Math 291-1: MENU Linear Algebra and Multivariable Calculus
Northwestern University, Fall 2021

Aaron Peterson

Last updated: November 24, 2021

Contents

Lecture 1: Introduction and Complex Numbers

Learning Objectives:

- Introduce the broad goals of the course.
- Introduce the basic notions of mathematical proof, and practice the elements of mathematical logic and style.
- Discuss the real and complex number fields, and prove basic properties of them.

Welcome to MATH 291! This quarter we will study linear algebra through rigorous proof.

What is Linear Algebra?

The answer to the question "What is linear algebra?" has evolved over time. Initially linear algebra was the study of systems of linear equations. Later, linear algebra was the study of the algebraic properties of vectors and matrices. Linear algebra is usually described now as the study of vector spaces (also called linear spaces) and linear transformations. Each viewpoint generalizes its predecessors, but the earlier viewpoints (and their relevant ideas) remain as useful today as ever. The modern notion of vector space is general enough to touch nearly all of modern mathematics, but the old techniques for solving systems of linear equations remain indispensable even in more general contexts.

In this course we will approach the subject through vectors and matrices, emphasizing geometric intuition and anticipating the transition to multivariable calculus later in the year. Along the way we will see how systems of linear equations can be modeled and analyzed using vectors and matrices, and we will also see the generalization to vector spaces.

What is Proof?

Besides linear algebra, you will also study mathematical proof-writing in MATH 291. This entails producing logically sound, stylistically appropriate arguments ("proofs") that establish the validity of a mathematical statement. These arguments are written as prose, often with mathematical notation replacing expressions that would be complicated to write out in words. For example, we would write

> If $x^2 = 5$, then $x = \sqrt{5}$ or $x = -\sqrt{5}$.

instead of

> If the square of an unknown number is five, then that number is either the square root of five, or the negative square root of five.

or

> $x^2 = 5 \Rightarrow x = \sqrt{5}, -\sqrt{5}$

The first of these statements is an efficient and correctly formed sentence that expresses a mathematical idea, and which employs some mathematical notation for shorthand. The second statement is identical to the first, but without the efficiency created by using some mathematical notation. The third is a bit more difficult to read as a sentence. I point this out to emphasize that you should strive to communicate all of your mathematical reasoning in a style similar to that of the first statement above, rather than the second and third. Proof is a contextual expression of logic, and we will learn how to wield standard logical devices and constructions. We will also learn the standard grammar of mathematics and master the art of communicating with other mathematicians through aesthetically pleasing logical proofs.

Real Numbers

The linear algebra is built on number systems known as *fields*. We will focus on two such number systems in this course: the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$. When a result, argument, or an example is valid for both of these, then we will write $\mathbb{K}$ to stand for either $\mathbb{R}$ or $\mathbb{C}$. In the context of linear algebra we refer to numbers in the field we are working with as **scalars**.

Informally, a field is a number system where addition, subtraction, multiplication, and division work exactly as you are used to. More precisely, we say that the real number system is a **field** because the operations of addition $+$ and multiplication $\cdot$ on $\mathbb{R}$ satisfy the following properties:

(Associativity) For every $x, y, z \in \mathbb{R}$, $(x + y) + z = x + (y + z)$ and $(xy)z = x(yz)$.

(Commutativity) For every $x, y \in \mathbb{R}$, $x + y = y + x$ and $xy = yx$.

(Identities) There exist unique numbers $0, 1 \in \mathbb{R}$ such that $0 \neq 1$ and for each $x \in \mathbb{R}$, $x + 0 = x$ and $x1 = x$. The number 0 is called the **additive identity**, and the number 1 is called the **multiplicative identity**.

(Inverses) Every $x \in \mathbb{R}$ has a unique **additive inverse** $-x \in \mathbb{R}$ satisfying $x + (-x) = 0$. Every nonzero $y \in \mathbb{R}$ has a unique **multiplicative inverse** $\frac{1}{y}$ satisfying $y\frac{1}{y} = 1$.

(Distributivity) For every $x, y, z \in \mathbb{R}$, $x(y + z) = xy + xz$ and $(x + y)z = xz + yz$.

General Notation: $\in$

For a set A, the notation "$x \in A$" means that the object x *is an element of* A, or that x is *in* A.

$\forall$: The Universal Quantifier

For a set A, the notation "$\forall x \in A$" reads "for every x in A" or "for all x in A". It precedes a claim involving x that could be either true or false, such as "x is a dog" or "x does not belong to the collection of rational numbers"). This is a **quantifier**, which is a condition that indicates to how many $x \in A$ the claim applies.

This quantifier is called **universal** because it indicates that the statement following it should hold for every x in the set A, without exception.

Occasionally there are multiple universal quantifiers in a given statement. For example, the statement that addition is commutative says that "For every $x, y \in \mathbb{R}$, $x + y = y + x$. We can write this symbolically as

$$\underbrace{\underbrace{(\forall x \in \mathbb{R})}_{\text{For all } x \text{ in } \mathbb{R},} \underbrace{(\forall y \in \mathbb{R})}_{\text{for all } y \text{ in } \mathbb{R},}}_{\text{Shortened: For all } x,y \in \mathbb{R},} (x + y = y + x).$$

∃: The Existential Quantifier

For a set A, the notation "$\exists x \in A$" reads "there exists x in A" or "there is x in A". This is another quantifier because it indicates to how many $x \in A$ the claim applies.

This quantifier is **existential** because it indicates that the statement following it should hold for at least one x in the set A. That is, that there exists at least one $x \in A$ to which the claim applies.

Occasionally you will see a mixture of universal and existential quantifiers in a given statement. For example, the statement about the existence of an additive identity says that "There exists $0 \in \mathbb{R}$ such that for every $x \in \mathbb{R}$, $x + 0 = x$." We can write this symbolically as

$$\underbrace{(\exists 0 \in \mathbb{R})}_{\text{There exists } 0 \text{ in } \mathbb{R},} \underbrace{(\forall x \in \mathbb{R})}_{\text{for all } x \text{ in } \mathbb{R},}(x + 0 = x).$$

The order here (i.e. that the existential claim comes before the universal claim) is important. As stated, there should be a number 0 with the property that for every $x \in \mathbb{R}$ we have $x + 0 = x$. In particular, this same 0 should work regardless of what x is, so we are not allowed to choose 0 based on x. On the other hand, if we change the order of the quantifiers to be "For every $x \in \mathbb{R}$, there exists $0 \in \mathbb{R}$ with $x + 0 = x$." then we would allow 0 to possibly depend on x, which is not quite strong enough to be useful.

For another example, consider the statement "For every person x, there exists a height y such that x is y meters tall." Note that here the existential quantifier comes after the universal quantifier, so that the height y associated to x is allowed to depend on x. If we switch the quantifiers, we obtain the patently false statement "There exists a height y such that for every person x, x is y meters tall," which says that every person is exactly the same height (y meters) as every other person!

These two examples illustrate how subtle differences in wording can drastically change the overall meaning of a logical expression, and therefore why it is so important to be precise.

The Complex Numbers

The complex numbers are defined in terms of real numbers. Here is one of many ways to do this.

Definition 1. A **complex number** is an expression of the form $a + ib$, where $a, b \in \mathbb{R}$ and the symbol i is called the **imaginary unit**. We denote the set of complex numbers by

$$\mathbb{C} \overset{def}{=} \{a + ib \ : \ a, b \in \mathbb{R}\}.$$

Addition and multiplication of complex numbers are defined via the formulas

$$(a + ib) + (c + id) \overset{def}{=} (a + c) + i(b + d) \quad \text{and} \quad (a + ib)(c + id) \overset{def}{=} (ac - bd) + i(ac + bd)$$

for every $a + ib, c + id \in \mathbb{C}$.

Remark 1. Note that if $a, b \in \mathbb{R}$ then we have

$$(a + 0i) + (b + 0i) = a + b + 0i \quad \text{and} \quad (a + 0i)(b + 0i) = ab + 0i,$$

so that we can think of the real numbers as the subset of the complex numbers where the real number a corresponds to the complex number $a + 0i$. To simplify notation with complex numbers, we will simply write a for $a + 0i$ and bi for $0 + bi$ going forward.

Remark 2. With the observation in the last remark, note that for $a, b \in \mathbb{R}$ we have

$$a + bi = (a + i0) + (b + i0)(0 + i1) = (a + i0) + (0 + ib) = a + ib,$$

so that the expression $a + bi$ represents the complex number $a + ib$. (We will use the two expressions interchangeably going forward.) The upshot here is that you can view the expression $a + ib$ not only as a complex number as defined above, but as a complex number built from adding the real number a to the product of the imaginary unit $i = 0 + i1$ and the real number b. This is a little pedantic, but again we are trying to be precise!

Remark 3. Note that

$$i^2 = (0 + 1i)(0 + 1i) = (0 - 1) + (0 - 0)i = -1.$$

In this sense, i functions as a square root of -1. There is no real number with this property, and therefore the presence of a square root of -1 distinguishes $\mathbb{C}$ from $\mathbb{R}$. It also implies far deeper (and more important) differences.

In particular, i is a solution to the polynomial equation $x^2 + 1 = 0$, which has no real solution. One amazing property of the complex numbers is that *every* polynomial equation $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$ (with $n \geq 1$ and $a_n \neq 0$) has a (complex) solution. This fact, known as the **Fundamental Theorem of Algebra**, is neither fundamental nor a theorem of algebra. You will see a proof of this result in a course in Complex Analysis, or Algebraic Topology, or (perhaps) Abstract Algebra (although proofs given in Abstract Algebra need something from analysis, usually the Intermediate Value Theorem). We'll say more about this result next quarter.

Remark 4. Let $z = a + ib$ be a complex number, where $a, b \in \mathbb{R}$. We call a and b (respectively) the **real part** and **imaginary part** of z, denoted by $\mathrm{Re}\,(z) \overset{def}{=} a$ and $\mathrm{Im}\,(z) \overset{def}{=} b$.

Note that if $z, w \in \mathbb{C}$, then $z = w$ exactly when $\mathrm{Re}\,(z) = \mathrm{Re}\,(w)$ and $\mathrm{Im}\,(z) = \mathrm{Im}\,(w)$.

The complex numbers are also a field.

Theorem 1. The complex numbers form a field, in the sense that the operations of addition and multiplication satisfy the following properties.

(Associativity) For every $z, w, u \in \mathbb{C}$, $(z + w) + u = z + (w + u)$ and $(zw)u = z(wu)$.

(Commutativity) For every $z, w \in \mathbb{C}$, $z + w = w + z$ and $zw = wz$.

(Identities) There exist unique numbers $0, 1 \in \mathbb{C}$ such that $0 \neq 1$ and for each $z \in \mathbb{C}$, $z + 0 = z$ and $z1 = z$. The number 0 is called the **additive identity**, and the number 1 is called the **multiplicative identity**.

(Inverses) Every $z \in \mathbb{C}$ has a unique **additive inverse** $-z \in \mathbb{C}$ satisfying $z + (-z) = 0$. Every nonzero $w \in \mathbb{C}$ has a unique **multiplicative inverse** $\frac{1}{w}$ satisfying $w\frac{1}{w} = 1$.

(Distributivity) For every $z, w, u \in \mathbb{C}$, $z(w + u) = zw + zu$ and $(z + w)u = zu + wu$.

Proof. We will prove the properties involving multiplication here, and you will prove the properties involving addition (including distributivity) on your homework.

For associativity of multiplication, let $x+iy, r+is, p+iq \in \mathbb{C}$. Then the definition of complex multiplication combined with associativity and distributivity of real number multiplication and commutativity of real number addition yields

$$\begin{aligned}((x+iy)(r+is))(p+iq) &= ((xr-ys)+i(xs+yr))(p+iq) \\ &= ((xr-ys)p-(xs+yr)q)+i((xr-ys)q+(xs+yr)p) \\ &= (xrp-ysp-xsq-yrq)+i(xrq-ysq+xsp+yrp) \\ &= (xrp-xsq-yrq-ysp)+i(xrq+xsp-yrp-ysq) \\ &= (x(rp-sq)-y(rq+sp))+i(x(rq+sp)+y(rp-sq)) \\ &= (x+iy)((rp-sq)+i(rq+sp)) \\ &= (x+iy)((r+is)(p+iq)).\end{aligned}$$

Proving Statements Involving Universal Quantifiers

To prove a statement with a universal quantifier, we must show that the claim holds for every possible case without exception. Therefore our proof should reflect this. For example, to prove that "Every Northwestern student is wearing purple today." we must somehow show that *every* student at Northwestern is wearing something purple today. Checking only a few students would not suffice.

When proving a statement involving a universal quantifier, we must start by supposing that we are dealing with a (generic) instance of the object in question. For example, our proof of the statement of commutativity of complex number multiplication should begin with "Let $z, w \in \mathbb{C}$." This indicates that the only thing we are assuming about z and w are that they are complex numbers. We will then argue that $zw = wz$, which will show that the product of any two complex numbers does not depend on the order in which we compute the product. There are several different standard phrases used to indicate a universal quantifier, but this is the most basic one. We will point out the variations (and how to deal with them when writing proofs) when we encounter them.

For commutativity of multiplication, let $x+iy, r+is \in \mathbb{C}$. Then the definition of complex multiplication and commutativity of real number multiplication yields

$$(x+iy)(r+is) = (xr-ys)+i(xs+yr) = (rx-sy)+i(sx+yr) = (r+is)(x+iy).$$

Proving Statements Involving Existential Quantifiers

To prove a statement with a existential quantifier, we must show that there is at least one object with the specified properties. Therefore we must somehow produce such an object, and then argue that the object we produced satisfies the claim. For example, to prove that "There is a Northwestern student who is wearing purple today." we must somehow wrangle up at least one student and demonstrate that they are wearing purple. A single student would do.

When proving a statement involving an existential quantifier, we must somehow produce the required object, and then argue that it satisfies the claim. For example, in the proof of the existence of an additive identity in $\mathbb{C}$, we will actually define the object that should be "0", and then prove that this object satisfies the "additive identity" property.

For the multiplicative identity, define $1 \stackrel{def}{=} 1 + 0i$. Let $x + iy \in \mathbb{C}$. Then

$$(x + iy)1 = (x1 - y0) + i(x0 + y1) = x + iy.$$

To show uniqueness of the multiplicative identity, we must show that there cannot be a different multiplicative identity. To this end, suppose that $1' \in \mathbb{C}$ is a multiplicative identity. Then $1' = 1'1 = 11' = 1$.

We now show existence and uniqueness of multiplicative inverses. Let $w = r + is \in \mathbb{C}$, and assume that $w \neq 0$. Then at least one of r or s is nonzero, so that $r^2 + s^2 > 0$. Define

$$\frac{1}{w} \stackrel{def}{=} \frac{r}{r^2 + s^2} + \frac{-s}{r^2 + s^2} i.$$

Then

$$w\frac{1}{w} = (r + si)\Big(\frac{r}{r^2 + s^2} + \frac{-s}{r^2 + s^2} i\Big) = \frac{r^2 + s^2}{r^2 + s^2} + \frac{-rs + sr}{r^2 + s^2} i = 1 + 0i = 1.$$

Therefore $\frac{1}{w}$ is a multiplicative inverse of w. For uniqueness, suppose that $b \in \mathbb{C}$ is another multiplicative inverse for w. Then

$$b = b1 = b\Big(w\frac{1}{w}\Big) = (bw)\frac{1}{w} = 1\frac{1}{w} = \frac{1}{w}.$$

□

The Real and Complex Numbers

The Real Numbers $\mathbb{R}$

The real number system $\mathbb{R}$ (with the usual $+$ and $\cdot$) is a **field**:

> **(Associativity)** For every $x, y, z \in \mathbb{R}$, $(x + y) + z = x + (y + z)$ and $(xy)z = x(yz)$.
>
> **(Commutativity)** For every $x, y \in \mathbb{R}$, $x + y = y + x$ and $xy = yx$.
>
> **(Identities)** There exist unique numbers $0, 1 \in \mathbb{R}$ such that $0 \neq 1$ and for each $x \in \mathbb{R}$, $x + 0 = x$ and $x1 = x$. The number 0 is called the **additive identity**, and the number 1 is called the **multiplicative identity**.
>
> **(Inverses)** Every $x \in \mathbb{R}$ has a unique **additive inverse** $-x \in \mathbb{R}$ satisfying $x + (-x) = 0$. Every nonzero $y \in \mathbb{R}$ has a unique **multiplicative inverse** $\frac{1}{y}$ satisfying $y\frac{1}{y} = 1$.
>
> **(Distributivity)** For every $x, y, z \in \mathbb{R}$, $x(y + z) = xy + xz$ and $(x + y)z = xz + yz$.

The Complex Numbers $\mathbb{C}$

Theorem 2. The complex numbers form a field, in the sense that the operations of addition and multiplication satisfy the following properties.

(Associativity) For every $z, w, u \in \mathbb{C}$, $(z + w) + u = z + (w + u)$ and $(zw)u = z(wu)$.

(Commutativity) For every $z, w \in \mathbb{C}$, $z + w = w + z$ and $zw = wz$.

(Identities) There exist unique numbers $0, 1 \in \mathbb{C}$ such that $0 \neq 1$ and for each $z \in \mathbb{C}$, $z + 0 = z$ and $z1 = z$. The number 0 is called the **additive identity**, and the number 1 is called the **multiplicative identity**.

(Inverses) Every $z \in \mathbb{C}$ has a unique **additive inverse** $-z \in \mathbb{C}$ satisfying $z+(-z) = 0$. Every nonzero $w \in \mathbb{C}$ has a unique **multiplicative inverse** $\frac{1}{w}$ satisfying $w\frac{1}{w} = 1$.

(Distributivity) For every $z, w, u \in \mathbb{C}$, $z(w + u) = zu + wu$ and $(z + w)u = zu + wu$.

The Complex Plane

Absolute Value and Conjugation

Proposition 1 (Properties of Conjugation)**.** The following hold for every $z, w \in \mathbb{C}$.

(a) $\bar{\bar{z}} = z$

(b) $\operatorname{Re}(z) = \dfrac{z+\bar{z}}{2}$ and $\operatorname{Im}(z) = \dfrac{z-\bar{z}}{2i}$

(c) $\bar{z} = z$ if, and only if, $\operatorname{Im}(z) = 0$ (i.e. if z is real).

(d) $z\bar{z} = |z|^2$

(e) $\overline{z+w} = \bar{z} + \bar{w}$, $\overline{zw} = \bar{z}\bar{w}$, and if $z \neq 0$, then $\overline{\left(\frac{1}{z}\right)} = \frac{1}{\bar{z}}$.

Lecture 2: More Complex Numbers

Learning Objectives:

- Discuss the real and complex number fields, and prove basic properties of them.
- Investigate first examples of several proof techniques, including contradiction and contraposition.

We can derive sophisticated properties of addition, subtraction, multiplication, and division using the fundamental ones above. We will not prove every possible property, but here are a couple examples to indicate what the arguments look like.

Remark 5. In class we looked at Example 1 and Example 4 as written, and our discussion led us to consider the other examples below as a means to illustrate a couple more proof techniques. This was a little awkward because the order of these extra results weren't planned, but I've written them up below in a more thoughtful ordering.

Example 1. For every $z \in \mathbb{K}$, $z0 = 0$.

Proof. Let $z \in \mathbb{K}$. Then since $0 = 0 + 0$ and by the distributivity of multiplication over addition,

$$z0 = z(0 + 0) = z0 + z0.$$

Therefore

$$0 = z0 + (-(z0)) = (z0 + z0) + (-(z0)) = z0 + (z0 + (-(z0))) = z0 + 0 = z0.$$

□

Example 2. The additive identity $0 \in \mathbb{K}$ does not have a multiplicative inverse.

Proof. We proceed by mean of contradiction. Suppose that 0 has a multiplicative inverse $w \in \mathbb{K}$. Then we have $0w = 1$. But the last result implies that $0w = 0$, so that $0 = 1$. But this contradicts the fact that $0 \neq 1$ in $\mathbb{K}$. Therefore 0 does not have a multiplicative inverse. □

Proof by Contradiction

One possible way of attempting to prove an implication of the form "If P, then Q." is to assume that P holds and that Q does **not** hold, and then argue that there is a contradiction (i.e. a statement that we know to be false because we know that its negation is true). Therefore if P holds, then Q must also hold because it is impossible for "not Q" to hold.
In the previous example, we attempted to prove an implication of the form "If $z = 0$, then z does not have a multiplcative inverse." Here P is the claim "$z = 0$", and Q is the claim "z does not have a multiplicative inverse." We assumed that $z = 0$ (P) and that z *does* have a multiplcative inverse (not Q), and showed that $0 = 1$, which contradicts the fact that $0 \neq 1$. Therefore it is impossible for (not Q) to hold if P holds, so that Q holds whenever P holds.

Contrapositive of a Logical Statement

Sometimes there are equivalent ways of phrasing the same logical statement. One standard example is the **contrapositive** of a logical implication. The contrapositive of an implication of the form “If P, then Q.” is the statement “If Q is false, then P is false.” For example, the statements

> ”If $z = 0$, then z does not have a multiplicative inverse.”

and

> If z has a multiplicative inverse, then $z \neq 0$.

are contrapositives of each other, and therefore are logically equivalent (i.e. that are each valid, or they are each invalid).
We can actually *prove* that an implication and its contrapositive are equivalent using contradiction. Assume that “If P, then Q.” holds, and suppose that Q does not hold. If P held, then Q would also hold. Because Q does not hold by assumption, we have a contradiction. Therefore P must not hold. In other words, “if Q is false then P is also false.” The same argument shows that if the implication “If Q is false, then P is false.” holds, then “If P, then Q.” must also hold.

Example 3. Let $z \in \mathbb{K}$. If z has a multiplicative inverse, then $z \neq 0$.

Proof. Note that the statement “If z has a multiplicative inverse, then $z = 0$.” is exactly the contrapositive of the claim “If $z = 0$, then z does not have a multiplicative inverse,” which we proved in Example 2. □

Example 4. Let $z, w \in \mathbb{K}$. If $zw = 0$, then either $z = 0$ or $w = 0$.

Proving Implications

Here we are asked to prove an implication of the form “If P, then Q,” where P is the statement “$zw = 0$” and Q is the statement “either $z = 0$ or $w = 0$.” To prove a statement of the form “If P, then Q”, we must show that Q holds whenever P holds. We are not proving that P holds, but rather we are proving that if we find ourselves in a situation where P holds, then we are justified in concluding that Q must also hold.

Proof. Suppose that $zw = 0$. If $w = 0$ then we are done, so suppose that $w \neq 0$. Then we multiply both sides of the equation $0 = zw$ by $\frac{1}{w}$ and apply the previous result and associativity of multiplication to obtain

$$0 = 0\frac{1}{w} = (zw)\frac{1}{w} = z\Big(w\frac{1}{w}\Big) = z1 = z.$$

□

Example 5. We give a second proof of the previous result: Let $z, w \in \mathbb{K}$. If $zw = 0$, then either $z = 0$ or $w = 0$.

Proof by Contraposition

Because the contrapositive "If Q is false, then P is false." is logically equivalent to "If P, then Q," we can establish a claim of the form "If P, then Q." by proving its contrapositive "If Q is false, then P is false." We will illustrate this here.

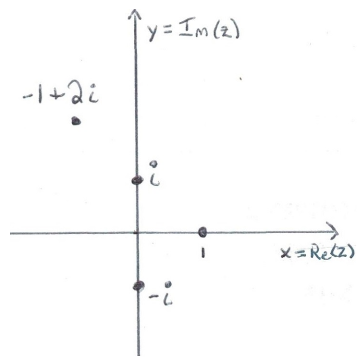
Proof. We proceed by contraposition, and will show that if neither $z = 0$ nor $w = 0$, then $zw \neq 0$. Suppose that neither $z = 0$ nor $w = 0$, so that $z \neq 0$ and $w \neq 0$. Because z, w are nonzero, they have multiplicative inverses $\frac{1}{z}$ and $\frac{1}{w}$. Therefore

$$zw\left(\frac{1}{z}\frac{1}{w}\right) = z\frac{1}{z} \cdot w\frac{1}{w} = 1 \cdot 1 = 1,$$

so that zw has a multiplicative inverse. The result in Example 3 then implies that $zw \neq 0$. □

The Complex Plane

Many of the ideas of complex analysis (including basic ones) can be understood geometrically via the identification of a complex number $z = x + iy$ with the point (x, y) in the usual Cartesian plane. In this way, we think of $\mathbb{C}$ as the **complex plane** (in analogy to your understanding of $\mathbb{R}$ as the real line). The x-axis of the complex plane is called the **real axis**, while the y-axis is called the **imaginary axis**.



Absolute Value and Conjugation

The analogy between complex numbers and points in the plane suggests that we should measure the 'size' $|z|$ of a complex number z, the **absolute value** (or sometimes called **modulus**) of z, as the distance from z to 0 as points in the plane. That is, we define

$$|z| = |x + iy| \stackrel{def}{=} \sqrt{x^2 + y^2}.$$

The properties of the absolute value for complex numbers mirror those for the absolute value of real numbers. In particular, we will soon prove that $|z_1 z_2| = |z_1||z_2|$ for complex numbers z_1, z_2. (We will actually prove a generalization of this result.)

For now, recall that

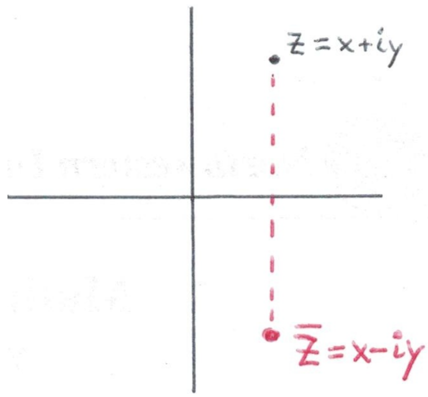
$$|x|^2 = x^2 \quad \text{(and therefore } |x| = \sqrt{x^2}\text{)} \quad \text{for every } x \in \mathbb{R}. \tag{1}$$

This cannot possibly hold for complex numbers, though, since $|i|^2 = 1 \neq -1 = i^2$. To generalize (1) to complex numbers, we need to introduce the notion of complex conjugation.

Definition 2. The **(complex) conjugate** $\bar{z}$ of a complex number $z = x + iy \in \mathbb{C}$ is defined to be

$$\bar{z} = \overline{x + iy} \stackrel{def}{=} x - iy.$$

When thought of in terms of the complex plane, conjugation $z \mapsto \bar{z}$ sends a number z to its reflection across the real axis, $\bar{z}$.



The basic properties of complex conjugation are summarized in the following proposition.

Proposition 2 (Properties of Conjugation)**.** The following hold for every $z, w \in \mathbb{C}$.

(a) $\bar{\bar{z}} = z$

(b) $\mathrm{Re}\,(z) = \dfrac{z + \bar{z}}{2}$ and $\mathrm{Im}\,(z) = \dfrac{z - \bar{z}}{2i}$

(c) $\bar{z} = z$ if, and only if, $\mathrm{Im}\,(z) = 0$ (i.e. if z is real).

(d) $z\bar{z} = |z|^2$

(e) $\overline{z + w} = \bar{z} + \bar{w}$, $\overline{zw} = \bar{z}\bar{w}$, and if $z \neq 0$, then $\overline{\left(\frac{1}{z}\right)} = \frac{1}{\bar{z}}$.

Proof. The proof is very short. Let $z, w \in \mathbb{C}$ and write $z = x + iy$ and $w = r + is$. For (a), we note that

$$\bar{\bar{z}} = \overline{x - iy} = x + iy = z.$$

For (b), we simply compute that

$$\frac{z + \bar{z}}{2} = \frac{x + iy + (x - iy)}{2} = \frac{2x}{2} = x = \mathrm{Re}\,(z) \quad \text{and} \quad \frac{z - \bar{z}}{2i} = \frac{x + iy - (x - iy)}{2i} = \frac{2iy}{2i} = y = \mathrm{Im}\,(z).$$

For (c), suppose that $\bar{z} = z$. Then by (b) we have $\mathrm{Im}\,(z) = \frac{z - \bar{z}}{2i} = 0$. Conversely, suppose that $\mathrm{Im}\,(z) = 0$. Then (b) gives $0 = \frac{z - \bar{z}}{2i}$, so that $z - \bar{z} = 0$, and therefore $z = \bar{z}$.

“If, and only if”

We will regularly see statements of the form “P if, and only if, Q.” (Symbolically: $P \Leftrightarrow Q$.) This is shorthand for the two separate statements “If Q then P.” ($Q \Rightarrow P$, or “P if Q”) and “If P then Q.” ($P \Rightarrow Q$, or “P only if Q”). To prove an “if and only if” statement, you need to prove both implications. Typically you must consider each implication separately. We illustrate this in the proof of Property (c) of Conjugation.

The statement “P if, and only if, Q” is sometimes stated that Q is “necessary and sufficient” for P to hold. “Necessary” means that if P holds, then Q necessarily holds; this is the “If P, then Q.” implication. “Sufficient” means that to conclude that P holds, it is sufficient to know that Q holds; this is the “If Q, then P.” implication.

We prove (d) by writing

$$z\bar{z} = (x+iy)(x-iy) = (x^2+y^2) + i(-xy+yx) = x^2+y^2 = |z|^2.$$

Finally, we turn to (e). The first and second formulas follow by writing

$$\overline{z+w} = \overline{(x+r)+i(y+s)} = (x+r)+i(-y-s) = (x-iy)+(r-is) = \bar{z}+\bar{w}$$

and

$$\overline{zw} = \overline{(xr-ys)+i(xs+yr)} = (xr-ys)-i(xs+yr) = (xr-(-y)(-s))+i(x(-s)+(-y)r) = (x-iy)(r-is) = \bar{z}\bar{w}.$$

For the third claim, suppose that $z \neq 0$. Then $\bar{z} \neq 0$ and the second claim implies that

$$\bar{z}\overline{\left(\frac{1}{z}\right)} = \overline{z\frac{1}{z}} = \bar{1} = 1.$$

By the uniqueness of multiplicative inverses, $\overline{\left(\frac{1}{z}\right)} = \frac{1}{\bar{z}}$. □

Remark 6. Note that part Property (d) of Conjugation gives us a better way to understand where the formula for $\frac{1}{w}$ comes from. If $w = r+is \in \mathbb{C}$ with $w \neq 0$, then at least one of $r, s \neq 0$ so that $|w|^2 = r^2+s^2 > 0$. Then we have

$$\frac{1}{w} = \frac{1}{w}\frac{\bar{w}}{\bar{w}} = \frac{\bar{w}}{|w|^2} = \frac{1}{r^2+s^2}(r-si) = \frac{r}{r^2+s^2} - \frac{s}{r^2+s^2}i,$$

which is exactly the formula we produced in the proof that complex numbers have multiplicative inverses!

There are many properties of real and complex numbers that we wish to use, but which we do not yet have the machinery to prove. For example, consider the following statement:

For every $n \in \mathbb{N}$ with $n \geq 2$ and every $z_1, \ldots, z_n \in \mathbb{C}$, $|z_1 z_2 \cdots z_n| = |z_1||z_2|\cdots|z_n|$.

Here we are not proving one single claim involving arbitrary complex numbers, but an *infinite number of claims*. The claim when $n = 2$ is that for every $z_1, z_2 \in \mathbb{C}$ we have $|z_1 z_2| = |z_1||z_2|$. The claim when $n = 3$ is that for every $z_1, z_2, z_3 \in \mathbb{C}$ we have $|z_1 z_2 z_3| = |z_1||z_2||z_3|$, and so on. You might try to tackle these claims individually, but the core difficulty is immediate: you will never prove an infinite number of claims one-by-one. We need some logical technique that will allow us to establish all of the claims at once. The technique is called the **Principle of Mathematical Induction**.

Lecture 3: Mathematical Induction

Learning Objectives:

- Apply the Principle of Mathematical Induction to prove families of logical statements.

There are many properties of real and complex numbers that we wish to use, but which we do not yet have the machinery to prove. For example, consider the following statement:

For every $n \in \mathbb{N}$ with $n \geq 2$ and every $z_1, \ldots, z_n \in \mathbb{C}$, $|z_1 z_2 \cdots z_n| = |z_1||z_2| \cdots |z_n|$.

Here we are not proving one single claim involving arbitrary complex numbers, but an *infinite number of claims*. The claim when $n = 2$ is that for every $z_1, z_2 \in \mathbb{C}$ we have $|z_1 z_2| = |z_1||z_2|$. The claim when $n = 3$ is that for every $z_1, z_2, z_3 \in \mathbb{C}$ we have $|z_1 z_2 z_3| = |z_1||z_2||z_3|$, and so on. You might try to tackle these claims individually, but the core difficulty is immediate: you will never prove an infinite number of claims one-by-one. We need some logical technique that will allow us to establish all of the claims at once. The technique is called the **Principle of Mathematical Induction**.

The idea behind mathematical induction is as follows. Suppose that $m \in \mathbb{Z}$ and that we have a sequence of logical statements $P(m)$, $P(m+1)$, $P(m+2)$, etc. In the example above, $m = 2$ and $P(2)$ is "For every $z_1, z_2 \in \mathbb{C}$, $|z_1 z_2| = |z_1||z_2|$," $P(3)$ is "For every $z_1, z_2, z_3 \in \mathbb{C}$ we have $|z_1 z_2 z_3| = |z_1||z_2||z_3|$," and so on. Then we can conclude that $P(n)$ holds for every integer $n \geq m$ if we show that

(i) $P(m)$ (i.e. the **base case**, or the "first" statement) holds, and

(ii) If $k \geq m$ and $P(k)$ holds, then $P(k+1)$ also holds.

Part (ii) is called the **induction step**, and the hypothesis that $P(k)$ holds is called the **induction hypothesis**.

Let's see how this plays out. Once we establish (i), then we know that $P(m)$ holds. Once we establish (ii), then we know that because $P(m)$ holds, $P(m+1)$ also holds. But then (ii) shows that because $P(m+1)$ holds, $P(m+2)$ must also hold, and so on. In short, the Principle of Mathematical Induction then implies that $P(n)$ holds for every integer $n \geq m$. If we view $P(m)$, $P(m+1)$, $P(m+2)$, etc. as rungs on a ladder, then the base case shows that we can get onto the bottom rung of the ladder, and the induction step shows that if we are standing on any given rung of the ladder, then we can climb onto the next rung. In this analogy, the conclusion is that we will eventually climb past each rung of the ladder. Here is an example.

Proposition 3. For every $n \in \mathbb{N}$ with $n \geq 2$ and every $z_1, \ldots, z_n \in \mathbb{C}$, $|z_1 z_2 \cdots z_n| = |z_1||z_2| \cdots |z_n|$.

Proof. We proceed by mathematical induction.

Base Case: For the case where $n = 2$, let $z_1, z_2 \in \mathbb{C}$ and write $z_1 = a + ib$ and $z_2 = c + id$. Then

$$\begin{aligned} |z_1 z_2| &= |(ac - bd) + i(ad + bc)| \\ &= \sqrt{(ac - bd)^2 + (ad + bc)^2} \\ &= \sqrt{a^2c^2 - 2abcd + b^2d^2 + a^2d^2 + 2abcd + b^2c^2} \\ &= \sqrt{a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2} \\ &= \sqrt{(a^2 + b^2)(c^2 + d^2)} \\ &= \sqrt{(a^2 + b^2)}\sqrt{(c^2 + d^2)} \\ &= |z_1||z_2|. \end{aligned}$$

Induction Step: Let $n \in \mathbb{N}$ with $n \geq 2$.
Induction Hypothesis: Assume that for every $z_1, \ldots, z_n \in \mathbb{C}$, $|z_1 \cdots z_n| = |z_1| \cdots |z_n|$.
Proof of the Induction Step: Let $z_1, \ldots, z_n, z_{n+1} \in \mathbb{C}$. Then the base case (applied to $(z_1 \cdots z_n)$ and z_{n+1}) followed by the induction hypothesis (applied to $z_1 \cdots z_n$) yield

$$|z_1 \cdots z_n z_{n+1}| = |z_1 \cdots z_n||z_{n+1}| = |z_1| \cdots |z_n||z_{n+1}|.$$

By the Principle of Mathematical Induction, the proof is complete. □

On your homework you will prove other properties of complex numbers similar to that in the previous example. Examples include:

- For every $n \in \mathbb{N}$ with $n \geq 2$,

  $$\operatorname{Re}(z_1 + \cdots + z_n) = \operatorname{Re}(z_1) + \cdots + \operatorname{Re}(z_n) \quad \text{and} \quad \operatorname{Im}(z_1 + \cdots + z_n) = \operatorname{Im}(z_1) + \cdots + \operatorname{Im}(z_n)$$

  for every $z_1, \ldots, z_n \in \mathbb{C}$.

- For every $n \in \mathbb{N}$ with $n \geq 2$,

  $$\overline{z_1 + \cdots + z_n} = \overline{z_1} + \cdots + \overline{z_n} \quad \text{and} \quad \overline{z_1 \cdots z_n} = \overline{z_1} \cdots \overline{z_n}$$

  for every $z_1, \ldots, z_n \in \mathbb{C}$.

- For every $n \in \mathbb{N}$ with $n \geq 2$, and every $z_1, \ldots, z_n \in \mathbb{C}$, if $z_1 z_2 \cdots z_n = 0$ then at least one of $z_1, \ldots, z_n = 0$.

Induction is also a great way to prove interesting algebraic formulas.

Proposition 4. For every $n \in \mathbb{N}$, $1 + 2 + 3 + \cdots + n = \dfrac{n(n+1)}{2}$.

Proof. We proceed by mathematical induction.
Base Case: For the base case where $n = 1$, we note that $1 = \frac{1(2)}{2} = \frac{1(1+1)}{2}$.
Induction Step: Let $n \in \mathbb{N}$.
Induction Hypothesis: Assume that $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$.
Proof of the Induction Step: We apply the induction hypothesis to see that

$$1 + \cdots + n + (n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \frac{(n+1)(n+2)}{2} = \frac{(n+1)((n+1)+1)}{2}.$$

By the Principle of Mathematical Induction, the proof is complete. □

Induction is a very general logical device, and can be used to prove all sorts of claims. Here are two.

Proposition 5. For every $n \in \mathbb{N}$, $2^{2n} - 1$ is divisible by 3.

Proof. We proceed by mathematical induction.
Base Case: For the case where $n = 1$, note that $2^{2(1)} - 1 = 4 - 1 = 3$ is divisible by 3.
Induction Step: Let $n \in \mathbb{N}$.
Induction Hypothesis: Assume that $2^{2n} - 1$ is divisible by 3.
Proof of the Induction Step: Because $2^{2n} - 1$ is divisible by 3, there is an integer m with $2^{2n} - 1 = 3m$. But then

$$2^{2(n+1)} - 1 = 4 \cdot 2^{2n} - 1 = 4 \cdot (2^{2n} - 1) + 3 = 4 \cdot 3m + 3 = 3(4m + 1).$$

Because $4m + 1$ is an integer, $2^{2(n+1)} - 1$ is divisible by 3.

By the Principle of Mathematical Induction, the proof is complete. □

Proposition 6. For every $n \in \mathbb{N}$ with $n \geq 4$, $n^2 > 3n$.

Proof. We proceed by mathematical induction.
Base Case: For the case where $n = 4$, note that $4^2 = 16 > 12 = 3(4)$.
Induction Step: Let $n \in \mathbb{N}$ with $n \geq 4$.
Induction Hypothesis: Assume that $n^2 > 3n$.
Proof of the Induction Step: We compute that

$$(n + 1)^2 = n^2 + 2n + 1 > 3n + 2n + 1 = 3n + 2 + 1 = 3(n + 1).$$

By the Principle of Mathematical Induction, the proof is complete. □

Remark 7. Mathematical induction can be taken as a basic principle of logical reasoning, but we can "prove" that induction holds if we instead adopt other basic principles. One of these, the **Well-Ordering Principle**, states that every nonempty subset of the natural numbers contains a least element. That is, for every $S \subseteq \mathbb{N}$, if $S \neq \emptyset$ then there exists $m \in S$ such that for every $k \in S$, $m \leq k$. As it turns out, not only can one prove that the Principle of Mathematical Induction is valid using the Well-Ordering Principle, but one can also prove the converse: if the Principle of Mathematical Induction is valid, then so is the Well-Ordering Principle. This equivalence is explored in one of your homework problems.

Remark 8. We may need a variant of the principle of mathematical induction this year called **strong induction**, where the induction hypothesis is that for each $k \in \mathbb{N}$, if $P(j)$ holds for every $1 \leq j \leq k$, then $P(k+1)$ also holds. In the "ladder" analogy, we are essentially assuming that we have been on the first k rungs of the ladder in order to show that we can climb onto the $k+1$-st rung. This is more than we assume for ordinary induction, where only the knowledge that we are on the k-th rung is used to show that we can climb onto the $k+1$-st rung. We will see an example discuss this more if it is needed.

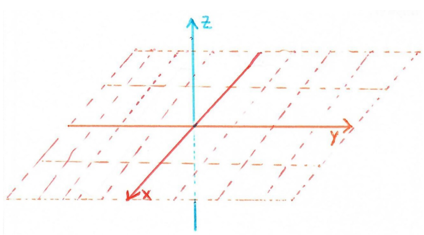
Lecture 4: Vectors

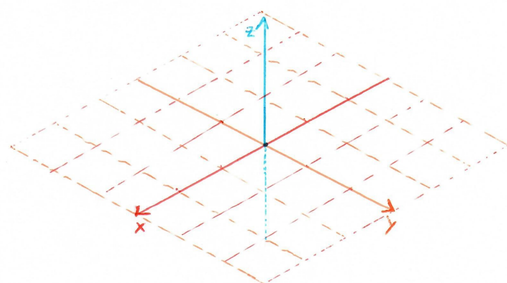
Learning Objectives:

- Define vectors as objects representing displacement in space.
- Interpret vector addition and scalar multiplication geometrically.
- Establish the basic algebraic properties of vector addition and scalar multiplication.
- Compute the length of a real vector, and show how scalar multiplication affects length.

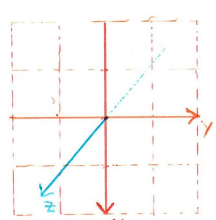
Definition 3. A **point** P in $\mathbb{R}^n$ is an ordered n-tuple $P \overset{def}{=} (a_1, \ldots, a_n)$, where $a_1, \ldots, a_n \in \mathbb{R}$. The numbers $a_1, \ldots, a_n$ are called the **standard coordinates** of P.

We can visualize this coordinate system as follows. In $\mathbb{R}^n$ we fix a point—the **origin**—and draw n mutually-perpendicular lines (copies of the real number line called the **coordinate axes**) passing through the origin. Here are two such drawings for $\mathbb{R}^3$ (with the coordinates of a point labeled (x, y, z)):





Depending on the application, we might sometimes was to change our vantage point when sketching the coordinate axes. For example, the two pictures above represent the same space from different vantage point: by rotating our view in the first picture slightly around the z-axis, we get the second view. Alternatively, we could have rotated the original view around the x-axis origin by 'pulling' the positive z-direction directly towards us to give the following picture:

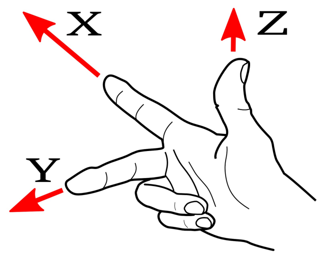


We will always label these axes so that each picture we draw looks like it has the same coordinate system, only perhaps viewed from a different vantage point. The convention is the **right-hand rule**: if you extend your right hand and position your wrist at the origin in such a way that both

- your fingers point along the positive x-axis, and

- when you curl your fingers by 90° they point along the positive y-axis,

then your thumb will point along the positive z-axis. Here is a picture[1]:



A coordinate system labeled in this way is called **right-handed**. To avoid confusion later on in the course, you should *always* draw your coordinate systems to be right-handed.

Location vs. Displacement

A point P is a *location* in $\mathbb{R}^n$, and the standard coordinates of P describe the location of P relative to the **origin** $O \stackrel{def}{=} (0, \ldots, 0)$. *Vectors* in $\mathbb{R}^n$ allow us to capture displacement from one point to another point. In this way, vectors capture *change of location*.

Definition 4. A **vector** $\vec{x}$ in $\mathbb{R}^n$ is a column of scalars $x_1, \ldots, x_n \in \mathbb{R}$, denoted by

$$\vec{x} \stackrel{def}{=} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The scalars $x_1, \ldots, x_n$ are called the **entries** of $\vec{x}$.

The relationship between points and vectors is captured by the following definition.

Definition 5. Let $P \stackrel{def}{=} (a_1, \ldots, a_n)$ and $Q \stackrel{def}{=} (b_1, \ldots, b_n)$ be points in $\mathbb{R}^n$, and let $\vec{v} \stackrel{def}{=} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ be a vector in $\mathbb{R}^n$. Define

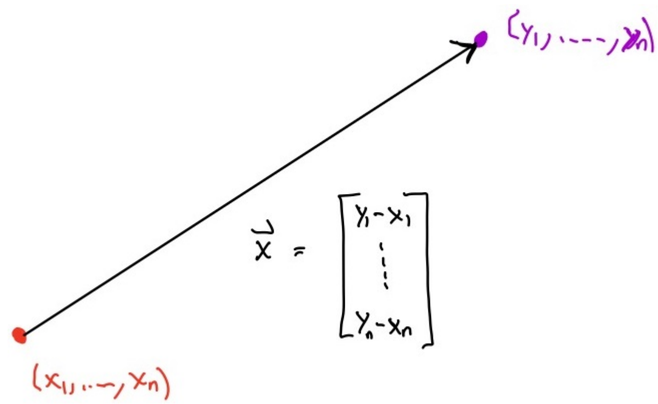
$$Q - P = (b_1, \ldots, b_n) - (a_1, \ldots, a_n) \stackrel{def}{=} \begin{bmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{bmatrix} \tag{2}$$

and

$$P + \vec{v} = (a_1, \ldots, a_n) + \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \stackrel{def}{=} (a_1 + x_1, \ldots, a_n + x_n). \tag{3}$$

[1] This was retrieved from https://stackoverflow.com/questions/19747082/how-does-coordinate-system-handedness-relate-to-rotation-direction-and-vertices.

Remark 9. Because a vector represents displacement from an initial location to a terminal location, we visualize it as a directed line segment (arrow) from the initial location to the terminal location.



As a measure of the size of the displacement represented by a vector, we define the length of a vector in $\mathbb{R}^n$ as follows.

Definition 6. Let $\vec{x} \in \mathbb{R}^n$, and write $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Then we define the **length** (or **norm** or **magnitude**) of $\vec{x}$ to be

$$\|\vec{x}\| \overset{def}{=} \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Remark 10. Note that for a vector $\vec{x}$ in $\mathbb{R}^1$, $\mathbb{R}^2$ and $\mathbb{R}^3$, the length of $\vec{x}$ is exactly the distance from the point $(x_1, \ldots, x_n)$ to the origin $(0, \ldots, 0)$.

Remark 11. Equations (2) and (3) ensure, for example, that

$$P + (Q - P) = (a_1, \cdots, a_n) + \begin{bmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{bmatrix} = (a_1 + (b_1 - a_1), \ldots, a_n + (b_n - a_n)) = (b_1, \ldots, b_n) = Q$$

for every pair of points P and Q in $\mathbb{R}^n$. The vector $Q - P$ describes the displacement from P to Q, and we obtain the point Q by adding the vector $Q - P$ to the point P.

The (seemingly pedantic) distinction between points and vectors will be important in multivariable calculus. Until then, it will be helpful to conflate the notions of point and vector using position vectors.

Definition 7. If $P = (a_1, \ldots, a_n)$ is a point in $\mathbb{R}^n$, then we define the **position vector** $\vec{p}$ of P as

$$\vec{p} \overset{def}{=} P - O = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

Going forward, we will simply write $\vec{p} \in \mathbb{R}^n$ to mean that $\vec{p}$ is a vector in $\mathbb{R}^n$. In certain contexts it will be helpful to interpret $\vec{p}$ as the position vector for a point in $\mathbb{R}^n$, in which case we should interpret $\vec{p}$ as representing this point.

Complex Vectors

Although we will exclusively limit our geometric considerations of vectors to vectors in $\mathbb{R}^n$, as algebraic objects (i.e. arrays of scalars) we are also interested in vectors with complex entries.

Definition 8. A **vector** $\vec{x}$ in $\mathbb{C}^n$ is a column of scalars $x_1, \ldots, x_n \in \mathbb{C}$, denoted by

$$\vec{x} \overset{def}{=} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The scalars $x_1, \ldots, x_n$ are called the **entries** of $\vec{x}$.

Remark 12. We will discuss the algebraic properties of vectors in the general context of $\mathbb{K}^n$ (i.e. either vectors with real entries or vectors with complex entries). We will still continue to rely on the case of $\mathbb{R}^n$ for geometric intuition, although many of our results and constructions immediately generalize to the case of $\mathbb{C}^n$ even if the geometric interpretation does not. Later on in the course we will learn how to understand $\mathbb{C}^n$ in terms of $\mathbb{R}^{2n}$. We will also see some ways that $\mathbb{R}^n$ is algebraically different than $\mathbb{C}^n$, but these will not arise until later.

Vector Arithmetic

There are two fundamental algebraic operations on vectors: vector addition and scalar multiplication.

Definition 9. Let $\vec{x}, \vec{y} \in \mathbb{K}^n$, and let $c \in \mathbb{K}$ be a scalar[2]

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \overset{def}{=} \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

and

$$c\vec{x} = c \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \overset{def}{=} \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix}.$$

Remark 13. In the real case, vector addition and scalar multiplication can be understood through our understanding of how vectors model displacement. For example, if $P = (a_1, \ldots, a_n)$ is a point in $\mathbb{R}^n$ and if $\vec{x}, \vec{y} \in \mathbb{R}^n$ are vectors, then the point $(P + \vec{x}) + \vec{y}$ obtained by displacing P by $\vec{x}$, and then displacing

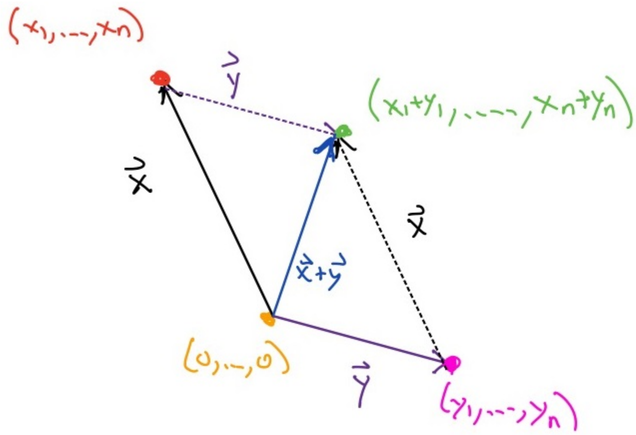
[2] Here and throughout, when $\mathbb{K}$ or $\mathbb{K}^n$ appear in a definition, result, or example, we mean that all instances of $\mathbb{K}$ and $\mathbb{K}^n$ must be interpreted as $\mathbb{R}$ or $\mathbb{R}^n$, or that all instances of $\mathbb{K}$ and $\mathbb{K}^n$ must be interpreted as $\mathbb{C}$ and $\mathbb{C}^n$. In the rare instances where we need to refer to real numbers in some parts and complex in others, we will not use the $\mathbb{K}$ notation.

the result by $\vec{y}$, can be described as displacing P by the single vector $\vec{x}+\vec{y}$, since

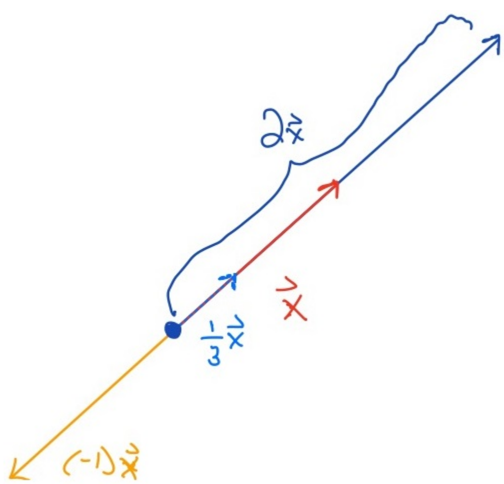
$$\begin{aligned}(P+\vec{x})+\vec{y} &= (a_1+x_1,\ldots,a_n+x_n)+\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}\\ &= ((a_1+x_1)+y_1,\ldots,(a_n+x_n)+y_n)\\ &= (a_1+(x_1+y_1),\ldots,a_n+(x_n+y_n)) = P+(\vec{x}+\vec{y}).\end{aligned}$$

Remark 14. For vectors in $\mathbb{R}^n$, addition and scalar multiplication can be interpreted geometrically[3] through our understanding of vectors as representing displacement.

For $\vec{x},\vec{y}\in\mathbb{R}^n$, the vector $\vec{x}+\vec{y}$ is the displacement obtained by displacing our location by $\vec{x}$, and then displacing our new location by $\vec{y}$. If we started at (say) the origin, then $\vec{x}+\vec{y}$ is the opposite corner of the parallelogram with one vertex at the origin and the vectors $\vec{x}$ and $\vec{y}$ (both drawn starting at the origin) forming adjacent edges. With this interpretation it is also clear that we should have $\vec{x}+\vec{y}=\vec{y}+\vec{x}$, but this is something that we will establish algebraically.



For scalar multiplication, note that if $c\in\mathbb{R}$ then the entries of the vector $c\vec{x}$ are exactly the entries of $\vec{x}$, but where each has been multiplied by c. Therefore displacement by $c\vec{x}$ should place us along the same 'line' through our starting point as would displacement by $\vec{x}$, but we expect that the 'magnitude' (or 'size') of the displacement would change.



[3] Here is a perfect instance of where we are relying on $\mathbb{R}^n$ (say for $n=2$ or $n=3$) for intuition because the same intuition doesn't easily hold for vectors in $\mathbb{C}^n$.

Indeed, note that

$$\|c\vec{x}\| = \sqrt{(cx_1)^2 + \cdots + (cx_n)^2} = \sqrt{c^2}\sqrt{x_1^2 + \cdots + x_n^2} = |c|\|\vec{x}\|,$$

so that scaling $\vec{x}$ by c results in a vector whose length is exactly $|c|$ times as long as the length of $\vec{x}$.

Note also that if $c < 0$ then the signs of the entries of $c\vec{x}$ would be reversed from those of $\vec{x}$, so that multiplication by a negative number should flip the direction of $\vec{x}$.

The algebraic properties of vector addition are as follows.

Proposition 7 (Properties of Vector Addition)**.** Vector addition satisfies the following properties.

(i) (Associativity) For every $\vec{x}, \vec{y}, \vec{z} \in \mathbb{K}^n$,

$$\vec{x} + (\vec{y} + \vec{z}) = (\vec{x} + \vec{y}) + \vec{z}.$$

(ii) (Commutativity) For every $\vec{x}, \vec{y} \in \mathbb{K}^n$,

$$\vec{x} + \vec{y} = \vec{y} + \vec{x}.$$

(iii) (Additive Identity) There is a unique vector $\vec{0} \in \mathbb{K}^n$ such that for every $\vec{x} \in \mathbb{K}^n$,

$$\vec{x} + \vec{0} = \vec{x}.$$

(iv) (Additive Inverses) For every $\vec{x} \in \mathbb{K}^n$ there exists a unique vector $-\vec{x} \in \mathbb{K}^n$ such that

$$\vec{x} + (-\vec{x}) = \vec{0}.$$

Proof. We prove (ii) and (iii) here; parts (i) and (iv) are on your homework assignment.

For (ii), let $\vec{x}, \vec{y} \in \mathbb{K}^n$. Write

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

By the commutativity of addition in $\mathbb{K}$,

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} = \begin{bmatrix} y_1 + x_1 \\ \vdots \\ y_n + x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \vec{y} + \vec{x}.$$

For (iii), define $\vec{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$. Let $\vec{x} \in \mathbb{K}^n$, and write

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Then because 0 is the additive identity of $\mathbb{K}$,

$$\vec{x} + \vec{0} = \begin{bmatrix} x_1 + 0 \\ \vdots \\ x_n + 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \vec{x}.$$

For uniqueness, suppose that $\vec{0}' \in \mathbb{K}$ is an additive identity. Then $\vec{0}' = \vec{0}' + \vec{0} = \vec{0}$ because $\vec{0}'$ and $\vec{0}$ are additive identities. □

There is a similar list of properties for Scalar Multiplication.

Proposition 8 (Properties of Scalar Multiplication)**.** Scalar multiplication satisfies the following properties.

(i) (Associativity) For every $\vec{x} \in \mathbb{K}^n$ and $a, b \in \mathbb{K}$,

$$a(b\vec{x}) = (ab)\vec{x}.$$

(ii) (Distributivity over Scalar Addition) For every $\vec{x} \in \mathbb{K}^n$ and $a, b \in \mathbb{K}$,

$$(a + b)\vec{x} = a\vec{x} + b\vec{x}.$$

(iii) (Distributivity over Vector Addition) For every $\vec{x}, \vec{y} \in \mathbb{K}$ and $a \in \mathbb{K}$,

$$a(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}.$$

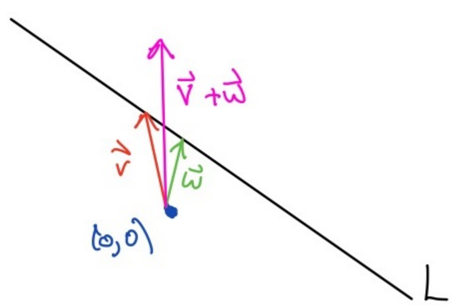
(iv) (Multiplicative Identity) For every $\vec{x} \in \mathbb{K}^n$,

$$1\vec{x} = \vec{x}.$$

Proof. The proofs of these result are all similar in flavor to the proofs of the Properties of Vector Addition, so we leave them as an exercise. □

The above properties of vector addition and scalar multiplication allow us to give algebraic proofs of geometric claims, as the following examples show.

Example 6. Fix $a, b, c \in \mathbb{R}$ with at least one of a, b nonzero, and consider the line L in $\mathbb{R}^2$ described by the equation $ax + by = c$. Let $\vec{v}, \vec{w} \in \mathbb{R}^2$ be two nonzero vectors whose endpoints lie on L. That is, the entries of $\vec{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ and $\vec{w} = \left[\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right]$ each satisfy the equation $ax + by = c$. Then the endpoint of $\vec{v} + \vec{w}$ lies on L if, and only if, $c = 0$ (i.e. exactly when L passes through the origin).

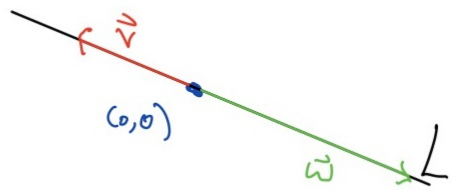


To prove this, note that $\vec{v} + \vec{w} = \left[\begin{smallmatrix} v_1+w_1 \\ v_2+w_2 \end{smallmatrix}\right]$, and

$$a(v_1 + w_1) + b(v_2 + w_2) = (av_1 + bv_2) + (aw_1 + bw_2) = c + c = 2c.$$

Therefore, if the endpoint of $\vec{v} + \vec{w}$ lies on L we must have $2c = c$, so that $c = 0$. On the other hand, if $c = 0$ then $2c = c$, so that the endpoint of $\vec{v} + \vec{w}$ lies on L.

Example 7. Fix $a, b \in \mathbb{R}$ with at least one of a, b nonzero, and consider the line L in $\mathbb{R}^2$ through the origin described by the equation $ax + by = 0$. Let $\vec{v}, \vec{w} \in \mathbb{R}^2$ be two nonzero vectors whose endpoints lie on L. That is, the entries of $\vec{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ and $\vec{w} = \left[\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right]$ each satisfy the equation $ax + by = 0$. Then each of $\vec{v}$ or $\vec{w}$ is a scalar multiple of the other.



To prove this, suppose that $c = 0$. Because one of a, b is nonzero, let us assume (without loss of generality) that $a \neq 0$. Because $av_1 + bv_2 = 0$, $v_1 = -\frac{b}{a}v_2$. If $v_2 = 0$ then $v_1 = 0$ as well, contradicting the assumption that $\vec{v} \neq \vec{0}$. Therefore $v_2 \neq 0$. Similarly, $w_1 = -\frac{b}{a}w_2$ and $w_2 \neq 0$. Because $w_2 \neq 0$, we can define $\lambda \overset{def}{=} \frac{v_2}{w_2}$. Then $v_2 = \lambda w_2$, and

$$v_1 = -\frac{b}{a}v_2 = -\frac{b}{a}\lambda w_2 = \lambda\Big(-\frac{b}{a}w_2\Big) = \lambda w_1,$$

so that $\vec{v} = \lambda\vec{w}$.

Properties of Vector Addition and Scalar Multiplication

Proposition (Properties of Vector Addition)**.** Vector addition satisfies the following properties.

(i) (Associativity) For every $\vec{x}, \vec{y}, \vec{z} \in \mathbb{K}^n$,

$$\vec{x} + (\vec{y} + \vec{z}) = (\vec{x} + \vec{y}) + \vec{z}.$$

(ii) (Commutativity) For every $\vec{x}, \vec{y} \in \mathbb{K}^n$,

$$\vec{x} + \vec{y} = \vec{y} + \vec{x}.$$

(iii) (Additive Identity) There is a unique vector $\vec{0} \in \mathbb{K}^n$ such that for every $\vec{x} \in \mathbb{K}^n$,

$$\vec{x} + \vec{0} = \vec{x}.$$

(iv) (Additive Inverses) For every $\vec{x} \in \mathbb{K}^n$ there exists a unique vector $-\vec{x} \in \mathbb{K}^n$ such that

$$\vec{x} + (-\vec{x}) = \vec{0}.$$

Proposition (Properties of Scalar Multiplication)**.** Scalar multiplication satisfies the following properties.

(i) (Associativity) For every $\vec{x} \in \mathbb{K}^n$ and $a, b \in \mathbb{K}$, $a(b\vec{x}) = (ab)\vec{x}$.

(ii) (Distributivity over Scalar Addition) For every $\vec{x} \in \mathbb{K}^n$ and $a, b \in \mathbb{K}$, $(a+b)\vec{x} = a\vec{x} + b\vec{x}$.

(iii) (Distributivity over Vector Addition) For every $\vec{x}, \vec{y} \in \mathbb{K}$ and $a \in \mathbb{K}$, $a(\vec{x}+\vec{y}) = a\vec{x} + a\vec{y}$.

(iv) (Multiplicative Identity) For every $\vec{x} \in \mathbb{K}^n$, $1\vec{x} = \vec{x}$.

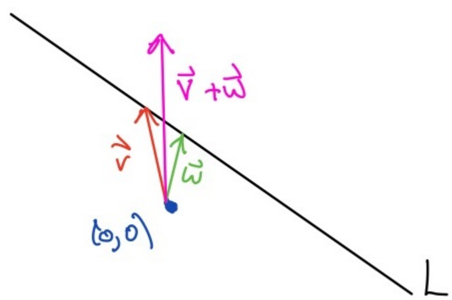
Lecture 5: Linear Combinations

Learning Objectives:

- Define the notion of linear combination of vectors.
- Express vectors and sets of vectors in terms of linear combinations.
- Determine when a list of vectors contains a redundant vector.

We start by discussing an example that was included in the notes from last time.

Example 8. Fix $a, b, c \in \mathbb{R}$ with at least one of a, b nonzero, and consider the line L in $\mathbb{R}^2$ described by the equation $ax + by = c$. Let $\vec{v}, \vec{w} \in \mathbb{R}^2$ be two nonzero vectors whose endpoints lie on L. That is, the entries of $\vec{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ and $\vec{w} = \left[\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right]$ each satisfy the equation $ax + by = c$. Then the endpoint of $\vec{v} + \vec{w}$ lies on L if, and only if, $c = 0$ (i.e. exactly when L passes through the origin).



To prove this, note that $\vec{v} + \vec{w} = \left[\begin{smallmatrix} v_1+w_1 \\ v_2+w_2 \end{smallmatrix}\right]$, and

$$a(v_1 + w_1) + b(v_2 + w_2) = (av_1 + bv_2) + (aw_1 + bw_2) = c + c = 2c.$$

Therefore, if the endpoint of $\vec{v} + \vec{w}$ lies on L we must have $2c = c$, so that $c = 0$. On the other hand, if $c = 0$ then $2c = c$, so that the endpoint of $\vec{v} + \vec{w}$ lies on L.

Linear Combinations

Vector addition and scalar multiplication allow us to use a small number of vectors in $\mathbb{K}^n$ to efficiently describe large sets of vectors. Indeed, the goal of understanding subsets of $\mathbb{K}^n$ as generated by sums and scalar multiples of a given set of vectors will dominate our efforts for the entire linear algebra portion of MATH 291. Let's give a precise name to the type of construction we have in mind.

Definition 10. Let $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$. A **linear combination** of $\vec{v}_1, \ldots, \vec{v}_m$ is a sum of the form

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m,$$

where $c_1, c_2, \ldots, c_m \in \mathbb{K}$ are scalars (sometimes called the **coefficients** of the linear combination).

Linear combinations are the foundation of most of the sophisticated ideas we will discuss this quarter, and the interesting questions in linear algebra often involve them. Let's see a few examples to help us understand what they represent and how they will be used.

Example 9. Fix $a, b, c \in \mathbb{R}$ with at least one of a, b nonzero, and consider the line L in $\mathbb{R}^2$ described by the equation $ax + by = c$. We will describe the set of vectors $\vec{v} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ whose endpoints lie on L.

Because one of a or b is nonzero, we'll address the case where $a \neq 0$ (the case where $b \neq 0$ is similar). Suppose that $\vec{v}$ is the position vector of a point on L. If we write $\vec{v} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$, then we have $ax + by = c$, or rather $x = \frac{c}{a} - \frac{b}{a}y$. In particular, we have

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{c}{a} - \frac{b}{a}y \\ y \end{bmatrix} = \begin{bmatrix} \frac{c}{a} \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{b}{a}y \\ y \end{bmatrix} = \begin{bmatrix} \frac{c}{a} \\ 0 \end{bmatrix} + y \begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix}.$$

Therefore, we have show that if $\vec{v}$ is the position vector of a point on L, then there exists $y \in \mathbb{R}$ such that

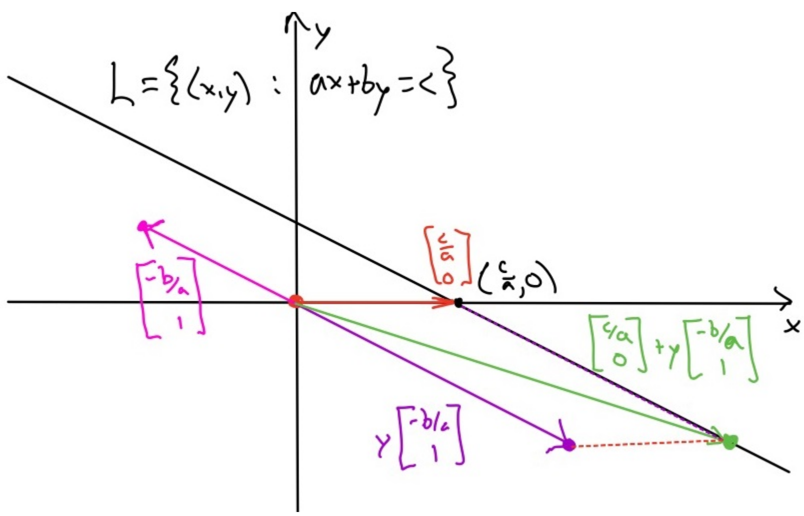
$$\vec{v} = \begin{bmatrix} \frac{c}{a} \\ 0 \end{bmatrix} + y \begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix}.$$

Moreover, if $\vec{v}$ has this form then the first entry of $\vec{v}$ is $\frac{c}{a} - \frac{b}{a}y$, so that $a\left(\frac{c}{a} - \frac{b}{a}y\right) + by = c - by + by = c$.

Therefore,

$$L = \left\{ \vec{v} \in \mathbb{R}^2 \ : \ \text{there exists } y \in \mathbb{R} \text{ with } \vec{v} = \begin{bmatrix} \frac{c}{a} \\ 0 \end{bmatrix} + y \begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix} \right\}.$$

In other words, we have written the position vector of each point on L as a linear combination of the two vectors $\left[\begin{smallmatrix} \frac{c}{a} \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} -\frac{b}{a} \\ 1 \end{smallmatrix}\right]$. The first of these is the position vectors of a point on L, and scalar multiples of the second vector are used to describe the other points on L in terms of their displacement from $(\frac{c}{a}, 0)$.

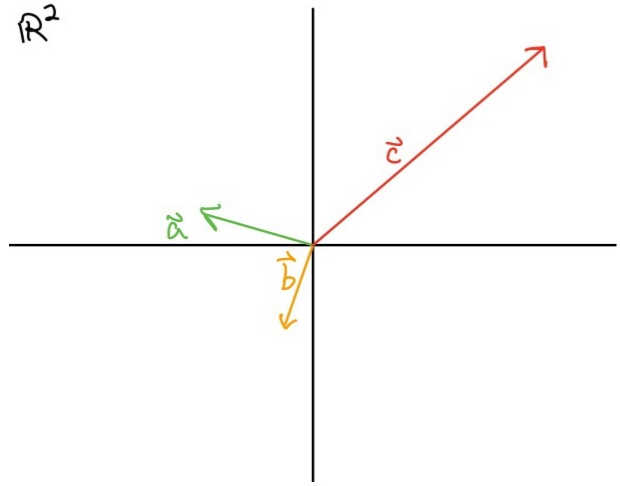


Remark 15. In the previous example, if $c = 0$ (i.e. if L passed through the origin), then we actually could represent L as

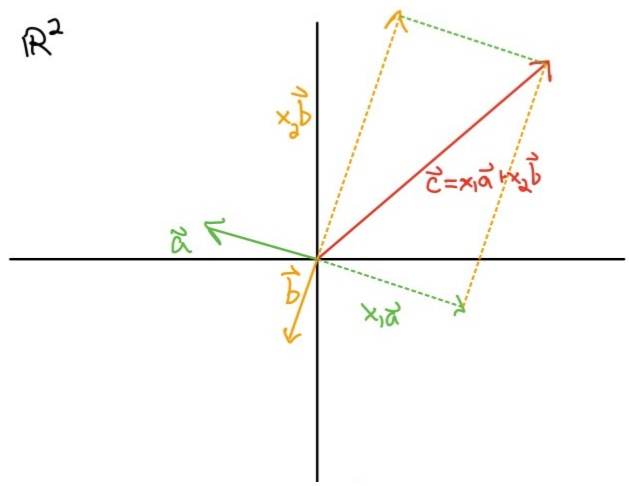
$$L = \left\{ \vec{v} \in \mathbb{R}^2 \ : \ \text{there exists } y \in \mathbb{R} \text{ with } \vec{v} = y \begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix} \right\}.$$

Remark 16. Note that the exact same arguments show that if $a, b, c \in \mathbb{C}$ with at least one of a, b nonzero, then the entries of $\vec{z} \in \mathbb{C}^2$ solve the equation $ax + by = c$ exactly when there exists $y \in \mathbb{C}$ with $\vec{z} = \left[\begin{smallmatrix} \frac{c}{a} \\ 0 \end{smallmatrix}\right] + y \left[\begin{smallmatrix} -\frac{b}{a} \\ 1 \end{smallmatrix}\right]$. Although the interpretation of this in terms of points and lines is no longer accessible, the algebraic result is still valid in the complex case!

Remark 17. Suppose that we are given $\vec{a}, \vec{b} \in \mathbb{R}^2$, and assume that neither of $\vec{a}$ or $\vec{b}$ is a scalar multiple of the other. For the vector $\vec{c}$ shown, is it possible to express $\vec{c}$ as a linear combination of $\vec{a}$, $\vec{b}$?



We expect the answer here to be "yes," as it appears that for suitable scalars $x_1, x_2 \in \mathbb{R}$, the vector $\vec{c}$ becomes the opposite corner of a parallelogram with one corner at the origin and with adjacent sides $x_1\vec{a}$ and $x_2\vec{b}$. In particular, we expect that $\vec{c} = x_1\vec{a} + x_2\vec{b}$.



This example was purely for intuition, but let us note that if $\vec{c} = \left[\begin{smallmatrix} c_1 \\ c_2 \end{smallmatrix}\right]$ and $\vec{a} = \left[\begin{smallmatrix} a_1 \\ a_2 \end{smallmatrix}\right]$ and $\vec{b} = \left[\begin{smallmatrix} b_1 \\ b_2 \end{smallmatrix}\right]$, then there exist scalars $x_1, x_2 \in \mathbb{R}$ such that $x_1\vec{a} + x_2\vec{b} = \vec{c}$ exactly if

$$x_1 \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + x_2 \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad \text{or rather} \quad \begin{bmatrix} a_1x_1 + b_1x_2 \\ a_2x_1 + b_2x_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

Because two vectors are equal exactly when their entries are equal, the scalars x_1, x_2 should be solutions of the following system of two equations:

$$\begin{aligned} a_1x_1 + b_1x_2 &= c_1 \\ a_2x_1 + b_2x_2 &= c_2. \end{aligned}$$

To summarize, $\vec{c}$ is a linear combination of $\vec{a}$,$\vec{b}$ when we can find scalars x_1, x_2 that solve the system of equations above. This is our first connection between vector arithmetic and systems of linear equations. We will develop general techniques for solving systems of linear equations soon, so this is all we will say about this example for now.

Remark 18. Note that it is impossible to write $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ as a linear combination of vectors of the form $\begin{bmatrix} a \\ b \\ 0 \end{bmatrix}, \begin{bmatrix} c \\ d \\ 0 \end{bmatrix}$. To see why, note that if there were scalars $m, n \in \mathbb{K}$ with

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = m \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} + n \begin{bmatrix} c \\ d \\ 0 \end{bmatrix} = \begin{bmatrix} am + cn \\ bm + dn \\ 0m + 0n \end{bmatrix} = \begin{bmatrix} am + cn \\ bm + dn \\ 0 \end{bmatrix},$$

then it would be true that $0 = 1$. Because $0 \neq 1$, there are no such scalars $m, n \in \mathbb{K}$.

These examples illustrate one of the fundamental questions of this quarter.

Fundamental Question: Representing a Vector as a Linear Combination of Other Vectors

Given vectors $\vec{v}_1, \ldots, \vec{v}_m, \vec{b} \in \mathbb{K}^n$, how can we determine whether $\vec{b}$ can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_m$?

We will spend a long time studying the linear combinations of a given a list of vectors $\vec{v}_1, \ldots, \vec{v}_m$. One might predict that including more vectors in this list will allow us to represent more vectors as linear combinations of the list, but that isn't always the case.

Proposition 9. Let $\vec{v}_1, \ldots, \vec{v}_m, \vec{u} \in \mathbb{K}^n$. Suppose that $\vec{b} \in \mathbb{K}^n$ is a linear combination of $\vec{v}_1, \ldots, \vec{v}_m, \vec{u}$. If $\vec{u}$ is a linear combination of $\vec{v}_1, \ldots, \vec{v}_m$, then $\vec{b}$ is also a linear combination of $\vec{v}_1, \ldots, \vec{v}_m$.

Proof. Suppose that $\vec{u}$ is a linear combination of $\vec{v}_1, \ldots, \vec{v}_m$. Choose scalars $a_1, \ldots, a_m \in \mathbb{K}$ such that $\vec{u} = a_1\vec{v}_1 + \cdots + a_m\vec{v}_m$. Because $\vec{b}$ is a linear combination of $\vec{v}_1, \ldots, \vec{v}_m, \vec{u}$, there are scalars $c_1, \ldots, c_m, d \in \mathbb{K}$ such that $\vec{b} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m + d\vec{u}$. Then we have

$$\begin{aligned} \vec{b} &= c_1\vec{v}_1 + \cdots + c_m\vec{v}_m + d\vec{u} \\ &= c_1\vec{v}_1 + \cdots + c_m\vec{v}_k + d(a_1\vec{v}_1 + \cdots + a_m\vec{v}_m) \\ &= (c_1 + da_1)\vec{v}_1 + \cdots + (c_m + da_m)\vec{v}_m. \end{aligned}$$

□

One interpretation of the previous proposition is that because we can represent exactly the same vectors whether we use linear combinations of $\vec{v}_1, \ldots, \vec{v}_m, \vec{u}$ and or $\vec{v}_1, \ldots, \vec{v}_m$, including the vector $\vec{u}$ would not allow the list $\vec{v}_1, \ldots, \vec{v}_m$ to represent any additional vectors as linear combinations beyond those it could already represent. For this reason, we think of $\vec{u}$ as redundant[4]. If none of the vectors in a list can be written as a linear combination of the other vectors in a list, then we think of each vector in the list as "adding something meaningful" to the list that can't be accounted for by the other vectors. This leads us to another fundamental question.

Fundamental Question

Given $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$, how can we verify that none of $\vec{v}_1, \ldots, \vec{v}_m$ can be written as a linear combination of the others?

The answer to this question involves a full investigation of *linear independence*, which we introduce next time.

[4]Your book gives a rigorous definition for this term, but we will not do so.

Lecture 6: Span and Linear Independence

Learning Objectives:

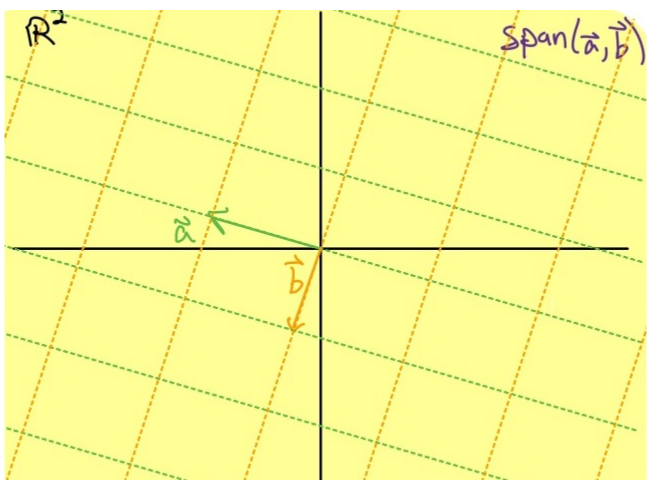
- Define the span of a set of vectors, and develop intuition for simple cases.
- Characterize geometrically meaningful sets as the span of a set of vectors.
- Define the notion of linear independence, and prove its basic properties.
- Develop intuition for linear independence.

Last time we introduced the term *linear combination* to describe vectors that can be obtained as sums of scalar multiples of other vectors. Because we wish to characterize which vectors can be obtained in this way from a given (small) list of vectors, we make the following definition.

Definition 11. Let $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$. The **span** of $\vec{v}_1, \ldots, \vec{v}_m$, denoted $\text{span}(\vec{v}_1, \ldots, \vec{v}_m)$, by

$$\text{span}(\vec{v}_1, \ldots, \vec{v}_m) \stackrel{def}{=} \{c_1\vec{v}_1 + \cdots + c_m\vec{v}_m \ : \ c_1, \ldots, c_m \in \mathbb{K}\}.$$

Example 10. Suppose that we are given $\vec{a}, \vec{b} \in \mathbb{R}^2$, and assume that neither of $\vec{a}$ or $\vec{b}$ is a scalar multiple of the other. Last time we argued that we expect that *every* vector $\vec{c} \in \mathbb{R}^2$ can be written as a linear combination of $\vec{a}, \vec{b}$. In other words, we expect that $\text{span}(\vec{a}, \vec{b}) = \mathbb{R}^2$. We have not proved this yet, but we will be able to do so soon.



The picture here illustrates why we use the word *span* here, as we picture the linear combinations of $\vec{a}, \vec{b}$ as "filling out" or "extending over" the entire plane $\mathbb{R}^2$.

Remark 19. The span of a set of vectors always includes the zero vector[5]. To see this, note that if $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^m$, then $\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_m$. Therefore $\vec{0} \in \text{span}(\vec{v}_1, \ldots, \vec{v}_m)$.

[5] As a matter of convention, we will define the span of the empty list of vectors as $\text{span}() \stackrel{def}{=} \{\vec{0}\}$. This is only notation, but it is notation that will make certain arguments and statements of results easier.

Example 11. Last time we saw that the vectors $\vec{v} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ whose entries solve the equation $ax + by = 0$ (where a is nonzero) all have the form $\vec{v} = \lambda \begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix}$ for some scalar $\lambda \in \mathbb{R}$. In other words, we saw that the set of vectors whose entries solve $ax + by = 0$ is exactly $\operatorname{span}\left(\begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix} \right)$.

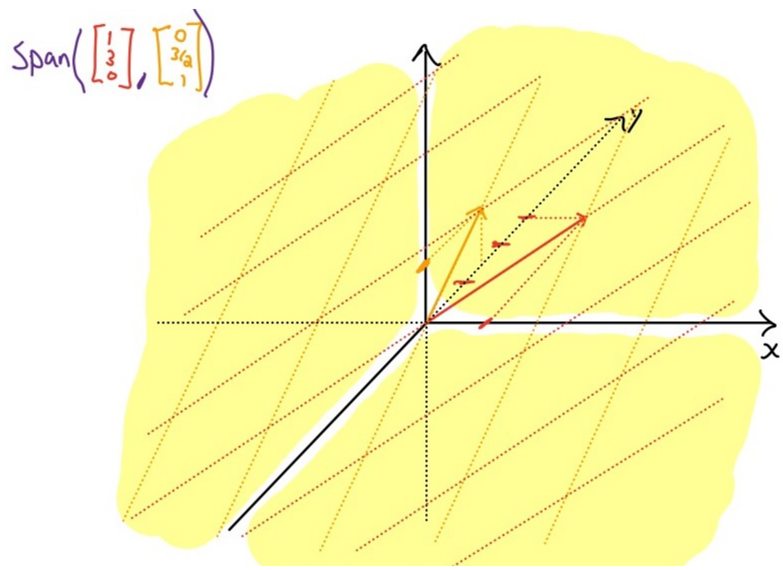
Example 12. Consider the equation $6x - 2y + 3z = 0$. The vectors $\vec{v} = \left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$ whose entries solve this equation all have the form (using the fact that, say, $y = 3x + \frac{3}{2}z$)

$$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ 3x + \frac{3}{2}z \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ \frac{3}{2} \\ 1 \end{bmatrix}$$

for any choice of scalars $x, z \in \mathbb{R}$. In other words, the vectors whose entries solve the equation $6x - 2y + 3z = 0$ are exactly

$$\operatorname{span}\left(\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \frac{3}{2} \\ 1 \end{bmatrix} \right).$$

Because neither of these vectors is a scalar multiple of the other, we expect that their span (which is also the set of solutions of $6x - 2y + 3z = 0$) is a flat surface in $\mathbb{R}^3$ that is parallel to both vectors:



In the picture, the surface passes through the origin at a slant and the positive x-axis, the positive z-axis, and the negative y-axis are all in front of the plane. We will call such a surface a **plane**. On your homework, you will explore why the collection of points (x, y, z) in $\mathbb{R}^3$ that solve an equation of the form $ax + by + cz = d$ (where at least one of a, b, c is nonzero) deserves to be called a plane.

Linear Independence

Last time we also discussed the problem of determining, for given vectors $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$, whether it is impossible to express one of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ as a linear combination of the others. To facilitate study, we want a characterization of this property that avoids the need to inspect every vector individually. Here is a step in the right direction.

Theorem 3. Let $m \geq 2$ and $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$. Then one of $\vec{v}_1, \ldots, \vec{v}_m$ can be written as a linear combination of the others if, and only if, there are scalars $c_1, \ldots, c_m \in \mathbb{K}$, at least one of which is nonzero, such that

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}.$$

Proof. ($\Rightarrow$) Suppose that for some j the vector $\vec{v}_j$ can be written as a linear combination of

$$\vec{v}_1, \ldots, \vec{v}_{j-1}, \vec{v}_{j+1}, \ldots, \vec{v}_m.$$

Then there exist scalars $c_1, \ldots, c_{j-1}, c_{j+1}, \ldots, c_m$ such that

$$\vec{v}_j = c_1\vec{v}_1 + \cdots + c_{j-1}\vec{v}_{j-1} + c_{j+1}\vec{v}_{j+1} + \cdots + c_m\vec{v}_m,$$

or rather (by subtracting $\vec{v}_j$ from both sides) we have

$$\vec{0} = c_1\vec{v}_1 + \cdots + (-1)\vec{v}_j + \cdots + c_m\vec{v}_m.$$

Because $-1 \neq 0$, we have therefore written $\vec{0}$ as a linear combination of $\vec{v}_1, \ldots, \vec{v}_m$ with at least one nonzero coefficient.

($\Leftarrow$) Suppose that there are scalars $c_1, \ldots, c_m \in \mathbb{K}$, at least one of which is nonzero, such that

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}.$$

Choose j such that $c_j \neq 0$. Then we can solve for $\vec{v}_j$ by writing

$$\vec{v}_j = -\frac{c_1}{c_j}\vec{v}_1 - \cdots - \frac{c_{j-1}}{c_j}\vec{v}_{j-1} - \frac{c_{j+1}}{c_j}\vec{v}_{j+1} - \cdots - \frac{c_m}{c_j}\vec{v}_m.$$

Therefore $\vec{v}_j$ can be written as a linear combination of the other vectors, and the proof is complete. □

Theorem 3 gives a condition under which a list of vectors contains at least one vector that can be written as a linear combination of the other vectors in the list. By restating this theorem using contraposition, we obtain a first answer to our question of how to tell whether *no* vector in a given list can be written as a linear combination of the others.

Theorem 4. Let $m \geq 2$ and $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$. Then none of $\vec{v}_1, \ldots, \vec{v}_m$ can be written as a linear combination of the others if, and only if, for every $c_1, \ldots, c_m \in \mathbb{K}$, if

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}$$

then $c_1 = \cdots = c_n = 0$.

This motivates one of the most important definitions in the course.

Definition 12. Let $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$. We call the set $\vec{v}_1, \ldots, \vec{v}_m$ **linearly independent**[6] if for every $c_1, \ldots, c_m \in \mathbb{K}$, if

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}$$

then $c_1 = c_2 = \cdots = c_m = 0$.

A set of vectors that is not linearly independent is called **linearly dependent**.

In other words, theorems 3 and 4 can be restated as follows.

Theorem 5 (Linear Independence and Linear Dependence)**.** Let $m \geq 2$ and $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$.

(a) $\vec{v}_1, \ldots, \vec{v}_m$ is a linearly independent set if, and only if, none of $\vec{v}_1, \ldots, \vec{v}_m$ can be written as a linear combination of the others.

(b) $\vec{v}_1, \ldots, \vec{v}_m$ is a linearly dependent set if, and only if, at least one of $\vec{v}_1, \ldots, \vec{v}_m$ can be written as a linear combination of the others.

Note that the definition of linear independence and dependence does not require that our set of vectors consists of two or more vectors.

Example 13. Let $\vec{v} \in \mathbb{K}^n$. Then $\vec{v}$ is a linearly independent set if, and only if, $\vec{v} \neq \vec{0}$.

Proof. ($\Rightarrow$) We proceed by contraposition. Suppose that $\vec{v} = \vec{0}$. Then $1\vec{v} = 1\vec{0} = \vec{0}$, so that $\vec{v}$ is a linearly dependent list of vectors.

($\Rightarrow$) Suppose that $\vec{v} \neq \vec{0}$. Let $c \in \mathbb{K}$ and suppose that $c\vec{v} = \vec{0}$. As you showed on your quiz, either $c = 0$ or $\vec{v} = \vec{0}$. Because $\vec{v} \neq \vec{0}$, it must be that $c = 0$. Therefore $\vec{v}$ is a linearly independent set. □

Remark 20. As a matter of convention, we will say that the empty set $\{\}$ of vectors is linearly independent. This may seem a little strange, but there are two good reasons for taking this convention. First, it is convenient for some arguments in proofs. Second, the empty set does *vacuously* satisfy the definition of linear independence, in the sense that because there are *no* ways to even take linear combinations of vectors in the empty set, it is technically true that if we had a linear combination of vectors in the empty set that equaled $\vec{0}$, then the coefficients in the linear combination would need to each be 0. Because there is no linear combination of vectors in the empty set to test this statement, the statement is valid.

For intuition, you should think of each vector $\vec{v}_1, \ldots, \vec{v}_m$ in a linearly independent set as adding a "dimension" to the span that wasn't present before. There are many results that one can state and prove that get at this, but here are a couple.

Proposition 10. Let $\vec{v}_1, \ldots, \vec{v}_m, \vec{u} \in \mathbb{K}^n$. Assume that $\vec{v}_1, \ldots, \vec{v}_m$ is linearly independent set, and that $\vec{u}$ cannot be written as a linear combination of $\vec{v}_1, \ldots, \vec{v}_m$. Then $\vec{v}_1, \ldots, \vec{v}_m, \vec{u}$ is a linearly independent set.

Proof. Suppose that $c_1, \ldots, c_m, c \in \mathbb{K}$ satisfy

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m + c\vec{u} = \vec{0}.$$

If $c \neq 0$, then we can subtract $c\vec{u}$ from both sides and divide by $-c$ to obtain

$$\vec{u} = -\frac{c_1}{c}\vec{v}_1 - \cdots - \frac{c_m}{c}\vec{v}_m,$$

[6] The definitions of linear independence and linear dependence here are the ones that are commonly used through mathematics. Your book does not define linear independence and linear dependence in this way, but instead chooses to define it in terms of the equivalent conditions listed in theorems 3 and 4. There are other instances where we will adopt standard definitions when the author of the textbook adopts non-standard definitions.

contradicting the assumption that $\vec{u}$ cannot be written as a linear combination of $\vec{v}_1, \ldots, \vec{v}_m$. Therefore $c = 0$, and we have

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}.$$

Because $\vec{v}_1, \ldots, \vec{v}_m$ is a linearly independent set, $c_1 = \cdots = c_m = 0$.

This shows that $\vec{v}_1, \ldots, \vec{v}_m, \vec{u}$ is a linearly independent set, and the proof is complete. □

Besides the geometric intuition described above (i.e. that $\vec{v}_1, \ldots, \vec{v}_m$ is a linearly independent set exactly when none of the vectors in the set can be written as a linear combination of the other vectors), the notion of linear independence can also be understood as a statement that $\vec{v}_1, \ldots, \vec{v}_m$ represent vectors in their span *efficiently* in the following sense.

Theorem 6. Let $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{K}^n$. Then the following are equivalent[7]:

(a) $\vec{v}_1, \ldots, \vec{v}_m$ is a linearly independent set.

(b) There is a unique choice of scalars $c_1, \ldots, c_m \in \mathbb{K}$ such that $\vec{0} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m$.

(c) For every $\vec{b} \in \operatorname{span}(\vec{v}_1, \ldots, \vec{v}_m)$, there is a unique choice of scalars $c_1, \ldots, c_m \in \mathbb{K}$ such that $\vec{b} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m$.

Proof. You will prove this result on your homework. □

[7]The phrase "The following are equivalent" (sometimes abbreviated TFAE) is similar to "if, and only if," but can be used to indicate that any number of statements are equivalent to each other. To prove a result of this type, one needs to show that every statements in the list implies every other statement in the list. For example, to say that three statements P, Q, R are equivalent, one would ultimately need to prove six implications: $P \Rightarrow Q$, $P \Rightarrow R$, $Q \Rightarrow P$, $Q \Rightarrow R$, $R \Rightarrow P$, and $R \Rightarrow Q$. Usually the proof can be done more efficiently, though. For example, in this theorem we prove that $P \Rightarrow R$, that $R \Rightarrow Q$, and that $Q \Rightarrow P$. These three implications are enough, because the other implications immediately follow from these. For example, the implication $Q \Rightarrow R$ follows from $Q \Rightarrow P$ and $P \Rightarrow R$.

Lecture 7: Linear Systems

Learning Objectives:

- Define linear systems and investigate elementary operations that preserve their solution sets.

Given $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{K}^m$ and $\vec{b} \in \mathbb{K}^m$, the question of whether (and how) $\vec{b}$ can be written as a linear combination of $\vec{v}_1, \ldots, \vec{v}_n$ boils down to determining whether there are scalars $x_1, \ldots, x_n \in \mathbb{K}$ such that

$$x_1\vec{v}_1 + \cdots + x_n\vec{v}_n = \vec{b}.$$

If we write $\vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ and, for each $k = 1, \ldots, n$, $\vec{v}_k = \begin{bmatrix} a_{1,k} \\ \vdots \\ a_{m,k} \end{bmatrix}$, then we can condense this equation by computing the left-hand-side as

$$\begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Because two vectors are equal exactly when their entries are equal, this gives us a system of m equations in n unknowns $x_1, \ldots, x_n$:

$$\begin{cases} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots & + & a_{2,n}x_n & = & b_2 \\ & & & & & & & \vdots & \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{cases}$$

We will therefore make the following definition.

Definition 13. Let $m, n \in \mathbb{N}$. An $m \times n$ (read "m by n") **linear system** is a system of m equations in n unknowns $x_1, \ldots, x_n$ of the form

$$☺ \begin{cases} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{cases}$$

Here the scalars $a_{j,k} \in \mathbb{K}$ are called the **coefficients** of ☺. A vector $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{K}^n$ whose entries satisfy ☺ is called a **solution** of ☺. The collection

$$\text{Sol}(☺) \stackrel{def}{=} \{\vec{x} \in \mathbb{K}^n \;:\; \vec{x} \text{ solves } ☺\}$$

is called the **solution set** of ☺.

Remark 21. The questions that motivated our discussion of span and linear independence over the last few days can be easily captured in terms of solutions to linear systems. In particular, the following results are immediate:

(a) $\vec{b} \in \text{span}(\vec{v}_1, \ldots, \vec{v}_n)$ if, and only if, $\text{Sol}(☺) \neq \emptyset$ (i.e. ☺ has at least one solution[8]).

(b) $\vec{v}_1, \ldots, \vec{v}_n$ is a linearly independent set if, and only if, $\text{Sol}(☺) = \{\vec{0}\}$ when $\vec{b} = \vec{0}$ (i.e. $\vec{x} = \vec{0}$ is the only solution of ☺ when $\vec{b} = \vec{0}$.).

Elementary System Operations

In high school you may have solved simple systems of linear equations (say 2×2 systems) by assuming that a solution exists and then attempting to findthe solutions using a substitution technique. This is insufficient for a general study of linear systems, as it is messy and also depends too strongly on the specifics of the system under consideration. Here, our approach will be to use a small collection of rules, called *elementary operations*, to transform one $m \times n$ system into another (hopefully simpler) $m \times n$ system with the same solution set. Through repeated applications of these elementary operations, we will be able to simplify our original system to the point where we can easily determine its solution set.

Remark 22. Before stating the elementary system operations we have in mind, we illustrate each with an example. Consider the system

$$☺ \begin{cases} 3x \qquad\quad - 2z = 2 \\ \qquad 6y - z = -2 \\ -x + 2y + \frac{1}{3}z = 1 \end{cases}$$

Each of the following systems ☹ (obtained as described) will result in a system that has the same solution set as ☺:

(i) ☹ is obtained by multiplying one of the equations in ☺ by a nonzero scalar (here we multiplied the second equation of ☺ by -3):

$$☹ \begin{cases} 3x \qquad\quad - 2z = 2 \\ \quad - 18y + 3z = 6 \\ -x + 2y + \frac{1}{3}z = 1 \end{cases}$$

(ii) ☹ is obtained by adding a scalar multiple of one of the equations in ☺ to another. Here we have added 2·(the second equation) to the first equation of ☺:

$$☹ \begin{cases} 3x + 12y - 4z = -2 \\ \qquad 6y - z = -2 \\ -x + 2y + \frac{1}{3}z = 1 \end{cases}$$

(iii) ☹ is obtained by swapping two of the equations in ☺. Here we have swapped the first and third equations of ☺:

$$☹ \begin{cases} -x + 2y + \frac{1}{3}z = 1 \\ \qquad 6y - z = -2 \\ 3x \qquad\quad - 2z = 2 \end{cases}$$

[8]The notation $\emptyset$ denotes the **empty set** $\{\}$, which is the set containing no elements.

The following result formally states (in their full generality) the elementary operations and establishes that they do not affect the solution set of a linear system.

Theorem 7 (Elementary System Operations)**.** Let $m, n \in \mathbb{N}$ and consider the linear system

$$\smiley \left\{ \begin{array}{ccccccc} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{array} \right.$$

where $a_{j,k}, b_j \in \mathbb{K}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$.

(i) (**Multiplying An Equation By A Nonzero Scalar**) Let $j \in \{1, \ldots, m\}$ and let $c \in \mathbb{K}$ with $c \neq 0$. Consider the system

$$\frownie \left\{ \begin{array}{ccccccc} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ ca_{j,1}x_1 & + & \cdots & + & ca_{j,n}x_n & = & cb_j \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{array} \right.$$

obtained by multiplying the j-th equation of $\smiley$ by c. Then $\mathrm{Sol}(\smiley) = \mathrm{Sol}(\frownie)$.

(ii) (**Adding A Scalar Multiple Of One Equation To Another**) Suppose $j, p \in \{1, \ldots, m\}$ with $j \neq p$, and let $c \in \mathbb{K}$. Consider the system

$$\frownie \left\{ \begin{array}{ccccccc} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ (a_{j,1} + ca_{p,1})x_1 & + & \cdots & + & (a_{j,n} + ca_{p,n})x_n & = & b_j + cb_p \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{array} \right.$$

obtained by adding the p-th equation (multiplied by c) to the j-th equation of $\smiley$. Then $\mathrm{Sol}(\smiley) = \mathrm{Sol}(\frownie)$.

(iii) (**Swapping Two Equations**) Suppose $j, p \in \{1, \ldots, m\}$ with $j \neq p$, and suppose (without loss of generality) that $j < p$. Consider the system

$$\frownie \left\{ \begin{array}{ccccccc} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ a_{p,1}x_1 & + & \cdots & + & a_{p,n}x_n & = & b_p \\ & & & & & \vdots & \\ a_{j,1}x_1 & + & \cdots & + & a_{j,n}x_n & = & b_j \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{array} \right.$$

obtained by swapping the j-th and p-th equations of $\smiley$. Then $\mathrm{Sol}(\smiley) = \mathrm{Sol}(\frownie)$.

Proof. We will prove part (i): multiplying an equation by a nonzero scalar does not change the solution set of the system. You will prove (ii) and (iii) on your homework. The heart of the argument is showing that, in each case, Sol(☺) = Sol(☹). Because we are showing that two *sets* are equal, we must show that each is a subset of the other. To do this, we show that if $\vec{x} \in$ Sol(☺) then $\vec{x} \in$ Sol(☹), and that if $\vec{x} \in$ Sol(☹) then $\vec{x} \in$ Sol(☺).

We proceed with the proof of (i). Let $j \in \{1, \dots, m\}$ and let $c \in \mathbb{K}$ with $c \neq 0$. Let ☹ be the system obtained by multiplying the j-th equation in ☺ by c.

Suppose[9] that $\vec{x} \in$ Sol(☺). Note that because $\vec{x}$ solves every equation in ☺, $\vec{x}$ solves every equation in ☹ except possibly the j-th equation, and therefore we need only verify that $\vec{x}$ solves the j-th equation of ☹. But because $\vec{x}$ satisfies the j-th equation of ☺, we have

$$ca_{j,1}x_1 + \cdots + ca_{j,n}x_n = c(a_{j,1}x_1 + \cdots + a_{j,n}x_n) = cb_j.$$

Therefore $\vec{x} \in$ Sol(☹).

Now suppose that $\vec{x} \in$ Sol(☹). As in the first case, we are done when we verify that $\vec{x}$ satisfies the j-th equation of ☺. To this end, note that because $c \neq 0$ we have

$$a_{j,1}x_1 + \cdots + a_{j,n}x_n = \frac{1}{c}(ca_{j,1}x_1 + \cdots + ca_{j,n}x_n) = \frac{1}{c}cb_j = b_j.$$

Therefore $\vec{x} \in$ Sol(☺), and (i) is proved. □

Remark 23. Note that in the proof of (i) above, we could have handled the proof that if $\vec{x} \in$ Sol(☹) then $\vec{x} \in$ Sol(☺) by simply noting that since ☺ is obtained from ☹ by multiplying the j-th equation of ☹ by the nonzero scalar $\frac{1}{c}$, if $\vec{x} \in$ Sol(☹) then the first half of the argument (interchanging the roles of ☺ and ☹, and replacing c with $\frac{1}{c}$) implies that $\vec{x} \in$ Sol(☺).

Now that we have elementary operations for manipulating linear systems, we must determine how to use these operations to characterize the solution set of a given system. Our plan of attack is as follows:

(i) Develop an algorithm for simplifying a linear system into a particularly "nice" form.

(ii) Describe how to read off the solution set of a linear system from this "nice" form.

(iii) Establish some rather surprising (and useful) results about the relationship between a linear system and its solution set.

This program will be made easier by introducing *matrices*, a generalization of vectors, as a way to condense the information contained in a system of linear equations. Our understanding of matrices will evolve quite a bit over the course of the year, but we first encounter them as a useful notational device. Don't be fooled, though: they are one of the central objects of study in linear algebra.

[9] Here, as a point of style, we use the word "Suppose" instead of "Let" because it may be the case that Sol(☺) is empty. The phrase "Suppose $\vec{x} \in$ Sol(☺)" reflects this, but the word "let" could (incorrectly) be interpreted as suggesting that the set under consideration is nonempty.

Elementary System Operations

Theorem (Elementary System Operations)**.** Let $m, n \in \mathbb{N}$ and consider the linear system

$$☺\begin{cases} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{cases}$$

where $a_{j,k}, b_j \in \mathbb{K}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$.

(i) (**Multiplying An Equation By A Nonzero Scalar**) Let $j \in \{1, \ldots, m\}$ and let $c \in \mathbb{K}$ with $c \neq 0$. Consider the system

$$☹\begin{cases} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ ca_{j,1}x_1 & + & \cdots & + & ca_{j,n}x_n & = & cb_j \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{cases}$$

obtained by multiplying the j-th equation of ☺ by c. Then $\text{Sol}(☺) = \text{Sol}(☹)$.

(ii) (**Adding A Scalar Multiple Of One Equation To Another**) Suppose $j, p \in \{1, \ldots, m\}$ with $j \neq p$, and let $c \in \mathbb{K}$. Consider the system

$$☹\begin{cases} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ (a_{j,1} + ca_{p,1})x_1 & + & \cdots & + & (a_{j,n} + ca_{p,n})x_n & = & b_j + cb_p \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{cases}$$

obtained by adding the p-th equation (multiplied by c) to the j-th equation of ☺. Then $\text{Sol}(☺) = \text{Sol}(☹)$.

(iii) (**Swapping Two Equations**) Suppose $j, p \in \{1, \ldots, m\}$ with $j \neq p$, and suppose (without loss of generality) that $j < p$. Consider the system

$$☹\begin{cases} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ a_{p,1}x_1 & + & \cdots & + & a_{p,n}x_n & = & b_p \\ & & & & & \vdots & \\ a_{j,1}x_1 & + & \cdots & + & a_{j,n}x_n & = & b_j \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{cases}$$

obtained by swapping the j-th and p-th equations of ☺. Then $\text{Sol}(☺) = \text{Sol}(☹)$.

Lecture 8: Matrices and Elementary Row Operations

Learning Objectives:

- Represent a linear system using an augmented matrix.
- Interpret elementary operations on systems as elementary row operations on matrices.
- Determine when a matrix is in reduced row-echelon form.
- Apply Gaussian elimination to transform a matrix into reduced row-echelon form.
- Determine the solution sets of a given linear system.

Systems of equations can be written more easily (and therefore studied more easily) if we represent them in terms of **matrices**.

Definition 14. Let $m, n \in \mathbb{N}$. An $m \times n$ **matrix**[10] A with entries in $\mathbb{K}$ is an array

$$A \overset{def}{=} \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

of scalars $a_{j,k} \in \mathbb{K}$, called the **entries** of A. We will sometimes write $A = [a_{j,k}]$ for short. The collection of all $m \times n$ matrices with entries in $\mathbb{K}$ is denoted $M_{m\times n}(\mathbb{K})$. Note that $a_{j,k}$ denotes the entry of A in row j and column k.

Example 14. Consider the system

$$☺ \begin{cases} 3x & & - & 2z & = & 6 \\ & 6y & - & z & = & 6 \\ -x & + \ 2y & + & \frac{1}{3}z & = & 0 \end{cases}$$

There are two interesting matrices associated to this system. The first is the **coefficient matrix** of ☺, which is the matrix whose entries are exactly the coefficients of the variables on the left-hand side. In particular, the coefficient matrix of ☺ is

$$\begin{bmatrix} 3 & 0 & -2 \\ 0 & 6 & -1 \\ -1 & 2 & \frac{1}{3} \end{bmatrix}.$$

In this case, the coefficient matrix of ☺ is 3×3 because ☺ was a 3×3 system of equations. Note that we have entered 0 into the matrix when the coefficient of the variable representing that spot is 0.

[10]Here $m \times n$ is read "m by n", and indicates that the matrix has m rows and n columns.

We will study coefficient matrices of systems later on in the quarter, but for now they are insufficient for our purposes because they omit crucial information about the system, namely the scalars on the right-hand sides of the equations in the system. If we wish to capture everything we will need to work with the **augmented matrix** of ☺, which in this case is the 3×4 matrix that we obtain by "augmenting" the coefficient matrix of ☺ with another column containing the scalars on the right-hand sides of the equations. The augmented matrix for ☺ is

$$\left[\begin{array}{ccc|c} 3 & 0 & -2 & 6 \\ 0 & 6 & -1 & 6 \\ -1 & 2 & \frac{1}{3} & 0 \end{array}\right].$$

Here we have inserted a vertical line between the first 3 columns of the matrix (which represents the coefficients of ☺) and the last column (which represents the "right-hand sides" of the equations in ☺.

When we represent a linear system in terms of its augmented matrix, each elementary system operations that we used to transform the linear system can be interpreted as an elementary *row* operation that transforms the augmented matrix. The elementary row operations are as follows. We will skimp on the notation here for ease of reading, as these were stated very carefully for systems of equations.

Definition 15. Let A be a matrix. The **elementary row operations** on A are

(i) Multiplying a row of A by a nonzero scalar.

(ii) Adding a scalar multiple of one row of A to another.

(iii) Swapping two rows of A.

Remark 24. Note that the elementary row operations (i), (ii), (iii), when performed on the augmented matrix of a linear system, correspond exactly to the associated elementary system operations.

Let's see an example.

Example 15. For this example, let's assume that we're just working with real numbers. Consider the linear system

$$☺\begin{cases} 3x & & - & 2z & = & 6 \\ & 6y & - & z & = & 6 \\ -x & + \ 2y & + & \frac{1}{3}z & = & 0 \end{cases}$$

I claim that ☺ can be transformed into the system

$$☹\begin{cases} x & & - & \frac{2}{3}z & = & 2 \\ & y & - & \frac{1}{6}z & = & 1 \\ & & & 0 & = & 0 \end{cases}$$

via elementary operations. To do this, we transform the augmented matrix of ☺ using elementary row operations (which is equivalent to transforming ☺ itself with elementary operations):

$$\left[\begin{array}{ccc|c} 3 & 0 & -2 & 6 \\ 0 & 6 & -1 & 6 \\ -1 & 2 & \frac{1}{3} & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & -\frac{2}{3} & 2 \\ 0 & 6 & -1 & 6 \\ -1 & 2 & \frac{1}{3} & 0 \end{array}\right] \quad (\text{Multiply row 1 by } \frac{1}{3})$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & -\frac{2}{3} & 2 \\ 0 & 6 & -1 & 6 \\ 0 & 2 & -\frac{1}{3} & 2 \end{array}\right] \quad (\text{Add 1(row 1) to row 3})$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & -\frac{2}{3} & 2 \\ 0 & 1 & -\frac{1}{6} & 1 \\ 0 & 2 & -\frac{1}{3} & 2 \end{array}\right] \quad (\text{Multiply row 2 by } \frac{1}{6})$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & -\frac{2}{3} & 2 \\ 0 & 1 & -\frac{1}{6} & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad (\text{Add } (-2)(\text{row 2}) \text{ to row 3}).$$

This last matrix is exactly the augmented matrix of ☹, as claimed.

The form of ☹ is particularly nice for reading off the solutions of ☺, as we can solve the first equation for x in terms of z as $x = 2 + \frac{2}{3}z$, and solve the second equation for y in terms of z as $y = 1 + \frac{1}{6}z$. There is nothing that forces us to choose any particular value for z, and once we pick z then x and y are determined. Because ☺ and ☹ have the same solutions, and $\vec{x} = \left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$ solves ☹ exactly when

$$\vec{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 + \frac{2}{3}z \\ 1 + \frac{1}{6}z \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} \frac{2}{3} \\ \frac{1}{6} \\ 1 \end{bmatrix},$$

We conclude that

$$\text{Sol}(☺) = \text{Sol}(☹) = \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} \frac{2}{3} \\ \frac{1}{6} \\ 1 \end{bmatrix} \;:\; s \in \mathbb{R} \right\}.$$

According to one of your homework problems this week, Sol(☺) is a line in $\mathbb{R}^3$! For another geometric interpretation of this result, note that the three equations in ☺ describe planes in $\mathbb{R}^3$, and therefore every solution of ☺ is (the position vector of) a point that lies on all three of these planes. Therefore we have shown that the intersection of the three planes whose equations make up ☺ is exactly a line. Neat!

To generalize the previous example, we need to answer two question:

(1) What is it about the system ☹ that made it possible to quickly read off the solutions of ☹?

(2) How can we (in a systematic way) transform a linear system into a system as simple to analyze at ☹ was?

The answer to (1) can be stated in terms of the augmented matrix of a system. Recall that the augmented matrix of ☹ was:

$$\left[\begin{array}{ccc|c} 1 & 0 & -\frac{2}{3} & 2 \\ 0 & 1 & -\frac{1}{6} & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

System ☹ is easy to solve because its augmented matrix is in **reduced row-echelon form**.

Definition 16. Let $B \in M_{n\times m}(\mathbb{K})$. Say B is in **reduced row-echelon form** if each of the following conditions are satisfied:

(i) The leading[11] nonzero entry (if there is one) in each row of B is 1.

(ii) If a column of B contains the leading 1 of some row, then all other entries in that column are 0.

(iii) If a row contains a leading 1, then every row above that row has a leading 1 that is further to the left.

Each leading 1 is called the **pivot** of its row.

Example 16. The augmented matrix for ☹

$$\left[\begin{array}{ccc|c} 1 & 0 & -\frac{2}{3} & 2 \\ 0 & 1 & -\frac{1}{6} & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

is in reduced row-echelon form. Note that only the first two rows have nonzero entries, and the leading nonzero entry in each of these rows is 1 (so that (i) is satisfied). Moreover, these leading entries are located in the first and second columns of the matrix, and all other entries in these columns are 0 (so that (ii) is satisfied). Finally, note that the first row (which has a leading 1) has no rows above it (so satisfies (iii), and that the leading 1 in the second row is to the right of the leading 1 in the first row (so satisfies (iii)), and that the third row does not have a leading 1 (and therefore satisfies (iii) in a rather uninteresting way).

The word "reduced row-echelon form" refers not only to the fact that the matrix has been reduced (simplified) via elementary row operations, and also that the rows are arranged so that the leading entries form "levels" (or *echelons*).

The answer to question (2) (i.e. a process for transforming a matrix into reduced row-echelon form with elementary row operations) is actually provided by an algorithm called *Gaussian Elimination.* This algorithm is used to prove part of the following theorem.

Theorem 8. Let $A \in M_{m\times n}(\mathbb{K})$. We can use elementary row operations to transform A into a matrix in reduced row-echelon form, called the **reduced row-echelon form** of A and denote $\operatorname{rref}(A)$.

Remark 25. The word 'the' in the theorem suggests that, given A, there is only one way to transform A into reduced row-echelon form. This is true, and when we prove (a more precise statement of) the theorem in a couple days, we will establish that there is indeed only one possibility for $\operatorname{rref}(A)$.

One process of **Gaussian Elimination**, which transforms A into a matrix in reduced row-echelon form, is quite mechanical. Given an $m \times n$ matrix A, proceed as follows. Let $j = 1$.

1. If every entry of Row j is 0, then proceed to Step 4. Otherwise, locate the leading nonzero entry c in Row j.

[11] Here 'leading' means 'first, starting from the left'.

2. Multiply Row j by $\frac{1}{c}$ so that the leading entry of Row j is 1.
3. Eliminate all other nonzero entries that lie in the same column as the leading entry of Row j.
4. If $j = m$ (i.e. if we were looking at the bottom row), then proceed to Step 5. Otherwise, replace j with $j+1$ and return to Step 1 (i.e. repeat with above steps with the next row).
5. Swap rows until all rows with leading entries are above all rows with all zeros, and such that each leading entry is either below and to the right or above and to the left of all other leading entries.

Example 17. Go back and review the computation we performed to transform the augmented matrix for the system ☺ into the augmented matrix for the system ☹. We used Gaussian elimination to perform this computation.

Example 18. Is $\left[\begin{smallmatrix} 3 \\ 16 \\ 5 \\ 3 \end{smallmatrix}\right] \in \operatorname{span}\Big(\left[\begin{smallmatrix} 0 \\ 2 \\ 1 \\ 3 \end{smallmatrix}\right], \left[\begin{smallmatrix} -1 \\ -4 \\ -1 \\ -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{smallmatrix}\right] \Big)$?

Note that the answer here is "yes" exactly if there exist $x, y, z \in \mathbb{K}$ such that

$$\begin{bmatrix} 3 \\ 16 \\ 5 \\ 3 \end{bmatrix} = x \begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix} + y \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix} + z \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0x - y + \frac{1}{2}z \\ 2x - 4y + 3z \\ x - y + z \\ 3x - y + z \end{bmatrix}.$$

Means that we are trying to determine whether there is a solution of the system

$$☺ \begin{cases} & - & y & + & \frac{1}{2}z & = & 3 \\ 2x & - & 4y & + & 3z & = & 16 \\ x & - & y & + & z & = & 5 \\ 3x & - & y & + & z & = & 3 \end{cases}$$

We use elementary operations to transform this system into one whose augmented matrix is in reduced row-echelon form. Note that Gaussian elimination is *one* way to perform this computation, but making some changes early on in the process can result in a simpler computation. The augmented matrix of the system we are given is

$$\left[\begin{array}{ccc|c} 0 & -1 & \frac{1}{2} & 3 \\ 2 & -4 & 3 & 16 \\ 1 & -1 & 1 & 5 \\ 3 & -1 & 1 & 3 \end{array}\right].$$

We start by multiplying Row 1 by 2 to remove the fractions, and then we swap the first and third rows to have a row where the leading entry is in the first column, and then proceed with Gaussian elimination as usual:

$$\left[\begin{array}{ccc|c} 0 & -1 & \frac{1}{2} & 3 \\ 2 & -4 & 3 & 16 \\ 1 & -1 & 1 & 5 \\ 3 & -1 & 1 & 3 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 5 \\ 2 & -4 & 3 & 16 \\ 0 & -2 & 1 & 6 \\ 3 & -1 & 1 & 3 \end{array}\right] \quad \text{(Multiply row 1 by 2, then swap row 1 and row 3)}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 5 \\ 0 & -2 & 1 & 6 \\ 0 & -2 & 1 & 6 \\ 0 & 2 & -2 & -12 \end{array}\right] \quad \text{(Add -2(row 1) to row 2, and -3(row 1) to row 4)}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & 2 \\ 0 & 1 & -\frac{1}{2} & -3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -6 \end{array}\right] \quad \begin{array}{l}\text{(Add (-1/2)(row 2) to row 1, -1(row 2) to row 3,} \\ \text{and 1(row 2) to row 4. Then multiply row 2 by (-1/2))}\end{array}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 6 \end{array}\right] \quad \begin{array}{l}\text{(Multiply row 4 by -1, then add (-1/2)(row 4) to row 1} \\ \text{and (1/2)(row 4) to row 2.)}\end{array}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \text{(Swap row 3 and row 4.).}$$

Therefore ☺ has the same solution set as the system

$$\begin{cases} x & & & = & -1 \\ & y & & = & 0 \\ & & z & = & 6 \\ & & 0z & = & 0 \end{cases}$$

and therefore

$$\text{Sol}(☺) = \left\{ \begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix} \right\}.$$

In terms of the question that was asked we can therefore conclude that

$$\begin{bmatrix} 3 \\ 16 \\ 5 \\ 3 \end{bmatrix} = -1 \begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix} + 0 \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix} + 6 \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix},$$

so that $\begin{bmatrix} 3 \\ 16 \\ 5 \\ 3 \end{bmatrix} \in \text{span}\left(\begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \right)$.

Example 19. Is $\begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix}$ a linearly independent set?

Note that the answer here is "yes" exactly if whenever we have $x, y, z \in \mathbb{K}$ such that

$$\begin{bmatrix} 0x - y + \frac{1}{2}z \\ 2x - 4y + 3z \\ x - y + z \\ 3x - y + z \end{bmatrix} = x \begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix} + y \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix} + z \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

then it must be that $x = y = z = 0$. Therefore we need to verify that the solution set of the system

$$☺ \begin{cases} \quad\; - \; y + \frac{1}{2}z = 3 \\ 2x - 4y + 3z = 16 \\ x - y + z = 5 \\ 3x - y + z = 3 \end{cases}$$

is exactly $\left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$.

We will solve this system using exactly the same steps as above.

$$\left[\begin{array}{ccc|c} 0 & -1 & \frac{1}{2} & 0 \\ 2 & -4 & 3 & 0 \\ 1 & -1 & 1 & 0 \\ 3 & -1 & 1 & 0 \end{array}\right].$$

We start by multiplying Row 1 by 2 to remove the fractions, and then we swap the first and third rows to have a row where the leading entry is in the first column, and then proceed with Gaussian elimination as usual:

$$\left[\begin{array}{ccc|c} 0 & -1 & \frac{1}{2} & 0 \\ 2 & -4 & 3 & 0 \\ 1 & -1 & 1 & 0 \\ 3 & -1 & 1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 2 & -4 & 3 & 0 \\ 0 & -2 & 1 & 0 \\ 3 & -1 & 1 & 0 \end{array}\right] \quad \text{(Multiply row 1 by 2, then swap row 1 and row 3)}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 2 & -2 & 0 \end{array}\right] \quad \text{(Add -2(row 1) to row 2, and -3(row 1) to row 4)}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{array}\right] \quad \begin{array}{l}\text{(Add (-1/2)(row 2) to row 1, -1(row 2) to row 3,} \\ \text{and 1(row 2) to row 4. Then multiply row 2 by (-1/2))}\end{array}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \quad \begin{array}{l}\text{(Multiply row 4 by -1, then add (-1/2)(row 4) to row 1} \\ \text{and (1/2)(row 4) to row 2.)}\end{array}.$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \text{(Swap row 3 and row 4.).}$$

Therefore ☺ has the same solution set as the system

$$\begin{cases} x & & & = & 0 \\ & y & & = & 0 \\ & & z & = & 0 \\ & & 0z & = & 0 \end{cases}$$

and therefore

$$\mathrm{Sol}(☺) = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

In terms of the question that was asked we can therefore conclude that

$$\begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix}$$

is a linearly independent set.

Example 20. Is $\begin{bmatrix} 4 \\ 16 \\ 5 \\ 3 \end{bmatrix} \in \mathrm{span}\left(\begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \right)$?

Note that the only difference between this example and the last is that we have changed the vector we are considering. By the same logic, we are led to consider solutions of the system

$$☺ \begin{cases} & - & y & + & \frac{1}{2}z & = & 4 \\ 2x & - & 4y & + & 3z & = & 16 \\ x & - & y & + & z & = & 5 \\ 3x & - & y & + & z & = & 3 \end{cases}$$

We repeat out computations from the last example to see that

$$\left[\begin{array}{ccc|c} 0 & -1 & \frac{1}{2} & 4 \\ 2 & -4 & 3 & 16 \\ 1 & -1 & 1 & 5 \\ 3 & -1 & 1 & 3 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 5 \\ 2 & -4 & 3 & 16 \\ 0 & -2 & 1 & 8 \\ 3 & -1 & 1 & 3 \end{array}\right] \quad \text{(Multiply row 1 by 2, then swap row 1 and row 3)}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 5 \\ 0 & -2 & 1 & 6 \\ 0 & -2 & 1 & 8 \\ 0 & 2 & -2 & -12 \end{array}\right] \quad \text{(Add -2(row 1) to row 2, and -3(row 1) to row 4)}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & 2 \\ 0 & 1 & -\frac{1}{2} & -3 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & -1 & -6 \end{array}\right] \quad \begin{array}{l}\text{(Add (-1/2)(row 2) to row 1, -1(row 2) to row 3,} \\ \text{and 1(row 2) to row 4. Then multiply row 2 by (-1/2).)}\end{array}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 6 \end{array}\right] \quad \begin{array}{l}\text{(Multiply row 4 by -1, then add (-1/2)(row 4) to row 1} \\ \text{and (1/2)(row 4) to row 2.)}\end{array}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 2 \end{array}\right] \quad \text{(Swap row 3 and row 4.).}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \quad \begin{array}{l}\text{(Multiply row 4 by (1/2), then add (-6)(row 4) to row 3,} \\ \text{and add 1(row 4) to row 1.)}\end{array}$$

Therefore ☺ has the same solution set as the system

$$\begin{cases} x & & & = & 0 \\ & y & & = & 0 \\ & & z & = & 0 \\ & & 0z & = & 1 \end{cases}$$

But there are no vectors $\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$ whose entries satisfy the equation $0 = 0z = 1$, so that ☺ does not have any solutions. That is,

$$\mathrm{Sol}(☺) = \emptyset.$$

Note that we could have seen this immediately after the third step above, where we inferred that the solution set of ☺ was the same as that of a system that included the equation $0z = 2$.

To answer the original question, $\left[\begin{smallmatrix} 4 \\ 16 \\ 5 \\ 3 \end{smallmatrix}\right] \notin \mathrm{span}\left(\left[\begin{smallmatrix} 0 \\ 2 \\ 1 \\ 3 \end{smallmatrix}\right], \left[\begin{smallmatrix} -1 \\ -4 \\ -1 \\ -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{smallmatrix}\right] \right)$

Definition 17. Let ☺ be an $m \times n$ linear system. If ☺ has no solutions, then ☺ is called **inconsistent**.

Lecture 9: Reduced Row-Echelon Form

Learning Objectives:

- Use matrices to represent linear combinations.
- Determine when two matrices are row equivalent.
- Show that every matrix has a unique reduced row-echelon form.

We start by introducing a bit more notation involving matrices.

Definition 18. Let $A \in M_{m\times n}(\mathbb{K})$, $A = [a_{j,k}]$. Then we call

$$\vec{a}_1 \overset{def}{=} \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix}, \quad \vec{a}_2 \overset{def}{=} \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{bmatrix}, \quad \cdots, \quad \vec{a}_n \overset{def}{=} \begin{bmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{bmatrix}$$

the **columns** (or **column vectors**) of A, and write $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$.

Remark 26. Note that if $A \in M_{m\times n}(\mathbb{K})$, then the columns of A are vectors in $\mathbb{K}^m$.

Matrices can be used to represent linear combinations of vectors in the following sense.

Definition 19. Let $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{K}^m$, and consider $A \overset{def}{=} \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{m\times n}(\mathbb{K})$. For $\vec{x} \in \mathbb{K}^n$, define

$$A\vec{x} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \overset{def}{=} x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n.$$

Remark 27. Note that $A\vec{x}$ is only defined with the number of columns of A agrees with the number of entries of $\vec{x}$.

Remark 28. As with scalar multiplication, the order here is very important.

Example 21. Recall that a linear system

$$☺ \begin{cases} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ & & & & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{cases}$$

is equivalent to the matrix equation

$$x_1 \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{m,1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1,n} \\ \vdots \\ a_{m,n} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Using the notation above, we can condense this further as

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix},$$

or more concisely as $A\vec{x} = \vec{b}$, where A is the coefficient matrix of ☺. This last expression is called the **matrix form** of the system ☺.

Example 22. Is $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 10 \end{bmatrix}$ a linearly independent set?

Note that the answer here is "yes" exactly if the only vector $\vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} \in \mathbb{K}^4$ that solves

$$\begin{bmatrix} 1 & 4 & 2 & 7 \\ 2 & 5 & 1 & 8 \\ 3 & 6 & 0 & 10 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + c_3 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + c_4 \begin{bmatrix} 7 \\ 8 \\ 10 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

is $\vec{c} = \vec{0}$. We will solve this system by row-reducing its augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & 4 & 2 & 7 & 0 \\ 2 & 5 & 1 & 8 & 0 \\ 3 & 6 & 0 & 10 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 4 & 2 & 7 & 0 \\ 0 & -3 & -3 & -6 & 0 \\ 0 & -6 & -6 & -11 & 0 \end{array}\right] \quad \text{(Add (-2)(row 1) to row 2 and add (-3)(row 1) to row 3)}$$

$$\longrightarrow \left[\begin{array}{cccc|c} 1 & 4 & 2 & 7 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & -6 & -6 & -11 & 0 \end{array}\right] \quad \text{(Multiply row 2 by (-1/3))}$$

$$\longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & -2 & -1 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] \quad \text{(Add (-4)(row 2) to row 1 and 6(row 2) to row 3)}$$

$$\longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & -2 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] \quad \text{(Add (1)(row 3) to row 1 and (-2)(row 3) to row 3).}$$

Therefore ☺ has the same solution set as the system

$$\begin{cases} c_1 \quad - 2c_3 = 0 \\ c_2 + c_3 = 0 \\ c_4 = 0 \end{cases}$$

It follows that $\vec{c}$ solves ☺ exactly when there is $s \in \mathbb{K}$ such that (setting $c_3 = s$ and solving for c_1 and c_2 in terms of c_3)

$$\vec{c} = \begin{bmatrix} 2s \\ -s \\ s \\ 0 \end{bmatrix} = s \begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \end{bmatrix}.$$

In other words,

$$\mathrm{Sol}(☺) = \mathrm{span}\left(\begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \end{bmatrix}\right).$$

In particular, note that this shows (among other things) that

$$2\begin{bmatrix}1\\2\\3\end{bmatrix} - \begin{bmatrix}4\\5\\6\end{bmatrix} + \begin{bmatrix}2\\1\\0\end{bmatrix} + 0\begin{bmatrix}7\\8\\10\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix},$$

so that $\left[\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right], \left[\begin{smallmatrix}4\\5\\6\end{smallmatrix}\right], \left[\begin{smallmatrix}2\\1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}7\\8\\10\end{smallmatrix}\right]$ is a *linearly dependent* set. The answer to the question is "no".

So far we have seen examples of systems where the solution set is infinite, where the system has a unique solution, and where the system is inconsistent (i.e. it has no solution). We will say a bit more about solution sets soon, but first we pause to prove our theorem about the existence and uniqueness of the reduced row-echelon form of a matrix.

Row Equivalence and Uniqueness of Reduced Row-Echelon Form

We start by making a definition.

Definition 20. Let $A, B \in M_{m\times n}(\mathbb{K})$. We say that A is **row equivalent** to B if A can be transformed into B via a finite sequence of elementary row operations.

Remark 29. Note that if A and B are matrices of different sizes, then A cannot be row equivalent to B because elementary row operations do not change the number of rows or columns in a matrix.

Row equivalence is not equality, but does *behave* like equality in the following very general sense.

Theorem 9. Let $A, B, C \in M_{m\times n}(\mathbb{K})$. Then the following hold.

- (Reflexivity) A is row equivalent to A.
- (Symmetry) If A is row equivalent to B, then B is row equivalent to A.
- (Transitivity) If A is row equivalent to B and if B is row equivalent to C, then A is row equivalent to C.

Proof. Because A can be transformed into itself by multiplying the first row of A by 1, A is row equivalent to A.

Suppose that A is row equivalent to B. Then A can be transformed into B via a finite sequence of elementary row operations. But each of the elementary row operations is invertible, in the sense that if $P, Q \in M_{m\times n}(\mathbb{K})$, then

(i) If we transform P into Q by multiplying row j (of P) by a nonzero scalar c, then we can transform Q into P by multiplying row j (of Q) by $\frac{1}{c}$.

(ii) If we transform P into Q by adding c(row j) to row k (of P), then we can transform Q into P by adding -c(row j) to row k (of Q).

(iii) If we transform P into Q by swapping row j and row k (of P), then we can transform Q into P by swapping row j and row k (of Q).

Therefore, performing on B the inverse of each elementary row operation used to transform A into B (in the reverse order, of course) transforms B into A. Therefore B is row equivalent to A.

Finally, suppose that A is row equivalent to B and that B is row equivalent to C. By first performing the elementary operations on A necessary to transform it into B, and then performing the elementary operations on B necessary to transform it into C, we will have performed a finite number of elementary operations on A to transform it into C. Therefore A is row equivalent to C. □

Remark 30. We can restate the relationship between linear systems and augmented matrices in terms of row equivalent as follows:

Theorem 10. Let ☺ and ☹ be $m \times n$ linear systems. Then ☺ can be transformed into ☹ by a finite sequence of elementary operations if, and only if, the augmented matrices of ☺ and ☹ are row equivalent.

Remark 31. Note that row equivalence is not limited to matrices that represent linear systems, as row operations and the definition of row equivalence apply even to matrices with a single column.

As mentioned last time, it is crucial for our ability to analyze linear systems that every matrix is row equivalent to a unique matrix in reduced row-echelon form. The existence of such a matrix follows from Gaussian elimination, but the uniqueness is far from obvious. The proof of uniqueness is typically more challenging than expected, but we will give a (relatively) simple proof due to W.H. Holtzmann. We start with a lemma[12] that says, roughly, that deleting a column from row equivalent matrices results in row equivalent matrices.

Lemma 1. Suppose $n \geq 2$, let $A, B \in M_{m\times n}(\mathbb{K})$, and fix $k = 1, \ldots, n$. Let $A', B' \in M_{m\times n-1}(\mathbb{K})$ be the matrices obtained from A and B by deleting the k-th column from each. If A is row equivalent to B, then A' is row equivalent to B'.

Proof. We proceed by induction on the number of elementary row operations necessary to transform A into B.

For the base case, we consider the three types of elementary row operations separately and show that if A is transformed into B via a single elementary row operation, then A' is transformed into B' by that same elementary row operation. This is quick (but notationally painful) to check by writing out the matrices A, B, A', and B', so we leave this step to the reader.

Now let $k \in \mathbb{N}$ and assume that A can be transformed into B via $k+1$ elementary operations, that D is the matrix obtained from A by applying the first k of these operations, and that $D' \in M_{m\times(n-1)}(\mathbb{K})$ be the matrix obtained from D by deleting the k-th column. For the induction hypothesis, we suppose that A' and D' are row equivalent. Because B is obtained from D by applying a single elementary row operation, the base case implies that D' is row equivalent to B'. By transitivity of row equivalence, A' is row equivalent to B'. By the Principle of Mathematical Induction, the result is proved. □

[12]A **lemma** is a result that is mostly important as a stepping stone to proving a more major result. We could include the proofs of lemmas in the proofs of major results, but we might choose to state and prove a lemma independently if either the lemma will be applicable to other results, if the lemma is interesting in its own right, or just to break up the proof of the larger result for ease of reading.

Lecture 10: More Reduced Row-Echelon Form

Learning Objectives:

- Show that every matrix has a unique reduced row-echelon form.

We can now prove (a more precise version of) the theorem[13] we stated last time.

Theorem 11. Let $A \in M_{m\times n}(\mathbb{K})$. Then there exists a unique $B \in M_{m\times n}(\mathbb{K})$ such that B is in reduced row-echelon form and A is row equivalent to B.

Proof. The existence of B is guaranteed by Gaussian elimination, so we need only show uniqueness.

Suppose $C \in M_{m\times n}(\mathbb{K})$ is a matrix in reduced row-echelon form that is row equivalent to A. By transitivity of row equivalence, C and B are row equivalent. We wish to show that $C = B$.

Suppose, to the contrary, that $C \neq B$. Let k denote the left-most column in which C and B differ. Because the first nonzero column of a matrix in reduced row-echelon form must be $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, and (because C and B differ in their k-th column) the k-th column of at least one of C and B is nonzero, there are two cases:

(a) All columns before the k-th column of C and B are zero, and the k-th column of one of these matrices is $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ and the k-th column of the other is $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

(b) $k > 1$ and among the first $k-1$ columns of C and B (which are equal) there is at least one pivot.

For case (a), note that none of the three elementary row operations transform a column with at least one nonzero entry into a column with all zero entries, and therefore the same is true for a finite sequence of elementary row operations. It follows that C and B are not row equivalent, a contradiction.

Now we consider case (b). Let p denote the number of columns before the k-th column of C and B that contain pivots. By assumption, $p \geq 1$. Apply the lemma repeatedly to delete each column of C and B beyond the k-th column, and also delete the columns before the k-th column which do not contain pivots. Let C' and B' denote the resulting matrices. Note that $C' \neq B'$ because C' and B' differ in their final columns. For example if

$$C = \begin{bmatrix} 1 & 2 & 0 & 3 & 5 \\ 0 & 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 0 & 7 & 9 \\ 0 & 0 & 1 & 8 & 9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

then

$$C' = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B' = \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & 8 \\ 0 & 0 & 0 \end{bmatrix}.$$

[13] This proof is based on a proof due to W.H. Holtzmann, but fills in a few logical gaps present in the original.

But then we have either

$$C' = \begin{bmatrix} 1 & 0 & \cdots & 0 & c_1 \\ 0 & 1 & \cdots & 0 & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & c_p \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \quad \text{or} \quad C' = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

and either

$$B' = \begin{bmatrix} 1 & 0 & \cdots & 0 & b_1 \\ 0 & 1 & \cdots & 0 & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & b_p \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \quad \text{or} \quad B' = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Note that because $k > 1$ and there was at least one column of B and C before the kth column containing a pivot, C' and B' have at least two columns. Interpret C' as the augmented matrix of a linear system ☺ and B' as the augmented matrix of a linear system ☹, both in the unknowns $x_1, \ldots, x_p$. Because the augmented matrices of ☺ and ☹ are row equivalent, ☺ and ☹ have the same solution sets.

Note that in the first case for C' we would have $\mathrm{Sol}(☺) = \left\{ \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \right\}$, and in the second case for C' the system ☺ would be inconsistent (since ☺ would contain the equation $0x_p = 1$). Similarly, the first case for B' we would have $\mathrm{Sol}(☹) = \left\{ \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix} \right\}$ and the second case for B' would yield that ☹ is inconsistent.

Because $\mathrm{Sol}(☺) = \mathrm{Sol}(☹)$, either both systems are inconsistent (and therefore $C' = B'$, contradicting the fact that $C' \neq B'$) or both sets are nonempty (in which case $\begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix}$, again contradicting the fact that $C' \neq B'$). In either case we have a contradiction, and the result is proved. □

Remark 32. Believe it or not, this was indeed a simpler proof of this result than the one that you will commonly find in textbooks.

In light of the last theorem, we make the following definition.

Definition 21. Let $A \in M_{m\times n}(\mathbb{K})$. The unique $m \times n$ matrix B in reduced row-echelon form that is row equivalent to A is called the **reduced row-echelon form of** A, and denoted $\mathrm{rref}(A)$.

Because the reduced row-echelon form of a matrix is unique, we can make the following definition (which will be useful going forward).

Definition 22. Let $A \in M_{m\times n}(\mathbb{K})$. The **rank** of A is the number of pivots in $\mathrm{rref}(A)$. We denote this number by $\mathrm{rank}(A)$.

Remark 33. Note that because each matrix has a unique reduced row-echelon form, there is no ambiguity in the definition of $\text{rank}(A)$ for $A \in M_{m\times n}(\mathbb{K})$.

Remark 34. Because $A \in M_{m\times n}(\mathbb{K})$ can have at most one pivot in each row, we have $0 \leq \text{rank}(A) \leq m$. But each column of A can also contain at most one pivot, so that $0 \leq \text{rank}(A) \leq n$ as well. Therefore $0 \leq \text{rank}(A) \leq \min(m, n)$.

Example 23. Note that for

$$A \overset{def}{=} \begin{bmatrix} 1 & 4 & 2 & 7 \\ 2 & 5 & 1 & 8 \\ 3 & 6 & 0 & 10 \end{bmatrix},$$

we have $\text{rank}(A) = 3$ because

$$\text{rref}(A) = \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

has three pivots.

Lecture 11: Solutions of Linear Systems

Learning Objectives:

- Investigate the relationship between span, linear independence, solutions of systems, reduced row-echelon form, and rank.

The idea that every matrix is row equivalent to exactly one matrix in reduced row-echelon form is extremely important, and allows us to determine sophisticated properties of linear systems by simply manipulating matrices. Let us summarize what we have learned about reduced row echelon form, rank, solutions of linear systems, span, and linear independence. Before we begin, we introduce some useful notation.

Definition 23. Let $n \in \mathbb{N}$ and k a natural number between 1 and n. Then $\vec{e}_k \in \mathbb{K}^n$ is defined to be the vector with 1 in the k-th entry, and 0 for all other entries.

Remark 35. Note that the notation for $\vec{e}_k$ only captures which entry is 1, and does *not* capture n. To avoid cluttering up the notation, in practice we will rely on context to tell us what is n. If there is any chance for ambiguity, then we will be more precise at the time.

The following lemma will also be useful.

Lemma 2. Let $A \in M_{m\times n}(\mathbb{K})$ and $\vec{b} \in \mathbb{K}^m$. Then there exists a unique $\vec{c} \in \mathbb{K}^m$ such that $\operatorname{rref}\left(\begin{bmatrix} A & \vec{b}\end{bmatrix}\right) = \begin{bmatrix}\operatorname{rref}(A) & \vec{c}\end{bmatrix}$. Moreover, $\vec{c} = \vec{0}$ if, and only if, $\vec{b} = \vec{0}$.

(Here $\begin{bmatrix} B & \vec{v}\end{bmatrix}$ denotes the $m \times (n+1)$ matrix whose first n columns are the columns of B, and whose final column is $\vec{v}$.)

Proof. We start with existence. By Gaussian elimination, there is a finite sequence of elementary row operations that transforms A into $\operatorname{rref}(A)$. Performing this same sequence of operations on $\begin{bmatrix} A & \vec{b}\end{bmatrix}$ transforms this matrix into $\begin{bmatrix}\operatorname{rref}(A) & \vec{c}'\end{bmatrix}$ for some $\vec{c}' \in \mathbb{K}^m$.

If the final column of $\begin{bmatrix}\operatorname{rref}(A) & \vec{c}'\end{bmatrix}$ does not contain the leading nonzero entry of a row, then $\begin{bmatrix}\operatorname{rref}(A) & \vec{c}'\end{bmatrix}$ is in reduced row-echelon form and, setting $\vec{c} = \vec{c}'$, we have $\operatorname{rref}\left(\begin{bmatrix} A & \vec{b}\end{bmatrix}\right) = \begin{bmatrix}\operatorname{rref}(A) & \vec{c}\end{bmatrix}$.

If the final column of $\begin{bmatrix}\operatorname{rref}(A) & \vec{c}'\end{bmatrix}$ does contain the leading nonzero entry of a row (say row j), then because $\operatorname{rref}(A)$ is in reduced row-echelon form, it must be that all pivots in $\operatorname{rref}(A)$ lie above row j. By using Gaussian elimination to turn this pivot in the final column into a 1, eliminate all other nonzero entries in the final column, and then swap rows in order to place the pivot in the row immediately below the last pivot of $\operatorname{rref}(A)$, we have reduced $\begin{bmatrix}\operatorname{rref}(A) & \vec{c}'\end{bmatrix}$ into reduced row-echelon form without changing any of the entries in the first n columns. Letting $\vec{c}$ denote the new final column (which will have a 1 for one entry, and 0 for all other entries), we conclude that $\operatorname{rref}\left(\begin{bmatrix} A & \vec{b}\end{bmatrix}\right) = \begin{bmatrix}\operatorname{rref}(A) & \vec{c}\end{bmatrix}$.

Uniqueness of $\vec{c}$ follows immediately by uniqueness of the reduced row-echelon form of a matrix. The claim that $\vec{c} = \vec{0}$ if, and only if, $\vec{b} = \vec{0}$ follows from our earlier observation that elementary row operation turn columns with all zero entries into columns with all zero entries, and that $\vec{c}$ and $\vec{b}$ are corresponding columns in row equivalent matrices. ∎

We start with span.

Theorem 12 (Span and Linear Systems)**.** Let $\vec{v}_1, \dots, \vec{v}_n, \vec{b} \in \mathbb{K}^m$, and let $A \in M_{m\times n}(\mathbb{K})$ be $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$.

The following are equivalent:

(a) $\vec{b} \in \operatorname{span}(\vec{v}_1, \dots, \vec{v}_n)$.

(b) The linear system $A\vec{x} = \vec{b}$ has at least one solution $\vec{x} \in \mathbb{K}^n$.

(c) $\operatorname{rref}\Big(\begin{bmatrix} A & \vec{b} \end{bmatrix}\Big)$ does not have a pivot in its final column.

(d) $\operatorname{rank}(A) = \operatorname{rank}\Big(\begin{bmatrix} A & \vec{b} \end{bmatrix}\Big)$.

Proof. We will prove that (a)⇒(b)⇒(c)⇒(a), and that (c)⇔(d).

((a)⇒(b)) Suppose (a) holds. Then there are scalars $x_1, \dots, x_n \in \mathbb{K}$ with

$$\vec{b} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Therefore $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is a solution of $A\vec{x} = \vec{b}$, and (b) holds.

((b)⇒(c)) We proceed by contraposition. Suppose $\operatorname{rref}\Big(\begin{bmatrix} A & \vec{b} \end{bmatrix}\Big)$ has a pivot in its final column. Then the linear system $A\vec{x} = \vec{b}$ can be transformed (by a finite sequence of elementary operations) into a system containing the equation $0 = 1$, and therefore is inconsistent.

((c)⇒(a)) Suppose (c) holds. By the lemma, there is a unique $\vec{c} \in \mathbb{K}^m$ with $\operatorname{rref}\Big(\begin{bmatrix} A & \vec{b} \end{bmatrix}\Big) = \begin{bmatrix} \operatorname{rref}(A) & \vec{c} \end{bmatrix}$. By (c), $\vec{c}$ does not contain the leading nonzero entry of any row. If $\vec{c} = \vec{0}$, then $\vec{b} = \vec{0}$ as well (by the lemma). Because $\vec{0} \in \operatorname{span}(\vec{v}_1, \dots, \vec{v}_n)$, (a) is proved. On the other hand, if $\vec{c} \neq \vec{0}$ (and the nonzero entries of $\vec{c}$ cannot be the leading nonzero entries in their rows) then $\operatorname{rref}(A)$ has at least one pivot. Let $\vec{v}_{k_1}, \dots, \vec{v}_{k_p}$ denote the columns of $\operatorname{rref}(A)$ that contain pivots. Note that for each $j = 1, \dots, p$, $\vec{v}_{k_j} = \vec{e}_j$, which is the vector in $\mathbb{K}^m$ with 1 in the j-th entry and 0 for all other entries. Because the final column of $\begin{bmatrix} \operatorname{rref}(A) & \vec{c} \end{bmatrix}$ does not contain a pivot, $\vec{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. Let $\vec{x} \in \mathbb{K}^n$ be the vector with entry c_j in the k_j-th row (for $j = 1, \dots, p$), and 0 for every other entry. Then $\operatorname{rref}(A)\vec{x} = c_1\vec{v}_{k_1} + \cdots + c_p\vec{v}_{k_p} = \vec{c}$. Therefore the linear system $\operatorname{rref}(A)\vec{x} = \vec{c}$ has a solution. Because $\begin{bmatrix} A & \vec{b} \end{bmatrix}$ and $\begin{bmatrix} \operatorname{rref}(A) & \vec{c} \end{bmatrix}$ are row equivalent, $\vec{x}$ also solves $A\vec{x} = \vec{b}$. This proves (a).

((c)⇒(d)) Suppose (c) holds. By the lemma, there is $\vec{c} \in \mathbb{K}^m$ such that $\operatorname{rref}(\begin{bmatrix} A & \vec{b} \end{bmatrix}) = \begin{bmatrix} \operatorname{rref}(A) & \vec{c} \end{bmatrix}$. By (c), the final column of this matrix does not have a pivot. Therefore all of the pivots of this matrix lie in the first n columns, and so are exactly the pivots of $\operatorname{rref}(A)$. Therefore $\operatorname{rank}(\begin{bmatrix} A & \vec{v} \end{bmatrix}) = \operatorname{rank}(A)$, so (d) holds.

((d)⇒(c)) Suppose (d) holds. By the lemma, there is $\vec{c} \in \mathbb{K}^m$ with $\operatorname{rref}(\begin{bmatrix} A & \vec{b} \end{bmatrix}) = \begin{bmatrix} \operatorname{rref}(A) & \vec{c} \end{bmatrix}$. By assumption, the number of pivots in the first n columns of this matrix ($\operatorname{rank}(A)$) is the same as the number of pivots in the entire matrix ($\operatorname{rank}\left(\begin{bmatrix} \operatorname{rref}(A) & \vec{c} \end{bmatrix} \right)$). Therefore the final column $\vec{c}$ does not contain the pivot of some row, and (c) holds. □

As a corollary, we have the following result.

Theorem 13 (Spanning Sets in $\mathbb{K}^m$)**.** Let $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{K}^m$, and let $A \in M_{m\times n}(\mathbb{K})$ be the matrix $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$.

The following are equivalent:

(a) $\operatorname{span}(\vec{v}_1, \ldots, \vec{v}_n) = \mathbb{K}^m$.

(b) For every $\vec{b} \in \mathbb{K}^m$, the system $A\vec{x} = \vec{b}$ has at least one solution $\vec{x} \in \mathbb{K}^n$.

(c) $\operatorname{rref}(A)$ has a pivot in every row.

(d) $\operatorname{rank}(A) = m$.

Proof. This is on your homework. □

Let's revisit some of our old examples in light of these theorems.

Example 24. Is $\begin{bmatrix} 3 \\ 16 \\ 5 \\ 3 \end{bmatrix} \in \operatorname{span}\left(\begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \right)$?

Because[14]

$$\operatorname{rref} \begin{bmatrix} 0 & -1 & \frac{1}{2} & 3 \\ 2 & -4 & 3 & 16 \\ 1 & -1 & 1 & 5 \\ 3 & -1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and this last matrix has no pivot in the final column, the Span and Linear Systems Theorem implies that

$$\begin{bmatrix} 3 \\ 16 \\ 5 \\ 3 \end{bmatrix} \in \operatorname{span}\left(\begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \right).$$

Example 25. Is $\begin{bmatrix} 4 \\ 16 \\ 5 \\ 3 \end{bmatrix} \in \operatorname{span}\left(\begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \right)$?

Because

$$\operatorname{rref} \begin{bmatrix} 0 & -1 & \frac{1}{2} & 3 \\ 2 & -4 & 3 & 16 \\ 1 & -1 & 1 & 5 \\ 3 & -1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

[14] Here and in the subsequent examples we are omitting the steps we used to compute the reduced row-echelon form because they are included in earlier examples, but you should make sure to show your work on quizzes, exams, and homework!

has a pivot in its final column, the Span and Linear Systems Theorem implies that

$$\begin{bmatrix} 4 \\ 16 \\ 5 \\ 3 \end{bmatrix} \notin \operatorname{span}\left(\begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \right).$$

We can also unpack the relationship between linear independence, linear systems, reduced row-echelon form, and rank.

Theorem 14 (Linear Independence and Linear Systems)**.** Let $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{K}^m$, and let $A \in M_{m\times n}(\mathbb{K})$ be $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$.

The following are equivalent:

(a) $\vec{v}_1, \ldots, \vec{v}_n$ is a linearly independent set.

(b) For every $\vec{b} \in \mathbb{K}^m$, the linear system $A\vec{x} = \vec{b}$ has at most one solution $\vec{x} \in \mathbb{K}^n$.

(c) $\operatorname{rref}(A)$ has a pivot in every column.

(d) $\operatorname{rank}(A) = n$.

Proof. We show that (a)⇒(b)⇒(c)⇒(a), and that (c)⇔(d).

((a)⇒(b)) Suppose (a) holds. Let $\vec{b} \in \mathbb{K}^m$. If $\vec{b} \notin \operatorname{span}(\vec{v}_1, \ldots, \vec{v}_n)$ then $A\vec{x} = \vec{b}$ has no solution. If $\vec{b} \in \operatorname{span}(\vec{v}_1, \ldots, \vec{v}_n)$, then Exercise 4 on Homework 2 implies that there is a unique choice of scalars $x_1, \ldots, x_n \in \mathbb{K}$ with

$$\vec{b} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n = A\vec{x}.$$

Therefore $A\vec{x} = \vec{b}$ has a unique solution, and (b) is proved.

((b)⇒(c)) We proceed by contraposition. Suppose that $\operatorname{rref}(A)$ has at least one column without a pivot. Note that $A\vec{0} = \vec{0}$. We produce $\vec{v} \neq \vec{0}$ such that $A\vec{v} = \vec{0}$. Let column k be the left-most column of A not containing a pivot. If $k = 1$, then the k-th column of $\operatorname{rref}(A)$ is $\vec{0}$ and $\vec{e}_1 \neq \vec{0}$ satisfies $\operatorname{rref}(A)\vec{e}_1 = \vec{0}$, so that $A\vec{e}_1 = \vec{0}$. If $k > 1$, then $\operatorname{rref}(A)$ has pivots in the first $k-1$ columns (and the first $k-1$ columns of $\operatorname{rref}(A)$ are $\vec{e}_1, \ldots, \vec{e}_{k-1}$), and the k-th column of $\operatorname{rref}(A)$ has the form $\vec{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_{k-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. Let $\vec{v} = \begin{bmatrix} -c_1 \\ -c_2 \\ \vdots \\ -c_{k-1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

Then $\vec{v} \neq \vec{0}$ (because the k-th entry of $\vec{v}$ is nonzero), and

$$\operatorname{rref}(A)\vec{v} = -c_1\vec{e}_1 - \cdots - c_{k-1}\vec{e}_{k-1} + 1\vec{c} = \vec{0}.$$

Because elementary operations preserve solution sets, $A\vec{v} = \vec{0}$ even though $\vec{v} \neq \vec{0}$, and (b) does not hold.

((c)⇒(a)) Suppose (c) holds. Suppose $x_1, \ldots, x_n \in \mathbb{K}$ satisfy $x_1\vec{v}_1 + \cdots + x_n\vec{v}_n = \vec{0}$. Then $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ solves the system $A\vec{x} = \vec{0}$. The lemma implies that $\operatorname{rref}\left(\begin{bmatrix} A & \vec{0} \end{bmatrix} \right) = \begin{bmatrix} \operatorname{rref}(A) & \vec{0} \end{bmatrix}$. Because $\operatorname{rref}(A)$ has a pivot in each column, this is the augmented matrix of the system $x_1 = 0, \ldots, x_n = 0$. Therefore (a) holds.

((c)⇒(d)) Suppose (c) holds. Because rref(A) has a pivot in each of its n columns, and there is at most one pivot in every column, rank$(A) = n$.

((d)⇒(c)) Suppose (d) holds. Then rref(A) has n pivots. Because there cannot be more than one pivot in each column, and because rref(A) has n columns, rref(A) has a pivot in every column. □

Example 26. Is $\begin{bmatrix}0\\2\\1\\3\end{bmatrix}, \begin{bmatrix}-1\\-4\\-1\\-1\end{bmatrix}, \begin{bmatrix}1/2\\3\\1\\1\end{bmatrix}$ a linearly independent set?

Because

$$\text{rref}\begin{bmatrix}0 & -1 & \frac{1}{2}\\ 2 & -4 & 3\\ 1 & -1 & 1\\ 3 & -1 & 1\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}$$

has a pivot in every column, the Linear Independence and Linear Systems Theorem implies that

$$\begin{bmatrix}0\\2\\1\\3\end{bmatrix}, \begin{bmatrix}-1\\-4\\-1\\-1\end{bmatrix}, \begin{bmatrix}1/2\\3\\1\\1\end{bmatrix}$$

a linearly independent set.

Example 27. Is $\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix}, \begin{bmatrix}7\\8\\10\end{bmatrix}$ a linearly independent set?

Because

$$\text{rref}\begin{bmatrix}1 & 4 & 2 & 7\\ 2 & 5 & 1 & 8\\ 3 & 6 & 0 & 10\end{bmatrix} = \begin{bmatrix}1 & 0 & -2 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

does not have a pivot in its third column, the Linear Independence and Linear Systems Theorem implies that

$$\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix}, \begin{bmatrix}7\\8\\10\end{bmatrix}$$

is a linearly dependent set.

Solutions of Linear Systems

Theorem (Span and Linear Systems)**.** Let $\vec{v}_1, \ldots, \vec{v}_n, \vec{b} \in \mathbb{K}^m$, and let $A \in M_{m\times n}(\mathbb{K})$ be $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$.

The following are equivalent:

(a) $\vec{b} \in \operatorname{span}(\vec{v}_1, \ldots, \vec{v}_n)$.

(b) The linear system $A\vec{x} = \vec{b}$ has at least one solution $\vec{x} \in \mathbb{K}^n$.

(c) $\operatorname{rref}\left(\begin{bmatrix} A & \vec{b} \end{bmatrix}\right)$ does not have a pivot in its final column.

(d) $\operatorname{rank}(A) = \operatorname{rank}\left(\begin{bmatrix} A & \vec{b} \end{bmatrix}\right)$.

Theorem (Linear Independence and Linear Systems)**.** Let $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{K}^m$, and let $A \in M_{m\times n}(\mathbb{K})$ be $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$.

The following are equivalent:

(a) $\vec{v}_1, \ldots, \vec{v}_n$ is a linearly independent set.

(b) For every $\vec{b} \in \mathbb{K}^m$, the linear system $A\vec{x} = \vec{b}$ has at most one solution $\vec{x} \in \mathbb{K}^n$.

(c) $\operatorname{rref}(A)$ has a pivot in every column.

(d) $\operatorname{rank}(A) = n$.

Lecture 12: More Solutions of Linear Systems

Learning Objectives:

- Summarize insights for solving linear systems.

Today we finished up our discussion of the results by last time by discussing the difficult step in the proof of the Linear Independence and Linear Systems Theorem. Last time we saw many theoretical results that link solution sets of equations to span, linear independence, reduced row-echelon form, and rank. You'll use those results on your homework, but before we move on from linear systems it is worth revisiting (and summarizing) the basic problem of determining the solution set of a linear system in light of our new results.

To determine the solution set of linear system $A\vec{x} = \vec{b}$:

1. Form the augmented matrix $\begin{bmatrix} A & \vec{b} \end{bmatrix}$.
2. Compute $\text{rref}\left(\begin{bmatrix} A & \vec{b} \end{bmatrix}\right) = \begin{bmatrix} \text{rref}(A) & \vec{c} \end{bmatrix}$. Because elementary operations preserve solution sets of systems of equations, the solution set of $\text{rref}(A)\vec{x} = \vec{c}$ is the same as the solution set of $A\vec{x} = \vec{b}$.
3. Write down the solution set of $\text{rref}(A)\vec{x} = \vec{c}$ either by

   3a. noting that the solution set is empty because the system is inconsistent (see Example 20, or Exercise 9 on Homework 3), or

   3b. solve for the leading variable in each equation of the system $\text{rref}(A)\vec{x} = \vec{c}$ in terms of the non-leading variables and entries of $\vec{c}$, and express each solution $\vec{x}$ in the form $\vec{x} = \vec{x}_p + c_1\vec{v}_1 + \cdots + c_q\vec{v}_q$ for appropriate vectors $\vec{x}_p, \vec{v}_1, \ldots, \vec{v}_q$ (see Examples 15, 18, 19, 22, Exercises 1 and 2 on Homework 2, Exercises 2 and 3 and 5 on Homework 3), or

   3c. (in light of Exercise 9 on Homework 2) if you have a particular solution $\vec{x}_p$ of $A\vec{x}_p = \vec{b}$, then follow the steps in (3b.) to write down all solutions $\vec{x}_h$ of the homogeneous system $\text{rref}(A)\vec{x}_h = \vec{0}$ in the form $c_1\vec{v}_1 + \cdots + c_q\vec{v}_k$ for appropriate vectors $\vec{v}_1, \ldots, \vec{v}_k$. Then each solution $\vec{x}$ of the original system has the form $\vec{x} = \vec{x}_p + c_1\vec{v}_1 + \cdots + c_q\vec{v}_k$.

Transition to Linear Transformations

So far in the course we have interpreted the equation $A\vec{x} = \vec{b}$ from two different viewpoints:

(i) (Geometrically) Expressing $\vec{b}$ as a linear combination of the column vectors of A.

(ii) (Algebraically) Encoding a system of m linear equations in m unknowns, where $\begin{bmatrix} A & \vec{b} \end{bmatrix}$ is the augmented matrix of the system.

We have studied each of these interpretations independently (e.g. our first problems on span and linear independence, and also our investigations of how to solve linear systems), and also used each interpretation to study the other (e.g. solution sets of linear systems can be represented in terms of linear combinations, and questions about linear combinations (e.g. involving linear independence, or whether a given vector is in the span of a given set of vectors) can be analyzed using linear systems).

There is a third interpretation that will be equally helpful to consider: viewing $\vec{x}$ as transformed by A to obtain $\vec{b}$, or rather as the operation that sends the vector $\vec{x}$ (in $\mathbb{K}^n$) to the vector $\vec{b} = A\vec{x}$ (in $\mathbb{K}^m$). This point of view involves *functions*.

Remark 36. From this point forward we will be doing many more computations where we compute a product $A\vec{x}$ of a matrix and a vector. Until now it was convenient to compute this product as a linear combination of the columns of A, but we will want a quicker way to compute these products going forward. As a computational device, note that we can think of the entry in the j-th row of $A\vec{x}$ as the sum of the products of the entries in the j-th row of A with the corresponding entries of $\vec{x}$:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{j,1} & a_{j,2} & \cdots & a_{j,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n \\ \vdots \\ a_{j,1}x_1 + a_{j,2}x_2 + \cdots + a_{j,n}x_n \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n \end{bmatrix}.$$

As practice, show that

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 4 & 3 & 2 \\ -3 & -2 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ -3 \end{bmatrix}.$$

Lecture 13: Linear Transformations

Learning Objectives:

- Define what it means for a transformation to be linear.
- Characterize linear transformations as exactly the matrix transformations.

Definition 24. Let A, B be sets. A **function** from A to B is a rule f that assigns to each element x of Aa unique element $f(x)$ of B. We typically denote that f is a function from A to B by writing $f : A \to B$.
The set A is called the **domain** of f, the set B is called the **codomain** of f. Each $x \in A$ is called an **input** of f. The element $f(x) \in B$ is called the **image of** x under f, or sometime the **output** of f associated to the input x.

Remark 37. We'll need more function terminology going forward, but we will introduce this terminology as-needed.

Example 28. The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) \overset{def}{=} x^2 + 1$ sends an input $x \in \mathbb{R}$ to $f(x) = x^2 + 1$. Note that because -1 is not an output of f (i.e. there is no real number x so that $-1 = f(x) = x^2 + 1$).

Example 29. Let $A \in M_{m\times n}(\mathbb{K})$. Then A induces a function $T : \mathbb{K}^n \to \mathbb{K}^m$, where $T(\vec{x}) = A\vec{x}$ for each $\vec{x} \in \mathbb{K}^n$. Such a function T (that is given by multiplying the input by a matrix) is called a **matrix transformation**.

Example 30. The function $I : \mathbb{K}^n \to \mathbb{K}^n$ given by $I(\vec{x}) = \vec{x}$ is called the **identity** function.

The identity function is a matrix transformation. To see why, note that for each $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{K}^n$,

$$I(\vec{x}) = \vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n = \begin{bmatrix} \vec{e}_1 & \cdots & \vec{e}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \vec{x}.$$

The $n \times n$ matrix $I_n \overset{def}{=} \begin{bmatrix} \vec{e}_1 & \cdots & \vec{e}_n \end{bmatrix}$ is called the $n \times n$ **identity matrix**.

Although we will be interested in all sorts of functions this year (especially once we get to multivariable calculus), our earlier investigations should make it clear that functions of the type identified in the previous example will play a central role in the course. The most important properties of these functions boil down to an algebraic fact that we haven't yet needed: matrix multiplication distributes over vector addition and scalar multiplication in the following sense, as you proved in your discussion.

Proposition 11. Let $A \in M_{m\times n}(\mathbb{K})$. Then for every $\vec{x}, \vec{y} \in \mathbb{K}^n$ and every $c \in \mathbb{K}$,

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} \quad \text{and} \quad A(c\vec{x}) = c(A\vec{x}).$$

Proof. Let $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{K}^m$ be the columns of A. Then by the distributive and commutative properties of vector addition and scalar multiplication we have

$$\begin{aligned} A(\vec{x}+\vec{y}) &= \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1+y_1 \\ \vdots \\ x_n+y_n \end{bmatrix} \\ &= (x_1+y_1)\vec{a}_1 + \cdots + (x_n+y_n)\vec{a}_n \\ &= (x_1\vec{a}_1 + \cdots + x_n\vec{a}_n) + (y_1\vec{a}_1 + \cdots + y_n\vec{a}_n) \\ &= A\vec{x} + A\vec{y} \end{aligned}$$

and

$$A(c\vec{x}) = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix} = (cx_1)\vec{a}_1 + \cdots + (cx_n)\vec{a}_n = c(x_1\vec{a}_1 + \cdots + x_n\vec{a}_n) = c(A\vec{x}).$$

□

Remark 38. Let $A \in M_{m\times n}(\mathbb{K})$, and let $T : \mathbb{K}^n \to \mathbb{K}^m$ be defined by $T(\vec{x}) = A\vec{x}$. Then the last proposition implies that for every $\vec{x}, \vec{y} \in \mathbb{K}^n$ and $c \in \mathbb{K}$,

$$T(\vec{x}+\vec{y}) = A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y} = T(\vec{x}) + T(\vec{y}) \quad \text{and} \quad T(c\vec{x}) = A(c\vec{x}) = c(A\vec{x}) = cT(\vec{x}).$$

The fact that matrix transformations preserve vector addition and scalar multiplication makes them enormously important in the context of linear algebra. Indeed, we make a definition that captures all functions which preserve our two basic vector operations.

Definition 25. Let $T : \mathbb{K}^n \to \mathbb{K}^m$. We call T **linear** (or a **linear transformation**) if the following two conditions are satisfied:

(i) For every $\vec{x}, \vec{y} \in \mathbb{K}^n$, $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$.

(ii) For every $c \in \mathbb{K}$ and every $\vec{x} \in \mathbb{K}^m$, $T(c\vec{x}) = cT(\vec{x})$.

Remark 39. Mathematicians are generally interested in studying functions that preserve whatever interesting structure they are studying. For example, once one introduces the notion of limit in single-variable calculus one immediately starts to study continuous functions. To see how continuous functions "preserve limits", consider the following theorem (which we will not prove, but which you'll see in an analysis course):

Let $f : \mathbb{R} \to \mathbb{R}$ and $L \in \mathbb{R}$. Then f is continuous at L if, and only if, for every $g : \mathbb{R} \to \mathbb{R}$ with $\lim_{x\to a} g(x) = L$, we have $\lim_{x\to a} f(g(x)) = f(L) = f(\lim_{x\to a} g(x))$.

Compare this idea with the definition of linear transformations, which are functions that preserve the basic vector operations of addition and scalar multiplication.

Because linear transformations preserve vector addition and scalar multiplication, they also preserve linear combinations.

Proposition 12. Let $T : \mathbb{K}^n \to \mathbb{K}^m$. Then T is linear if, and only if, for every $k \in \mathbb{N}$, and every $c_1, \ldots, c_k \in \mathbb{K}$ and $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{K}^m$,

$$T(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k) = c_1T(\vec{v}_1) + \cdots + c_kT(\vec{v}_k).$$

Proof. First suppose that T preserves linear combinations. Let $\vec{x}, \vec{y} \in \mathbb{K}^n$ and $c \in \mathbb{K}$. Then

$$T(\vec{x}+\vec{y}) = T(1\vec{x}+1\vec{y}) = 1T(\vec{x}) + 1T(\vec{y}) = T(\vec{x}) + T(\vec{y}) \quad \text{and} \quad T(c\vec{x}) = cT(\vec{x}).$$

Therefore T is linear.

To prove the converse, suppose that T is linear. We proceed by induction. Let $\vec{v} \in \mathbb{K}^m$ and $c \in \mathbb{K}$. Then $T(c\vec{v}) = cT(\vec{v})$ by property (ii) of linearity.

Now let $k \in \mathbb{N}$ and assume that the result holds for linear combinations of k vectors. Let $c_1, \ldots, c_{k+1} \in \mathbb{K}$ and $\vec{v}_1, \ldots, \vec{v}_{k+1} \in \mathbb{K}^n$. Then we have

$$\begin{aligned} T(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k + c_{k+1}\vec{v}_{k+1}) &= T((c_1\vec{v}_1 + \cdots + c_k\vec{v}_k) + c_{k+1}\vec{v}_{k+1}) \\ &= T(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k) + T(c_{k+1}\vec{v}_{k+1}) \\ &= c_1T(\vec{v}_1) + \cdots + c_kT(\vec{v}_k) + c_{k+1}T(\vec{v}_{k+1}), \end{aligned}$$

where in the second step we used property (i) of linear transformations, and in the third step we used the induction hypothesis and the base case. By the Principle of Mathematical Induction, the proof is complete. □

As another example, linear transformations send the additive identity of $\mathbb{K}^n$ to the additive identity of $\mathbb{K}^m$.

Proposition 13. Let $T : \mathbb{K}^n \to \mathbb{K}^m$ be linear. Then $T(\vec{0}) = \vec{0}$.

Proof. The proof is almost a triviality, since $T(\vec{0}) = T(0\vec{0}) = 0T(\vec{0}) = \vec{0}$. □

Example 31. The function $S : \mathbb{K}^2 \to \mathbb{K}^3$ given by $S(\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]) = \begin{bmatrix} 2x_1-3x_2 \\ 5x_1+x_1^2 \\ x_2x_1+2 \end{bmatrix}$ is not linear. There are many ways to verify this. The simplest one is to note that $T(\vec{0}) = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \neq \vec{0}$. Alternatively, note that

$$T(2\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]) = T(\left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right]) = \begin{bmatrix} 4 \\ 14 \\ 2 \end{bmatrix} \neq \begin{bmatrix} 4 \\ 12 \\ 4 \end{bmatrix} = 2\begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix} = 2T(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]).$$

We can spend time determining exactly why the function S in the last example fails to be linear (e.g. the x_1^2 in the second entry is a problem, as is the x_2x_1 in the third entry, as is the $+2$ in the third entry), but it would be quicker to simply prove the remarkable fact that *every* linear transformation from $\mathbb{K}^n$ to $\mathbb{K}^m$ is a matrix transformation.

Theorem 15. Let $T : \mathbb{K}^n \to \mathbb{K}^m$. Then T is linear if, and only if, T is a matrix transformation. Moreover, the matrix $A \in M_{m\times n}(\mathbb{K})$ satisfying $T(\vec{x}) = A\vec{x}$ for every $\vec{x} \in \mathbb{K}^n$ is unique, and it is given by

$$A = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) & \cdots & T(\vec{e}_n) \end{bmatrix}.$$

Proof. Suppose there is a matrix $A \in M_{m\times n}(\mathbb{K})$ such that $T(\vec{x}) = A\vec{x}$ for every $\vec{x} \in \mathbb{K}^n$. The argument in Remark 38 shows that T is linear.

Now suppose that T is linear. Let $\vec{x} \in \mathbb{K}^n$. Then $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$, so because T is linear we apply Remark 12 to see that

$$T(\vec{x}) = T(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) = x_1T(\vec{e}_1) + \cdots + x_nT(\vec{e}_n) = \begin{bmatrix} T(\vec{e}_1) & \cdots & T(\vec{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} T(\vec{e}_1) & \cdots & T(\vec{e}_n) \end{bmatrix} \vec{x}.$$

Therefore $T(\vec{x}) = A\vec{x}$ for every $\vec{x} \in \mathbb{K}^n$, where $A = \begin{bmatrix} T(\vec{e}_1) & \cdots & T(\vec{e}_n) \end{bmatrix} \in M_{m \times n}(\mathbb{K})$.

For uniqueness, suppose that $B \in \mathbb{M}_{m\times n}(\mathbb{K})$ satisfies $T(\vec{x}) = B\vec{x}$ for every $\vec{x} \in \mathbb{K}^n$. Let $\vec{b}_1, \ldots, \vec{b}_n$ be the columns of B. Then for each $1 \le j \le n$, $T(\vec{e}_j) = B\vec{e}_j = \vec{b}_j$, so that $B = \begin{bmatrix} T(\vec{e}_1) & \cdots & T(\vec{e}_n) \end{bmatrix}$. This completes the proof. $\square$

Lecture 14: More Linear Transformations

Learning Objectives:

- Show that various transformations are linear, and produce the standard matrix of the transformation.
- Establish several standard examples of linear transformations.

We start today with another helpful characterization of linearity, and an observation about the relationship between linearity and linear independence.

Proposition 14. Let $T : \mathbb{K}^n \to \mathbb{K}^m$. The T is linear if, and only if, for every $\vec{x}, \vec{y} \in \mathbb{K}^n$ and every $\lambda \in \mathbb{K}$, $T(\lambda\vec{x} + \vec{y}) = \lambda T(\vec{x}) + T(\vec{y})$.

Proof. Suppose that T is linear. Let $\vec{x}, \vec{y} \in \mathbb{K}^n$ and $\lambda \in \mathbb{K}$. Then by Remark 12, $T(\lambda\vec{x}+\vec{y}) = T(\lambda\vec{x}+1\vec{y}) = \lambda T(\vec{x}) + 1T(\vec{y}) = \lambda T(\vec{x}) + T(\vec{y})$.

Now suppose that the second condition holds. Let $\vec{x}, \vec{y} \in \mathbb{K}$ and $\lambda \in \mathbb{K}$. Then

$$T(\vec{x} + \vec{y}) = T(1\vec{x} + \vec{y}) = 1T(\vec{x}) + T(\vec{y}) = T(\vec{x}) + T(\vec{y})$$

and, since $T(\vec{0}) = T(1\vec{0} + \vec{0}) = 1T(\vec{0}) + T(\vec{0}) = T(\vec{0}) + T(\vec{0})$ (and therefore $T(\vec{0}) = \vec{0}$),

$$T(\lambda\vec{x}) = T(\lambda\vec{x} + \vec{0}) = \lambda T(\vec{x}) + T(\vec{0}) = \lambda T(\vec{x}) + \vec{0} = \lambda T(\vec{x}).$$

Therefore T is linear. □

Example 32 (Warm-Up)**.** Let $T : \mathbb{K}^n \to \mathbb{K}^m$ be linear, and let $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{K}^n$. Show that if $T(\vec{v}_1), \ldots, T(\vec{v}_k)$ is linearly independent, then $\vec{v}_1, \ldots, \vec{v}_k$ is linearly independent. Is the converse necessarily true?

Suppose that $T(\vec{v}_1), \ldots, T(\vec{v}_k)$ is linearly independent. Suppose $c_1, \ldots, c_k \in \mathbb{K}$ such that $c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = \vec{0}$. Then

$$\vec{0} = T(\vec{0}) = T(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k) = c_1T(\vec{v}_1) + \cdots + c_kT(\vec{v}_k).$$

Because $T(\vec{v}_1), \ldots, T(\vec{v}_k)$ is linearly independent, $c_1 = \cdots = c_k = 0$. Therefore $\vec{v}_1, \ldots, \vec{v}_k$ is linearly independent.

The converse is not necessarily true. There are many counterexamples. For a counterexample, consider $T : \mathbb{K}^n \to \mathbb{K}^m$ given by $T(\vec{x}) = \vec{0} = O_{m\times n}\vec{x}$ for each $\vec{x} \in \mathbb{K}^n$. (Here $O_{m\times n}$ is the $m \times n$ **zero matrix**, which has each entry equal to 0.) Then $\vec{e}_1$ is a linearly independent set in $\mathbb{K}^n$, but since $T(\vec{e}_1) = \vec{0}$, $T(\vec{e}_1)$ is a linearly dependent set in $\mathbb{K}^m$.

In light of the theorem from last time, we make the following definition.

Definition 26. Let $T : \mathbb{K}^n \to \mathbb{K}^m$ be a linear transformation. The unique matrix $A \in M_{m\times n}(\mathbb{K})$ satisfying $T(\vec{x}) = A\vec{x}$ for every $\vec{x} \in \mathbb{K}^n$ is called the (**standard**) **matrix** of T.

The previous theorem is quite useful in practice. As a first observation, the theorem gives us a way to show that a given transformation is linear (i.e. show that it is a matrix transformation).

Example 33. Let $a \in \mathbb{K}$, and consider the transformation $T : \mathbb{K}^n \to \mathbb{K}^n$ defined by $T(\vec{x}) \stackrel{def}{=} a\vec{x}$.

Then T is linear (by Proposition 14), because for each $\vec{x}, \vec{y} \in \mathbb{K}^n$ and $c \in \mathbb{K}$,

$$T(c\vec{x} + \vec{y}) = a(c\vec{x} + \vec{y}) = ac\vec{x} + a\vec{y} = cT(\vec{x}) + T(\vec{y}).$$

The standard matrix of T is $n \times n$, and for each $1 \le k \le n$ the k-th column of T is $T(\vec{e}_k) = a\vec{e}_k$. Therefore

$$T(\vec{x}) = \begin{bmatrix} a & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a \end{bmatrix} \vec{x} \quad \text{for each } \vec{x} \in \mathbb{K}^n.$$

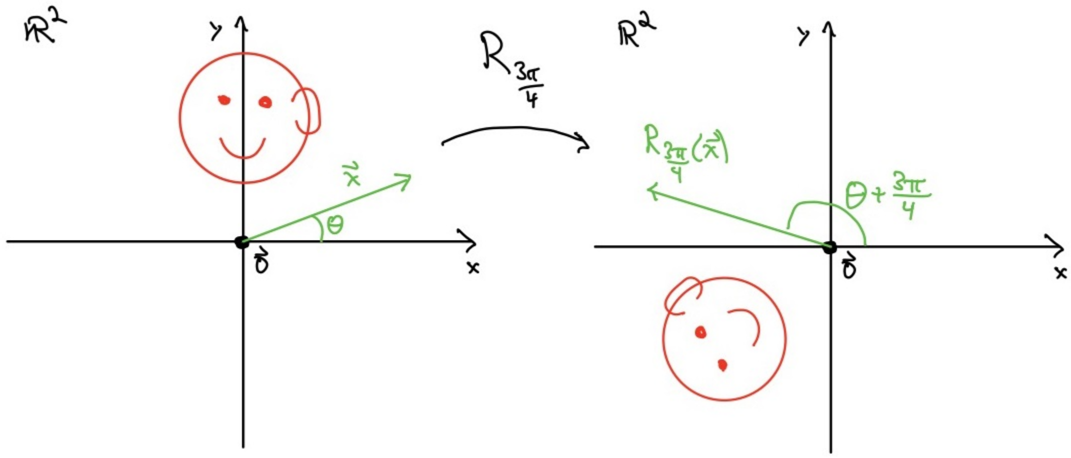
Example 34. There is a natural way to think about a vector $\vec{x} \in \mathbb{K}^n$ as a matrix in $M_{n\times 1}(\mathbb{K})$. Suppose that $\vec{x} \in \mathbb{K}^n$, $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Define $T : \mathbb{K}^1 \to \mathbb{K}^n$ by $T([a]) = a\vec{x}$. Then note that since, for every $[a], [b] \in \mathbb{K}$ and $\lambda \in \mathbb{K}$ we have

$$T(\lambda[a] + [b]) = (\lambda a + b)\vec{x} = \lambda(a\vec{x}) + b\vec{x} = \lambda T([a]) + T([b]),$$

T is linear. Note that in $\mathbb{K}^1$, $\vec{e}_1 = [1]$, so that the matrix of T is $\begin{bmatrix} T(\vec{e}_1) \end{bmatrix} = \begin{bmatrix} \vec{x} \end{bmatrix}$. Then we can consider $\vec{x}$ as corresponding to the matrix $X \in M_{n\times 1}(\mathbb{K})$ given by $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. In particular, note that we have, for each $a \in \mathbb{K}$,

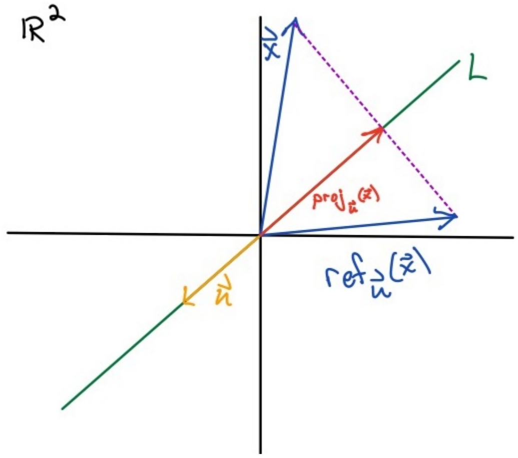
$$X[a] = T([a]) = a\vec{x}.$$

Example 35. Let $\alpha \in \mathbb{R}$. Let $R_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates $\mathbb{R}^2$ by α radians about the origin.



On your homework this week, you will rigorously define R_α in terms of polar coordinates on $\mathbb{R}^2$, and you will show that R_α is a linear transformation by explicitly producing a matrix $A_\alpha \in M_{2\times 2}(\mathbb{R})$ such that $R_\alpha(\vec{x}) = A_\alpha \vec{x}$.

Example 36. Let $\vec{u} \in \mathbb{R}^2$ with $\vec{u} \neq \vec{0}$. Let $L = \text{span}(\vec{u})$. Then L is the line through $(0, 0)$ that is parallel to $\vec{u}$. On your homework you will explore two linear transformations related to L: the orthogonal projection onto L and the orthogonal reflection across L:



We will explore the concept of orthogonality in MATH 291-2, but for now it suffices to say that "orthogonal" is a generalization of "perpendicular".

Example 37. There is no linear transformation $T : \mathbb{K}^3 \to \mathbb{K}^3$ such that

$$T\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right) = \begin{bmatrix}8\\0\\0\end{bmatrix}, \quad T\left(\begin{bmatrix}2\\-2\\1\end{bmatrix}\right) = \begin{bmatrix}8\\0\\2\end{bmatrix}, \quad T\left(\begin{bmatrix}3\\-1\\1\end{bmatrix}\right) = \begin{bmatrix}4\\0\\2\end{bmatrix}.$$

To see why, note that $\left[\begin{smallmatrix}3\\-1\\1\end{smallmatrix}\right] = \left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right] + \left[\begin{smallmatrix}2\\-2\\1\end{smallmatrix}\right]$, so if $T\left(\left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right]\right) = \left[\begin{smallmatrix}8\\0\\0\end{smallmatrix}\right]$ and $T\left(\left[\begin{smallmatrix}2\\-2\\1\end{smallmatrix}\right]\right) = \left[\begin{smallmatrix}8\\0\\2\end{smallmatrix}\right]$ then we must have

$$T\left(\begin{bmatrix}3\\-1\\1\end{bmatrix}\right) = T\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right) + T\left(\begin{bmatrix}2\\-2\\1\end{bmatrix}\right) = \begin{bmatrix}8\\0\\0\end{bmatrix} + \begin{bmatrix}8\\0\\2\end{bmatrix} = \begin{bmatrix}16\\0\\2\end{bmatrix} \neq \begin{bmatrix}4\\0\\2\end{bmatrix}.$$

Example 38. Let's determine whether there exists a linear transformation $T : \mathbb{K}^3 \to \mathbb{K}^3$ such that

$$T\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right) = \begin{bmatrix}8\\0\\0\end{bmatrix}, \quad T\left(\begin{bmatrix}2\\-2\\1\end{bmatrix}\right) = \begin{bmatrix}8\\0\\2\end{bmatrix}, \quad T\left(\begin{bmatrix}3\\-1\\-1\end{bmatrix}\right) = \begin{bmatrix}4\\0\\2\end{bmatrix}$$

Let's first suppose that such a transformation T exists. We will either derive a contradiction (and conclude that such a T cannot exist), or determine what T must be (and then verify that the transformation we produce does the trick).

To this end, let $A \in M_{3\times3}(\mathbb{K})$, $A = \begin{bmatrix}\vec{a}_1 & \vec{a}_2 & \vec{a}_3\end{bmatrix}$ be the matrix of T, so that $T(\vec{x}) = A\vec{x}$ for every $\vec{x} \in \mathbb{K}^3$. Then $\vec{a}_k = T(\vec{e}_k)$ for $k = 1, 2, 3$. The problem, of course, is that we only know the images of $\left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}2\\-2\\1\end{smallmatrix}\right]$, and $\left[\begin{smallmatrix}3\\-1\\-1\end{smallmatrix}\right]$ under T. However, if we can write $\vec{e}_1, \vec{e}_2, \vec{e}_3$ as linear combinations of $\left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}2\\-2\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}3\\-1\\-1\end{smallmatrix}\right]$, then we should be able to use linearity to compute T.

To write $\vec{e}_1$ as a linear combination of the three given vectors, we need to find a solution of

$$\begin{bmatrix}1 & 2 & 3\\1 & -2 & -1\\0 & 1 & -1\end{bmatrix}\vec{x} = \begin{bmatrix}1\\0\\0\end{bmatrix}.$$

To this end, we reduce the augmented matrix of the system to find

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 1 & -2 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & -4 & -4 & -1 \\ 0 & 1 & -1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 5 & 1 \\ 0 & 0 & -8 & -1 \\ 0 & 1 & -1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3/8 \\ 0 & 0 & 1 & 1/8 \\ 0 & 1 & 0 & 1/8 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3/8 \\ 0 & 1 & 0 & 1/8 \\ 0 & 0 & 1 & 1/8 \end{array}\right],$$

or rather that

$$\vec{e}_1 = \frac{3}{8}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{8}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} + \frac{1}{8}\begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}.$$

Similar computations yield that

$$\vec{e}_2 = \frac{5}{8}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{8}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \frac{1}{8}\begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}$$

and

$$\vec{e}_3 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}.$$

Therefore we must have

$$\vec{a}_1 = T(\vec{e}_1) = \frac{3}{8}\begin{bmatrix} 8 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{8}\begin{bmatrix} 8 \\ 0 \\ 2 \end{bmatrix} + \frac{1}{8}\begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 9/2 \\ 0 \\ 1/2 \end{bmatrix}$$

and

$$\vec{a}_2 = T(\vec{e}_2) = \frac{5}{8}\begin{bmatrix} 8 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{8}\begin{bmatrix} 8 \\ 0 \\ 2 \end{bmatrix} - \frac{1}{8}\begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 7/2 \\ 0 \\ -1/2 \end{bmatrix}$$

and

$$\vec{a}_3 = T(\vec{e}_1) = \frac{1}{2}\begin{bmatrix} 8 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 8 \\ 0 \\ 2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}.$$

Therefore the only possibility is that $A = \begin{bmatrix} 9/2 & 7/2 & 6 \\ 0 & 0 & 0 \\ 1/2 & -1/2 & 0 \end{bmatrix}$. We then verify that $T(\vec{x}) = A\vec{x}$ satisfies

$$T\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 9/2 & 7/2 & 6 \\ 0 & 0 & 0 \\ 1/2 & -1/2 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 8 \\ 0 \\ 0 \end{bmatrix}$$

and

$$T\left(\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 9/2 & 7/2 & 6 \\ 0 & 0 & 0 \\ 1/2 & -1/2 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 0 \\ 2 \end{bmatrix}$$

and

$$T\left(\begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 9/2 & 7/2 & 6 \\ 0 & 0 & 0 \\ 1/2 & -1/2 & 0 \end{bmatrix}\begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix},$$

as desired.

Lecture 15: Matrix Algebra

Learning Objectives:

- Define the basic matrix operations of addition, multiplication, and scalar multiplication and explore their basic properties and limitations.

The connection between linear transformations and matrices allow us to define matrix addition, scalar multiplication, and matrix multiplication that correspond with natural ways of combining linear transformations. Before defining these algebraic operations on matrices, we will first prove a result about linear transformations that shows that there is really only one reasonable way to define them.

Theorem 16. Let $T, S : \mathbb{K}^n \to \mathbb{K}^m$ and $R : \mathbb{K}^m \to \mathbb{K}^p$ be linear transformations, and let $A, B \in M_{m\times n}(\mathbb{K})$ and $C \in M_{p\times m}(\mathbb{K})$ be the standard matrices of T, S, and R (respectively). Write $A = \begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}$ and $B = \begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}$.

(i) The function $(T+S) : \mathbb{K}^n \to \mathbb{K}^m$, $(T+S)(\vec{x}) \overset{def}{=} T(\vec{x}) + S(\vec{x})$ is a linear transformation, and its $(m \times n)$ standard matrix is $\begin{bmatrix}\vec{a}_1 + \vec{b}_1 & \cdots & \vec{a}_n + \vec{b}_n\end{bmatrix}$.

(ii) For each $\alpha \in \mathbb{K}$, the function $\alpha T : \mathbb{K}^n \to \mathbb{K}^m$, $(\alpha T)(\vec{x}) \overset{def}{=} \alpha T(\vec{x})$ is a linear transformation, and its $(m \times n)$ standard matrix is $\begin{bmatrix}\alpha\vec{a}_1 & \cdots & \alpha\vec{a}_n\end{bmatrix}$.

(iii) The composition $R \circ T : \mathbb{K}^n \to \mathbb{K}^p$, $(R \circ T)(\vec{x}) \overset{def}{=} R(T(\vec{x}))$ is a linear transformation, and its $(p \times n)$ standard matrix is $\begin{bmatrix}C\vec{a}_1 & \cdots & C\vec{a}_n\end{bmatrix}$.

Proof. Let $\alpha \in \mathbb{K}$. To be efficient, we will prove the linearity of $T + S$, αT, and $R \circ T$ simultaneously.

Let $\vec{x}, \vec{y} \in \mathbb{K}^n$ and $\lambda \in \mathbb{K}$. Then

$$\begin{aligned}(T+S)(\lambda\vec{x}+\vec{y}) &= T(\lambda\vec{x}+\vec{y}) + S(\lambda\vec{x}+\vec{y}) \\ &= \lambda T(\vec{x}) + T(\vec{y}) + \lambda S(\vec{x}) + S(\vec{y}) \\ &= \lambda(T(\vec{x}) + S(\vec{x})) + T(\vec{y}) + S(\vec{y}) \\ &= \lambda(T+S)(\vec{x}) + (T+S)(\vec{y})\end{aligned}$$

and

$$(\alpha T)(\lambda\vec{x}+\vec{y}) = \alpha T(\lambda\vec{x}+\vec{y}) = \alpha(\lambda T(\vec{x}) + T(\vec{y})) = (\alpha\lambda)T(\vec{x}) + \alpha T(\vec{y}) = \lambda(\alpha T)(\vec{x}) + (\alpha T)(\vec{y})$$

and

$$(R\circ T)(\lambda\vec{x}+\vec{y}) = R(T(\lambda\vec{x}+\vec{y})) = R(\lambda T(\vec{x}) + T(\vec{y})) = \lambda R(T(\vec{x})) + R(T(\vec{y})) = \lambda(R\circ T)(\vec{x}) + (R\circ T)(\vec{y}).$$

Therefore Proposition 14 implies that $T + S$, αT, and $R \circ T$ are linear.

For the matrix formulas, we simply note that for each integer k from 1 to n, then k-th column of the standard matrices of $T+S$, αT, and $R \circ T$ are (respectively)

$$(T+S)(\vec{e}_k) = T(\vec{e}_k) + S(\vec{e}_k) = \vec{a}_k + \vec{b}_k$$

and

$$(\alpha T)(\vec{e}_k) = \alpha T(\vec{e}_k) = \alpha \vec{a}_k$$

and

$$(R \circ T)(\vec{e}_k) = R(T(\vec{e}_k) = R(\vec{a}_k) = C\vec{a}_k.$$

This completes the proof. □

In light of the previous theorem, we make the following definitions.

Definition 27. Let $A, B \in M_{m\times n}(\mathbb{K})$, $C \in M_{p\times m}(\mathbb{K})$, and $\alpha \in \mathbb{K}$. Write $A = \begin{bmatrix} a_{j,k} \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$ and $B = \begin{bmatrix} b_{j,k} \end{bmatrix}$. Define

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} + \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{bmatrix} \overset{def}{=} \begin{bmatrix} a_{1,1}+b_{1,1} & \cdots & a_{1,n}+b_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1}+b_{m,1} & \cdots & a_{m,n}+b_{m,n} \end{bmatrix},$$

$$\alpha A = \alpha \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \overset{def}{=} \begin{bmatrix} \alpha a_{1,1} & \cdots & \alpha a_{1,n} \\ \vdots & \ddots & \vdots \\ \alpha a_{m,1} & \cdots & \alpha a_{m,n} \end{bmatrix},$$

and

$$CA = C \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \overset{def}{=} \begin{bmatrix} C\vec{a}_1 & \cdots & C\vec{a}_n \end{bmatrix}.$$

Remark 40. Suppose that $A \in M_{m\times n}(\mathbb{K})$ and $B \in M_{k\times p}(\mathbb{K})$. Then note that AB is only defined when $n = k$ (and in this case, it is a $m \times p$ matrix). This is, of course, tied to the fact that matrix multiplication comes from composition of linear transformations, and therefore the outputs of the first linear transformation ($S(\vec{x}) = B\vec{x}$) must be suitable inputs for the second linear transformation ($T(\vec{y}) = A\vec{y}$). As an additional wrinkle, note that BA may be undefined (i.e. $p \neq m$)) even if AB is defined ($n = k$).

Remark 41. Because the columns of the product

$$CA = C \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} = \begin{bmatrix} C\vec{a}_1 & \cdot & C\vec{a}_n \end{bmatrix}$$

are exactly the matrix C applied to the columns of A, our computational device for computing the product of a matrix and a vector can be used to compute the product of two matrices. In particular, the entry in the j-th row and k-th column of CA is exactly the entry in the j-th row of $C\vec{a}_k$, so that

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{j,1} & c_{j,2} & \cdots & c_{j,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p,1} & c_{p,2} & \cdots & c_{p,m} \end{bmatrix} \begin{bmatrix} a_{1,1} & \cdots & a_{1,k} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,k} & \cdots & a_{2,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,k} & \cdots & a_{m,n} \end{bmatrix} = \begin{bmatrix} \ddots & \vdots & \iddots \\ \cdots & c_{j,1}a_{1,k} + c_{j,2}a_{2,k} \cdots + c_{j,m}a_{m,k} & \cdots \\ \iddots & \vdots & \ddots \end{bmatrix}.$$

Remark 42. It is immediate to check (either by associativity and commutativity of real number addition or of vector addition) that $A+(B+C)=(A+B)+C$ and $A+B=B+A$ when A,B,C are matrices of the same size.

Remark 43. The sum of matrices is straightforward to compute: we simply add the corresponding entries of the two matrices. For example,

$$\begin{bmatrix}1 & 2\\ 3 & 4\\ 5 & 6\end{bmatrix}+\begin{bmatrix}0 & -1\\ 2 & 2\\ 1 & 6\end{bmatrix}=\begin{bmatrix}1 & 1\\ 5 & 6\\ 6 & 12\end{bmatrix}.$$

Note that because addition of linear transformations is only defined when the transformations have the same domain and the same codomain, the notion of matrix addition only makes sense for matrices with the same number of rows and columns. Moreover, if $A,B\in M_{m\times n}(\mathbb{K})$, then in light of part (i) in Theorem 16 the definition of matrix addition can be interpreted as

$$(A+B)\vec{x}=A\vec{x}+B\vec{x}\quad\text{for every } \vec{x}\in\mathbb{K}^n.$$

Therefore, the definition of matrix addition ensures that matrix addition distributes over multiplication by a vector. We can even strengthen this by noting that if $C=\begin{bmatrix}\vec{c}_1 & \cdots & \vec{c}_p\end{bmatrix}\in M_{n\times p}(\mathbb{K})$, then we have

$$\begin{aligned}(A+B)C &= \begin{bmatrix}(A+B)\vec{c}_1 & \cdots & (A+B)\vec{c}_p\end{bmatrix}\\ &= \begin{bmatrix}A\vec{c}_1+B\vec{c}_1 & \cdots & A\vec{c}_p+B\vec{c}_p\end{bmatrix}\\ &= \begin{bmatrix}A\vec{c}_1 & \cdots & A\vec{c}_p\end{bmatrix}+\begin{bmatrix}B\vec{c}_1 & \cdots & B\vec{c}_p\end{bmatrix}\\ &= AC+BC,\end{aligned}$$

so that matrix addition distributes over matrix multiplication. (The case $A(B+C)=AB+AC$ for appropriately sized A,B,C follows by a similar argument.)

Remark 44. The scalar multiple of a matrix is also easy to compute: we simply scale the entries of the matrix. For example,

$$-3\begin{bmatrix}0 & -1\\ 2 & 2\\ 1 & 6\end{bmatrix}=\begin{bmatrix}0 & 3\\ -6 & -6\\ -3 & -18\end{bmatrix}.$$

If $A\in M_{m\times n}(\mathbb{K})$ and $\alpha\in\mathbb{K}$, then in light of part (ii) in Theorem 16 and the linearity of $T(\vec{x})=A\vec{x}$, scalar multiplication of a matrix satisfies

$$(\alpha A)\vec{x}=\alpha(A\vec{x})=A(\alpha\vec{x})\quad\text{for every } \vec{x}\in\mathbb{K}^n.$$

Therefore, the scalar multiplication for matrices is "associative" in this sense. Of as in the last remark, if $C=\begin{bmatrix}\vec{c}_1 & \cdots & \vec{c}_p\end{bmatrix}\in M_{n\times p}(\mathbb{K})$ then

$$(\alpha A)C=\begin{bmatrix}(\alpha A)\vec{c}_1 & \cdots & (\alpha A)\vec{c}_p\end{bmatrix}=\begin{bmatrix}\alpha(A\vec{c}_1) & \cdots & \alpha(A\vec{c}_p)\end{bmatrix}=\alpha(AC)$$

and

$$(\alpha A)C=\begin{bmatrix}(\alpha A)\vec{c}_1 & \cdots & (\alpha A)\vec{c}_p\end{bmatrix}=\begin{bmatrix}A(\alpha\vec{c}_1) & \cdots & A(\alpha\vec{c}_p)\end{bmatrix}=A(\alpha C),$$

so that $(\alpha A)C=\alpha(AC)=A(\alpha C)$ as expected. Similarly, once can verify that $\alpha(A+B)=\alpha A+\alpha B$, $(\alpha+\beta)A=\alpha A+\beta A$, $1A=A$, and $\alpha(\beta A)=(\alpha\beta)A$ for each $\alpha,\beta\in\mathbb{K}$, and each pair of matrices A,B of the same size.

Example 39. Note that the linear transformation $T : \mathbb{K}^n \to \mathbb{K}^n$ explored in Example 33 is exactly $T(\vec{x}) = \alpha\vec{x} = \alpha I(\vec{x}) = (\alpha I)(\vec{x})$, where $\alpha \in \mathbb{K}$ is a fixed scalar. Therefore Theorem 16 necessitates that the matrix of this transformation is

$$aI_n = a \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} a & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a \end{bmatrix},$$

which (of course) is exactly what we obtained in that example by computing directly.

Matrix addition and scalar multiplication have many standard properties. We showed above that matrix multiplication distributes over matrix multiplication.

Example 40. Note that the definition of matrix product ensures that for all $A \in M_{m\times n}(\mathbb{K})$ and $B \in M_{n\times p}(\mathbb{K})$ and $\vec{x} \in \mathbb{K}^p$, we have

$$(AB)\vec{x} = A(B\vec{x}),$$

so that matrix multiplication is "associative" in this sense. Indeed, we can extend this by noting that if $C = \begin{bmatrix} \vec{c}_1 & \cdots & \vec{c}_q \end{bmatrix} \in M_{p\times q}(\mathbb{K})$, then

$$(AB)C = \begin{bmatrix} (AB)\vec{c}_1 & \cdots & (AB)\vec{c}_q \end{bmatrix} = \begin{bmatrix} A(B\vec{c}_1) & \cdots & A(B\vec{c}_q) \end{bmatrix} = A \begin{bmatrix} B\vec{c}_1 & \cdots & B\vec{c}_q \end{bmatrix} = A(BC),$$

so that matrix multiplication is associative in the expected sense.

Lecture 16: Invertibility

Learning Objectives:

- Explore the basic notions of invertibility for matrices and linear transformations.
- Characterize invertibility of matrices in terms of solutions of linear systems.

Remark 45. Even when the products AB and BA are defined, they may not be equal to one another. That is, matrix multiplication fails to be commutative in general. For a counterexample, consider $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$. Then $AB = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ while $BA = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$.

Because matrix multiplication is not commutative, the idea of "multiplicative identity" is more complicated as well.

Example 41. Let $A \in M_{m\times n}(\mathbb{K})$, $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$. Then note that $I_m A$ and AI_n are defined (since I_m has the same number of columns as A has rows, and I_n has the same number of rows at A has columns). Moreover, we have

$$I_m A = I_m \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} = \begin{bmatrix} I_m\vec{a}_1 & \cdots & I_m\vec{a}_n \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} = A$$

and

$$AI_n = A \begin{bmatrix} \vec{e}_1 & \cdots & \vec{e}_n \end{bmatrix} = \begin{bmatrix} A\vec{e}_1 & \cdots & A\vec{e}_n \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} = A.$$

Therefore the identity matrices act as "multiplicative identities" for matrices. What is complicated here is that there are infinitely many identity matrices (one I_k for each $k = 1, 2, 3, \ldots$), and $I_k A = A$ if A has k rows, while $AI_j = A$ when A has j columns.

The next natural question, of course, is to ask in what sense a matrix may have a multiplicative inverse.

Definition 28. Let $A \in M_{n\times n}(\mathbb{K})$. Then we say that A is **invertible** if there exists $B \in M_{n\times n}(\mathbb{K})$ such that $AB = I_n$ and $BA = I_n$.

Note that here we need both of the equations $AB = I_n$ and $BA = I_n$ because matrix multiplication is not commutative. Even with this complication, just as for the other algebraic objects we have seen there is only one possible inverse of an invertible matrix.

Proposition 15. Let $A \in M_{n\times n}(\mathbb{K})$. If A is invertible, then the matrix $B \in M_{n\times n}(\mathbb{K})$ satisfying $AB = I_n$ and $BA = I_n$ is unique.

Proof. Suppose that $C \in M_{n\times n}(\mathbb{K})$ is another such matrix. Then

$$C = CI_n = C(AB) = (CA)B = I_nB = B.$$

□

We are therefore justified in making the following definition.

Notation 1. Let $A \in M_{n\times n}(\mathbb{K})$. If A is invertible, then we denote by A^{-1} the unique $n \times n$ matrix such that $AA^{-1} = I_n$ and $A^{-1}A = I_n$. A^{-1} is called the **inverse** of A.

Remark 46. It might seem that this definition of invertability is a little too restrictive. Arguably we should have considered matrices that were not square, and we should have said that a matrix $A \in M_{m\times n}(\mathbb{K})$ is invertible if there are matrices $B, C \in M_{n\times m}(\mathbb{K})$ such that $BA = I_n$ and $AC = I_m$.

As it turns out, though, this generalization does not give us a more general notion of invertibility. On your homework, you will show that if $A \in M_{m\times n}(\mathbb{K})$ is invertible in this more general sense, then $m = n$ (so that A must be square) and that $B = C$ (so that there is no need to consider distinct "left inverse" and "right inverse"). In other words, any matrix that satisfies this more general notion of invertibility must satisfy the more restrictive notion that we have defined above.

Example 42. Recall the question asked in Example 38: Does there exists a linear transformation $T : \mathbb{K}^3 \to \mathbb{K}^3$ such that

$$T\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right) = \begin{bmatrix}8\\0\\0\end{bmatrix}, \quad T\left(\begin{bmatrix}2\\-2\\1\end{bmatrix}\right) = \begin{bmatrix}8\\0\\2\end{bmatrix}, \quad T\left(\begin{bmatrix}3\\-1\\-1\end{bmatrix}\right) = \begin{bmatrix}4\\0\\2\end{bmatrix}?$$

Because linear transformations are matrix transformations, this is equivalent to asking whether there exists a matrix $A \in M_{3\times 3}(\mathbb{K})$ such that

$$A\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}8\\0\\0\end{bmatrix}, \quad A\begin{bmatrix}2\\-2\\1\end{bmatrix} = \begin{bmatrix}8\\0\\2\end{bmatrix}, \quad A\begin{bmatrix}3\\-1\\-1\end{bmatrix} = \begin{bmatrix}4\\0\\2\end{bmatrix}.$$

In terms of matrix products, this is equivalent to asking whether there exists $A \in M_{3\times 3}(\mathbb{K})$ such that

$$A\begin{bmatrix}1 & 2 & 3\\1 & -2 & -1\\0 & 1 & -1\end{bmatrix} = \begin{bmatrix}8 & 8 & 4\\0 & 0 & 0\\0 & 2 & 2\end{bmatrix}. \tag{4}$$

To solve this problem, we first noted that

$$\begin{bmatrix}1 & 2 & 3\\1 & -2 & -1\\0 & 1 & -1\end{bmatrix}\begin{bmatrix}3/8\\1/8\\1/8\end{bmatrix} = \vec{e}_1, \quad \begin{bmatrix}1 & 2 & 3\\1 & -2 & -1\\0 & 1 & -1\end{bmatrix}\begin{bmatrix}5/8\\-1/8\\-1/8\end{bmatrix} = \vec{e}_2, \quad \text{and} \quad \begin{bmatrix}1 & 2 & 3\\1 & -2 & -1\\0 & 1 & -1\end{bmatrix}\begin{bmatrix}1/2\\1/2\\-1/2\end{bmatrix} = \vec{e}_3.$$

In terms of matrices, we showed that

$$\begin{bmatrix}1 & 2 & 3\\1 & -2 & -1\\0 & 1 & -1\end{bmatrix}\begin{bmatrix}3/8 & 5/8 & 1/2\\1/8 & -1/8 & 1/2\\1/8 & -1/8 & -1/2\end{bmatrix} = I_3.$$

One can check that

$$\begin{bmatrix}3/8 & 5/8 & 1/2\\1/8 & -1/8 & 1/2\\1/8 & -1/8 & -1/2\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\1 & -2 & -1\\0 & 1 & -1\end{bmatrix} = I_3$$

as well, so that $\left[\begin{smallmatrix}1 & 2 & 3\\ 1 & -2 & -1\\ 0 & 1 & -1\end{smallmatrix}\right]$ is invertible and

$$\begin{bmatrix}3/8 & 5/8 & 1/2\\ 1/8 & -1/8 & 1/2\\ 1/8 & -1/8 & -1/2\end{bmatrix} = \begin{bmatrix}1 & 2 & 3\\ 1 & -2 & -1\\ 0 & 1 & -1\end{bmatrix}^{-1}.$$

Because $\left[\begin{smallmatrix}1 & 2 & 3\\ 1 & -2 & -1\\ 0 & 1 & -1\end{smallmatrix}\right]$ is invertible, the only matrix A that can satisfy equation (4) is

$$A = \begin{bmatrix}8 & 8 & 4\\ 0 & 0 & 0\\ 0 & 2 & 2\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\ 1 & -2 & -1\\ 0 & 1 & -1\end{bmatrix}^{-1} = \begin{bmatrix}8 & 8 & 4\\ 0 & 0 & 0\\ 0 & 2 & 2\end{bmatrix}\begin{bmatrix}3/8 & 5/8 & 1/2\\ 1/8 & -1/8 & 1/2\\ 1/8 & -1/8 & -1/2\end{bmatrix} = \begin{bmatrix}9/2 & 7/2 & 6\\ 0 & 0 & 0\\ 1/2 & -1/2 & 0\end{bmatrix},$$

which is what we found in Example 38!

Remark 47. Soon we will prove a result that will allows us to conclude that if $A, B \in M_{n\times n}(\mathbb{K})$ satisfy $AB = I_n$, then A is invertible and $B = A^{-1}$.

Invertibility of a matrix (and our ability to compute the inverse of a matrix) is, like everything else we have studied this quarter, intimately tied to systems of linear equations. We will soon explore this in full detail next time, but here is a first result.

Proposition 16. Let $A \in M_{n\times n}(\mathbb{K})$. Then A is invertible if, and only if, for each $\vec{y} \in \mathbb{K}^n$ the system $A\vec{x} = \vec{y}$ has exactly one solution.

Proof. Suppose that A is invertible. Let $\vec{y} \in \mathbb{K}^n$. Since $\vec{y} = A(A^{-1}\vec{y})$, $\vec{x} = A^{-1}\vec{y}$ is a solution of $A\vec{x} = \vec{y}$. If $\vec{z}$ also satisfies $A\vec{z} = \vec{y}$, then $\vec{z} = A^{-1}(A\vec{z}) = A^{-1}\vec{y}$, so that $A^{-1}\vec{y}$ is the only solution of $A\vec{x} = \vec{y}$.

Now assume that for each $\vec{y} \in \mathbb{K}^n$ the linear system $A\vec{x} = \vec{y}$ has a unique solution. For each $1 \le j \le n$, let $\vec{b}_j$ be the unique solution of $A\vec{x} = \vec{e}_j$. Then $B = \begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_n\end{bmatrix}$ satisfies

$$AB = \begin{bmatrix}A\vec{b}_1 & \cdots & A\vec{b}_n\end{bmatrix} = \begin{bmatrix}\vec{e}_1 & \cdots & \vec{e}_n\end{bmatrix} = I_n.$$

Now now that $A = I_nA = (AB)A = A(BA)$, so that $A(I_n - BA) = O_{n\times n}$. Let $I_n - BA = \begin{bmatrix}\vec{c}_1 & \cdots & \vec{c}_n\end{bmatrix}$. Then for each $1 \le j \le n$, $A\vec{c}_j = \vec{0}$, so that $\vec{c}_j = \vec{0}$ because $A\vec{c}_j = \vec{0}$ has a unique solution (and $\vec{0}$ is one solution, so must be the only one). Therefore $I_n - BA = O_{n\times n}$, so that $BA = I_n$ as well. Therefore A is invertible, and the proof is complete. □

Remark 48. Proposition 16 is the doorway to one of the most famous and widely used results in linear algebra, which we will discuss next time!

Lecture 17: More Invertibility

Learning Objectives:

- Determine various conditions under which a matrix or linear transformation are invertible.
- Compute the inverse of a matrix.

We can also define a notion of invertibility for linear transformations as well.

Definition 29. Let $T : \mathbb{K}^n \to \mathbb{K}^n$ be linear. We say that T is **invertible** if there is a linear transformation $S : \mathbb{K}^n \to \mathbb{K}^n$ such that $T(S(\vec{y})) = \vec{y}$ for every $\vec{y} \in \mathbb{K}^n$ and $S(T(\vec{x})) = \vec{x}$ for every $\vec{x} \in \mathbb{K}^n$.

Remark 49. Note that the conditions on S above can be more efficiently restated as $T \circ S = I$ and $S \circ T = I$, where $I : \mathbb{K}^n \to \mathbb{K}^n$ is the identity map.

Just as for the inverse of a matrix, the inverse of a linear transformation (when it exists) is unique.

Proposition 17. Let $T : \mathbb{K}^n \to \mathbb{K}^n$ be linear. If T is invertible, then the linear transformation $S : \mathbb{K}^n \to \mathbb{K}^n$ satisfying $T(S(\vec{y})) = \vec{y}$ for every $\vec{y} \in \mathbb{K}^n$ and $S(T(\vec{x})) = \vec{x}$ for every $\vec{x} \in \mathbb{K}^n$ is unique.

Proof. Suppose that $R : \mathbb{K}^n \to \mathbb{K}^n$ is another such transformation. Let $\vec{y} \in \mathbb{K}^n$. Then $R(\vec{y}) = R(T(S(\vec{y}))) = S(\vec{y})$, so $R = S$ as functions. □

Definition 30. Let $T : \mathbb{K}^n \to \mathbb{K}^n$ be linear. If T is invertible, then we denote by $T^{-1} : \mathbb{K}^n \to \mathbb{K}^n$ the unique linear transformation such that $T(T^{-1}(\vec{y})) = \vec{y}$ for every $\vec{y} \in \mathbb{K}^n$ and $T^{-1}(T(\vec{x})) = \vec{x}$ for every $\vec{x} \in \mathbb{K}^n$. T^{-1} is called the **inverse** of T.

Remark 50. The existence of an inverse for a linear transformation implies that the transformation has certain special properties as a function. To be able to state this properly, we make a definition.

Definition 31. Let A, B be sets and let $f : A \to B$ be a function. We say that f is **injective** if whenever $f(a) = f(c)$ then $a = c$. We say that f is **surjective** if for every $b \in B$ there is at least one $a \in A$ such that $f(a) = b$. We say that f is **bijective** if f is both injective and surjective.

Remark 51. The definition of surjectivity has a clear interpretation: $f : A \to B$ is surjective every element of B is the image under f of at least one element of A. For injectivity, it may be easier to understand the (logically equivalent) contrapositive of the definition: $f : A \to B$ is injective if any two distinct inputs $a \neq c$ in A yield distinct outputs $f(a) \neq f(c)$ in B.

Remark 52. On your homework, you prove that every bijective linear map is invertible.

Remark 53. We will shortly prove that invertible linear transformations have invertible standard matrices, and that invertible matrices give rise to invertible transformations.

The Invertibility Theorem

We now come to one of the most celebrated theorems in linear algebra: the Invertibility Theorem. This theorem usually goes by the name "Invertible Matrix Theorem", but because some of the statements in our version involve linear transformations rather than matrices, we will give it the more general name. The statement of the theorem is quite long, but because we have done so much work over the last few weeks, but proof is shockingly short.

Theorem 17 (Invertibility Theorem)**.** Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{n\times n}(\mathbb{K})$ and let $T : \mathbb{K}^n \to \mathbb{K}^n$ be the linear transformation $T(\vec{x}) = A\vec{x}$. The following statements are equivalent.

(a-1) A is invertible.

(a-2) For every $\vec{y} \in \mathbb{K}^n$, the linear system $A\vec{x} = \vec{y}$ has a unique solution.

(a-3) $\text{rank}(A) = n$

(a-4) $\text{rref}(A) = I_n$

(a-5) T is invertible.

(a-6) T is bijective.

(b-1) $\text{span}(\vec{a}_1, \ldots, \vec{a}_n) = \mathbb{K}^n$

(b-2) For every $\vec{y} \in \mathbb{K}^n$, the linear system $A\vec{x} = \vec{y}$ has at least one solution.

(b-3) $\text{rref}(A)$ has a pivot in every row.

(b-4) There is $B \in M_{n\times n}(\mathbb{K})$ with $AB = I_n$.

(b-5) T is surjective.

(b-6) There is a linear transformation $S : \mathbb{K}^n \to \mathbb{K}^n$ such that $T(S(\vec{y})) = \vec{y}$ for every $\vec{y} \in \mathbb{K}^n$.

(c-1) $\vec{a}_1, \ldots, \vec{a}_n$ is a linearly independent set.

(c-2) The solution set of the linear system $A\vec{x} = \vec{0}$ is $\{\vec{0}\}$.

(c-3) For every $\vec{y} \in \mathbb{K}^n$, the linear system $A\vec{x} = \vec{y}$ has at most one solution.

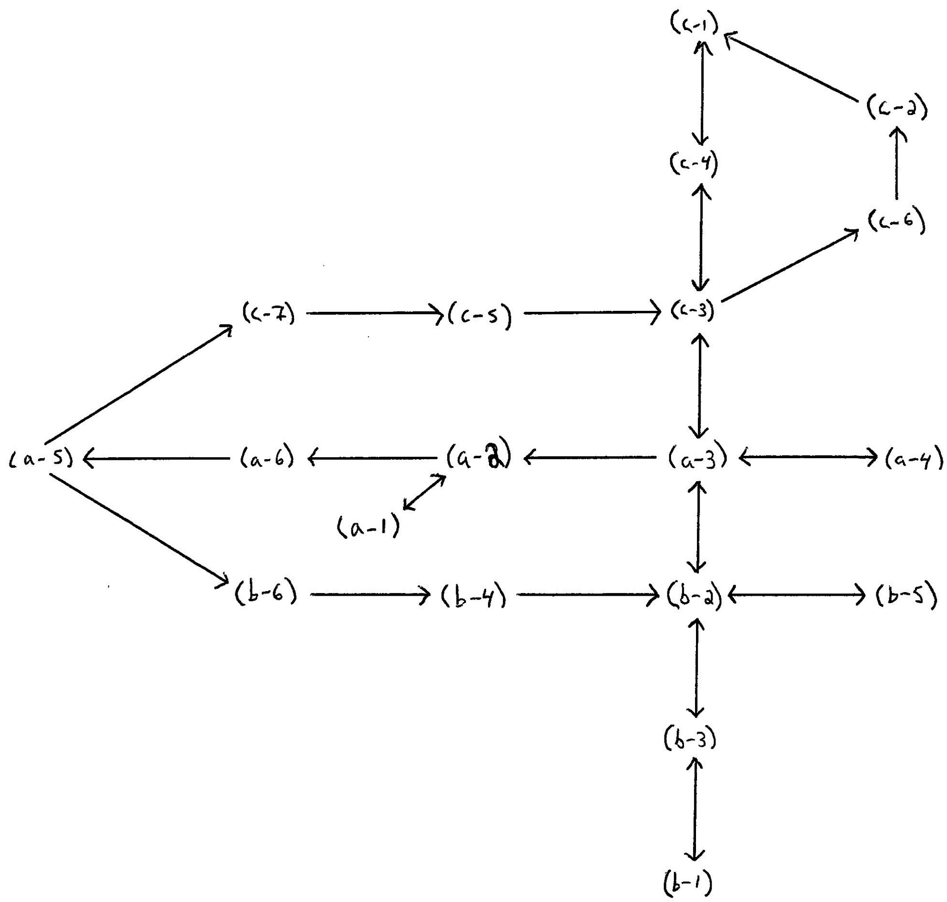
(c-4) $\text{rref}(A)$ has a pivot in every column.

(c-5) There is $B \in M_{n\times n}(\mathbb{K})$ with $BA = I_n$.

(c-6) T is injective.

(c-7) There is a linear transformation $S : \mathbb{K}^n \to \mathbb{K}^n$ with $S(T(\vec{x})) = \vec{x}$ for every $\vec{x} \in \mathbb{K}^n$.

Proof. We will prove the equivalence of the statements in the following way:



Proposition 16 implies that $(a-1) \Leftrightarrow (a-2)$.
By the Spanning Set in $\mathbb{K}^m$ Theorem:

$$(b-1) \Leftrightarrow (b-2) \Leftrightarrow (b-3) \Leftrightarrow (a-3)$$

By the Linear Independence and Linear Systems Theorem:

$$(c-1) \Leftrightarrow (c-3) \Leftrightarrow (c-4) \leftrightarrow (a-3)$$

$(a-3) \Leftrightarrow (a-4)$ is immediate because I_n has rank n and is the only $n \times n$ matrix with n pivots.

$(b-2) \Leftrightarrow (b-5)$ is immediate from the definition of surjectivity.

$(c-3) \Rightarrow (c-6)$: Assume $(c-3)$. Suppose that $T(\vec{x}) = T(\vec{z})$. Let $\vec{y}$ be this common value. Then $A\vec{x} = \vec{y} = A\vec{z}$, so that $\vec{x} = \vec{z}$.

$(c-6) \Rightarrow (c-2)$: Assume $(c-6)$. If $\vec{0} = A\vec{x} = T(\vec{x})$, then since $\vec{0} = A\vec{0} = T(\vec{0})$, $(c-6)$ implies that $\vec{x} = \vec{0}$.

$(c-2) \Rightarrow (c-1)$: Assume $(c-2)$. Let $c_1, \ldots, c_n \in \mathbb{K}$, and suppose $c_1\vec{a}_1 + \cdots + c_n\vec{a}_n = \vec{0}$. Then $A \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \vec{0}$. By $(c-2)$, $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \vec{0}$, so that $c_1 = \cdots = c_n = 0$.

$(a-3) \Rightarrow (a-2)$: We have $(a-3) \Rightarrow [(b-2)$ and $(c-3)] \Rightarrow (a-2)$.

$(a-2) \Rightarrow (a-6)$: We have $(a-2) \Rightarrow [(b-2)$ and $(c-3)] \Rightarrow [(b-5)$ and $(c-6)] \Rightarrow (a-6)$.

$(a-6) \Rightarrow (a-5)$: This is Exercise 5(a,b,c) in Homework 5.

$(a-5) \Rightarrow (c-7)$ and $(a-5) \Rightarrow (b-6)$ are immediate with $S = T^{-1}$.

$(c-7) \Rightarrow (c-5)$ and $(b-6) \Rightarrow (b-4)$ follow by letting B be the standard matrix of S.

$(c-5) \Rightarrow (c-3)$[15]: Assume $(c-5)$. Let $\vec{y} \in \mathbb{K}^n$. Suppose $\vec{x}$ solves $A\vec{x} = \vec{y}$. Then $\vec{x} = I_n\vec{x} = B(A\vec{x}) = B\vec{y}$, so this is the only possible solution of $A\vec{x} = \vec{y}$.
$(b-4) \Rightarrow (b-2)$: Assume $(b-4)$. Let $\vec{y} \in \mathbb{K}^n$. Then $\vec{y} = I_n\vec{y} = A(B\vec{y})$, so that $\vec{x} = B\vec{y}$ is a solution of $A\vec{x} = \vec{y}$.

This completes the proof. □

Remark 54. As a consequence of this theorem, note that a linear transformation $T : \mathbb{K}^n \to \mathbb{K}^n$ is invertible if, and only if, its standard matrix A is invertible. Moreover, it is immediate to check that if $S : \mathbb{K}^n \to \mathbb{K}^n$ is the inverse of T if, and only if, the standard matrix of S is A^{-1}.

Remark 55. As a consequence of the Invertibility Theorem, we have an excellent computational technique for determining when a square matrix is invertible, with the bonus that by the time we know that the matrix is invertible, we've already computed its inverse!

Theorem 18 (Computing Inverses)**.** Let $A \in M_{n\times n}(\mathbb{K})$. Let $\begin{bmatrix} A & I_n \end{bmatrix}$ be the $n \times (2n)$ matrix whose first n columns are the columns of A, and whose last n columns are the columns of I_n. Then A is invertible if, and only if, $\operatorname{rref}(\begin{bmatrix} A & I_n \end{bmatrix}) = \begin{bmatrix} I_n & B \end{bmatrix}$ for some matrix $B \in M_{n\times n}(\mathbb{K})$. Moreover, in this case we have $B = A^{-1}$.

Proof. Suppose that A is invertible. By the Invertibility Theorem, A is row equivalent to I_n. Performing a sequence of elementary row operations on $\begin{bmatrix} A & I_n \end{bmatrix}$ that transforms A into I_n, we transform $\begin{bmatrix} A & I_n \end{bmatrix}$ into $\begin{bmatrix} I_n & B \end{bmatrix}$ for some $B \in M_{n\times n}(\mathbb{K})$. Because $\begin{bmatrix} I_n & B \end{bmatrix}$ is in reduced row-echelon form, $\operatorname{rref}(\begin{bmatrix} A & I_n \end{bmatrix}) = \begin{bmatrix} I_n & B \end{bmatrix}$.

Conversely, suppose that there is $B \in M_{n\times n}(\mathbb{K})$ such that $\operatorname{rref}(\begin{bmatrix} A & I_n \end{bmatrix}) = \begin{bmatrix} I_n & B \end{bmatrix}$. Then $\begin{bmatrix} A & I_n \end{bmatrix}$ is row equivalent to $\begin{bmatrix} I_n & B \end{bmatrix}$. By Lemma 1 (applied n times, each time to the right-most column), we see that A is row equivalent to I_n, and therefore $\operatorname{rref}(A) = I_n$. By the Invertibility Theorem, A is invertible.

Now suppose that A is invertible, and let $B \in M_{n\times n}(\mathbb{K})$ be the matrix such that $\operatorname{rref}(\begin{bmatrix} A & I_n \end{bmatrix}) = \begin{bmatrix} I_n & B \end{bmatrix}$. Then $\begin{bmatrix} A & I_n \end{bmatrix} \vec{☺} = \vec{0}$ if, and only if, $\begin{bmatrix} I_n & B \end{bmatrix} \vec{☺} = \vec{0}$. Let $\vec{x} \in \mathbb{K}^n$, define $\vec{y} = A\vec{x}$, and let $\vec{☺} = \begin{bmatrix} \vec{x} \\ -\vec{y} \end{bmatrix} \in \mathbb{K}^{2n}$ be the vector whose first n entries are those of $\vec{x}$, and whose final n entries are those of $-\vec{y}$. Then

$$\begin{bmatrix} A & I_n \end{bmatrix} \vec{☺} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n - y_1\vec{e}_1 - \cdots - y_n\vec{e}_n = A\vec{x} - \vec{y} = \vec{0},$$

so that

$$\vec{0} = \begin{bmatrix} I_n & B \end{bmatrix} \vec{☺} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n - y_1\vec{b}_1 - \cdots - y_n\vec{b}_n = \vec{x} - B\vec{y} = \vec{x} - (BA)\vec{x},$$

and therefore $BA\vec{x} = \vec{x}$. Taking $\vec{x} = \vec{e}_j$ for each $j = 1, \dots, n$, we see that $BA = I_n$. By the Invertibility Theorem, there is $C \in M_{n\times n}(\mathbb{K})$ with $AC = I_n$. But then $C = I_nC = (BA)C = B(AC) = BI_n = B$, so that $AB = I_n$. Therefore A is invertible with $A^{-1} = B$. □

[15] The argument here shows that for every $\vec{x}$ that solves $A\vec{x} = \vec{y}$, $\vec{x} = B\vec{y}$. Therefore $B\vec{y}$ is the only possible solution of $A\vec{x} = \vec{y}$. In particular if $\vec{x}_1$ and $\vec{x}_2$ are two solutions of $A\vec{x} = \vec{y}$, then $\vec{x}_1 = B\vec{y} = \vec{x}_2$. This argument generated some confusion in class, so here is an equivalent (but hopefully more clear) alternate argument: Let $\vec{y} \in \mathbb{K}^n$. Suppose that $\vec{x}_1, \vec{x}_2 \in \mathbb{K}^n$ are solutions of $A\vec{z} = \vec{y}$. Then $A(\vec{x}_1 - \vec{x}_2) = A\vec{x}_1 - A\vec{x}_2 = \vec{y} - \vec{y} = \vec{0}$. Therefore $\vec{0} = B\vec{0} = BA(\vec{x}_1 - \vec{x}_2) = \vec{x}_1 - \vec{x}_2$, so that $\vec{x}_1 = \vec{x}_2$.

Remark 56. The previous theorem says that in order to tell whether $A \in M_{n\times n}(\mathbb{K})$ is invertible, we simply find the reduced row-echelon form of the augmented matrix $\begin{bmatrix} A & I_n \end{bmatrix}$. If the first n columns reduce to I_n, then A is invertible and the last n columns are A^{-1}. If the first n columns do not reduce to I_n, then A is not invertible!

Example 43. Let's revisit the example of whether $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -2 & -1 \\ 0 & 1 & -1 \end{bmatrix}$ is invertible. Then since

$$\begin{aligned} \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & -2 & -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \end{bmatrix} &\longrightarrow \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -4 & -4 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \end{bmatrix} \\ &\longrightarrow \begin{bmatrix} 1 & 0 & 5 & 1 & 0 & -2 \\ 0 & 0 & -8 & -1 & 1 & 4 \\ 0 & 1 & -1 & 0 & 0 & 1 \end{bmatrix} \\ &\longrightarrow \begin{bmatrix} 1 & 0 & 5 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1/8 & -1/8 & -1/2 \\ 0 & 1 & -1 & 0 & 0 & 1 \end{bmatrix} \\ &\longrightarrow \begin{bmatrix} 1 & 0 & 0 & 3/8 & 5/8 & 1/2 \\ 0 & 0 & 1 & 1/8 & -1/8 & -1/2 \\ 0 & 1 & 0 & 1/8 & -1/8 & 1/2 \end{bmatrix} \\ &\longrightarrow \begin{bmatrix} 1 & 0 & 0 & 3/8 & 5/8 & 1/2 \\ 0 & 1 & 0 & 1/8 & -1/8 & 1/2 \\ 0 & 0 & 1 & 1/8 & -1/8 & -1/2 \end{bmatrix}. \end{aligned}$$

Because the first three columns of this matrix are I_3, we conclude that A is invertible and

$$A^{-1} = \begin{bmatrix} 3/8 & 5/8 & 1/2 \\ 1/8 & -1/8 & 1/2 \\ 1/8 & -1/8 & -1/2 \end{bmatrix}.$$

Very slick!

Invertibility Theorem

Theorem (Invertibility Theorem)**.** Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{n\times n}(\mathbb{K})$ and let $T : \mathbb{K}^n \to \mathbb{K}^n$ be the linear transformation $T(\vec{x}) = A\vec{x}$. The following statements are equivalent.

(a-1) A is invertible.

(a-2) For every $\vec{y} \in \mathbb{K}^n$, the linear system $A\vec{x} = \vec{y}$ has a unique solution.

(a-3) $\operatorname{rank}(A) = n$

(a-4) $\operatorname{rref}(A) = I_n$

(a-5) T is invertible.

(a-6) T is bijective.

(b-1) $\operatorname{span}(\vec{a}_1, \ldots, \vec{a}_n) = \mathbb{K}^n$

(b-2) For every $\vec{y} \in \mathbb{K}^n$, the linear system $A\vec{x} = \vec{y}$ has at least one solution.

(b-3) $\operatorname{rref}(A)$ has a pivot in every row.

(b-4) There is $B \in M_{n\times n}(\mathbb{K})$ with $AB = I_n$.

(b-5) T is surjective.

(b-6) There is a linear transformation $S : \mathbb{K}^n \to \mathbb{K}^n$ such that $T(S(\vec{y})) = \vec{y}$ for every $\vec{y} \in \mathbb{K}^n$.

(c-1) $\vec{a}_1, \ldots, \vec{a}_n$ is a linearly independent set.

(c-2) The solution set of the linear system $A\vec{x} = \vec{0}$ is $\{\vec{0}\}$.

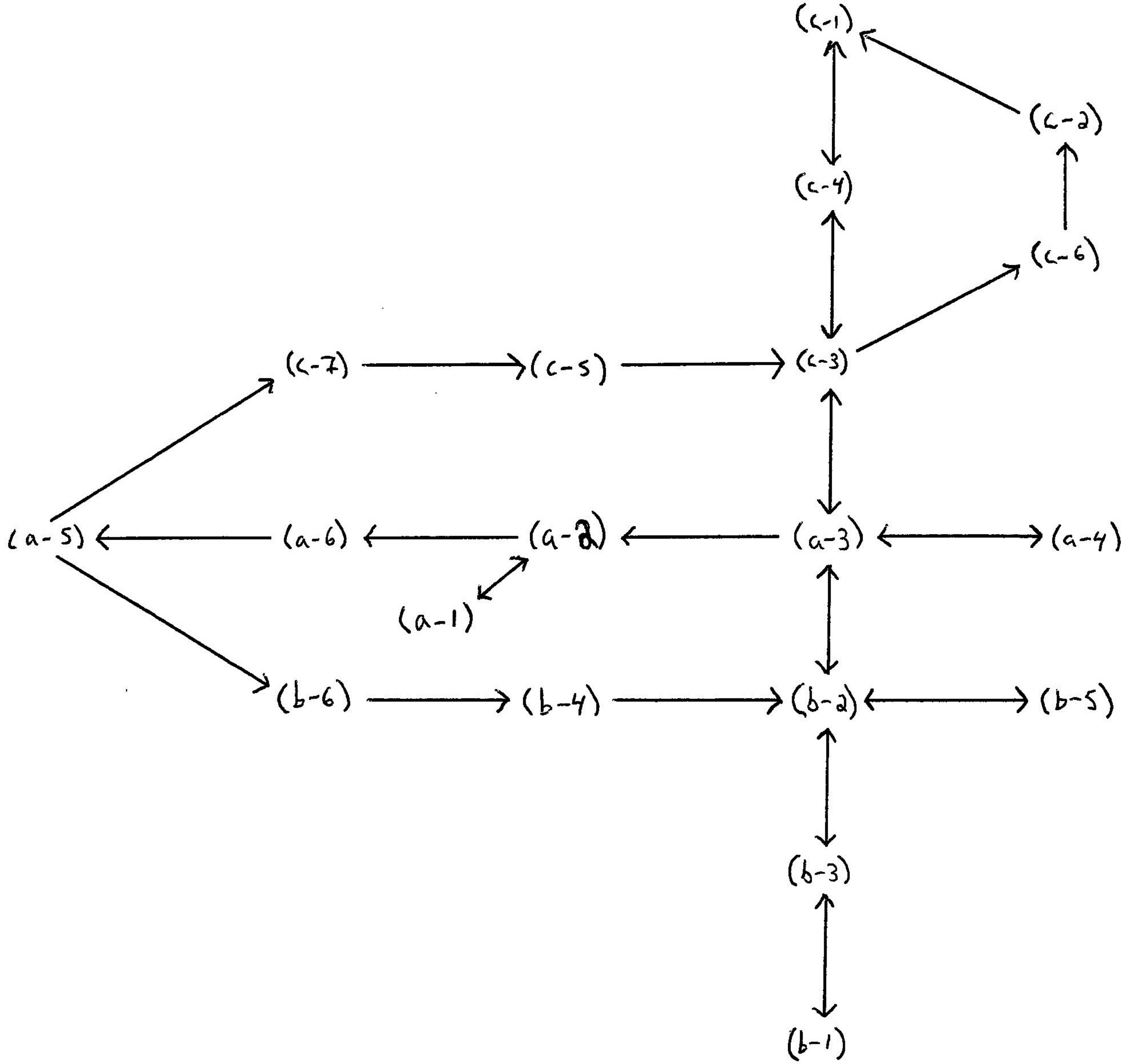
(c-3) For every $\vec{y} \in \mathbb{K}^n$, the linear system $A\vec{x} = \vec{y}$ has at most one solution.

(c-4) $\operatorname{rref}(A)$ has a pivot in every column.

(c-5) There is $B \in M_{n\times n}(\mathbb{K})$ with $BA = I_n$.

(c-6) T is injective.

(c-7) There is a linear transformation $S : \mathbb{K}^n \to \mathbb{K}^n$ with $S(T(\vec{x})) = \vec{x}$ for every $\vec{x} \in \mathbb{K}^n$.



(c-1)
(c-2)
(c-4)
(c-6)
(c-7)
(c-5)
(c-3)
(a-5)
(a-6)
(a-2)
(a-3)
(a-4)
(a-1)
(b-6)
(b-4)
(b-2)
(b-5)
(b-3)
(b-1)

Lecture 18: Vector Spaces

Learning Objectives:

- Generalize the properties of $\mathbb{K}^n$ to formulate a definition for a vector space over $\mathbb{K}$.
- Determine when a set with notions of addition and scalar multiplication is a vector space.
- Establish several standard examples of vector spaces.

At the beginning of the course we discussed how linear algebra began with the study of linear systems, and then generalized to encompass the study of vectors in $\mathbb{R}^n$ and $\mathbb{C}^n$. The first of these topics is algebraic, and the latter is geometric. However, we have now seen that each of these two aspects of linear algebra heavily inform the other, and we have built up an impressive base of results about vectors, linear systems, and linear transformations.

The time has come to address the more modern formulation of linear algebra as the study of *vector spaces*[16] which are sets of objects with notions of addition and scalar multiplication that mimics those those of $\mathbb{K}^n$. Here is the definition.

[16]Some mathematicians (including the author of your book) prefer the term *linear space* to vector space, as there is a feeling that vector space heavily connotes the very specific example of $\mathbb{K}^n$. I am not one of those mathematicians, though, and so we will use the vector space exclusively here.

Definition 32. Let $\mathbb{K}$ be a field. A set V is called a **vector space over** $\mathbb{K}$ if V is equipped with notions of addition and scalar multiplication that satisfy the following properties.

(1) For every $u, v \in V$ and $a \in \mathbb{K}$, $u + v \in V$ and $av \in V$.

(2) (Associativity of Vector Addition) For every $u, v, w \in V$,

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

(3) (Commutativity of Vector Addition) For every $u, v \in V$, $u + v = v + u$.

(4) (Additive Identity) There exists an element $0 \in V$ such that for every $v \in V$, $v + 0 = v$.

(5) (Additive Inverses) For every $v \in V$ there is $-v \in V$ such that $v + (-v) = 0$.

(6) (Associativity of Scalar Multiplication) For every $v \in V$ and $a, b \in \mathbb{K}$, $a(bv) = (ab)v$.

(7) (Distributivity of Scalar Multiplication over Scalar Addition) For every $v \in V$ and $a, b \in \mathbb{K}$,

$$(a + b)v = av + bv.$$

(8) (Distributivity of Scalar Multiplication over Vector Addition) For every $u, v \in V$ and $a \in \mathbb{K}$,

$$a(u + v) = au + av.$$

(9) (Scalar Multiplicative Identity) For every $v \in V$, $1v = v$.

If $\mathbb{K} = \mathbb{R}$, then call V a **real vector space**. If $\mathbb{K} = \mathbb{C}$, then call V a **complex vector space**.

Remark 57. Note that property (1) simply says that the notion of addition combine two vectors in V and return another vector in V, and that the notion of scalar multiplcation combines a scalar in $\mathbb{K}$ and a vector in V to give another vector in V.

Remark 58. Note that (2)-(3)-(4)-(5) exactly generalize the Properties of Vector Addition for $\mathbb{K}^n$, and that (6)-(7)-(8)-(9) exactly generalize the Properties of Scalar Multiplication for $\mathbb{K}^n$.

Example 44. $\mathbb{K}^n$, equipped with its usual notion of vector addition and scalar multiplcation, is a vector space over $\mathbb{K}$.

As a vector space over $\mathbb{R}$, $\mathbb{R}^n$ is a real vector space. Similarly, as a vector space over $\mathbb{C}$, $\mathbb{C}^n$ is a complex vector space.

Example 45. Let $n, m \in \mathbb{N}$. Then $M_{m\times n}(\mathbb{K})$, with the usual notions of matrix addition and scalar multiplcation, is a vector space over $\mathbb{K}$.

(1) The definition of matrix multiplcation and scalar multiplcation ensures that for every $A, B \in M_{m\times n}(\mathbb{K})$ and $a \in \mathbb{K}$, $A + B \in M_{m\times n}(\mathbb{K})$ and $aA \in M_{m\times n}(\mathbb{K})$.

(2)-(3) For every $A, B, C \in M_{m\times n}(\mathbb{K})$, Remark 42 shows that $A+(B+C) = (A+B)+C$ and $A+B = B+A$.

(4) The first problem on Quiz 4 shows that for every $A \in M_{m\times n}(\mathbb{K})$, $A + O_{m\times n} = A$.

(5) The second quiz problem on Quiz 4 shows that for every $A \in M_{m\times n}(\mathbb{K})$ there exists $-A \in M_{m\times n}(\mathbb{K})$ such that $A + (-A) = O_{m\times n}$.

(6)-(9) The method for verifying the properties of scalar multiplcation of matrices was addressed in Remark 44.

As a vector space over $\mathbb{R}$, $M_{m\times n}(\mathbb{R})$ is a real vector space. As a vector space over $\mathbb{C}$, $M_{m\times n}(\mathbb{C})$ is a complex vector space.

Several important examples of vector spaces can be viewed as special cases of the following (extremely general) examples.

Example 46. Let X be a nonempty set. Define $F(X, \mathbb{K})$ to be the space of functions from X into $\mathbb{K}$. That is,

$$F(X,\mathbb{K}) \overset{def}{=} \{f \ : \ f : X \to \mathbb{K}\}.$$

For $f, g \in F(X, \mathbb{K})$ and $a \in \mathbb{K}$, define

$$(f+g) : X \to \mathbb{K}, \quad (f+g)(x) \overset{def}{=} f(x) + g(x) \qquad \text{and} \qquad af : X \to \mathbb{K}, \quad (af)(x) \overset{def}{=} af(x).$$

We verify[17] that $F(X, \mathbb{K})$ is a vector space over $\mathbb{K}$.

(1) Let $f, g \in F(X, \mathbb{K})$ and $a \in \mathbb{K}$. The definition of $f + g$ and af implies that $f + g, af \in F(X, \mathbb{K})$.

(2) Let $f, g, h \in F(X, \mathbb{K})$. Then for every $x \in X$,

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

so that $(f + g) + h = f + (g + h)$ as functions from X to $\mathbb{K}$.

(3) Let $f, g \in F(X, \mathbb{K})$. Then for every $x \in X$,

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

so that $f + g = g + f$.

(4) Define $0 : X \to \mathbb{K}$ by $0(x) \overset{def}{=} 0$. Let $f \in F(X, \mathbb{K})$. Then for every $x \in X$,

$$(f+0)(x) = f(x)+0(x) = f(x)+0 = f(x),$$

so that $f + 0 = f$.

(5) Let $f \in F(X, \mathbb{K})$. Define $-f : X \to \mathbb{K}$ by $(-f)(x) \overset{def}{=} -f(x)$ for each $x \in X$. Then for each $x \in X$,

$$(f+(-f))(x) = f(x)-f(x) = 0 = 0(x),$$

so that $f + (-f) = 0$.

[17] Throughout this proof, we use that fact that for two functions ☺, ☹ $: X \to \mathbb{K}$, ☺ $=$ ☹ exactly exactly when ☺$(x) =$ ☹(x) for every $x \in X$.

(6) Let $f \in F(X,\mathbb{K})$ and $a,b \in \mathbb{K}$. Then for every $x \in X$,

$$(a(bf))(x) = a(bf)(x) = a(bf(x)) = (ab)f(x) = ((ab)f)(x),$$

so that $a(bf) = (ab)f$.

(7) Let $f \in F(X,\mathbb{K})$ and $a,b \in \mathbb{K}$. Then for every $x \in X$,

$$((a+b)f)(x) = (a+b)f(x) = af(x) + bf(x) = (af)(x) + (bf)(x) = (af+bf)(x),$$

so that $(a+b)f = af + bf$.

(8) Let $f,g \in F(X,\mathbb{K})$ and $a \in \mathbb{K}$. Then for every $x \in X$,

$$(a(f+g))(x) = a(f+g)(x) = a(f(x)+g(x)) = af(x)+ag(x) = (af)(x)+(ag)(x) = (af+ag)(x),$$

so that $a(f+g) = af + ag$.

(9) Let $f \in F(X,\mathbb{K})$. Then for every $x \in X$,

$$(1f)(x) = 1f(x) = f(x),$$

so that $1f = f$.

Therefore $F(X,\mathbb{K})$ is a vector space over $\mathbb{K}$.

Example 47. The argument above remains unchanged if we replace $F(X,\mathbb{K})$ with $F(X,V) = \{f : f : X \to V\}$ where V is a vector space over $\mathbb{K}$. Therefore if X is a nonempty set and V is a vector space over $\mathbb{K}$, then $F(X,V)$ is a vector space over $\mathbb{K}$.

Vector Spaces

Definition. Let $\mathbb{K}$ be a field. A set V is called a **vector space over** $\mathbb{K}$ if V is equipped with notions of addition and scalar multiplication that satisfy the following properties.

(1) For every $u, v \in V$ and $a \in \mathbb{K}$, $u + v \in V$ and $av \in V$.

(2) (Associativity of Vector Addition) For every $u, v, w \in V$,

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

(3) (Commutativity of Vector Addition) For every $u, v \in V$, $u + v = v + u$.

(4) (Additive Identity) There exists an element $0 \in V$ such that for every $v \in V$, $v + 0 = v$.

(5) (Additive Inverses) For every $v \in V$ there is $-v \in V$ such that $v + (-v) = 0$.

(6) (Associativity of Scalar Multiplication) For every $v \in V$ and $a, b \in \mathbb{K}$, $a(bv) = (ab)v$.

(7) (Distributivity of Scalar Multiplication over Scalar Addition) For every $v \in V$ and $a, b \in \mathbb{K}$,

$$(a + b)v = av + bv.$$

(8) (Distributivity of Scalar Multiplication over Vector Addition) For every $u, v \in V$ and $a \in \mathbb{K}$,

$$a(u + v) = au + av.$$

(9) (Scalar Multiplicative Identity) For every $v \in V$, $1v = v$.

If $\mathbb{K} = \mathbb{R}$, then call V a **real vector space**. If $\mathbb{K} = \mathbb{C}$, then call V a **complex vector space**.

Lecture 19: More Vector Spaces

Learning Objectives:

- Generalize various definition from $\mathbb{K}^n$ to general vector spaces.
- Review the elementary properties of $\mathbb{K}^n$ that remain true for general vector spaces (with the same proofs).

Remark 59. As one concrete instance of Example 46, note that in your single-variable calculus course you spent most of your time studying $F(\mathbb{R},\mathbb{R}) = \{f \ : \ f : \mathbb{R} \to \mathbb{R}\}$, the real vector space of real-valued functions on $\mathbb{R}$.

Remark 60. When you take a course in Complex Analysis, you will spend a lot of time studying $F(\mathbb{R},\mathbb{C})$ and $F(\mathbb{C},\mathbb{C})$, both of which are complex vector spaces.

Example 48. One important example that will provide counterexamples of many upcoming questions is $\mathbb{K}^\infty$, the space of sequences of scalars in $\mathbb{K}$. That is,

$$\mathbb{K}^\infty = \{(a_1, a_2, a_3, a_4, \ldots) \ : \ a_j \in \mathbb{K} \text{ for every } j \in \mathbb{N}\}.$$

Here, we define

$$(a_1, a_2, a_3, \ldots) + (b_1, b_2, b_3, \ldots) \overset{def}{=} (a_1 + b_1, a_2 + b_2, a_3 + b_3, \ldots)$$

and

$$c(a_1, a_2, a_3, \ldots) \overset{def}{=} (ca_1, ca_2, ca_3, \ldots)$$

for every $(a_1, a_2, a_3, \ldots), (b_1, b_2, b_3, \ldots) \in \mathbb{K}^\infty$ and every $c \in \mathbb{K}$. You will show on your homework that $\mathbb{K}^\infty$ is a vector space over $\mathbb{K}$.

It is sometimes (perhaps surprisingly) helpful to take what is usually a complex vector space, but view it as a real vector space. You'll explore this more deeply on a future homework assignment, but here is an example.

Example 49. $\mathbb{C}^n$ is a vector space over $\mathbb{R}$ with the usual notion of vector addition, and where we define

$$a\vec{v} \overset{def}{=} (a + i0)\vec{v} \quad \text{for every } a \in \mathbb{R},\ \vec{v} \in \mathbb{C}^n.$$

One can immediately verify that this satisfies the definition of vector space over $\mathbb{R}$. This viewpoint is useful when we wish to analyze complex vector spaces using results that are only valid for real vector spaces. (We will see several results of this form in MATH 291-2 and MATH 291-3.)

Properties of Vector Spaces

Vector spaces over $\mathbb{K}$ share many basic algebraic properties with $\mathbb{K}^n$. Indeed, everything that we proved about $\mathbb{K}^n$ using only the Properties of Vector Addition and Properties of Scalar Multiplcation (and, of course, the properties of the field $\mathbb{K}$) remain true for vector spaces over $\mathbb{K}$ with exactly the same proofs. (This is exactly why we spent so much time focusing on proving results without using the specific definitions of vector additon and scalar multiplication!) In particular, if V is a vector space over $\mathbb{K}$, then the following results hold (with exactly the same proofs that we gave earlier in the course):

- The additive identity $0 \in V$ is unique.
- For every $v \in V$, the additive inverse $-v$ of v is unique.
- $(-1)v = -v$ and $-(-v) = v$ for every $v \in V$.
- $0v = 0$ and $a0 = 0$ for every $v \in V$ and $a \in \mathbb{K}$.
- If $a \in \mathbb{K}$ and $v \in V$ satisfy $av = 0$, then $a = 0$ or $v = 0$.
- The distributive properties apply to finite sums. In particular, for every $k \in \mathbb{N}$,

  $$(a_1 + a_2 + \cdots + a_k)v = a_1v + a_2v + \cdots + a_kv \quad \text{for every } a_1, a_2, \ldots, a_k \in \mathbb{K} \text{ and } v \in V$$

  and

  $$a(v_1 + v_2 + \cdots + v_k) = av_1 + av_2 + \cdots + av_k \quad \text{for every } a \in \mathbb{K} \text{ and } v_1, v_2, \ldots, v_k \in V.$$

The elementary notions of linear combination, span, and linear independence remain exactly the same as in the case of $\mathbb{K}^n$.

Definition 33. Let V be a vector space over $\mathbb{K}$, and let $v_1, \ldots, v_n \in V$.

- A **linear combination** of $v_1, \ldots, v_n$ is a sum of the form

  $$c_1v_1 + c_2v_2 + \cdots + c_nv_n,$$

  where $c_1, c_2, \ldots, c_n \in \mathbb{K}$ are scalars (sometimes called the **coefficients** of the linear combination).
- Define the **span** of $v_1, \ldots, v_n$, denoted $\operatorname{span}(v_1, \ldots, v_n)$, by

  $$\operatorname{span}(v_1, \ldots, v_n) \overset{def}{=} \{c_1v_1 + \cdots + c_nv_n \ : \ c_1, \ldots, c_n \in \mathbb{K}\}.$$
- We call the set $v_1, \ldots, v_n$ **linearly independent** if for every $c_1, \ldots, c_n \in \mathbb{K}$, if

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

  then $c_1 = c_2 = \cdots = c_n = 0$. A set of vectors that is not linearly independent is called **linearly dependent**.

Many of the fundamental results about span and linear independence remain true (with the same proofs!) in the more general setting. Here are a couple such results. The proofs (in $\mathbb{K}^n$) can be found earlier in the notes (and, for the last result, in one of our quizzes).

Proposition 18 (Proposition 9)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \dots, v_n, u \in V$. If $u \in \operatorname{span}(v_1, \dots, v_n)$, then $\operatorname{span}(v_1, \dots, v_n, u) = \operatorname{span}(v_1, \dots, v_n)$.

Theorem 19 (Linear Independence and Linear Dependence)**.** Let V be a vector space over $\mathbb{K}$, let $m \geq 2$, and let $v_1, \dots, v_m \in V$.

(a) $v_1, \dots, v_m$ is a linearly independent set if, and only if, none of $v_1, \dots, v_m$ can be written as a linear combination of the others.

(b) $v_1, \dots, v_m$ is a linearly dependent set if, and only if, at least one of $v_1, \dots, v_m$ can be written as a linear combination of the others.

Proposition 19 (Proposition 10)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \dots, v_m, u \in V$. Assume that $v_1, \dots, v_m$ is linearly independent set, and that $u \notin \operatorname{span}(v_1, \dots, v_m)$. Then $v_1, \dots, v_m, u$ is a linearly independent set.

Theorem 20 (Theorem 6)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \dots, v_m \in V$. Then the following are equivalent:

(a) $v_1, \dots, v_m$ is a linearly independent set.

(b) There is a unique choice of scalars $c_1, \dots, c_m \in \mathbb{K}$ such that $0 = c_1 v_1 + \cdots + c_m v_m$.

(c) For every $u \in \operatorname{span}(v_1, \dots, v_m)$, there is a unique choice of scalars $c_1, \dots, c_m \in \mathbb{K}$ such that $u = c_1 v_1 + \cdots + c_m v_m$.

Proposition 20 (Problem 2, Quiz 2)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \dots, v_m, u_1, \dots, u_n \in V$. If $v_1, \dots, v_m, u_1, \dots, u_n$ is a linearly independent set, then $v_1, \dots, v_m$ is a linearly independent set.

Example 50. Consider $\mathbb{C}^2$ as a vector space over $\mathbb{R}$. Then note that if $\vec{x} = \begin{bmatrix} w \\ z \end{bmatrix} \in \mathbb{C}^2$, then writing $w = a + ib$ and $z = c + id$ for $a, b, c, d \in \mathbb{R}$, we have

$$\vec{x} = \begin{bmatrix} a + ib \\ c + id \end{bmatrix} = \begin{bmatrix} a \\ 0 \end{bmatrix} + \begin{bmatrix} ib \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ c \end{bmatrix} + \begin{bmatrix} 0 \\ id \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} i \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 1 \end{bmatrix} + d \begin{bmatrix} 0 \\ i \end{bmatrix}.$$

Therefore (as a vector space over $\mathbb{R}$) we have $\mathbb{C}^2 = \operatorname{span}(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} i \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ i \end{smallmatrix}\right])$. We also have that $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} i \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ i \end{smallmatrix}\right]$ is a linearly independent set, for if $c_1, c_2, c_3, c_4 \in \mathbb{R}$ satisfy

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} i \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + c_4 \begin{bmatrix} 0 \\ i \end{bmatrix} = \begin{bmatrix} c_1 + ic_2 \\ c_3 + ic_4 \end{bmatrix},$$

then $0 = c_1 + ic_2$ and $0 = c_3 + ic_4$, so that $c_1 = c_2 = c_3 = c_4 = 0$.

This might seem strange that we have a set of four vectors in $\mathbb{C}^2$ that is linearly independent, but the crucial point here is that we are considering $\mathbb{C}^2$ as a vector space over $\mathbb{R}$. When we consider $\mathbb{C}^2$ as a vector space over $\mathbb{C}$, then it would have been impossible to have a linearly independent set of four vectors. This illustrates the important point that when we consider a set V as a vector space over $\mathbb{K}$, our choice of $\mathbb{K}$ actually does affects the algebraic properties of V. We will see some surprising instances of this phenomenon next quarter.

Generalization of Definitions and Results

- The additive identity $0 \in V$ is unique.
- For every $v \in V$, the additive inverse $-v$ of v is unique.
- $(-1)v = -v$ and $-(-v) = v$ for every $v \in V$.
- $0v = 0$ and $a0 = 0$ for every $v \in V$ and $a \in \mathbb{K}$.
- If $a \in \mathbb{K}$ and $v \in V$ satisfy $av = 0$, then $a = 0$ or $v = 0$.
- The distributive properties apply to finite sums. In particular, for every $k \in \mathbb{N}$,

$$(a_1 + a_2 + \cdots + a_k)v = a_1 v + a_2 v + \cdots + a_k v \quad \text{for every } a_1, a_2, \ldots, a_k \in \mathbb{K} \text{ and } v \in V$$

and

$$a(v_1 + v_2 + \cdots + v_k) = av_1 + av_2 + \cdots + av_k \quad \text{for every } a \in \mathbb{K} \text{ and } v_1, v_2, \ldots, v_k \in V.$$

Definition. Let V be a vector space over $\mathbb{K}$, and let $v_1, \ldots, v_n \in V$.

- A **linear combination** of $v_1, \ldots, v_n$ is a sum of the form

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

where $c_1, c_2, \ldots, c_n \in \mathbb{K}$ are scalars (sometimes called the **coefficients** of the linear combination).

- Define the **span** of $v_1, \ldots, v_n$, denoted $\text{span}(v_1, \ldots, v_n)$, by

$$\text{span}(v_1, \ldots, v_n) \overset{def}{=} \{c_1 v_1 + \cdots + c_n v_n \ : \ c_1, \ldots, c_n \in \mathbb{K}\}.$$

- We call the set $v_1, \ldots, v_n$ **linearly independent** if for every $c_1, \ldots, c_n \in \mathbb{K}$, if

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

then $c_1 = c_2 = \cdots = c_n = 0$. A set of vectors that is not linearly independent is called **linearly dependent**.

Proposition (Proposition 9)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \ldots, v_n, u \in V$. If $u \in \operatorname{span}(v_1, \ldots, v_n)$, then $\operatorname{span}(v_1, \ldots, v_n, u) = \operatorname{span}(v_1, \ldots, v_n)$.

Theorem (Linear Independence and Linear Dependence)**.** Let V be a vector space over $\mathbb{K}$, let $m \geq 2$, and let $v_1, \ldots, v_m \in V$.

(a) $v_1, \ldots, v_m$ is a linearly independent set if, and only if, none of $v_1, \ldots, v_m$ can be written as a linear combination of the others.

(b) $v_1, \ldots, v_m$ is a linearly dependent set if, and only if, at least one of $v_1, \ldots, v_m$ can be written as a linear combination of the others.

Proposition (Proposition 10)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \ldots, v_m, u \in V$. Assume that $v_1, \ldots, v_m$ is linearly independent set, and that $u \notin \operatorname{span}(v_1, \ldots, v_m)$. Then $v_1, \ldots, v_m, u$ is a linearly independent set.

Theorem (Theorem 6)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \ldots, v_m \in V$. Then the following are equivalent:

(a) $v_1, \ldots, v_m$ is a linearly independent set.

(b) There is a unique choice of scalars $c_1, \ldots, c_m \in \mathbb{K}$ such that $0 = c_1 v_1 + \cdots + c_m v_m$.

(c) For every $u \in \operatorname{span}(v_1, \ldots, v_m)$, there is a unique choice of scalars $c_1, \ldots, c_m \in \mathbb{K}$ such that $u = c_1 v_1 + \cdots + c_m v_m$.

Proposition (Problem 2, Quiz 2)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \ldots, v_m, u_1, \ldots, u_n \in V$. If $v_1, \ldots, v_m, u_1, \ldots, u_n$ is a linearly independent set, then $v_1, \ldots, v_m$ is a linearly independent set.

Lecture 20: Subspaces

Learning Objectives:

- Determine when a subset of a vector space is a vector space in its own right.
- Generate several examples of interesting vector spaces that can be viewed as subspaces of other vector spaces.

Remark 61. Note that the empty set $\emptyset$ is *not* a vector space over $\mathbb{K}$ with any notion of addition or scalar multiplcation, since property (4) of vector spaces (i.e. that there is an additive identity 0) implies that a vector space must contain at least one element (i.e. an additive identity).

We have now seen quite a few examples of vector spaces over $\mathbb{K}$: $\mathbb{K}^n$, $M_{m\times n}(\mathbb{K})$, $\mathbb{K}^\infty$, and even more abstract examples like $F(X, \mathbb{K})$ (for a nonempty set X) and $F(X, V)$ (for a nonempty set X and a vector space V over $\mathbb{K}$). We even saw an example of how to think of the complex vector space $\mathbb{C}^n$ as a vector space over $\mathbb{R}$. (You will explore this last example in more detail on your next homework.)

Many of the examples that we care about in applications are not in the list above, but rather are *subspaces* of the spaces listed above. To make this precise, we give a definition.

Definition 34. Let V be a vector space over $\mathbb{K}$. A subset[18] $W \subseteq V$ is called a **subspace** of V if W is a vector space over $\mathbb{K}$, where addition and scalar multiplication in W are the same as in V.

Here are some examples (proofs to follow).

Example 51. Let V be a vector space over $\mathbb{K}$. Then V is a subspace of itself.

Example 52. Let V be a vector space over $\mathbb{K}$, and let 0 denote the additive identity in V. Then $\{0\}$ is a subspace of V (called the **trivial subspace**).

Example 53. Let V be a vector space over $\mathbb{K}$, and let $v_1, \dots, v_m \in V$. Then $\operatorname{span}(v_1, \dots, v_m)$ is a subspace of V.

Example 54. The set $P_n(\mathbb{K})$ of polynomials over $\mathbb{K}$ with degree no more n, that is

$$P_n(\mathbb{K}) \overset{def}{=} \{f : \mathbb{K} \to \mathbb{K} \ : \ \exists a_0, a_1, \dots, a_n \in \mathbb{K} \text{ such that } \forall x \in \mathbb{K},\ f(x) = a_0 + a_1 x + \cdots + a_n x^n\},$$

is a vector space over $\mathbb{K}$ with the usual notions of addition and scalar multiplcation of functions. Note that $P_n(\mathbb{K}) \subseteq F(\mathbb{K}, \mathbb{K})$.

[18]If A, B are sets, the notation $A \subseteq B$, read "A is a subset of B", indicates that every element of A is also an element of B.

Example 55. Let $I \subseteq \mathbb{R}$ be an interval. Then the space

$$C^0(I, \mathbb{R}) \overset{def}{=} \{f : I \to \mathbb{R} \ : \ f \text{ is continuous}\}$$

of continuous, real-valued functions on I is a subspace of $F(I, \mathbb{R})$.

If V is a vector space over $\mathbb{K}$ and if $W \subseteq V$, then (in principle, at least) showing that W is a subspace of V necessitates verifying that W, with its notions of addition and scalar multiplication, satisfies properties (1)-(9) of the definition of vector space over $\mathbb{K}$. Of course, the fact that the notions of addition and scalar multiplication in W are exactly the same as those in V should mean that most of the properties of addition and scalar multiplication in W (like associativity, commutativity, etc.) should follow immediately from those in V. As one might expect, it turns out that we can simplify this verification down to a few crucial points. This observation gives us the following theorem.

Theorem 21 (Subspace Criteria)**.** Let V be a vector space over $\mathbb{K}$, and let $W \subseteq V$. Equip W with the notions of addition and scalar multiplication from V. Then W is a subspace of V if, and only if, the following criteria hold:

(i) $0_V \in W$, where 0_V is the additive identity of V.

(ii) W is **closed under vector addition**, in the sense that for every $v, w \in W$, $v + w \in W$.

(iii) W is **closed under scalar multiplication**, in the sense that for every $v \in W$ and $a \in \mathbb{K}$, $av \in W$.

Moreover, in this case the additive identity 0_V of V is also the additive identity of W.

Proof. Suppose that W is a subspace of V. Criteria (ii) and (iii) follow immediately from property (1) of vector spaces. Because W is a vector space, W has an identity element 0_W. By property (1) of vector spaces, $00_W = 0_V \in W$. Therefore (i) holds. Moreover, because $0_V \in W$, $0_W = 0_W + 0_V = 0_V$.

Now assume that (i),(ii),(iii) hold. We verify that W satisfies the properties of a vector space. Throughout, let $v, w, u \in W$ and $a, b \in \mathbb{K}$. Because $v, w, u \in W$ and $W \subseteq V$, $v, w, u \in V$ as well.

(1) This follows immediately from criteria (ii) and (iii).

(2) Because addition in V is associative, $v + (u + w) = (v + u) + w$.

(3) Because addition in V is commutative, $v + u = u + v$.

(4) By (i), $0_V \in W$. Because $v + 0_V = v$, 0_V is also an additive identity of W. Hence, take $0_W = 0_V$.

(5) By (iii), $-v = (-1)v \in W$. Moreover, $v + (-v) = 0_V = 0_W$.

(6) Because scalar multiplcation in V is associative, $a(bv) = (ab)v$.

(7) Because scalar multiplcation in V distributes over scalar addition, $(a + b)v = av + bv$.

(8) Because scalar multiplication in V distributes over vector addition, $a(v + u) = av + au$.

(9) Because of the scalar multiplicative identity property of V, $1v = v$.

This completes the proof that W, with the notions of addition and scalar multiplcation inherited from V, is a vector space over $\mathbb{K}$. ☐

With this theorem in hand, producing interesting example of vector spaces is a breeze! Let's revisit the examples we listed earlier.

Proposition 21 (Spans are Subspaces)**.** Let V be a vector space over $\mathbb{K}$, and let $v_1, \ldots, v_m \in V$. Then $W \stackrel{def}{=} \operatorname{span}(v_1, \ldots, v_m)$ is a subspace of V.

Proof. Note that $0 = 0v_1 + \cdots + 0v_m \in W$. Let $u, w \in W$ and $a \in \mathbb{K}$. Choose $c_1, \ldots, c_m, d_1, \ldots, d_m \in \mathbb{K}$ with

$$u = c_1 v_1 + \cdots + c_m v_m \quad \text{and} \quad w = d_1 v_1 + \cdots + d_m v_m.$$

Then

$$u + v = (c_1 + d_1)v_1 + \cdots + (c_m + d_m)v_m \in W \quad \text{and} \quad au = (ac_1)v_1 + \cdots + (ac_m)v_m \in W.$$

By the Subspace Criteria, W is a subspace of V. □

Lecture 21: More Subspaces

Learning Objectives:

- Generate several examples of interesting vector spaces that can be viewed as subspaces of other vector spaces.

Example 56. Let V be a vector space over $\mathbb{K}$, and let 0 denote the additive identity in V. Then $\{0\}$ is a subspace of V (called the **trivial subspace**).

Proof. Because $\{0\} = \text{span}(0) \subseteq V$, the Spans are Subspaces proposition implies that $\{0\}$ is a subspace of V. $\square$

Example 57. For $n \in \mathbb{N} \cup \{0\}$, the set $P_n(\mathbb{K})$ of polynomials over $\mathbb{K}$ with degree no more n, that is

$$P_n(\mathbb{K}) \overset{def}{=} \{f : \mathbb{K} \to \mathbb{K} \ : \ \exists a_0, a_1, \ldots, a_n \in \mathbb{K} \text{ such that } \forall x \in \mathbb{K},\ f(x) = a_0 + a_1 x + \cdots + a_n x^n\},$$

is a vector space over $\mathbb{K}$ with the usual notions of addition and scalar multiplcation of functions.

Proof. For each $k = 0, \ldots, n$, let $p_k : \mathbb{K} \to \mathbb{K}$ be the polynomial $p_k(x) = x^k$ (where $p_0(x) \equiv 1$). Then $P_n(\mathbb{K}) = \text{span}(p_0, p_1, \ldots, p_n) \subset F(\mathbb{K}, \mathbb{K})$, so by the Spans are Subspaces Proposition, $P_n(\mathbb{K})$ is a subspace of $F(\mathbb{K}, \mathbb{K})$. $\square$

The past couple examples get some milage out of the fact that spans of finite sets of vectors are subspaces. For more general examples where the subsets are defined by some other property, we may need to get our hands dirty.

Example 58. The set $P(\mathbb{K})$ of polynomials over $\mathbb{K}$,

$$P(\mathbb{K}) \overset{def}{=} \{f : \mathbb{K} \to \mathbb{K} \ : \ \exists n \in \mathbb{N} \text{ and } \exists a_0, a_1, \ldots, a_n \in \mathbb{K} \text{ such that } \forall x \in \mathbb{K},\ f(x) = a_0 + a_1 x + \cdots + a_n x^n\},$$

is a vector space over $\mathbb{K}$ with the usual notions of addition and scalar multiplcation of functions.

Proof. Note that $P(\mathbb{K}) \subseteq F(\mathbb{K}, \mathbb{K})$. Moreover, note that $P_n(\mathbb{K}) \subseteq P(\mathbb{K})$ for every $n \in \mathbb{N} \cup \{0\}$. Because $P_0(\mathbb{K}) \subseteq P(\mathbb{K})$ and $0 \in P_0(\mathbb{K})$, $0 \in P(\mathbb{K})$. Now suppose $f, g \in P(\mathbb{K})$ and $a \in \mathbb{K}$. Then there exist $n, m \in \mathbb{N} \cup \{0\}$ with $f \in P_n(\mathbb{K})$ and $g \in P_m(\mathbb{K})$. Then $f + g \in P_{\max(n,m)}(\mathbb{K}) \subseteq P(\mathbb{K})$ and $af \in P_n(\mathbb{K}) \subseteq P(\mathbb{K})$. By the Subspace Criteria, $P(\mathbb{K})$ is a subspace of $F(\mathbb{K}, \mathbb{K})$. $\square$

Example 59. Let $I \subseteq \mathbb{R}$ be an interval. Then the space

$$C^0(I, \mathbb{R}) \overset{def}{=} \{f : I \to \mathbb{R} \ : \ f \text{ is continuous}\}$$

of continuous, real-valued functions on I is a subspace of $F(I, \mathbb{R})$.

Proof. Note that because the zero function $0(x) \equiv 0$ is continuous, $0 \in C^0(I, \mathbb{R})$. Now suppose that $f, g \in C^0(I, \mathbb{R})$ and $a \in \mathbb{K}$. Then since $f + g : I \to \mathbb{R}$ and $af : I \to \mathbb{R}$ are continuous[19], $f + g, af \in C^0(I, \mathbb{R})$. By the Subspace Criteria, $C^0(I, \mathbb{R})$ is a subspace of $F(I, \mathbb{R})$. $\square$

Besides investigating specific examples, we might also prove additional abstract facts about subspaces.

Proposition 22. Let V be a vector space over $\mathbb{K}$, and let W be a subspace of V. Let $w_1, \ldots, w_n \in W$. Then $w_1 \ldots, w_n$ is a linearly independent set in W if, and only if, $w_1, \ldots, w_n$ is a linearly independent set in V.

Proof. Suppose that $w_1, \ldots, w_n$ is linearly independent in W. Suppose $c_1, \ldots, c_n \in \mathbb{K}$ satisfy $c_1 w_1 + \cdots + c_n w_n = 0_V$. Then $c_1 w_1 + \cdots + c_n v_n = 0_W$ (since $0_W = 0_V$), and therefore $c_1 = \cdots = c_n = 0$. Therefore $w_1, \ldots, w_n$ is linearly independent in V. The proof of the reverse direction is almost identical. $\square$

Proposition 23. Let V be a vector space over $\mathbb{K}$, and let W and U be subspaces of V. Then the intersection[20] $W \cap U$ is a subspace of V.

Proof. Because W and U are subspaces, $0 \in W$ and $0 \in U$. Therefore $0 \in W \cap U$. Now suppose $v, z \in W \cap U$ and $a \in \mathbb{K}$. Because $v, z \in W \cap U$, $v, z \in W$ and $v, z \in U$. Because W and U are subspaces, $v + z, av \in W$ and $v + z, av \in U$. Therefore $v + z, az \in W \cap U$. By the Subspace Criteria, $W \cap U$ is a subspace of V. $\square$

Remark 62. A natural question to follow the previous example is whether the *union*[21] of two subspaces of a vector space V must form a subspace of V.

The answer to this question is "no". For a counterexample, consider the subspaces W, U of $\mathbb{K}^2$ given by $W = \text{span}(\vec{e}_1)$ and $U = \text{span}(\vec{e}_2)$. Then $W \cup U = \{\vec{x} \in \mathbb{K}^2 \ : \ \exists a \in \mathbb{K} \text{ such that } \vec{x} = a\vec{e}_1 \text{ or } \vec{x} = a\vec{e}_2\}$. But then $\vec{e}_1, \vec{e}_2 \in W \cup U$, but $\vec{e}_1 + \vec{e}_2 \notin W \cup U$. To see this note that if $a \in \mathbb{K}$ satisfied $\vec{e}_1 + \vec{e}_2 = a\vec{e}_1$, then $(1 - a)\vec{e}_1 + \vec{e}_2 = \vec{0}$ and therefore (because $\vec{e}_1, \vec{e}_2$ is a linearly independent set) $1 = 0$ (which we know to be false). Similarly, we cannot write $\vec{e}_1 + \vec{e}_2 = a\vec{e}_2$ for some $a \in \mathbb{K}$.

On your homework, you will explore the exact conditions under which the union of two subspaces of a vector space is a subspace.

Two particular examples of subspaces, investigated (without names) earlier in the quarter, are worth revisiting now.

---

[19] This was proved in your calculus course. Let $x_0 \in I$. Then because f and g are continuous at x_0, we have $\lim_{x \to x_0} (f + g)(x) = \lim_{x \to x_0} (f(x) + g(x)) = f(x_0) + g(x_0) = (f + g)(x_0)$, so $f + g$ is continuous at x_0. Similarly, $\lim_{x \to x_0} (af)(x) = \lim_{x \to x_0} af(x) = af(x_0) = (af)(x_0)$, so that af is continuous at x_0. Therefore $f + g$ and af are continuous on I.

[20] If A and B are sets, the **intersection** of A and B, denoted $A \cap B$, is $A \cap B = \{x \ : \ x \in A \text{ and } x \in B\}$. That is, $A \cap B$ is the set of elements that are common to both A and B.

[21] If A and B are sets, the **union** of A and B, denotes $A \cup B$, is $A \cup B = \{x \ : \ x \in A \text{ or } x \in B\}$. That is, $A \cup B$ is the set of elements that are in either A or B (or both).

Definition 35. Let $A \in M_{m\times n}(\mathbb{K})$, and let $\vec{a}_1, \dots, \vec{a}_n \in \mathbb{K}^m$ be the columns of A. The **column space** of A, denoted $\operatorname{col}(A)$, is defined to be

$$\operatorname{col}(A) \overset{def}{=} \operatorname{span}(\vec{a}_1, \dots, \vec{a}_n).$$

The **nullspace** of A, denoted $\operatorname{null}(A)$, is defined to be

$$\operatorname{null}(A) \overset{def}{=} \{\vec{x} \in \mathbb{K}^n \ : \ A\vec{x} = \vec{0}\}.$$

Proposition 24. Let $A \in M_{m\times n}(\mathbb{K})$

(a) $\operatorname{col}(A)$ is a subspace of $\mathbb{K}^m$, and $\operatorname{col}(A) = \{A\vec{x} \in \mathbb{K}^m \ : \ \vec{x} \in \mathbb{K}^n\}$.

(b) $\operatorname{null}(A)$ is a subspace of $\mathbb{K}^n$.

Proof. Because $\operatorname{col}(A)$ is the span of n vectors in $\mathbb{K}^n$, $\operatorname{col}(A)$ is a subspace of $\mathbb{K}^n$. The equivalence of $\operatorname{span}(\vec{a}_1, \dots, \vec{a}_n)$ and $\{A\vec{x} \in \mathbb{K}^m \ : \ \vec{x} \in \mathbb{K}^n\}$ is immediate from the definition of multiplication of matrices and vectors, since $A\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n$ for each $x_1, \dots, x_n \in \mathbb{K}$. This proves (a).

(b) follows immediately from Exercise 8 on Homework 2. However, we'll prove it again here (with far better notation!). Note that since $A\vec{0} = \vec{0}$, $\vec{0} \in \operatorname{null}(A)$. If $\vec{x}, \vec{y} \in \operatorname{null}(A)$ and $a \in \mathbb{K}$, then

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} = \vec{0} + \vec{0} = \vec{0} \quad \text{and} \quad A(a\vec{x}) = a(A\vec{x}) = a\vec{0} = \vec{0},$$

so that $\vec{x} + \vec{y}, a\vec{x} \in \operatorname{null}(A)$. By the Subspace Criteria, $\operatorname{null}(A)$ is a subspace of $\mathbb{K}^n$. □

Remark 63. At this point, it might be good to pause and note that we can now add a couple more items to the Invertibility Theorem. In particular, we have the following result.

Theorem 22 (Invertibility Theorem, cont'd)**.** Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{n\times n}(\mathbb{K})$ and let $T : \mathbb{K}^n \to \mathbb{K}^n$ be the linear transformation $T(\vec{x}) = A\vec{x}$. The following statements are equivalent.

(a-1) A is invertible.

$\vdots$

(a-7) There is $B \in M_{n\times n}(\mathbb{K})$ such that $\operatorname{rref}(\begin{bmatrix} A & I_n \end{bmatrix}) = \begin{bmatrix} I_n & B \end{bmatrix}$.

(a-8) A^T is invertible.

$\vdots$

(b-7) $\operatorname{col}(A) = \mathbb{K}^n$

$\vdots$

(c-8) $\operatorname{null}(A) = \{\vec{0}\}$

Moreover, in (a-7), (b-4), and (c-5) we have $B = A^{-1}$.

Proof. The equivalence of (a-1) with (a-7) is the Computing Inverses Theorem (which also shows that $B = A^{-1}$). The equivalence of (a-1) with (a-8) is Exercise 5(d) from Homework 6. The equivalence of (a-1) and (b-7) follows immediately from the definition $\text{col}(A) = \text{span}(\vec{a}_1, \dots, \vec{a}_n)$ and the equivalence of (a-1) with (b-1). The equivalence of (a-1) and (c-8) follows immediately from the equivalence of (a-1) and (c-2), since $\text{null}(A)$ is the solution set of $A\vec{x} = \vec{0}$.

The fact that a right or left inverse of A must be A^{-1} is Exercise 2(a) on Homework 6. □

Lecture 22: Bases and Dimension

Learning Objectives:

- Determine whether a vector space is finite-dimensional or infinite-dimensional.
- Determine whether a set of vectors is a basis for a finite-dimensional vector space.
- Examine the properties of bases in terms of representing vectors in the space using linear combinations.

The rest of the course (and much of the next quarter) will be devoted to studying the geometric properties of vector spaces. Because we have already done so much work studying $\mathbb{K}^n$, it will be important to distinguish the vector spaces whose geometric structure most closely resembles that of $\mathbb{K}^n$ (for some n). To this end, we make the following definition.

Definition 36. Let V be a vector space over $\mathbb{K}$. Then we say that V is **finite-dimensional** if there is a finite set of vectors $v_1, \dots, v_n \in V$ such that $\operatorname{span}(v_1, \dots, v_n) = V$. We say that V is **infinite-dimensional** if V is not finite-dimensional.

Example 60. $\mathbb{K}^n$ is finite-dimensional, since $\mathbb{K}^n = \operatorname{span}(\vec{e}_1, \dots, \vec{e}_n)$.

Example 61. $M_{m\times n}(\mathbb{K})$ is finite-dimensional. For each $1 \le j \le m$ and $1 \le k \le n$, let $A_{j,k} \in M_{m\times n}(\mathbb{K})$ be the matrix with entry 1 in the j,k-th spot, and all other entries equal to 0. Then for each $A = [a_{j,k}] \in M_{m\times n}(\mathbb{K})$, we have

$$A = a_{1,1}A_{1,1} + \cdots + a_{1,n}A_{1,n} + \cdots + a_{m,1}A_{m,1} + \cdots + a_{m,n}A_{m,n},$$

so that $A \in \operatorname{span}(A_{1,1}, \dots, A_{1,n}, \dots, A_{m,1}, \dots, A_{m,n})$. Because the span of any collection of vectors in $M_{m\times n}(\mathbb{K})$ is a subset of $M_{m\times n}(\mathbb{K})$, we therefore have $M_{m\times n}(\mathbb{K}) = \operatorname{span}(A_{1,1}, \dots, A_{1,n}, \dots, A_{m,1}, \dots, A_{m,n})$.

Example 62. Because $P_n(\mathbb{K}) = \operatorname{span}(1, x, x^2, \dots, x^n)$, $P_n(\mathbb{K})$ is finite-dimensional.

Example 63. If V is a vector space over $\mathbb{K}$, then the trivial subspace $W = \{0_V\}$ is finite-dimensional, as $W = \operatorname{span}(0_V)$.

Example 64. The subspace $V = \left\{ A \in M_{2\times 2}(\mathbb{R}) \ : \ \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} A = A \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \right\}$ of $M_{2\times 2}(\mathbb{R})$ is finite-dimensional.

First note that V is indeed a subspace of $M_{2\times 2}(\mathbb{R})$ by the Subspace Criteria, since $0_{2\times 2} \in V$ and if $A, B \in V$ and $c \in \mathbb{R}$, then

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}(A+B) = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} A + \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} B = A \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} + B \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = (A+B) \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

and
$$\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}(cA) = c\Big(\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}A\Big) = c\Big(A\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}\Big) = (cA)\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}.$$

Note that if $A = \begin{bmatrix}a & b\\ c & d\end{bmatrix} \in V$ then
$$\begin{bmatrix}a+2c & b+2d\\ 2a+c & 2b+d\end{bmatrix} = \begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}\begin{bmatrix}a & b\\ c & d\end{bmatrix} = \begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix} = \begin{bmatrix}a+2b & 2a+b\\ c+2d & 2c+d\end{bmatrix},$$
so that $a+2c = a+2b$, so $b = c$. Moreover, $2a+c = c+2d$, so $a = d$. Therefore $A = \begin{bmatrix}a & b\\ b & a\end{bmatrix}$ for some $a, b \in \mathbb{R}$. On the other hand, if $A = \begin{bmatrix}a & b\\ b & a\end{bmatrix}$ for some $a, b \in \mathbb{R}$, then
$$\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}A = \begin{bmatrix}a+2b & 2a+b\\ 2a+b & a+2b\end{bmatrix} = A\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}.$$
Therefore $V = \Big\{\begin{bmatrix}a & b\\ b & a\end{bmatrix} \;:\; a, b \in \mathbb{R}\Big\}$. Let $A \in V$. Then $A = \begin{bmatrix}a & b\\ b & a\end{bmatrix}$ for some $a, b \in \mathbb{R}$. Moreover,
$$A = a\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + b\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}.$$
Because linear combinations of $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$ are also in V, we therefore have $V = \operatorname{span}\Big(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\Big)$.

Remark 64. Not every vector space is finite dimensional. Indeed, $P(\mathbb{K})$, $\mathbb{K}^\infty$, and $C^0(\mathbb{R}, \mathbb{R})$ are each infinite-dimensional (and there are many other examples!). We do not yet have the machinery to prove that these spaces are infinite-dimensional, but we will soon.

Bases

It is usually convenient to study the geometric structure of a vector space in terms of a fixed set of vectors that exactly captures the structure of the space. Such a set is called a *basis* for the space.

Definition 37. Let V be a finite-dimensional vector space over $\mathbb{K}$. Let $v_1, \dots, v_n \in V$. Then the set $v_1, \dots, v_n$ is a **basis** for V if $v_1, \dots, v_n$ is linearly independent and if $\operatorname{span}(v_1, \dots, v_n) = V$.

Example 65. Because $\mathbb{K}^n = \operatorname{span}(\vec{e}_1, \dots, \vec{e}_n)$ and $\vec{e}_1, \dots, \vec{e}_n$ is a linearly independent set, $\vec{e}_1, \dots, \vec{e}_n$ is a basis for $\mathbb{K}^n$. We call this the **standard basis** for $\mathbb{K}^n$.

Example 66. Because $P_n(\mathbb{K}) = \operatorname{span}(1, x, x^2, \dots, x^n)$ and $1, x, x^2, \dots, x^n$ is a linearly independent set[22], $1, x, x^2, \dots, x^n$ is a basis for $P_n(\mathbb{K})$.

[22] Technically you proved this result only when $\mathbb{K} = \mathbb{R}$, but the result is true (with the same proof!) when $\mathbb{K} = \mathbb{C}$. Alternatively, we can prove that the polynomials $1, z, z^2, \dots, z^n$ are also linearly independent in $P(\mathbb{C})$ by noting that if $a_0, \dots, a_n \in \mathbb{C}$ and we write $a_j = r_j + it_j$, $r_j, t_j \in \mathbb{R}$ for each $0 \le j \le n$, then if $0 = a_0 + a_1 z + \cdots + a_n z^n$ for each $z \in \mathbb{C}$, then taking $z = x + i0$ for $x \in \mathbb{R}$ implies that $0 = (r_0 + r_1 x + \cdots + r_n x^n) + i(t_0 + t_1 x + \cdots + t_n x^n)$ for every $x \in \mathbb{R}$. Because the real and imaginary parts of a complex number are unique, we see that $0 = r_0 + r_1 x + \cdots + r_n x^n$ and $0 = t_0 + t_1 x + \cdots + t_n x^n$ for every $x \in \mathbb{R}$. Therefore (by the result you proved in discussion) $r_0 = \cdots = r_n = 0$ and $t_0 = \cdots = t_n = 0$. Therefore $a_0 = \cdots = a_n = 0$.

Example 67. In the notation of Example 61, because $M_{m\times n}(\mathbb{K}) = \operatorname{span}(A_{1,1}, \ldots, A_{1,n}, \ldots, A_{m,1}, \ldots, A_{m,n})$, and if $c_{1,1}, \ldots, c_{1,n}, \ldots, c_{m,1}, \ldots, c_{m,n} \in \mathbb{K}$ satisfy

$$0_{m\times n} = c_{1,1}A_{1,1} + \cdots + c_{1,n}A_{1,n} + \cdots + c_{m,1}A_{m,1} + \cdots + c_{m,n}A_{m,n} = \begin{bmatrix} c_{j,k} \end{bmatrix}$$

then $c_{j,k} = 0$ for each j, k, we see that $A_{1,1}, \ldots, A_{1,n}, \ldots, A_{m,1}, \ldots, A_{m,n}$ is a basis for $M_{m\times n}(\mathbb{K})$.

To capture the idea that bases describe the vector space V in the most efficient way possible, we prove a (now almost trivial) theorem that will be absolutely crucial to the observations we will use throughout the rest of the linear algebra portion of MATH 291.

Theorem 23. Let V be a finite-dimensional vector space over $\mathbb{K}$. Let $v_1, \ldots, v_n \in V$.

(a) $v_1, \ldots, v_n$ is linearly independent if, and only if, for every $b \in V$ there is at most one choice of scalars $c_1, \ldots, c_n \in \mathbb{K}$ with $b = c_1v_1 + \cdots + c_nv_n$.

(b) $\operatorname{span}(v_1, \ldots, v_n) = V$ if, and only if, for each $b \in V$ there is at least one choice of scalars $c_1, \ldots, c_n \in \mathbb{K}$ with $b = c_1v_1 + \cdots + c_nv_n$.

(c) $v_1, \ldots, v_n$ is a basis for V if, and only if, for each $b \in V$ there is exactly one choice of scalars $c_1, \ldots, c_n \in \mathbb{K}$ with $b = c_1v_1 + \cdots + c_nv_n$.

Proof. For (a), suppose that $v_1, \ldots, v_n$ is linearly independent. Let $b \in V$. If $b \notin \operatorname{span}(v_1, \ldots, v_n)$, then there is no choice of scalars $c_1, \ldots, c_n$ with $b = c_1v_1 + \cdots + c_nv_n$. If $b \in \operatorname{span}(v_1, \ldots, v_n)$, then a past result implies that there is a unique (and therefore at most one) choice of scalars $c_1, \ldots, c_n$ with $b = c_1v_1 + \cdots + c_nv_n$. Conversely, suppose that for each $b \in V$ there is at most one choice of scalars $c_1, \ldots, c_n$ with $b = c_1v_1 + \cdots + c_nv_n$. Taking $b = 0$, and noting that $0 = 0v_1 + \cdots + 0v_n$, we have that if $0 = c_1v_1 + \cdots + c_nv_n$ then $c_1 = \cdots = c_n = 0$, so that $v_1, \ldots, v_n$ is a linearly independent set.

Part (b) follows immediately from the definition of span, and part (c) follows immediately from (a), (b), and the definition of basis. □

Remark 65. For infinite-dimensional vector spaces, the definition of basis must be slightly more general. If V is a (not necessarily finite-dimensional) vector space over $\mathbb{K}$, and if $\mathfrak{B} \subseteq V$, then we say that $\mathfrak{B}$ is a **basis** for V if

(i) for every $n \in \mathbb{N}$, every set $v_1, \ldots, v_n$ of n elements of $\mathfrak{B}$ is linearly independent, and

(ii) for every $b \in V$, there is $m \in \mathbb{N}$, $w_1, \ldots, w_m \in \mathfrak{B}$, and $c_1, \ldots, c_m \in \mathbb{K}$ with $b = c_1w_1 + \cdots + c_mw_m$.

Part (i) captures the idea that the (possibly infinite) set $\mathfrak{B}$ is linearly independent, and part (ii) captures the idea that the (possibly infinite) set $\mathfrak{B}$ spans V. Note that in (ii), the exact choice of m and the vectors $w_1, \ldots, w_m$ (as well as the scalars $c_1, \ldots, c_m$) might depend on b.

We will not study bases in infinite-dimensional vector spaces this year, so this definition is provided merely for fun. I'll add more information about how our results for finite-dimensional vectors spaces generalize to the infinite-dimensional case when it can be done without going too far into the weeds.

Lecture 23: More Bases and Dimension

Learning Objectives:

- Determine the relationship between the sizes of linearly independent sets and spanning sets.
- Construct basis for finite-dimensional vector spaces.
- Show that a vector space is infinite-dimensional.

Example 68. Let V be a vector space over $\mathbb{K}$. As a matter of convention, we say that the empty set of vectors is a basis for the trivial subspace $\{0_V\}$ of V. This is because of the fact that the empty set of vectors is (vacuously) linearly independent, and because of the convention that $\text{span}() \stackrel{def}{=} \{0_V\}$.

Existence of Bases

We turn to two questions that may have flown under the radar until now: when does a vector space V actually *have* a basis? And if V has a basis, then how can we actually *produce* a basis for V? The answers to both of these questions (for finite-dimensional vector spaces) boils down to an important result that we'll call the Independence vs. Span Theorem, which compares the sizes of linearly independent sets and spanning sets.

Remark 66. To set the stage, we discuss a special case of the Independence vs. Span Theorem that we already understand. Recall that in $\mathbb{K}^n$, a linearly independent set $\vec{v}_1, \ldots, \vec{v}_m$ cannot have more than n vectors (because, by the Linear Independence and Linear Systems Theorem, the reduced row-echelon form of the $n \times m$ matrix $\begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{bmatrix}$ must have a pivot in every column), so that $m \leq n$. We also saw that if $\text{span}(\vec{w}_1, \ldots, \vec{w}_\ell) = \mathbb{K}^n$, then (by the Spanning Set in $\mathbb{K}^m$ Theorem, since the reduced row-echelon form of the $n \times \ell$ matrix $\begin{bmatrix} \vec{w}_1 & \cdots & \vec{w}_\ell \end{bmatrix}$ must have a pivot in every row) $n \leq \ell$. In particular, $m \leq \ell$, or rather that every linearly independent set in $\mathbb{K}^n$ must necessarily have no more vectors than any spanning set in $\mathbb{K}^n$. Intuitively, this can be explained using the idea (which we will make rigorous soon) that $\mathbb{K}^n$ is "n dimensional". Here, it is impossible to have more than n "independent directions" in $\mathbb{K}^n$, and yet we need at least n "directions" to span $\mathbb{K}^n$. The Independence vs. Span Theorem generalizes this idea (that linearly independent sets must no larger than spanning sets) to all finite-dimensional vector spaces.

Theorem 24 (Independence vs. Span)**.** Let V be a vector space over $\mathbb{K}$. Let $v_1, \ldots, v_m \in V$ and $w_1, \ldots, w_\ell \in V$. If $v_1, \ldots, v_m$ is a linearly independent set and $\text{span}(w_1, \ldots, w_\ell) = V$, then $m \leq \ell$.

Proof. Because $w_1, \ldots, w_\ell$ span V, for each $k = 1, \ldots, m$ there are scalars $u_{1,k}, \ldots, u_{\ell,k} \in \mathbb{K}$ such that

$$v_k = u_{1,k} w_1 + \cdots + u_{\ell,k} w_\ell.$$

Let $U \in M_{\ell \times m}(\mathbb{K})$ be $U = \begin{bmatrix} u_{j,k} \end{bmatrix}$. We claim that the columns of U are linearly independent. To see this, suppose $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \in \mathbb{K}^m$ satisfies $U\vec{x} = \vec{0}$. Then for each $1 \leq j \leq \ell$, $\underbrace{u_{j,1}x_1 + \cdots + u_{j,m}x_m}_{j-th \text{ entry of } U\vec{x}} = 0$.

Therefore

$$\begin{aligned}
0_V &= 0w_1 + \cdots + 0w_\ell \\
&= (u_{1,1}x_1 + \cdots + u_{1,m}x_m)w_1 + \cdots + (u_{\ell,1}x_1 + \cdots + u_{\ell,m}x_m)w_\ell \\
&= x_1(u_{1,1}w_1 + \cdots + u_{\ell,1}w_\ell) + \cdots + x_m(u_{1,m}w_1 + \cdots + u_{\ell,m}w_\ell) \\
&= x_1v_1 + \cdots + x_mv_m.
\end{aligned}$$

Because $v_1, \ldots, v_m$ is a linearly independent set, $x_1 = \cdots = x_m = 0$. Therefore the columns of U are linearly independent. By the Linear Independence and Linear Systems Theorem, rref(U) has a pivot in every column. Therefore $U \in M_{\ell \times m}(\mathbb{K})$ has at least as many rows as it has columns, so $m \leq \ell$. $\square$

We can immediate apply the Independence vs. Span Theorem to prove that our examples of infinite-dimensional vector spaces are indeed infinite dimensional. The argument for each is almost the same.

Example 69. $P(\mathbb{K})$ is infinite-dimensional, for if there were $p_1, \ldots, p_\ell \in P(\mathbb{K})$ with $\text{span}(p_1, \ldots, p_\ell) = P(\mathbb{K})$, then every linearly independent set in $P(\mathbb{K})$ would have no more than ℓ elements. But you showed on your discussion worksheet that $1, x, \ldots, x^\ell$ is a linearly independent set of $\ell + 1$ polynomials in $P(\mathbb{K})$, and $\ell + 1 > \ell$. Therefore no finite set of polynomials spans $P(\mathbb{K})$, so that $P(\mathbb{K})$ is infinite-dimensional.

Example 70. Because $P(\mathbb{R})$ is a subspace of $C^0(\mathbb{R}, \mathbb{R})$, and because there are linearly independent sets of every (finite) size in $P(\mathbb{R})$, there are linearly independents sets of every (finite) size in $C^0(\mathbb{R}, \mathbb{R})$. Therefore the same arguments as above show that $C^0(\mathbb{R}, \mathbb{R})$ is infinite-dimensional.

Example 71. $\mathbb{K}^\infty$ is infinite-dimensional. To see this, for each $n \in \mathbb{N}$ let $E_n = (\delta_{j,n})_{j=1}^\infty$, where $\delta_{n,n} = 1$ and $\delta_{j,n} = 0$ if $j \neq n$. That is, E_n is the sequence with entry 1 in the n-th spot, and 0 for every other entry. Suppose that there were $S_1, \ldots, S_\ell \in \mathbb{K}^\infty$ with $\text{span}(S_1, \ldots, S_\ell) = \mathbb{K}^\infty$. Then every linearly indpendent set in $\mathbb{K}^\infty$ has at most ℓ elements. But $E_1, \ldots, E_{\ell+1}$ is linearly indpenedent, for if $c_1, \ldots, c_{\ell+1} \in \mathbb{K}$ satisfies

$$0_{\mathbb{K}^\infty} = c_1E_1 + \cdots + c_{\ell+1}E_{\ell+1} = (c_1, c_2, \ldots, c_{\ell+1}, 0, 0, \ldots),$$

then $c_1 = \cdots = c_{\ell+1} = 0$. Therefore $\mathbb{K}^\infty$ is not spanned by a finite number of vectors.

Example 72. The set $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is a basis for $\mathbb{K}^3$. To see why, note that since

$$\text{rref} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \text{rref} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -4 & -1 \\ 0 & -2 & 0 \end{bmatrix} = \text{rref} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & -2 & 0 \end{bmatrix} = \text{rref} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

the Invertibility Theorem implies that $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ both spans $\mathbb{K}^3$ and is linearly independent.

Remark 67. The last example suggests that we should add another line to the Invertibility Theorem. Let's do that now.

Theorem (Invertibility Theorem, cont'd)**.** Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{n\times n}(\mathbb{K})$ and let $T : \mathbb{K}^n \to \mathbb{K}^n$ be the linear transformation $T(\vec{x}) = A\vec{x}$. The following statements are equivalent.

(a-1) A is invertible.

$\vdots$

(a-9) $\vec{a}_1, \ldots, \vec{a}_n$ is a basis for $\mathbb{K}^n$.

$\vdots$

Proof. Suppose that A is invertible. By our first version of the Invertibility Theorem, $\vec{a}_1, \ldots, \vec{a}_n$ is a linearly independent and $\operatorname{span}(\vec{a}_1, \ldots, \vec{a}_n) = \mathbb{K}^n$. Therefore $\vec{a}_1, \ldots, \vec{a}_n$ is a basis for $\mathbb{K}^n$.

Suppose that $\vec{a}_1, \ldots, \vec{a}_n$ is a basis for $\mathbb{K}^n$. Then $\vec{a}_1, \ldots, \vec{a}_n$ is a linearly independent set. ☐

Perhaps the most important application of the Independence vs. Span Theorem is that it allows us to explicitly describe how to construct bases.

Theorem 25 (Constructing Bases)**.** Let V be a finite-dimensional vector space over $\mathbb{K}$.

(a) Every spanning set for V contains a basis for V, in the sense that if $v_1, \ldots, v_n \in V$ and $\operatorname{span}(v_1, \ldots, v_n) = V$, then there are $1 \leq k_1 < \cdots < k_m \leq n$ such that $v_{k_1}, \ldots, v_{k_m}$ is a basis for V.

Here we allow $m = 0$ in the case where the empty set of vectors is a basis for V.

(b) Every linearly independent set in V can be extended to a basis for V, in the sense that if $v_1, \ldots, v_n \in V$ and $v_1, \ldots, v_n$ is a linearly independent set, then there are $w_1, \ldots, w_m \in V$ such that $v_1, \ldots, v_n, w_1, \ldots, w_m$ is a basis for V.

Here we allow the cases where $n = 0$ (where the linearly independent set we start with is the empty set) or $m = 0$ (where $v_1, \ldots, v_n$ already spans V).

In particular, if V is finite-dimensional then V has a basis.

Remark 68. The proof of this result should be very intuitive, but writing the proof to emphasize the intuition results in an overly technical argument. Therefore, to supplement the proof, we will also give an informal argument that captures the the intuition of the rigorous proof (without the complication).

"Proof" (Enlightening, Not Rigorous). For (a), suppose that $v_1, \ldots, v_n \in V$ and $\operatorname{span}(v_1, \ldots, v_n) = V$. If $v_1, \ldots, v_n$ is linearly independent, then we are done. If not, then one of the vectors (say v_j) is a linear combination of the others. Therefore by a past result we have

$$V = \operatorname{span}(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_n) = \operatorname{span}(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n).$$

We therefore get a set $v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n$ of $n - 1$ vectors that still span V. We repeat this argument, at each stage removing a vector that is a linear combination of the others in the set, until we arrive at a linearly independent set (which could very well be the empty set). Eventually we have a (perhaps empty) subcollection $v_{k_1}, \dots, v_{k_m}$ of $v_1, \dots, v_n$ that is linearly independent and spans V.

For (b), because V is finite-dimensional, there is $N \in \mathbb{N}$ and a set of N vectors in V that spans V. Suppose $v_1, \dots, v_n$ is a linearly independent set in V. If $\text{span}(v_1, \dots, v_n) = V$, then we are done (this is the case where $m = 0$). If not, then there is $w_1 \in V$ with $w_1 \notin \text{span}(v_1, \dots, v_n)$. By a past result, $v_1, \dots, v_n, w_1$ is a linearly independent set of $n+1$ vectors. We repeat this argument as many times as possible. Because a linearly independent set in V cannot have more than N vectors (due to the Independence vs. Span Theorem), for some m with $n + m \leq N$ we will have a linearly independent list $v_1, \dots, v_n, w_1, \dots, w_m$ which (by our inability to repeat the above argument) also spans V. □

Proof (Rigorous, Not Enlightening). We start with (a). Suppose $v_1, \dots, v_n \in V$ and $\text{span}(v_1, \dots, v_n) = V$. If $V = \{0_V\}$, then the empty subset of $v_1, \dots, v_n$ is a linearly independent set that spans V. Suppose that V is not trivial. Then at least one of the v_j is nonzero, so $v_1, \dots, v_n$ contains a linearly independent set consisting of one vector. Define

$$S = \{j \;:\; 1 \leq j \leq n \text{ and there are } 1 \leq k_1 < \cdots < k_j \leq n \text{ with } v_{k_1}, \dots, v_{k_m} \text{ linearly indpendent}\}.$$

Then $1 \in S$ and $S \subseteq \{1, \dots, n\}$. Let m be the maximum[23] number in S. Then there is $1 \leq k_1 < \cdots < k_m \leq n$ with $v_{k_1}, \dots, v_{k_m}$. Note that $\text{span}(v_{k_1}, \dots, v_{k_m}) \subseteq V$. If there were j with $v_j \notin \text{span}(v_{k_1}, \dots, v_{k_m})$, then $v_{k_1}, \dots, v_{k_m}, v_j$ would be a linearly independent subset of $v_1, \dots, v_n$ with $m + 1$ vectors which contradicts the definition of m. Therefore $v_j \in \text{span}(v_{k_1}, \dots, v_{k_m})$ for each j, so that $V = \text{span}(v_1, \dots, v_n) \subseteq \text{span}(v_{k_1}, \dots, v_{k_m})$. Therefore $v_{k_1}, \dots, v_{k_m}$ is a linearly independent set with $V = \text{span}(v_{k_1}, \dots, v_{k_m})$, so that $v_{k_1}, \dots, v_{k_m}$ is a basis for V.

We now turn to (b). Because V is finite-dimensional, there is $N \in \mathbb{N}$ such that some set of N vectors in V spans V. Let $v_1, \dots, v_n \in V$ with $v_1, \dots, v_n$ linearly independent. (If we are starting with the empty set, then $n = 0$.) By the Independence vs. Span Theorem, $n \leq N$. Let

$$S = \{k \;:\; 0 \leq k \leq N - n \text{ and there are } w_1, \dots, w_k \in V \text{ with } v_1, \dots, v_n, w_1, \dots, w_k \text{ linearly independent}\}.$$

Because $v_1, \dots, v_n$ is linearly independent, $0 \in S$. Let m be the maximum number in S. Let $w_1, \dots, w_m \in V$ with $v_1, \dots, v_n, w_1, \dots, w_m$ linearly independent. Then we have $\text{span}(v_1, \dots, v_n, w_1, \dots, w_m) \subseteq V$. Let $b \in V$. If $b \notin \text{span}(v_1, \dots, v_n, w_1, \dots, w_m)$, then a past lemma implies that $v_1, \dots, v_n, w_1, \dots, w_m, b$ is a linearly indpendent set, contradicting the choice of m. Therefore $b \in \text{span}(v_1, \dots, v_n, w_1, \dots, w_M)$. This shows that $\text{span}(v_1, \dots, v_n, w_1, \dots, w_m) = V$, so that $v_1, \dots, v_n, w_1, \dots, w_m$ is a basis for V. □

Remark 69. The Constructing Bases Theorem only applies to finite-dimensional vector spaces, and therefore you might still wonder whether an infinite-dimensional vector space V must have a basis (and, of course, how to find such a basis if one exists). It is possible to prove that every infinite-dimensional vector space has a basis (called a **Hamel Basis**) using the logical sledgehammer known as Zorn's Lemma. Because the proof gives no blueprint to how to actually produce a basis (in the way that the Constructing Bases Theorem does), there is no insight that allows us to actually find such a basis.

---

[23]

> **Lemma 3.** For every $n \in \mathbb{N}$, if $S \subseteq \{1, \dots, n\}$ is nonempty, then S has a maximum.

Proof. We proceed by induction. The only nonempty $S \subseteq \{1\}$ is $S = \{1\}$, and the maximum of $\{1\}$ is 1. Let $n \in \mathbb{N}$ and suppose that the result holds for subsets of $\{1, \dots, n\}$. Let $S \subseteq \{1, \dots, n, n+1\}$ be nonempty. If $n + 1 \in S$, then since $k \leq n + 1$ for every $k \in S$, $n + 1$ is the maximum of S. If $n + 1 \notin S$, then $S \subseteq \{1, \dots, n\}$ and the induction hypothesis implies that S has a maximum. By the Principle of Mathematical Induction, the proof is complete. □

Invertibility Theorem, revisited.

Theorem (Invertibility Theorem, cont'd)**.** Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{n\times n}(\mathbb{K})$ and let $T : \mathbb{K}^n \to \mathbb{K}^n$ be the linear transformation $T(\vec{x}) = A\vec{x}$. The following statements are equivalent.

(a-1) A is invertible.

$\vdots$

(a-7) There is $B \in M_{n\times n}(\mathbb{K})$ such that $\mathrm{rref}(\begin{bmatrix} A & I_n \end{bmatrix}) = \begin{bmatrix} I_n & B \end{bmatrix}$.

(a-8) A^T is invertible.

(a-9) $\vec{a}_1, \ldots, \vec{a}_n$ is a basis for $\mathbb{K}^n$.

$\vdots$

(b-4) There is $B \in M_{n\times n}(\mathbb{K})$ with $AB = I_n$.

$\vdots$

(b-7) $\mathrm{col}(A) = \mathbb{K}^n$

$\vdots$

(c-5) There is $B \in M_{n\times n}(\mathbb{K})$ with $BA = I_n$.

$\vdots$

(c-8) $\mathrm{null}(A) = \{\vec{0}\}$

Moreover, in (a-7), (b-4), and (c-5) we have $B = A^{-1}$.

Lecture 24: Even More Bases and Dimension

Learning Objectives:

- Define the notion of dimension of a finite-dimensional vector spaces in terms of the number of vectors in a basis.
- Compute the dimension of several standard finite-dimensional vector spaces.
- Show that the dimension of a finite-dimensional vector space V behaves similarly to the notion of dimension in $\mathbb{K}^n$.
- Establish results about dimension that confirm our intuition about what dimension should measure.

Another easy consequence of the Independence vs. Span Theorem is the observation that, for a finite-dimensional vector space V, there is only one possiblity for the number of vectors that appear in a basis for V.

Proposition 25. Let V be a finite-dimensional vector space. Suppose that $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ are bases for V. Then $m = n$.

Proof. Because $v_1, \ldots, v_n$ is linearly independent and $\text{span}(w_1, \ldots, w_m) = V$, the Independence vs. Span Theorem implies that $n \leq m$. Because $w_1, \ldots, w_m$ is linearly independent and $\text{span}(v_1, \ldots, v_n) = V$, the Independence vs. Span Theorem implies that $m \leq n$. Therefore $m = n$. $\square$

Because the number of vectors in a basis for a finite-dimensional vector space V captures the number of "independent direction" needed to describe V, the number of vectors in a basis for V should capture what we intuitively think of as the "dimension" of V. We therefore make the following definition.

Definition 38. Let V be a finite-dimensional vector space. Let n be the number of vectors in a basis for V. We call n the **dimension** of V, and write $\dim(V) = n$.

Example 73. Because $\vec{e}_1, \ldots, \vec{e}_n$ is a basis for $\mathbb{K}^n$, $\dim(\mathbb{K}^n) = n$.

Example 74. Because $1, x, x^2, \ldots, x^n$ is a basis for $P_n(\mathbb{K})$, $\dim(P_n(\mathbb{K})) = n + 1$.

Example 75. Because $A_{1,1}, \ldots, A_{1,n}, \ldots, A_{m,1}, \ldots, A_{m,n}$ is a basis for $M_{m\times n}(\mathbb{K})$ (where $A_{j,k} \in M_{m\times n}(\mathbb{K})$ has entry 1 in the j, k-th spot, and all other entries 0), $\dim(M_{m\times n}(\mathbb{K})) = mn$.

Example 76. Let V be a vector space over $\mathbb{K}$. Because the empty set is a basis for the trivial subspace of V, $\dim(\{0_V\}) = 0$.

Example 77. Because $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is a basis for

$$V = \left\{ A \in M_{2\times 2}(\mathbb{R}) \;:\; \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} A = A \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \right\},$$

we conclude that $\dim(V) = 2$.

Example 78. Because $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} i \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ i \end{bmatrix}$ is a basis for $\mathbb{C}^2$ (considered as a vector space over $\mathbb{R}$), we have that $\dim_{\mathbb{R}}(\mathbb{C}^2) = 4$ (here the subscript $\dim_{\mathbb{R}}$ indicates that we are computing the dimension of $\mathbb{C}^2$ as a vector space over $\mathbb{R}$, whereas $\dim_{\mathbb{C}}(\mathbb{C}^2) = 2$).

Example 79. Let $a, b, c \in \mathbb{R}$ with at least one of a, b, c nonzero. On Exercise 2 of Homework 2, you showed that the plane P through the origin of $\mathbb{R}^3$ given by $ax + by + cz = 0$ is spanned by a linearly independent set of two vectors $\vec{v}_1, \vec{v}_2$. Therefore, as a subspace of $\mathbb{R}^3$, $\dim(P) = 2$. This agrees with our notion that a plane is a "two-dimensional object".

Example 80. Let $\vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$. Because the line L through $\vec{0}$ that is parallel to $\vec{v}$ is given by $L = \operatorname{span}(\vec{v})$, and because the set $\vec{v}$ is linearly indpendent (because $\vec{v} \neq \vec{0}$), we see that, as a subspace of $\mathbb{R}^n$, $\dim(L) = 1$. This agrees with our intuition that a line is a "one-dimensional object".

The fact that the dimension of a finite-dimensional vector space is well-defined allows us to recover at least a portion of the Invertibility Theorem (of course, without any reference to invertibility) for finite-dimensional vector spaces.

Theorem 26. Let V be a finite-dimensional vector space over $\mathbb{K}$, and let $n = \dim(V)$. Let $v_1, \ldots, v_n \in V$. Then the following are equivalent.

(a) $v_1, \ldots, v_n$ is linearly independent.

(b) $\operatorname{span}(v_1, \ldots, v_n) = V$.

(c) $v_1, \ldots, v_n$ is a basis for V.

Proof. (a)⇒(c). Suppose $v_1, \ldots, v_n$ is linearly independent. By the Constructing Bases Theorem, $v_1, \ldots, v_n$ extends to a basis for V. But since $\dim(V) = n$, this extended set must also consist of n vectors, and is therefore $v_1, \ldots, v_n$ itself.

(b)⇒(c). Suppose $\operatorname{span}(v_1, \ldots, v_n) = V$. By the Constructing Bases Theorem, $v_1, \ldots, v_n$ contains a basis for V. But since $\dim(V) = n$, this basis must also consist of n vectors, and is therefore $v_1, \ldots, v_n$ itself.

(c)⇒(a),(b). This is immediate from the definition of basis. □

The Constructing Bases Theorem (and its proof) are the keys to proving all sorts of intuitive results about dimensionality. Here are a couple results.

Theorem 27 (Characterization of Infinite-Dimensionality)**.** Let V be a vector space over $\mathbb{K}$. Then V is infinite-dimensional if, and only if, for every $n \in \mathbb{N}$ there are vectors $v_1, \ldots, v_n \in V$ with $v_1, \ldots, v_n$ linearly indpendent.

Proof. ($\Rightarrow$) Assume that V is infinite-dimensional. We proceed by induction. Because the trivial vector space is finite-dimensional, V contains a nonzero vector v_1. Because v_1 is nonzero, the set v_1 is linearly independent. Now let $n \in \mathbb{N}$ and assume that there are $v_1, \ldots, v_n \in V$ with $v_1, \ldots, v_n$ linearly independent. Because V is infinite-dimensional, there exists $v_{n+1} \in V$ with $v_{n+1} \notin \operatorname{span}(v_1, \ldots, v_n)$. Therefore $v_1, \ldots, v_n, v_{n+1}$ is linearly independent. By the Principle of Mathematical Induction, the result is proved.

($\Leftarrow$) Assume that for every $n \in \mathbb{N}$ there are vectors $v_1, \ldots, v_n \in V$ with $v_1, \ldots, v_n$ linearly independent. Suppose to the contrary that V were finite-dimensional. Then there is $N \in \mathbb{N}$ and a set of N vectors in V that spans V. But by hypothesis there is a linearly independent set of $N+1$ vectors in V. By the Independence vs. Span Theorem, we obtain the absurd result that $N + 1 \leq N$. $\square$

Remark 70. Note that the "$\Leftarrow$" part of the proof of the Characterization of Infinite-Dimensionality was exactly the argument that we gave when we showed that $P(\mathbb{K})$, $\mathbb{K}^\infty$, $C^0(\mathbb{R}, \mathbb{R})$, etc. were infinite dimensional!

The next result confirms our intuition that a finite-dimensional vector space cannot have an infinite-dimensional subspace, and that any proper subspace of a finite-dimensional vector space must have smaller dimension than the space itself.

Theorem 28 (Dimension and Subspaces)**.** Let V be a finite-dimensional vector space over $\mathbb{K}$, and let W be a subspace of V. Then W is finite-dimensional and $\dim(W) \leq \dim(V)$. Moreover, $\dim(W) = \dim(V)$ if, and only if, $W = V$.

Proof. Let $n = \dim(V)$. Let $m \in \mathbb{N}$, let $w_1, \ldots, w_m \in W$, and suppose that $w_1, \ldots, w_m$ is linearly independent (in W). Then $w_1, \ldots, w_m$ is also linearly independent in V, and therefore (by the Constructing Bases Theorem) extends to a basis $w_1, \ldots, w_m, v_1, \ldots, v_\ell$ for V. Since $m + \ell = n$, we must have $m \leq n$. Therefore W is finite-dimensional, and if we take $m = \dim(W)$ then we have $\dim(W) \leq \dim(V)$. If $m = n$ then there is a basis $w_1, \ldots, w_n$ for W that extends (by the Constructing Bases Theorem) to a basis for V. But since a basis for V must consist of n vectors, $w_1, \ldots, w_n$ was already a basis for V, so that $W = \operatorname{span}(w_1, \ldots, w_n) = V$. Conversely, if $W = V$ then every basis for W is also a basis for V, so that $\dim(W) = \dim(V)$. $\square$

Lecture 25: General Linear Transformations

Learning Objectives:

- Determine when a map from one vector space to another is linear.
- Establish the basic properties of linear transformations that generalize from earlier results.
- Characterize the existence and uniqueness of a linear transformation with finite-dimensional domain in terms of its values on a basis.

Now that we have generalized the notion of vector space from $\mathbb{K}^n$ to general vector spaces over $\mathbb{K}$, our next task is to generalize the notion of linear transformations between vector spaces. Just as with vector spaces, we will see that almost everything about linear transformation from $\mathbb{K}^n$ to $\mathbb{K}^m$ will generalize (in a suitable way) to linear transformations between two vector spaces over $\mathbb{K}$ (especially then the vector spaces involved are finite-dimensional).

Definition 39. Let V and W be vector spaces over $\mathbb{K}$, and let $T : V \to W$. We call T **linear** (or a **linear transformation**) if the following two conditions are satisfied:

(i) For every $u, v \in V$, $T(u + v) = T(u) + T(v)$.

(ii) For every $c \in \mathbb{K}$ and every $v \in V$, $T(cv) = cT(v)$.

Remark 71. Note that here we require both V and W to be vector spaces over the same field $\mathbb{K}$, as this is the only way that property (ii) makes sense.

Remark 72. Note that the operations of addition and scalar multiplication that appear are those in V and W, respectively. We mustn't become confused and attempt, for example, to add an element of V and an element of W. If we wanted to be very pedantic here, for (i) and (ii) we should write something along the lines of $T(u \oplus_V v) = T(u) \oplus_W T(v)$ and $T(c \odot_V v) = c \odot_W T(v)$, where $\oplus_V, \odot_V$ are the notions of vector addition and scalar multiplication in V, and $\oplus_W, \odot_W$ are the notions of vector addtiion and scalar multiplication in W. As an act of mercy, let's avoid this level of pendantry unless it become absolutely necessary in order to avoid confusion.

Example 81. The map $S : M_{m\times n}(\mathbb{K}) \to M_{n\times m}(\mathbb{K})$, $S(A) \stackrel{def}{=} A^T$ is a linear transformation, since for every $A, B \in M_{m\times n}(\mathbb{K})$ and $c \in \mathbb{K}$,

$$S(A + B) = (A + B)^T = A^T + B^T = S(A) + S(B) \quad \text{and} \quad S(cA) = (cA)^T = cA^T = cS(A).$$

Here is an example that illustrates why it is useful (sometimes, at least) to be a little pedantic with our notation.

Example 82. Note that $\mathbb{R}$ is a vector space over $\mathbb{R}$ with the usual notions of addition and (scalar) multiplication. Let $\mathbb{R}^+ = (0, +\infty)$ be the collection of positive real numbers with the notions of addition $\oplus$ and scalar multiplication $\oplus$ given by

$$x \oplus y \stackrel{def}{=} xy \quad \text{and} \quad c \odot x \stackrel{def}{=} x^c.$$

On your homework you showed that $\mathbb{R}^+$ (with these operations) is a vector space over $\mathbb{R}$ (with additive identity $0_{\mathbb{R}^+} = 1$, and where the additive inverse of $x \in \mathbb{R}^+$ is $\frac{1}{x}$).

The transformation $\exp : \mathbb{R} \to \mathbb{R}^+$, $\exp(x) = e^x$ is linear, since for each $x, y \in \mathbb{R}$ and $c \in \mathbb{R}$,

$$\exp(x+y) = e^{x+y} = e^x e^y = \exp(x) \oplus \exp(y) \quad \text{and} \quad \exp(cx) = e^{cx} = (e^x)^c = c \odot e^x = c \odot \exp(x).$$

Example 83. The map $D : P(\mathbb{R}) \to P(\mathbb{R})$, $D(p(x)) \overset{def}{=} p'(x)$ is linear, since for each $p, q \in P(\mathbb{R})$ and $c \in \mathbb{R}$ we have

$$D(p(x) + q(x)) = (p(x) + q(x))' = p'(x) + q'(x) \quad \text{and} \quad D(cp(x)) = (cp(x))' = cp'(x) \quad \text{for every } x \in \mathbb{R}.$$

In fact, differentiation can be viewed as a linear map between various vector spaces (with the same proof!):

$$D : P_n(\mathbb{R}) \to P_n(\mathbb{R}), \quad D : P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R}), \quad D : C^1(\mathbb{R}, \mathbb{R}) \to C^0(\mathbb{R}, \mathbb{R}).$$

The only thing to check here (other than that differentiation is linear, which has an identical proof as the one above) is that the derivative of a polynomial of degree no more than n is a polynomial of degree no more than $n-1$, and that the derivative of a differentiable function with continuous derivative is a continuous function.

Example 84. Consider $\mathbb{C}^n$ as a vector space over $\mathbb{R}$, and define $T : \mathbb{R}^{2n} \to \mathbb{C}^n$ by

$$T\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{bmatrix}\right) \overset{def}{=} \begin{bmatrix} x_1 + iy_1 \\ \vdots \\ x_n + iy_n \end{bmatrix}.$$

Then T is linear, since for each $\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{bmatrix}, \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^{2n}$ and $c \in \mathbb{R}$,

$$T\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{bmatrix}\right) = T\left(\begin{bmatrix} x_1 + a_1 \\ \vdots \\ x_n + a_n \\ y_1 + b_1 \\ \vdots \\ y_n + b_n \end{bmatrix}\right) = \begin{bmatrix} x_1 + a_1 + i(y_1 + b_1) \\ \vdots \\ x_n + a_n + i(y_n + b_n) \end{bmatrix} = \begin{bmatrix} x_1 + iy_1 \\ \vdots \\ x_n + iy_n \end{bmatrix} + \begin{bmatrix} a_1 + ib_1 \\ \vdots \\ a_n + ib_n \end{bmatrix}$$

$$= T\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{bmatrix}\right) + T\left(\begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{bmatrix}\right)$$

and

$$T\left(c\begin{bmatrix}x_1\\ \vdots\\ x_n\\ y_1\\ \vdots\\ y_n\end{bmatrix}\right) = T\left(\begin{bmatrix}cx_1\\ \vdots\\ cx_n\\ cy_1\\ \vdots\\ cy_n\end{bmatrix}\right) = \begin{bmatrix}cx_1+icy_1\\ \vdots\\ cx_n+icy_n\end{bmatrix} = c\begin{bmatrix}x_1+iy_1\\ \vdots\\ x_n+iy_n\end{bmatrix} = cT\left(\begin{bmatrix}x_1\\ \vdots\\ x_n\\ y_1\\ \vdots\\ y_n\end{bmatrix}\right).$$

Example 85. Here is an example that will be extremely important next quarter. Fix $\vec{a} = \begin{bmatrix}a_1\\ a_2\end{bmatrix} \in \mathbb{K}^2$, and define $T : \mathbb{K}^2 \to \mathbb{K}$ by

$$T\left(\begin{bmatrix}x_1\\ x_2\end{bmatrix}\right) \overset{def}{=} x_1a_2 - x_2a_1.$$

The significance of this map is that $T(\vec{x})$ is the determinant of the 2×2 matrix $\begin{bmatrix}x_1 & a_1\\ x_2 & a_2\end{bmatrix}$, and therefore (as you showed on your homework) $T(\vec{x}) = 0$ if, and only if, $\begin{bmatrix}x_1 & a_1\\ x_2 & a_2\end{bmatrix}$ is not invertible.

The map T is linear, since for $\vec{x}, \vec{y} \in \mathbb{K}^2$ and $c \in \mathbb{K}$,

$$T(\vec{x}+\vec{y}) = T\left(\begin{bmatrix}x_1+y_1\\ x_2+y_2\end{bmatrix}\right) = (x_1+y_1)a_2 - (x_2+y_2)a_1 = (x_1a_2 - x_2a_1) + (y_1a_2 - y_2a_1) = T(\vec{x}) + T(\vec{y})$$

and

$$T(c\vec{x}) = T\left(\begin{bmatrix}cx_1\\ cx_2\end{bmatrix}\right) = (cx_1)a_2 - (cx_2)a_2 = c(x_1a_2 - x_2a_1) = cT(\vec{x}).$$

Properties of Linear Transformations

Many of the properties of linear transformations $T : V \to W$ are already familiar to us from the special case where $V = \mathbb{K}^n$ and $W = \mathbb{K}^m$. Here are a couple such results.

Proposition 26. Let V and W be vector spaces over $\mathbb{K}$, and let $T : V \to W$ be linear. Then $T(0_V) = 0_W$.

Proof. By linearity, we have $T(0_V) = T(00_V) = 0T(0_V) = 0_W$. □

Proposition 27. Let V and W be vector spaces over $\mathbb{K}$, and let $T : V \to W$ be a transformation. Then the following are equivalent.

(a) T is linear.

(b) For every $n \in \mathbb{N}$, $v_1, \ldots, v_n \in V$, and $c_1, \ldots, c_n \in \mathbb{K}$,

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

(c) For every $u, v \in V$ and $c \in \mathbb{K}$, $T(cu + v) = cT(u) + T(v)$.

Proof. We prove that $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

Suppose (a) holds. We proceed by induction. If $v_1 \in V$ and $c_1 \in \mathbb{K}$, then by linearity we have $T(c_1v_1) = c_1T(v_1)$. Now let $n \in \mathbb{N}$ and suppose the result of (b) holds. Let $c_1, \ldots, c_{n+1} \in \mathbb{K}$ and $v_1, \ldots, v_n, v_{n+1} \in V$. Then linearity and the induction hypothesis imply that

$$\begin{aligned} T(c_1v_1 + \cdots + c_nv_n + c_{n+1}v_{n+1}) &= T((c_1v_1 + \cdots + c_nv_n) + c_{n+1}v_{n+1}) \\ &= T(c_1v_1 + \cdots + c_nv_n) + T(c_{n+1}v_{n+1}) \\ &= c_1T(v_1) + \cdots + c_nT(v_n) + c_{n+1}T(v_{n+1}). \end{aligned}$$

By the Principle of Mahtematical Induction, (b) holds.

Suppose (b) holds. Let $u, v \in V$ and $c \in \mathbb{K}$. Then $T(cu + v) = T(cu + 1v) = cT(u) + 1T(v) = cT(u) + T(v)$. Therefore (c) holds.

Suppose (c) holds. Note first that since $T(0_V) = T(10_V + 0_V) = 1T(0_V) + T(0_V) = T(0_V) + T(0_V)$, so adding $-T(0_V)$ to both sides yields $0_W = T(0_V)$. Let $u, v \in V$ and $c \in \mathbb{K}$. Then we have

$$T(u + v) = T(1u + v) = 1T(u) + T(v) = T(u) + T(v)$$

and

$$T(cu) = T(cu + 0_V) = cT(u) + T(0_V) = cT(u) + 0_W = cT(u).$$

Therefore (a) holds. □

Recall that the standard matrix A of a linear transformation $T : \mathbb{K}^n \to \mathbb{K}^m$ is completely determined by the values $T(\vec{e}_1), \ldots, T(\vec{e}_n)$, since these end up being the columns of A. As a generalization of this, we prove that a linear transformation (defined on a finite-dimensional vector space) is completely determined by its values on a basis for the space (and, moreover, that we can always get a linear transformation with specified outputs associated to the input basis).

Theorem 29 (Constructing Linear Maps)**.** Let V and W be vector spaces over $\mathbb{K}$, assume V is finite-dimensional, let $v_1, \ldots, v_n \in V$, and assume that $v_1, \ldots, v_n$ is a basis for V.

Then for every $w_1, \ldots, w_n \in W$, there is a unique linear transformation $T : V \to W$ such that

$$T(v_1) = w_1, \quad T(v_2) = w_2, \quad \ldots, \quad T(v_n) = w_n.$$

Proof. Define $T : V \to W$ as follows. Let $x \in V$. Because $v_1, \ldots, v_n$ is a basis, there are unique $c_1, \ldots, c_n \in \mathbb{K}$ with $x = c_1v_1 + \cdots + c_nv_n$. Then take $T(x) \overset{def}{=} x_1w_1 + \cdots + x_nw_n$. For a given $x \in V$, the existence of the scalars $c_1, \ldots, c_n$ shows that T is defined on V. Because these scalars are unique (and therefore that there is no ambiguity about how $T(x)$ is defined), T is well-defined.

For linearity, let $x, y \in V$ and $\lambda \in \mathbb{K}$. Let $c_1, \ldots, c_n, d_1, \ldots, d_n \in \mathbb{K}$ be the unique scalars such that $x = c_1v_1 + \cdots + c_nv_n$ and $y = d_1v_1 + \cdots + d_nv_n$. Then

$$\begin{aligned} T(x + y) &= T((c_1 + d_1)v_1 + \cdots + (c_n + d_n)v_n) \\ &= (c_1 + d_1)w_1 + \cdots + (c_n + d_n)w_n \\ &= (c_1w_1 + \cdots + c_nw_n) + (d_1w_1 + \cdots + d_nw_n) \\ &= T(x) + T(y) \end{aligned}$$

and

$$T(\lambda x) = T((\lambda c_1)v_1 + \cdots + (\lambda c_n)v_n) = (\lambda c_1)w_1 + \cdots + (\lambda c_n)w_n = \lambda(c_1w_1 + \cdots + c_nw_n) = \lambda T(x).$$

Therefore T is linear.

For uniqueness, suppose that $S : V \to W$ is another linear transformation with $S(v_j) = w_j$ for $j = 1, \dots, n$. Let $x \in V$, and choose $c_1, \dots, c_n \in \mathbb{K}$ with $x = c_1 v_1 + \cdots + c_n v_n$. Then

$$T(x) = c_1 w_1 + \cdots + c_n w_n = S(c_1 v_1 + \cdots + c_n v_n) = S(x).$$

Therefore $S = T$ as functions from V to W. $\square$

Example 86. Recall Example 38, wherein we were asked to determine whether there exists a linear transformation $T : \mathbb{K}^3 \to \mathbb{K}^3$ such that

$$T\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right) = \begin{bmatrix}8\\0\\0\end{bmatrix}, \qquad T\left(\begin{bmatrix}2\\-2\\1\end{bmatrix}\right) = \begin{bmatrix}8\\0\\2\end{bmatrix}, \quad T\left(\begin{bmatrix}3\\-1\\-1\end{bmatrix}\right) = \begin{bmatrix}4\\0\\2\end{bmatrix}.$$

We have now solved this problem in two ways: first in Example 38, and second (using invertibility) in Example 42. Given the machinery we have at our disposal now, answering this question is very easy. First note that because

$$\operatorname{rref}\begin{bmatrix}1 & 2 & 3\\1 & -2 & -1\\0 & 1 & -1\end{bmatrix} = \operatorname{rref}\begin{bmatrix}1 & 2 & 3\\0 & -4 & -4\\0 & 1 & -1\end{bmatrix} = \operatorname{rref}\begin{bmatrix}1 & 0 & 1\\0 & 1 & 1\\0 & 0 & -2\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix} = I_3,$$

the Invertibility Theorem implies that $\left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}2\\-2\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}3\\-1\\-1\end{smallmatrix}\right]$ is a basis for $\mathbb{K}^3$. By the Invertibility Theorem, there is a (unique!) linear transformation $T : \mathbb{K}^3 \to \mathbb{K}^3$ with

$$T\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right) = \begin{bmatrix}8\\0\\0\end{bmatrix}, \qquad T\left(\begin{bmatrix}2\\-2\\1\end{bmatrix}\right) = \begin{bmatrix}8\\0\\2\end{bmatrix}, \quad T\left(\begin{bmatrix}3\\-1\\-1\end{bmatrix}\right) = \begin{bmatrix}4\\0\\2\end{bmatrix}.$$

Of course, while this approach allows us to answer the qualitative question of whether such a linear transformation exists, it does not actually give us the (more refined) information about how to compute the outputs of such a transformation, other than that the value of $T(\vec{x})$ is completely determined by linearity and the fact that $\vec{x}$ can be written uniquely as a linear combination of $\left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}2\\-2\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}3\\-1\\-1\end{smallmatrix}\right]$.

Lecture 26: Images and Kernels

Learning Objectives:

- Investigate the notions of injectivity, surjectivity, bijectivity, and invertibility for linear transformations.
- Establish basic properties of the image and kernel of a linear transformation.

In addition to the properties we summarized last time, we note that the definitions of injective, surjective, and bijective remain the same when we consider general linear transformations.

Definition 40. Let V and W be vector spaces over $\mathbb{K}$, and suppose that $T : V \to W$ is a linear transformation.

- We call T **injective** (or **one-to-one**) if for every $u, v \in V$, if $T(u) = T(v)$ then $u = v$.
- We call T **surjective** (or **onto**) if for every $w \in W$ there is $v \in V$ with $T(v) = w$.
- We call T **bijective** if T is both injective and surjective.

The definition of invertibility also generalizes in the expected way.

Definition 41. Let V and W be vector spaces over $\mathbb{K}$, and let $T : V \to W$ be a linear transformation. We call T **invertible** if there is a linear transformation $S : W \to V$ with $S(T(v)) = v$ and $T(S(w)) = w$ for every $v \in V$ and $w \in W$. We sometimes call an invertible linear transformation an **(linear) isomorphism between V and W**.

We'll say more about why we are introducing the term *isomorphism* in a couple days. The argument that you gave on Exercise 5 of Homework 5 shows that the inverse function of a bijective linear transformation is linear. When combined with the argument we gave in Proposition 17 (i.e. that the inverse of an invertible linear transformaton is unique), we immediately have the following result.

Theorem 30. Let V and W be vector spaces over $\mathbb{K}$, and let $T : V \to W$ be a linear transformation. Then T is bijective if, and only if, T is invertible.
Moreover, inverse function T^{-1} of T is the unique function satisfying $T^{-1}(T(v)) = v$ and $T(T^{-1}(w)) = w$ for every $v \in V$ and $w \in W$.

Images and Kernels

Recall that for $A \in M_{m\times n}(\mathbb{K})$ we had two natural subspaces associated with A: the column space $\mathrm{col}(A) \subseteq \mathbb{K}^m$, and the nullspace $\mathrm{null}(A) \subseteq \mathbb{K}^n$. In terms of linear transformations, we have the following.

Definition 42. Let V and W be vector spaces over $\mathbb{K}$, and assume that $T : V \to W$ is a linear transformation. Define the **image** of T, denoted image(T), to be the collection of outputs of T:

$$\text{image}(T) \stackrel{def}{=} \{T(v) \in W \ : \ v \in V\} = \{w \in W \ : \ \text{there is } v \in V \text{ with } w = T(v)\}.$$

Define the **kernel** of T, denotes ker(T), to be

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

Remark 73. Note that if $T : \mathbb{K}^n \to \mathbb{K}^m$ is a linear transformation with $T(\vec{x}) = A\vec{x}$ for every $\vec{x} \in \mathbb{K}^n$, then

$$\text{image}(T) = \{T(\vec{x}) \ : \ \vec{x} \in \mathbb{K}^n\} = \{A\vec{x} \ : \ \vec{x} \in \mathbb{K}^n\} = \text{col}(A)$$

and

$$\ker(T) = \{\vec{x} \in \mathbb{K}^n \ : \ T(\vec{x}) = \vec{0}\} = \{\vec{x} \in \mathbb{K}^n \ : \ A\vec{x} = \vec{0}\} = \text{null}(A),$$

so that the image and kernel of a linear transformation generalize the notions of column space and nullspace.

Just as for the column space and nullspace of a matrix, the image and kernel of a linear transformation are subspaces of the domain and codomain (respectively) of the transformation.

Proposition 28. Let V and W be vector spaces over $\mathbb{K}$, and let $T : V \to W$ be a linear transformation. Then the following hold.

(a) image(T) is a subspace of W,

(b) ker(T) is a subspace of V,

(c) If V is finite-dimensional then

 (c-i) ker(T) and image(T) are finite-dimensional, and

 (c-ii) if $V = \text{span}(v_1, \dots, v_n)$, then $\text{image}(T) = \text{span}(T(v_1), \dots, T(v_n))$.

Proof. We start with the image. Note that since $T(0_V) = 0_W$, $0_W \in \text{image}(T)$. Suppose $u, w \in \text{image}(T)$ and $c \in \mathbb{K}$. Then there are $v, z \in V$ with $T(v) = u$ and $T(z) = w$, so that

$$u + w = T(v) + T(z) = T(v + z) \in \text{image}(T) \quad \text{and} \quad cu = cT(v) = T(cv) \in \text{image}(T).$$

By the Subspace Criteria, image(T) is a subspace of W.

Now we show that ker(T) is a subspace of V. Note that since $T(0_V) = 0_W$, $0_V \in \ker(T)$. Suppose that $v, z \in \ker(T)$ and $c \in \mathbb{K}$. Then

$$T(v + z) = T(v) + T(z) = 0_W + 0_W = 0_W \quad \text{and} \quad T(cv) = cT(v) = c0_W = 0_w,$$

so that $v + z \in \ker(T)$ and $cv \in \ker(T)$. By the Subspace Criteria, ker(T) is a subspace of V.

Finally, assume that V is finite-dimensional. By the Subspaces and Dimension Theorem, ker(T) is finite-dimensional. Moreover, there are $v_1, \dots, v_n \in V$ with $V = \text{span}(v_1, \dots, v_n)$. Because $T(v_1), \dots, T(v_n) \in \text{image}(T)$ and image(T) is a subspace of W, $\text{span}(T(v_1), \dots, T(v_n)) \subseteq \text{image}(T)$. On the other hand, if

$w \in \text{image}(T)$ then there is $v \in V$ with $T(v) = w$. Write $v = c_1v_1 + \cdots + c_nv_n$ for some $c_1, \ldots, c_n \in \mathbb{K}$. Then $w = T(c_1v_1 + \cdots + c_nv_n) = c_1T(v_1) + \cdots + c_nT(v_n) \in \text{span}(T(v_1), \ldots, T(v_n))$. Therefore $\text{image}(T) = \text{span}(T(v_1), \ldots, T(v_n))$ and $\text{image}(T)$ is finite-dimensional. □

As an application, we have our most general version of a result you first proved in Exercise 9 of Homework 2.

Theorem 31 (Solution Sets of Linear Equations)**.** Let V and W be vector spaces, let $T : V \to W$ be a linear transformation, and let $w_0 \in W$. If there is $v_0 \in V$ such that $T(v_0) = w$, then the set of solutions of $T(v) = w_0$ is exactly

$$v_0 + \ker(T) \stackrel{def}{=} \{v_0 + v_h \ : \ v_h \in \ker(T)\}.$$

Proof. Suppose first that $v \in \{v_0 + v_h \ : \ v_h \in \ker(T)\}$. Choose $v_h \in \ker(T)$ with $v = v_0 + v_h$. Then $T(v) = T(v_0 + v_h) = T(v_0) + T(v_h) = w_0 + 0_W = w_0$.

Conversely, suppose that $v \in V$ and $T(v) = w_0$. Then $v = v_0 + (v - v_0)$, and, since $T(v - v_0) = T(v) - T(v_0) = w_0 - w_0 = 0_W$, $v - v_0 \in \ker(T)$. Therefore $v \in \{v_0 + v_h \ : \ v_h \in \ker(T)\}$, and the proof is complete. □

The Solution Sets of Linear Equations Theorem has wide-ranging applications. Here is one from analysis (where we can solve a *linear partial differential equation*).

Example 87. Find all twice-differentiable functions $f : \mathbb{R} \to \mathbb{R}$ that solve the differential equation $f''(x) + f(x) = 2x^2 - x$.

Solution. Note first that if f is a solution to the above equation, then since f is differentiable (and therefore continuous) we have that $f''(x) = 2x^2 - x - f(x)$ is also continuous, and therefore $f \in C^2(\mathbb{R}, \mathbb{R})$. Therefore it suffices to look for solutions in $C^2(\mathbb{R}, \mathbb{R})$. Consider the map $L : C^2(\mathbb{R}, \mathbb{R}) \to C^0(\mathbb{R}, \mathbb{R})$ given by $L(f(x)) = f''(x) + f(x)$. Then L is linear, for if $f, g \in C^2(\mathbb{R}, \mathbb{R})$ and $c \in \mathbb{R}$, then

$$\begin{aligned} L(f(x) + g(x)) &= (f(x) + g(x))'' + f(x) + g(x) \\ &= f''(x) + g''(x) + f(x) + g(x) \\ &= f''(x) + f(x) + g''(x) + g(x) \\ &= L(f(x)) + L(g(x)) \end{aligned}$$

and

$$L(cf(x)) = (cf(x))'' + cf(x) = cf''(x) + cf(x) = c(f''(x) + f(x)) = cL(f(x))$$

for every $x \in \mathbb{R}$. Note that we are trying to find solutions of the equation $T(f(x)) = 2x^2 - x$. We use whatever techniques we wish to first find a single solution f_0 of this equation. For one approach, let's guess (given the form of the right-hand side) that there is a solution f_0 of the form $f_0(x) = ax^2 + bx + c$. Then we must have

$$2x^2 - x = T(f_0(x)) = f_0''(x) + f_0(x) = 2a + ax^2 + bx + c, \quad \text{so that } 2 = a, \ -1 = b, \ 0 = 2a + c.$$

Therefore $a = 2$, $b = -1$, and $c = -2a = -4$, so that $f_0(x) = 2x^2 - x - 4$. (We quickly verify that $L(f_0(x)) = 4 + 2x^2 - x - 4 = 2x^2 - x$, so that the f_0 we produced really does solve the equation.) By the Solution Sets of Linear Equations Theorem, every solution of this differential equation has the form $f_0(x) + f_h(x)$, where $f_h \in \ker(L)$. But on your homework you showed that

$$\ker(L) = \{f : \mathbb{R} \to \mathbb{R} \ : \ f''(x) + f(x) = 0 \text{ for every } x \in \mathbb{R}\} = \text{span}(\cos(x), \sin(x)),$$

so that the solution set of the differential equation is

$$\{2x^2 - x - 4 + c_1 \cos(x) + c_2 \sin(x) \ : \ c_1, c_2 \in \mathbb{R}\}.$$

Lecture 27: More Images and Kernels

Learning Objectives:

- Use the kernel and image of a linear transformation to characterize whether the function is injective and/or surjective.
- Compute the kernel, image, nullity, and rank of a linear transformation.

One important use of images and kernels is that they characterize surjectivity and injectivity of linear transformations, much as, in light of the Linear Independence and Linear Systems and Spanning Sets in $\mathbb{K}^m$ Theorems, column space and nullspace did for (the linear transformations associated to) matrices.

Proposition 29. Let V and W be vector spaces over $\mathbb{K}$, and let $T : V \to W$ be linear.

(a) T is injective if, and only if, $\ker(T) = \{0_V\}$.

(b) T is surjective if, and only if, $\text{image}(T) = W$.

Proof. We start with (a). Suppose T is injective. Because $T(0_V) = 0_W$, $\{0_V\} \subseteq \ker(T)$. Suppose $v \in \ker(T)$. Since $T(v) = 0_W = T(0_V)$ and T is injective, $v = 0_V \in \{0_V\}$. Therefore $\ker(T) = \{0_V\}$. Conversely, suppose $\ker(T) = \{0_V\}$. Let $u, v \in V$ and suppose $T(u) = T(v)$. Then $0_W = T(u) - T(v) = T(u - v)$, so that $u - v \in \ker(T) = \{0_V\}$. Therefore $u - v = 0_V$, so that $u = v$. In other words, T is injective.

We now prove (b). Suppose T is surjective. The inclusion $\text{image}(T) \subseteq W$ is automatic from the deifnition of T. Suppose $w \in W$. Because T is surjective, there is $v \in V$ with $T(v) = w$. Therefore $w \in \text{image}(T)$, so that $\text{image}(T) = W$. Conversely suppose $\text{image}(T) = W$. Let $w \in W$. Because $w \in W = \text{image}(T)$, there is $v \in V$ with $T(v) = w$. Therefore T is surjective. □

In the finite-dimensional case, it turns out that there are some interesting things that we can say about the dimensions of the image and kernel of a linear transformation. To motivate this, consider the following observations (including some results that you will prove on your homework).

Remark 74. Let $T : \mathbb{K}^7 \to \mathbb{K}^4$, be a linear transformation with matrix $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_7 \end{bmatrix} \in M_{4\times 7}(\mathbb{K})$, and assume that

$$\text{rref}(A) = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 & -3 & 5 \\ 0 & 0 & 0 & 1 & 0 & 4 & -7 \\ 0 & 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then the pivot columns of A are $\vec{a}_2, \vec{a}_4, \vec{a}_5$ and the redundant columns of A are $\vec{a}_1, \vec{a}_3, \vec{a}_6, \vec{a}_7$. Indeed, from $\text{rref}(A)$ we see that we can write the redundant columns as linear combinations of the preceding pivot columns as

$$\vec{a}_1 = \vec{0}, \qquad \vec{a}_3 = 2\vec{a}_2, \qquad \vec{a}_6 = -3\vec{a}_2 + 4\vec{a}_4 + 3\vec{a}_5, \qquad \vec{a}_7 = 5\vec{a}_2 - 7\vec{a}_4 + 0\vec{a}_5,$$

or rather

$$\vec{0} = 1\vec{a}_1, \qquad \vec{0} = -2\vec{a}_2 + 1\vec{a}_3, \qquad \vec{0} = 3\vec{a}_2 - 4\vec{a}_4 - 3\vec{a}_5 + 1\vec{a}_6, \qquad \vec{0} = -5\vec{a}_2 + 7\vec{a}_4 + 0\vec{a}_5 + 1\vec{a}_7. \tag{5}$$

By Exercise 4 on Homework 4, you proved that the pivot columns of A form a linearly independent set. On your last homework assignment, you will show that the pivot columns of A actually form a basis for the column space of A (which is also the image of T). That is, you will prove that

$$\text{image}(T) = \text{col}(A) = \text{span}(\vec{a}_2, \vec{a}_4, \vec{a}_5)$$

(and that $\vec{a}_2, \vec{a}_4, \vec{a}_5$ is a linearly independent set), so that $\dim(\text{image}(T)) = 3$.

On the other hand, the equations in (5) (interpreted as linear combinations of the columns of A) tell us that

$$\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4 \overset{def}{=} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 0 \\ -4 \\ -3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -5 \\ 0 \\ 7 \\ 0 \\ 0 \\ 1 \end{bmatrix} \in \text{null}(A) = \ker(T),$$

so that (by the Spans are Subspaces Theorem)

$$\text{span}\left(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\right) \subseteq \text{null}(A) = \ker(T).$$

On the other hand, if $\vec{x} = \left[\begin{smallmatrix} x_1 \\ \vdots \\ x_7 \end{smallmatrix}\right] \in \text{null}(A)$, then from $\text{rref}(A)$ we see that we have $x_2 = -2x_3 + 3x_6 - 5x_7$, $x_4 = -4x_6 + 7x_7$, and $x_5 = -3x_6$, so that

$$\vec{x} = \begin{bmatrix} x_1 \\ -2x_3 + 3x_6 - 5x_7 \\ x_3 \\ -4x_6 + 7x_7 \\ -3x_6 \\ x_6 \\ x_7 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} 0 \\ 3 \\ 0 \\ -4 \\ -3 \\ 1 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 0 \\ -5 \\ 0 \\ 7 \\ 0 \\ 0 \\ 1 \end{bmatrix} \in \text{span}\left(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\right),$$

and therefore $\ker(T) = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4)$. On the other hand, note that if $c_1, c_2, c_3, c_4 \in \mathbb{K}$ and

$$\vec{0} = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + c_4\vec{v}_4 = \begin{bmatrix} c_1 \\ -2c_2 + 3c_3 - 5c_4 \\ c_2 \\ -4c_3 + 7c_4 \\ -3c_3 \\ c_3 \\ c_4 \end{bmatrix},$$

then $c_1 = c_2 = c_3 = c_4 = 0$. Therefore $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ is a basis for $\ker(T)$, so that $\dim(\ker(T)) = 4$.

Note that $\dim(\text{image}(T)) + \dim(\ker(T)) = 3 + 4 = 7$, which is the dimension of the domain of T (i.e. $\mathbb{K}^7$). This is no accident. Here, one could predict this from the fact that the pivot columns of A form a basis for $\text{image}(T)$, and there is a basis for $\ker(T)$ consisting of one vector for each redundant columns of A (where the vector encodes how to represent that redundant columns as a linear combination of the preceding pivot columns). Since A has 7 columns, and each column is either a pivot column or a redundant column, the sum of the number of pivot columns (rank(A), which is also dim(image(T))) and the number of redundant columns (dim(ker(T))) must be 7.

In light of the connection between the dimensions of ker(T) and image(T) and rank(A) and the dimension of null(A), we therefore give the following definitions.

Definition 43. Let V and W be vector spaces with V finite-dimensional, and assume that $T : V \to W$ is linear.

(i) The **rank** of T, denoted rank(T), is defined by

$$\text{rank}(T) \overset{def}{=} \dim(\text{image}(T)).$$

(ii) The **nullity** of T, denoted nullity(T), is defined by

$$\text{nullity}(T) \overset{def}{=} \dim(\ker(T)).$$

Remark 75. In the last example, we had $\text{rank}(T) = 3$ and $\text{nullity}(T) = 4$.

The Rank-Nullity Theorem

The observation in the previous example—that if $T : V \to W$ then nullity(T) measures the dimension of the subspace of V that is sent to 0_W or "killed off" by T, and that rank(T) measures the dimension of the space of outputs of T—leads us to the intuition that if $\dim(V) = n$ (and so there are n "independent directions" needed to account for V), and since T kills off nullity(T)-many directions by sending ker(T) to $\{0_W\}$, then we would expect the remaining $n - \text{nullity}(T)$ directions to "fill out" image(T), so that $n - \text{nullity}(T) = \text{rank}(T)$, or $\dim(V) = \text{rank}(T) + \text{nullity}(T)$. This is good intuition, and leads us to the so-called **Rank-Nullity Theorem**.

Theorem 32 (Rank-Nullity)**.** Let V and W be vector spaces over $\mathbb{K}$, and assume that V is finite-dimensional. Let $T : V \to W$ be a linear transformation. Then image(T) and ker(T) are finite-dimensional and

$$\dim(V) = \text{rank}(T) + \text{nullity}(T).$$

We will prove this next time.

Lecture 28: Rank-Nullity Theorem

Learning Objectives:

- Use the kernel and image of a linear transformation to characterize whether the function is injective and/or surjective.
- Compute the kernel, image, nullity, and rank of a linear transformation.

We start today by proving the Rank-Nullity Theorem.

Theorem 33 (Rank-Nullity)**.** Let V and W be vector spaces over $\mathbb{K}$, and assume that V is finite-dimensional. Let $T : V \to W$ be a linear transformation. Then image(T) and ker(T) are finite-dimensional and

$$\dim(V) = \operatorname{rank}(T) + \operatorname{nullity}(T).$$

Proof. We have already proved that image(T) and ker(T) are finite-dimensional because V is finite-dimensional. Therefore we need only show that the desired relationship between the dimensions of V and these spaces holds.

By the Constructing Bases Theorem, the empty set extends to a basis $v_1, \ldots, v_p$ for $\ker(T)$. Extend the linearly independent list $v_1, \ldots, v_p$ to a basis $v_1, \ldots, v_p, u_1, \ldots, u_\ell$ for V. (Here we allow the case $p = 0$ if $\ker(T) = \{0_V\}$, and we allow $\ell = 0$ if $\ker(T) = V$.) Then $\dim(V) = p + \ell$ and $\operatorname{nullity}(T) = p$.

We will finish the proof by showing that $T(u_1), \ldots, T(u_\ell)$ is a basis for image(T) (so that $\operatorname{rank}(T) = \ell$). Note first that since $T(u_1), \ldots, T(u_\ell) \in \operatorname{image}(T)$, $\operatorname{span}(T(u_1), \ldots, T(u_\ell)) \subseteq \operatorname{image}(T)$. Let $w \in \operatorname{image}(T)$. Choose $v \in V$ with $T(v) = w$. Choose $c_1, \ldots, c_p, d_1, \ldots, d_\ell \in \mathbb{K}$ with

$$v = c_1 v_1 + \cdots + c_p v_p + d_1 u_1 + \cdots + d_\ell u_\ell.$$

Then since $v_1, \ldots, v_p \in \ker(T)$,

$$T(v) = c_1 T(v_1) + \cdots + c_p T(v_p) + d_1 T(u_1) + \cdots + d_\ell T(u_\ell) = d_1 T(u_1) + \cdots + d_\ell T(u_\ell) \in \operatorname{span}(T(u_1), \ldots, T(u_\ell))$$

and therefore $\operatorname{image}(T) = \operatorname{span}(T(u_1), \ldots, T(u_\ell))$.

It remains to show that $T(u_1), \ldots, T(u_\ell)$ is a linearly independent set. Let $d_1, \ldots, d_\ell \in \mathbb{K}$ and assume that $0_W = d_1 T(u_1) + \cdots + d_\ell T(u_\ell)$. Because T is linear, $0_W = T(d_1 u_1 + \cdots + d_\ell u_\ell)$, so that $d_1 u_1 + \cdots + d_\ell u_\ell \in \ker(T)$. Therefore there are $c_1, \ldots, c_p \in \mathbb{K}$ with

$$d_1 u_1 + \cdots + d_\ell u_\ell = c_1 v_1 + \cdots + c_p v_p, \quad \text{or } 0_V = c_1 v_1 + \cdots + c_p v_p - d_1 u_1 - \cdots - d_\ell u_\ell.$$

Because $v_1, \ldots, v_p, u_1, \ldots, u_\ell$ is a basis for V (and therefore linearly independent), $c_1 = \cdots = c_p = d_1 = \cdots = d_\ell = 0$. Therefore $T(u_1), \ldots, T(u_\ell)$ is a linearly independent set and the proof is complete. □

Remark 76. On your homework you will show that the Rank-Nullity Theorem still holds if we replace the assumption that V is finite-dimensional with the assumption that both ker(T) and image(T) are finite-dimensional. One consequence of this is that if V is infinite-dimensional, then at least one of ker(T) or image(T) must be infinite-dimensional as well.

Example 88. Consider the space $M_{n\times n}(\mathbb{C})$ as a vector space over $\mathbb{R}$, and consider the set of **skew-Hermitian** matrices

$$W = \{A \in M_{n\times n}(\mathbb{C}) \ : \ A^* = -A\}.$$

Show that W is a (real) subspace of $M_{n\times n}(\mathbb{C}$ and compute its dimension.

We could attempt to mimic the computation in your homework where you showed that the space

$$H = \{A \in M_{n\times n}(\mathbb{C}) \ : \ A^* = A\}$$

of Hermitian matrices is a real subspace of $M_{n\times n}(\mathbb{C})$ of dimension[24] n^2.

For a more creative approach, consider the transformation

$$T : M_{n\times n}(\mathbb{C}) \to M_{n\times n}(\mathbb{C}), \quad T(A) \stackrel{def}{=} A - A^*.$$

Then note that for $A = [a_{j,k}]$ and $B = [b_{j,k}] \in M_{n\times n}(\mathbb{C})$ and $c \in \mathbb{R}$,

$$T(A+B) = T([a_{j,k}+b_{j,k}]) = [a_{j,k}+b_{j,k}] - [\overline{a_{k,j}} + \overline{b_{k,j}}] = [a_{j,k}] - [\overline{a_{k,j}}] + [b_{j,k}] - [\overline{b_{k,j}}] = T(A) + T(B)$$

and (since $c \in \mathbb{R}$, so that $\overline{c} = c$)

$$T(cA) = T([ca_{j,k}]) = [ca_{j,k}] - [\overline{ca_{k,j}}] = c[a_{j,k}] - [\overline{ca_{k,j}}] = c[a_{j,k}] - c[\overline{a_{k,j}}] = c[a_{j,k}] - [\overline{a_{k,j}}] = cT(A).$$

Note that $T(A) = 0_{n\times n}$ if, and only if, $A^* = A$, so that $\ker(T) = H$ is the real subspace of Hermitian matrices. Therefore $\text{nullity}(T) = \dim(H) = n^2$.

We claim that $W = \text{image}(T)$. To see this, note that if $A \in W$ then $A^* = -A$. Then

$$T\Big(\frac{1}{2}A\Big) = \frac{1}{2}T(A) = \frac{1}{2}(A - A^*) = \frac{1}{2}(A + A) = A,$$

so $A \in \text{image}(T)$. On the other hand, if $B \in \text{image}(T)$ then there is $A \in M_{n\times n}(\mathbb{C})$ such that $B = T(A) = A - A^*$. Then

$$B^* = (A - A^*)^* = A^* - (A^*)^* = A^* - A = -(A - A^*) = -B,$$

so that $B \in W$. Therefore $W = \text{image}(T)$. It is therefore immediate that W is a (real) subspace of $M_{n\times n}(\mathbb{C})$.

Because the dimension of $M_{n\times n}(\mathbb{C})$ (as a real vector space) is $2n^2$, the Rank-Nullity Theorem implies that

$$2n^2 = \dim(M_{n\times n}(\mathbb{C})) = \dim(\ker(T)) + \dim(\text{image}(T)) = \dim(H) + \dim(W) = n^2 + \dim(W),$$

so that $\dim(W) = n^2$.

The Rank-Nullity Theorem has many interesting practical applications of the type displayed in the previous example, but its important theoretical applications allow us to recover some of the conclusions of the Invertibility Theorem for linear maps between finite-dimensional spaces.

[24] The basis you produced for this space had n diagonal matrices, and $2(n-1)+2(n-2)+\cdots+2(2)+2 = 2\frac{(n-1)n}{2} = n^2-n$ other matrices, and $n + (n^2 - n) = n^2$.

Theorem 34. Let V and W be finite-dimensional vector spaces over $\mathbb{K}$, and let $T : V \to W$ be a linear transformation.

(a) If T is injective, then $\dim(V) \leq \dim(W)$.

(b) If T is surjective, then $\dim(V) \geq \dim(W)$.

(c) If T is bijective, then $\dim(V) = \dim(W)$.

(d) If $\dim(V) = \dim(W)$, then T is injective if, and only if, T is surjective.

Proof. By the Rank-Nullity Theorem, we have

$$\dim(V) = \dim(\text{image}(T)) + \dim(\ker(T)). \tag{6}$$

Suppose T is injective. Then $\ker(T) = \{0_V\}$, so $\dim(\ker(T)) = 0$. Because $\text{image}(T)$ is a subspace of W, (6) and the Subspaces and Dimension Theorem imply that $\dim(V) = \dim(\text{image}(T)) \leq \dim(W)$.

Suppose T is surjective. Then $\text{image}(T) = W$, so that $\dim(\text{image}(T)) = \dim(W)$. Equation (6) gives

$$\dim(V) = \dim(\ker(T)) + \dim(\text{image}(T)) = \dim(\ker(T)) + \dim(W) \geq \dim(W).$$

Part (c) follows immediately from (a) and (b).

Now assume that $\dim(V) = \dim(W)$. Then (6) gives

$$\dim(\text{image}(T)) + \dim(\ker(T)) = \dim(W).$$

If T is injective then $\dim(\ker(T)) = 0$, so $\dim(\text{image}(T)) = \dim(W)$. By the Subspaces and Dimension Theorem, $\text{image}(T) = W$, so that T is surjective. Conversely, if T is surjective then $\dim(\text{image}(T)) = \dim(W)$, so that $\dim(\ker(T)) = 0$ and therefore $\ker(T) = \{0_V\}$. Therefore T is injective. □

Remark 77. One consequence of the previous theorem is that if $T : V \to W$ is linear and if V and W are finite-dimensional with $\dim(V) = \dim(W)$, then T is invertible if, and only if, T is either injective or surjective.

Example 89. The conclusions above do not hold in general for infinite-dimensional vector spaces. For example, consider the maps

$$D : P(\mathbb{R}) \to P(\mathbb{R}), \quad D(p(x)) = p'(x)$$

and

$$M : P(\mathbb{R}) \to P(\mathbb{R}), \quad M(p(x)) = xp(x).$$

Then D is not injective because $D(1) = 0$ (so $\ker(D) \neq \{0(x)\}$), yet D is surjective because, for if $p(x) = a_0 + a_1x + \cdots + a_nx^n \in P(\mathbb{R})$ we have

$$D\Big(a_0x + \frac{a_1}{2}x^2 + \cdots + \frac{a_n}{n+1}x^{n+1}\Big) = a_0 + a_1x + \cdots + a_nx^n = p(x).$$

You will investigate the map M on your homework.

Intuitively, the reasons why the conclusions of the last theorem do not hold for infinite-dimensional vectors spaces is that because an infinite set can have a proper subset that is also infinite.

Lecture 29: Isomorphisms

Learning Objectives:

- Determine when two vector spaces are isomorphic.
- Explore the properties of isomorphisms.

Example 90. Consider $\mathbb{C}^n$ as a vector space over $\mathbb{R}$, and consider the map

$$T : \mathbb{R}^{2n} \to \mathbb{C}^n, \quad T\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{bmatrix}\right) = \begin{bmatrix} x_1 + iy_1 \\ \vdots \\ x_n + iy_n \end{bmatrix}.$$

We showed in Example 84 that T is a linear transformation. Because complex numbers are uniquely determined by their real and imaginary parts, it is immediate to show that T is bijective (and therefore invertible). Therefore it has an inverse maps $T^{-1} : \mathbb{C}^n \to \mathbb{R}^{2n}$ that is also linear.

Note that because T is bijective, T gives a one-to-one correspondence between elements $\vec{v} \in \mathbb{R}^{2n}$ and $T(\vec{v}) \in \mathbb{C}^n$. Because T is linear (and therefore preserve linear combinations), we therefore expect that T should preserve anything that we might characterize using linear combinations (i.e. subspaces, span, linear independence, bases, dimension, etc.). Because there is a bijective map between $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ that preserves linear structure, we expect that, as vector spaces over $\mathbb{R}$, $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ should have exactly the "same" algebraic structure.

In light of the above example, we recall a definition.

Definition 44. Let V and W be vector spaces over $\mathbb{K}$. An invertible linear transformation $T : V \to W$ is called an *isomorphism* from V to W. If such an isomorphism exists, then we say that V is is **isomorphic** to W.

Remark 78. Note that if V is isomorphic to W, then there is an invertible linear map $T : V \to W$, and therefore $T^{-1} : W \to V$ is also an invertible linear map so that W is isomorphic to V. Therefore we can simply say that V and W "are isomorphic" rather than "V is isomorphic to W and W is isomorphic to V."

Example 91. The space $P_n(\mathbb{K})$ is isomorphic to $\mathbb{K}^{n+1}$. There are many isomorphisms, but one straightforward candidate is to use the unique linear map $T : P_n(\mathbb{K}) \to \mathbb{K}^{n+1}$ that satisfies

$$T(1) = \vec{e}_1, \quad T(x) = \vec{e}_2, \quad T(x^2) = \vec{e}_3, \quad \dots \quad , \quad T(x^n) = \vec{e}_{n+1}.$$

Note that such a map exists (and is unique!) by the Constructing Linear Maps Theorem because $1, x, x^2, \ldots, x^{n+1}$ is a basis for $P_n(\mathbb{K})$. By linearity of T, we have that

$$T(a_0 + a_1x + \cdots + a_nx^n) = a_0\vec{e}_1 + a_1\vec{e}_2 + \cdots + a_n\vec{e}_{n+1} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \text{for every } a_0 + a_1x + \cdots + a_nx^n \in P_n(\mathbb{K}).$$

By one of your homework problems, because T is a linear map that sends a basis $(1, x, x^2, \ldots, x^n)$ for $P_n(\mathbb{K})$ to a basis $(\vec{e}_1, \ldots, \vec{e}_{n+1})$ for $\mathbb{K}^{n+1}$, T is an isomorphism.

Example 92. An argument similar to that in the last example shows that $M_{m\times n}(\mathbb{K})$ and $\mathbb{K}^{mn}$ are isomorphic, with one isomorphism given by

$$T : M_{m\times n}(\mathbb{K}) \to \mathbb{K}^{mn}, \quad T([a_{j,k}]) = \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{m,1} \\ a_{1,2} \\ \vdots \\ a_{m,2} \\ \vdots \\ a_{1,n} \\ \vdots \\ a_{m,n} \end{bmatrix}.$$

Isomorphisms give a way to say two objects are "equivalent" when they are not "equal".

Example 93. Let $p \le n$, and consider the subspace $V_p = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_p \\ 0 \\ \vdots \\ 0 \end{bmatrix} : x_1, \ldots, x_p \in \mathbb{K} \right\} = \operatorname{span}(\vec{e}_1, \ldots, \vec{e}_p)$ of $\mathbb{K}^n$. Because $\vec{e}_1, \ldots, \vec{e}_p$ (in $\mathbb{K}^n$) is a basis for V_p, the Constructing Bases Theorem implies that there is a unique linear map $T : V_p \to \mathbb{K}^p$ with $T(\vec{e}_j) = \vec{e}_j$ for each $1 \le j \le p$. By linearity, this T is given by

$$T\left(\begin{bmatrix} x_1 \\ \vdots \\ x_p \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}.$$

Because the linear transformation T sends a basis of V_p to a basis for $\mathbb{K}^p$, one of your homework problems implies that T is an isomorphism.

Therefore, while $\mathbb{K}^p$ is not a subset of $\mathbb{K}^n$ if $n \neq p$ (since $\mathbb{K}^p$ consists of vectors with p entries, while $\mathbb{K}^n$ consists of vectors with n entries), we can think of V_p as an *isomorphic copy* of $\mathbb{K}^p$ that is a subspace of $\mathbb{K}^n$.

Just as "row equivalence" was a type of equivalence between matrices (that is different than ordinary equality), "isomorphism" is a type of equivalence between vector spaces. In particular, we have the following properties.

Proposition 30 (Isomorphism is an Equivalence Relation)**.** Let V, W, U be vector spaces over $\mathbb{K}$.

- (Reflexivity) V is isomorphic to V.
- (Symmetry) If V is isomorphic to W, then W is isomorphic to V.
- (Transitivity) If V is isomorphic to W and W is isomorphic to U, then V is isomorphic to U.

Proof. Note that $I : V \to V$ is a bijective linear transformation, and therefore an isomorphism. This proves reflexivity.

Suppose V is isomorphic to W. Let $T : V \to W$ be an isomorphism. Then T is an invertible linear transformation, so the inverse $T^{-1} : W \to V$ is an invertible linear transformation (and therefore also an isomorphism). This proves symmetry.

Suppose V is isomorphic to W and that W is isomorphic to V. Let $T : V \to W$ and $S : W \to U$ be isomorphism. We claim that $S \circ T : V \to U$ is an isomorphism. Let $u \in U$. Since S is surjective, there is $w \in W$ with $S(w) = u$. Since T is surjective, there is $v \in V$ with $T(v) = w$. Then $(S \circ T)(v) = S(T(v)) = S(w) = u$, so that $S \circ T$ is surjective. Now let $v_1, v_2 \in V$, and suppose that $(S \circ T)(v_1) = (S \circ T)(v_2)$. Since $S(T(v_1)) = S(T(v_2))$ and S is injective, $T(v_1) = T(v_2)$. Since T is injective, $v_1 = v_2$. Therefore $S \circ T$ is injective. Since the composition of linear maps is linear, $S \circ T$ is linear. Therefore $S \circ T$ is an isomorphism, and the result is proved. □

Example 94. The identity map $I : \mathbb{K}^n \to \mathbb{K}^n$ from $\mathbb{K}^n$ to itself is an isomorphism, but there are many other isomorphisms from $\mathbb{K}^n$ to itself. Indeed, because a linear transformation is an isormophism exactly when it is invertible, every linear $T : \mathbb{K}^n \to \mathbb{K}^n$ with invertible standard matrix A is an isomorphism. If we denote the columns of A by $\vec{a}_1, \ldots, \vec{a}_n$, then this is exactly the linear map from $\mathbb{K}^n$ to itself that sends the standard basis to the basis $\vec{a}_1, \ldots, \vec{a}_n$, in the sense that

$$T(\vec{e}_1) = \vec{a}_1, \quad \ldots \quad , \quad T(\vec{e}_n) = \vec{a}_n.$$

Example 95. You showed on your homework that if $T : \mathbb{K}^n \to \mathbb{K}^m$ is invertible, then it must be that $m = n$. Therefore $\mathbb{K}^n$ and $\mathbb{K}^m$ are isomorphic if, and only if, $m = n$.

Isomorphisms preserve almost any algebraic structure that you might wish to preserve. To hammer this home, consider the following theorem. To explain some notation that appears in this theorem: if A, B are sets and $f : A \to B$, then for a subset $C \subseteq A$ we write the **image of C under f** as

$$f(C) \stackrel{def}{=} \{f(c) \ : \ c \in C\} \subseteq B.$$

Theorem 35 (Properties of Isomorphisms)**.** Let V, W be vector spaces over $\mathbb{K}$, and let $T : V \to W$ be an isomorphism. Let $v_1, \dots, v_n \in V$, and let $U \subseteq V$.

(a) $V = \operatorname{span}(v_1, \dots, v_n)$ if, and only if, $W = \operatorname{span}(T(v_1), \dots, T(v_n))$.

(b) $v_1, \dots, v_n$ is a linearly independent set if, and only if, $T(v_1), \dots, T(v_n)$ is a linearly independent set.

(c) $v_1, \dots, v_n$ is a basis for V if, and only if, $T(v_1), \dots, T(v_n)$ is a basis for W.

(d) V is finite-dimensional if, and only if W is finite-dimensional. Moreover, in this case we have $\dim(V) = \dim(W)$.

(e) U is a subspace of V if, and only if, $T(U)$ is a subspace of W.

Proof. You will prove the $\Rightarrow$ direction (c) on your homework (by, of course, proving the $\Rightarrow$ directions of (a) and (b)). The $\Leftarrow$ directions of (a),(b),(c) follow by applying the $\Rightarrow$ directions of these results to $T^{-1} : W \to V$.

Both directions of (d) follows from (a), and the statement that $\dim(V) = \dim(W)$ if one (and therefore both) of V and W are finite-dimensional follows from Theorem 34.

Suppose that U is a subspace of V. By the Subspace Criteria, $0_V \in U$, so that $0_W = T(0_V) \in T(U)$. Let $h_1, h_2 \in T(U)$ and $c \in \mathbb{K}$. Choose $u_1, u_2 \in U$ with $T(u_1) = h_1$ and $T(u_2) = h_2$. Then $h_1 + h_2 = T(u_1) + T(u_2) = T(u_1 + u_2) \in T(U)$ and $ch_1 = cT(u_1) = T(cu_1) \in T(U)$. By the Subspace Criteria, $T(U)$ is a subspace of W. The reverse direction follows from the same argument applied to T^{-1} once we show that $U = T^{-1}(T(U))$. Let $u \in U$. Then $u = T^{-1}(T(u)) \in T^{-1}(T(U))$. Now suppose that $u \in T^{-1}(T(U))$. Then there is $h \in T(U)$ with $u = T^{-1}(h)$. There is $u' \in U$ with $T(u') = h$. Then $u = T^{-1}(T(u')) = u' \in U$. $\square$

Example 96. One upshot of the Properties of Isomorphism Theorem is that isomorphisms allow us to answer questions about one vector space using techniques developed for a different vector space. For an illustration, we will show that for each $a \in \mathbb{K}$, the set $1,\ x - a,\ (x-a)^2,\ (x-a)^3$ is a basis for $P_3(\mathbb{K})$.

To see this, let $T : P_3(\mathbb{K}) \to \mathbb{K}^4$ be the isomorphism $T(a_0 + a_1 x + a_2 x^2 + a_3 x^3) = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$. Then

$$T(1) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad T(x-a) = \begin{bmatrix} -a \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad T((x-a)^2) = T(a^2 - 2ax + x^2) = \begin{bmatrix} a^2 \\ -2a \\ 1 \\ 0 \end{bmatrix},$$

and

$$T((x-a)^3) = T(-a^3 + 3a^2 x - 3ax^2 + x^3) = \begin{bmatrix} -a^3 \\ 3a^2 \\ -3a \\ 1 \end{bmatrix}.$$

Because

$$\text{rref}\begin{bmatrix} T(1) & T(x-a) & T((x-a)^2) & T((x-a)^3) \end{bmatrix} = \text{rref}\begin{bmatrix} 1 & -a & a^2 & -a^3 \\ 0 & 1 & -2a & 3a^2 \\ 0 & 0 & 1 & -3a \\ 0 & 0 & 0 & 1 \end{bmatrix} = I_4,$$

The Invertibility Theorem implies that $T(1), T(x-a), T((x-a)^2), T((x-a)^3)$ is a basis for $\mathbb{K}^4$, and therefore the Properties of Isomorphisms Theorem implies that $1, x-a, (x-a)^2, (x-a)^3$ is a basis for $P_3(\mathbb{K})$.