has the form

\[
A = \begin{bmatrix}
\lambda & 0 & 0 & \cdots & 0 & a_{1(k+1)} & a_{1(k+2)} & \cdots & a_{1n} \\
0 & \lambda & 0 & \cdots & 0 & a_{2(k+1)} & a_{2(k+2)} & \cdots & a_{2n} \\
0 & 0 & \lambda & \cdots & 0 & a_{3(k+1)} & a_{3(k+2)} & \cdots & a_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \lambda & a_{k(k+1)} & a_{k(k+2)} & \cdots & a_{kn} \\
0 & 0 & 0 & \cdots & 0 & a_{(k+1)(k+1)} & a_{(k+1)(k+2)} & \cdots & a_{(k+1)n} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & a_{n(k+1)} & a_{n(k+2)} & \cdots & a_{nn}
\end{bmatrix}.
\]

Don’t let this giant matrix scare you. All it is saying is that there is a diagonal block with \( \lambda \) along its diagonal, taking up the top \( k \times k \) portion, and beneath this block there are only 0s. The rest of the matrix is unspecified (in the same sense that the action of \( L \) on \( \vec{v}_{k+1}, \vec{v}_{k+2}, \ldots, \vec{v}_n \) is unspecified). If we take \( \det(A - tI) \) and expand by minors down the first column, then we will get \((\lambda - t)\) times a minor for the first term, and all other terms zero. Therefore, \( \phi(t) \) is the product of \((\lambda - t)\) and the determinant of \( A - tI \) with its first row and column removed. Then we do expansion by minors down the first column of this, and get the same result, and repeat, until we have removed all \( k \) of the first rows and columns. We are left with \( \phi(t) \) being of the form \((\lambda - t)^k\) times some \((n - k) \times (n - k)\) minor which yields a polynomial in \( t \). This says precisely that we can factor out at least \( k \) copies of \((\lambda - t)\) from \( \phi(t) \), or equivalently that we can factor out at least \( k \) copies of \((t - \lambda)\) from \( \phi(t) \). But since \( \text{mult}(\phi(t), \lambda) \) is the largest number of copies of \((t - \lambda)\) which can be factored out of \( \phi(t) \), we can conclude that \( k \leq \text{mult}(\phi(t), \lambda) \). Since \( k = \dim(V(L, \lambda)) \), this proves the theorem.

### 5.5 Nondiagonalizability

Recall from the first section in the chapter, the concept of a diagonal mapping. This means that, with respect to a basis \( \{ \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \} \), the mapping \( L \) has a diagonal matrix; it sends each basis element \( \vec{v}_i \) to a scalar multiple of itself, \( \lambda_i \vec{v}_i \).

**Definition.** A linear mapping \( L: V \rightarrow V \) is called **diagonalizable** if there exists a basis for \( V \) with respect to which \( L \) is a diagonal mapping. Otherwise, \( L \) is called **nondiagonalizable**.

That was the coordinate-free way of defining the concept. However, in practice, one usually is given the mapping in the form of a matrix \( A \) with respect to a basis \( \{ \vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n \} \). What would diagonalizability say about a matrix, when we do not have the diagonalizing basis handy? We have to consider all representations of the linear mapping \( L: V \rightarrow V \), \( L(\vec{v}) = A\vec{v} \), in other bases. We saw on Day 3 that when we perform a change of basis \( \{ \vec{u}_1, \ldots, \vec{u}_n \} \rightarrow \{ \vec{v}_1, \ldots, \vec{v}_n \} \), the linear mapping \( L \) gets represented by a new matrix of the form \( BAB^{-1} \), where \( B \) is an invertible matrix which essentially represents the new basis in terms of the old basis. Thus we get the following definition.
**Definition.** A square matrix $A$ is said to be *diagonalizable* if for some invertible matrix $B$, the matrix $BAB^{-1}$ is diagonal. Otherwise, $A$ is called *nondiagonalizable*.

I hope that the following is clear from the pictures that I have drawn in the preceding sections, but just to recap, I will point it out. The linear mappings (or matrices) which are diagonalizable are precisely those which admit the extremely simple geometrical description from the first subsection of today. They have a basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ on which they act just by “stretching the axes”, and they act on all other vectors in the only way which respects these simple stretches.

Since we are trying to understand all matrices, this begs the question, which ones are described by this picture? Which ones are diagonalizable? It also leads us to ask, which matrices are not diagonalizable, and what does their geometrical action look like? We will develop the machinery necessary to answer each of these questions for any particular matrix.

Let us first look at the reasons that a matrix can fail to be diagonalizable.

- complex eigenvalues. notable examples include $SO(2) \cong S^1$.
- $\dim V(\lambda) < \text{mult}(\phi(t), \lambda)$, and nilpotence
- any others i’m missing?

exponential matrix, applications (diff eqs)
6 applications/solidification

markov chains.

graph theory.
7 inner products

dot product, transpose, orthogonality/normality, ortho projs, Gram-Schmidt, least squares, orthogonal transforms, symmetric maps.

**exploration:** unitary, hermitian; spectral thm.
8 applications/solidification

asdf

9 more explorations

projective geometry.