MATH 235 Class 1:
Review: Definition and Examples of Vector Spaces

September 9, 2021

If you asked a mathematician what linear algebra is all about, they would tell you that the goal is to study and understand *vector spaces*. Loosely speaking, a vector space is any set of mathematical objects that you can add together and multiply by scalars, in such a way that these operations behave in familiar ways. The conditions we impose on these operations are consistent with our experience working with vectors in $\mathbb{R}^n$ and $\mathbb{C}^n$ in previous linear algebra courses.

Once a vector space is defined, there are all sorts of bells and whistles you can study along with it. The nonempty subsets of vector spaces preserving the addition and scalar multiplication structure will be called *subspaces*. The functions between vector spaces that preserve the addition and scalar multiplication will be called *linear transformations*. Vector spaces can also be equipped with *inner products*, which allow us to define the *length* of objects in a vector space, and also the *angle* between two objects. Looking at linear transformations that preserve the inner product between vectors opens a whole new subject area for exploration.

In broad strokes, our course will be all about studying these ideas. In your previous linear algebra course, you have already encountered many of these ideas in the setting of $\mathbb{R}^n$ and $\mathbb{C}^n$. To begin this course, it's worth spending some time defining and exploring new examples of vector spaces.

Vector Spaces: The Definition

The formal mathematical definition of a vector space is given below. When you read it over, keep in mind the intuition that we want a set of objects that we can add together and multiply by scalars, holding the familiar examples of $\mathbb{R}^n$ and $\mathbb{C}^n$ in mind as well.

Definition 1.1: Vector Space

Let $\mathbb{F}$ denote either the set of real numbers, or complex numbers. A *vector space* over $\mathbb{F}$ is a set V, equipped with two operations, *addition* and *scalar multiplication*. The elements of V are usually called *vectors*.

The addition operation is often written $+$, and it takes two elements $v, w \in V$ and outputs a new element $v + w \in V$, called the *sum* of the two vectors. Addition of vectors obeys four key properties:

1. For all $v, w, x \in V$, we have $(v + w) + x = v + (w + x)$. (Associativity of addition).
2. There is a vector $0 \in V$, called a *zero vector*, such that $v + 0 = v$ for all $v \in V$. (Existence of a zero vector).
3. For all $v \in V$, there is a vector $-v \in V$ such that $v + (-v) = 0$. (Existence of an *additive inverse*).
4. For all $v, w \in V$, we have $v + w = w + v$. (Commutativity of addition).

The scalar multiplication operation is often written $\cdot$, and it takes an element $c \in \mathbb{F}$ (usually called a *scalar*), along with a vector $v \in V$, and outputs a new element $c \cdot v \in V$, called the *scalar multiplication of v by c*. We often leave out the multiplication symbol and write cv instead of $c \cdot v$. Scalar multiplication of vectors also obeys four key properties:

1. For all scalars $c_1, c_2 \in \mathbb{F}$ and all vectors $v \in V$, we have $c_1(c_2 v) = (c_1 c_2)v$.

2. For all vectors $v \in V$, we have $1 \cdot v = v$.
3. For all scalars $c_1, c_2 \in \mathbb{F}$ and all vectors $v \in V$, we have $(c_1 + c_2)v = c_1 v + c_2 v$.
4. For all scalars $c \in \mathbb{F}$ and all vectors $v, w \in V$, we have $c(v + w) = cv + cw$.

In this course, the set of scalars will always be either the real numbers or the complex numbers. However, the definition of vector space works perfectly well if the set of scalars is replaced with the rational numbers, $\mathbb{Q}$, or even the integers modulo p, $\mathbb{Z}_p$, where p is a prime.

All in all, the only thing we ask of our set of scalars is that the scalars form a *field*. Loosely speaking, a field is a set where you can add, subtract, multiply, and divide elements by each other in "familiar" ways. We won't say much more about this, but you should be aware that the set of scalars in a vector space does not always have to be $\mathbb{R}$ or $\mathbb{C}$!

Vector Spaces: Fundamental Examples

Let's now introduce several important examples of vector spaces, ones that we will be working with throughout the course. We begin with the example that motivated the definition: vectors in $\mathbb{R}^n$ and $\mathbb{C}^n$. Throughout the rest of these notes, we will let $\mathbb{F}$ stand in for either the set of real numbers or the set of complex numbers (with both working equally well in each context where $\mathbb{F}$ appears).

Example 1.1

The set of $n \times 1$ column vectors with entries in $\mathbb{F}$, denoted $\mathbb{F}^n$, forms a vector space. The elements of $\mathbb{F}^n$ take the form

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

where all of $v_1, v_2, \ldots, v_n$ belong to $\mathbb{F}$. Vector addition is defined component-wise, so that

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix},$$

for all vectors $\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{F}^n$. Finally, given a scalar $c \in \mathbb{F}$ and a vector $\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{F}^n$, scalar multiplication is defined by

$$c \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{pmatrix}.$$

Strictly speaking, we still have to verify that vector addition and scalar multiplication satisfy the eight conditions described in Definition 1.1. Most of that checking will be left to you (though the facts should not be surprising!). Note that a zero vector is $\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, and given that $v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$, we have that $-v = \begin{pmatrix} -v_1 \\ -v_2 \\ \vdots \\ -v_n \end{pmatrix}$.

We verify commutativity of addition and the third condition on scalar multiplication. Given $v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ and $w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ in $\mathbb{F}^n$, note that

$$v + w = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} = \begin{pmatrix} w_1 + v_1 \\ w_2 + v_2 \\ \vdots \\ w_n + v_n \end{pmatrix} = w + v.$$

Similarly, given scalars $c_1, c_2 \in \mathbb{F}$ and $v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ in $\mathbb{F}^n$, we have

$$(c_1 + c_2)v = \begin{pmatrix} (c_1 + c_2)v_1 \\ (c_1 + c_2)v_2 \\ \vdots \\ (c_1 + c_2)v_n \end{pmatrix} = \begin{pmatrix} c_1v_1 + c_2v_1 \\ c_1v_2 + c_2v_2 \\ \vdots \\ c_1v_n + c_2v_n \end{pmatrix} = \begin{pmatrix} c_1v_1 \\ c_1v_2 \\ \vdots \\ c_1v_n \end{pmatrix} + \begin{pmatrix} c_2v_1 \\ c_2v_2 \\ \vdots \\ c_2v_n \end{pmatrix} = c_1v + c_2v.$$

Another related example is the vector space of matrices with entries in $\mathbb{F}$.

Example 1.2

Recall that for any positive integers m and n, an $m \times n$ *matrix with entries in* $\mathbb{F}$ is an array of the form

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

where all the *entries* a_{ij} of the matrix belong to $\mathbb{F}$. The set of all such matrices will be denoted by $M_{m\times n}(\mathbb{F})$.

Here, vector addition is the ordinary entry-wise addition of matrices:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{pmatrix}.$$

Scalar multiplication is also performed entry-wise:

$$c \cdot \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{pmatrix}.$$

You are strongly encouraged to check that this makes $M_{m\times n}(\mathbb{F})$ into a vector space, according to Definition 1.1.

Next, we introduce a couple examples that were *not* studied at length in your previous linear algebra course.

Example 1.3

A *polynomial function* with coefficients in $\mathbb{F}$ is a function $p : \mathbb{F} \to \mathbb{F}$ of the form

$$p(x) = a_0 + a_1 x + \ldots + a_n x^n,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{F}$. If a_n is non-zero, then we say that the *degree* of $p(x)$ is equal to n, and we write $n = \deg(p(x))$.

If $p(x) = a_0$ for some $a_0 \in \mathbb{F}$, we say that $p(x)$ is a *constant polynomial*. If $p(x) = 0$, then we say that $p(x)$ is the *zero polynomial*. We do not assign a degree to the zero polynomial.

We let $P_n(\mathbb{F})$ denote the set of polynomial functions with coefficients in $\mathbb{F}$, having degree at most n, and we let $P(\mathbb{F})$ denote the set of all polynomial functions with coefficients in $\mathbb{F}$, without any restrictions on degree. We can turn either $P_n(\mathbb{F})$ or $P(\mathbb{F})$ into a vector space by defining

$$(a_0 + a_1 x + \cdots + a_n x^n) + (b_0 + b_1 x + \cdots + b_n x^n) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$$

for vector addition, and

$$c(a_0 + a_1 x + \cdots + a_n x^n) = (ca_0) + (ca_1)x + \cdots + (ca_n)x^n$$

for scalar multiplication.

Again, you are strongly encouraged to check that this makes $P_n(\mathbb{F})$ and $P(\mathbb{F})$ into vector spaces (what is a zero vector, and what is the additive inverse of a given polynomial?)

Example 1.4

For an example that greatly generalizes Example 1.3, consider the set $\mathcal{F}$ of all functions $f : \mathbb{R} \to \mathbb{R}$. In other words, the vectors consist of all functions that take elements of $\mathbb{R}$ as input, and give elements of $\mathbb{R}$ as output. Some examples of vectors in $\mathcal{F}$ are $e^x, \sin x$, and also all of $P(\mathbb{R})$ from Example 2.3.

Vector addition and scalar multiplication are both defined point-wise. In other words, given $f, g \in \mathcal{F}$, we set

$$(f+g)(x) = f(x) + g(x)$$

for all $x \in \mathbb{R}$, and given a scalar $c \in \mathbb{R}$, we set

$$(cf)(x) = c \cdot f(x)$$

for all $x \in \mathbb{R}$.

Once more, please check that this turns $\mathcal{F}$ into a vector space (again, what is the zero vector, and what is the additive inverse of a given function?)

Vector Spaces: Unusual Examples

We now move on to a couple more interesting examples of vector spaces. The first will illustrate checking the full set of vector space conditions, and the second will hopefully illustrate the usefulness of looking at this more general definition of vector space.

Example 1.5

Consider the set V of all positive real numbers. We define a vector addition operation $\oplus$ and a scalar multiplication operation $\odot$ on V by setting $x \oplus y = xy$ and $c \odot x = x^c$ for all $x, y \in V$ and all $c \in \mathbb{R}$. Note that xy is a positive real number when x and y are, and x^c is a positive real number for all exponents c when x is a positive real number, so vector addition and scalar multiplication both always return elements of V, as required by the definition.

Let us check that these operations satisfy the conditions required of a vector space. Given $x, y, z \in V$, first note that

$$(x \oplus y) \oplus z = (xy) \oplus z = (xy)z = x(yz) = x \oplus (yz) = x \oplus (y \oplus z)$$

$$x \oplus y = xy = yx = y \oplus x,$$

showing that addition in V is associative and commutative. The zero vector is the element $1 \in V$ in this case, because for any $x \in V$, we have

$$x \oplus 1 = (x)(1) = x.$$

With this in mind, the additive inverse of a vector $x \in V$ is the element $\frac{1}{x} \in V$, since

$$x \oplus (1/x) = x(1/x) = 1.$$

(Why is $\frac{1}{x} \in V$ when $x \in V$?)

As for the scalar multiplication conditions, if we are given $x, y \in V$ and scalars $c_1, c_2 \in \mathbb{R}$, note that

$$\begin{aligned}
c_1 \odot (c_2 \odot x) &= c_1 \odot (x^{c_2}) = (x^{c_2})^{c_1} = x^{c_1 c_2} = (c_1 c_2) \odot x \\
1 \odot x &= x^1 = x \\
(c_1 + c_2) \odot x &= x^{c_1 + c_2} = (x^{c_1})(x^{c_2}) = x^{c_1} \oplus x^{c_2} = (c_1 \odot x) \oplus (c_2 \odot x) \\
c \odot (x \oplus y) &= c \odot (xy) = (xy)^c = (x^c)(y^c) = x^c \oplus y^c = (c \odot x) \oplus (c \odot y).
\end{aligned}$$

This verifies that V, equipped with the operations given above, is a vector space.

For those going on to study the subject of *abstract algebra*, the next example is a special case of a very useful idea:

Example 1.6

Consider the set $\mathbb{C}$ of complex numbers, but as a vector space over the *real numbers*. In other words, rather than considering $\mathbb{C}$ as a vector space with complex scalars, as we usually do, we restrict our scalars to be real numbers only.

Vector addition is defined to be the usual addition in the complex numbers:

$$(a+bi)+(c+di)=(a+c)+(b+d)i.$$

Scalar multiplication of a complex number by a real number is also defined "as expected". In other words, for any $a+bi \in \mathbb{C}$ and any $c \in \mathbb{R}$, we take

$$c(a+bi)=ca+(cb)i.$$

You can verify that $\mathbb{C}$ is a vector space (with real scalars) under the definition given here.

MATH 235 Class 2:
Properties of Vector Spaces, Introduction to Subspaces

September 14, 2021

One of the big recurring themes in theoretical mathematics is *generalization*. We begin with some known familiar examples (like $\mathbb{R}^n$ and $\mathbb{C}^n$) and attempt to come up with a list of key properties, or *axioms*, that capture their "essence" (like we see in the definition of a vector space). If we have done this well, then many other interesting structures satisfy these axioms, and these structures can be manipulated in ways that "feel" a lot like our manipulations of the more familiar structures.

In particular, it is often possible to prove a lot of properties of the familiar examples directly from the axioms. This tells us that *all* structures that satisfy the axioms share these familiar properties, which adds to the ways in which these structures resemble the familiar ones. To begin these notes, we will derive several familiar properties of $\mathbb{R}^n$ and $\mathbb{C}^n$ directly from the vector space axioms, both as an illustration of how the axioms can be used, and also as an illustration of the properties that all vector spaces have in common.

Properties of Vector Spaces

If you look back at the conditions in the definition of a vector space, you will notice that there is no claim about *uniqueness* of a zero vector, or of the inverse of a vector under vector addition. How do we know that a vector space cannot have multiple zero vectors, or that a given vector cannot have multiple inverses? Here, we set about to show the uniqueness of both of these objects from the properties of vector spaces.

Theorem 2.1

In any vector space, the zero vector is unique. In other words, if there are two vectors $0, z \in V$, having the property that $v + 0 = v$ and $v + z = v$ for all $v \in V$, then $z = 0$.

Proof. Suppose V is a vector space having two vectors $0, z$ that satisfy the condition given in the statement of the theorem. Knowing that $v+0 = v$ for all $v \in V$, we can take $z = v$ here to get that $z+0 = z$. Similarly, knowing that $v + z = v$ for all $v \in V$, we can take $v = 0$ to get $0 + z = 0$. By commutativity of addition in V,

$$z = z + 0 = 0 + z = 0,$$

proving that $z = 0$. $\square$

Given that the zero vector in a vector space is unique, we usually use the symbol 0 to denote the zero vector, and no ambiguity will arise about which vector we are talking about!

Theorem 2.2

In any vector space, the additive inverse of a vector is unique. In other words, for each $v \in V$, if there are vectors $w, x \in V$ such that $v + w = 0$ and $v + x = 0$, then $w = x$.

Proof. Let V be a vector space, and suppose we are given $v \in V$. Suppose also that we have $w, x \in V$ such that $v + w = 0$ and $v + x = 0$. Applying the definition of zero vector, associativity of addition, and commutativity of addition,

$$\begin{aligned}
x &= x + 0 \\
&= x + (v + w) \\
&= (x + v) + w \\
&= (v + x) + w \\
&= 0 + w \\
&= w + 0 \\
&= w.
\end{aligned}$$

□

Similarly to the zero vector, knowing that the additive inverse of a vector $v \in V$ is unique, we usually use the symbol $-v$ to denote the inverse of v. Again, no ambiguity will arise regarding which vector we are talking about. This uniqueness also enables us to define vector subtraction:

Definition 2.1: Subtraction of Vectors

Let V be a vector space, and suppose we are given $v, w \in V$. The *difference* of these two vectors, denoted $v - w$, is equal to the vector $v + (-w)$, obtained by adding v to the unique additive inverse of w.

Next, we prove a couple facts regarding the way scalar multiplication behaves:

Lemma 2.1

Let V be a vector space. The following properties hold:

(1) For all $v \in V$, we have $0 \cdot v = 0$. (Note: the zero on the left-hand side is the zero scalar in $\mathbb{F}$, and the zero on the right-hand side is the zero vector in V.)

(2) For all scalars $c \in \mathbb{F}$, we have $c \cdot 0 = 0$. (Here, the zero on both sides is the zero vector in V.)

(3) Given $c \in \mathbb{F}$ and $v \in V$ such that $c \cdot v = 0$, either $c = 0$ or $v = 0$.

Proof.

Proof of (1): Let $v \in V$ be arbitrary. Applying properties of the scalar 0 and one of the distributive properties of vector spaces, we have

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

Taking $0 \cdot v = 0 \cdot v + 0 \cdot v$ and adding $-(0 \cdot v)$ on both sides, we get

$$\begin{aligned}
0 \cdot v - 0 \cdot v &= 0 \cdot v + 0 \cdot v - 0 \cdot v \\
0 &= 0 \cdot v,
\end{aligned}$$

as we needed to show.

Proof of (2): Similar to the proof of (1), and left as an excellent exercise for you!

Proof of (3): Suppose we have $c \in \mathbb{F}$ and $v \in V$ such that $c \cdot v = 0$. If $c = 0$, then we are done. Otherwise, we know that $\frac{1}{c} \in \mathbb{F}$ exists, and we can multiply both sides of the equation by $\frac{1}{c}$ and apply part (2) of this lemma:

$$\begin{aligned}
\frac{1}{c} \cdot (c \cdot v) &= \frac{1}{c} \cdot 0 \\
(\frac{1}{c} \cdot c) \cdot v &= 0 \\
1 \cdot v &= 0 \\
v &= 0.
\end{aligned}$$

This completes the proof. □

For our final example of vector space axiom proofs (for now), let's establish a couple of properties regarding the additive inverses of vectors:

Lemma 2.2

Let V be a vector space. We have the following:

(1) For all $v \in V$, we have that $(-1) \cdot v = -v$, the additive inverse of v.

(2) More generally, for all $c \in \mathbb{F}$ and $v \in V$, we have $c \cdot (-v) = (-c) \cdot v = -(c \cdot v)$.

(3) Scalar multiplication distributes over subtraction: for all $c \in \mathbb{F}$ and $v, w \in V$, we have $c(v - w) = cv - cw$.

Proof.

Proof of (1): By Theorem 2.2, the additive inverse of a vector is unique. Thus, if we can prove that $(-1) \cdot v$ is an additive inverse of v, it will follow immediately that $(-1) \cdot v = -v$. This can be done by direct computation, together with the vector space axioms:

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

where the last equality follows from part (1) of Lemma 2.1. This computation verifies that $(-1) \cdot v$ is an additive inverse of v, as required.

Proof of (2): Similarly to part (1), this can be proved by showing that $c \cdot (-v)$ and $(-c) \cdot v$ are both additive inverses of $c \cdot v$. The proof of this is left as another excellent exercise for you!

Proof of (3): Suppose we are given $c \in \mathbb{F}$ and vectors $v, w \in V$. Proceeding by definition of subtraction, and using part (2) of this lemma, we get

$$c(v - w) = c(v + (-w)) = cv + c(-w) = cv + (-cw) = cv - cw.$$

□

Subspaces: Definition and Examples

Next, we move on to the idea of a *subspace* of a vector space. In words, a subset S of a vector space V is a subspace when S also satisfies all the conditions required of a vector space, using the same addition and scalar multiplication operations as we have on V.

Definition 2.2: Subspace

Let V be a vector space over $\mathbb{F}$. A non-empty subset S of V is called a *subspace* of V if S, equipped with the same addition and scalar multiplication operations as in V, itself forms a vector space.

As you may already appreciate, it can be tedious to check all eight of the vector space axioms. Happily, for subspaces there is a notable shortcut, which we now seek to prove before giving any examples of subspaces. You may remember from your previous linear algebra course that a subspace of $\mathbb{F}^n$ was a nonempty subset that is closed under addition and scalar multiplication. As it turns out, this fact is preserved in general, and makes for a convenient Subspace Test:

Theorem 2.3

(The Subspace Test). Let V be a vector space (over $\mathbb{F}$). A subset S of V is a subspace of V if and only if the following three conditions hold:

(1) The zero vector of V belongs to S.

(2) The set S is *closed under addition*: given $v, w \in S$, we have that $v + w \in S$.

(3) The set S is *closed under scalar multiplication:* given $v \in S$ and $c \in \mathbb{F}$, we have that $cv \in S$.

Proof. First, assume that S is a subspace of V. By definition, this means that S is a non-empty subset of V that satisfies all of the conditions given in the definition of a vector space, with the same addition and scalar multiplication as in V. In particular, when two vectors in S are added together, the result must be in S, so condition (2) is satisfied. Similarly, when a vector in S is multiplied by a scalar, the result must be in S, so condition (3) is satisfied.

Finally, since S is non-empty, there is a vector $v \in S$. Using closure under scalar multiplication, we know that $0 \cdot v$ must belong to S. But by Lemma 2.1, part (1), this is equal to the zero vector in V, so condition (1) is satisfied.

Next, suppose that S is a subset of V satisfying the three conditions of this theorem. By condition (1), the set S is non-empty. By conditions (2) and (3), the result of adding together two vectors in S, or multiplying a vector in S by a scalar, is another vector in S. Knowing this, let's look at the four properties of vector addition in V:

1. For all $v, w, x \in V$, we have $(v + w) + x = v + (w + x)$. (Associativity of addition).
2. There is a vector $0 \in V$, called a *zero vector*, such that $v + 0 = v$ for all $v \in V$. (Existence of a zero vector).
3. For all $v \in V$, there is a vector $-v \in V$ such that $v + (-v) = 0$. (Existence of an *additive inverse*).
4. For all $v, w \in V$, we have $v + w = w + v$. (Commutativity of addition).

Note that associativity and commutativity of addition automatically hold for all vectors in S, because they hold for all vectors in the larger set V. By condition (1), the zero vector of V belongs to S, and for all vectors $v \in S$, we have $v + 0 = v$, so existence of a zero vector holds in S. By closure under scalar multiplication, given any vector $v \in S$, we know that $(-1) \cdot v \in S$. By Lemma 2.2, part (1), this is equal to $-v$, the additive inverse of v in V. Thus existence of additive inverses holds in S as well.

Finally, let's look at the four properties of scalar multiplication in V:

1. For all scalars $c_1, c_2 \in \mathbb{F}$ and all vectors $v \in V$, we have $c_1(c_2 v) = (c_1 c_2)v$.
2. For all vectors $v \in V$, we have $1 \cdot v = v$.

3. For all scalars $c_1, c_2 \in \mathbb{F}$ and all vectors $v \in V$, we have $(c_1 + c_2)v = c_1 v + c_2 v$.
4. For all scalars $c \in \mathbb{F}$ and all vectors $v, w \in V$, we have $c(v + w) = cv + cw$.

Note that all of these properties already hold for all vectors in V and all scalars in $\mathbb{F}$, so they hold for all vectors in the subset S automatically, meaning S satisfies these four conditions as well.

In conclusion, if S a subset of V satisfying the three conditions outlined in this theorem, then S is a subspace of V. □

Let's now put the Subspace Test to use in verifying a few examples of subspaces. First, let's recall some old definitions:

Definition 2.3: Matrix Transpose

Recall that given a matrix $A \in M_{m\times n}(\mathbb{F})$ with entries

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

the *transpose* matrix is the matrix $A^T \in M_{n\times m}(\mathbb{F})$ with entries

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

Informally, the transpose matrix is obtained by turning all rows of A into columns and all columns into rows. In particular, note that when A is an $m \times n$ matrix, then its transpose is an $n \times m$ matrix.

For example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

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

Now that the transpose has been re-introduced, we proceed to define a type of matrix that will turn out to be very important in this course:

Definition 2.4: Symmetric Matrices

A square matrix $A \in M_{n\times n}(\mathbb{F})$ is called *symmetric* if $A^T = A$.

For example, the matrix $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}$ is symmetric, as is $\begin{pmatrix} 1 & 0 & 1+i \\ 0 & 1 & 0 \\ 1+i & 0 & 1 \end{pmatrix}$. One of the reasons we're interested in symmetric matrices at this point is that they provide a great example of a subspace.

Example 2.1

Let V denote the vector space $M_{n\times n}(\mathbb{F})$, and let S denote the set of symmetric matrices in V. We will verify that S is a subspace of V using the Subspace Test.

First, let's check that the zero vector belongs to S. In this case, the zero vector is the zero matrix Z, with all entries equal to 0. You can check immediately that $Z^T = Z$, so that the zero matrix is symmetric.

Next, we need to check that S is closed under addition and scalar multiplication. These follow from more general properties of matrix transposes (which you are encouraged to verify!). More specifically, if A and B are $m \times n$ matrices and $c \in \mathbb{F}$ is a scalar, we have

$$(A+B)^T = A^T + B^T$$
$$(cA)^T = cA^T.$$

Applying these, assume that $A, B \in S$ and that $c \in \mathbb{F}$ is a scalar. By assumption, this means $A^T = A$ and $B^T = B$. Thus

$$(A+B)^T = A^T + B^T = A + B,$$

proving that $A + B \in S$. Hence S is closed under addition. Similarly,

$$(cA)^T = cA^T = cA,$$

proving that $cA \in S$. Thus S is closed under scalar multiplication.

By the Subspace Test, we have now shown that S is a subspace of V.

Next, we'll take up another example, from the vector space of polynomials over $\mathbb{F}$. Let's make another definition that we'll use in the example.

Definition 2.5: Evaluation of a Polynomial

Suppose we are given $p(x) \in P(\mathbb{F})$, and suppose

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

for some $a_0, a_1, \ldots, a_n \in \mathbb{F}$. Given $c \in \mathbb{F}$, we define the *evaluation of p at c* to be

$$p(c) = a_0 + a_1 c + \cdots + a_n c^n \in \mathbb{F}.$$

You are encouraged to verify that evaluation satisfies some familiar properties. For instance, the evaluation of $p(x) + q(x)$ at c is equal to $p(c) + q(c)$, and given a scalar $a \in \mathbb{F}$, the evaluation of $ap(x)$ at c is $ap(c)$. This helps greatly with our next example:

Example 2.2

Consider the vector space $V = P(\mathbb{F})$, and let $S = \{p(x) \in V : p(1) = 0\}$. We will verify that S is a subspace of V using the Subspace Test.

First, we verify that the zero vector belongs to S. Here, the zero vector is the zero polynomial, defined by $z(x) = 0$ for all $x \in \mathbb{F}$. This certainly satisfies $z(1) = 0$, so S contains the zero vector.

Next, we verify that S is closed under addition and scalar multiplication. If we are given $p(x), q(x) \in S$ and a scalar c, then evaluating $p(x) + q(x)$ at 1 gives

$$p(1) + q(1) = 0 + 0 = 0$$

and evaluating $cp(x)$ at 1 gives

$$cp(1) = c(0) = 0.$$

This proves that $p(x) + q(x) \in S$ and that $cp(x) \in S$, establishing the two closure properties.

MATH 235 Class 3:
Review: Matrix Multiplication and Systems of Linear Equations

September 16, 2021

In your previous linear algebra course, matrices were defined and studied for two main reasons. The first of these reasons is that matrices can be used to encode the coefficients in a linear system of equations. Once this is done, there are systematic ways to manipulate the matrix of coefficients via *elementary row operations*, in order to determine the set of solutions in a standard and efficient manner. We will review these procedures here, but our review will be quick. You are encouraged to consult your notes from your previous linear algebra course for additional examples and details!

The second of these reasons is that $m \times n$ matrices can be associated with linear transformations from $\mathbb{F}^n$ to $\mathbb{F}^m$. In a very precise way, the matrix associated to a linear transformation tells you everything you need to know about manipulating that transformation. Moreover, there is a way to define multiplication of matrices that is directly compatible with composition of the corresponding linear transformations. We will recap the correspondence between linear transformations and matrices in the reading for Class 7. For now, we focus only on re-introducing the definition of matrix multiplication, as well as some of its main properties.

Matrix Multiplication

Just as we can add two matrices of the same size, we can also multiply two matrices of compatible size. The official definition is captured below.

Definition 3.1: Matrix Multiplication

Let $A \in M_{m\times n}(\mathbb{F})$, and let $B \in M_{n\times p}(\mathbb{F})$. (Notice here that the number of columns of A is equal to the number of rows of B.) In this case, we can define a *matrix product* $C = AB$, which is a matrix in $M_{m\times p}(\mathbb{F})$. If the entries of A are a_{ij} (where a_{ij} is the entry in row i and column j), and the entries of B are b_{ij}, then the entries c_{ij} of the matrix C are determined from those of A and B by

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

This formula holds for all pairs of indices (i, j), with $1 \le i \le m$ and $1 \le j \le p$.

Let's take a look at a couple computational examples, just to get another feel for the definition.

Example 3.1

Suppose we set $A = \begin{pmatrix} 2 & 1 \\ 3 & -1 \\ 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$. Since A is a 3×2 matrix and B is a 2×2 matrix, the product AB will be defined and be a 3×2 matrix. According to the definition, the entry c_{11} of the matrix $C = AB$ should be

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} = (2)(1) + (1)(2) = 4.$$

Similarly, we compute the other five entries of AB:

$$\begin{aligned}
c_{12} &= a_{11}b_{12} + a_{12}b_{22} = (2)(2) + (1)(-1) = 3\\
c_{21} &= a_{21}b_{11} + a_{22}b_{21} = (3)(1) + (-1)(2) = 1\\
c_{22} &= a_{21}b_{12} + a_{22}b_{22} = (3)(2) + (-1)(-1) = 7\\
c_{31} &= a_{31}b_{11} + a_{32}b_{21} = (0)(1) + (2)(2) = 4\\
c_{32} &= a_{31}b_{12} + a_{32}b_{22} = (0)(2) + (2)(-1) = -2.
\end{aligned}$$

Altogether, we get $AB = \begin{pmatrix} 4 & 3 \\ 1 & 7 \\ 4 & -2 \end{pmatrix}$.

In practice, we often write out the computation of the matrix product all in one go, as in the next example.

Example 3.2

Suppose $A = \begin{pmatrix} 1 & i \\ 1+i & 2 \end{pmatrix}$. Let's compute A^2, the matrix product of A with itself.

$$\begin{aligned}
A^2 &= \begin{pmatrix} 1 & i \\ 1+i & 2 \end{pmatrix}\begin{pmatrix} 1 & i \\ 1+i & 2 \end{pmatrix}\\
&= \begin{pmatrix} (1)(1) + i(1+i) & (1)(i) + i(2) \\ (1+i)(1) + 2(1+i) & (1+i)i + 2(2) \end{pmatrix}\\
&= \begin{pmatrix} i & 3i \\ 3+3i & 3+i \end{pmatrix}.
\end{aligned}$$

One important special case of a matrix product is the product of a matrix $A \in M_{m\times n}(\mathbb{F})$ with a vector $\mathbf{x} \in \mathbb{F}^n$, where we consider $\mathbf{x}$ to be an $n \times 1$ matrix. If $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, then notice that $A\mathbf{x}$ will be an $m \times 1$ matrix (i.e. a vector in $\mathbb{F}^m$). The entries of this matrix are

$$A\mathbf{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.$$

As we will see below, this conveniently allows us to convert systems of linear equations into matrix-vector form, so that we can use the various properties of matrix multiplication to manipulate the system.

Before we get to that, however, let's state several useful (and probably expected) properties of matrix multiplication. As part of this theorem, we will need to recall a key definition:

Definition 3.2: Identity Matrix

The $n \times n$ *identity matrix* I_n (sometimes written I, when n is already clear from context), is the

$n \times n$ matrix with 1s on the diagonal, and 0 entries everywhere else. In other words,

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

And now, here are some properties of matrix multiplication that we will use frequently:

Theorem 3.1

(1) For all matrices $A \in M_{m\times n}(\mathbb{F})$, $B \in M_{n\times p}(\mathbb{F})$, and $C \in M_{p\times r}(\mathbb{F})$, we have $(AB)C = A(BC)$. (Matrix multiplication is associative.)

(2) For all matrices $A \in M_{m\times n}(\mathbb{F})$ and $B, C \in M_{n\times p}(\mathbb{F})$, we have $A(B + C) = AB + AC$. (Left distributivity).

(3) For all matrices $A, B \in M_{m\times n}(\mathbb{F})$, and $C \in M_{n\times p}(\mathbb{F})$, we have $(A + B)C = AC + BC$. (Right distributivity).

(4) For all matrices $A \in M_{m\times n}(\mathbb{F})$, we have $I_m A = A$ and $AI_n = A$, where I_m and I_n are identity matrices as defined above. We also have $Z_m A = Z_m$ and $AZ_n = Z_n$, where Z_m and Z_n are $m \times m$ and $n \times n$ zero matrices.

(5) For all matrices $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times p}(\mathbb{F})$ and all scalars $c \in \mathbb{F}$, we have

$$c(AB) = (cA)B = A(cB).$$

(6) If $A, B \in M_{n\times n}(\mathbb{F})$, it is **not** necessarily true that $AB = BA$.

Proof. We will prove (2), part of (4), and (6), leaving the rest as exercises for you.

Proof of (2): Given matrices A, B, C as indicated, suppose their entries are given by a_{ij}, b_{ij}, c_{ij}, respectively. Note that both $A(B + C)$ and $AB + AC$ will be $m \times p$ matrices. For $1 \le i \le m$ and $1 \le j \le p$, we will verify that the (i, j) entry of $A(B + C)$ agrees with the same entry of $AB + AC$.

By definition of matrix multiplication, the (i, j) entry of $A(B + C)$ will be

$$\begin{aligned} \sum_{k=1}^{n} a_{ik}(B + C)_{kj} &= \sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj}) \\ &= \sum_{k=1}^{n} (a_{ik}b_{kj} + a_{ik}c_{kj}) \\ &= \sum_{k=1}^{n} a_{ik}b_{kj} + \sum_{k=1}^{n} a_{ik}c_{kj} \\ &= (AB)_{ij} + (AC)_{ij}, \end{aligned}$$

which is the (i, j) entry of $AB + AC$. This verifies that $A(B + C) = AB + AC$.

Proof of (4), identity matrix portion: Suppose A is an $m \times n$ matrix. If we use the definition of matrix multiplication to compute the (i,j) entry of I_mA, for $1 \le i \le m$ and $1 \le j \le n$, we get

$$\sum_{k=1}^{m} (I_m)_{ik} a_{kj}.$$

Now, if $i \neq k$, then the (i,k) entry of the identity matrix I_m is equal to 0, and if $i = k$, then the (i,k) entry is equal to 1. Hence all the terms in the sum above are equal to 0 except for the term where $k = i$, and that term is equal to $1 \cdot a_{ij} = a_{ij}$. Hence the (i,j) entry of I_mA is equal to the (i,j) entry of A, for all applicable indices i and j. This proves that $I_mA = A$. A similar argument shows that $AI_n = A$.

Proof of (6): There are many counter-examples to commutativity of multiplication for square matrices. Picking one 2×2 example, notice that if $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, observe that

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = B,$$

while

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

the 2×2 zero matrix. □

Solving Systems of Linear Equations

One of the most common computational tasks in linear algebra is solving a system of linear equations.

Definition 3.3: System of Linear Equations; Homogeneous and Inhomogeneous System

A *system of linear equations* is a simultaneous set of equations of the form

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

where the quantities $x_1, \ldots, x_n$ are the *unknowns*, and the numbers a_{ij} and b_i are given elements of $\mathbb{F}$. We would like to determine what assignment of values in $\mathbb{F}$ to $x_1, \ldots, x_n$, if any, satisfy all m equations. Any such assignment of values is referred to as a *solution* to the system.

If all of the numbers $b_1, \ldots, b_m$ are zero, the system is called *homogeneous*. Otherwise, the system is called *inhomogeneous*.

Example 3.3

As an illustration of the definition above, the system

$$\begin{cases} 2x_1 - x_2 + x_3 = 0 \\ x_1 + 4x_2 - 2x_3 = 0 \end{cases}$$

is a homogeneous linear system of 2 equations in 3 unknowns, and

$$\begin{cases} x_1 + 2x_2 = 0 \\ x_1 - x_2 = -1 \\ 2x_1 + x_2 = 1 \end{cases}$$

is an inhomogeneous linear system of 3 equations in 2 unknowns.

Remarks

Notice that the system of equations (1) in Definition 3.3 above can be re-written in matrix-vector form. If we set

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in M_{m\times n}(\mathbb{F}),$$

and set

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$$

and

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{F}^m,$$

then the system of equations can be written as $A\mathbf{x} = \mathbf{b}$, since the matrix product $A\mathbf{x}$ has the left-hand side of the m equations as its entries. In this context, A is often called the *coefficient matrix* of the system.

Applying this translation to the examples given in Example 3.3, we end up with

$$\begin{pmatrix} 2 & -1 & 1 \\ 1 & 4 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 2 \\ 1 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

When it comes to solving a system of linear equations, this can be done systematically by manipulating the *augmented coefficient matrix* of the system. In the notation of Definition 3.3, the augmented coefficient matrix would be

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right)$$

More specifically, we seek to perform *elementary row operations* on this matrix to put it into either *row echelon form* (REF) or *reduced row echelon form* (RREF). Since this is review from your previous linear algebra course, we will only recall the definitions here, without explicitly describing the algorithm to put a matrix into one of these forms. For that, please review the concept of *Gauss-Jordan elimination* from your previous linear algebra course!

Definition 3.4: Elementary Row Operation

Suppose $A \in M_{m \times n}(\mathbb{F})$. An *elementary row operation* is one of three types of operation we can perform on the entries of A. The three types of operation are given as follows:

(1) Swap two distinct rows of A, say row i and row j.

(2) Multiply all entries of a fixed row of A by a **non-zero** scalar $c \in \mathbb{F}$.

(3) Take some row of A, say row j, and add c times row i to it, for some scalar $c \in \mathbb{F}$ and where $i \neq j$.

All of these operations are *invertible*; in other words, having obtained a matrix A_1 from A by applying one of these operations, there is another elementary row operation, of the same type, that can be applied to A_1 to get A back (why?)

If B is a matrix that can be obtained from A in a finite number of elementary row operations, then we say that A and B are *row equivalent* to each other.

The important fact here is that if A and B are row equivalent to each other, then the systems of equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same solutions $\mathbf{x} \in \mathbb{F}^n$. An informal argument can be given as follows: any solution to a system of equations $A\mathbf{x} = \mathbf{0}$ remains a solution to the system $A_1\mathbf{x} = \mathbf{0}$, where A_1 is obtained from A via a single elementary row operation (why?).

Iterating this, if B can be obtained from A by a finite number of elementary row operations, any solution to $A\mathbf{x} = \mathbf{0}$ remains a solution to $B\mathbf{x} = \mathbf{0}$. But since elementary operations are reversible, we can obtain A by applying a finite number of elementary row operations to B, and so any solution to $B\mathbf{x} = \mathbf{0}$ also gives a solution to $A\mathbf{x} = \mathbf{0}$.

Hence, it makes sense to define some "standard forms" for matrices that can be obtained through row operations, and for which the solution to a system of equations can be obtained from the matrix very easily. This is the point of the definitions given below:

Definition 3.5: Row Echelon Form; Reduced Row Echelon Form

Suppose $R \in M_{m \times n}(\mathbb{F})$. We say that R is in *row echelon form* (REF) if:

1. The rows of R that contain only zero entries (if any) are below the rows with non-zero entries.
2. In any non-zero row of R (where some entry is non-zero), the first non-zero entry of the row (called a *pivot*) has only zero entries below it in its column.

We say that R is in *reduced row echelon form* (RREF) if:

1. The rows of R that contain only zero entries (if any) are below the rows with non-zero entries.
2. In any non-zero row of R, the pivot is equal to 1, and the pivot is the **only** non-zero entry in its column.

To illustrate, let's apply elementary row operations to the coefficient matrices coming from Example 3.3, putting the matrices into REF and RREF, and reading off the general solution to the system of equations (if it exists).

Example 3.4

Let's solve the system

$$\begin{cases} 2x_1 - x_2 + x_3 = 0 \\ x_1 + 4x_2 - 2x_3 = 0. \end{cases}$$

Here, the augmented coefficient matrix is

$$\left(\begin{array}{ccc|c} 2 & -1 & 1 & 0 \\ 1 & 4 & -2 & 0 \end{array}\right)$$

Since this is a homogeneous system, no matter what row operations we apply, the 0 entries at the end of each row always remain 0. Hence, it is customary to omit the last entry, and to work with the ordinary coefficient matrix instead of the augmented one. We begin with

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

Let's begin by swapping the two rows, since this will put a 1 in the top-left corner of the matrix, which we can use as a pivot in the RREF.

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

To clear out the other entries in the first column, we now add -2 times the first row to the second row, which gives

$$\begin{pmatrix} 1 & 4 & -2 \\ 0 & -9 & 5 \end{pmatrix}.$$

This matrix is now in row echelon form, and it is possible to read off the general solution from here. However, we will proceed to take the reduced row echelon form. We multiply the second row by $-\frac{1}{9}$ to get

$$\begin{pmatrix} 1 & 4 & -2 \\ 0 & 1 & -5/9 \end{pmatrix}.$$

Now, the pivot in the second row is also a 1. The only step left is to clear out the other entry of the second column by adding -4 times the second row to the first row. The result is

$$\begin{pmatrix} 1 & 0 & 2/9 \\ 0 & 1 & -5/9 \end{pmatrix}.$$

This is a matrix in reduced row echelon form, row equivalent to A. Thus, the original system of equations has the same solutions as the system

$$\begin{cases} x_1 + (2/9)x_3 = 0 \\ x_2 - (5/9)x_3 = 0. \end{cases}$$

From this form of the equation, it becomes clear that we can set x_3 to be whatever we like, and then x_1 and x_2 are completely determined by that choice, since $x_1 = -(2/9)x_3$ and $x_2 = (5/9)x_3$. If we let t be our choice for the variable x_3, a vector form for the general solution thus becomes

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -(2/9)t \\ (5/9)t \\ t \end{pmatrix} = t \begin{pmatrix} -2/9 \\ 5/9 \\ 1 \end{pmatrix},$$

where $t \in \mathbb{F}$ is arbitrary, showing that all solutions are scalar multiples of the fixed solution vector $\begin{pmatrix} -2/9 \\ 5/9 \\ 1 \end{pmatrix}$.

Example 3.5

Now, let's solve the system

$$\begin{cases} x_1 + 2x_2 = 0 \\ x_1 - x_2 = -1 \\ 2x_1 + x_2 = 1. \end{cases}$$

Just like in the homogeneous case, we can work with the augmented coefficient matrix, and put that matrix into reduced row echelon form. The solutions to the system of equations with that augmented coefficient matrix will match up with the solutions to the system we started with (why?)

Setting up the matrix, we have

$$\begin{pmatrix} 1 & 2 & | & 0 \\ 1 & -1 & | & -1 \\ 2 & 1 & | & 1 \end{pmatrix}.$$

The top-left entry already has a 1, so that will be our first pivot. We now add -1 times the first row to the second row, and -2 times the first row to the third row, to get

$$\begin{pmatrix} 1 & 2 & | & 0 \\ 0 & -3 & | & -1 \\ 0 & -3 & | & 1 \end{pmatrix}.$$

Next, to clear out the bottom entry of the second column, we add -1 times the second row to the third row:

$$\begin{pmatrix} 1 & 2 & | & 0 \\ 0 & -3 & | & -1 \\ 0 & 0 & | & 2 \end{pmatrix}$$

This augmented matrix is now in row echelon form, but let's keep going. If we multiply the last row by $\frac{1}{2}$, we have

$$\begin{pmatrix} 1 & 2 & | & 0 \\ 0 & -3 & | & -1 \\ 0 & 0 & | & 1 \end{pmatrix},$$

and adding that third row to the second clears out all the non-pivot entries in the final column:

$$\begin{pmatrix} 1 & 2 & | & 0 \\ 0 & -3 & | & 0 \\ 0 & 0 & | & 1 \end{pmatrix}.$$

Finally, we multiply the second row by $-1/3$, and then add -2 times that new second row to the first row to clean things up:

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

This matrix is in RREF, and the simplified system corresponding to this augmented matrix is

$$\begin{cases} x_1 = 0 \\ x_2 = 0 \\ 0 = 1. \end{cases}$$

The last equation is a contradiction, hence the system has no solution (i.e. the system is *inconsistent*). Notice that we could have reached this contradictory conclusion much sooner in our calculations, without going all the way to the RREF. In this example, we carried out the full calculation only as one more illustration of RREF.

We conclude by reviewing a couple structural results about linear systems of equations:

Theorem 3.2

Suppose $A \in M_{m\times n}(\mathbb{F})$, and let $\mathbf{b}$ be a non-zero vector in $\mathbb{F}^m$.

(1) The set of solutions $\mathbf{x} \in \mathbb{F}^n$ to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{F}^n$.

(2) If $\mathbf{x}_p$ is a single solution to the inhomogeneous system $A\mathbf{x} = \mathbf{b}$, then every solution to the inhomogeneous system has the form $\mathbf{y} + \mathbf{x}_p$, where $A\mathbf{y} = \mathbf{0}$, and conversely, every vector of the form $\mathbf{y} + \mathbf{x}_p$, where $A\mathbf{y} = \mathbf{0}$, is a solution to $A\mathbf{x} = \mathbf{b}$.

Proof. We will only prove one direction of part (2) of this theorem, leaving the other direction of the proof and the proof of part (1) as excellent exercises. So, suppose we are given a single solution $\mathbf{x}_p$ to the inhomogeneous system $A\mathbf{x} = \mathbf{b}$, and suppose $\mathbf{z}$ is another solution. Then $A\mathbf{x}_p = \mathbf{b}$ and $A\mathbf{z} = \mathbf{b}$. From properties of matrix multiplication, it follows that

$$A(\mathbf{z} - \mathbf{x}_p) = A\mathbf{z} - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0},$$

so that $\mathbf{y} = \mathbf{z} - \mathbf{x}_p$ is a solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$. Re-writing this, we have $\mathbf{z} = \mathbf{y} + \mathbf{x}_p$, where $\mathbf{y}$ has just been identified as a solution to the homogeneous system. $\square$

MATH 235 Class 4:
Linear Independence, Span, Basis, and Dimension

September 21, 2021

Of all the definitions attached to the study of vector spaces, the definition of a basis is possibly the most important. The notion of a basis in linear algebra leads directly to the concepts of *dimension* and *coordinates* in a vector space. In turn, we need these definitions in place in order to represent linear transformations in matrix form, and to work with those matrix representations.

The definition of a basis for an arbitrary vector space is similar to the definition in $\mathbb{F}^n$ and in subspaces of $\mathbb{F}^n$. In particular, it requires two further definitions as components: linear independence and spanning. One main difference at play in the more general setting is that a vector space no longer must have a *finite* basis. In particular, it will be necessary to discuss linear independence and spanning for *infinite* sets of vectors. These slightly more general definitions will be the focus of the first part of these notes, followed by a discussion of the definition of basis, some examples, and some theoretical results about bases (including why they must exist).

Following that, we build on the definition of basis to officially define the dimension of a vector space. This will allow us to define the coordinate matrix of a vector with respect to a chosen basis in a finite-dimensional vector space a few lessons from now, once we've taken up the topic of linear transformations.

Linear Independence

Let's begin with the official definition of linear independence. As mentioned above, the definition is generalized to take into account the fact that we might be working in a general vector space, and also that we might be working with an infinite set of vectors.

Definition 4.1: Linear Independence

Let V be a vector space, and let S denote a subset of V (either finite or infinite). We say that S is *linearly independent* if the only way to obtain the zero vector as a finite linear combination of vectors in S is when all the coefficients are zero. In other words, S is linearly independent if the following implication is true: for all distinct vectors $v_1, \ldots, v_n \in S$, if

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0$$

for some choice of scalars $c_1, \ldots, c_n \in \mathbb{F}$, then $c_1 = c_2 = \cdots = c_n = 0$.

At this point, a word of caution is in order. A common mistake is to get the implication backwards in the definition of linear independence! In other words, many people will (correctly) observe that *if* we take $c_1 = c_2 = \ldots = c_n = 0$, then for any choice of vectors $v_1, v_2, \ldots, v_n \in S$, we have

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0.$$

They will then go on to say that this means S is linearly independent. **This is not correct.** By that reasoning, *every* set S would be linearly independent, which would make this a useless definition to have!

Example 4.1

In the vector space $V = M_{2\times 2}(\mathbb{F})$, the set

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

is linearly independent. Indeed, suppose we have scalars $c_1, c_2, c_3, c_4 \in \mathbb{F}$ such that

$$c_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c_3 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + c_4 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Simplifying the linear combination, we end up with

$$\begin{pmatrix} c_1 & c_2 \\ c_3 & c_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This immediately implies $c_1 = c_2 = c_3 = c_4 = 0$. According to the new definition of linear independence we've given, we have to check similar linear combinations for *every* subset of S (not just the unique subset of size 4). For instance, we'd also need to argue why

$$c_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + c_3 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

implies $c_1 = c_2 = c_3 = 0$. However, this is not necessary: those verifications follow from the single verification we've already done (why?)

Example 4.2

In the vector space $V = P(\mathbb{F})$, the infinite set of vectors

$$S = \{1, x, x^2, x^3, \dots\}$$

is linearly independent. To see why, suppose we are given a finite set of vectors in S. In particular, there is a vector of largest degree, and we may assume without loss of generality that we are given a set of elements of the form $1, x, x^2, \dots, x^n$ (why can this be assumed?). If we have scalars $c_0, c_1, c_2, \dots, c_n \in \mathbb{F}$ such that

$$c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n = 0,$$

we wish to show that $c_0 = c_1 = \cdots = c_n = 0$. Suppose, for a contradiction, that $c_n \neq 0$ (if this were not true, we could omit x^n from the set and take the first nonzero coefficient instead). In this case, the polynomial $c_0 + c_1 x + \cdots + c_n x^n$ is zero when evaluated at all values in $\mathbb{F}$. In particular, this is a polynomial of finite degree having infinitely many roots. However, it is a well-known fact that polynomials of degree n over $\mathbb{F}$ have at most n distinct roots; only the zero polynomial has infinitely many roots. This is a contradiction, so we conclude that $c_0 = c_1 = \cdots = c_n = 0$. Hence S is linearly independent.

Example 4.3

For a more computational example, let's check that the set of three matrices

$$S = \left\{ \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} \right\}$$

is linearly independent in $V = M_{2\times 2}(\mathbb{F})$.

By definition, we assume we have scalars $c_1, c_2, c_3 \in \mathbb{F}$ such that

$$c_1 \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} + c_2 \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} + c_3 \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

We would like to show that $c_1 = c_2 = c_3 = 0$. Combining the left-hand side into a single matrix, we get

$$\begin{pmatrix} c_1 + 2c_2 & -c_1 + 3c_3 \\ c_1 + 2c_2 + c_3 & -c_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Comparing entries leads to a system of four equations in the three unknowns c_1, c_2, c_3:

$$\begin{cases} c_1 + 2c_2 = 0 \\ -c_1 + 3c_3 = 0 \\ c_1 + 2c_2 + c_3 = 0 \\ -c_3 = 0. \end{cases}$$

We set up the coefficient matrix and put it into RREF to solve, using $\sim$ to denote row equivalences:

$$\begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 3 \\ 1 & 2 & 1 \\ 0 & 0 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

From the RREF, we find that $c_1 = c_2 = c_3 = 0$ must hold, as we needed. In conclusion, the original set of matrices is linearly independent.

The Span of a Set

In your previous linear algebra course, the span of a finite set of vectors may have been introduced to you as the set of all linear combinations of those vectors. Here, we introduce a seemingly different definition, that will apply more easily to infinite sets of vectors.

Definition 4.2: The Subspace Spanned by a Set

Let V be a vector space over $\mathbb{F}$, and let S be a subset of V. The *subspace spanned by* S, denoted $\operatorname{Span} S$, is the smallest subspace of V containing S, with respect to set containment. In other words, $\operatorname{Span} S$ is determined by the following two properties:

(1) $\operatorname{Span} S$ is a subspace of V, and $S \subseteq \operatorname{Span} S$.

(2) If W is a subspace of V such that $S \subseteq W$, then $\operatorname{Span} S \subseteq W$.

When the definition is given in this way, it is not immediately clear why $\operatorname{Span} S$ must always exist, let alone agree with the definition that takes the span of a finite set of vectors to be the set of all their linear combinations. We seek to address both of these questions in the theorem below:

Theorem 4.1

Let V be a vector space, and let S be a subset of V. Then:

(1) If $\operatorname{Span} S$ exists, then it is unique.

(2) $\operatorname{Span} S$ exists.

(3) If $S = \{v_1, \ldots, v_n\}$, for some vectors $v_1, \ldots, v_n \in \mathbb{F}$, then

$$\operatorname{Span} S = \{c_1 v_1 + c_2 v_2 + \cdots + c_n v_n : c_1, c_2, \ldots, c_n \in \mathbb{F}\}.$$

(4) More generally, $\operatorname{Span} S$ is equal to the set of all linear combinations of finitely many vectors in S. In other words, every element of $\operatorname{Span} S$ is of the form $c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$, where $v_1, \ldots, v_n$ are in S, and $c_1, \ldots, c_n \in \mathbb{F}$. Conversely, every vector of this form is in $\operatorname{Span} S$.

Proof.
Proof of (1): Suppose that $\operatorname{Span} S$ exists, and suppose W_1 and W_2 are each subspaces of V satisfying the definition of $\operatorname{Span} S$. By part (1) of Definition 4.2, both W_1 and W_2 contain the vectors in S. By part (2) of the definition applied to W_2, knowing that W_1 is a subspace containing S gives us $W_2 \subseteq W_1$. Applying the same part of the definition to W_1 and using that W_2 is a subspace containing S, we get $W_1 \subseteq W_2$. These two facts together imply that $W_1 = W_2$, so that $\operatorname{Span} S$ is unique if it exists.

Proof of (2): We claim that $\operatorname{Span} S$ is equal to the intersection of all the subspaces of V that contain S. In other words, our claim is that the set

$$W_0 = \{v \in V : v \in W \text{ for all subspaces } W \text{ such that } S \subseteq W\}$$

satisfies the definition of $\operatorname{Span} S$.

First, we check that W_0 is a subspace using the Subspace Test. Note that $0 \in W_0$, because 0 belongs to every subspace of V, including the ones (like V itself) that contain S. Now, suppose $v_1, v_2 \in W_0$ and that $c \in \mathbb{F}$. We wish to show that $v_1 + v_2 \in W_0$ and that $cv_1 \in W_0$.

For this, let W be a subspace of V that contains S. By definition, both v_1 and v_2 belong to W. Since W is a subspace, this means that both $v_1 + v_2$ and cv_1 belong to W. But W was an arbitrary subspace of V containing S, so it follows that $v_1 + v_2$ and cv_1 belong to W_0 as well.

This proves that W_0 is a subspace of V, and it contains all the vectors in S because each vector in S satisfies the condition defining W_0 by construction. Similarly, W_0 is contained in every subspace W such that $S \subseteq W$, by definition. Thus, W_0 satisfies the definition of $\operatorname{Span} S$, as we needed to show.

Proof of (3): By the uniqueness of $\operatorname{Span} S$, as given in part (1), we show that if $S = \{v_1, v_2, \ldots, v_n\}$ for some vectors $v_1, \ldots, v_n \in V$, then

$$W_0 = \{c_1 v_1 + c_2 v_2 + \cdots + c_n v_n : c_1, c_2, \ldots, c_n \in \mathbb{F}\}$$

satisfies the definition of $\operatorname{Span} S$. The proof that W_0 is a subspace of V (using the Subspace Test) is left as an excellent exercise for you! We show here that if W is a subspace of V containing all of $v_1, v_2, \ldots, v_n$, then $W_0 \subseteq W$.

Thus, let $x = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$ be an arbitrary element of W_0, where $c_1, c_2, \ldots, c_n$ belong to $\mathbb{F}$. We show that $x \in W$. Note that by closure of W under scalar multiplication, and the fact that W contains $v_1, v_2, \ldots, v_n$, we have that $c_1 v_1, c_2 v_2, \ldots, c_n v_n$ all belong to W. Then, by closure of W under addition, and the fact that $c_1 v_1, c_2 v_2 \in W$, we get that $c_1 v_1 + c_2 v_2 \in W$. Since $c_3 v_3 \in W$ as well, closure under addition again yields that $c_1v_1 + c_2v_2 + c_3v_3$ belongs to W.

Continuing in this fashion, we find that $x = c_1v_1 + c_2v_2 + \cdots + c_nv_n \in W$, as desired. This proves that $W_0 \subseteq W$, which allows us to conclude that W_0 satisfies the definition of $\operatorname{Span} S$.

Proof of (4): The proof of this is very similar to the proof of (3), and is therefore left as a very helpful exercise for you to attempt! $\square$

Let's apply this new definition in a few examples:

Example 4.4

(1) If $S = \varnothing$, the empty set, then $\operatorname{Span} S = \{0\}$, the zero subspace. (Why?)

(2) If S is already a subspace of V, then $\operatorname{Span} S = S$. (Why?)

(3) If $V = P_2(\mathbb{F})$, we claim that $\operatorname{Span}\{1, (x-1), (x-1)^2\} = V$. Since $\operatorname{Span}\{1, (x-1), (x-1)^2\} \subseteq V$ follows by definition of the span, it remains to show that $V \subseteq \operatorname{Span}\{1, (x-1), (x-1)^2\}$. For this, we must show that for any choice of polynomial $a + bx + cx^2 \in V$, there are scalars $c_0, c_1, c_2 \in \mathbb{F}$ such that
$$c_0 \cdot 1 + c_1(x-1) + c_2(x-1)^2 = a + bx + cx^2.$$
Expanding out the left-hand side and collecting the coefficients, this becomes
$$(c_0 - c_1 + c_2) + (c_1 - 2c_2)x + c_2x^2 = a + bx + cx^2.$$
Equating the coefficients on both sides, we need to have
$$\begin{cases} c_0 - c_1 + c_2 = a \\ c_1 - 2c_2 = b \\ c_2 = c. \end{cases}$$
We could write out the augmented matrix for this system and put it into RREF, but it will be easier to argue directly instead. The last equation means we must have $c_2 = c$, so that the second equation tells us $c_1 = b + 2c_2 = b + 2c$. The first equation then gives us $c_0 = c_1 - c_2 + a = (b + 2c) - c + a = a + b + c$. Altogether, this means
$$(a + b + c) \cdot 1 + (b + 2c)(x-1) + c(x-1)^2 = a + bx + cx^2,$$
which tells us that $V \subseteq \operatorname{Span}\{1, (x-1), (x-1)^2\}$. In conclusion, $V = \operatorname{Span}\{1, (x-1), (x-1)^2\}$, as desired.

Basis of a Vector Space

Simply put, a basis for a vector space represents the combination of the two ingredients we've just introduced above.

Definition 4.3: Basis of a Vector Space

Let V be a vector space. A subset B of V is called a *basis* for V if:

(1) B is a linearly independent set.

(2) $\operatorname{Span} B = V$.

Example 4.5

(1) The set $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right\}$ is a basis for $\mathbb{F}^n$, called the *standard basis of* $\mathbb{F}^n$. (Verify this!)

(2) Let $V = M_{m\times n}(\mathbb{F})$, and for each index i with $1 \leq i \leq m$ and index j with $1 \leq j \leq n$, let E_{ij} be the matrix with a 1 in the (i,j)th entry, and a 0 everywhere else. Then the set of matrices $S = \{E_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ is a basis for V, called the *standard basis of* $M_{m\times n}(\mathbb{F})$. (Again, it's good practice to verify this!)

(3) For your first example of a vector space without a finite basis, let $V = P(\mathbb{F})$, and let $S = \{1, x, x^2, \dots\}$, the set of all powers of x. We claim that S is a basis for V. We already showed that S is linearly independent in Example 4.2, so it only remains to check that $\operatorname{Span} S = V$.

By definition, any vector in V can be represented as a finite linear combination

$$c_0 + c_1 x + \cdots + c_n x^n,$$

for some choice of scalars $c_0, c_1, \dots, c_n$ and some integer $n \geq 0$. However, since $\operatorname{Span} S$ is a subspace of V that contains all of $1, x, \dots, x^n$, it is easy to show using closure under scalar multiplication and addition that $\operatorname{Span} S$ contains $c_0 + c_1 x + \cdots + c_n x^n$ as well (how?). This verifies that every element of V is in $\operatorname{Span} S$, so that S is a spanning set for V, and hence a basis.

Example 4.6

For a more computational example, consider the set

$$S = \left\{ \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} \right\}$$

of matrices in $M_{2\times 2}(\mathbb{F})$ from Example 4.3. We showed that this set is linearly independent, but it is not a basis for $M_{2\times 2}(\mathbb{F})$; we will show that S does not span all of $M_{2\times 2}(\mathbb{F})$. Indeed, it will be enough to pick a single matrix and show that this matrix is not in the span of $M_{2\times 2}(\mathbb{F})$.

Note that $\operatorname{Span} S$ consists of all matrices of the form

$$a \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} + b \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix},$$

where $a, b, c \in \mathbb{F}$. Combining this all together, we see that

$$\operatorname{Span} S = \left\{ \begin{pmatrix} a + 2b & -a + 3c \\ a + 2b + c & -c \end{pmatrix} : a, b, c \in \mathbb{F} \right\}.$$

If we can show there are 2×2 matrices that are not of the form given above, we will have shown that $\operatorname{Span} S$ is not equal to all of $M_{2\times 2}(\mathbb{F})$. We claim, in fact, that $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ is not in $\operatorname{Span} S$. Indeed, if it *were* in $\operatorname{Span} S$, then we would be able to find $a, b, c \in \mathbb{F}$ such that

$$\begin{cases} a + 2b = 0 \\ -a + 3c = 1 \\ a + 2b + c = 0 \\ -c = 1. \end{cases}$$

Again, we could write out the augmented matrix and use that to show this system has no solution, but it will be easier to argue directly from the equations. Subtracting the first equation from the third tells us that $c = 0$ must hold, while the fourth equation tells us $c = -1$ must hold. These two conditions are contradictory, so we conclude there is no solution after all.

Later in this lesson, we will see why adding the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ into the set S gives a basis for $M_{2\times 2}(\mathbb{F})$.

Other ways to describe a basis for V are as a maximal linearly independent subset of V and a minimal spanning set of vectors for V. We prove these characterizations of a basis immediately below:

Theorem 4.2

Let V be a vector space. Then for a subset B of V, the following are equivalent:

(1) B is a basis of V.

(2) B is a linearly independent subset of V, and if S is a linearly independent subset of V with $B \subseteq S$, then $B = S$.

(3) B is a spanning set for V (i.e. $\operatorname{Span} B = V$), and if S is a spanning set for V with $S \subseteq B$, then $S = B$.

Proof.
(1) $\Rightarrow$ (2): Suppose B is a basis of V. By definition, this means B is a linearly independent subset of V. Suppose in addition that S is a linearly independent subset of V with $B \subseteq S$, and suppose, towards a contradiction, that $B \neq S$. In particular, there must be some vector $w \in S$ such that $w \notin B$. On the other hand, since $\operatorname{Span} B = V$, we know that $w \in \operatorname{Span} B$. By part (4) of Theorem 4.1, this means we have vectors $v_1, \dots, v_n \in B$ and scalars $c_1, \dots, c_n \in \mathbb{F}$ such that

$$w = c_1v_1 + c_2v_2 + \cdots + c_nv_n.$$

Written differently, this says

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n + (-1)w = 0.$$

But $v_1, v_2, \dots, v_n, w \in S$, and the above equation contradicts the statement that S is linearly independent. We conclude that our assumption that there is a vector w in S that is not in B must be incorrect. It follows immediately that $B = S$, as desired.

(2) $\Rightarrow$ (3): Suppose B is a linearly independent subset of V, not properly contained in any other linearly independent subset of V. We first argue that B is also a spanning set for V. Assume for a contradiction that $\operatorname{Span} B \neq V$. In this case, there must be some vector $w \in V$ such that $w \notin \operatorname{Span} B$. We then claim that the set $B \cup \{w\}$ is a linearly independent subset of V as well. Indeed, suppose some linear combination of vectors in $B \cup \{w\}$ is equal to the zero vector. If w is not used as one of those vectors, we already know all the coefficients must be zero by linear independence of B. Hence we may assume that there are $v_1, \dots, v_n \in B$ and scalars $c_1, \dots, c_n, d \in \mathbb{F}$ such that

$$c_1v_1 + \cdots + c_nv_n + dw = 0.$$

If $d = 0$, then $c_1 = \cdots = c_n = 0$ follows immediately from linear independence of B. Hence, we may assume that $d \neq 0$. But then

$$w = (-c_1/d)v_1 + (-c_2/d)v_2 + \cdots + (-c_n/d)v_n,$$

showing that $w \in \operatorname{Span} B$. This is a contradiction. Therefore, $\operatorname{Span} B = V$ after all.

Now, suppose also that S is a spanning set for V with $S \subseteq B$, and assume for a contradiction that $S \neq B$. In particular, there is some vector $w \in B$ such that $w \notin S$. Since $\operatorname{Span} S = V$, there are vectors $v_1, \ldots, v_n \in S$ and scalars $c_1, \ldots, c_n \in \mathbb{F}$ such that

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = w.$$

But again, writing this as

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n - w = 0,$$

we have a contradiction to the fact that B is linearly independent. Thus, $S = B$ after all.

(3) $\Rightarrow$ (1): Assume that B is a spanning set for V that does not properly contain any spanning set for V. To prove that B is a basis for V, it only remains to show that B is linearly independent. For a contradiction, suppose that there are vectors $v_1, \ldots, v_n \in B$ and scalars $c_1, \ldots, c_n \in \mathbb{F}$, not all zero, such that

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

Without loss of generality, we may assume that $c_1 \neq 0$, so that we may write

$$v_1 = (-c_2/c_1)v_2 + \cdots + (-c_n/c_1)v_n.$$

From this, we claim that the set $S = B \setminus \{v_1\}$, obtained by removing v_1 from B, still spans V. Indeed, we already know every vector $x \in V$ is a linear combination of vectors in B. If v_1 is not used in that linear combination, then x is in fact a linear combination of vectors in S. On the other hand, if v_1 *is* used in that linear combination, we can use the equation above to replace v_1 with a linear combination of vectors in S, showing that x is also a linear combination of vectors in S.

By assumption then, the set S, being a spanning set for V contained in B, must be equal to B. But this is a contradiction, since v_1 belongs to B but not to S. We conclude that our initial assumption was false, and B is in fact linearly independent. $\square$

Some of the arguments used in the proof of Theorem 4.2 are worth extracting as results of their own, since they allow us to either build up to a basis from a linearly independent set, or cut down to a basis from a spanning set. The proofs of these statements are left as exercises; they can almost be copied out word-for-word from appropriate portions of the proof above.

Theorem 4.3

Let V be a vector space, and let S be a subset of V.

(1) If S is linearly independent and there is a vector $w \in V$ such that $w \notin \operatorname{Span} S$, then $S \cup \{w\}$ is linearly independent.

(2) If $\operatorname{Span} S = V$, and $w \in S$ can be written as a linear combination of other vectors in S, then $S \setminus \{w\}$ also spans V.

Dimension, and Basis Examples

Now that we've defined what a basis is and explored some of its characterizations, it is worth asking: does every vector space have a basis? The answer is yes, but in this course, we only have the means to prove it in a special case. This motivates a new definition:

Definition 4.4: Finite-Dimensional Vector Space

A vector space V is said to be *finite-dimensional* if V can be spanned by a finite set of vectors.

Example 4.7

By Example 4.5, parts (1) and (2), both $\mathbb{F}^n$ and $M_{m\times n}(\mathbb{F})$ are finite-dimensional, with the standard basis for each forming a finite spanning set for the vector space.

On the other hand, $P(\mathbb{F})$ is *not* finite-dimensional. Indeed, if $p_1, \ldots, p_n$ are finitely many polynomials in $P(\mathbb{F})$, then there is some integer $m \geq 0$ such that all of $p_1, \ldots, p_n$ belong to $P_m(\mathbb{F})$ (why?). It follows that any linear combination of $p_1, \ldots, p_n$ is also within $P_m(\mathbb{F})$, which means that any polynomial of degree larger than m does not belong to $\operatorname{Span}\{p_1, \ldots, p_n\}$. We conclude immediately that $P(\mathbb{F})$ is not spanned by finitely many vectors.

Our first goal is to prove that every finite-dimensional vector space has a basis. The full proof that every vector space has a basis requires the use of something called *Zorn's lemma*. This result is equivalent to a famous axiom of mathematics called the *axiom of choice*, which (roughly speaking) says that we can always make infinitely many choices at once in a well-defined way. Those of you who are curious are encouraged to look it up, though it falls outside the scope of our course.

To aid with the proof of our result, we begin with a useful lemma:

Lemma 4.1

Let V be a vector space, and suppose that V is spanned by a set $S = \{v_1, \ldots, v_m\}$. If T is a linearly independent subset of V, then T has at most m vectors in it.

Proof. Suppose that $T = \{w_1, \ldots, w_n\}$ is a finite linearly independent subset of V; our goal is to show that $n \leq m$. Knowing that $\operatorname{Span} S = V$, each vector in T can be written as a linear combination of vectors in S, say

$$w_j = \sum_{i=1}^{m} a_{ij} v_i,$$

for some scalars $a_{ij} \in \mathbb{F}$, where $1 \leq i \leq m$ and $1 \leq j \leq n$.

If we assume to the contrary that $n > m$, then the system of equations with coefficient matrix $A = (a_{ij})$ has more unknowns than equations, i.e. A has more columns than rows. Therefore, in the RREF of A, the number of pivot columns will be smaller than the total number of columns, which means that at least one variable in the system $A\mathbf{x} = \mathbf{0}$ will be a free variable. This means we can find scalars $x_1, \ldots, x_n \in \mathbb{F}$, not all zero, such that $A\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{0}$. In other words, we get the equations $\sum_{j=1}^{n} a_{ij} x_j = 0$.

We now claim that the linear combination

$$x_1 w_1 + x_2 w_2 + \cdots + x_n w_n = 0,$$

which would contradict the linear independence of T. We can verify this directly:

$$\begin{aligned}\sum_{j=1}^{n} x_j w_j &= \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij} v_i\right)\\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j\right) v_i\\ &= \sum_{i=1}^{m} 0 \cdot v_i\\ &= 0.\end{aligned}$$

This contradiction to linear independence shows that we must have $n \leq m$ after all. □

Theorem 4.4

Let V be a vector space. If V is finite-dimensional, then V has a basis.

Proof. Let V be a finite-dimensional, non-zero vector space. (The result is trivial for the zero vector space – why?) By definition, this means there is some finite set S such that $\operatorname{Span} S = V$. By Lemma 4.1, any linearly independent subset of V has finitely many elements, limited by $|S|$. Among the subsets of V, there are certainly linearly independent ones; any singleton subset $\{v\}$ with $v \neq 0$ will work. In particular, we can choose a linearly independent subset of V that has the most elements out of all possible linearly independent subsets of V. Let B be such a subset; we claim that B is a basis for V.

We will show that B is a basis by verifying that it meets the criteria of Theorem 4.2. Certainly, we already know that B is linearly independent by construction, so let C denote a linearly independent subset of V such that $B \subseteq C$. Since B has the most linearly independent elements that any subset of V can have, it must be true that $B = C$. This completes the proof that B is a basis of V. □

Apart from the existence of a basis, the other piece of groundwork we need is that the number of elements in a basis is unique. In other words, every basis for a finite-dimensional vector space has the same number of elements.

Theorem 4.5

Let V be a finite-dimensional vector space, and suppose B_1 and B_2 are bases for V. Then $|B_1| = |B_2|$, i.e. B_1 and B_2 have the same number of elements.

Proof. Suppose we are given bases B_1 and B_2 for V. In particular, B_2 spans V and B_1 is a linearly independent subset of V. It follows by Lemma 4.1 that $|B_1| \leq |B_2|$. On the other hand, since B_1 spans V and B_2 is a linearly independent subset of V, Lemma 4.1 again guarantees that $|B_2| \leq |B_1|$. We conclude that $|B_1| = |B_2|$, as desired. □

This directly leads into the definition of dimension in linear algebra:

Definition 4.5: Dimension

Let V be a finite-dimensional vector space. The *dimension* of V, denoted $\dim V$, is equal to the number of elements in a basis for V (which is well-defined by Theorem 4.5).

Example 4.8

Continuing from Example 4.5, we can see immediately that $\dim \mathbb{F}^n = n$ and that $\dim M_{m\times n}(\mathbb{F}) = mn$. Indeed, the standard basis for $\mathbb{F}^n$ is seen immediately to have n vectors in it, and there are mn standard basis vectors E_{ij} for $M_{m\times n}(\mathbb{F})$, since we have m choices for the first index i and n choices for the second index j.

One very helpful thing about knowing the dimension of a vector space is that it simplifies the verification that a given set is a basis, in the following way:

Theorem 4.6

Suppose that V is a finite-dimensional vector space, that $\dim V = n$, and that S is a subset of V with n vectors in it. If S is linearly independent, then S is a basis for V. Similarly, if S spans V, then S is a basis for V.

Proof. Let B denote a known basis for V. We know that $|B| = n$ by definition of dimension. First, assume that S is a linearly independent set of n vectors in V. If S did not span all of V, then there would be a vector $w \in V$ not belonging to $\operatorname{Span} S$. But by Theorem 4.3 part (1), this would mean $S \cup \{w\}$ is linearly independent, forming a linearly independent subset of V with $n+1$ vectors in it. But since V is spanned by B, which has only n vectors, there cannot be a linearly independent subset of size $n+1$ by Lemma 4.1. It follows that S spans V after all.

Now suppose S is a spanning set for V with n vectors in it. If S was not linearly independent, then we could find a vector $w \in S$ that is a linear combination of other vectors in S (why?). In particular, $S \setminus \{w\}$ would also span V by part (2) of Theorem 4.3. But then, since $S \setminus \{w\}$ has only $n-1$ vectors in it, Lemma 4.1 would tell us that there can be no linearly independent subset of V with more than $n-1$ vectors, contradicting the existence of the basis B. It follows that S is linearly independent after all. $\square$

Let's apply this theorem to construct a non-standard basis for $M_{2\times 2}(\mathbb{F})$:

Example 4.9

In Example 4.6, we saw that

$$S = \left\{ \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} \right\}$$

was a linearly independent subset of $M_{2\times 2}(\mathbb{F})$ that is not a basis. Using what we know now, we can conclude that S is not a basis for this space in a different way, because $\dim M_{2\times 2}(\mathbb{F}) = 2 \times 2 = 4$, and S has only three vectors in it.

We also found in Example 4.6 that $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ did not belong to $\operatorname{Span} S$. By Theorem 4.3 part (1), the set

$$S' = \left\{ \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\},$$

obtained by adding this matrix to S, is linearly independent. But now S' is a linearly independent set of four matrices in the four-dimensional vector space $M_{2\times 2}(\mathbb{F})$, so S' must be a basis for $M_{2\times 2}(\mathbb{F})$ by Theorem 4.6.

MATH 235 Class 5: Introduction to Linear Transformations

September 23, 2021

Throughout mathematics, once we have defined a new type of object (like vector spaces), we usually seek to study how different objects of the same type relate to each other. One of the common ways to do that is by studying the functions between different objects of that type. Furthermore, we will want the functions to preserve whatever operations and structures we have. When this is applied to the vector space structure, the resulting functions are called *linear transformations*. Informally speaking, these are the functions between vector spaces that preserve the vector addition and scalar multiplication operations.

In this lesson, we formally define linear transformations, give several examples, and then state and prove some basic properties of these functions. We conclude by introducing two important subspaces connected with a linear transformation: the kernel and the range.

Definition and Examples of Linear Transformations

In your previous linear algebra course, you have probably seen the definition of a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$. The general definition has many of the same features.

Definition 5.1: Linear Transformation

Let V and W be vector spaces over $\mathbb{F}$. Under a slight abuse of notation, we use the same notation for vector addition and scalar multiplication in both V and W. We say that a function $T : V \to W$ is a *linear transformation* if:

(1) T preserves addition of vectors: for all $v_1, v_2 \in V$, we have $T(v_1 + v_2) = T(v_1) + T(v_2)$. (Note that the addition $v_1 + v_2$ is taking place in V, and $T(v_1) + T(v_2)$ is taking place in W.)

(2) T preserves scalar multiplication: for all vectors $v \in V$ and all scalars $c \in \mathbb{F}$, we have $T(cv) = cT(v)$. (Again, note that the scalar multiplication cv is taking place in V, while the scalar multiplication $cT(v)$ is taking place in W.)

In practice, we can combine these two conditions into a single verification, captured by the following theorem:

Theorem 5.1

Let V and W be vector spaces over $\mathbb{F}$. A function $T : V \to W$ is a linear transformation if and only if, for all vectors $v_1, v_2 \in V$ and all scalars $c_1, c_2 \in \mathbb{F}$, we have

$$T(c_1v_1 + c_2v_2) = c_1T(v_1) + c_2T(v_2).$$

Proof. First, suppose T is a linear transformation as according to Definition 5.1. Now, suppose we have vectors $v_1, v_2 \in V$ and scalars $c_1, c_2 \in \mathbb{F}$. Using the fact that T preserves addition of vectors, we know that

$$T(c_1v_1 + c_2v_2) = T(c_1v_1) + T(c_2v_2).$$

Now, since T also preserves scalar multiplication, we have $T(c_1v_1) = c_1T(v_1)$ and $T(c_2v_2) = c_2T(v_2)$. Therefore, we get

$$T(c_1v_1 + c_2v_2) = c_1T(v_1) + c_2T(v_2),$$

as desired.

Conversely, suppose T is a function from V to W satisfying the property that for all vectors $v_1, v_2 \in V$ and all scalars $c_1, c_2 \in \mathbb{F}$, we have $T(c_1v_1 + c_2v_2) = c_1T(v_1) + c_2T(v_2)$. To prove that T preserves addition, we take $c_1 = c_2 = 1$ and let v_1, v_2 be arbitrary. The result is that

$$T(1 \cdot v_1 + 1 \cdot v_2) = 1 \cdot T(v_1) + 1 \cdot T(v_2).$$

By the vector space axiom regarding scalar multiplication by 1, we immediately get

$$T(v_1 + v_2) = T(v_1) + T(v_2),$$

proving that T preserves addition.

To prove that T preserves scalar multiplication, we take $c_1 \in \mathbb{F}$, and $v_1, v_2 \in V$ to be arbitrary, while taking $c_2 = 0$. This gives us

$$T(c_1v_1 + 0v_2) = c_1T(v_1) + 0T(v_2),$$

which simplifies down to

$$T(c_1v_1) = c_1T(v_1),$$

since the scalar 0 multiplied by any vector is the zero vector by Lemma 2.1. Because c_1 and v_1 were arbitrary, we conclude that T preserves scalar multiplication. □

We will often use Theorem 5.1 to show that a function is a linear transformation instead of the actual definition, because using the theorem tends to be slightly simpler. This is illustrated in the examples below:

Example 5.1

(1) Given a vector space V, the function $I : V \to V$ given by $I(v) = v$ for all $v \in V$ is a linear transformation. This mapping is usually called the *identity map*. It is almost an instant verification to check that I satisfies the conditions of Theorem 5.1.

(2) Given any two vector spaces V and W, the function $Z : V \to W$ given by $Z(v) = 0$ for all $v \in V$ is a linear transformation, called the *zero transformation*. Indeed, note that for any vectors $v_1, v_2 \in V$ and scalars $c_1, c_2 \in \mathbb{F}$, we have

$$Z(c_1v_1 + c_2v_2) = 0 = c_1 \cdot 0 + c_2 \cdot 0 = c_1Z(v_1) + c_2Z(v_2).$$

(3) For an important type of example, one that you have seen in your previous linear algebra course, let $A \in M_{m\times n}(\mathbb{F})$ and consider the matrix-multiplication function $T_A : \mathbb{F}^n \to \mathbb{F}^m$ given by $T_A(\mathbf{x}) = A\mathbf{x}$ for each vector $\mathbf{x} \in \mathbb{F}^n$. This is also a linear transformation. Indeed, given $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{F}^n$ and $c_1, c_2 \in \mathbb{F}$, we get

$$T_A(c_1\mathbf{x}_1+c_2\mathbf{x}_2) = A(c_1\mathbf{x}_1+c_2\mathbf{x}_2) = A(c_1\mathbf{x}_1)+A(c_2\mathbf{x}_2) = c_1(A\mathbf{x}_1)+c_2(A\mathbf{x}_2) = c_1T_A(\mathbf{x}_1)+c_2T_A(\mathbf{x}_2),$$

using the various properties of matrix multiplication given in Theorem 3.1.

Example 5.2

Suppose that $A \in M_{n\times n}(\mathbb{F})$. We define a function, called the *trace*, from $M_{n\times n}(\mathbb{F})$ to $\mathbb{F}$. If A has entries a_{ij}, where $1 \le i, j \le n$, this function $\operatorname{tr} : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is given by $\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$. In other words, $\operatorname{tr} A$ computes the sum of the diagonal entries of A.

For example, $\operatorname{tr}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 1 + 4 = 5$, and $\operatorname{tr}(I_n) = n$, where I_n is the $n \times n$ identity matrix.

We can check that tr is in fact a linear transformation. Indeed, if $A, B \in M_{n\times n}(\mathbb{F})$, and $c_1, c_2 \in \mathbb{F}$, where A has entries a_{ij} and B has entries b_{ij}, the entries of $c_1A + c_2B$ are given by $c_1a_{ij} + c_2b_{ij}$. So, if we add up the diagonal entries, we get

$$\begin{aligned}
\operatorname{tr}(c_1A + c_2B) &= \sum_{i=1}^{n}(c_1a_{ii} + c_2b_{ii}) \\
&= \sum_{i=1}^{n}(c_1a_{ii}) + \sum_{i=1}^{n}(c_2b_{ii}) \\
&= c_1\sum_{i=1}^{n}a_{ii} + c_2\sum_{i=1}^{n}b_{ii} \\
&= c_1\operatorname{tr}(A) + c_2\operatorname{tr}(B).
\end{aligned}$$

Example 5.3

Now, set $V = P_n(\mathbb{F})$. Two important linear transformations involving V are the *evaluation* mappings, and the *differentiation* mappings.

Given any scalar $a \in \mathbb{F}$, we define a map $\mathrm{ev}_a : V \to \mathbb{F}$ by taking $\mathrm{ev}_a(p) = p(a)$ for any $p \in V$. It follows immediately from properties of evaluating a polynomial at a point that ev_a is linear for any $a \in \mathbb{F}$ (can you verify this?).

Next, we may define the differentiation map $D : V \to V$ as follows. If $p = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, then we define

$$D(p) = a_1 + 2a_2x + \cdots + na_nx^{n-1}.$$

This agrees with the definition of derivative given in calculus when $\mathbb{F} = \mathbb{R}$, but the definition may be new to you when working with the case $\mathbb{F} = \mathbb{C}$!

Let's quickly verify that D is linear, with the help of Theorem 5.1. If $p = a_0 + a_1x + \cdots + a_nx^n$ and $q = b_0 + b_1x + \cdots + b_nx^n$, and $\alpha, \beta \in \mathbb{F}$ are scalars, then

$$\begin{aligned}
D(\alpha p + \beta q) &= D((\alpha a_0 + \beta b_0) + (\alpha a_1 + \beta b_1)x + (\alpha a_2 + \beta b_2)x^2 + \cdots + (\alpha a_n + \beta b_n)x^n) \\
&= (\alpha a_1 + \beta b_1) + 2(\alpha a_2 + \beta b_2)x + \cdots + n(\alpha a_n + \beta b_n)x^{n-1} \\
&= (\alpha a_1 + 2\alpha a_2x + \cdots + n\alpha a_nx^{n-1}) + (\beta b_1 + 2\beta b_2x + \cdots + n\beta b_nx^{n-1}) \\
&= \alpha(a_1 + 2a_2x + \cdots + na_nx^{n-1}) + \beta(b_1 + 2b_2x + \cdots + nb_nx^{n-1}) \\
&= \alpha D(p) + \beta D(q).
\end{aligned}$$

Properties of Linear Transformations

One of the fundamental properties of linear transformations is that they map zero vectors to zero vectors. This is one of the ways they distinguish themselves from the "linear functions" studied in calculus, which don't necessarily map the number 0 to itself!

Theorem 5.2

Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. Then $T(0_V) = 0_W$, where 0_V is the zero vector of V, and 0_W is the zero vector of W.

Proof. We can prove this using the fact that T preserves addition. Since $0_V = 0_V + 0_V$, note that

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

Now, applying the result of Question 2 on Assignment 1, we see that the equation $T(0_V) = T(0_V) + T(0_V)$ in W implies that $T(0_V) = 0_W$. $\square$

Another incredibly useful property of linear transformations is that they can be completely determined from what they do to the vectors in a basis for the domain of the function. We first illustrate with an example:

Example 5.4

In Example 4.9, we showed that the set

$$B = \left\{ \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}$$

is a basis for $V = M_{2\times 2}(\mathbb{F})$. So, suppose we have a linear transformation $T : V \to \mathbb{F}^2$ such that

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

Can we use this information to work out $T \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, for example? The answer is yes! Since B is a basis for $M_{2\times 2}(\mathbb{F})$, we know that there will be scalars $a, b, c, d \in \mathbb{F}$ for which

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = a \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} + b \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} + d \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Once those scalars are found, applying T to both sides and using the linearity properties of T gives us

$$\begin{aligned} T \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} &= aT \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} + bT \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} + cT \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} + dT \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \\ &= a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 0 \end{pmatrix} + c \begin{pmatrix} 0 \\ 1 \end{pmatrix} + d \begin{pmatrix} -1 \\ -1 \end{pmatrix}. \end{aligned}$$

(Side note: why does T applied to the right-hand side result in the expression we wrote down in the first line? Theorem 5.1 only allows us to do this for linear combinations of two vectors at a time – how do we extend it to four vectors?)

For practice, let's find those scalars a, b, c, d and explicitly compute the value of $T \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Combining all the scalars together into a single matrix, we need to have

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} a + 2b & -a + 3c + d \\ a + 2b + c & -c + d \end{pmatrix}.$$

This leads to a system of four equations having augmented matrix

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ -1 & 0 & 3 & 1 & 1 \\ 1 & 2 & 1 & 0 & 1 \\ 0 & 0 & -1 & 1 & 1 \end{array} \right).$$

Putting this matrix into reduced-row echelon form, we get

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ -1 & 0 & 3 & 1 & 1 \\ 1 & 2 & 1 & 0 & 1 \\ 0 & 0 & -1 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ 0 & 2 & 3 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{array}\right)$$
$$\sim \left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$
$$\sim \left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$
$$\sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$
$$\sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

We conclude that $a = c = 0$, that $b = 1/2$, and that $d = 1$. (Note: we could have found this solution by inspection, if we were looking for it. However, this gave us more practice with how to find the scalars in general!)

In particular, from our work above,

$$T\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 1 \end{pmatrix} + d\begin{pmatrix} -1 \\ -1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 \\ 0 \end{pmatrix} + 1\begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

Now, we move on to stating and proving the theorem that captures the more general principle at work here. As part of the argument, we will use a lemma that will be very useful to us far beyond this theorem. It is often called the *unique representation theorem* for a basis, and can be applied even to vector spaces with an infinite basis:

Lemma 5.1

Let V be a vector space over $\mathbb{F}$, and let B be a basis for V. Then every vector $v \in V$ can be expressed *uniquely* as a linear combination of vectors in B. In other words, there are vectors $v_1, \ldots, v_n \in B$ and scalars $c_1, \ldots, c_n \in \mathbb{F}$ such that

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n,$$

and if there are any other scalars $d_1, \ldots, d_n \in \mathbb{F}$ such that

$$v = d_1 v_1 + d_2 v_2 + \cdots + d_n v_n,$$

then $c_1 = d_1, c_2 = d_2, \ldots, c_n = d_n$.

Proof. Given an arbitrary vector $v \in V$, the fact that there are vectors $v_1, \ldots, v_n \in B$ and scalars $c_1, \ldots, c_n \in$

$\mathbb{F}$ such that $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$ follows immediately from the fact that B spans V. To show the uniqueness part, suppose there are also scalars $d_1, \ldots, d_n \in \mathbb{F}$ such that $v = d_1v_1 + d_2v_2 + \cdots + d_nv_n$. This means we have an equality

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = d_1v_1 + d_2v_2 + \cdots + d_nv_n,$$

and moving everything to one side gives

$$(c_1 - d_1)v_1 + (c_2 - d_2)v_2 + \cdots + (c_n - d_n)v_n = 0.$$

By linear independence of the set B, we conclude that $c_1 - d_1 = c_2 - d_2 = \cdots = c_n - d_n = 0$. In other words, $c_1 = d_1, c_2 = d_2, \ldots, c_n = d_n$, as desired. □

Now, we extend the ideas at play in Example 5.4:

Theorem 5.3

Suppose V and W are vector spaces over $\mathbb{F}$, and that V is finite-dimensional. If $B = \{v_1, \ldots, v_n\}$ is a basis for V, and $w_1, \ldots, w_n \in W$ are any choice of vectors, then there is a unique linear transformation $T : V \to W$ such that $T(v_1) = w_1, T(v_2) = w_2, \ldots, T(v_n) = w_n$.

Proof. First, we prove that a linear transformation T satisfying the given condition exists. Let $v \in V$ be arbitrary. By Lemma 5.1, there are *unique* scalars $c_1, \ldots, c_n \in \mathbb{F}$ such that

$$v = c_1v_1 + c_2v_2 + \cdots + c_nv_n.$$

Given this representation, we *define*

$$T(v) = c_1w_1 + c_2w_2 + \cdots + c_nw_n.$$

The function $T : V \to W$ we are defining in this way is *well-defined* because of the uniqueness of $c_1, \ldots, c_n$. In other words, there is no ambiguity about what the vector $T(v)$ should be.

We now have to verify that:

- T is a linear transformation.
- We have $T(v_1) = w_1, T(v_2) = w_2, \ldots, T(v_n) = w_n$.

To check linearity, we again appeal to Theorem 5.1. Suppose we have vectors $v, x \in V$, and scalars $a, b \in \mathbb{F}$. We wish to show that $T(av+bx) = aT(v)+bT(x)$. First, we pick the unique scalars $c_1, \ldots, c_n, d_1, \ldots, d_n \in \mathbb{F}$ such that

$$\begin{aligned} v &= c_1v_1 + c_2v_2 + \cdots + c_nv_n \\ x &= d_1v_1 + d_2v_2 + \cdots + d_nv_n. \end{aligned}$$

By definition of T, we then get

$$\begin{aligned} T(v) &= c_1w_1 + c_2w_2 + \cdots + c_nw_n \\ T(x) &= d_1w_1 + d_2w_2 + \cdots + d_nw_n. \end{aligned}$$

Direct computation then shows that

$$aT(v)+bT(x) = a(c_1w_1+c_2w_2+\cdots+c_nw_n)+b(d_1w_1+d_2w_2+\cdots+d_nw_n) = (ac_1+bd_1)w_1+(ac_2+bd_2)w_2+\cdots+(ac_n+bd_n)w_n.$$

On the other hand,

$$av+bx = a(c_1v_1+c_2v_2+\cdots+c_nv_n)+b(d_1v_1+d_2v_2+\cdots+d_nv_n) = (ac_1+bd_1)v_1+(ac_2+bd_2)v_2+\cdots+(ac_n+bd_n)v_n.$$

The definition of T then immediately implies

$$T(av + bx) = (ac_1 + bd_1)w_1 + (ac_2 + bd_2)w_2 + \cdots + (ac_n + bd_n)w_n = aT(v) + bT(w),$$

as we needed to show.

Now, to check that $T(v_1) = w_1, T(v_2) = w_2$, and so on, let's first note that $v_1 = 1 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_n$. The definition of T thus gives

$$T(v_1) = 1 \cdot w_1 + 0 \cdot w_2 + \cdots + 0 \cdot w_n = w_1.$$

Arguing similarly, we get $T(v_2) = w_2, \ldots, T(v_n) = w_n$.

Now that we know such a transformation T exists, let's explain why T is unique. Suppose $S : V \to W$ is also a linear transformation such that $S(v_1) = w_1, S(v_2) = w_2, \ldots, S(v_n) = w_n$. Our goal is to show that for any vector $v \in V$, we have $T(v) = S(v)$. This will show that $T = S$ as functions. Given such a vector $v \in V$, choose the unique scalars $c_1, \ldots, c_n \in \mathbb{F}$ such that $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$. If we apply S to both sides, and use the linearity properties of S, we get

$$\begin{aligned}
S(v) &= S(c_1v_1 + c_2v_2 + \cdots + c_nv_n) \\
&= S(c_1v_1) + S(c_2v_2) + \cdots + S(c_nv_n) \\
&= c_1S(v_1) + c_2S(v_2) + \cdots + c_nS(v_n) \\
&= c_1w_1 + c_2w_2 + \cdots + c_nw_n.
\end{aligned}$$

However, this last line exactly matches the definition of $T(v)$. Therefore, we conclude that $S = T$, so that T is the unique linear transformation from V to W satisfying the condition that $T(v_1) = w_1, T(v_2) = w_2, \ldots, T(v_n) = w_n$. □

The Vector Space of Linear Transformations

Now that we've defined and briefly studied linear transformations, they afford us with yet another example of a vector space. In order to do this, we must define what addition and scalar multiplication of linear transformations look like:

Definition 5.2: Addition and Scalar Multiplication of Linear Transformations

Given two vector spaces V and W, let $\mathcal{L}(V, W)$ denote the set of all linear transformations from V to W. Given two linear transformations $T, S \in \mathcal{L}(V, W)$, we can define their *sum* $T + S \in \mathcal{L}(V, W)$ to be

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

for all vectors $v \in V$. Similarly, given a scalar $c \in \mathbb{F}$, we can define a *scalar multiple* transformation $cT \in \mathcal{L}(V, W)$ to be

$$(cT)(v) = c(T(v)).$$

for all vectors $v \in V$.

Of course, as part of this definition, it is necessary to check that if T, S are both linear transformations from V to W, and $c \in \mathbb{F}$ is a scalar, then $T + S$ is a linear transformation, and cT is a linear transformation. These are left as excellent exercises for you!

Furthermore, we have

Theorem 5.4

For any vector spaces V, W over $\mathbb{F}$, the set $\mathcal{L}(V, W)$ is also a vector space over $\mathbb{F}$, with the scalar multiplication as given above.

The proof of this theorem is also left as an exercise for you. (What is the zero vector here? What is the additive inverse of a given transformation?)

The Kernel and Range of a Linear Transformation

Once a linear transformation is defined from a vector space V to a vector space W, there are two subspaces of particular interest that we associate with it. One is a subspace of V, called the *kernel*, and the other is a subspace of W, called the *range*. At least in some form, both concepts should be familiar to you from your previous work with linear algebra and with functions.

Definition 5.3: Kernel and Range

Let V and W be vector spaces over $\mathbb{F}$, and let $T : V \to W$ be a linear transformation. The *kernel* of T, often denoted $\ker T$, consists of all the vectors in V that map to the zero vector in W. More symbolically,

$$\ker T = \{v \in V : T(v) = 0_W\}.$$

The *range* of T, often denoted $\operatorname{ran} T$, consists of all the vectors in W that are of the form $T(v)$ for some vector $v \in V$. More symbolically,

$$\operatorname{ran} T = \{w \in W : w = T(v) \text{ for some } v \in V\}.$$

As promised, let's first verify that both the kernel and range of a linear transformation are subspaces:

Theorem 5.5

For any vector spaces V and W and any linear transformation $T : V \to W$, $\ker T$ is a subspace of V and $\operatorname{ran} T$ is a subspace of W.

Proof. Suppose we are given a linear transformation $T : V \to W$. We will apply the Subspace Test to show that both $\ker T$ and $\operatorname{ran} T$ are subspaces. First, we check that $\ker T$ is a subspace of V. Note that $0_V \in V$, because $T(0_V) = 0_W$ by Theorem 5.2. Now, suppose we are given vectors $v_1, v_2 \in \ker T$, and a scalar $c \in \mathbb{F}$. By definition of $\ker T$, we know that $T(v_1) = T(v_2) = 0_W$. By linearity of T, this tells us

$$\begin{aligned} T(v_1 + v_2) &= T(v_1) + T(v_2) = 0_W + 0_W = 0_W \\ T(cv_1) &= cT(v_1) = c \cdot 0_W = 0_W. \end{aligned}$$

These two verifications imply that $v_1 + v_2 \in \ker T$ and $cv_1 \in \ker T$, so that $\ker T$ is closed under both vector addition and scalar multiplication. By the Subspace Test, we have now shown that $\ker T$ is a subspace of V.

Next, we show that $\operatorname{ran} T$ is a subspace of W. The equation $T(0_V) = 0_W$ also shows that $0_W \in \operatorname{ran} T$, so $\operatorname{ran} T$ contains the zero vector. Next, suppose we have $w_1, w_2 \in \operatorname{ran} T$ and $c \in \mathbb{F}$. By definition of $\operatorname{ran} T$, this means there are vectors $v_1, v_2 \in V$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. We wish to show that $w_1 + w_2 \in \operatorname{ran} T$ and that $cw_1 \in \operatorname{ran} T$. By direct computation, we see that

$$\begin{aligned} w_1 + w_2 &= T(v_1) + T(v_2) = T(v_1 + v_2) \\ cw_1 &= cT(v_1) = T(cv_1), \end{aligned}$$

which is enough to establish that both $w_1 + w_2$ and cw_1 belong to the range of T. Thus $\operatorname{ran} T$ is closed under vector addition and scalar multiplication, so the Subspace Test again applies, to prove that $\operatorname{ran} T$ is a subspace of W. $\square$

Now, let's take a look at the kernel and range in the case of several of the examples considered earlier:

Example 5.5

(1) Suppose we consider the identity map $I : V \to V$ introduced in Example 5.1, part (1). Note that $\ker I = \{0_V\}$, the zero subspace of V, because every vector is mapped to itself, and so only the zero vector of V is mapped to 0_V. On the other hand, $\operatorname{ran} I = V$, because for each vector $v \in V$, we have $I(v) = v$.

(2) Now, suppose V and W are vector spaces and we consider the zero map $Z : V \to W$ introduced in Example 5.1, part (2). Here, we claim that $\ker Z = V$. Indeed, for every vector $v \in V$, we have $Z(v) = 0_W$ by definition, so that $v \in \ker Z$. On the other hand, we note that $\operatorname{ran} Z = \{0_W\}$, because every vector in V is mapped to 0_W by definition.

Example 5.6

Consider the trace function $\operatorname{tr} : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ introduced in Example 5.2. Here, more or less by definition, $\ker \operatorname{tr}$ is equal to the set of matrices $A = (a_{ij})$ for which the sum of the diagonal entries, $\sum_{i=1}^n a_{ii}$, is equal to 0. On the other hand, $\operatorname{ran} \operatorname{tr} = \mathbb{F}$. To see why, note that for any scalar $a \in \mathbb{F}$, we have

$$\operatorname{tr}\begin{pmatrix} a & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} = a + 0 + 0 + \cdots + 0 = a,$$

which means that $a \in \operatorname{ran} \operatorname{tr}$.

Example 5.7

For a more involved computational example, suppose we wished to find the kernel and range of the linear transformation $T : M_{2\times 2}(\mathbb{F}) \to \mathbb{F}^2$ introduced in Example 5.4. We will start with $\operatorname{ran} T$, because that is the easier of the two computations. Just from knowing the value of T on the four basis vectors for $M_{2\times 2}(\mathbb{F})$, we know in particular that $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are both in $\operatorname{ran} T$, because they are mapped to by two of the vectors in B. Now, $\operatorname{ran} T$ is a subspace of $\mathbb{F}^2$ that contains $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, so by definition of the span, we find that $\operatorname{ran} T$ contains $\operatorname{Span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}$, which is equal to all of $\mathbb{F}^2$ (why?). So we conclude that $\operatorname{ran} T = \mathbb{F}^2$.

As for $\ker T$, we look for which linear combinations of basis vectors in B map to the zero vector. We saw in Example 5.4 that

$$T\left(a\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} + b\begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} + d\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right) = a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 1 \end{pmatrix} + d\begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} a-d \\ c-d \end{pmatrix}.$$

We would like to know when the output of T is the zero vector. This amounts to the conditions $a - d = 0$ and $c - d = 0$ on the coefficients in the linear combination. These reduce to $a = d$ and $c = d$, leaving b and d as our free parameters. From this, we conclude that

$$\begin{aligned} \ker T &= \left\{d\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} + b\begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} + d\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} : b, d \in \mathbb{F}\right\} \\ &= \left\{b\begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} + d\begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} : b, d \in \mathbb{F}\right\} \\ &= \operatorname{Span}\left\{\begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix}\right\}. \end{aligned}$$