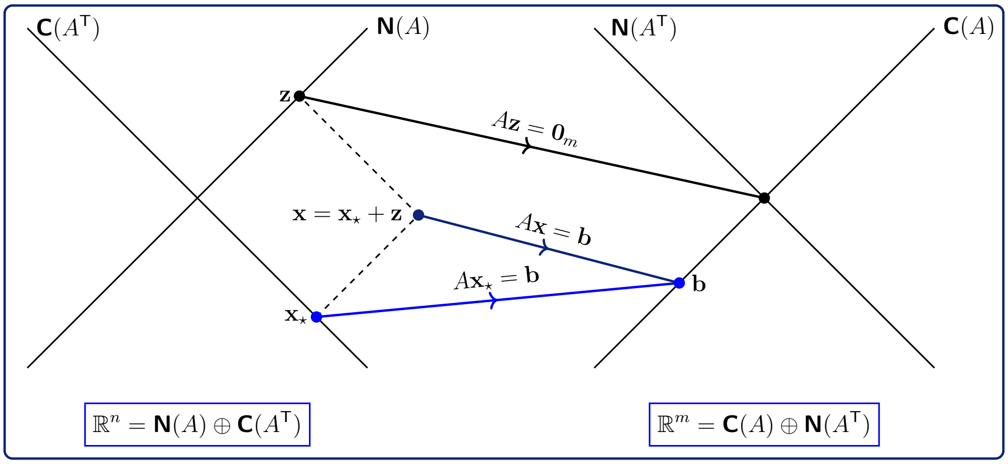
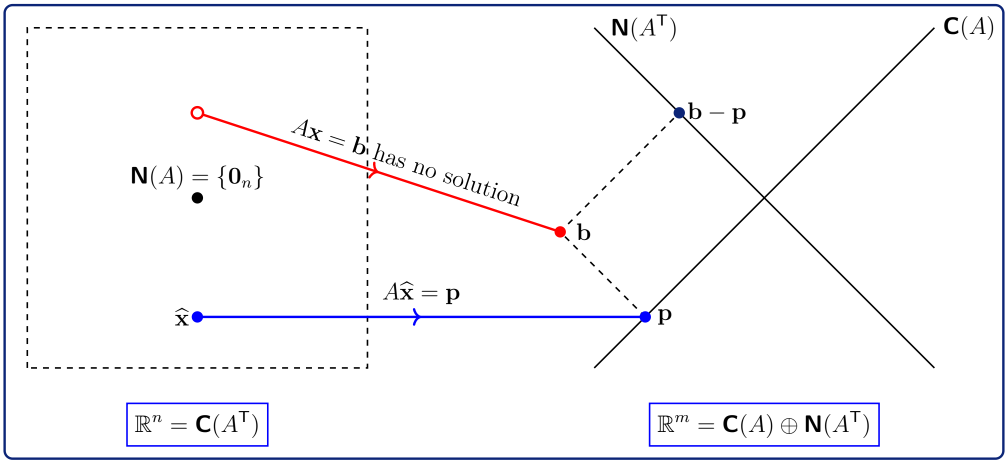
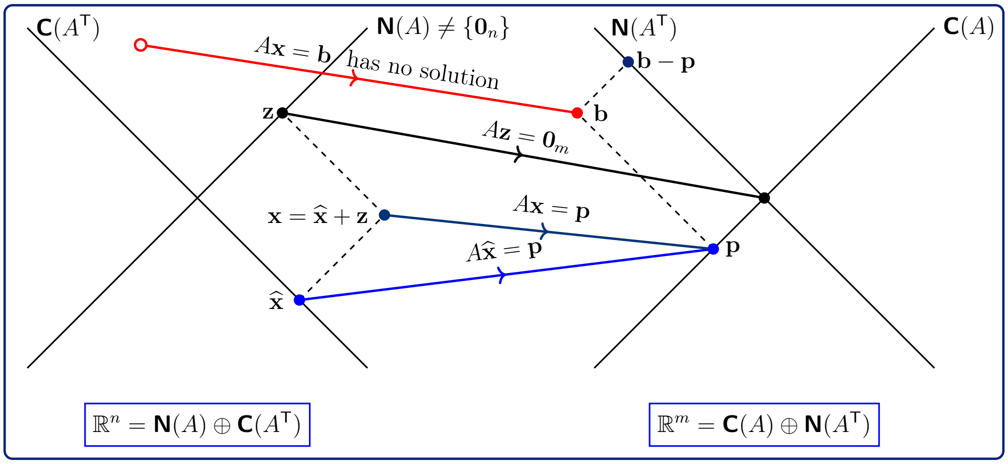
MATH 3260: LINEAR ALGEBRA I

Timothy E. Faver

April 8, 2026







Contents

1. Vectors and Matrices

1.1. Introduction: who cares?

One of the most important questions to ask in mathematics is not *What does this mean?* or *How do I do it?* or *Why is this true?* but *Who cares?* You should care about linear algebra because it is used *everywhere*. Linear algebra will show up in virtually every math problem that you could pose—even calculus, and especially multivariable calculus. And unlike much of the calculus you learned prior to this course, there is a very good chance that you will use linear algebra in your career.

Nearly every part of linear algebra has an application. As we tend to do in mathematics, we will start this course from the very beginning, so applications might not be obvious right away. That being said, here are some meaningful applications that are particularly relevant to contemporary applied mathematics.

- How do we visualize, encode, or arrange a large set of data, say, "n-dimensional" data with $n \geq 4$? How do we optimize the storage of that data, possibly by extracting the most important parts or principal components, or otherwise approximating it?

- How do we fit a line, or polynomial, or some kind of well-behaved "curve" to data? Do we want the curve merely to pass through all of the data points or do we want the curve to be a good approximation to the "behavior" of the data (even if it does not pass through any of the data points at all)?

- How do we find the minimum distance between objects, possibly where those objects are sets of other objects? Or how do we find the best approximation to an object within a specialized class of other objects?

- How do machine learning models work? How does data science work? (All the cool kids sure do want to be data scientists these days.)

We will discuss some of these applications later, but we need to build up quite a body of knowledge first. Specifically, to get to the answers, we need to introduce and study two fundamental objects: vectors and matrices. This course *might* start off simple. While the majority of the calculations that we do will remain "simple" for the sake of being able to do them by hand—the numbers will be much nicer than they would be in any "real" application, for which you would be using a computer to do the calculations—the course *will not* remain that way. Concepts quickly build on each other, and it is easy to get overwhelmed. *Vocabulary* will be a major part of this course, possibly much more so than you are used to from prior mathematics classes. *Do not slack off.*

Most of the punchy applications boil down to solving linear systems of equations; these are incredibly versatile problems, much more so than they might look at first glance. More accurately, our task is to *understand* linear systems of equations. Having a solution formula for something is not the same as understanding that thing. The challenge is finding the right system to model your application and then understanding it.

1.2. A toy problem will illustrate many techniques.

We can tease out a tremendous amount of structure and theory from very simple motivating examples, and here will be our favorite for some time. We solve the **LINEAR SYSTEM**

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11. \end{cases} \tag{1.2.1}$$

This is a system of equations because there is more than one equation, and it is linear because the unknowns only appear as the "linear powers" x and y, not x^2 or xy or $\cos(x+y)$.

This is far from the most challenging or profound problem that we could pose, and we probably already know how to solve it, but imagine if the system had 50 variables and 50 equations. Then we would probably want a precise and systematic way of approaching it.

1.2.1 Problem (!). Try to solve (1.2.1).

Before we do anything to (1.2.1), here are some questions that we should ask.

1. Does it have a solution? That is, do there *exist* numbers x and y that make the two equalities in (1.2.1) true?

2. If not, why not? Can we quantify or qualify *failure* to solve a linear system?

3. And if there is no solution, can we somehow *approximate* the problem by one that does have a solution? Will that approximation be meaningful or helpful?

4. Is there only one solution? Is there only one way to choose the values of x and y to make the two equalities in (1.2.1) true? That is, is the solution *unique*?

5. If not, why not? Can we quantify or qualify why a linear system might have more than one solution? Is there a *best* solution to choose from among many?

We will solve (1.2.1) by transforming it into an "equivalent" system of equations that is much easier to solve—actually, several "equivalent" systems. We say that two systems are **EQUIVALENT** if they have precisely the same solutions. We accomplish this transformation via algebra—more precisely, two tricks with algebra. (A trick is just a technique that does not feel natural yet.)

Recall that if a and b are real numbers, then

$$a = b \iff ca = cb \text{ for all } c \neq 0.$$

That is, if we know $a = b$, then we also know $ac = bc$ for all nonzero c (when $c = 0$, this is still true, as $a0 = b0 = 0$). And if we know $ac = ab$ for all nonzero c (actually, for just *one* nonzero c), then we can divide by c to get $a = b$ (we want $c \neq 0$ so you can divide by c). In the context of linear systems, scaling *both* sides of the *same* equation by the *same* **nonzero** number does not change things. We multiply the first equation by the very convenient number $c = -3$:

$$\left\{\begin{array}{rcrcr} x & - & 2y & = & 1 \\ 3x & + & 2y & = & 11 \end{array}\right. \iff \left\{\begin{array}{rcrcr} -3x & + & 6y & = & -3 \\ 3x & + & 2y & = & 11. \end{array}\right. \tag{1.2.2}$$

That was the first trick.

Now we use another property of algebra. Recall that if a and b are real numbers, then

$$a = b \iff a + c = b + c \text{ for all } c.$$

That is, if we know $a = b$, then we can add c to both sides to get $a + c = b + c$. And if we know $a+c = b+c$ for all c, then just subtract c from both sides (or add $-c$ to both sides) to get $a = b$. (Actually, we just need $a + c = b + c$ for one particular c, not for all c, to allow us to do this subtraction. But if we know $a + c = b + c$ for all c, just take $c = 0$ to get $a = b$.)

So, we can add any c that we like to both sides of the second equation in the system on the right in (1.2.2) to find the equivalence

$$\left\{\begin{array}{rcrcr} x & - & 2y & = & 1 \\ 3x & + & 2y & = & 11 \end{array}\right. \iff \left\{\begin{array}{rcrcr} -3x & + & 6y & = & -3 \\ 3x & + & 2y + c & = & 11 + c \end{array}\right.$$

Is there a particularly helpful choice of c? We work backwards (a great direction to work in math!).

If we know that x and y solve

$$\left\{\begin{array}{rcrcr} -3x & + & 6y & = & -3 \\ 3x & + & 2y + c & = & 11 + c, \end{array}\right.$$

then we must have $-3x + 6y = -3$. After all, that is just the first equation here. So, we could take $c = -3$, which is the same as $c = -3x + 6y$. Since the variables show up on the left, we use $c = -3x + 6y$ on the left and $c = -3$ on the right.

We get

$$3x + 2y + (-3x + 6y) = 8y \quad \text{and} \quad 11 + (-3) = 8.$$

Then we have the arrow going one way:

$$\left\{\begin{array}{rcrcr} x & - & 2y & = & 1 \\ 3x & + & 2y & = & 11 \end{array}\right. \implies \left\{\begin{array}{rcrcr} -3x & + & 6y & = & -3 \\ & & 8y & = & 8 \end{array}\right. \tag{1.2.3}$$

The second equation on the right in (1.2.3) is pretty nice: $8y = 8$, so $y = 1$. What is really nice is that $8y = 8$ has only the unknown y, not both x and y. The first equation then becomes $-3x + 6 = -3$, so $-3x = -9$, and therefore $x = 3$. It looks like we solved our original problem (1.2.1). Did we? We can always plug $x = 3$ and $y = 1$ into (1.2.1) and make sure that everything is equal.

1.2.2 Problem (!). Do that.

Did we really have to do that? Without actually knowing the values of x and y, can we figure out why any solution to the second system in (1.2.3) is also a solution to the original problem (1.2.1)? First, it is helpful to notice that

$$\left\{\begin{array}{rcrcr} -3x & + & 6y & = & -3 \\ & & 8y & = & 8 \end{array}\right. \iff \left\{\begin{array}{rcrcr} x & - & 2y & = & 1 \\ & & 8y & = & 8. \end{array}\right.$$

That is, keeping the factor of -3 on each of the terms in the first equation was unnecessary. We only put that -3 in to help simplify the second equation; that -3 actually makes the first equation worse.

We also know

$$\begin{cases} x - 2y = 1 \\ 8y = 8 \end{cases} \iff \begin{cases} x - 2y = 1 \\ 8y + c = 8 + c. \end{cases}$$

for any choice of c. Nothing new here, right?

If we know that x and y solve

$$\begin{cases} x - 2y = 1 \\ 8y + c = 8 + c. \end{cases}$$

then the first equation says $x - 2y = 1$, so multiplying both sides of that by 3 gives $3x - 6y = 3$ (which looks familiar). Then we could take $c = 3x - 6y$ on the left in the second equation and $c = 3$ on the right in the second equation to get back to where we were:

$$\begin{cases} x - 2y = 1 \\ 8y = 8 \end{cases} \Longrightarrow \begin{cases} x - 2y = 1 \\ 3x + 2y = 11. \end{cases} \tag{1.2.4}$$

This is the reverse of the arrow in (1.2.3).

When we combine (1.2.2), (1.2.3), and (1.2.4), we get

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff \begin{cases} x - 2y = 1 \\ 8y = 8 \end{cases} \iff \begin{cases} x = 3 \\ y = 1. \end{cases} \tag{1.2.5}$$

This is an existence and uniqueness result for (1.2.1): there exists a solution ($x = 3$ and $y = 1$), and it is the only solution. Specifically, existence is the logic of $\Longleftarrow$: plug these values in and check that true equalities result. Uniqueness is the logic of $\Longrightarrow$: if x and y solve the problem, then we must have $x = 3$ and $y = 1$.

The preceding work illustrates two incredibly important operations in solving linear systems: multiply both sides of one equation by the same number, and subtract (or add) a multiple of one equation to another equation. There is a third operation—interchanging two equations, which sounds silly but actually is worthwhile—that we will meet later. Eventually we will encode and view these operations at pretty high and abstract levels.

The preceding work also illustrates something that is incredibly *unimportant* about linear systems: what we call the variables. As long as we are consistent, w could write x and y, or x_1 and x_2, or α and β, and so on. What matters are the *coefficients* on the variables and the *numbers on the right.*

We are going to stack these numbers together as **COLUMN VECTORS**, which we just call "lists of numbers" right now. Here are the three important vectors in (1.2.1), and we also write them as ordered pairs to make typesetting easier:

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} = (1, 3), \qquad \begin{bmatrix} -2 \\ 2 \end{bmatrix} = (-2, 2), \qquad \text{and} \qquad \begin{bmatrix} 1 \\ 11 \end{bmatrix} = (1, 11).$$

But

$$\begin{bmatrix}1\\3\end{bmatrix} \neq \begin{bmatrix}1 & 3\end{bmatrix}.$$

The object on the right is a **ROW VECTOR**, and we will talk about that eventually. We typeset column vectors in bold, say,

$$\mathbf{v} = \begin{bmatrix}1\\3\end{bmatrix} = (1, 3).$$

We will do a lot of arithmetic with (column) vectors, and much of it will happen "componentwise." We add vectors by adding their corresponding components, so

$$\begin{bmatrix}1\\3\end{bmatrix} + \begin{bmatrix}-2\\2\end{bmatrix} = \begin{bmatrix}1+(-2)\\3+2\end{bmatrix} = \begin{bmatrix}-1\\5\end{bmatrix}. \tag{1.2.6}$$

1.2.3 Remark. *Strictly speaking, we are overworking the role of the symbol + in* (1.2.6)*. The + in*

$$\begin{bmatrix}1\\3\end{bmatrix} + \begin{bmatrix}-2\\2\end{bmatrix}$$

is the addition of column vectors, while the + in

$$\begin{bmatrix}1+(-2)\\3+2\end{bmatrix}$$

is the addition of real numbers. Nobody ever thinks like that in practice, but as you open yourself to new ideas in this course, you should be aware that the same symbol can mean different things, depending on context.

1.2.4 Problem (!). Compute

$$\begin{bmatrix}1\\0\\1\end{bmatrix} + \begin{bmatrix}0\\1\\0\end{bmatrix}.$$

Then we can rewrite the original problem (1.2.1) as

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff \begin{bmatrix}x\\3x\end{bmatrix} + \begin{bmatrix}-2y\\2y\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}.$$

Big deal, right? All we have done is introduced some new notation; this tells us absolutely nothing about solving (1.2.1) that we did not already know. We do one more bit of arithmetic. There are "common factors" of x and y in some of those vectors, and our gut instinct should be to factor them out.

So, we define multiplication of a vector by a number (we do *not* multiply two vectors) componentwise:

$$2\begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}2(1)\\2(3)\end{bmatrix} = \begin{bmatrix}2\\6\end{bmatrix}.$$

We often refer to this kind of multiplication as "scalar" multiplication to emphasize that one of the factors is a "scalar"—which is to say, a real number. When multiplying a vector by a number, we always write the number first:

$$2\begin{bmatrix}1\\3\end{bmatrix}, \text{ not } \begin{bmatrix}1\\3\end{bmatrix}2 \quad \text{and} \quad c\mathbf{v}, \text{ not } \mathbf{v}c.$$

1.2.5 Problem (!). Compute

$$(-1)\begin{bmatrix}1\\0\\1\end{bmatrix} \quad \text{and} \quad 0\begin{bmatrix}0\\1\\0\end{bmatrix}.$$

Content from Strang's *ILA* 6E. See the pictures on pp. v–vi for how to interpret vector addition and scalar multiplication in two dimensions. I'll talk about this later. Page 3 has another good picture that contrasts $\mathbf{v}$ and $-\mathbf{v}$ as a kind of "reflection." Also look at Figure 1.2 (a) on p. 8 to see the effect of "averaging" the sum of two vectors. There is more componentwise arithmetic on pp. 1–2.

We rewrite (1.2.1) once again as

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff x\begin{bmatrix}1\\3\end{bmatrix} + y\begin{bmatrix}-2\\2\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}.$$

Again, this offers absolutely no insights into actually solving (1.2.1)—yet.

The expression

$$x\begin{bmatrix}1\\3\end{bmatrix} + y\begin{bmatrix}-2\\2\end{bmatrix}$$

is something that we will see often: it is a **LINEAR COMBINATION** of the vectors $(1,3)$ and $(-2,2)$. Many important ideas can be phrased in the language of linear combinations.

Content from Strang's *ILA* 6E. Page 3 has some pictures of linear combinations. See also a linear system on p. 3 that is written in vector form and then solved with elimination, as we did (1.2.1).

1.3. Vectors store data.

Here are some more precise definitions of concepts from our first pass at linear systems and vectors. Throughout, we use the following set-theoretic terminology as a convenient abbreviation: if S is a set and x is an element of S, then we write $x \in S$. If a set has only finitely many elements, we may write those elements out between curly braces; the order or repetition of the elements doesn't matter. For example, $\{1,2,3\} = \{2,1,3\} = \{1,1,2,3\}$ and $1 \in \{1,2,3\}$. In particular, we denote by $\mathbb{R}$ the set of all real numbers, so $1 \in \mathbb{R}$. For a set S, we write $x \notin S$ to mean that x is not an element of S; for example, $0 \notin \{1,2,3\}$. Going forward, we often call real numbers **SCALARS**.

1.3.1 Undefinition. *Let $n \geq 1$ be an integer.*

(i) *A **COLUMN VECTOR** of length n is an "ordered list" of n real numbers, which we call the **ENTRIES** or the **COMPONENTS** of $\mathbf{v}$. If $\mathbf{v}$ is a column vector of length n with entries $v_1, \ldots, v_n$ in that order, then we write*

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \qquad \text{or} \qquad \mathbf{v} = (v_1, \ldots, v_n).$$

The number v_j here specifically is the jth entry (or component) of $\mathbf{v}$.

(ii) *The set of all column vectors of length n is $\mathbb{R}^n$, and we write*

$$\mathbb{R}^n = \left\{ \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \;\middle|\; v_1, \ldots, v_n \in \mathbb{R} \right\}.$$

We typically work with $n \geq 2$, and we do not typically distinguish $\mathbb{R}^1$ and $\mathbb{R}$, so $\mathbb{R}^1 = \mathbb{R}$.

(iii) *If $\mathbf{v} \in \mathbb{R}^n$, then the **LENGTH** or **SIZE** of $\mathbf{v}$ is the integer n. (We might want to call n the "dimension" of $\mathbf{v}$, but this will conflict with another important use of the word "dimension" in the future.)*

(iv) *Two vectors $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ are **EQUAL** if and only if their corresponding entries are equal:*

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \iff v_j = w_j, \; j = 1, \ldots, n.$$

(v) *We define vector addition and multiplication by real numbers componentwise, regardless of the length of the vectors. In particular, if $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then $\mathbf{v}+\mathbf{w} \in \mathbb{R}^n$ is given by adding the corresponding components of $\mathbf{v}$ and $\mathbf{w}$, and $c\mathbf{v} \in \mathbb{R}^n$ is given by multiplying each component of $\mathbf{v}$ by c. However, if $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^m$ with $n \neq m$, then $\mathbf{v} + \mathbf{w}$ is not defined.*

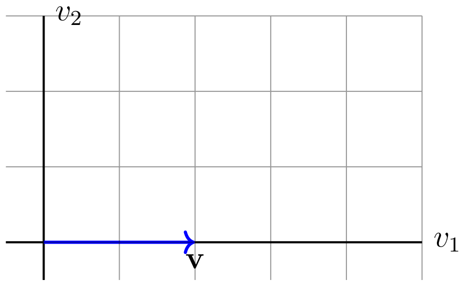
1.3.2 Problem (!). Does the expression

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

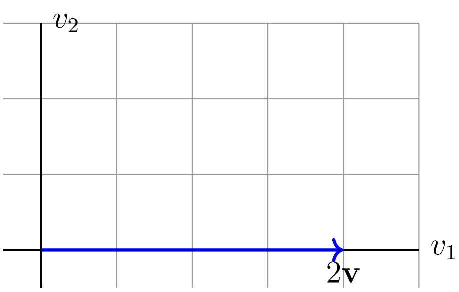
make sense?

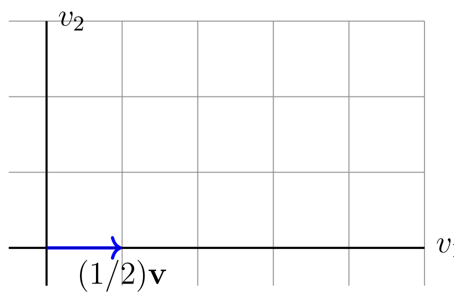
When $n \geq 4$, we have no (natural) way of drawing vectors. Even $n = 3$ can be hard, but sometimes drawing pictures for $n = 2$ can lead to insight. When we do so, we follow the convention that the vector $\mathbf{v} = (v_1, v_2)$ is represented by the "directed line segment" (= arrow, but fancier) from the origin $(0, 0)$ to the point (v_1, v_2). This leads to some nice geometric interpretations of vector addition and scalar multiplication.

1.3.3 Example. **(i)** Here is a drawing of the vector $\mathbf{v} = (2, 0)$.



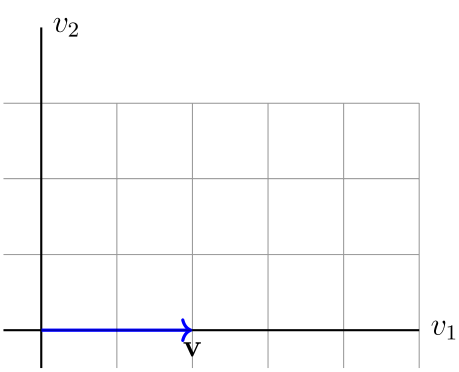
And here are drawings of $2\mathbf{v} = (4, 0)$ and $(1/2)\mathbf{v} = (1, 0)$.

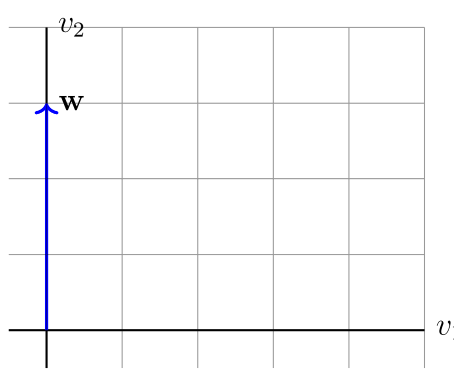




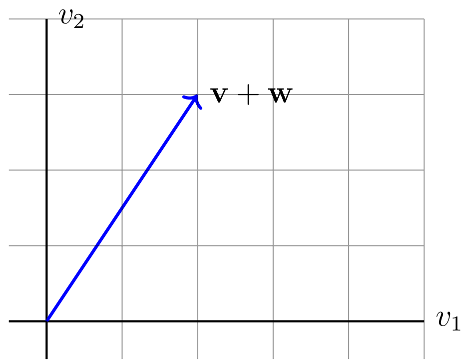
It should feel that $2\mathbf{v}$ is a "stretching" of $\mathbf{v}$ and $(1/2)\mathbf{v}$ is a "shrinking" (which is also a kind of stretching).

(ii) Here are side-by-side drawings of the vectors $\mathbf{v} = (2, 0)$ from before and $\mathbf{w} = (0, 3)$, which is new.

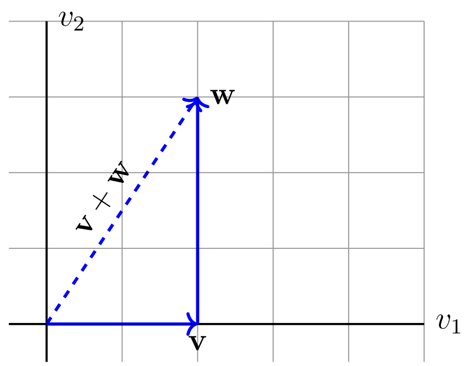




And here is the sum $\mathbf{v} + \mathbf{w} = (2, 3)$.



A cartoonish, but helpful, way to visualize the action of vector addition is that we placed the "tip" of $\mathbf{w}$ at the "tail" (the arrow end) of $\mathbf{v}$ to get the sum $\mathbf{v} + \mathbf{w}$.



1.3.4 Example. We compute

$$0 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0(1) \\ 0(2) \\ 0(3) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Hopefully it feels intuitive why we want to call the vector on the right the "zero vector in $\mathbb{R}^3$."

1.3.5 Definition. *The* **ZERO VECTOR** *in* $\mathbb{R}^n$ *is the vector* $\mathbf{0}$ *whose entries are all* 0*. Often we will write* $\mathbf{0}_n$ *to emphasize that this is the zero vector with* n *entries. A vector is* **NONZERO** *if it has at least one nonzero entry (but a nonzero vector may have some zero entries).*

For example,

$$\mathbf{0}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{but} \quad \mathbf{0}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

while both

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

are nonzero vectors.

1.3.6 Problem (!). **(i)** Let $\mathbf{v} \in \mathbb{R}^n$. What is $\mathbf{v} + \mathbf{0}_n$?

(ii) Does $\mathbf{0}_2 + \mathbf{0}_3$ make sense?

(iii) Generalize Example 1.3.4 by computing $0\mathbf{v}$ for an arbitrary $\mathbf{v} \in \mathbb{R}^n$.

(iv) Suppose that $c\mathbf{v} = \mathbf{0}_n$ for some $\mathbf{v} \in \mathbb{R}^n$ with $\mathbf{v} \neq \mathbf{0}_n$. Why must it be the case that $c = 0$? [Hint: *at least one component of* $\mathbf{v}$ *is nonzero.*]

1.3.7 Problem (!). **(i)** Let $\mathbf{v} \in \mathbb{R}^n$. What is $1\mathbf{v}$?

(ii) Let $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Explain why $c(\mathbf{v} + \mathbf{w}) = c\mathbf{v} + c\mathbf{w}$.

(iii) Let $\mathbf{v} \in \mathbb{R}^n$ and c_1, $c_2 \in \mathbb{R}$. Explain why $(c_1 + c_2)\mathbf{v} = c_1\mathbf{v} + c_2\mathbf{v}$.

The most important (and only) arithmetical operations that we have defined for vectors are "vector addition" of two vectors in $\mathbb{R}^n$ to get a third vector in $\mathbb{R}^n$ and "scalar multiplication" of a number (a "scalar") in $\mathbb{R}$ and a vector in $\mathbb{R}^n$ to get a second vector in $\mathbb{R}^n$. When we do (possibly) both simultaneously, we get a new structure.

1.3.8 Definition. *Let* $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ *and* $c_1, \ldots, c_n \in \mathbb{R}$. *The* ***LINEAR COMBINATION*** *of* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ ***WEIGHTED*** *by* $c_1, \ldots, c_n$ *is the vector* $\mathbf{v} \in \mathbb{R}^m$ *defined by*

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n. \tag{1.3.1}$$

We may also express this in sigma notation:

$$\mathbf{v} = \sum_{j=1}^{n} c_j\mathbf{v}_j.$$

If $\mathbf{v} \in \mathbb{R}^m$ *has the form* (1.3.1), *then we often say that* $\mathbf{v}$ *is a linear combination of* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *without mentioning the weights. If* $n = 1$, *then the linear combination of* $\mathbf{v}_1$ *weighted by* c_1 *is just the scalar multiplication* $c_1\mathbf{v}_1$.

1.3.9 Problem (!). Convince yourself that, in the notation of the previous definition, we do indeed have $\mathbf{v} \in \mathbb{R}^m$. Also, what are the integers m and n encoding in that definition?

1.3.10 Example. We have

$$\begin{bmatrix}2\\3\\0\end{bmatrix} = \begin{bmatrix}2\\0\\0\end{bmatrix} + \begin{bmatrix}0\\3\\0\end{bmatrix} = 2\begin{bmatrix}1\\0\\0\end{bmatrix} + 3\begin{bmatrix}0\\1\\0\end{bmatrix},$$

and so $(2,3,0)$ is a linear combination of $(1,0,0)$ and $(0,1,0)$. So is $(1,0,0)$:

$$\begin{bmatrix}1\\0\\0\end{bmatrix} = 1\begin{bmatrix}1\\0\\0\end{bmatrix} + 0\begin{bmatrix}0\\1\\0\end{bmatrix}.$$

1.3.11 Problem (!). Here is a generalization of this example. Let

$$\mathbf{e}_1 := \begin{bmatrix}1\\0\\0\end{bmatrix}, \qquad \mathbf{e}_2 := \begin{bmatrix}0\\1\\0\end{bmatrix}, \qquad \text{and} \qquad \mathbf{e}_3 := \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Explain why any $\mathbf{v} \in \mathbb{R}^3$ is a linear combination of $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$.

Content from Strang's *ILA* 6E. There are examples of linear combinations with $n = 2$ on p. vi and p. 2.

1.3.12 Remark. *Why is Undefinition 1.3.1 an undefinition, not a definition? Because we did not give a precise definition of "ordered list."*

We could think of column vectors of length n as functions from the set $\{1, \ldots, n\}$ to $\mathbb{R}$. That is, if $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$, then $\mathbf{v}$ is the same as the function $f\colon \{1, \ldots, n\} \to \mathbb{R}$ such that $f(j) = v_j$ for $j = 1, \ldots, n$. And since functions are really sets of ordered pairs, $f = \{(j, v_j)\}_{j=1}^n$. No one ever thinks like this in practice.

So far, none of this (mostly) more precise terminology tells us anything new about solving linear systems, and, honestly, none of the following is going to help, either. The goal is to build more terminology so that we can ask questions about linear systems *in the right language.*

1.4. Matrices store more data.

Here is a major step toward that right language. Recall that our original problem (1.2.1) can be written as a system of linear equations or as a vector equation involving a linear combination:

$$\left\{\begin{array}{rcrcr} x_1 & - & 2x_2 & = & 1 \\ 3x_1 & + & 2x_2 & = & 11 \end{array}\right. \iff x_1\begin{bmatrix}1\\3\end{bmatrix} + x_2\begin{bmatrix}-2\\2\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}.$$

We put the coefficient vectors together into a **MATRIX**:

$$A := \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}.$$

This is a **SQUARE** matrix: it has the same number of columns and rows (two each). We most often think of matrices in terms of columns (though rows are sometimes useful). If we put

$$\mathbf{a}_1 := \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{a}_2 := \begin{bmatrix} -2 \\ 2 \end{bmatrix},$$

then we will also write A as

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}.$$

This is sort of a "row vector" of column vectors.

Here is where we are going with all of this. Abbreviate $\mathbf{x} = (x_1, x_2)$ and $\mathbf{b} = (1, 11)$. Our goal is to define a notion of "matrix-vector multiplication" so that if $A\mathbf{x}$ is the "product" of A and $\mathbf{x}$, then our original problem compresses to

$$A\mathbf{x} = \mathbf{b}.$$

First, of course, we need some more terminology. We control the "sizes" or "dimensions" of matrices by counting the numbers of rows and the numbers of columns—and we always list *rows before columns*. We write $A \in \mathbb{R}^{2\times 2}$ for the matrix A above, and, for example,

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 3} \quad \text{and} \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \in \mathbb{R}^{3\times 2}.$$

More generally, we say the following.

1.4.1 Undefinition. *Let m, $n \geq 1$ be integers.*

(i) *An $m \times n$ **MATRIX** is a rectangular array of numbers with m rows and n columns.*

(ii) *We denote the set of all $m \times n$ matrices by $\mathbb{R}^{m\times n}$.*

(iii) *We say that the **SIZE** of a matrix $A \in \mathbb{R}^{m\times n}$ is $m \times n$, pronounced "m by n." (We might be tempted to say that the "dimension" of A is $m \times n$, but this will conflict with another important use of the word "dimension" in the future.)*

(iv) *Since a matrix with m rows and 1 column is really just an ordered list of m numbers, we will typically not distinguish $\mathbb{R}^{m\times 1}$ and $\mathbb{R}^m$, and so usually $\mathbb{R}^{m\times 1} = \mathbb{R}^m$. Also, $\mathbb{R}^{1\times 1} = \mathbb{R}$. (Occasionally we make irritating exceptions to this practice.) But we do not equate $\mathbb{R}^{1\times n}$ and $\mathbb{R}^n$, and we often call matrices in $\mathbb{R}^{1\times n}$ **ROW VECTORS**.*

(v) *The (i, j)-**ENTRY** (sometimes the (i, j)-**COMPONENT**) of a matrix is the entry in row i, column j of that matrix. Sometimes we will write A_{ij} for the (i, j)-entry of A, although with large matrices it might be clearer to write $A_{i,j}$, or maybe even $A(i, j)$.*

(vi) *Matrices A, $B \in \mathbb{R}^{m \times n}$ are* **EQUAL**, *written $A = B$, if the (i,j)-entry of A equals the (i,j)-entry of B for all i and j. (If $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$ with either $m \neq p$ or $n \neq q$, then we define $A \neq B$.)*

(vii) *The* **ZERO MATRIX** *in $\mathbb{R}^{m \times n}$ is the $m \times n$ matrix whose entries are all zero. There is no standard notation for the zero matrix as for the zero vector.*

(viii) *A matrix is* **NONZERO** *if it has at least one nonzero entry. (We allow some entries in a nonzero matrix to be zero.)*

1.4.2 Example.

(i) Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

Then $A \in \mathbb{R}^{3 \times 2}$. The $(1,2)$-entry of A is 2, and the $(2,1)$-entry of A is 3.

(ii) The matrix

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

is the zero matrix in $\mathbb{R}^2$, but the matrix

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

is nonzero because the $(1,1)$-entry is nonzero (although all of the other entries are zero).

Content from Strang's *ILA* 6E. A 3×2 matrix appears on p. vi, and a larger one (what size?) on p. vii.

Regarding our choice not to identify $\mathbb{R}^{1 \times n}$ and $\mathbb{R}^n$, we have things like

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = (1, 2, 3).$$

1.4.3 Problem (!). Reread that sentence until it makes sense.

1.4.4 Remark. *As with column vectors, our attempt at defining a matrix is really an undefinition because we did not rigorously define "rectangular array" of numbers. We could think of $A \in \mathbb{R}^{m \times n}$ as the function $f \colon I \to \mathbb{R}$ such that $f(i,j) = A_{ij}$, where $I = \{(i,j) \mid i = 1, \ldots, m, \ j = 1, \ldots, n\}$. Or as the function $g \colon \{1, \ldots, n\} \to \mathbb{R}^m$ such that*

$g(j) = \mathbf{a}_j$, where $\mathbf{a}_j$ is the jth column of A, i.e., $A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}$. (Strictly speaking, f and g are not the same function, as they have different domains and ranges!) Neither way of thinking will make any of the following any easier.

And as with column vectors, we add matrices and multiply them by real numbers componentwise.

1.4.5 Problem (!). Compute

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ -1 & -1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \qquad \text{and} \qquad 2\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

We conclude here with some practice in reading definitions and proof styles.

1.4.6 Lemma. *Let A, $B \in \mathbb{R}^{m \times n}$. Then $A = B$ if and only if the jth column of A equals the jth column of B (in the sense of part (iv) of Definition 1.3.1) for $j = 1, \ldots, n$. That is,*

$$\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix} = \begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_n\end{bmatrix} \iff \mathbf{a}_j = \mathbf{b}_j, \ j = 1, \ldots, n.$$

Proof. This is the first of many "if and only if"-type statements that we will meet in this course. We need to show that each side of the "if and only if" implies the other.

($\Longrightarrow$) Suppose that $A = B$, and so we know that the (i,j)-entry of A equals the (i,j)-entry of B for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. We need to show that the jth column of A equals the jth column of B. These columns are vectors in $\mathbb{R}^m$, and so by part (iv) of Definition 1.3.1, we need to show that each of their corresponding m components are equal. The ith component of column j of A is the (i,j)-entry of A. The hypothesis $A = B$ means that the (i,j)-entry of A equals the (i,j)-entry of B. And the (i,j)-entry of B is the ith component of column j of B.

($\Longleftarrow$) Suppose that the jth column of A equals the jth column of B for $j = 1, \ldots, n$. We need to show that the (i,j)-entry of A equals the (i,j)-entry of B for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. The hypothesis implies that the ith component of the jth column of A equals the ith component of the jth column of B for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. And the ith component of the jth column of A is the (i,j)-entry of A, while the ith component of the jth column of B is the (i,j)-entry of B. ∎

1.5. Matrices act on vectors by multiplication.

We are finally ready to think about linear systems again. With

$$A := \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

we are going to define the symbol $A\mathbf{x}$ so that

$$\left\{\begin{array}{ccccc} x_1 & - & 2x_2 & = & 1 \\ 3x_1 & + & 2x_2 & = & 11 \end{array}\right. \iff x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix} \iff A\mathbf{x} = \mathbf{b}.$$

The answer is pretty much staring us in the face: we should put

$$A\mathbf{x} := x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix}.$$

This is something new. This is not a componentwise definition of multiplication. *Instead, the idea behind matrix-vector multiplication is that we take a linear combination of the columns of the matrix weighted by the entries of the vector.* If we write

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}, \qquad \mathbf{a}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad \mathbf{a}_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix},$$

then we are saying

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} := x_1\mathbf{a}_1 + x_2\mathbf{a}_2.$$

We wrote := above to indicate that we were making a definition; until now, we had never specified what the symbol $A\mathbf{x}$ should mean. We will use this := notation often in this course and write $X := Y$ to indicate that we defining a new object X in terms of the hopefully familiar object Y.

We do some computations with this definition of matrix-vector multiplication in words first: take the linear combination of the columns of the matrix with the weights as the entries from the vector, all appearing in order.

1.5.1 Problem (!). Convince yourself that for this to work, the number of columns of the matrix has to equal the number of entries of the vector.

1.5.2 Example. (i) Let

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 3} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^3.$$

Then

$$A\mathbf{v} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0\begin{bmatrix} 3 \\ 4 \end{bmatrix} + 1\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1+0+5 \\ 2+0+6 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}.$$

(ii) Let

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ -1 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

Then

$$A\mathbf{v} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + 4 \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} + \begin{bmatrix} 8 \\ 0 \\ -4 \end{bmatrix} = \begin{bmatrix} 2+8 \\ 0+0 \\ (-2)+(-4) \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \\ -6 \end{bmatrix}.$$

And now for the full definition in symbols.

1.5.3 Definition. *Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{v} \in \mathbb{R}^n$ with*

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

The ***MATRIX-VECTOR PRODUCT*** *of A and $\mathbf{v}$ is*

$$A\mathbf{v} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} := v_1\mathbf{a}_1 + \cdots + v_n\mathbf{a}_n = \sum_{j=1}^{n} v_j\mathbf{a}_j.$$

Again, in words, the matrix-vector product $A\mathbf{v}$ is the linear combination of the columns of the matrix A weighted by the entries of the vector $\mathbf{v}$.

Content from Strang's *ILA* 6E. Page 1 has examples of matrix-vector multiplication.

1.5.4 Problem (!). Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{v} \in \mathbb{R}^n$. Use the definition of $A\mathbf{v}$ from Definition 1.5.3 to count the number of entries in $A\mathbf{v}$.

1.5.5 Problem ($\star$). Let $A \in \mathbb{R}^{m \times n}$. Use the definition of $A\mathbf{0}_n$ from Definition 1.5.3 to show that $A\mathbf{0}_n = \mathbf{0}_m$.

Every linear system compresses as a matrix-vector equation. Suppose there are m equations in n unknowns. Let $\mathbf{x}$ be the column vector of length n that contains all of these unknowns. Let A be the $m \times n$ matrix containing all of the coefficients, so the (i, j)-entry of A is the coefficient on the jth unknown in the ith equation. Let $\mathbf{b}$ be the column vector of length m that contains the right sides of these equations. Then the problem is

$$A\mathbf{x} = \mathbf{b}.$$

This is the *right way* to view systems of linear equations.

1.5.6 Example. Here is a review of how all of this works for our toy problem (1.2.1). We have

$$\begin{cases} x_1 & - & 2x_2 & = & 1 \\ 3x_1 & + & 2x_2 & = & 11 \end{cases} \overset{(1)}{\iff} \begin{bmatrix} x_1 - 2x_2 \\ 3x_1 + 2x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

$$\overset{(2)}{\iff} \begin{bmatrix} x_1 \\ 3x_1 \end{bmatrix} + \begin{bmatrix} -2x_2 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

$$\overset{(3)}{\iff} x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

$$\overset{(4)}{\iff} \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

- Equality (1) is the componentwise definition of vector equality.
- Equality (2) is the componentwise definition of vector addition.
- Equality (3) is the componentwise definition of scalar multiplication.
- Equality (4) is the definition of matrix-vector multiplication.

1.5.7 Problem (!). Rewrite each linear system below as a matrix-vector equation $A\mathbf{x} = \mathbf{b}$ for some matrix $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^n$. Specify the values of m and n in each case.

(i) $\begin{cases} x_1 & + & 2x_2 & & & + & 3x_4 & = & 1 \\ & & & & x_3 & + & 4x_4 & = & 2 \end{cases}$

(ii) $\begin{cases} x_1 & + & 2x_2 & + & x_3 & + & 7x_4 & = & 1 \\ 2x_1 & + & 4x_2 & + & 2x_3 & + & 14x_4 & = & 2 \\ & & & & 2x_3 & + & 8x_4 & = & 3 \end{cases}$

(iii) $\begin{cases} x_1 & & & = & 1 \\ 2x_1 & & & = & 2 \\ & x_2 & & = & 3 \\ & & x_3 & = & 4 \end{cases}$

(iv) $\begin{cases} x_1 & + & 2x_2 & & & = & 4 \\ 2x_1 & + & 4x_2 & & & = & 3 \\ x_1 & + & 2x_2 & + & 2x_3 & = & 2 \\ 7x_1 & + & 14x_2 & + & 8x_3 & = & 1 \end{cases}$

1.5.8 Problem ($\star$). Compute each matrix-vector product and then describe in words the effect of this multiplication. For your description in words, pretend that you are talking out loud to a classmate about this multiplication, and you do not have any paper or board to write on; try to use as few symbols as possible in your description.

(i) $\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

(ii) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ for any c, x_1, x_2, $x_3 \in \mathbb{R}$

(iii) $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ for any x_1, x_2, $x_3 \in \mathbb{R}$

(iv) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ for any x_1, $x_2 \in \mathbb{R}$

This course is called linear algebra. Our equations are linear because the unknowns appear only linearly, as linear powers. This is a static way of viewing our problems. Linearity is also dynamic: this is how matrix-vector multiplication behaves. Specifically, it behaves *linearly*.

What does this mean? Broadly, an operation is linear if the operation applied to a sum of inputs yields the sum of the operation applied to the individual inputs, and if the operation applied to a multiple of an input yields the multiple of the operation applied to that input.

1.5.9 Example. Differentiation and integration in calculus are linear operations.

(i) if f and g are differentiable functions on an interval I and if $c \in \mathbb{R}$, then $f + g$ and cf are differentiable on I, and their derivative satisfy

$$(f+g)' = f' + g' \quad \text{and} \quad (cf)' = cf'.$$

(ii) If f and g are continuous on the interval $[a, b]$ and if $c \in \mathbb{R}$, then

$$\int_a^b (f(x) + g(x))\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx \quad \text{and} \quad \int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx.$$

1.5.10 Example. Squaring a real number is not a linear operation:

$$(x+y)^2 = x^2 + 2xy + y^2 \neq x^2 + y^2$$

if $x \neq 0$ and $y \neq 0$. And likewise

$$(cx)^2 = c^2x^2 \neq cx^2$$

if $c \neq 1$.

1.5.11 Theorem. *Matrix-vector multiplication is* **LINEAR** *in the following sense: if* $A \in \mathbb{R}^{m \times n}$, $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$, *and* $c \in \mathbb{R}$, *then*

$$A(\mathbf{v}+\mathbf{w}) = A\mathbf{v} + A\mathbf{w} \quad \text{and} \quad A(c\mathbf{v}) = c(A\mathbf{v}).$$

Proof. We prove this for $n = 2$ to keep the calculations concrete. We do not need to specify the value of m. So, let $A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}$, $\mathbf{v} = (v_1, v_2)$, and $\mathbf{w} = (w_1, w_2)$. Then $\mathbf{v}+\mathbf{w} = (v_1+w_1, v_2+w_2)$ and $c\mathbf{v} = (cv_1, cv_2)$. We compute

$$\begin{aligned} A(\mathbf{v}+\mathbf{w}) &= \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}\begin{bmatrix}v_1+w_1 \\ v_2+w_2\end{bmatrix} \\ &= (v_1+w_1)\mathbf{a}_1 + (v_2+w_2)\mathbf{a}_2 \\ &= (v_1\mathbf{a}_1 + v_2\mathbf{a}_2) + (w_1\mathbf{a}_1 + w_2\mathbf{a}_2) \\ &= \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}\begin{bmatrix}v_1 \\ v_2\end{bmatrix} + \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}\begin{bmatrix}w_1 \\ w_2\end{bmatrix} \\ &= A\mathbf{v} + A\mathbf{w} \end{aligned}$$

and

$$\begin{aligned} A(c\mathbf{v}) &= \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}\begin{bmatrix}cv_1 \\ cv_2\end{bmatrix} \\ &= (cv_1)\mathbf{a}_1 + (cv_2)\mathbf{a}_2 \\ &= c(v_1\mathbf{a}_1 + v_2\mathbf{a}_2) \\ &= c\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}\begin{bmatrix}v_1 \\ v_2\end{bmatrix}\right) \\ &= c(A\mathbf{v}). \end{aligned}$$

■

Our original questions remain the same—how to solve a linear system, how to understand failure to solve it. The new question right now probably should be *Why is writing a linear system as* $A\mathbf{x} = \mathbf{b}$ *any better than the original way that we wrote it?* This is a valid question, to which we do not have a convincing answer right now, and a major goal of this course is to articulate and defend the payoff of the form $A\mathbf{x} = \mathbf{b}$.

Content from Strang's *ILA* 6E. Read all of p. 2 right now.

1.6. The dot product performs matrix-vector multiplication.

The goal of the course is the same as always: understand, and maybe even solve, the problem $A\mathbf{x} = \mathbf{b}$. Eventually this will take us into understanding just A, apart from any linear systems. For now, we should try to understand $A\mathbf{x}$ as best as we can. There is another way of computing matrix-vector products in addition to Definition 1.5.3. We tease it out in an example.

1.6.1 Example. We compute

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

This is just checking that $x_1 = 3$ and $x_2 = 1$ solves our original problem

$$\begin{cases} x_1 & - & 2x_2 & = & 1 \\ 3x_1 & + & 2x_2 & = & 11 \end{cases}$$

from (1.2.1), right?

Here is another way of looking at this arithmetic:

$$3 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3(1) + 1(-2) \\ 3(3) + 1(2) \end{bmatrix}.$$

Think about how the vectors $(3, 1)$ and $(1, -2)$ appear in the first component on the right. And about how $(3, 1)$ and $(3, 2)$ appear in the second component. One might say that the vector by which we're multiplying the matrixand the rows of the matrix *viewed as column vectors* are doing all of the arithmetic.

We introduce a new structure: the **DOT PRODUCT** of vectors in $\mathbb{R}^2$. Put

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} := v_1 w_1 + v_2 w_2.$$

So we have

$$\begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 1(3) + (-2)(1) = 3 - 2 = 1$$

and

$$\begin{bmatrix} 3 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 3(3) + 2(1) = 9 + 2 = 11.$$

Here is the takeaway in words: we can compute a matrix-vector product by taking the dot product of the rows of the matrix—*viewed as column vectors*—with the vector in the product.

Content from Strang's *ILA* 6E. Equation (1) on p. 9 defines the dot product of vectors in $\mathbb{R}^2$. See the box above on p. 9 for more dot products.

We generalize this example.

1.6.2 Definition. *The dot product of* $\mathbf{v} = (v_1, \ldots, v_n)$, $\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n$ *is*

$$\mathbf{v} \cdot \mathbf{w} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} := v_1 w_1 + \cdots + v_n w_n = \sum_{j=1}^{n} v_j w_j.$$

Content from Strang's *ILA* 6E. This is equation (2) on p. 9. We won't talk about anything else from Section 1.2 for quite a while. The dot product turns out to be the key to a deeper *geometric* understanding of $\mathbb{R}^n$, in particular an understanding of *angles*, but we won't need that for some time.

1.6.3 Example. $\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 3(1) + 4(0) + 5(0) = 3$

Since the dot product in $\mathbb{R}^1 = \mathbb{R}$ is just ordinary multiplication, we can still use the symbol $\cdot$ for multiplication of real numbers if we want.

1.6.4 Problem ($\star$). Prove that the dot product is **COMMUTATIVE** in the sense that $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$ for all $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$. This is how we expect multiplication to behave, that $xy = yx$ for all x, $y \in \mathbb{R}$. (To keep things transparent, do this for $n = 3$.)

We can use the dot product to "extract" components of a vector. This will be a hugely useful operation.

1.6.5 Example. Put

$$\mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{e}_3 := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

These are the **STANDARD BASIS VECTORS** for $\mathbb{R}^3$, and we will use them a lot. We claim that if $\mathbf{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$, then

$$\mathbf{v} \cdot \mathbf{e}_1 = v_1, \qquad \mathbf{v} \cdot \mathbf{e}_2 = v_2, \qquad \text{and} \qquad \mathbf{v} \cdot \mathbf{e}_3 = v_3.$$

We basically did the first equality in Example 1.6.3, so here is the second:

$$\mathbf{v} \cdot \mathbf{e}_2 = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = v_1(0) + v_2(1) + v_3(0) = v_2. \tag{1.6.1}$$

Now here is another nice identity: start with $\mathbf{v}$ and "expand it":

$$\begin{aligned} \mathbf{v} &= \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \\ &= \begin{bmatrix} v_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ v_3 \end{bmatrix} \\ &= v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3 \\ &= (\mathbf{v} \cdot \mathbf{e}_1)\mathbf{e}_1 + (\mathbf{v} \cdot \mathbf{e}_2)\mathbf{e}_2 + (\mathbf{v} \cdot \mathbf{e}_3)\mathbf{e}_3. \end{aligned} \tag{1.6.2}$$

This is a really clean representation of a vector in terms of its components and some other, simpler vectors (and it also solves Problem 1.3.11, right?). Among other things, it tells us that the ith component of $\mathbf{v}$ is $\mathbf{v} \cdot \mathbf{e}_i$. We will return to such representations many times in the future.

The vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ from the previous example are phenomenally useful.

1.6.6 Definition. *The* **STANDARD BASIS VECTORS IN** $\mathbb{R}^n$ *are the vectors* $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ *defined as follows: the components of* $\mathbf{e}_j$ *are all* 0*, except for the component in row* j*, which is* 1.

We will only ever use the symbol $\mathbf{e}_j$ in this course for a standard basis vector. Unfortunately, the notation does not make clear what the value of n is, so $\mathbf{e}_1$ could be $(1, 0)$ or $(1, 0, 0, 0)$. This will usually be clear from context.

1.6.7 Problem ($\star$). **(i)** Write out the standard basis vectors in $\mathbb{R}^5$. You should make clear what all of their entries are.

(ii) Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis vectors in $\mathbb{R}^n$. Show that

$$\mathbf{e}_j \cdot \mathbf{e}_k = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

Your proof should work regardless of the underlying choice of n. This generalizes the calculation in (1.6.1).

(iii) Let $\mathbf{v} \in \mathbb{R}^n$. Prove that

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{e}_1)\mathbf{e}_1 + \cdots + (\mathbf{v} \cdot \mathbf{e}_n)\mathbf{e}_n. \tag{1.6.3}$$

This generalizes the calculation in (1.6.2).

Now that we have an understanding of the mechanics of dot product calculations, we can examine how the dot product arises in matrix-vector multiplication. All of the ideas are in Example 1.6.1. We work with a matrix with three columns to see this a little more abstractly. Let $A \in \mathbb{R}^{m \times 3}$ and write

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ * & * & * \end{bmatrix}.$$

The symbols $*$ below the first row denote the remaining $m - 1$ rows of A, and the exact values of the entries in those rows are wholly unimportant right now. We will often use $*$ to denote parts of a matrix—entries, whole rows, entire columns—whose exact values we can ignore.

Let $\mathbf{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$. Then

$$A\mathbf{v} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ * & * & * \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = v_1 \begin{bmatrix} a_{11} \\ * \end{bmatrix} + v_2 \begin{bmatrix} a_{12} \\ * \end{bmatrix} + v_3 \begin{bmatrix} a_{13} \\ * \end{bmatrix} = \begin{bmatrix} v_1 a_{11} + v_2 a_{12} + v_3 a_{13} \\ * \end{bmatrix}.$$

Again, we are using $*$ in multiple objects to denote components whose exact values we do not need to know. We have shown that the first component of $A\mathbf{v}$ is

$$v_1 a_{11} + v_2 a_{12} + v_3 a_{13} = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} \cdot \mathbf{v},$$

which is the dot product of the first row of A *viewed as a column vector* with $\mathbf{v}$.

This generalizes substantially; the proof is just good bookkeeping and good notation.

1.6.8 Theorem. *Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{v} \in \mathbb{R}^n$. The ith component of $A\mathbf{v}$ is the dot product of row i of A viewed as a column vector and $\mathbf{v}$.*

1.6.9 Example. We compute

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1(1)+4(0)+7(1) \\ 2(1)+5(0)+8(1) \\ 3(1)+6(0)+9(1) \end{bmatrix} = \begin{bmatrix} 8 \\ 10 \\ 12 \end{bmatrix}.$$

What do you get if you use Definition 1.5.3?

Content from Strang's *ILA* 6E. Read about the "row picture" and the "column picture" on p. 19; the matrix is A given on p. 18. Strang says it best: to *compute* $A\mathbf{v}$ by hand for "small" A and $\mathbf{v}$, use dot products, but to *understand* $A\mathbf{v}$, use the "linear combination of columns" definition. (There will be at least one exception to this, when we study orthogonality. I'll bring it up when it arises.) This is morally similar to the derivative: to compute it by hand, use the product rule or chain rule or something like that, but to understand it, use the limit definition.

1.6.10 Problem ($\star$). Go back and redo each of the matrix-vector products in Example 1.5.2 and Problem 1.5.8 with dot products. What do you find easier for work by hand: Definition 1.5.3 or Theorem 1.6.8?

1.7. Matrices are dynamic.

We started thinking about matrices *statically*: they encode data, specifically the coefficients of a linear system of equations. Now that we can multiply matrices and vectors, we can think *dynamically*: matrices act on vectors to produce new vectors. We might even associate a matrix $A \in \mathbb{R}^{m \times n}$ with a "map" ("function"?) that associates each vector $\mathbf{v} \in \mathbb{R}^n$ with a new vector $A\mathbf{v} \in \mathbb{R}^m$.

Matrix-vector multiplication tells us useful things about matrices, not just vectors. First, matrix-vector multiplication can "extract" the columns of a matrix. Start small with $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \in \mathbb{R}^{m \times 3}$. We compute

$$A\mathbf{e}_1 = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1\mathbf{a}_1 + 0\mathbf{a}_2 + 0\mathbf{a}_3 = \mathbf{a}_1 + \mathbf{0}_m + \mathbf{0}_m = \mathbf{a}_1.$$

In words, *multiplying by the first standard basis vector extracts the first column of the matrix.*

1.7.1 Problem (!). With $A \in \mathbb{R}^{m \times 3}$ as above, show that $A\mathbf{e}_2 = \mathbf{a}_2$ and $A\mathbf{e}_3 = \mathbf{a}_3$.

This generalizes nicely.

1.7.2 Theorem. *The jth column of $A \in \mathbb{R}^{m \times n}$ is $A\mathbf{e}_j$.*

1.7.3 Problem (!). Prove it. [Hint: *use the definition of matrix-vector multiplication, not dot products, and the definition of the jth standard basis vector.*]

1.7.4 Example. Let $I_n \in \mathbb{R}^{n \times n}$ be the matrix whose jth column is $\mathbf{e}_j$. That is,

$$I_n := \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

For $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$, we compute

$$I_n \mathbf{v} = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = v_1 \mathbf{e}_1 + \cdots + v_n \mathbf{e}_n = \mathbf{v}.$$

Does this remind you of (1.6.3)? It should.

Because multiplying any vector by I_n just returns that vector, we call I_n the **IDENTITY MATRIX** in $\mathbb{R}^n$.

Content from Strang's *ILA* 6E. Look at the four matrices on p. 18: identity, diagonal, triangular, symmetric. Why are the last three called what they are?

We will often say in this course that *what things do defines what things are*. And what matrices do is multiply vectors! Part (vi) of Undefinition 1.4.1 and Lemma 1.4.6 give "static" ways of viewing matrix equality via data storage: two matrices are equal if all of their corresponding entries (equivalently, all of their corresponding columns) are equal. Here is a "dynamic" on matrix equality: two matrices are equal if they always do the same thing to the same vector. And what matrices do is multiply vectors!

1.7.5 Theorem. *Let A, $B \in \mathbb{R}^{m \times n}$. Then $A = B$ if and only if $A\mathbf{v} = B\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$.*

Proof. ($\Longrightarrow$) Suppose that $A = B$. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix}$. Since $A = B$, Lemma 1.4.6 gives $\mathbf{a}_j = \mathbf{b}_j$ for all j. Let $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$. Then

$$A\mathbf{v} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = v_1\mathbf{a}_1 + \cdots + v_n\mathbf{a}_n = v_1\mathbf{b}_1 + \cdots + v_n\mathbf{b}_n = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = B\mathbf{v}.$$

($\Longleftarrow$) Suppose that $A\mathbf{v} = B\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$, then $A = B$. The key words here are "for all." We can pick any $\mathbf{v} \in \mathbb{R}^n$ that we like, and we will have the equality $A\mathbf{v} = B\mathbf{v}$. If we want to extract data about A and B, there are good, specific choices for $\mathbf{v}$: take $\mathbf{v} = \mathbf{e}_j$ for $j = 1, \ldots, n$. Then $A\mathbf{e}_j = B\mathbf{e}_j$ for each j, and so the jth column of A equals the jth column of B. Since this is true for all j, Lemma 1.4.6 implies $A = B$. ■

1.7.6 Problem ($\star$). A matrix $A \in \mathbb{R}^{n \times n}$ is **DIAGONAL** if $A_{ij} = 0$ for $i \neq j$.

(i) What does an arbitrary diagonal matrix $A \in \mathbb{R}^{4\times 4}$ look like? Write it down. [Hint: *you need at most* 5 *numbers here.*]

(ii) Describe in words the effect of multiplying a vector by a diagonal matrix. What happens to the components of that vector?

1.8. The column space controls the existence of solutions.

Remember that our goal in this course is to understand the problem $A\mathbf{x} = \mathbf{b}$ as best as we can. Our work so far has focused on understanding $A\mathbf{x}$. Now it is time to relate $\mathbf{b}$ to A.

By definition, $A\mathbf{x}$ is a linear combination of the columns of A weighted by the entries of $\mathbf{x}$. To have $A\mathbf{x} = \mathbf{b}$, we therefore want to be able to express $\mathbf{b}$ as a linear combination of the columns of A. The set of all $\mathbf{b}$ that can be expressed in this way has a special name.

1.8.1 Definition. *The* **COLUMN SPACE** *of $A \in \mathbb{R}^{m\times n}$ is the set of all linear combinations of the columns of A. We denote it by $\mathbf{C}(A)$, and every vector in $\mathbf{C}(A)$ is a vector in $\mathbb{R}^m$. Equivalently,*

$$\mathbf{C}(A) := \{A\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^n\}.$$

For a vector $\mathbf{b} \in \mathbb{R}^m$, we write $\mathbf{b} \in \mathbf{C}(A)$ to mean that $\mathbf{b}$ is a vector in $\mathbf{C}(A)$; equivalently, $\mathbf{b}$ has the form $\mathbf{b} = A\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^n$.

We say column *space*, not column *set*. The *set* of columns of $A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix} \in \mathbb{R}^{m\times n}$ is just the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ of at most n vectors (maybe fewer than n, if some of the columns of A are repeated). We will see, and appreciate, how $\mathbf{C}(A)$ is a “dynamic” object that has some more structure than a plain old set of vectors, which is we call it a *space*.

Content from Strang's *ILA* 6E. The column space is defined at the bottom of p. 20.

1.8.2 Example. Let

$$A = \begin{bmatrix}2 & 0\\0 & 3\end{bmatrix}.$$

Then

$$\mathbf{C}(A) = \{A\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^2\} = \left\{v_1\begin{bmatrix}2\\0\end{bmatrix} + v_2\begin{bmatrix}0\\3\end{bmatrix} \;\middle|\; v_1,\ v_2 \in \mathbb{R}\right\} = \left\{\begin{bmatrix}2v_1\\3v_2\end{bmatrix} \;\middle|\; v_1,\ v_2 \in \mathbb{R}\right\}.$$

To be able to solve $A\mathbf{x} = \mathbf{b}$ for as many $\mathbf{b}$ as possible, we want $\mathbf{C}(A)$ to be as “large” as possible. Ideally (perhaps) we would have $\mathbf{C}(A) = \mathbb{R}^m$. What does “=” mean here? (We have only seen “=” for equality of numbers, vectors, and matrices, and equality of the latter two really boiled down to equality of numbers, but now we are talking about the equality of two sets, $\mathbf{C}(A)$ and $\mathbb{R}^m$.) This means that every vector in $\mathbf{C}(A)$ is a vector in $\mathbb{R}^m$ (that's true by definition of $\mathbf{C}(A)$: boring!), and, more excitingly, that every vector in $\mathbb{R}^m$ is a vector in $\mathbf{C}(A)$. So then every $\mathbf{b} \in \mathbb{R}^m$ would be a linear combination of the columns of A, and so we could solve $A\mathbf{x} = \mathbf{b}$ for all $\mathbf{b} \in \mathbb{R}^m$.

1.8.3 Example. With A as in Example 1.8.2, we claim that $\mathbf{C}(A) = \mathbb{R}^2$. We need to take an arbitrary $\mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$ and show $\mathbf{b} \in \mathbf{C}(A)$. That is, we need to find $\mathbf{v} \in \mathbb{R}^2$ such that $A\mathbf{v} = \mathbf{b}$. From Example 1.8.2, it suffices to find v_1, $v_2 \in \mathbb{R}$ such that

$$\begin{bmatrix} 2v_1 \\ 3v_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Looking at componentwise equalities, this is equivalent to

$$2v_1 = b_1 \quad \text{and} \quad 3v_2 = b_2,$$

and that is the same as

$$v_1 = \frac{b_1}{2} \quad \text{and} \quad v_2 = \frac{b_2}{3}.$$

This tells us what $\mathbf{v}$ should be for us to have $\mathbf{b} = A\mathbf{v}$, and we get something more: there is only one way to define $\mathbf{v}$ in terms of $\mathbf{b}$, because there is only one way to define v_1 and v_2 in terms of b_1 and b_2.

This approach to understanding $\mathbf{C}(A)$ told us nothing new about the mechanics of solving $A\mathbf{x} = \mathbf{b}$. In fact, to show that every $\mathbf{b} \in \mathbb{R}^2$ is in $\mathbf{C}(A)$, we just solved $A\mathbf{x} = \mathbf{b}$. With more work, and patience, and maybe some trust, the payoff will be that we can control $\mathbf{C}(A)$ *without explicitly* solving $A\mathbf{x} = \mathbf{b}$.

It is worth being precise about set equality, as we will have to contend with this concept many times in the future.

1.8.4 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be sets of vectors in $\mathbb{R}^m$. We say that $\mathcal{V} = \mathcal{W}$ if $\mathcal{V}$ and $\mathcal{W}$ contain precise the same vectors. That is, if $\mathbf{v} \in \mathcal{V}$, then $\mathbf{v} \in \mathcal{W}$, and if $\mathbf{w} \in \mathcal{W}$, then $\mathbf{w} \in \mathcal{V}$.*

1.8.5 Example. Let

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}.$$

For $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$, we have

$$A\mathbf{v} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2v_1 \\ 0 \end{bmatrix}.$$

So, if $\mathbf{b} = (b_1, b_2) \in \mathbf{C}(A)$, then $b_2 = 0$. Thus $\mathbf{C}(A) \neq \mathbb{R}^2$, as $\mathbf{e}_2 = (0, 1) \in \mathbb{R}^2$ but $\mathbf{e}_2 \notin \mathbf{C}(A)$.

We can be a little more precise about what $\mathbf{C}(A)$ is, rather than what it is not. We showed above that

$$\mathbf{C}(A) = \left\{ \begin{bmatrix} 2v_1 \\ 0 \end{bmatrix} \;\middle|\; v_1 \in \mathbb{R} \right\},$$

and we can be even simpler:

$$\mathbf{C}(A) = \{c\mathbf{e}_1 \mid c \in \mathbb{R}\} =: \mathcal{V}.$$

That is, $\mathbf{C}(A)$ is the set of all scalar multiples of $\mathbf{e}_1$. Here is why, using Definition 1.8.4.

First let $\mathbf{b} \in \mathbf{C}(A)$. Then $\mathbf{b} = (2v_1, 0)$ for some $v_1 \in \mathbb{R}$. This says $\mathbf{b} = 2v_1\mathbf{e}_1 \in \mathcal{V}$, if we take $c = 2v_1$.

Now let $\mathbf{b} \in \mathcal{V}$. Then $\mathbf{b} = c\mathbf{e}_1 = (c, 0)$ for some $c \in \mathbb{R}$. If we take $v_1 = c/2$, then $\mathbf{b} = (2v_1, 0) \in \mathbf{C}(A)$.

1.8.6 Problem ($\star$). **(i)** Prove that

$$\mathbf{C}\left(\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}\right) = \mathbb{R}^2.$$

[Hint: *repeat the work that gave us the equivalent systems in* (1.2.5)*, but instead of having the right side of that system be* (1, 11)*, use an arbitrary* $\mathbf{b} = (b_1, b_2)$.]

(ii) What is

$$\mathbf{C}\left(\begin{bmatrix} 1 & -2 & 4 \\ 3 & 2 & 5 \end{bmatrix}\right)?$$

[Hint: *You know the column space from the previous part, and you know that this column space is the set of all linear combinations of the form*

$$v_1\begin{bmatrix} 1 \\ 3 \end{bmatrix} + v_2\begin{bmatrix} -2 \\ 2 \end{bmatrix} + v_3\begin{bmatrix} 4 \\ 5 \end{bmatrix}.$$

Is there an "easy" value that you can pick for v_3 *to relate this linear combination to what would appear in the previous part?*]

1.8.7 Problem ($\star$). Let $A \in \mathbb{R}^{m \times n}$.

(i) Explain why $\mathbf{0}_m \in \mathbf{C}(A)$. [Hint: *you want to find* $\mathbf{x} \in \mathbb{R}^n$ *such that* $A\mathbf{x} = \mathbf{0}_m$*. Is there an* $\mathbf{x}$ *that immediately comes to mind?*]

(ii) Suppose that $\mathbf{C}(A) = \{\mathbf{0}_m\}$. What does this tell you about A? [Hint: *for any standard basis vector* $\mathbf{e}_j$ *of* $\mathbb{R}^n$*, what do you know about* $A\mathbf{e}_j$*?*]

Failure in math and life teaches us a lot, and there is a lot to be learned from what happens when $\mathbf{C}(A) \neq \mathbb{R}^m$ for $A \in \mathbb{R}^{m \times n}$. Here are some problematic A.

1.8.8 Example.

(i) Earlier we saw that if

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix},$$

then $\mathbf{C}(A)$ is the set of all scalar multiples of $\mathbf{e}_1 = (1,0)$, and that is not $\mathbb{R}^2$.

(ii) Let

$$A = \begin{bmatrix} 1 & -2 \\ 3 & -6 \end{bmatrix}.$$

For v_1, $v_2 \in \mathbb{R}$, we have

$$v_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + v_2 \begin{bmatrix} -2 \\ -6 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} - 2v_2 \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (v_1 - 2v_2) \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

The last equality is true by distribution:

$$c_1\mathbf{v} + c_2\mathbf{v} = (c_1 + c_2)\mathbf{v} \quad \text{for any} \quad c_1,\ c_2 \in \mathbb{R},\ \mathbf{v} \in \mathbb{R}^n.$$

This calculation says that every $\mathbf{b} \in \mathbf{C}(A)$ is a multiple of $(1,3)$. Is every vector in $\mathbb{R}^2$ a multiple of $(1,3)$? Surely not: something like $(0,1)$ cannot be written as

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = c \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

What goes wrong in an equality like that?

(iii) Let

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}.$$

Any $\mathbf{b} = (b_1, b_2, b_3) \in \mathbf{C}(A)$ has $b_3 = 0$; the deadly thing is that row of all 0. To check this, we compute

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 1(v_1) + 0(v_2) + 3v_3 \\ 0(v_1) + 2v_2 + 4v_3 \\ 0(v_1) + 0(v_2) + 0(v_3) \end{bmatrix} = \begin{bmatrix} v_1 + 3v_3 \\ 2v_2 + 4v_3 \\ 0 \end{bmatrix}.$$

What really was going on in the previous example? The rows of zeros in the first and third matrices were problematic, but the column space is about *columns*.

1.8.9 Example. We reexamine the matrices from Example 1.8.8.

(i) Starting with

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

might be easiest. The immediate observation should be that the second column is -2 times the first column.

(ii) This helps us recognize that the second column of

$$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

is 0 times the first column. Even though these matrices have two columns, only one matters—somehow there is "redundant" data in the matrix!

(iii) Is there redundancy in

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}?$$

If so, the redundancy is subtler than the situations of the first two matrices, because no column is a multiple of another. For example, if the second column is a multiple of the first, then

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

Equate the first components to get $c = 0$, so the second column must be the zero vector, which is certainly false. Similar mechanics can help us check the general claim that no column is a multiple of another.

The subtle redundancy is that we can reconstruct the third column from the first two:

$$\begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}.$$

That is, the third column is a linear combination of the first two.

In fact, the third column just disappears when looking at linear combinations of all of the columns:

$$\begin{aligned} v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} &= v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} + v_3 \left(3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \right) \\ &= (v_1 + 3v_3) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + (v_2 + 2v_3) \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}. \end{aligned}$$

And so to describe the column space of this matrix, we do not need to know the third column *at all*. This is redundancy.

1.8.10 Problem (!). Draw pictures in $\mathbb{R}^2$ of the column spaces of the matrices in parts (i) and (ii) of Example 1.8.8. Explain how these pictures illustrate why you cannot "reach" any vector in $\mathbb{R}^2$ via a linear combinations of the columns of these matrices.

1.9. Dependence characterizes redundant data.

Here was the problem with the matrices in Examples 1.8.8 and 1.8.9: one of their columns was a linear combination of the other columns. Informally, from the point of view of the column structure of the matrices, there was redundant data. This is bad by itself, as it means that more data is being stored than necessary. But things worse from the point of view of solving linear systems, as somehow redundant data prevented the column space from being as large as possible. Our job is to understand why.

The problem with the matrices in Examples 1.8.8 and 1.8.9 was twofold. First, these were matrices in $\mathbb{R}^{m \times m}$ but their column spaces were not all of $\mathbb{R}^m$ (so we could not always solve $A\mathbf{x} = \mathbf{b}$ for all $\mathbf{b}$). Second, one of their columns was in the span of the others. Somehow these problems are related. We first give a name to the latter situation and then make a conjecture.

1.9.1 Definition. *The columns of a matrix $A \in \mathbb{R}^{m \times n}$ are* ***DEPENDENT*** *if (at least) one column is a linear combination of the other columns. If $n = 1$ and the matrix only has one column, we say its column is dependent if it is the zero vector. The columns of a matrix are* ***INDEPENDENT*** *if they are not dependent.*

1.9.2 Remark. *We do not say that "the matrix is dependent" or that the column that is a linear combination of the others is "dependent" or "dependent on the other columns." We only talk about the dependence of the columns in totality, relative to the whole matrix.*

The inclusion of the special case of the zero vector when there is only one column (and when it does not make sense to talk about "span of the *others*," because there are no "other" columns when $n = 1$) is a bit of a technicality that will be helpful later. For $n \geq 2$, here is the importance of quantifiers: all that it takes for a matrix to have dependent columns is for *one* column to be "bad." Also, dependence is a relative thing: we talk about vectors being dependent in the context of the rest of the columns of a matrix.

We will mostly focus on the "negative" case of dependence here and later think more positively about independence.

1.9.3 Example.

(i) The columns of

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

are dependent because the third column is the linear combination

$$\begin{bmatrix}3\\4\\0\end{bmatrix} = 3\begin{bmatrix}1\\0\\0\end{bmatrix} + 2\begin{bmatrix}0\\2\\0\end{bmatrix}.$$

(ii) The columns of the matrix

$$\begin{bmatrix}1 & 0\\0 & 2\\0 & 0\end{bmatrix}$$

are not dependent because neither is a linear combination of the other. Here is why: if the first column is a linear combination of the second column, then

$$\begin{bmatrix}1\\0\\0\end{bmatrix} = c\begin{bmatrix}0\\2\\0\end{bmatrix}$$

for some $c \in \mathbb{R}$. But then

$$\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}0\\2c\\0\end{bmatrix},$$

and equating the first components gives $1 = 0$. The same sort of argument shows that the second column is not a linear combination of the first column.

1.9.4 Problem ($\star$). Show that the columns of

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

are dependent. [Hint: $\mathbf{v} = 1\mathbf{v}$. *What does this say about a repeated column being a linear combination of the other columns?*]

1.9.5 Problem ($\star$). Explain why if $A \in \mathbb{R}^{m\times n}$ satisfies one of the following conditions, its columns are dependent.

(i) The same column appears at least twice in A.

(ii) The zero vector (in $\mathbb{R}^m$) is a column of A.

(iii) One column of A is a multiple of another column.

Dependence is exactly the right condition to encode redundancy from the point of view of the column space. If a matrix has dependent columns, then we do not need all of those columns to describe its column space.

1.9.6 Theorem (Removal). *Let $A \in \mathbb{R}^{m \times n}$ with $n \geq 2$ and suppose that one of the columns of A is a linear combination of the others. Let $\widetilde{A} \in \mathbb{R}^{m \times (n-1)}$ contain the $n-1$ columns of A except the one that is the linear combination of the others. Then $\mathbf{C}(A) = \mathbf{C}(\widetilde{A})$.*

Proof. To keep the notation simple, let $n = 3$; this is large enough to show the ideas of the proof but not so large that the notation is *too* cumbersome. Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \in \mathbb{R}^{m \times 3}$. To make the ideas of the proof even more apparent, assume that $\mathbf{a}_3$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. That is, there are $x_1, x_2 \in \mathbb{R}$ such that $\mathbf{a}_3 = x_1\mathbf{a}_1 + x_2\mathbf{a}_2$. Put $\widetilde{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$. We want to show $\mathbf{C}(A) = \mathbf{C}(\widetilde{A})$, and we check the conditions of Definition 1.8.4.

1. *The proof that* $\mathbf{b} \in \mathbf{C}(A) \implies \mathbf{b} \in \mathbf{C}(\widetilde{A})$. If $\mathbf{b} \in \mathbf{C}(A)$, then $\mathbf{b} = A\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^3$. We have

$$\begin{aligned}
\mathbf{b} &= A\mathbf{v} \\
&= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \\
&= v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3\mathbf{a}_3 \\
&= v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3(x_1\mathbf{a}_1 + x_2\mathbf{a}_2) \\
&= v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3x_1\mathbf{a}_1 + c_3x_2\mathbf{a}_2 \\
&= (v_1 + v_3x_1)\mathbf{a}_1 + (v_2 + v_3x_2)\mathbf{a}_2 \\
&= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} v_1 + v_3x_1 \\ v_2 + v_3x_2 \end{bmatrix} \\
&= \widetilde{A} \begin{bmatrix} v_1 + v_3x_1 \\ v_2 + v_3x_2 \end{bmatrix} \in \mathbf{C}(\widetilde{A}).
\end{aligned}$$

2. *The proof that* $\widetilde{\mathbf{b}} \in \mathbf{C}(\widetilde{A}) \implies \widetilde{\mathbf{b}} \in \mathbf{C}(A)$. If $\widetilde{\mathbf{b}} \in \mathbf{C}(\widetilde{A})$, then $\widetilde{\mathbf{b}} = \widetilde{A}\widetilde{\mathbf{v}}$ for some $\widetilde{\mathbf{v}} \in \mathbb{R}^2$. We have

$$\begin{aligned}
\widetilde{\mathbf{b}} &= \widetilde{A}\widetilde{\mathbf{v}} \\
&= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} \widetilde{v}_1 \\ \widetilde{v}_2 \end{bmatrix} \\
&= \widetilde{v}_1\mathbf{a}_1 + \widetilde{v}_2\mathbf{a}_2 \\
&= \widetilde{v}_1\mathbf{a}_1 + \widetilde{v}_2\mathbf{a}_2 + \mathbf{0}_m \\
&= \widetilde{v}_1\mathbf{a}_1 + \widetilde{v}_2\mathbf{a}_2 + 0\mathbf{a}_3 \\
&= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} \widetilde{v}_1 \\ \widetilde{v}_2 \\ 0 \end{bmatrix} \\
&= A \begin{bmatrix} \widetilde{v}_1 \\ \widetilde{v}_2 \\ 0 \end{bmatrix} \in \mathbf{C}(A).
\end{aligned}$$

■

1.9.7 Example. We can "iterate" this theorem several times. Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The zero vector is always a linear combination of any other vectors under consideration:

$$\mathbf{0}_4 = 0\mathbf{a}_2 + 0\mathbf{a}_3 + 0\mathbf{a}_4 + 0\mathbf{a}_5 + 0\mathbf{a}_6.$$

So

$$\mathbf{C}(A) = \mathbf{C}(\begin{bmatrix}\mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5\end{bmatrix}).$$

Next, $\mathbf{a}_3 = 2\mathbf{a}_2 = 2\mathbf{a}_2 + 0\mathbf{a}_4 + 0\mathbf{a}_5$, so

$$\mathbf{C}(\begin{bmatrix}\mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5\end{bmatrix}) = \mathbf{C}(\begin{bmatrix}\mathbf{a}_2 & \mathbf{a}_4 & \mathbf{a}_5\end{bmatrix}).$$

Finally, $\mathbf{a}_5 = 3\mathbf{a}_2 + 4\mathbf{a}_4$, so

$$\mathbf{C}(\begin{bmatrix}\mathbf{a}_2 & \mathbf{a}_4 & \mathbf{a}_5\end{bmatrix}) = \mathbf{C}(\begin{bmatrix}\mathbf{a}_2 & \mathbf{a}_4\end{bmatrix}).$$

Thus

$$\mathbf{C}\left(\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right) = \mathbf{C}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\right).$$

Certainly $\mathbf{a}_2$ is not a multiple of $\mathbf{a}_4$, nor is $\mathbf{a}_4$ a multiple of $\mathbf{a}_2$, so the columns of

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

are not dependent, and therefore they are independent.

Now that we have some experience with dependent columns, we can make the conjecture that Examples 1.8.8 and 1.8.9 motivated.

1.9.8 Conjecture. *If the columns of $A \in \mathbb{R}^{m \times m}$ are dependent, then $\mathbf{C}(A) \neq \mathbb{R}^m$, and so we cannot always solve $A\mathbf{x} = \mathbf{b}$.*

We do not yet have the tools to prove this conjecture. Worse, even if we know that it is true, we probably want a way of verifying that the columns of a matrix are dependent—hopefully a more systematic way than just "getting lucky" and noticing that one column is a linear combination of the others.

1.9.9 Problem (!). We can talk about a nonsquare matrix with dependent columns, but the conjecture was only for a square matrix. Here is why. Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Show that $\mathbf{C}(A) = \mathbb{R}^2$ and that the columns of A are dependent.

Our current definition encapsulates the inefficient redundancies in Examples 1.8.8 and 1.8.9. However, it singles out a particular column for blame. One column has to be "guilty" and expressible as a linear combination of the other columns. Finding that guilty column could be hard if the matrix is large. Fortunately, there is another test for dependence.

1.9.10 Example. Let

$$A = \begin{bmatrix} 1 & -2 \\ 3 & -6 \end{bmatrix}.$$

We studied A in Examples 1.8.8 and 1.8.9. We know that the second column of A is -2 times the first column. With $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$, we have $\mathbf{a}_2 = -2\mathbf{a}_1$. Then

$$\mathbf{0}_2 = 2\mathbf{a}_1 + \mathbf{a}_2 = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = A\mathbf{x}, \qquad \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

We solved the problem $A\mathbf{x} = \mathbf{0}_2$ in an "interesting" way—by finding a nonzero solution $\mathbf{x}$. (Of course $\mathbf{x} = \mathbf{0}_2$ always meets $A\mathbf{x} = \mathbf{0}_2$. Now we have an extra, "nontrivial" solution.

This "nontrivial" solution to the "homogeneous" problem $A\mathbf{x} = \mathbf{0}$ turns out to "characterize" dependence.

1.9.11 Theorem. *The columns of a matrix $A \in \mathbb{R}^{m\times n}$ are dependent if and only if there exists a vector $\mathbf{x} \in \mathbb{R}^n$ such that both $\mathbf{x} \neq \mathbf{0}_n$ and $A\mathbf{x} = \mathbf{0}_m$.*

Proof. This is an "if and only if" statement, so we need to give proofs going in both directions.

($\Longrightarrow$) Start by assuming that the columns of A are dependent. We want to find a solution $\mathbf{x} \neq \mathbf{0}_n$ to $A\mathbf{x} = \mathbf{0}_m$, and we know that one column is a linear combination of the others.

1. Here is how this works for a small matrix. Let

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix} \in \mathbb{R}^{m\times 4} \quad \text{with} \quad \mathbf{a}_3 = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_4\mathbf{a}_4$$

for some x_1, x_2, $x_4 \in \mathbb{R}$. By the way, the number of rows m here is irrelevant.

Now rearrange:

$$\mathbf{0}_m = (x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_4) - \mathbf{a}_3 = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + (-1)\mathbf{a}_3 + x_4\mathbf{a}_4 = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ -1 \\ x_4 \end{bmatrix}.$$

So with $\mathbf{x} = (x_1, x_2, -1, x_4)$, we have $A\mathbf{x} = \mathbf{0}_m$. And we definitely have $\mathbf{x} \neq \mathbf{0}_4$ because of that entry of -1. Maybe $x_1 = x_2 = x_4 = 0$, but still $\mathbf{x} \neq \mathbf{0}_4$.

2. Here is how this works in general. Say that the jth column of A is a linear combination of the other columns. So

$$\mathbf{a}_j = \sum_{\substack{k=1 \\ k \neq j}}^{n} x_k \mathbf{a}_k$$

for some $x_k \in \mathbb{R}$. Then

$$\mathbf{0}_m = \sum_{\substack{k=1 \\ k \neq j}}^{n} x_k \mathbf{a}_k + (-1)\mathbf{a}_j = A\mathbf{x},$$

where $\mathbf{x}$ is the vector whose kth entry is x_k for $k \neq j$, and whose jth entry is -1. That jth entry is nonzero, so $\mathbf{x} \neq \mathbf{0}_n$.

($\Longleftarrow$) Now we assume that $A\mathbf{x} = \mathbf{0}_m$ for some $\mathbf{x} \in \mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{0}_n$. We want to show that (at least) one column of A is a linear combination of the others.

1. Here is how this works for a small matrix. Suppose that $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix} \in \mathbb{R}^{m \times 4}$, $\mathbf{x} = (x_1, x_2, x_3, x_4)$, and $x_2 \neq 0$. Then we have

$$\mathbf{0}_m = A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + x_4\mathbf{a}_4,$$

and so

$$-x_2\mathbf{a}_2 = x_1\mathbf{a}_1 + x_3\mathbf{a}_3 + x_4\mathbf{a}_4.$$

Since $x_2 \neq 0$, we may divide to find

$$\mathbf{a}_2 = \left(-\frac{x_1}{x_2}\right)\mathbf{a}_1 + \left(-\frac{x_3}{x_2}\right)\mathbf{a}_3 + \left(-\frac{x_4}{x_2}\right)\mathbf{a}_4.$$

And so $\mathbf{a}_2$ is a linear combination of the other columns.

2. Here is how this works in general. Let $A\mathbf{x} = \mathbf{0}_m$ and suppose that the jth entry of $\mathbf{x}$ is nonzero. Then

$$\mathbf{0}_m = A\mathbf{x} = \sum_{k=1}^{n} x_k \mathbf{a}_k = x_j\mathbf{a}_j + \sum_{\substack{k=1 \\ k \neq j}}^{n} x_k \mathbf{a}_k,$$

and so the jth column

$$\mathbf{a}_j = \sum_{\substack{k=1 \\ k \neq j}}^{n} \left(-\frac{x_k}{x_j}\right)\mathbf{a}_k$$

is a linear combination of the other columns. ■

What is special here is not that we have a solution to the homogeneous problem $A\mathbf{x} = \mathbf{0}_m$, as we always do ($\mathbf{x} = \mathbf{0}_n$). What is special is that this solution is *not* the zero vector in $\mathbb{R}^n$.

Content from Strang's *ILA* 6E. The blurb on "Independent columns" at the bottom of p. 30 effectively proves this theorem. Forget about the matrix A in that discussion and just think that it's telling you about the (in)dependence of the columns of a matrix C.

This theorem transfers the burden of guilt from a particular column of the matrix to a solution to the matrix-vector equation $A\mathbf{x} = \mathbf{0}_m$. Since this course is all about understanding (and maybe sometimes even solving) $A\mathbf{x} = \mathbf{0}_m$, that should make us happy.

1.10. The null space controls uniqueness of solutions.

The collection of all solutions to $A\mathbf{x} = \mathbf{0}_m$ has a special name and in many ways serves as the moral complement of the column space.

1.10.1 Definition. *The* **NULL SPACE** *of* $A \in \mathbb{R}^{m\times n}$ *is the collection of all solutions to* $A\mathbf{x} = \mathbf{0}_m$*. We denote it by* $\mathbf{N}(A)$*, and every vector in* $\mathbf{N}(A)$ *is a vector in* $\mathbb{R}^n$*. Equivalently,*

$$\mathbf{N}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}_m\}.$$

The null space is "implicitly" defined: vectors in the null space are defined by what they do, but we do not immediately have a formula for them. By contrast, the column space is "explicitly" defined: every vector in $\mathbf{C}(A)$ has the form $A\mathbf{v}$.

Eventually we will develop a detailed procedure for figuring out exactly what vectors are in $\mathbf{N}(A)$. This will be part of our broader program of developing a detailed procedure for solving $A\mathbf{x} = \mathbf{b}$, which should be unsurprising, since studying the null space amounts to studying $A\mathbf{x} = \mathbf{b}$ in the special case of $\mathbf{b} = \mathbf{0}_m$.

1.10.2 Example. $\mathbf{N}(I_n) = \{\mathbf{0}_n\}$ for any n. If $I_n\mathbf{x} = \mathbf{0}_n$, then $\mathbf{x} = \mathbf{0}_n$.

1.10.3 Problem (!). At the very least we always know one vector in the null space. Let $A \in \mathbb{R}^{m\times n}$. Explain why $\mathbf{0}_n \in \mathbf{N}(A)$. Do we have $\mathbf{0}_m \in \mathbf{N}(A)$?

1.10.4 Problem ($\star$). Describe as precisely as possible the null spaces of

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & -2 \\ 3 & -6 \end{bmatrix}.$$

[Hint: *for the first matrix, some of your work in Problem 1.8.6 might help.*]

Now we can recast Theorem 1.9.11 and facts about dependence in the language of the null space.

1.10.5 Corollary. *The columns of a matrix* $A \in \mathbb{R}^{m\times n}$ *are dependent if and only if* $\mathbf{N}(A) \neq \{\mathbf{0}_n\}$*. (Since* $\mathbf{0}_n \in \mathbf{N}(A)$ *always, saying* $\mathbf{N}(A) \neq \{\mathbf{0}_n\}$ *means there is some* $\mathbf{v} \in \mathbf{N}(A)$ *with*

$\mathbf{v} \neq \mathbf{0}_n$.*)*

1.10.6 Problem (!). Prove it.

1.10.7 Problem (!). Let

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}.$$

Find a nonzero vector $\mathbf{x} \in \mathbb{R}^3$ such that $A\mathbf{x} = \mathbf{0}_3$. [Hint: *part (iii) of Example 1.8.9.*]

We know that the column space of A controls existence of solutions to $A\mathbf{x} = \mathbf{b}$: we can solve this problem precisely when $\mathbf{b} \in \mathbf{C}(A)$. We know that $\mathbf{N}(A)$ controls solutions to $A\mathbf{x} = \mathbf{0}_m$, but does $\mathbf{N}(A)$ tell us anything useful about the more general problem $A\mathbf{x} = \mathbf{b}$?

Yes! Remember that when we are solving equations, we want to know about uniqueness along with existence. That is, if given $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, we have a solution $\mathbf{x} \in \mathbb{R}^n$ to $A\mathbf{x} = \mathbf{b}$, is this the only $\mathbf{x}$? Could we have two different choices for $\mathbf{x}$?

Here is the trick. Let $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathbb{R}^n$ solve $A\mathbf{x} = \mathbf{b}$. That is,

$$A\mathbf{x}_1 = \mathbf{b} \quad \text{and} \quad A\mathbf{x}_2 = \mathbf{b}.$$

For solutions to be unique, we want $\mathbf{x}_1 = \mathbf{x}_2$. Is this true? If $\mathbf{x}_1 = \mathbf{x}_2$, then $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0}_n$. What if solutions are not unique? Then $\mathbf{x}_1 \neq \mathbf{x}_2$, and so $\mathbf{x}_1 - \mathbf{x}_2 \neq \mathbf{0}_n$. And what is $\mathbf{x}_1 - \mathbf{x}_2$ doing?

All we know about about $\mathbf{x}_1$ and $\mathbf{x}_2$ is how they behave under multiplication by A, and so we consider how $\mathbf{x}_1 - \mathbf{x}_2$ behaves under multiplication by A. We compute

$$A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0}_m. \tag{1.10.1}$$

Thus $\mathbf{x}_1 - \mathbf{x}_2 \in \mathbf{N}(A)$. And if $\mathbf{x}_1 \neq \mathbf{x}_2$, then $\mathbf{N}(A)$ is bigger than just $\{\mathbf{0}_n\}$. By the way, the first equality here, that of $A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2$, was linearity of matrix-vector multiplication.

This argument started by looking at nonunique solutions to $A\mathbf{x} = \mathbf{b}$ and deduced information about $\mathbf{N}(A)$. Start instead with $\mathbf{N}(A)$ and suppose that it is bigger than just $\{\mathbf{0}_n\}$. Let $\mathbf{z} \in \mathbf{N}(A)$ with $\mathbf{z} \neq \mathbf{0}_n$, so $A\mathbf{z} = \mathbf{0}_m$. Suppose that we also have a solution $\mathbf{x}_\star$ to $A\mathbf{x}_\star = \mathbf{b}$. Finally, let $c \in \mathbb{R}$. Then

$$A(\mathbf{x}_\star + c\mathbf{z}) = A\mathbf{x}_\star + cA\mathbf{z} = \mathbf{b} + c\mathbf{0}_m = \mathbf{b} + \mathbf{0}_m = \mathbf{b}. \tag{1.10.2}$$

Again, as in (1.10.1), the linearity of matrix-vector multiplication makes the first equality possible. For each c, we get a new vector $\mathbf{x}_\star + c\mathbf{z}$. And so from (1.10.2) we get a new solution to $A\mathbf{x} = \mathbf{b}$.

1.10.8 Problem (!). Prove it: if $c_1 \neq c_2$, explain why $\mathbf{x}_\star + c_1\mathbf{z} \neq \mathbf{x}_\star + c_2\mathbf{z}$. [Hint: *what happens if, instead,* $\mathbf{x}_\star + c_1\mathbf{z} = \mathbf{x}_\star + c_2\mathbf{z}$*? Here it is important that* $\mathbf{z} \neq \mathbf{0}_n$.]

These arguments lead to an important result.

1.10.9 Theorem. *Let $A \in \mathbb{R}^{m \times n}$. The following are equivalent.*

(i) *The columns of A are dependent.*

(ii) $\mathsf{N}(A) \neq \{\mathbf{0}_n\}$.

(iii) *There exists $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}_n$ and $\mathbf{x} \neq \mathbf{0}_n$.*

(iv) *If for some $\mathbf{b} \in \mathbb{R}^m$ the problem $A\mathbf{x} = \mathbf{b}$ has a solution, then this solution is not unique.*

The equivalence of the third and fourth parts here illustrate how linearity reduces the question of uniqueness of solutions for the general problem $A\mathbf{x} = \mathbf{b}$ to the very specific homogeneous problem $A\mathbf{x} = \mathbf{0}_m$. Knowledge of just this one problem $A\mathbf{x} = \mathbf{0}_m$ has ramifications for the infinite family of problems $A\mathbf{x} = \mathbf{b}$.

1.11. Independence encodes the essential data.

Our discussion of the null space has focused on (non)uniqueness of solutions to $A\mathbf{x} = \mathbf{b}$, and we have concluded that dependent columns prevent uniqueness of solutions. This still has not resolved Conjecture 1.9.8 on how dependent columns affect *existence*. That will take more work, and the work starts with thinking more positively. The opposite of dependence, which seems to be bad, is independence: the columns of a matrix $A \in \mathbb{R}^{m \times n}$ are independent if they are not dependent. Independence is a "relative" concept—a column is not independent, or dependent, by itself; whether a column is independent or dependent is a matter of how it interacts with all of the other columns.

If "dependent" means "at least one column is a linear combination of the others," then "not dependent" has to mean "no column is a linear combination of the others." And that is what "independent" means. Likewise, if dependence means "one column is guilty" (of being a linear combination of the others, of being linearly redundant), then independence means "all columns are innocent." Either way, that is a lot to check—find one guilty column, or prove that every column is innocent. Fortunately, a restatement of Corollary 1.10.5, that great equalizer, offers a different take.

1.11.1 Corollary. *The columns of a matrix $A \in \mathbb{R}^{m \times n}$ are* ***INDEPENDENT*** *if and only if $\mathsf{N}(A) = \{\mathbf{0}_n\}$, that is, precisely when the only solution to $A\mathbf{x} = \mathbf{0}_m$ is $\mathbf{x} = \mathbf{0}_n$.*

1.11.2 Example. The columns of the $n \times n$ identity matrix $I_n = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix}$ are independent, for if $I_n\mathbf{x} = \mathbf{0}_n$, then since $I_n\mathbf{x} = \mathbf{x}$, we immediately have $\mathbf{x} = \mathbf{0}_n$.

1.11.3 Problem ($\star$). Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$. Prove that the columns of A are independent if and only if whenever $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n = \mathbf{0}_m$, we must have $c_j = 0$ for $j = 1, \ldots, n$. [Hint: *use the definition of matrix-vector multiplication.*]

We can recast Theorem 1.10.9 from the point of view of independence.

1.11.4 Theorem. *Let $A \in \mathbb{R}^{m \times n}$. The following are equivalent.*

(i) *The columns of A are independent.*

(ii) $\mathbf{N}(A) = \{\mathbf{0}_n\}$.

(iii) *The only solution to $A\mathbf{x} = \mathbf{0}_m$ is $\mathbf{x} = \mathbf{0}_n$.*

(iv) *If for some $\mathbf{b} \in \mathbb{R}^m$ the problem $A\mathbf{x} = \mathbf{b}$ has a solution, then this solution is unique.*

1.11.5 Problem (!). Prove it (by heavy reference to Theorem 1.10.9.

Remember that we started the study of dependent columns because we observed that square matrices with dependent columns seemed to have too small of a column space. That led us to make Conjecture 1.9.8. Here is the more upbeat analogue for a square matrix with independent columns.

1.11.6 Conjecture. *If the columns of $A \in \mathbb{R}^{m \times m}$ are independent, then $\mathbf{C}(A) = \mathbb{R}^m$.*

1.11.7 Problem (!). Again, the conjecture is only for square matrices. Explain why the columns of

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

are independent but $\mathbf{C}(A) \neq \mathbb{R}^3$.

As with Conjecture 1.9.8, we do not yet have the tools to prove Conjecture 1.11.6, nor do we have a particularly efficient way of verifying that a matrix's columns are independent beyond the definition. Both conjectures beg the question of which columns in a matrix really matter—which ones are redundant and which ones are essential for describing the column space. This will lead to a more general definition of dependence and independence that can be given beyond the context of matrices. So, what columns really matter?

Content from Strang's *ILA* 6E. Answer: the "independent ones," as alluded to on pp. 20–22. This will require us to broaden the definition of independence to allow only *some* of the columns of the matrix to be independent—that is, some of the columns of a matrix with dependent columns can still be independent, if we define "independent" correctly. Now is also a good time to (re)read pp. v–vii up to, but not including, the $A = CR$ section.

1.12. Lists and spans make language more precise.

To describe what columns really matter, it will help to have a variation on our notions of dependence and independence that frees us from thinking solely about vectors as columns of matrices. Naturally, we start by thinking about vectors as columns of matrices.

1.12.1 Example. We have some experience telling us that the columns of the matrix

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

are dependent—recall part (iii) of Example 1.8.8, part (iii) of Example 1.8.9, part (i) of Example 1.9.3, and Problem 1.10.7. Among other ways to see this, we have $\mathbf{a}_3 = 3\mathbf{a}_1 + 2\mathbf{a}_2$, and from this the removal theorem gives $\mathsf{C}(A) = \mathsf{C}(\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix})$. But we really *need* those columns: $\mathsf{C}(A) \neq \mathsf{C}(\begin{bmatrix} \mathbf{a}_1 \end{bmatrix})$ and $\mathsf{C}(A) \neq \mathsf{C}(\begin{bmatrix} \mathbf{a}_2 \end{bmatrix})$. (Why?)

We also saw in part (ii) of Example 1.9.3 that the columns of

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$$

are independent. It is no accident that we needed *just* these columns to control $\mathsf{C}(A)$, but it turns out that we had other options: $\mathsf{C}(A) = \mathsf{C}(\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_3 \end{bmatrix})$. Unsurprisingly, the columns of $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_3 \end{bmatrix}$ are independent and $\mathsf{C}(A) \neq \mathsf{C}(\begin{bmatrix} \mathbf{a}_3 \end{bmatrix})$.

1.12.2 Problem (!). Fill in the gaps from Example 1.12.1 as follows.

(i) Prove that $\mathsf{C}(A) = \mathsf{C}(\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_3 \end{bmatrix}) = \mathsf{C}(\begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix})$. [Hint: *use the removal theorem.*]

(ii) Prove that the columns of $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$, $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_3 \end{bmatrix}$, and $\begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$ are independent.

(iii) Prove that $\mathsf{C}(A) \neq \mathsf{C}(\begin{bmatrix} \mathbf{a}_1 \end{bmatrix})$, $\mathsf{C}(A) \neq \mathsf{C}(\begin{bmatrix} \mathbf{a}_2 \end{bmatrix})$, and $\mathsf{C}(A) \neq \mathsf{C}(\begin{bmatrix} \mathbf{a}_3 \end{bmatrix})$.

The important columns in A from Example 1.12.1 were $\mathbf{a}_1$ and $\mathbf{a}_2$. (Or maybe $\mathbf{a}_1$ and $\mathbf{a}_3$. Or maybe $\mathbf{a}_2$ and $\mathbf{a}_3$.) We could describe $\mathsf{C}(A)$ using no more, and no less, than these two, and when we stuck them in a matrix, that matrix had independent columns. Something about our language is clunky here—that last sentence. We do not need to "stick vectors in a matrix" to appreciate their independence. Going forward, it will save us some time if we can think about (in)dependence of vectors without introducing a matrix.

1.12.3 Definition. *A list of vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ *is* ***DEPENDENT*** *if one of the following conditions holds:*

(i) $n = 1$ *and* $\mathbf{v}_1 = \mathbf{0}_m$.

(ii) *$n \geq 2$ and (at least) one vector in this list is a linear combination of the others.*

The list is ***INDEPENDENT*** *if it is not dependent.*

We separate the cases $n = 1$ and $n \geq 2$ because if a list has length 1, it does not make sense to say that that one vector is a linear combination of the others in the list—there *are no* other vectors in the list in that case.

1.12.4 Example. In this example we think about the standard basis vectors in $\mathbb{R}^3$.

(i) The list $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_1 + 2\mathbf{e}_2$ is dependent because the third vector is a linear combination of the first two.

(ii) The list $\mathbf{e}_1$, $\mathbf{e}_2$, $2\mathbf{e}_1$ is dependent because the third vector is a linear combination of the first two: $2\mathbf{e}_1 = 2\mathbf{e}_1 + 0\mathbf{e}_2$.

(iii) The list $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_1$ is dependent because the third vector is, again, a linear combination of the first two: $\mathbf{e}_1 = 1\mathbf{e}_1 + 0\mathbf{e}_2$. *Any list with a repeated vector is dependent.*

(iv) The list $\mathbf{e}_1$, $\mathbf{0}_3$, $\mathbf{e}_2$ is dependent because $\mathbf{0}_3 = 0\mathbf{e}_1 + 0\mathbf{e}_2$. *Any list containing the zero vector is dependent.*

There are lots of equivalent ways for a list to be (in)dependent, and the logic is basically the same as with columns of a matrix.

1.12.5 Theorem (Equivalent conditions for a dependent list). *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$. The following are equivalent (so if one of the things below is true, both are true; if one is false, both are false).*

(i) *The list is dependent.*

(ii) *There exist scalars $c_1, \ldots, c_n \in \mathbb{R}$ such that at least one of the scalars is nonzero and $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}_m$.*

1.12.6 Problem (!). Reread the proof of Theorem 1.9.11 and use that proof to explain why Theorem 1.12.5 is true. [Hint: *the list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is dependent if and only if the columns of $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$ are dependent.*]

Here is the happier version, for independence.

1.12.7 Corollary (Equivalent conditions for an independent list). *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$. The following are equivalent (so if one of the things below is true, both are true; if one is false, both are false).*

(i) *The list is independent.*

(ii) *Let $c_1, \ldots, c_n \in \mathbb{R}$ satisfy $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}_m$. Then $c_j = 0$ for $j = 1, \ldots, n$. That is, the only linear combination of the list that adds up to the zero vector is the "trivial" linear combination with weights all zero.*

1.12.8 Example. In this example we think again about the standard basis vectors in $\mathbb{R}^3$.

(i) The list $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ is independent because if $c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3 = \mathbf{0}_3$, then

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

and so $c_1 = c_2 = c_3$.

(ii) We check the (in)dependence of the list $\mathbf{e}_1$, $\mathbf{e}_1 + \mathbf{e}_2$, $\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$. Assume $c_1\mathbf{e}_1 + c_2(\mathbf{e}_1 + \mathbf{e}_2) + c_3(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3) = \mathbf{0}_3$. This reads

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

and that is a system of linear equations (for which we still lack a systematic solution procedure!). But this is not too hard to solve: we want

$$\begin{bmatrix} c_1 + c_2 + c_3 \\ c_2 + c_3 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad (1.12.1)$$

and looking at the third entries, that says $c_3 = 0$, thus $0 = c_2 + c_3 = c_2$, and last $0 = c_1 + c_2 + c_3 = c_1$.

By the way, the *vector* equation (1.12.1) is equivalent to the *matrix-vector* equation

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \mathbf{0}_3.$$

This is the dream: an *upper-triangular* matrix (so-called because below the diagonal all entries are 0) with nonzero entries on the diagonal. Such a system is easy because we can "back-solve" or "back-substitute" starting from the bottom, which we just did ($c_3 = 0$ to $c_2 = 0$ to $c_1 = 0$). Very soon we will develop versatile procedures for converting *any* problem $A\mathbf{x} = \mathbf{b}$ to "upper-triangular" form.

Content from Strang's *ILA* 6E. Pages 116–117 (just read Example 1 on p. 117) define independence and give examples. I know this is a big jump ahead in the book, but you can read it now. Ignore for now the remark at the bottom of p. 116 about the "free variable" and "special solution."

1.12.9 Problem (!). Give answers to the following questions that are a little more direct than just repeating the definition of (in)dependence.

(i) When is a list of length 1 dependent? Independent?

(ii) When is a list of length 2 dependent? Independent?

1.12.10 Problem ($\star$). If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a list in $\mathbb{R}^m$, we'll say that $\mathbf{w}_1, \ldots, \mathbf{w}_n$ is a **REORDERING** of that list if for each j between 1 and n, there is a unique k between 1 and n such that $\mathbf{w}_j = \mathbf{v}_k$. For example, the list $\mathbf{e}_2$, $\mathbf{e}_1$, $\mathbf{e}_3$ is a reordering of $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, but the list $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_2$ is not.

(i) Explain why a list and any reordering of that list have the same span. [Hint: *does the order in which you add vectors ever matter? Nope.*]

(ii) Suppose that a list is dependent. Prove that any reordering of that list is dependent, too. [Hint: *to get a sense of how the argument should go, assume that* $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$ *is dependent, and then explain why* $\mathbf{v}_4$, $\mathbf{v}_3$, $\mathbf{v}_2$, $\mathbf{v}_1$ *is also dependent.*]

(iii) Do the same for independence.

(iv) Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a list in $\mathbb{R}^m$ with $n \le m$, and suppose that the vectors in this list are (some of) the standard basis vectors for $\mathbb{R}^m$. No standard basis vector appears two or more times in this list (and not every standard basis vector has to be in the list: maybe $n < m$). Prove that this list is independent.

We introduced lists so that we can talk about (in)dependence without having to bring up a matrix all the time. In turn, this involves talking a lot about linear combinations of lists.

1.12.11 Definition. *The* ***SPAN*** *of the vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ *is the set of all linear combinations of these vectors, and we denote it by* $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. *If a set* $\mathcal{V}$ *of vectors in* $\mathbb{R}^m$ *equals* $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, *then we say that the list* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ ***SPANS*** $\mathcal{V}$.

Content from Strang's *ILA* 6E. The span of a list of vectors is defined in the box on p. 21.

1.12.12 Example.

(i) Let $\mathbf{v} \in \mathbb{R}^m$. Then $\operatorname{span}(\mathbf{v}) = \{c\mathbf{v} \mid c \in \mathbb{R}\}$. So the span of a single vector is the set of all scalar multiples of that vector.

(ii) With $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ as the standard basis vectors for $\mathbb{R}^3$ (context matters!), Example 1.6.5

gives $\text{span}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \mathbb{R}^3$.

(iii) Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. Then $\mathbf{C}\left(\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}\right) = \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. So every column space is a span, and every span is a column space—but talking about spans cuts down on some chatter and avoids unnecessarily introducing a matrix just to have a column space.

1.12.13 Problem (!). Prove that $\text{span}(\mathbf{0}) = \{\mathbf{0}\}$. That is, the only vector in the span of $\mathbf{0}$ is $\mathbf{0}$ itself.

1.12.14 Problem ($\star$). Let $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$.

(i) Prove that $\mathbf{0}_n \in \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$.

(ii) Prove that $\mathbf{v}_j \in \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$ for each $j = 1, \ldots, p$. [Hint: *make most coefficients* 0 *and one special coefficient* 1.]

(iii) Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{v} \in \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$. Prove that $A\mathbf{v} \in \text{span}(A\mathbf{v}_1, \ldots, A\mathbf{v}_p)$.

1.13. Independence, lists, and spans are friends.

There is another way to check for independence that works well by hand for "small" matrices and that more generally reinforces the underlying idea of linear combinations. The statement of this independence test is bit technical, so we will do some examples before the proof.

1.13.1 Lemma (Linear independence). *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a list in $\mathbb{R}^m$. The following are equivalent.*

(i) *The list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is independent.*

(ii) *The following two conditions hold. First, $\mathbf{v}_1 \neq \mathbf{0}_m$. Second, if $n \geq 2$, then for $j = 2, \ldots, n$, the jth vector in the list is not a linear combination of the first $j-1$ vectors. In more symbols and fewer words, this means $\mathbf{v}_1 \neq \mathbf{0}_m$ and, if $n \geq 2$, then $\mathbf{v}_j \notin \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1})$ for $j = 2, \ldots, n$.*

This result is often called the "linear independence lemma," although there certainly are plenty of other tests for linear independence. By violating either of the two conditions in the second part, we get linear dependence, so we could just as well call this the "linear dependence" lemma.

1.13.2 Example. Let

$$\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix}2\\3\\0\end{bmatrix}, \qquad \text{and} \qquad \mathbf{v}_3 = \begin{bmatrix}4\\5\\6\end{bmatrix}.$$

Certainly

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \neq \mathbf{0}_3.$$

Next, we want to check that $\mathbf{v}_2 \notin \text{span}(\mathbf{v}_1)$. Otherwise, we would have

$$\begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = c \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

for some $c \in \mathbb{R}$. Equating the second components, this would mean $3 = 0$. Last, we want to check that $\mathbf{v}_3 \notin \text{span}(\mathbf{v}_1, \mathbf{v}_2)$. Otherwise, we would have

$$\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}.$$

Equating the third components, we would have $6 = 0$. And so the list $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ is independent.

1.13.3 Problem (!). Use Lemma 1.13.1 to explain why the columns of the matrix

$$\begin{bmatrix} 1 & 1 & 2 & 3 \\ 1 & 0 & 2 & 4 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

are dependent.

1.13.4 Problem ($\star$). Give an example of a dependent list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ with $n \geq 2$ such that $\mathbf{v}_j \notin \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1})$ for $2 \leq j \leq n$.

Content from Strang's *ILA* 6E. Reread p. 20, this time paying attention to dependence and independence. Then work through the (in)dependence tests on p. 21 for the matrices A_4 and A_5. There is one thing here that we have not yet discussed: what does it mean for only "some" of the columns of A to be independent?

Now we prove the linear independence lemma.

Proof (Of Lemma 1.13.1). Proofs involving linear (in)dependence often work best when done by contradiction: what goes wrong if the result that we want to be true is false?

($\Longrightarrow$) First suppose that the list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is independent. We want to prove that the second, technical part of the lemma is true.

Since the list is independent, no vector can be the zero vector, or the list would be dependent. In particular, $\mathbf{v}_1 \neq \mathbf{0}_m$. Then if $n = 1$, we are done. Otherwise, suppose $n \geq 2$. If the jth vector for some $j \geq 2$ is a linear combination of the previous $j-1$ vector, then that vector is a linear combination of *all of the other* vectors in the list. Just put the weights on the remaining vectors to be zero. And then the list is dependent, which is wrong.

1. Here is how this works for a short list. Say $n = 4$ and $\mathbf{v}_3 \in \text{span}(\mathbf{v}_1, \mathbf{v}_2)$. Then $\mathbf{v}_3 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ for some c_1, $c_2 \in \mathbb{R}$. But then by the great trick of adding zero,

$$\mathbf{v}_3 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \mathbf{0}_m = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + 0\mathbf{v}_4,$$

and so $\mathbf{v}_3$ is in the span of the other vectors in the list. This is linear dependence.

We also want to rule out the cases $\mathbf{v}_2 \in \text{span}(\mathbf{v}_1)$ and $\mathbf{v}_4 \in \text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Can you adapt the argument in the preceding paragraph to show why either of these cases imply dependence?

2. Here is how this works in general. Say that for some $j \geq 2$ we have $\mathbf{v}_j \in \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1})$. So

$$\mathbf{v}_j = c_1\mathbf{v}_1 + \cdots + c_{j-1}\mathbf{v}_{j-1}$$

for some $c_1, \ldots, c_{j-1} \in \mathbb{R}$. Rearrange this to read

$$c_1\mathbf{v}_1 + \cdots + c_{j-1}\mathbf{v}_{j-1} + (-1)\mathbf{v}_j = \mathbf{0}_m.$$

If $j = n$, then the list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is dependent. If $j < n$, add zero to find

$$c_1\mathbf{v}_1 + \cdots + c_{j-1}\mathbf{v}_{j-1} + (-1)\mathbf{v}_j + 0\mathbf{v}_{j+1} + \cdots + 0\mathbf{v}_n = \mathbf{0}_m,$$

and so the list is again dependent.

($\Longleftarrow$) Now suppose that the technical condition in the second part of the lemma holds. We need to prove that the list is independent. If the list has only one vector, then that vector is nonzero, so we are done. Say that the list has at least two vectors. One way to prove independence is to assume $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}_m$ and show $c_j = 0$ for all j. What goes wrong if at least one coefficient is nonzero?

If $c_2 = \cdots = c_n = 0$, then we get $c_1\mathbf{v}_1 = \mathbf{0}_m$. Since at least one coefficient has to be nonzero, that has to be c_1 here: $c_1 \neq 0$. Then $\mathbf{v}_1 = \mathbf{0}_m$, a contradiction.

Otherwise, suppose that a coefficient with index 2 or higher is nonzero (maybe there multiple nonzero coefficients). The trick is to look at the *highest-indexed* coefficient that is nonzero.

1. Say that $n = 4$ for simplicity. What if $c_4 \neq 0$? Then we have $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + c_4\mathbf{v}_4 = \mathbf{0}_m$, and this rearranges to give

$$\mathbf{v}_4 = \left(-\frac{c_1}{c_4}\right)\mathbf{v}_1 + \left(-\frac{c_2}{c_4}\right)\mathbf{v}_2 + \left(-\frac{c_3}{c_4}\right)\mathbf{b}_3 \in \text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3).$$

This is a contradiction.

Now what if $c_4 = 0$ but $c_3 \neq 0$? Then $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}_3$, and, as above,

$$\mathbf{v}_3 = \left(-\frac{c_1}{c_3}\right)\mathbf{v}_1 + \left(-\frac{c_2}{c_3}\right)\mathbf{v}_2 \in \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2).$$

That is a contradiction.

Do you see what to do if $c_3 = c_4 = 0$ but $c_2 \neq 0$?

2. Here is how this works in general. Let j be the largest index such that $c_j \neq 0$. (In the case above, we had $j = 4 = n$, but maybe $j < n$.) If $j = 1$, this means that $c_2 = 0, \ldots, x_n = 0$, and then $\mathbf{0}_m = c_1\mathbf{v}_1$. Since $c_1 \neq 0$, we get $\mathbf{v}_1 = \mathbf{0}_m$, a contradiction. Otherwise, if $2 \leq j \leq n$, then we rearrange

$$\mathbf{0}_m = c_1\mathbf{v}_1 + \cdots + c_{j-1}\mathbf{v}_{j-1} + c_j\mathbf{v}_j$$

into

$$\mathbf{v}_j = \left(-\frac{c_1}{c_j}\right)\mathbf{v}_1 + \cdots + \left(-\frac{c_{j-1}}{c_j}\right)\mathbf{v}_{j-1} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}),$$

yet another contradiction. ■

We could view the linear independence lemma as a "sweeping" tool: look at the list from left to right. If the first vector is nonzero, keep going. If each successive vector is not in the span of its predecessors, keep going. Either we run out of vectors, and the list is independent, or, somewhere sweeping through the list, we find that one vector *is* in the span of its predecessors, you have a dependent list. The linear independence lemma makes dependence a little more algorithmic: to find the "guilty" vector that is a combination of the others, sweep the list from left to right until we have the vector that is a combination of its predecessors (and therefore a combination of everything else in the list by weighting some vectors with coefficients of zero, as needed).

Easier said than done, perhaps. How do we know that $\mathbf{v}_j \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1})$? We need to find coefficients $c_1, \ldots, c_{j-1} \in \mathbb{R}$ such that $\mathbf{v}_j = c_1\mathbf{v}_1 + \cdots + c_{j-1}\mathbf{v}_{j-1}$. Equivalently, we need to solve $c_1\mathbf{v}_1 + \cdots + c_{j-1}\mathbf{v}_{j-1} + (-1)\mathbf{v}_j = \mathbf{0}_m$. This is a linear system of equations:

$$\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_{j-1} & \mathbf{v}_j \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_{j-1} \\ -1 \end{bmatrix} = \mathbf{0}_m.$$

And we still do not have an algorithm for *solving* this problem—just lots of tools for *understanding* it.

The linear independence lemma also allows us to winnow down a list to its essential components from the point of view of spans. The following concept and lemma yield the list analogue of (repeated iterations of) the removal theorem for matrices (Theorem 1.9.6).

1.13.5 Problem ($\star$). If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a list in $\mathbb{R}^m$, and if $p \leq n$, then a **SUBLIST** of the original list is a list of the form $\mathbf{v}_{j_1}, \ldots, \mathbf{v}_{j_p}$, where $1 \leq j_1 < j_2 < \cdots < j_p \leq n$. This is a painful definition, so here is an example. Start with the list $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$. Then $\mathbf{v}_2$, $\mathbf{v}_4$ is a sublist, as is the list $\mathbf{v}_1$ with one entry, but the list $\mathbf{v}_1$, $\mathbf{v}_1$, $\mathbf{v}_2$ is not a sublist, nor is the list $\mathbf{v}_2$, $\mathbf{v}_1$.

(i) Suppose that a list is independent. Prove that any sublist is independent, too. [Hint: *if the list* $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ *is independent and the sublist is* $\mathbf{v}_1$, $\mathbf{v}_2$, *you want to show that if* $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$, *then* $c_1 = c_2 = 0$. *But you only know stuff about linear combinations of the whole list. How can you get* $\mathbf{v}_3$ *to show up in* $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$? *What is the right coefficient to slap on* $\mathbf{v}_3$?]

(ii) Suppose that a list contains a dependent sublist. Prove that the whole list is dependent, too.

(iii) Is every sublist of a dependent list always dependent, too?

1.13.6 Lemma. *Let* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *be a list in* $\mathbb{R}^m$, *and suppose there is at least one nonzero vector in the list. Then this list has an independent sublist with the same span: there is a sublist* $\mathbf{v}_{j_1}, \ldots, \mathbf{v}_{j_r}$ *of* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *such that* $\operatorname{span}(\mathbf{v}_{j_1}, \ldots, \mathbf{v}_{j_r}) = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

Proof (of Lemma 1.13.6). We reduce the list as follows. Let $\mathbf{v}_{j_1}$ be the first nonzero vector in the list. (At least one exists.) So $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j_1}) = \operatorname{span}(\mathbf{v}_{j_1})$. Also, the list $\mathbf{v}_{j_1}$ is independent because $\mathbf{v}_{j_1} \neq \mathbf{0}$.

If $j_1 = n$ or if $\mathbf{v}_j \in \operatorname{span}(\mathbf{v}_{j_1})$ for $j > j_1$, stop. Otherwise, let $j_2 > j_1$ such that $\mathbf{v}_{j_2}$ is the first vector in the list that is a multiple of $\mathbf{v}_{j_1}$, i.e., $\mathbf{v}_{j_2} \notin \operatorname{span}(\mathbf{v}_{j_1})$. So $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j_2}) = \operatorname{span}(\mathbf{v}_{j_1}, \mathbf{v}_{j_2})$. Also, the list $\mathbf{v}_{j_1}$, $\mathbf{v}_{j_2}$ is independent because $\mathbf{v}_{j_1} \neq \mathbf{0}$ and $\mathbf{v}_{j_2} \notin \operatorname{span}(\mathbf{v}_{j_1})$.

If $j_2 = n$ or if $\mathbf{v}_j \in \operatorname{span}(\mathbf{v}_{j_1}, \mathbf{v}_{j_2})$ for $j > j_2$, stop. Otherwise, let $j_3 > j_2$ such that $\mathbf{v}_{j_3}$ is the first vector in the list that is not in $\operatorname{span}(\mathbf{v}_{j_1}, \mathbf{v}_{j_2})$. So $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j_3}) = \operatorname{span}(\mathbf{v}_{j_1}, \mathbf{v}_{j_2}, \mathbf{v}_{j_3})$. And the list $\mathbf{v}_{j_1}$, $\mathbf{v}_{j_2}$, $\mathbf{v}_{j_3}$ is independent since $\mathbf{v}_{j_3} \notin \operatorname{span}(\mathbf{v}_{j_1}, \mathbf{v}_{j_2})$.

Now turn the crank and keep going: eventually we run out of vectors in the list. More precisely, either $\mathbf{v}_n$ is not in the span of the previous vectors in the list, and we take $j_r = n$, or for some $j_r < n$, all of the vectors $\mathbf{v}_{j_r+1}, \ldots, \mathbf{v}_n$ are in the span of the sublist stopping with $\mathbf{v}_{j_r}$. ■

1.13.7 Problem (!). Go back to Example 1.9.7 and think of the columns of the matrix there as the original list (so $n = 5$). Convince yourself that in the notation of Lemma 1.13.6, the results of Example 1.9.7 gave $r = 2$, $j_1 = 2$, and $j_2 = 4$.

Content from Strang's *ILA* 6E. Think once more about the matrices A_1 through A_5 on pp. 20–21. Apply the algorithm in Lemma 1.13.6 to extract the linearly independent columns that span the column spaces.

1.13.8 Remark. *Reading this is probably not a good use of time.*

(i) *We keep using the word "list," but we never defined it. A list is not quite a set: the* *list* $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_1$ *in* $\mathbb{R}^3$ *has three entries, although the third repeats the first (so the list is what, independent or dependent?). But the set* $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1\}$ *is the same as* $\{\mathbf{e}_1, \mathbf{e}_2\}$*, because repeating an element when describing a set doe not yield a different set. Also,* $\{\mathbf{e}_1, \mathbf{e}_2\} = \{\mathbf{e}_2, \mathbf{e}_1\}$*, because changing the order of elements when describing a set does not change that set. Thus*

$$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1\} = \{\mathbf{e}_1, \mathbf{e}_2\} = \{\mathbf{e}_2, \mathbf{e}_1\}.$$

So the concept of a list should both allow repetition and encode order: when we talk about the columns of the matrix $\begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_1\end{bmatrix}$*, we want to respect their order:*

$$\begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_1\end{bmatrix} \neq \begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_1\end{bmatrix} \neq \begin{bmatrix}\mathbf{e}_2 & \mathbf{e}_1\end{bmatrix}.$$

(ii) *Perhaps the most precise (which is not the same as most useful) way to think of a list is as a set of ordered pairs: the list* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *is the set* $\{(k, \mathbf{v}_k)\}_{k=1}^n$*. This encodes order (*$\mathbf{v}_1$ *comes before* $\mathbf{v}_2$*) and allows repetition: the list* $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_1$ *is the set* $\{(1, \mathbf{e}_1), (2, \mathbf{e}_2), (3, \mathbf{e}_1)\}$*, which is not the same as the list* $\{(1, \mathbf{e}_1), (2, \mathbf{e}_2)\}$ *or the list* $\{(1, \mathbf{e}_2), (2, \mathbf{e}_1)\}$*.*

(iii) *It might be preferable to write a list enclosed in parentheses: instead of talking about the list* $\mathbf{v}_1, \ldots, \mathbf{v}_n$*, we would talk about the list* $(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \{(k, \mathbf{v}_k)\}_{k=1}^n$*. This could lead to a weird overworking of ordered pairs in the case of a list of length 2:* $(\mathbf{v}_1, \mathbf{v}_2) = \{(1, \mathbf{v}_1), (2, \mathbf{v}_2)\}$*. And a list of length 1 would be* $(\mathbf{v}_1) = \{(1, \mathbf{v}_1)\}$*. The upshot, though, is that drilling into what it means for two sets to be equal, we have* $(\mathbf{v}_1, \ldots, \mathbf{v}_n) = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ *if and only if* $n = m$ *and* $\mathbf{v}_k = \mathbf{w}_k$ *for* $k = 1, \ldots, n$*; that is, two lists are equal precisely when they have the same length and when the corresponding entries are equal.*

(iv) *This is all very similar to how we talked about vectors as sets of ordered pairs in Remark 1.3.12 and even more similar to one of the definitions of a matrix that we attempted in Remark 1.4.4. In fact, the former remark suggested defining a vector* $\mathbf{v} = (v_1, \ldots, v_n)$ *as a list in* $\mathbb{R}$ *(which means that a list of vectors would be a list of lists!), and the latter remark suggested defining the matrix* $A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix} \in \mathbb{R}^{m \times n}$ *as* $A = \{(k, \mathbf{a}_k)\}_{k=1}^n$*, and by the definition of list above, that would have* $A = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$*. That is, a matrix would just be the list of its columns…*

1.14. The pivot columns are the essential data of a matrix.

We have seen several times that we may not need all of the columns of a matrix to span the column space, just some, but that there is a threshold below which we cannot dip—we seem to need a certain *minimum number* of columns to span the column space. That number seems to be connected to the minimum length of an independent list that spans the column space.

1.14.1 Example. We studied the matrix

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{e}_1 & 2\mathbf{e}_1 & \mathbf{e}_2 & (3\mathbf{e}_1 + 4\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5 \end{bmatrix}.$$

in Example 1.9.7 and saw that $\mathbf{C}(A) = \text{span}(\mathbf{a}_2, \mathbf{a}_4)$ and that the list $\mathbf{a}_2$, $\mathbf{a}_4$ is independent. We might also notice that any independent list of columns of A of length 2 also spans $\mathbf{C}(A)$ (there are several such lists, like $\mathbf{a}_2$, $\mathbf{a}_5$), no list of columns of length 1 spans $\mathbf{C}(A)$ (try it: there are only five such lists), and any list of columns of length 3 or greater will be dependent (there are many such lists, so just look at a few examples to be convinced). Something similar happened in Problem 1.12.2 for a different matrix.

But there is something special about the list $\mathbf{a}_2$, $\mathbf{a}_4$: these are the "first" independent columns of A to appear going from left to right. This is how we worked in Example 1.9.7, and this is also how the linear independence lemma operates.

More precisely, here is what we should appreciate about the list $\mathbf{a}_2$, $\mathbf{a}_4$.

- $\mathbf{a}_2$ is the first nonzero column of A.

- $\mathbf{a}_3 \in \text{span}(\mathbf{a}_2)$.

- $\mathbf{a}_4$ is the first column in A not in $\text{span}(\mathbf{a}_2)$. That is, $\mathbf{a}_4 \notin \text{span}(\mathbf{a}_2)$ but $\mathbf{a}_j \in \text{span}(\mathbf{a}_2)$ for $j = 1, 2, 3$.

- $\mathbf{a}_5 \in \text{span}(\mathbf{a}_2, \mathbf{a}_4)$.

- And after that there are no more columns of A. Also, every column of A is in $\text{span}(\mathbf{a}_2, \mathbf{a}_4)$, and so from the point of view of the column space, only 40% of the original data of A was important.

Here is the abstraction of the situation with the special list $\mathbf{a}_2$, $\mathbf{a}_4$ in the previous example.

1.14.2 Definition. *Let $A \in \mathbb{R}^{m \times n}$ be a nonzero matrix. (This means that at least one entry of A is not the scalar 0, and so at least one column of A is not the zero vector $\mathbf{0}_m$.) Define a list of columns of A recursively as follows.*

1. *The first entry in this list is the first nonzero column of A.*

2. *If every column of A is a multiple of the first nonzero column of A, then this list has only one entry: the first nonzero column of A.*

3. *Otherwise, keep going: for $i \geq 2$, the ith column in this list is the first column of A that is not in the span of the previous $i - 1$ columns in this list.*

The columns in this list are the **PIVOT COLUMNS** *of A, and the length of this list is the* **RANK** *of A, denoted* $\text{rank}(A)$. *We define the rank of the zero matrix to be 0.*

1.14.3 Remark. *In much more precise, but painful, notation, the list of pivot columns* $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}$ *of* A *satisfies the following.*

(i) *The first pivot column is the first nonzero column of* A*: if* $j < j_1$*, then* $\mathbf{a}_j = \mathbf{0}_m$*.*

(ii) *The first pivot column is the first nonzero column of* A*:* $\mathbf{a}_{j_1} \neq \mathbf{0}_m$*.*

(iii) *If there are at least two pivot columns, then the* i*th pivot column is not in the span of the previous* $i-1$ *pivot columns: if* $r \geq 2$ *and* $i \geq 2$*, then* $\mathbf{a}_{j_i} \notin \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$*.*

(iv) *If there are at least two pivot columns, then the* i*th pivot column is the first column of* A *not in the span of the previous* $i-1$ *pivot columns: if* $r \geq 2$*,* $i \geq 2$*, and* $j < j_i$*, then* $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$*, and if* $j > j_r$*, then* $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$*.*

1.14.4 Example. The pivot columns of

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

are $\mathbf{a}_2$, $\mathbf{a}_4$, and $\operatorname{rank}(A) = 2$.

1.14.5 Example. Each column of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$$

is a multiple of the first, and the first column is nonzero, so $\operatorname{rank}(A) = 1$ and the list of pivot columns of A is just $\mathbf{a}_1$. In general in a rank-1 matrix, every column is a multiple of just one other column (maybe not the first, if the first column is zero).

1.14.6 Problem (!). Give an example of a matrix $A \in \mathbb{R}^{m \times n}$ with $n \geq 2$ such that $\operatorname{rank}(A) = 1$ but not every column of A is a multiple of the first column.

Content from Strang's *ILA* 6E. Work through the example on p. 23 with the matrix A_6. We won't talk about this for some time in class, but the "row rank = column rank" calculations for 2×2 and 3×3 rank-1 matrices are good practice, so check the details yourself.

1.14.7 Example. Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$

The first pivot column of A is $\mathbf{a}_2$ since $\mathbf{a}_1 = \mathbf{0}_3$. Since $\mathbf{a}_3 = 2\mathbf{a}_2$ but $\mathbf{a}_4 \notin \text{span}(\mathbf{a}_2)$, the second pivot column of A is $\mathbf{a}_4$. (More precisely, if $\mathbf{a}_4 = c\mathbf{a}_2$ for some $c \in \mathbb{R}$, look at the third components to get $0 = 2$.) Finally, it turns out that $\mathbf{a}_5 = 3\mathbf{a}_2 + 4\mathbf{a}_4$, so $\mathbf{a}_5 \in \text{span}(\mathbf{a}_2, \mathbf{a}_4)$, and therefore $\mathbf{a}_5$ is not a pivot column of A. One way to see this is to set up the system $c_1\mathbf{a}_2 + c_2\mathbf{a}_4 = \mathbf{a}_5$ and grind out $c_1 = 3$ and $c_2 = 4$. Thus the pivot columns of A are the list $\mathbf{a}_2$, $\mathbf{a}_4$, and so $\text{rank}(A) = 2$.

This work shows that every column of A is in the span of the pivot columns. In particular, the relationships among the columns of A here are exactly the same as they were in Example 1.14.1, except there the matrix had more zero entries and so was simpler. To recap:

- Part (i) of Problem 1.12.14 gives $\mathbf{a}_1 \in \text{span}(\mathbf{a}_2, \mathbf{a}_4)$.
- Part (ii) of Problem 1.12.14 gives $\mathbf{a}_2$, $\mathbf{a}_4 \in \text{span}(\mathbf{a}_2, \mathbf{a}_4)$.
- We just saw $\mathbf{a}_3 \in \text{span}(\mathbf{a}_2)$, specifically with $\mathbf{a}_3 = 2\mathbf{a}_2$, so

$$\mathbf{a}_3 = 2\mathbf{a}_2 + \mathbf{0}_3 = 2\mathbf{a}_2 + 0\mathbf{a}_4 \in \text{span}(\mathbf{a}_2, \mathbf{a}_4).$$

- And we also saw (or at least claimed) that $\mathbf{a}_5 \in \text{span}(\mathbf{a}_2, \mathbf{a}_4)$.

The removal theorem, going exactly as we did in Example 1.9.7, then tells us $\mathbf{C}(A) = \text{span}(\mathbf{a}_2, \mathbf{a}_4)$. *The pivot columns span the column space of A.*

1.14.8 Problem (!). Let

$$A = \begin{bmatrix} 2 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 3 & 6 & 6 \\ 0 & 4 & 8 & 8 \end{bmatrix}.$$

Find the pivot columns of A and calculate $\text{rank}(A)$.

1.14.9 Problem (!). What is the rank of a diagonal matrix? (Say something more profound than "the number of pivot columns.")

We are observing patterns in these examples, and right now we do have the tools to confirm at least some of them.

1.14.10 Theorem. *Let $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) \geq 1$. Then the list of pivot columns of A is independent and spans $\mathbf{C}(A)$.*

Proof. Call this list $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}$.

1. *The proof of independence.* If $r = 1$, then this list has only one entry: $\mathbf{a}_{j_1} \neq \mathbf{0}_m$, and so this list is independent. Otherwise, if $r \geq 2$, then $\mathbf{a}_{j_1} \neq \mathbf{0}_m$ and, for $i \geq 2$,

$\mathbf{a}_{j_i} \notin \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$. The linear independence lemma therefore implies that the list is independent.

2. *The proof of spanning.* This is the removal theorem iterated to remove any column of A that is not a pivot column. ■

Out of these examples we distill the following conjecture as to how rank is a "threshold" for controlling the column space efficiently.

1.14.11 Conjecture. *Let $A \in \mathbb{R}^{m \times n}$ with $r := \operatorname{rank}(A) \geq 1$.*

(i) *Any independent list of length r in $\mathbf{C}(A)$ spans $\mathbf{C}(A)$.*

(ii) *No list of less than length r in $\mathbf{C}(A)$ can span $\mathbf{C}(A)$.*

(iii) *Any list of length greater than r in $\mathbf{C}(A)$ is dependent.*

Part of developing the tools to prove this conjecture will be developing a mechanism for *easily identifying* the pivot columns of a matrix. Right now it is very much a column-by-column process.

1.14.12 Problem (!). Let $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(A) = r \geq 1$. We are conjecturing that a list of length greater than r in $\mathbf{C}(A)$ is always dependent. Is a list of length r or fewer vectors in $\mathbf{C}(A)$ always independent?

1.14.13 Problem ($\star$). Let $A \in \mathbb{R}^{m \times n}$.

(i) Explain why $\operatorname{rank}(A) \leq n$. Eventually we will show that $\operatorname{rank}(A) \leq m$, too.

(ii) Suppose that $\operatorname{rank}(A) = n$. Prove that the columns of A are independent.

(iii) Suppose that the columns of A are independent. If $n = 1$, what goes wrong if $\operatorname{rank}(A) = 0$? If $n \geq 2$ and $\operatorname{rank}(A) < n$, explain why a column of A must be a linear combination of the other columns. Why is that a problem? Conclude that if the columns of A are independent, then $\operatorname{rank}(A) = n$.

(iv) Prove that the columns of A are dependent if and only if $\operatorname{rank}(A) < n$.

1.15. Here is a summary of everything so far.

This problem is a summary of everything that we have built so far.

1.15.1 Problem (!). Let $A \in \mathbb{R}^{m \times n}$. The goal of life is to solve $A\mathbf{x} = \mathbf{b}$ given $\mathbf{b} \in \mathbb{R}^m$. Fill in the blanks below.

(i) We can always solve $A\mathbf{x} = \mathbf{b}$ when $\mathbf{C}(A) =$ [what: $\mathbb{R}^n$, $\mathbb{R}^m$, $\{\mathbf{0}_n\}$, $\{\mathbf{0}_m\}$, or something else entirely?], where

$$\mathbf{C}(A) = \text{[what is the definition?]}$$

is the column space. So for existence of solutions we want $\mathbf{C}(A)$ to be as [large or small?] as possible.

(ii) Matrix-vector multiplication is a linear combination of [what kind of stuff?]. So $\mathbf{C}(A)$ is the set of all linear combinations of [what?]. To describe $\mathbf{C}(A)$ as efficiently as possible, we only want to use [what kind of columns?] If the columns of A are [independent or dependent?] and A is square, then we expect that solutions [will or will not?] always exist.

(iii) If a solution to $A\mathbf{x} = \mathbf{b}$ exists, then it is unique precisely when $\mathbf{N}(A) =$ [what: $\mathbb{R}^n$, $\mathbb{R}^m$, $\{\mathbf{0}_n\}$, $\{\mathbf{0}_m\}$, or something else entirely?], where

$$\mathbf{N}(A) = \text{[what is the definition?]}$$

is the null space. So for uniqueness of solutions we want $\mathbf{N}(A)$ to be as [large or small?] as possible. If the columns of A are [independent or dependent], then solutions, if they exist, [will or will not be?] unique [and is this just a conjecture or did we prove it already?].

2. Solving Linear Systems

2.1. Introduction: why multiply?

The time has come for a technological leap forward. The fundamental goal of the course is understanding the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ and maybe even solving it. Consideration of this goal has led us to appreciate the need for a deep awareness of the data contained in A and how that data is related to $\mathbf{b}$. It would be helpful if we could develop both an algorithmic procedure for solving $A\mathbf{x} = \mathbf{b}$ (when this problem has a solution) and for extracting and representing useful data about A in meaningful ways.

Remarkably, we can achieve both goals with the same tool: matrix multiplication. Previously we multiplied a vector and a scalar and got another vector. Then we multiplied a matrix and a vector (not just any matrix and any vector—their sizes had to be compatible) and got another vector, in the process, perhaps, multiplying a bunch of vectors by scalars. Now we will multiply two (compatible) matrices and get another matrix.

We defined matrix-vector multiplication in a *meaningful* way: the product $A\mathbf{x}$ represents one side of a system of equations, and the mechanics of taking this product are compatible with natural componentwise operations on vectors. (Reread Example 1.5.6.) We will do the same for matrix-matrix multiplication. But what should this new operation encode? Ideally, two seemingly disparate things: the algorithmic steps of solving $A\mathbf{x} = \mathbf{b}$, and some kind of useful data about A.

We have focused much more on the data contained in A than the mechanics of solving $A\mathbf{x} = \mathbf{b}$, and so we stay with the former for now. Prior life experience has guided us to reverse-engineer multiplication to reveal useful data. We factored integers into products of powers of primes:

$$12 = 2^2(3).$$

And we factored polynomials into simpler polynomials:

$$x^2 - 4x + 4 = (x-2)^2.$$

Both kinds of factorizations reveal (potentially) useful information: what the essential components of an integer are, how to find zeros and maybe graph polynomials. If we know how to multiply matrices, perhaps we can factor them so that only the most important information comes out in the factorization.

2.1.1 Example. We know that the pivot columns of

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

are $\mathbf{a}_2$, $\mathbf{a}_4$ and that every other column is a linear combination of these columns. From the point of view of $\mathbf{C}(A)$, the pivot columns are the most essential data from A. We store them in one matrix:

$$C := \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_4 \end{bmatrix}.$$

Suppose that we store how to build the other columns of A in a "coefficient matrix": since

$$\begin{aligned}
\mathbf{a}_1 &= 0\mathbf{a}_2 + 0\mathbf{a}_4 \\
\mathbf{a}_2 &= 1\mathbf{a}_2 + 0\mathbf{a}_4 \\
\mathbf{a}_3 &= 2\mathbf{a}_2 + 0\mathbf{a}_4 \\
\mathbf{a}_4 &= 0\mathbf{a}_2 + 1\mathbf{a}_4 \\
\mathbf{a}_5 &= 3\mathbf{a}_2 + 4\mathbf{a}_4,
\end{aligned}$$

put

$$R := \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{r}_4 & \mathbf{r}_5 \end{bmatrix}$$

We might think of R as the "recipe" matrix that tells us how to build A out of C. Specifically, the jth column of A is the familiar matrix-vector product

$$\mathbf{a}_j = C\mathbf{r}_j.$$

Our experience with building integers out of products of primes and polynomials out of products of simpler polynomials might make us wonder if there is a way to build A out of the product of C and R. Is there a way to define the matrix product CR so that $A = CR$? If so, then the jth column of CR should be the jth column of A, which is $\mathbf{a}_j = C\mathbf{r}_j$. That is, the jth column of CR should be C times the jth column of R.

2.1.2 Problem ($\star$). **(i)** Use the work of Example 1.14.7 to find matrices C and R such that if

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix},$$

then maybe we could have $A = CR$.

(ii) Why might we expect

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}?$$

Going forward, this is how we will define matrix products: the jth column of the matrix product AB is the matrix-vector product of the matrix A and the jth column (vector) of B. For this to make sense, the jth column of B needs to have the same number of rows as A does columns, and so we will not be able to multiply *any* two matrices.

We have already seen that matrices are both *static* and *dynamic*. They *statically* encode data about linear systems (= life), and they *dynamically* act on vectors to produce new vectors. Now we will see how matrices continue to be usefully static—matrix factorizations will encode data about matrices—and dynamic—matrices will act on other matrices to produce new matrices.

Content from Strang's *ILA* 6E. The real goal from now on is to answer the questions posed at the end of p. 22. We'll get there.

2.2. Matrices act on matrices via multiplication.

Here is our new tool.

2.2.1 Definition. *Let $A \in \mathbb{R}^{m \times n}$ and $B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_p \end{bmatrix} \in \mathbb{R}^{n \times p}$. The* **MATRIX PRODUCT** *AB is*

$$AB := \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_p \end{bmatrix} \in \mathbb{R}^{m \times p}.$$

One reason for the restriction on the sizes of A and B is that we want this definition to return the usual definition of matrix-vector multiplication when B is a column vector. Take $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^n$. Then we know how to compute $A\mathbf{b} \in \mathbb{R}^m$. Think about the matrix $\begin{bmatrix} \mathbf{b} \end{bmatrix} \in \mathbb{R}^{n \times 1}$. If the product $A \begin{bmatrix} \mathbf{b} \end{bmatrix}$ has any meaning *as a matrix*, it probably should be $\begin{bmatrix} A\mathbf{b} \end{bmatrix} \in \mathbb{R}^{m \times 1}$. So the matrix-matrix product $A \begin{bmatrix} \mathbf{b} \end{bmatrix}$ will just be the matrix whose only column is the vector $A\mathbf{b}$.

More broadly, if $A \in \mathbb{R}^{m \times n}$ and $B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_p \end{bmatrix}$, and if we want the jth column of AB to be $A\mathbf{b}_j$, then we need $\mathbf{b}_j \in \mathbb{R}^n$ for $A\mathbf{b}_j$ to be defined. This is why we require $B \in \mathbb{R}^{n \times p}$. Then AB has p columns, and the jth column is $A\mathbf{b}_j \in \mathbb{R}^m$, thus $AB \in \mathbb{R}^{m \times p}$.

Content from Strang's *ILA* 6E. Matrix multiplication is defined in equation (1) on p. 27. Work through the examples on that page and p. 28, noting the appearance of the dot product.

2.2.2 Example. **(i)** Let

$$A = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}.$$

Both A, $B \in \mathbb{R}^{2 \times 2}$, so the product AB is defined and $AB \in \mathbb{R}^{2 \times 2}$ as well. We compute

$$A\mathbf{b}_1 = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and

$$A\mathbf{b}_2 = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}.$$

Thus

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & -2 \end{bmatrix}.$$

(ii) Let

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}.$$

Since $A \in \mathbb{R}^{2\times 2}$ and $B \in \mathbb{R}^{2\times 3}$, the product AB is defined and $AB \in \mathbb{R}^{2\times 3}$. We compute

$$A\mathbf{b}_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix},$$

$$A\mathbf{b}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix},$$

and

$$A\mathbf{b}_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}.$$

Thus

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix}.$$

(iii) Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Since $A \in \mathbb{R}^{3\times 3}$ and $B \in \mathbb{R}^{3\times 2}$, the product AB is defined and $AB \in \mathbb{R}^{3\times 2}$. We compute

$$A\mathbf{b}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}$$

and

$$A\mathbf{b}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \\ 18 \end{bmatrix}.$$

Thus

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 4 & 10 \\ 9 & 18 \end{bmatrix}.$$

2.2.3 Problem (!). Describe in words the effects of computing the three products in the previous example. [Hint: *for part* **(i)**, *think about subtraction.*] Compare your response to patterns that you observed in Problem 1.5.8.

Coming out of these examples is a nice fact that helps when computing "small" products AB by hand.

2.2.4 Theorem. *Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. Then the (i, j)-entry of AB is the dot product of row i of A (considered as a column vector in $\mathbb{R}^n$) with column j of B.*

Proof. We know what AB is at the level of columns: column j of AB is the matrix-vector product of A with column j of B. So the entry in row i of column j of AB is the dot product of row i of A (considered as a column vector in $\mathbb{R}^n$) with column j of B. ■

2.2.5 Problem (!). Redo the matrix products in Example 2.2.2 using dot products.

2.2.6 Problem ($\star$). Suppose that A and B are matrices such that the product AB is defined.

(i) If a whole row of A has all zero entries, what do you know about AB?

(ii) If a whole column of B has all zero entries, what do you know about AB?

2.2.7 Problem ($\star$). Suppose that A and B are matrices such that the product AB is defined.

(i) Prove that if $\mathbf{v} \in \mathbf{C}(AB)$, then $\mathbf{v} \in \mathbf{C}(A)$.

(ii) Give an example of A and B for which $\mathbf{C}(AB) \neq \mathbf{C}(A)$.

Here is something less nice. We expect that the order in which we multiply real numbers yields the same result: if $x, y \in \mathbb{R}$, then $xy = yx$. Not so for matrices: $AB \neq BA$ in general.

2.2.8 Problem ($\star$). **(i)** Explain why even if the matrix product AB is defined, the product BA may not be defined. What do you need to know about A and B for both products AB and BA to be defined?

(ii) Use the matrices A and B from part **(i)** of Example 2.2.2 to show that we may have $AB \neq BA$ even when these products are both defined.

Is this that big a deal? Is our definition of matrix multiplication wrong because it doesn't commute ($AB \neq BA$, typically, even when both products are defined)? No We will shortly see that the factorization of Example 2.1.1 extends to all matrices with this definition of matrix multiplication. And when we develop our algorithm for solving $A\mathbf{x} = \mathbf{b}$, we will see how to encode it as a sequence of matrix products, again with this definition of matrix multiplication. These forthcoming successes will vindicate Definition 2.2.1 despite some of the attendant strangeness. Anyway, most dynamic actions in life do not "commute"—order matters when putting on shoes vs. socks.

Content from Strang's *ILA* 6E. Check the multiplication in equation (6) on p. 28 for further reinforcement that $AB \neq BA$ in general. Answer the question at the bottom of the page.

2.2.9 Problem ($\star$). The noncommutativity of matrix multiplication ($AB \neq BA$ in general, even when both products are defined) was probably something of a surprise. After all, for a, $b \in \mathbb{R}$, we always have $ab = ba$. Here is another surprise: AB can equal the zero matrix even when both A and B are nonzero matrices.

(i) Cook up an example of this yourself by working with 2×2 matrices. [Hint: *you can do this with diagonal matrices if you play the entries off each other carefully.*]

(ii) Suppose that $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ with AB equal to the zero matrix. Explain why if $\mathbf{v} \in \mathbf{C}(B)$, then $\mathbf{v} \in \mathbf{N}(A)$ as well.

2.3. The CR-factorization of a matrix extracts essential columns.

Now that we know how to multiply matrices, we can start factoring matrices into meaningful products. All of the ideas here basically come from Example 2.1.1, which gave us matrix multiplication in the first place.

Let $A \in \mathbb{R}^{m \times n}$ with $r := \operatorname{rank}(A) \geq 1$, so A is not the zero matrix, and therefore A has at least one nonzero column. Suppose that the list of pivot columns of A is $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}$. Put these columns into the matrix

$$C := \begin{bmatrix} \mathbf{a}_{j_1} & \cdots & \mathbf{a}_{j_r} \end{bmatrix} \in \mathbb{R}^{m \times r}.$$

We know that $\mathbf{C}(A)$ equals the span of the pivot columns, so $\mathbf{C}(A) = \mathbf{C}(C)$. In particular, every column of A is in the span of the pivot columns. For each $j = 1, \ldots, n$, then, there is $\mathbf{x}_j \in \mathbb{R}^r$ such that

$$\mathbf{a}_j = C\mathbf{x}_j.$$

In particular, if $\mathbf{a}_j$ is the ith pivot column of A, then $\mathbf{a}_j$ is the ith column of C, so we can take $\mathbf{x}_j = \mathbf{e}_i \in \mathbb{R}^r$.

Put these "recipe" vectors into the matrix

$$R := \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} \in \mathbb{R}^{r \times n}.$$

Then

$$CR = C\begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} C\mathbf{x}_1 & \cdots & C\mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = A.$$

So, every (nonzero) matrix $A \in \mathbb{R}^{m \times n}$ has this CR-FACTORIZATION, where C consists of the pivot columns of A (the absolutely essential columns needed to get the column space) and R is the "recipe" factor that puts them together to get back all of the columns of A. In particular, the columns of C are independent, and if the jth column of A is the ith pivot column, then the jth column of R is the standard basis vector $\mathbf{e}_i \in \mathbb{R}^r$.

2.3.1 Example. **(i)** Let

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}.$$

The first column is nonzero, so $\mathbf{a}_1$ is a pivot column, and $\mathbf{a}_2 \notin \operatorname{span}(\mathbf{a}_1)$, so $\mathbf{a}_2$ is a pivot column. But $\mathbf{a}_3 = 3\mathbf{a}_1 + 2\mathbf{a}_2$, so $\mathbf{a}_3$ is not a pivot column, and therefore the pivot columns are $\mathbf{a}_1$, $\mathbf{a}_2$. Since $\mathbf{a}_1 = 1\mathbf{a}_1 + 0\mathbf{a}_2$ and $\mathbf{a}_2 = 0\mathbf{a}_1 + 1\mathbf{a}_2$, the CR-factorization is then

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \end{bmatrix}$$

(ii) In Example 1.14.7 we saw that the pivot columns of

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}$$

are $\mathbf{a}_2$ and $\mathbf{a}_4$, and that the other columns of A satisfy

$$\mathbf{a}_1 = 0\mathbf{a}_2 + 0\mathbf{a}_4, \qquad \mathbf{a}_3 = 2\mathbf{a}_2 + 0\mathbf{a}_4, \qquad \text{and} \qquad \mathbf{a}_5 = 3\mathbf{a}_2 + 4\mathbf{a}_4.$$

The CR-factorization of A is therefore

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

(iii) Let

$$A = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & 0 & 4 & 6 \\ 3 & 0 & 6 & 9 \end{bmatrix}.$$

The first column of A is nonzero, so it is a pivot column, and every other column of A is a multiple of the first: $\mathbf{a}_2 = 0\mathbf{a}_1$, $\mathbf{a}_3 = 2\mathbf{a}_1$, and $\mathbf{a}_4 = 3\mathbf{a}_1$. Then the CR-factorization of A is

$$\begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & 0 & 4 & 6 \\ 3 & 0 & 6 & 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 3 \end{bmatrix}.$$

2.3.2 Problem (!). Show that for any $\mathbf{a} \in \mathbb{R}^m$ and $r_1, \ldots, r_{n-1} \in \mathbb{R}$, we have

$$\begin{bmatrix} \mathbf{a} & r_1\mathbf{a} & \cdots & r_{n-1}\mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{a} \end{bmatrix} \begin{bmatrix} 1 & r_1 & \cdots & r_{n-1} \end{bmatrix}.$$

2.3.3 Problem (!). Suppose that the columns of $A \in \mathbb{R}^{m \times n}$ are independent. What are the factors C and R in the CR-factorization of A?

2.3.4 Problem ($\star$). Let $A \in \mathbb{R}^{m \times n}$ be nonzero and let $A = CR$ be its CR-factorization. Why is the jth column of A the zero vector (in $\mathbb{R}^m$) precisely when the jth column of R is the zero vector?

Here is a summary of the important properties of the CR-factorization.

2.3.5 Theorem. *Let $A \in \mathbb{R}^{m \times n}$ with $r := \operatorname{rank}(A) \geq 1$. Let $C \in \mathbb{R}^{m \times r}$ be the matrix of pivot columns of A.*

(i) *There is a unique matrix $R \in \mathbb{R}^{r \times n}$ such that $A = CR$.*

(ii) *For $i = 1, \ldots, r$, the ith pivot column of R is $\mathbf{e}_i \in \mathbb{R}^r$.*

(iii) *The pivot columns of A and R occur in the same locations: if $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}$ are the pivot columns of A, then $\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r}$ are the pivot columns of R.*

Proof. We give the proof in the following steps (not exactly corresponding to the three results of the theorem).

1. We proved existence above; here we prove uniqueness of R. (There is no question of uniqueness of C, as the pivot columns are unique by construction.) Suppose that $A = CR$ and $A = C\widetilde{R}$ for $R, \widetilde{R} \in \mathbb{R}^{r \times n}$. Then for any $\mathbf{v} \in \mathbb{R}^r$, we have $C(R - \widetilde{R})\mathbf{v} = \mathbf{0}_m$. (Check that yourself, please.) And so $(R - \widetilde{R})\mathbf{v} \in \mathbf{N}(C) = \{\mathbf{0}_r\}$, since the columns of C are independent. Thus $R\mathbf{v} = \widetilde{R}\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$, and so $R = \widetilde{R}$.

2. Suppose that $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}$ are the pivot columns of A. The columns of $C = \begin{bmatrix} \mathbf{a}_{j_1} & \cdots & \mathbf{a}_{j_r} \end{bmatrix}$ are independent, and $C\mathbf{r}_j = \mathbf{a}_j$ for all j. In particular, $C\mathbf{r}_{j_i} = \mathbf{a}_{j_i}$, and also $C\mathbf{e}_i = \mathbf{a}_{j_i}$. Then $C(\mathbf{r}_{j_i} - \mathbf{e}_i) = \mathbf{a}_{j_i} - \mathbf{a}_{j_i} = \mathbf{0}_m$, so $\mathbf{r}_{j_i} - \mathbf{e}_i \in \mathbf{N}(C) = \{\mathbf{0}_m\}$ by independence. Thus $\mathbf{r}_{j_i} = \mathbf{e}_i$.

3. Now we show that $\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r}$ are the pivot columns of R. Here we use the technically painful, but precise, definition of pivot columns from Remark 1.14.3. (If the notation here is overwhelming, look at the concrete case of part (ii) of Example 2.3.1.)

(i) If $j < j_1$, then $\mathbf{0}_m = \mathbf{a}_j = C\mathbf{r}_j$, so by independence $\mathbf{r}_j = \mathbf{0}_r$, and certainly $\mathbf{r}_{j_1} = \mathbf{e}_1 \neq \mathbf{0}_r$.

(ii) If $r \geq 2$ and $i \geq 2$, then $\mathbf{r}_{j_i} = \mathbf{e}_i \notin \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_{i-1}) = \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_{i-1}})$.

(iii) Suppose that $r \geq 2$, $i \geq 2$, and $j < j_i$. Then $C\mathbf{r}_j = \mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$. But $C\mathbf{r}_j \in \mathbf{C}(C) = \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$. By independence, the entries of $\mathbf{r}_j$ must be zero in rows i through r. Then $\mathbf{r}_j \in \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_{i-1}) = \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_{i-1}})$.

(iv) Since $R \in \mathbb{R}^{r \times n}$, every column of R is in $\operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_r) = \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r})$, and so this is true in particular of columns $\mathbf{r}_j$ with $j > r$.

4. Finally, suppose that we only know the locations of the pivot columns of R; assume that they are $\mathbf{r}_{k_1}, \ldots, \mathbf{r}_{k_r}$. We show that $\mathbf{a}_{k_1}, \ldots, \mathbf{a}_{k_r}$ are the pivot columns of A. Again, we are using the framework of Remark 1.14.3.

(i) Let $k < k_1$. Then $\mathbf{r}_k = \mathbf{0}_r$, so $\mathbf{a}_k = C\mathbf{r}_k = \mathbf{0}_m$. Since the columns of C are independent and $\mathbf{r}_{k_1} \neq \mathbf{0}_r$, likewise $\mathbf{a}_{k_1} = C\mathbf{r}_{k_1} \neq \mathbf{0}_m$.

(ii) Now let $r \geq 2$ and $i \geq 2$. If $\mathbf{a}_{k_i} \in \operatorname{span}(\mathbf{a}_{k_1}, \ldots, \mathbf{a}_{k_{i-1}})$, then $C\mathbf{r}_{k_i} \in \operatorname{span}(C\mathbf{r}_{k_1}, \ldots, C\mathbf{r}_{k_{i-1}})$. Since the columns of C are independent, it follows that $\mathbf{r}_{k_i} \in \operatorname{span}(\mathbf{r}_{k_1}, \ldots, \mathbf{r}_{k_{i-1}})$, a contradiction.

(iii) Continue to assume $r \geq 2$ and $i \geq 2$ and now suppose $k < k_i$. Then $\mathbf{a}_k = C\mathbf{r}_k \in \operatorname{span}(C\mathbf{r}_{k_1}, \ldots, C\mathbf{r}_{k_{i-1}}) = \operatorname{span}(\mathbf{a}_{k_1}, \ldots, \mathbf{a}_{k_{i-1}})$.

(iv) Last, let $k > k_r$. Then $\mathbf{a}_k = C\mathbf{r}_k \in \operatorname{span}(\mathbf{r}_{k_1}, \ldots, \mathbf{r}_{k_r}) = \operatorname{span}(\mathbf{a}_{k_1}, \ldots, \mathbf{a}_{k_r})$. ■

2.3.6 Problem (!). Use the matrix

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

to answer the following questions.

(i) If $A = CR$ is the CR-factorization of a matrix A and a column of R is a standard basis vector, is that column necessarily a pivot column of R?

(ii) Is a pivot column of R necessarily equal to a pivot column of A?

2.3.7 Problem ($\star$). Let $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(A) \geq 1$ and let $A = CR$ be its CR-factorization.

(i) Explain why R has no row with all zero entries.

(ii) Show that $\operatorname{rank}(A) = \operatorname{rank}(C) = \operatorname{rank}(R)$.

(iii) Prove that $\mathbf{C}(C) = \mathbf{C}(A)$ and $\mathbf{C}(R) = \mathbb{R}^r$, $r := \operatorname{rank}(A)$.

Content from Strang's *ILA* 6E. Read and work through all of the calculations on pp. 29–30 under "Rank One Matrices and $A = CR$." Then read "C Contains the First r Independent Columns of A" on p. 30 and "Matrix Multiplication C times R on pp. 31–32. Check the calculations in Example 2, equation (10), equation (11), and the box on p. 32. Also jump ahead to Example 5 on pp. 34–35 (you don't have to read about that "columns × rows" way of multiplying matrices). For yet another example, go back to "Matrix Multiplication $A = CR$ on p. vii. You do not have to feel that you could see these CR-factorizations immediately; you should agree that the given matrix multiplication works out.

If you're curious, read pp. 32–33 to learn more about computing R. Feel free to skip that for now. We will revisit this in extensive detail in the future.

Our discussion of the CR-factorization so far has been glib and existential. We know that a nonzero matrix A has pivot columns: just scan the matrix left to right and find them. We know that the "recipe" factor R exists: just write each column as the appropriate linear combination of the pivot columns. But this is a lot of work, and it involves solving linear systems (to represent columns as linear combinations of pivot columns) or proving that linear systems have no solutions (to check that a column really is a pivot column). We lack easy algorithms for doing this right now. We will develop them, eventually.

2.4. We can multiply more than two matrices.

So far, we have only defined the product of two matrices. Why stop there? We can multiply more than two numbers together in one go.

Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, and $C \in \mathbb{R}^{p \times q}$. Then $AB \in \mathbb{R}^{m \times p}$ is defined, so $(AB)C \in \mathbb{R}^{m \times q}$ is defined. But we could also say that $BC \in \mathbb{R}^{n \times q}$ is defined, and then $A(BC) \in \mathbb{R}^{m \times q}$ is defined. From our experience with arithmetic, we certainly expect

$$(AB)C = A(BC) \tag{2.4.1}$$

when all of the products involved (there are four) are defined; this is **ASSOCIATIVITY** of matrix multiplication. But we probably also expected commutativity of matrix multiplication, so maybe (2.4.1) does not happen.

It does! We start small. Take $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, and $\mathbf{v} = (v_1, \ldots, v_p) \in \mathbb{R}^p$. Then $AB \in \mathbb{R}^{m \times p}$ is defined, so $(AB)\mathbf{v} \in \mathbb{R}^m$ is defined. Also, $B\mathbf{v} \in \mathbb{R}^n$ is defined, so $A(B\mathbf{v}) \in \mathbb{R}^m$ is defined. Of course, we hope that $(AB)\mathbf{v} = A(B\mathbf{v})$. Here is why:

$$\begin{aligned}
A(B\mathbf{v}) &= A(v_1\mathbf{b}_1 + \cdots + v_p\mathbf{b}_p) \text{ by definition of the matrix-vector product } B\mathbf{v} \\
&= v_1A\mathbf{b}_1 + \cdots + v_pA\mathbf{b}_p \text{ by linearity of matrix-vector multiplication} \\
&= \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_p \end{bmatrix}\mathbf{v} \text{ by definition of the matrix-vector product } \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_p \end{bmatrix}\mathbf{v} \\
&= (AB)\mathbf{v} \text{ by definition of the matrix-matrix product } AB.
\end{aligned}$$

This is enough to get (2.4.1) because the definition of matrix multiplication hinges on columns. (When all else fails in math, suck it up and go back to the definition.) Back to $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, and $C = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_q \end{bmatrix} \in \mathbb{R}^{p \times q}$. We have

$$\begin{aligned}
(AB)C &= \begin{bmatrix} (AB)\mathbf{c}_1 & \cdots & (AB)\mathbf{c}_q \end{bmatrix} \text{ by definition of the matrix-matrix product } (AB)C \\
&= \begin{bmatrix} A(B\mathbf{c}_1) & \cdots & A(B\mathbf{c}_q) \end{bmatrix} \text{ by the identity above that } A(B\mathbf{c}_j) = (AB)\mathbf{c}_j
\end{aligned}$$

$$= A \begin{bmatrix} B\mathbf{c}_1 & \cdots & B\mathbf{c}_q \end{bmatrix} \text{ by definition of the matrix-matrix product } A \begin{bmatrix} B\mathbf{c}_1 & \cdots & B\mathbf{c}_q \end{bmatrix}$$

$$= A(B \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_q \end{bmatrix}) \text{ by definition of the matrix-matrix product } BC$$

$$= A(BC).$$

2.4.1 Theorem. *Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, and $C \in \mathbb{R}^{p \times q}$. Then $(AB)C = A(BC)$.*

This theorem says that the order in which we *group* matrices during multiplication does not matter: matrix multiplication is **ASSOCIATIVE**. Thus we just write ABC and eliminate the parentheses. The *order* still totally matters, and we should not expect $ABC = ACB$ or some nonsense like that.

Associativity of matrix multiplication is more than just an "expected" algebraic fact—it confirms that our definition of matrix multiplication is the "right" one! Here is why. Say that we did not know how to define AB but we knew what it should *do*: for matrices A and B, we want AB to satisfy $(AB)\mathbf{v} = A(B\mathbf{v})$ whenever the matrix product AB and the matrix-vector products $(AB)\mathbf{v}$, $B\mathbf{v}$, and $A(B\mathbf{v})$ are defined. Remember, what things do defines what things are. If $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{q \times p}$, we need to take $\mathbf{v} \in \mathbb{R}^p$ for $B\mathbf{v}$ to be defined. Then we need $n = q$ for $A(B\mathbf{v})$ to be defined, so the matrix product AB can only be defined for $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$.

Now, if we have $(AB)\mathbf{v} = A(B\mathbf{v})$ for *all* $\mathbf{v} \in \mathbb{R}^p$, we can pick $\mathbf{v}$ cleverly: take $\mathbf{v} = \mathbf{e}_j$. Then $(AB)\mathbf{e}_j = A(B\mathbf{e}_j)$ for $j = 1, \ldots, p$. The jth column of AB is $(AB)\mathbf{e}_j$, and the jth column of B is $B\mathbf{e}_j$. So the jth column of the matrix-matrix product AB is the matrix-vector product of A and the jth column of B.

2.4.2 Problem ($\star$). If $A \in \mathbb{R}^{m \times m}$, then we can multiply A and A:

$$A^2 := AA \in \mathbb{R}^{m \times m}.$$

For an integer $k \geq 2$, we put $A^k := A^{k-1}A^k$. Let $D \in \mathbb{R}^{m \times m}$ be diagonal. Describe in words the matrix D^k.

Content from Strang's *ILA* 6E. Read "AB times $C = A$ times BC" on p. 29.

Here is an example of how multiplying more than two matrices at a time can clean up a CR-factorization. We computed the CR-factorization

$$\underbrace{\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}}_{R}$$

back in Example 2.1.1. If we look at the second factor (the R-factor), we might notice the two standard basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$ for $\mathbb{R}^2$ as columns. What if they fell out differently, and the second factor was

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 1 & 0 & 0 & 4 \end{bmatrix}?$$

Here the columns of the identity are shuffled to the front of the matrix, and we have the nice **BLOCK** structure

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 1 & 0 & 0 & 4 \end{bmatrix} = \begin{bmatrix} I_2 & F \end{bmatrix}, \qquad F := \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix}.$$

We will sometimes take the point of view that a matrix is a **BLOCK MATRIX** whose entries are other matrices. This can help us zoom out and focus on important "global" features of a matrix, rather than viewing it only entry-by-entry.

However, the problem here is that $R \neq \begin{bmatrix} I_2 & F \end{bmatrix}$, although the block structure of $\begin{bmatrix} I_2 & F \end{bmatrix}$ is probably cleaner than the "jumbled up" version above. We need to think dynamically: how do we reorder the columns in a matrix? First, how do we *extract* the columns of a matrix? Multiply by a standard basis vector! If $B \in \mathbb{R}^{p \times q}$, then the jth column of B is $B\mathbf{e}_j$, where $\mathbf{e}_j \in \mathbb{R}^q$ is the jth standard basis vector. So to reorder the columns of B, multiply B by the matrix $P \in \mathbb{R}^{q \times q}$ whose columns are the standard basis vectors for $\mathbb{R}^q$ *in the order that we want the columns of B to go.* That is, P is a "permutation" of the identity matrix I_q.

2.4.3 Definition. *A* ***PERMUTATION MATRIX*** $P \in \mathbb{R}^{n \times n}$ *is a matrix whose columns are the standard basis vectors for* $\mathbb{R}^n$ *in some order. Each standard basis vector appears once, and only once, as a column of* P*. Equivalently, we form a permutation matrix by reordering (some, maybe all) of the columns of the identity matrix* I_n*.*

Multiplying on the right by a permutation matrix to shuffle the order of columns of a given matrix is our first example of how matrices act on other matrices in meaningful ways: *by multiplication.* Whenever we try to do something in this course from now on, we should always ask ourselves if we can do it via matrix multiplication.

Back to the concrete situation above, we like the matrix $B = \begin{bmatrix} I_2 & F \end{bmatrix}$, and we want to reorder its columns to match those of R. So we want to reorder the five columns of B into columns 3, 1, 4, 2, 5.

2.4.4 Problem (!). Check that

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 1 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} \mathbf{e}_3 & \mathbf{e}_1 & \mathbf{e}_4 & \mathbf{e}_2 & \mathbf{e}_5 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix},$$

where $\mathbf{e}_j$ here is the jth standard basis vector in $\mathbb{R}^5$.

We conclude

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = C \begin{bmatrix} I_2 & F \end{bmatrix} P, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad F = \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.4.2}$$

Here is the general truth. Let $A \in \mathbb{R}^{m \times n}$ have the CR-factorization $A = CR$. Suppose that $r := \operatorname{rank}(A)$ with $1 \le r < n$ and suppose that columns $j_1, \ldots, j_r$ of A are the pivot columns. Then the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r \in \mathbb{R}^r$ appear (at least once) in R; we want them to come first, so we want a permutation matrix $P \in \mathbb{R}^{n \times n}$ such that $R = \begin{bmatrix} I_r & F \end{bmatrix} P$ for some "junky" block F. This matrix P should be the permutation matrix such that for any $B \in \mathbb{R}^{r \times n}$, columns $j_1, \ldots, j_r$ of BP are the first r columns of B. That is, column j_i of P should be $\mathbf{e}_i \in \mathbb{R}^n$, and otherwise column j of P can just be $\mathbf{e}_j$. We conclude that we can write A in the form

$$A = C \begin{bmatrix} I_r & F \end{bmatrix} P, \tag{2.4.3}$$

where C contains the pivot columns of A and P is a permutation matrix. It is indeed possible that the block I_r does come first in R, in which case $P = I_n$.

Content from Strang's *ILA* 6E. The inclusion of this permutation matrix P in the factorization is what Strang means by the parenthetical remark "in correct order" on p. in the displayed equations after "$A = CR$ becomes."

2.4.5 Problem (!). Here is some practice with block structure for matrices. Let

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 3 & 3 \\ 2 & 2 & 3 & 3 \end{bmatrix}.$$

There are multiple ways to break A up into blocks, some of which are more informative than others. The convention is always that blocks in the same row need to have the same number of rows, and blocks in the same column need to have the same number of columns. We allow both row and column vectors to count as blocks, and occasionally there are 1×1 ($=$ scalar) blocks, too.

(i) Find A_1, $A_2 \in \mathbb{R}^{3 \times 2}$ such that

$$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}.$$

(ii) Find $A_1 \in \mathbb{R}^{3 \times 3}$ and $A_2 \in \mathbb{R}^{3 \times 1}$ such that

$$\begin{bmatrix} A_1 & A_2 \end{bmatrix}.$$

(iii) Find A_{11}, $A_{12} \in \mathbb{R}^{1 \times 2}$ and A_{21}, $A_{22} \in \mathbb{R}^{2 \times 2}$ such that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

(iv) Find A_{11}, $A_{12} \in \mathbb{R}^{2\times 2}$ and A_{21}, $A_{22} \in \mathbb{R}^{1\times 2}$ such that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

2.4.6 Problem ($\star$). We can do matrix-vector multiplication with block matrices, provided that the blocks match appropriately.

(i) Suppose that we want the products

$$\begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} = A_1\mathbf{v}_1 + A_2\mathbf{v}_2 \quad \text{and} \quad \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \mathbf{v} = \begin{bmatrix} B_1\mathbf{v} \\ B_2\mathbf{v} \end{bmatrix}$$

to make sense. What are the sizes of the matrices A_1, A_2, B_1, B_2 and the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}$?

(ii) Find matrices A_1 and A_2 and vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ such that the matrix-vector product

$$\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 3 & 3 \\ 2 & 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

equals

$$A_1\mathbf{v}_1 + A_2\mathbf{v}_2$$

Do all of the arithmetic to make sure that the equality is true. Then find another pair of matrices and vectors—of different sizes than your first pairs—such that the product equality is still true.

(iii) Find matrices B_1 and B_2 such that the matrix-vector product

$$\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 3 & 3 \\ 2 & 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

equals

$$\begin{bmatrix} B_1\mathbf{v} \\ B_2\mathbf{v} \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

Do all of the arithmetic to make sure that the equality is true. Then find another pair of matrices—of different sizes than your first pairs—such that the product equality is still true.

2.4.7 Problem (⋆). The block F and the permutation matrix P from (2.4.3) need not be unique. Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Find two different matrices F_1, $F_2 \in \mathbb{R}^{2\times 3}$ and two different permutation matrices P_1, $P_2 \in \mathbb{R}^{5\times 5}$ such that

$$A = C \begin{bmatrix} I_2 & F_1 \end{bmatrix} P_1 \quad \text{and} \quad A = C \begin{bmatrix} I_2 & F_2 \end{bmatrix} P_2,$$

where in both cases C contains the pivot columns of A.

Content from Strang's *ILA* 6E. At this point we have learned all the matrix-vector mechanics that we need to actually solve linear systems (and to understand our failure when we can't solve them). Just to be safe, read "Review of AB on p. 29 and make sure you have no doubts there. Then read "Thoughts on Chapter 1" on p. 38 for a summary of everything that we've done and a hint of what's to come.

2.4.8 Problem (⋆). Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbf{C}(A)$. Let $\mathbf{v} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$. Justify the following steps to show that $\mathbf{v} \in \mathbf{C}(A)$.

(i) There is $\mathbf{x} \in \mathbb{R}^p$ such that $\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_p \end{bmatrix} \mathbf{x}$.

(ii) There are $\mathbf{w}_j \in \mathbb{R}^m$ such that $\mathbf{v}_j = A\mathbf{w}_j$ for $j = 1, \ldots, p$.

(iii) We have $\mathbf{v} = A \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_p \end{bmatrix} \mathbf{x} \in \mathbf{C}(A)$.

2.4.9 Problem (⋆). Let $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = n$ and let $\mathbf{v} \in \mathbb{R}^m$. By part (ii) of Problem 1.14.13, the columns of A are independent, so if there are $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathbb{R}^n$ such that $A\mathbf{x}_1 = A\mathbf{x}_2$, then $\mathbf{x}_1 = \mathbf{x}_2$ by Theorem 1.11.4. Use this to establish the following. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_p$, $\mathbf{v} \in \mathbb{R}^n$. If $A\mathbf{v} \in \operatorname{span}(A\mathbf{v}_1, \ldots, A\mathbf{v}_p)$, then $\mathbf{v} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$. [Hint: *use $A\mathbf{v}$ in place of $\mathbf{v}$ from Problem 2.4.8 and consider the formula from part (iii) of that problem.*]

2.4.10 Problem (+). Here is a situation that will arise from time to time. Let $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = n$ and let $B \in \mathbb{R}^{n\times p}$; suppose that B is nonzero. If the pivot columns of B are $\mathbf{b}_{j_1}, \ldots, \mathbf{b}_{j_r}$, then the pivot columns of AB are $A\mathbf{b}_{j_1}, \ldots, A\mathbf{b}_{j_r}$. Consequently, $\operatorname{rank}(AB) = \operatorname{rank}(B)$.

Use Problem 2.4.9 (and the reminders therein) and the precise properties of pivot columns from Remark 1.14.3 to prove this by establishing the following.

(i) If $j < j_1$, then $A\mathbf{b}_j = \mathbf{0}_m$. [Hint: *what is $\mathbf{b}_j$ here?*]

(ii) $A\mathbf{b}_{j_1} \neq \mathbf{0}_m$. [Hint: rank$(A) = n$, *and what do you know about* $\mathbf{b}_{j_1}$*?*]

(iii) If $r \geq 2$ and $i \geq 2$, then $A\mathbf{b}_{j_i} \notin \text{span}(A\mathbf{b}_{j_1}, \ldots, A\mathbf{b}_{j_{i-1}})$. [Hint: *if* $A\mathbf{b}_{j_i} \in \text{span}(A\mathbf{b}_{j_1}, \ldots, A\mathbf{b}_{j_{i-1}})$*, what does Problem 2.4.9 say?*]

(iv) If $r \geq 2$, $i \geq 2$, and $j < j_i$, then $A\mathbf{b}_j \in \text{span}(A\mathbf{b}_{j_1}, \ldots, A\mathbf{b}_{j_{i-1}})$. [Hint: *use part (iii) of Problem 1.12.14. What do you know about* $\mathbf{b}_j$ *here?*] If $r \geq 2$, $i \geq 2$, and $j > j_r$, then $A\mathbf{b}_j \in \text{span}(A\mathbf{b}_{j_1}, \ldots, A\mathbf{b}_{j_r})$. [Hint: *again, what do you know about* $\mathbf{b}_j$ *here?*]

2.5. There are only three possible solution behaviors.

We are almost ready to start solving linear systems. It will be helpful to know where we are going before we get there, so we (briefly) pause from matrix manipulations and look at three linear systems, each of which is in a very nice form, and which together illustrate the scope of possibilities for solution behavior to $A\mathbf{x} = \mathbf{b}$.

2.5.1 Example.

(i) We consider

$$\begin{bmatrix} 1 & -2 \\ 0 & 8 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}.$$

As a linear system, this reads

$$\begin{cases} x_1 - 2x_2 = 1 \\ \phantom{x_1 - {}} 8x_2 = 8 \end{cases}$$

Look familiar? This was our very first problem!

Of course, we "back-solve" or "back-substitute" to get first $x_2 = 1$ and then $x_1 - 2 = 1$, so $x_1 = 3$. The problem has only one solution:

$$\mathbf{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

(ii) We consider

$$\begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}.$$

Write it out:

$$\begin{cases} x_1 - 2x_2 = 1 \\ 0 = 8. \end{cases}$$

Of course this system has no solution, because $0 \neq 8$.

(iii) We consider

$$\begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Write it out:

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

There is really not much to do, since the second equation is both true and does not involve unknowns. There is not much more we can do with the first equation, since we hav no specific value for x_2.

Here is the right, if not obvious, thing to do: rewrite $x_1 = 1 + 2x_2$. This says that every choice of $x_2 \in \mathbb{R}$ gives x_1 via this formula. We can pick any x_2, so there are infinitely many solutions. At the level of vectors, we could write

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

Every value of x_2 gives a different solution, and so this problem has infinitely many solutions.

Content from Strang's *ILA* 6E. Work through the three systems on p. 40, which have the same properties as the three above.

The three examples above are paradigmatic in the sense that a linear system has only one of three general solution "behaviors": only one solution, no solution, or infinitely many solutions. This is actually very easy to prove using matrix notation—which is why we use that notation, to make our lives easier. But the other thing to take from this example is that the *structure* of the linear systems was very nice: all of the matrices were "upper-triangular" in the sense that their entries were 0 below the diagonal. This made back-solving/substituting very, very easy.

Content from Strang's *ILA* 6E. For a very broad overview of where we're going, read p. 39. It's okay if you don't understand everything on a first pass. Then read the first three paragraphs on p. 83.

We formalize the situations of Example 2.5.1. The method of proof here has a lot in common with Theorem 1.10.9; now is a good time to pause and reread that theorem and the discussion preceding it.

2.5.2 Theorem. *Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then one, and only one, of the following is true.*

(i) *There exists a unique solution $\mathbf{x} \in \mathbb{R}^n$ to the problem $A\mathbf{x} = \mathbf{b}$. That is, we can solve the problem, and if $A\mathbf{x}_1 = \mathbf{b}$ and $A\mathbf{x}_2 = \mathbf{b}$ for some $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathbb{R}^n$, then $\mathbf{x}_1 = \mathbf{x}_2$.*

(ii) *There is no solution to the problem $A\mathbf{x} = \mathbf{b}$. That is, $A\mathbf{x} \neq \mathbf{b}$ for every $\mathbf{x} \in \mathbb{R}^n$.*

(iii) *There are infinitely many solutions to $A\mathbf{x} = \mathbf{b}$.*

Proof. We want one, and only, one, of three possibilities to hold. Certainly they are all mutually exclusive—if one is true, then the others have to be false. If there is a unique solution, it cannot be the case that no solutions exist or that infinitely many solutions exist. If there is no solution, there cannot exist a unique solution or infinitely many solutions. And if there are infinitely many solutions, then there cannot be a unique solution nor the absence of any solution.

Why, then, must one of these three possibilities be true in the first place? What if they are all false? Assume that $A\mathbf{x} = \mathbf{b}$ has a solution (so the second part is false) but this solution is not unique (so the first part is false). We show that the third part must be true, and so all three parts cannot be false.

We are assuming that there are $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ such that $A\mathbf{x}_1 = \mathbf{b}$, $A\mathbf{x}_2 = \mathbf{b}$, and $\mathbf{x}_1 \neq \mathbf{x}_2$. Our goal is to find infinitely many different $\mathbf{x} \in \mathbb{R}^n$ that satisfy $A\mathbf{x} = \mathbf{b}$. Here is the trick. (Like most tricks in math, it may not be obvious at first glance, so you should reread this proof until it becomes obvious.)

Put $\mathbf{z} := \mathbf{x}_1 - \mathbf{x}_2$. Then $\mathbf{z} \neq \mathbf{0}_n$, since $\mathbf{x}_1 \neq \mathbf{x}_2$. And

$$A\mathbf{z} = A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0}_m.$$

The second equality is the linearity of matrix-vector multiplication.

Now let $c \in \mathbb{R}$ be arbitrary and $\mathbf{y}_c = \mathbf{x}_1 + c\mathbf{z}$. Then

$$A\mathbf{y}_c = A(\mathbf{x}_1 + c\mathbf{z}) = A\mathbf{x}_1 + A(c\mathbf{z}) = A\mathbf{x}_1 + c(A\mathbf{z}) = \mathbf{b} + c\mathbf{0}_m = \mathbf{b} + \mathbf{0}_m = \mathbf{b}.$$

The second and third equalities are, again, the linearity of matrix-vector multiplication.

So why does this give infinitely many solutions? Each different $c \in \mathbb{R}$ generates a different $\mathbf{y}_c = \mathbf{x}_1 + c\mathbf{z} \in \mathbb{R}^n$. This was Problem 1.10.8. ■

2.5.3 Problem (!). By considering the vector $\mathbf{z} = (2, -1)$, explain how the proof of this theorem generalizes the situation in part (iii) of Example 2.5.1.

Content from Strang's *ILA* 6E. After you do the problem above, reread Example 3 on p. 40. The vector that Strang calls $\boldsymbol{X}$ is what I call $\mathbf{z}$.

2.6. Elimination solves linear systems.

The time has come to systematically solve linear systems! We go all the way back to our very first example, in which we showed that

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}\mathbf{x} = \begin{bmatrix} 1 \\ 11 \end{bmatrix} \iff \begin{bmatrix} 1 & -2 \\ 0 & 8 \end{bmatrix}\mathbf{x} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}. \tag{2.6.1}$$

The latter system was easy to solve with "back-substitution":

$$\begin{bmatrix} 1 & -2 \\ 0 & 8 \end{bmatrix}\mathbf{x} = \begin{bmatrix} 1 \\ 8 \end{bmatrix} \iff \left\{ \begin{array}{rcrcl} x_1 & - & 2x_2 & = & 1 \\ & & 8x_2 & = & 8 \end{array} \right. \iff \left\{ \begin{array}{rcrcl} x_1 & - & 2x_2 & = & 1 \\ & & x_2 & = & 1 \end{array} \right.$$

$$\iff \begin{cases} x_1 \quad - \quad 2 \quad = \quad 1 \\ \qquad\qquad x_2 \quad = \quad 1 \end{cases} \iff \begin{cases} x_1 = 3 \\ x_2 = 1. \end{cases}$$

We start by identifying the nice form of the second system in (2.6.1).

2.6.1 Definition. *A matrix $U \in \mathbb{R}^{m \times m}$ is* ***UPPER-TRIANGULAR*** *if all of the entries of U below the diagonal are 0. That is, the (i, j)-entry of U is 0 when $i > j$.*

2.6.2 Example. Each matrix below is upper-triangular:

$$\begin{bmatrix} 1 & -2 \\ 0 & 8 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \text{and} \qquad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}.$$

2.6.3 Problem (!). Is the rank of an upper-triangular matrix always equal to the number of nonzero entries on its diagonal?

Content from Strang's *ILA* 6E. For a longer example of why upper-triangular matrices are nice for back-substitution, read p. 41 through the "Special note" in the box. I expect that you are comfortable with this back-substitution method for solving linear systems, and I will not do examples with it here.

How do we rewrite linear systems as we did (2.6.1)? How do we "convert" $A \in \mathbb{R}^{m \times m}$ into an upper-triangular matrix U so that we have the equivalence of the problems

$$A\mathbf{x} = \mathbf{b} \quad \text{and} \quad U\mathbf{x} = \mathbf{c}$$

for some appropriate $\mathbf{c}$? The point is that the arrows go *both* ways:

$$A\mathbf{x} = \mathbf{b} \Longrightarrow U\mathbf{x} = \mathbf{c} \quad \text{and} \quad U\mathbf{x} = \mathbf{c} \Longrightarrow A\mathbf{x} = \mathbf{b}.$$

This is special, because an arrow that goes one way in math does not have to go the other way.

The good news is that we already know how to do this, and it is all contained in the manipulations that we did on our very first problem at the level of equations and variables. The big idea was subtracting a multiple of one equation from another. We can do all of this at the level of matrices (and cut out the variables) by subtracting a multiple of one row of a matrix from another. And the bigger idea is that we encode this via matrix multiplication.

1. First, why is this new problem $U\mathbf{x} = \mathbf{c}$ so nice? Because U is upper-triangular, which (when the diagonal entries of U are nonzero) permits us to solve $U\mathbf{x} = \mathbf{c}$ by *back-substitution.* Going down, each equation in $U\mathbf{x} = \mathbf{c}$ has one fewer unknown than the preceding equation, and at the bottom we have an equation in just one unknown. We can solve that because we know how to do algebra. Then we know one of the two unknowns in the equation above that, so that second-to-last equation is also an equation in one unknown. Turn the crank. . .

2. Second, how do we go from $A\mathbf{x} = \mathbf{b}$ to $U\mathbf{x} = \mathbf{c}$? We compress the operations of *Gaussian elimination* into a matrix E, so that EA is upper-triangular. Let $U := EA$ and $\mathbf{c} := E\mathbf{b}$. Then if $A\mathbf{x} = \mathbf{b}$, we can apply E to both sides to get $EA\mathbf{x} = E\mathbf{b}$, thus $U\mathbf{x} = \mathbf{c}$. Any solution to our original problem solves this new problem.

3. Finally, why does solving the easier problem $U\mathbf{x} = \mathbf{c}$ yield a solution to $A\mathbf{x} = \mathbf{b}$? We will show that E is *invertible*: there is a matrix E^{-1} such that $E^{-1}E = I_n$. So if $U\mathbf{x} = \mathbf{c}$, then $E^{-1}U\mathbf{x} = E^{-1}\mathbf{c}$. From what U and $\mathbf{c}$ are, this says $E^{-1}EA\mathbf{x} = E^{-1}E\mathbf{b}$, and so $A\mathbf{x} = \mathbf{b}$. Thus any solution to the new problem solves the original problem—the thing we actually care about.

We revisit (2.6.1) from the point of view of matrices. To turn

$$A = \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}$$

into

$$U = \begin{bmatrix} 1 & -2 \\ 0 & 8 \end{bmatrix},$$

we want to subtract 3 times the first row of A from the second row of A. The innovation of linear algebra is that we can encode this via matrix multiplication. Whenever we want to "do something" in this class, we should ask ourselves how we can accomplish this by multiplying by a suitable matrix.

What matrix E satisfies

$$EA = U?$$

At the very least we need $E \in \mathbb{R}^{m\times 2}$ since $A \in \mathbb{R}^{2\times 2}$. And we really want $m = 2$ since $EA = U \in \mathbb{R}^{2\times 2}$. So, $E \in \mathbb{R}^{2\times 2}$.

Here is where it is wise to think about matrix multiplication as E times the columns of A. What is E doing to each column? We want

$$E \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 - 3v_1 \end{bmatrix}. \tag{2.6.2}$$

How can we view the vector on the right as a linear combination with weights given by v_1 and v_2? The vectors in that linear combination will be the columns of E.

So, work backwards:

$$\begin{bmatrix} v_1 \\ v_2 - 3v_1 \end{bmatrix} = \begin{bmatrix} v_1 \\ -3v_1 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ -3 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

If we put

$$E := \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix},$$

then we have the desired equality (2.6.2).

2.6.4 Problem (!). Check that. Then compute $EA = U$ with A and U as above.

Here is our thinking. Assume $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} = (1, 11)$. Then $EA\mathbf{x} = E\mathbf{b}$. Compute $EA = U$ with U as above and $E\mathbf{b} = (1, 8) =: \mathbf{c}$. Then solve $U\mathbf{x} = \mathbf{c}$. That *should* give a solution to the original problem $A\mathbf{x} = \mathbf{b}$, and we can always plug it in and check that it does.

Going in reverse requires a little more thought. Why does solving $EA\mathbf{x} = E\mathbf{b}$ give a solution to $A\mathbf{x} = \mathbf{b}$? It would be nice if we could "cancel" the factor of E from both sides.

2.6.5 Problem ($\star$). In fact, you can do that right now. Put

$$F := \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}.$$

First explain in words the effect of multiplying $F\mathbf{w}$ for some $\mathbf{w} \in \mathbb{R}^2$. Then check that $FE\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^2$. Finally, suppose that $EA\mathbf{x} = E\mathbf{b}$, multiply both sides by F, and explain why $A\mathbf{x} = \mathbf{b}$.

It might feel as though we are doing "elimination" twice: we multiplied EA and then $E\mathbf{b}$ separately. We can combine all of the data of our problem into one "augmented" matrix: put

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 3 & 2 & 11 \end{array}\right].$$

We draw a line separating the $\mathbf{b}$ column from the A-block to emphasize that A and $\mathbf{b}$ appear in different places in the problem and that $\mathbf{b}$ is not a column of A. Then do one matrix multiplication:

$$E\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} EA & E\mathbf{b} \end{bmatrix} = \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 8 & 8 \end{array}\right] = \begin{bmatrix} U & \mathbf{c} \end{bmatrix}.$$

From here, solve $U\mathbf{x} = \mathbf{c}$ by back-substitution.

Here is the cartoon for $A \in \mathbb{R}^{3\times3}$. We want to turn A into an upper-triangular matrix U by multiplying A by the "right" matrices. In the "ideal" case, at the level of rows, we are going to subtract multiples of row 1 to create 0 entries in rows 2 and below of column 1. Specifically, the multiples will be based on the $(1, 1)$-entry, which for now we hope is nonzero.

So we have the conversion

$$\begin{bmatrix} \circledast & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \to \begin{bmatrix} * & * & * \\ {\color{blue}0} & {\color{blue}*} & {\color{blue}*} \\ {\color{blue}0} & {\color{blue}*} & {\color{blue}*} \end{bmatrix}.$$

The changed entries are in blue. Now subtract a multiple of the second row from the third row to create zeros in the second column below the second row. Again, in the "ideal" case, the multiple will be based on the $(2, 2)$-entry, which we should hope is nonzero:

$$\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & * & * \end{bmatrix} \to \begin{bmatrix} * & * & * \\ 0 & \circledast & * \\ 0 & {\color{blue}0} & {\color{blue}*} \end{bmatrix}.$$

Again, the blue entries are new or changed. Because both the second and third rows had 0 in their first column, subtracting a multiple of the second row from the third row did not destroy that 0 in the first column of the third row. This is the nice upper-triangular structure that is ideal for back-solving.

How do we accomplish this multiplication? Let $A \in \mathbb{R}^{m \times n}$ and $\ell \in \mathbb{R}$. To subtract ℓ times row j of A from row i of A (with $i \neq j$), multiply A by the **ELIMINATION MATRIX** $E_{ij} \in \mathbb{R}^{m \times m}$ whose entries are 1 on the diagonal, $-\ell$ in the (i,j)-position, and 0 elsewhere. So, E_{ij} is "almost" the identity matrix, except for the (i,j)-entry.

2.6.6 Problem (!).

(i) Prove that this formula for E_{ij} works by computing the following very special case and explaining the effect in words:

$$E_{21}\mathbf{v}, \qquad \text{where} \qquad E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -\ell & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{v} := \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

Then spend at least five minutes thinking about how using dot products could help you prove the more general result stated in the paragraph above this problem.

(ii) Write down a formula for the elimination matrix that subtracts 5 times row 2 of a matrix in $\mathbb{R}^{4 \times 4}$ from row 4 of that matrix.

We do an example in glacially slow detail.

2.6.7 Example. Let

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{bmatrix}.$$

We want to multiply A by "elimination" matrices like the 2×2 situation above so that 0 appears in the second and third rows of the first column. To get 0 in the $(2,1)$-entry, we should subtract 2 times the first row from the second. The matrix

$$E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

accomplishes this, and here is what we get:

$$E_{21}A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 8 & 7 & 9 \end{bmatrix}.$$

We use the idiosyncratic notation

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{bmatrix} \xrightarrow[E_{21}]{\mathsf{R2} \mapsto \mathsf{R2}-2\times\mathsf{R1}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 8 & 7 & 9 \end{bmatrix}, \qquad E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

to represent this. Saying R2 $\mapsto$ R2 $-$ 2 $\times$ R1 means that row 2 is replaced by row 2 minus 2 times row 1.

Now we want to clear out the $(3,1)$-entry, and we can do this by subtracting 4 times row 1 from row 3. So, we multiply

$$\begin{bmatrix}2&1&1\\0&1&1\\8&7&9\end{bmatrix} \xrightarrow[E_{31}]{\text{R3} \mapsto \text{R3}-4\times\text{R1}} \begin{bmatrix}2&1&1\\0&1&1\\0&3&5\end{bmatrix}, \qquad E_{31} := \begin{bmatrix}1&0&0\\0&1&0\\-4&0&1\end{bmatrix}.$$

Finally, we want to clear out the 3 in the $(3,2)$-entry:

$$\begin{bmatrix}2&1&1\\0&1&1\\0&3&5\end{bmatrix} \xrightarrow[E_{32}]{\text{R3} \mapsto \text{R3}-3\times\text{R2}} \begin{bmatrix}2&1&1\\0&1&1\\0&0&2\end{bmatrix}, \qquad E_{32} := \begin{bmatrix}1&0&0\\0&1&0\\0&-3&1\end{bmatrix}.$$

And we are done! Let's abbreviate $E = E_{32}E_{31}E_{21}$. The product

$$EA = \begin{bmatrix}2&1&1\\0&1&1\\0&0&2\end{bmatrix} =: U$$

is upper-triangular. If we wanted to solve $A\mathbf{x} = \mathbf{b}$ for some $\mathbf{b} \in \mathbb{R}^3$, it would suffice to solve $U\mathbf{x} = E\mathbf{b}$ instead.

Content from Strang's *ILA* 6E. Read and work through everything on p. 42 right now. This is hugely important. Then read p. 45 up to and including equation (7). This is another example of elimination. Last, read all of p. 49 (but don't worry about inverses for now).

The process in Example 2.6.7 is called **Gaussian elimination**, and we are going to use it (and sometimes do it) a lot. We start with a matrix A, multiply A by a bunch of matrices that we collect in one product E, and find that $U = EA$ has zeros in some very nice places. This is a leitmotif of our subject. Per the magisterial *Numerical Linear Algebra* by Trefethen & Bau, "The algorithms of numerical linear algebra are mainly built upon one technique used over and over again: putting zeros into matrices" (p. 191).

We are going to focus on "reducing" A to an upper-triangular form and less on back-substitution. This is mostly just a longer version of part (i) of Example 2.5.1.

2.6.8 Problem (!). Use the method of Example 2.6.7 to solve $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (0,1,6)$. Specifically, do the three elimination steps of that example on the augmented matrix $\begin{bmatrix}A & \mathbf{b}\end{bmatrix}$ to get

$$E\begin{bmatrix}A & \mathbf{b}\end{bmatrix} = \begin{bmatrix}U & \mathbf{c}\end{bmatrix}$$

with U as we found above and $\mathbf{c} = E\mathbf{b}$. (Multiplying the factors of E together to get a formula for E is a bad idea.) Now solve $U\mathbf{x} = \mathbf{c}$.

2.6.9 Problem ($\star$). We prefer upper-triangular matrices, in part for consistency, but "lower-triangular" matrices can be equally nice. Solve

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

2.6.10 Problem (+). We usually expect that matrix multiplication is not commutative. However, sometimes it is.

(i) Let $\ell_1, \ell_2 \in \mathbb{R}$ and put

$$E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -\ell_1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E_{31} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\ell_2 & 0 & 1 \end{bmatrix}.$$

Explain in words what E_{21} and E_{31} "do" (i.e., what is the effect of multiplying $E_{21}\mathbf{v}$ and $E_{31}\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^3$?). Then explain why you think this means that $E_{21}E_{31} = E_{31}E_{21}$. Do the actual matrix multiplication to convince yourself that this is true.

(ii) Let $\ell_3 \in \mathbb{R}$ and

$$E_{32} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\ell_3 & 1 \end{bmatrix}.$$

Without doing any calculations, explain why you should expect E_{31} from above and E_{32} *not* to commute. Then do the multiplication to check $E_{31}E_{32} \neq E_{32}E_{31}$.

2.6.11 Remark. *Associativity of matrix multiplication (Theorem 2.4.1) is key to how matrices act. We defined the matrix product AB in such a way that $(AB)\mathbf{v} = A(B\mathbf{v})$, at least when A, B, and $\mathbf{v}$ are sized appropriately so that all of the products involved are defined. If we think about matrices as dynamic objects, we could have the matrix B act on the vector $\mathbf{v}$ first to get the vector $B\mathbf{v}$, and then we could have the matrix A act on the vector $B\mathbf{v}$ to get the vector $A(B\mathbf{v})$. Or we could have the matrix A act on the matrix B all at once, and we get the new matrix AB. Then the matrix AB acts on the vector $\mathbf{v}$ to get the vector $(AB)\mathbf{v}$. Our choice of the definition for the symbol AB ensures that the two vectors $(AB)\mathbf{v}$ and $A(B\mathbf{v})$ are the same.*

At the level of Gaussian elimination, this allows us to collect all of the elimination matrices (and their forthcoming relatives) into one big matrix that acts on A all at once. Associativity of matrix multiplication ensures that the order in which we group those factors does not matter.

2.7. Sometimes elimination breaks down.

Elimination can break down in two ways. The first is not so bad and just requires a new kind of matrix to correct things. The second is worse and will prevent us from solving the linear system.

2.7.1 Example. What if at the jth step of elimination, the (j, j)-entry is 0, but an entry further down in column j is not 0? All hope is not lost. Consider

$$\underbrace{\begin{bmatrix} 2 & 2 & 1 \\ 4 & 4 & 3 \\ 8 & 9 & 9 \end{bmatrix}}_{A} \xrightarrow[E_{21}]{\text{R2} \mapsto \text{R2}-2\times\text{R1}} \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 8 & 9 & 9 \end{bmatrix} \xrightarrow[E_{31}]{\text{R3} \mapsto \text{R3}-4\times\text{R1}} \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 5 \end{bmatrix}.$$

The matrices E_{21} and E_{31} are the same as before in Example 2.6.7.

The problem is that the $(2, 2)$-entry is now 0, and that will not help us eliminate the 3 in the $(3, 2)$-entry: subtracting any multiple of row 2 from row 3 will not turn that 9 into a 0. But if we could "flip" rows 2 and 3, we would win. After all, the four problems

$$A\mathbf{x} = \mathbf{b}, \qquad \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 5 \end{bmatrix} \mathbf{x} = \mathbf{b}, \qquad \begin{cases} 2x_1 + 2x_2 + x_3 = b_1 \\ x_3 = b_2 \\ x_2 + 5x_3 = b_3, \end{cases}$$

$$\text{and} \quad \begin{cases} 2x_1 + 2x_2 + x_3 = b_1 \\ x_2 + 5x_3 = b_3 \\ x_3 = b_2 \end{cases}$$

are really the same

If only there were a matrix $P \in \mathbb{R}^{3\times 3}$ such that

$$P \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix}.$$

What we really want is that

$$P \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_3 \\ v_2 \end{bmatrix}.$$

We can get P by working backwards and thinking of matrix-vector multiplication as a linear combination:

$$\begin{bmatrix} v_1 \\ v_3 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

Here is the result:

$$\begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 5 \end{bmatrix} \xrightarrow[P_{23}]{\text{R3} \mapsto \text{R2, R2} \mapsto \text{R3}} \begin{bmatrix} 2 & 2 & 1 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix}, \qquad P_{23} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

We call this matrix P_{23} to emphasize that we get it by interchanging columns 2 and 3 of the identity matrix. By the way, P_{23} is a permutation matrix: each standard basis vector for $\mathbb{R}^3$ appears exactly once as a column of P_{23}. What we get is that

$$EA = U, \qquad E := P_{23}E_{31}E_{21}$$

with U upper-triangular. The matrix E is now a little more complicated than in Example 2.6.7, as we have to include a factor of a permutation matrix, not just an elimination matrix.

In general, to interchange rows i and j of $A \in \mathbb{R}^{m\times m}$, multiply $P_{ij}A$, where $P_{ij} \in \mathbb{R}^{m\times m}$ is the matrix whose columns are those of the $m \times m$ identity matrix with columns i and j interchanged. Such a matrix P_{ij} is, again, a permutation matrix—just a pretty simple one, because only two columns of I_m appear out of order. So, if at some stage of elimination, the diagonal entry that we want to use to eliminate entries below is 0, but other entries in that column are nonzero, we just "permute" the rows to bring that nonzero entry up to the row that you want. Then eliminate as usual in the remaining rows.

Also, now we see two very different actions of a permutation matrix. Let $A \in \mathbb{R}^{m\times n}$. If $P \in \mathbb{R}^{m\times m}$ is a permutation matrix, then PA reorders the rows of A. But if $\widetilde{P} \in \mathbb{R}^{n\times n}$ is a permutation matrix, then $A\widetilde{P}$ reorders the columns of A. If you forget the pattern, just take

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and compute PA and AP.

Content from Strang's *ILA* 6E. Read "Possible breakdown of elimination" on p. 43 up to but not including the "Caution!" paragraph. Then read p. 45 after equation (1) and look at the calculation in "Exchange rows 2 and 3." These P_{ij} permutation matrices are special cases of a more general permutation matrix structure, which is the identity matrix with its columns (equivalently, rows) rearranged in various ways. See pp. 64–65. We won't need those more general permutation matrices for a while.

2.7.2 Problem (!). Explain in words (no need for any calculations) why $P_{ij}A = P_{ji}A$.

2.7.3 Problem ($\star$). Let $P_{13} \in \mathbb{R}^{3\times 3}$ be the permutation matrix that interchanges columns 1 and 3 of the 3×3 identity matrix. Compute $P_{13}A$ and AP_{13} for an arbitrary $A \in \mathbb{R}^{3\times 3}$. Then conjecture about what the different effects of multiplying $P_{ij}A$ and AP_{ij} are for an arbitrary $A \in \mathbb{R}^{m\times m}$ and an arbitrary permutation matrix $P_{ij} \in \mathbb{R}^{m\times m}$ that interchanges columns i and j of the $m \times m$ identity matrix. (You do not have to prove your conjecture.)

2.7.4 Problem (+). Let $A \in \mathbb{R}^{m\times n}$ and let $S \in \mathbb{R}^{n\times d}$ be a matrix whose columns are some of the columns of the $n \times n$ identity matrix. Here $d \geq 1$ is any integer, and the columns of the identity may be repeated, and some columns of the identity may not appear at all. Describe in words the structure of the matrix AS. [Hint: *the letter S might stand for "selection" matrix—what is being "selected" here?*]

Here is the nastier breakdown of elimination: what if at some step, the diagonal entry that we want to use to eliminate entries below is 0 and *all* other entries in that column are 0, too? Good news is that we do not have to do any more elimination on entries in that column, as they are already 0. Bad news is that will not be able to solve $A\mathbf{x} = \mathbf{b}$ for all $\mathbf{b}$.

2.7.5 Example. The matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

is problematic. We eliminate:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 0 & 0 & 5 \end{bmatrix} \xrightarrow[E_{21}]{\text{R2} \mapsto \text{R2}-2\times\text{R1}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \qquad E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This may not look so problematic right now. We would want to use the $(2,2)$-entry in $E_{21}A$ to eliminate the $(3,2)$-entry, but the $(3,2)$-entry is already 0. So, $E_{21}A$ is already upper-triangular! Why is this not enough for us to be happy?

What if we actually try to solve $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$ arbitrary? If $A\mathbf{x} = \mathbf{b}$, then $E_{21}A\mathbf{x} = E_{21}\mathbf{b} = (b_1, b_2 - 2b_1, b_3)$. Thus we want

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix} \mathbf{x} = \begin{bmatrix} b_1 \\ b_2 - 2b_1 \\ b_3 \end{bmatrix}.$$

At the level of actual equations, this is

$$\begin{cases} x_1 + 2x_2 + 3x_3 = b_1 \\ 0 = b_2 - 2b_1 \\ 5x_3 = b_3. \end{cases}$$

Look at that second equation: it says $b_2 - 2b_1 = 0$, equivalently, $b_2 = 2b_1$. Think about the logic here. We assumed that $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} = (b_1, b_2, b_3)$, and we deduced that $b_2 = 2b_1$. This means that $\mathbf{b}$ cannot be just any vector in $\mathbb{R}^3$; it has to satisfy this "solvability condition" of $b_2 = 2b_1$. Surely not every vector in $\mathbb{R}^3$ does this—for example, take $\mathbf{b} = (1, 0, 0)$. So we cannot always solve $A\mathbf{x} = \mathbf{b}$.

This is worth interpreting in the context of the column space. Look at the structure of A: the second row is twice the first row. More precisely,

$$A\mathbf{x} = \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ 2x_1 + 4x_2 + 6x_3 \\ 5x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ 2(x_1 + 2x_2 + 3x_3) \\ 5x_3 \end{bmatrix}.$$

So, if $\mathbf{b} = (b_1, b_2, b_3) \in \mathbf{C}(A)$, then $b_2 = 2b_1$. This is exactly the solvability condition that we deduced from elimination.

2.7.6 Problem ($\star$). Does the arrow "go the other way"? The previous example showed

$$\mathbf{b} \in \mathbf{C}(A) \Longrightarrow b_2 = 2b_1.$$

Do we have

$$b_2 = 2b_1 \Longrightarrow \mathbf{b} \in \mathbf{C}(A)?$$

Yes! If $b_2 = 2b_1$, then $A\mathbf{x} = \mathbf{b}$ is the system

$$\begin{cases} x_1 + 2x_2 + 3x_3 = b_1 \\ 2x_1 + 4x_2 + 6x_3 = 2b_1 \\ 5x_3 = b_3. \end{cases}$$

Use the third equation to solve for x_3, take x_2 to be any number that you like, and then use the first equation to write x_1 in terms of the values forced on x_3 and chosen for x_2. Why does this also satisfy the second equation automatically?

Content from Strang's *ILA* 6E. Read the rest of "Possible Breakdown of Elimination" on p. 43 starting with "Caution!"

2.8. Gaussian elimination and upper-triangular matrices are friends.

These results will follow and support us for the rest of the course and beyond. Here is an abstraction of our elimination procedure.

2.8.1 Theorem (Gaussian elimination). *Let $A \in \mathbb{R}^{m \times m}$. Then there exist matrices E, $U \in \mathbb{R}^{m \times m}$ with the following properties.*

(i) $EA = U$.

(ii) *U is upper-triangular.*

(iii) *E is the product of elimination matrices E_{ij} and/or permutation matrices P_{ij}.*

Proof. If the $(1,1)$-entry of A is nonzero, multiply A by elimination matrices $E_{21}, \ldots, E_{m1}$ to subtract multiples of row 1 of A from rows 2 through m of A. Call the product of these elimination matrices E_1. If $m = 2$, then $E_1 A$ is upper-triangular. If $m \geq 3$ and the $(2,2)$-entry of $E_1 A$ is nonzero, multiply $E_1 A$ by elimination matrices $E_{32}, \ldots, E_{m2}$ to subtract multiples of row 2 of $E_1 A$ from rows 3 through m of $E_1 A$. Call the product of these elimination matrices E_2. If $m = 3$, then $E_2 E_1 A$ is upper-triangular. Otherwise, turn the crank and keep going.

If at any stage the (j, j)-entry is zero and the entries in column j in rows $j + 1$ through m are zero, just proceed to the next step and consider the $(j + 1, j + 1)$-entry. If the (j, j)-entry is zero and some entry in rows $j + 1$ through m of column j is nonzero, multiply by a permutation matrix so that this nonzero entry is now the (j, j)-entry. Then eliminate as before. Call the product of the elimination matrices and the permutation matrices E_j. ■

What this result says is that if $A\mathbf{x} = \mathbf{b}$, then $EA\mathbf{x} = E\mathbf{b}$, and so $U\mathbf{x} = E\mathbf{b}$. The upper-triangular system $U\mathbf{x} = E\mathbf{b}$ is much easier to solve, and so we like it. At least, we like it *when the diagonal entries of U are nonzero.*

2.8.2 Theorem. *Let $U \in \mathbb{R}^{m \times m}$ be an upper-triangular matrix whose diagonal entries are nonzero. Then for any $\mathbf{c} \in \mathbb{R}^m$, there exists a unique $\mathbf{x} \in \mathbb{R}^m$ such that $U\mathbf{x} = \mathbf{c}$.*

Proof. This is really back-substitution in the abstract. Here is the proof for $m = 3$. Take

$$U = \begin{bmatrix} u_{11} & * & * \\ 0 & u_{22} & * \\ 0 & 0 & u_{33} \end{bmatrix},$$

where u_{11}, u_{22}, and u_{33} are nonzero. So to solve $U\mathbf{x} = \mathbf{c}$ with $\mathbf{c} = (c_1, c_2, c_3)$, first we consider

$$u_{33} x_3 = c_3.$$

Since $u_{33} \neq 0$, we can divide to find that x_3 must be

$$x_3 = \frac{c_3}{u_{33}}.$$

Go back up a step and look at

$$u_{22} x_2 + \text{stuff depending on } x_3 = c_2.$$

The point is that we know what this "stuff" is because we know x_3 exactly. Solve this as

$$x_2 = \frac{c_2 - \text{stuff}}{u_{22}}.$$

This is the only choice for x_2. Do the same for x_1. ■

But are we really sure that if $EA = U$, then a solution to $U\mathbf{x} = E\mathbf{b}$ is also a solution to $A\mathbf{x} = \mathbf{b}$? For small problems, we can check it by plug-and-chug, but why is this true in general?

2.9. Matrix inverses move us forward and backward.

The time has come to be sure that we can "invert" E, and this is a good reason to study matrix inverses in general. We will overall be much more concerned with properties of inverses than formulas for inverses. There is an algorithm that will let you do that, and we will see it briefly, but we mostly abide by the slogan "What things do defines what things are."

Content from Strang's *ILA* 6E. Read the first two paragraphs on p. 50.

Here is what we want: why does $EA\mathbf{x} = E\mathbf{b}$ imply $A\mathbf{x} = \mathbf{b}$? More abstractly, if $E \in \mathbb{R}^{m\times m}$ and $E\mathbf{v} = E\mathbf{b}$ for some $\mathbf{v}$, $\mathbf{b} \in \mathbb{R}^m$, do we necessarily have $\mathbf{v} = \mathbf{b}$? It would be nice if we could "undo" the "action" of E by multiplying by another matrix. Is there $F \in \mathbb{R}^{m\times m}$ such that $F(E\mathbf{w}) = \mathbf{w}$ for all $\mathbf{w} \in \mathbb{R}^m$? If so, then assuming $E\mathbf{v} = E\mathbf{b}$ gives $F(E\mathbf{v}) = F(E\mathbf{b})$, and thus $\mathbf{v} = \mathbf{b}$ as desired.

Look more closely at the equation $F(E\mathbf{w}) = \mathbf{w}$. This just says $(FE)\mathbf{w} = \mathbf{w}$. What does that tell us about the matrix product FE? If $(FE)\mathbf{w} = \mathbf{w}$ for all $\mathbf{v} \in \mathbb{R}^m$, then we could take $\mathbf{w} = \mathbf{e}_j$ as the standard basis vectors. We find $(FE)\mathbf{e}_j = \mathbf{e}_j$, and so the jth column of FE must be $\mathbf{e}_j$: the jth column of the $m \times m$ identity matrix. That is, we want $FE = I_m$.

We are actually going to ask for a little bit more in the following definition: that $EF = I_m$ as well. This is an artifact of our intuition from multiplication of real numbers (if $ab = 1$ for $a, b \in \mathbb{R}$, then of course $ba = 1$), but it is necessary to require here since matrix multiplication is not commutative. (That is, just because we have $FE = I_m$ *should not* automatically imply that $EF = I_m$. Surprisingly, and gloriously, it does, but that takes some work.)

2.9.1 Definition. *A matrix $E \in \mathbb{R}^{m\times m}$ is* **INVERTIBLE** *if there exists a matrix $F \in \mathbb{R}^{m\times m}$ such that*

$$FE = I_m \quad \text{and} \quad EF = I_m. \tag{2.9.1}$$

2.9.2 Example. **(i)** Let

$$E = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$$

be the elimination matrix that subtracts 2 times the first row from the second row. Can we invert E? We're done if we find $F \in \mathbb{R}^{2\times 2}$ such that $EF = FE = I_2$. What *should* F be? This is where it might help to think about E *dynamically*: what does E do? We just said it: E subtracts 2 times the first row from the second row. So undoing E should add two times the first row to the second row. That is,

$$E\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} v_1 \\ (v_2 - 2v_1) + 2v_1 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

So maybe

$$F = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$

works. Check it yourself.

(ii) Let

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

be the permutation matrix that interchanges rows 1 and 2. Undoing P should interchange those rows again: we want $F \in \mathbb{R}^{2\times 2}$ such that if

$$P \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_2 \\ v_1 \end{bmatrix}, \quad \text{then} \quad F \begin{bmatrix} v_2 \\ v_1 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

This looks like we should just take $F = P$. Check that $P^2 = I_2$. By the way, we are using "power" notation for matrix multiplication: $P^2 = PP$.

Content from Strang's *ILA* 6E. Read Examples 4 and 5 on p. 52 about inverting elimination matrices. Skip the remarks about the inverse of FE in Example 5 for now.

Example 2.9.2 should be comforting in that it suggests that elimination and permutation matrices are invertible. We probably want to say that their "inverses" are what we expect: invert subtracting by adding, invert permuting by permuting again. What gives us the right to say that a matrix has only one inverse? A (nonzero) real number has only one reciprocal to undo multiplication, but why is this true for matrices?

Here is why. Suppose that E has "two" inverses F_1 and F_2, so

$$F_1E = EF_1 = F_2E = EF_2 = I_m. \tag{2.9.2}$$

We need to show that $F_1 = F_2$. Here is a great trick: multiply by 1. We know that $1x = x$ for any $x \in \mathbb{R}$, and the same is true for matrices.

2.9.3 Problem (!). Check that $AI_m = I_mA = A$ for any $A \in \mathbb{R}^{m\times m}$.

So,

$$F_1 = F_1I_m = F_1(EF_2) = (F_1E)F_2 = I_mF_2 = F_2. \tag{2.9.3}$$

Here is the formal result.

2.9.4 Theorem. *Let $E \in \mathbb{R}^{m\times m}$. There exists at most one $F \in \mathbb{R}^{m\times m}$ satisfying* (2.9.1).

Content from Strang's *ILA* 6E. This is Note 2 on p. 50.

We can now talk about "the" inverse of a matrix.

2.9.5 Definition. *Let $E \in \mathbb{R}^{m\times m}$ be invertible. The* **INVERSE** *of E is the unique matrix F satisfying*

$$FE = EF = I_m,$$

and we write $F = E^{-1}$.

We generalize Example 2.9.2.

2.9.6 Theorem. *Elimination and permutation matrices are invertible.*

(i) *Let* $E_{ij} \in \mathbb{R}^{m \times m}$ *be the elimination matrix that subtracts* ℓ *times row* j *from row* i *(so* 1*'s along the diagonal,* $-\ell$ *in the* (i,j)*-entry, and* 0 *everywhere else). Then* E_{ij} *is invertible, and* E_{ij}^{-1} *is the elimination matrix that adds* ℓ *times row* j *to row* i *(so* 1*'s along the diagonal,* ℓ *in the* (i,j)*-entry, and* 0 *everywhere else).*

(ii) *Let* $P_{ij} \in \mathbb{R}^{m \times m}$ *be the permutation matrix that interchanges rows* i *and* j *(so* P_{ij} *is the* $m \times m$ *identity matrix with columns* i *and* j *interchanged). Then* P_{ij} *is invertible and* $P_{ij}^{-1} = P_{ij}$.

Definition 2.4.3 gave a more general definition of permutation matrix. Eventually we will figure out how to invert an arbitrary permutation matrix (short answer: if an arbitrary permutation matrix interchanges a bunch of rows, interchange them back), but the only permutation matrices that we need for Gaussian elimination interchange only two rows at a time.

2.9.7 Problem (!). We probably expect that undoing the undoing of an action does that action. More precisely, if $E \in \mathbb{R}^{m \times m}$ is invertible, we should expect that E^{-1} is also invertible and $(E^{-1})^{-1} = E$. Prove this by showing that E satisfies the definition of inverse for E^{-1}. *What things do defines what things are.*

Now go back and look very carefully at the calculation in (2.9.3). We did not use all of the equalities in (2.9.2). Instead, we only needed that $F_1E = I_m$ and $EF_2 = I_m$. We might call F_1 a **LEFT INVERSE** and F_2 a **RIGHT INVERSE**. Here is what we have proved.

2.9.8 Corollary. *Let* $E \in \mathbb{R}^{m \times m}$ *have left and right inverses in the sense that there are* F_1, $F_2 \in \mathbb{R}^{m \times m}$ *such that*

$$F_1E = I_m \quad \text{and} \quad EF_2 = I_m.$$

Then E *is invertible and* $F_1 = F_2 = E^{-1}$.

Proof. Here is a summary of what we have already done. The calculation in (2.9.3) shows $F_1 = F_2$. Put $F = F_1$. Then the hypotheses give $FE = F_1E = I_m$ and $EF = EF_2 = I_m$, and so F satisfies Definition 2.9.5. ■

2.9.9 Problem (!). Here is a situation in which having a left inverse *and* some more information imply invertibility. Let E, $A \in \mathbb{R}^{m \times m}$. Suppose that $EA = I_m$ and E is invertible. Prove that A is invertible, too.

We are particularly interested in inverting a matrix that is a product of elimination matrices and permutation matrices. We know that any elimination or permutation matrix is invertible. More generally, is the product of invertible matrices invertible?

Yes. Suppose that A, $B \in \mathbb{R}^{m \times m}$ are invertible. We will show that AB is invertible. Think about action: first we do B to a vector $\mathbf{v}$ by multiplying $B\mathbf{v}$, and then we do A by multiplying $A(B\mathbf{v}) = (AB)\mathbf{v}$. To undo AB, we probably want to undo A first and then B. (Getting dressed, socks go on first, then shoes; getting undressed, shoes come off first, then socks.) So we might guess that $(AB)^{-1} = B^{-1}A^{-1}$. The good news is that we can check this using the definition:

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_mB = B^{-1}B = I_m.$$

2.9.10 Problem (!). Check that $(AB)(B^{-1}A^{-1}) = I_m$ as well.

Here is the formal result.

2.9.11 Theorem. *Let A, $B \in \mathbb{R}^{m \times m}$ be invertible. Then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.*

Content from Strang's *ILA* 6E. Read "The Inverse of a Product AB" on pp. 51–52. Then go back to Example 5 on p. 52. The point for our larger story is that multiplying elimination matrices together when getting $EA = U$ is not the best of ideas, whereas computing E^{-1} is more meaningful.

This seems to be everything that we want. Theorem 2.8.1 tells us that for any $A \in \mathbb{R}^{m \times m}$, we can always find a product of elimination and/or permutation matrices, which we call E, such that $EA = U$ is upper-triangular. Now we know that E is invertible. Given $\mathbf{b} \in \mathbb{R}^m$, it is usually easier to solve $U\mathbf{x} = E\mathbf{b}$, and then we have $E^{-1}U\mathbf{x} = E^{-1}(E\mathbf{b})$, where

$$E^{-1}U = E^{-1}(EA) = (E^{-1}E)A = I_mA = A \quad \text{and} \quad E^{-1}(E\mathbf{b}) = (E^{-1}E)\mathbf{b} = \mathbf{b}.$$

Thus $A\mathbf{x} = \mathbf{b}$, which is what we always wanted to be sure of.

Invertibility is another way of asking about solvability of linear systems. Suppose that $A \in \mathbb{R}^{m \times m}$ is invertible. We claim that $A\mathbf{x} = \mathbf{b}$ always has a solution, and that solution is unique. For uniqueness, work backwards and assume $A\mathbf{x} = \mathbf{b}$; then $A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b}$, and so $\mathbf{x} = A^{-1}\mathbf{b}$. To check that this is actually a solution, plug in: $A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_m\mathbf{b} = \mathbf{b}$.

2.9.12 Theorem. *Let $A \in \mathbb{R}^{m \times m}$ be invertible and $\mathbf{b} \in \mathbb{R}^m$. Then the problem $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.*

Content from Strang's *ILA* 6E. This is Note 3 on p. 50.

2.9.13 Problem (!). Hugely important: prove the following for an invertible $A \in \mathbb{R}^{m \times m}$.

(i) $\mathbf{N}(A) = \{\mathbf{0}_m\}$. (So if $\mathbf{N}(A)$ is bigger than $\{\mathbf{0}_m\}$, then A is not invertible.)

(ii) $\mathbf{C}(A) = \mathbb{R}^m$. (So if $\mathbf{C}(A)$ is smaller than $\mathbb{R}^m$, then A is not invertible.)

(iii) The columns of A are independent. (So if the columns of A are dependent, then A is not invertible.)

(iv) Every column of A is a pivot column and $\text{rank}(A) = m$.

Does the logic go the other way? If $A\mathbf{x} = \mathbf{b}$ always has a unique solution, is A invertible? If $\mathbf{N}(A) = \{\mathbf{0}_m\}$ (uniqueness guaranteed, maybe not existence), is A invertible? If $\mathbf{C}(A) = \mathbb{R}^m$ (existence guaranteed, maybe not uniqueness), is A invertible? If the columns of A are independent, is A invertible? Yes, yes, yes, and yes. But establishing all of that needs some preparation.

2.9.14 Problem ($\star$). Often knowing that a matrix is invertible is more useful than having a formula for that inverse. Here is a situation in which the presence of an invertible matrix "keeps things the same." Let $A \in \mathbb{R}^{m \times n}$ be any matrix and let $B \in \mathbb{R}^{n \times n}$ be invertible. Show that $\mathbf{C}(AB) = \mathbf{C}(A)$ as follows. First, explain why $AB\mathbf{v} \in \mathbf{C}(A)$ for any $\mathbf{v} \in \mathbb{R}^n$. Next, justify the equality $A\mathbf{x} = (AB)(B^{-1}\mathbf{x})$ and explain how that shows that anything in $\mathbf{C}(A)$ is in $\mathbf{C}(AB)$.

Content from Strang's *ILA* 6E. I am not going to talk about determinants now, or much later (I hope!), but you should read Note 6 on p. 50 and Example 2 on p. 51 and also think about the four 2×2 matrices in Example 3 on p. 51. Determinants are a quick and easy way of understanding 2×2 matrices, which arise in a lot of applications (e.g., ordinary differential equations). Try using Note 6 to solve our original problem

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

Using the solution formula $\mathbf{x} = A^{-1}\mathbf{b}$ from Theorem 2.9.12 in practice requires us to compute A^{-1}. This turns out to be "expensive" computationally, rather more so than elimination and back-substitution.

Content from Strang's *ILA* 6E. Read "The Cost of Elimination" on pp. 57–58. The following link to a section from the fifth edition elaborates on this:

https://math.mit.edu/ gs/linearalgebra/ila5/linearalgebra5_11-1.pdf.

The point is that using A^{-1} to solve $A\mathbf{x} = \mathbf{b}$ for $A \in \mathbb{R}^{m \times m}$ might take around m^3 arithmetical operations, but using elimination would take only around $m^3/3$ operations. If this excites you, take a numerical linear algebra class. Read the beautiful book by Trefethen & Bau, too.

2.9.15 Problem (!). Let $A \in \mathbb{R}^{m \times m}$ be invertible. What are the factors C and R in the CR-factorization of A? Compare your answer to your result for Problem 2.3.3.

2.10. Gauss–Jordan elimination inverts matrices.

We go back to elimination in the context of inverses. How does being able to solve a linear system $A\mathbf{x} = \mathbf{b}$ via elimination say anything about the invertibility of A?

We start with the nicest case: upper-triangular. We can eliminate "upwards" on an upper-triangular matrix with nonzero diagonal entries to find an invertible matrix E such that $EU = I_m$. Then $U = E^{-1}$, and so U is invertible. Here is how this works.

2.10.1 Example. Let

$$U = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

We met this matrix in Example 2.6.7. We turn U into I_3 starting from the bottom.

The first thing to do is to make that entry of 2 in the $(3,3)$-slot into a 1. This requires division by 2 in the third row. Of course we want to encode this, like everything else, via matrix multiplication. What matrix $D \in \mathbb{R}^{3\times 3}$ does that? We want

$$D \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3/2 \end{bmatrix}.$$

Expand the vector on the right as a linear combination weighted by v_1, v_2, and v_3 to see that D should be

$$D_{33} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}.$$

We call it D_{33} now because the action is happening in the $(3,3)$-entry.

So, we have the transformation

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} \xrightarrow[D_{33}]{\text{R3} \mapsto (1/2)\times\text{R3}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad D_{33} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}.$$

This **scaling matrix** D_{33}, along with the elimination and permutation matrices, is the last of the so-called **elementary matrices** that we need to encode "row operations" on matrices.

Now we eliminate "upwards." We want the other entries in column 3 to be 0, so we subtract multiples of row 3 from rows 1 and 2. (Well, multiples of 1.) We get

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[E_{23}]{\text{R2} \mapsto \text{R2}-\text{R3}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{23} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[E_{13}]{\text{R1} \mapsto \text{R1}-\text{R3}} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{13} := \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

And then we subtract a multiple of row 2 from row 1 to make that $(1,2)$-entry 0:

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[E_{12}]{\text{R1} \mapsto \text{R1}-\text{R2}} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{12} := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Last, we rescale the first row:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[D_{11}]{\text{R1} \mapsto (1/2)\times\text{R1}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3, \qquad D_{11} := \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We conclude

$$D_{11}E_{12}E_{13}E_{23}D_{33}U = I_3,$$

so putting

$$E := D_{11}E_{12}E_{13}E_{23}D_{33}$$

gives $EU = I_3$. Certainly E is invertible, as all elimination matrices are invertible, and scaling matrices are invertible when their diagonal entries are nonzero. Then $U = E^{-1}I_3 = E^{-1}$, and so U is invertible with $U^{-1} = E$.

2.10.2 Problem ($\star$). Let $D \in \mathbb{R}^{m\times m}$ be **DIAGONAL**: the (i,j)-entry of D is 0 for $i \neq j$. Prove that if all of the diagonal entries of D are nonzero, then D is invertible; give an explicit formula for D^{-1}.

2.10.3 Example. Here is how all of the elementary matrices work. Consider the problem

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2, \end{cases}$$

which compresses as the matrix-vector equation

$$A\mathbf{x} = \mathbf{b}, \qquad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

and also as the augmented matrix

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \left[\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{array}\right].$$

(i) Suppose that $a_{11} \neq 0$. We want to reduce A to upper-triangular form. We subtract a_{21}/a_{11} times the first equation from the second equation:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases} \iff \begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ 0 + (a_{22} - a_{21}/a_{11})x_2 = b_2 - (a_{21}/a_{11})b_2. \end{cases}$$

Equivalently, we multiply the augmented matrix by an elimination matrix:

$$E_{21}\begin{bmatrix}A & \mathbf{b}\end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (-a_{21}/a_{11}) & 1 \end{bmatrix}\left[\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{array}\right]$$
$$= \left[\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ 0 & (a_{22} - a_{21}/a_{11})x_2 & b_2 - (a_{21}/a_{11})b_2 \end{array}\right].$$

(ii) Suppose that we want to interchange the equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases} \iff \begin{cases} a_{21}x_1 + a_{22}x_2 = b_2 \\ a_{11}x_1 + a_{12}x_2 = b_1. \end{cases}$$

We multiply the augmented matrix by a permutation matrix:

$$P_{12}\begin{bmatrix}A & \mathbf{b}\end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\left[\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{array}\right] = \left[\begin{array}{cc|c} a_{21} & a_{22} & b_2 \\ a_{11} & a_{12} & b_1 \end{array}\right].$$

(iii) Suppose that $a_{12} = 0$ and $a_{22} \neq 0$. We want to rescale the second equation:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ 0 + a_{22}x_2 = b_2 \end{cases} \iff \begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ 0 + (a_{22}/a_{21})x_2 = b_2/a_{21}. \end{cases}$$

Equivalently, we multiply the augmented matrix by a scaling matrix:

$$D_{22}\begin{bmatrix}A & \mathbf{b}\end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & (1/a_{22}) \end{bmatrix}\left[\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ 0 & a_{22} & b_2 \end{array}\right] = \left[\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ 0 & 1 & b_2/a_{22} \end{array}\right].$$

The arithmetic in Example 2.10.1 is called **Gauss–Jordan elimination**. Here is how this works in the abstract.

2.10.4 Theorem (Gauss–Jordan elimination). *Let $U \in \mathbb{R}^{m\times m}$ be upper-triangular with nonzero diagonal entries. Then there exists an invertible matrix $E \in \mathbb{R}^{m\times m}$, which is the product of elimination and/or scaling matrices (but not permutation matrices), such that $EU = I_m$. Consequently, $U = E^{-1}$ is invertible.*

Proof. This should feel basically the same as the proof of Theorem 2.8.1. Multiply U by a scaling matrix D_{mm} to divide row m by $u_{mm} \neq 0$ so that the (m,m)-entry of $D_{mm}U$ is 1. Then subtract multiples of row m from rows $m-1$ through 1 to create zeros in rows $m-1$ through 1 of column m. Go to the $(m-1,m-1)$-entry: rescale to make it 1, create zeros in rows $m-2$ through 1 of column $m-1$ through elimination. Repeat. Let E be the product of all of the scaling and/or elimination matrices used, in the order that you use them from the bottom up at each stage. No need for permutation matrices because all of the diagonal entries are nonzero. ■

2.10.5 Remark. *Previously we used "Gaussian elimination" on an arbitrary $A \in \mathbb{R}^{m\times m}$ to find an invertible matrix $E \in \mathbb{R}^{m\times m}$ such that $EA = U$ with U upper-triangular. Now, in the special case that the diagonal entries of U were nonzero, we used "Gauss–Jordan elimination" to find another invertible matrix $\widetilde{E}$ such that $\widetilde{E}U = I_m$, thus $\widetilde{E}EA = I_m$, so A is invertible with $A^{-1} = (\widetilde{E}E)^{-1}$.*

Content from Strang's *ILA* 6E. Page 57 offers an algorithm for computing A^{-1} by hand if you really need to do it for a small A. I will never ask you to do that, and Strang gives a few problems asking for an explicit calculation (Problems 29, 31, 32 in Section 2.2 if you're curious)—that's how deprecated the method is. Far better to *understand* A^{-1} than have a general formula for it.

2.10.6 Problem (!). Explain why the matrix A from Example 2.6.7 is invertible. What is A^{-1}? (No one really cares what the exact formula is, so just express A^{-1} as the product of the inverses of a bunch of elimination, scaling, and/or permutation matrices.)

2.10.7 Problem ($\star$). All matrices in this problem are in $\mathbb{R}^{3\times 3}$. Let E_{21} be the elimination matrix that subtracts 2 times row 1 from row 2. Let $\widetilde{E}_{21}$ be the elimination matrix that subtracts row 1 from row 2. Let D_{11} be the scaling matrix that multiplies row 1 by 2. Show that $E_{21} = D_{11}^{-1}\widetilde{E}_{21}D_{11}$.

2.11. The LU-factorization is computationally practical.

We have learned a lot about invertible matrices—in particular that we can always solve $A\mathbf{x} = \mathbf{b}$ with $\mathbf{x} = A^{-1}\mathbf{b}$ when A is invertible, but that we probably should not because computing A^{-1} is computationally expensive. The alternative is that we do elimination on A so that $U := EA$ is upper-triangular with nonzero diagonal entries, and then we solve $U\mathbf{x} = E\mathbf{b}$ via back-substitution. That requires us to compute $E\mathbf{b}$, too. There is a variation on this approach that is still computationally less expensive than computing A^{-1} and that gives us some new insights into matrix multiplication, so it's worth learning. We start with a very concrete example.

2.11.1 Example. Let

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{bmatrix}.$$

We saw in Example 2.6.7 that

$$EA = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} = U,$$

where

$$E := E_{32}E_{31}E_{21}$$

and

$$E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{31} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad E_{32} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}.$$

We went further in Example 2.10.1 and found $\widetilde{E}$ such that $\widetilde{E}U = I_3$, but that's less important here. Rather, the new thing to focus on is the *factorization*

$$A = E^{-1}U.$$

Recall that we originally talked about multiplying matrices with the goal of *factoring* matrices: breaking matrices into products of simpler matrices to reveal meaningful properties. What is simpler about the matrices E^{-1} and U, and what is meaningful about the factorization $A = E^{-1}U$?

Certainly U is simpler than A because U is upper-triangular: U has a nice structure with a lot of simple data—many zero entries. What about E^{-1}? A *bad* idea is to compute E as the product $E = E_{32}E_{31}E_{21}$ and then try to compute E^{-1} from that; we have never actually computed the inverse of a matrix as complicated as that E will be. But we do know that

$$E^{-1} = (E_{32}E_{31}E_{21})^{-1} = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}.$$

And we know what each of these inverses are because they are inverses of elimination matrices:

$$E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}, \qquad E_{31}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad E_{21}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Now think about what they are doing. Multiplying by E_{31}^{-1} says "Add 4 times row 1 to row 3":

$$E_{31}^{-1}E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix}.$$

says add 4 times row 1 to row 3. Multiplying by E_{21}^{-1} says "Add 2 times row 1 to row 2":

$$E^{-1} = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = E_{21}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix} =: L.$$

Look at that matrix L. It is **LOWER-TRIANGULAR**, because every entry above the diagonal is 0. And the entries below the diagonal are the negatives of the multipliers from the original elimination step. *This is no accident.*

How does this factorization $A = LU$ help? We solve $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} = (0, 1, 5)$ (which appeared in Problem 2.6.8). The problem $A\mathbf{x} = \mathbf{b}$ is the same as $LU\mathbf{x} = \mathbf{b}$. Now here is the trick: abbreviate $\mathbf{c} := U\mathbf{x}$. Then we want $L\mathbf{c} = \mathbf{b}$. The clever idea is to view $\mathbf{c}$ as an unknown; then we can solve $L\mathbf{c} = \mathbf{b}$ using back-substitution, and then we solve $U\mathbf{x} = \mathbf{c}$ with another round of back-substitution. Nowhere does elimination hit $\mathbf{b}$.

Here we go: $L\mathbf{c} = \mathbf{b}$ is the system

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix} \mathbf{c} = \begin{bmatrix} 0 \\ 1 \\ 5 \end{bmatrix},$$

equivalently

$$\begin{cases} c_1 & & & = & 0 \\ 2c_1 & + & c_2 & & = & 1 \\ 4c_1 & + & 3c_2 & + & c_3 & = & 5 \end{cases}$$

The first equation immediately gives $c_1 = 0$, so the second reduces to $c_2 = 1$, and then the third is $3 + c_3 = 5$, thus $c_3 = 2$. Hence

$$\mathbf{c} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

Next, $U\mathbf{x} = \mathbf{c}$ is the system

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix},$$

equivalently

$$\begin{cases} 2x_1 + x_2 + x_3 = 0 \\ \phantom{2x_1 + {}} x_2 + x_3 = 1 \\ \phantom{2x_1 + x_2 + {}} 2x_3 = 2 \end{cases}$$

The third equation is $2x_3 = 2$, thus $x_3 = 1$. Then the second equation is $x_2 + 1 = 1$, so $x_2 = 0$. And the first equation is then $2x_1 + 0 + 1 = 0$, so $2x_1 = -1$, and therefore $x_1 = -1/2$. That is, $\mathbf{x} = (-1/2, 0, 1)$.

This example has a number of lessons for us. First, if we can factor $A = LU$, with L lower-triangular and U upper-triangular, and where both L and U have all nonzero entries on their diagonals, then we can solve $A\mathbf{x} = \mathbf{b}$ easily by back-substitution *and without doing any elimination calculations on* $\mathbf{b}$. Second, we *might* be able to achieve this "LU-factorization" if we can reduce A to upper-triangular form using only elimination, not permutation, matrices. In particular, finding that factor of L involved inverting the product of elimination matrices that governed that reduction—but we did not multiply all those elimination matrices together and then calculate the inverse; instead, we used properties of inverses of products and *what elimination matrices do.* (Brute force may not always be the best force.)

All of this turns out to be more generally true.

2.11.2 Theorem (LU-factorization). *Suppose that $A \in \mathbb{R}^{m \times m}$ can be reduced to upper-triangular form using only elimination, not permutation, matrices. That is, there is $E \in \mathbb{R}^{m \times m}$ such that $EA = U$, where U is upper-triangular and E is a product of only elimination matrices. Then $L := E^{-1}$ is lower-triangular, the diagonal entries of L are all 1, and $A = LU$. Moreover, for any $\mathbf{b} \in \mathbb{R}^m$, there is $\mathbf{x} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{b}$ if and only if there is $\mathbf{c} \in \mathbb{R}^m$ such that*

$$\begin{cases} L\mathbf{c} = \mathbf{b} \\ U\mathbf{x} = \mathbf{c}. \end{cases} \tag{2.11.1}$$

Proof. We are only going to prove the last sentence. The proof that L is lower-triangular when E is a product of only elimination matrices is essentially an abstraction of the calculation in Example 2.11.1. (Try replacing the multipliers 2, 4, and 3 with arbitrary ℓ_{21}, ℓ_{31}, $\ell_{32} \in \mathbb{R}$ and watch the same lower-triangular structure appear.)

Here is the proof of that last sentence, assuming that we have the factorization $A = LU$. First, if there is $\mathbf{x} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{b}$, then $LU\mathbf{x} = \mathbf{b}$. Put $\mathbf{c} = U\mathbf{x}$ to find $L\mathbf{c} = \mathbf{b}$. So, both equations in (2.11.1) are true.

Now suppose that both equations in (2.11.1) are true. Work backwards:

$$\mathbf{b} = L\mathbf{c} = L(U\mathbf{x}) = (LU)\mathbf{x} = A\mathbf{x}.$$

By the way, the proof of that last sentence did not use *at all* the fact that L and U are triangular or that L has diagonal entries equal to 1. However, if we wanted to start by solving (2.11.1) and end up with a solution to $A\mathbf{x} = \mathbf{b}$, it would be necessary for L and U to have all nonzero diagonal entries. ■

2.11.3 Problem ($\star$). This is yet another perspective on our original toy problem (1.2.1). Let

$$A = \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}.$$

Find matrices L, $U \in \mathbb{R}^{2 \times 2}$ such that L is lower-triangular, U is upper-triangular, and $A = LU$. Let $\mathbf{b} = (1, 11)$. Solve $A\mathbf{x} = \mathbf{b}$ by first solving $L\mathbf{c} = \mathbf{b}$ for some $\mathbf{c} \in \mathbb{R}^2$ and then solving $U\mathbf{x} = \mathbf{c}$ for $\mathbf{x} \in \mathbb{R}^2$.

Content from Strang's *ILA* 6E. Here are sketches of the existence of the LU-factorization. First, reread Example 5 on p. 52 to see again how inverting products of elimination matrices works. Think carefully about the two bold sentences on "feels an effect" and "feels no effect." Do you understand exactly what this means? Then read p. 53 and contrast the calculations in equations (10) and (11). Which do you like better? Read all of p. 59—and think about the last paragraph on p. 58: "A proof means that we have not just seen that pattern and believed it and liked it, but understood it." *This is why we prove things.* Another proof of LU appears on p. 60, using the matrix multiplication technique discussed on p. 34.

So who cares? The work in Example 2.11.1 probably felt no more efficient than a routine back-substitution approach (which you did in Problem 2.6.8, right?) Maybe it felt more inefficient!

What if we need to solve $A\mathbf{x} = \mathbf{b}_j$ for many $\mathbf{b}_j$? For a a finite number of $\mathbf{b}_j$, maybe we could work with a large augmented matrix $\begin{bmatrix} A & \mathbf{b}_1 & \cdots & \mathbf{b}_p \end{bmatrix}$, do elimination on A via the matrix E, so $EA = U$, and then study $\begin{bmatrix} U & E\mathbf{b}_1 & \cdots & E\mathbf{b}_p \end{bmatrix}$. Then we would have to solve $U\mathbf{x} = E\mathbf{b}_j$ by back-substitution. However, it is arguably less computationally expensive to solve $LU = \mathbf{b}_j$ by the two-step process above. In particular, it may be the case that solving $A\mathbf{x} = \mathbf{b}_j$ is part of a larger *iterative* process: at the jth step, we get a new $\mathbf{b}_j$, but A stays the same. If we want to keep doing this *indefinitely*, the elimination calculations $E\mathbf{b}_j$ may become expensive. Doing the elimination just once to get $A = LU$, and then solving $LU\mathbf{x} = \mathbf{b}_j$ via the two-step process, may be less expensive.

The LU-factorization works when no row interchanges are needed, i.e., when we can write $EA = U$ with U upper-triangular and E as a product only of elimination matrices, not permutation matrices. Basically, it is possible to "almost" commute permutation and elimination matrices so that we have $PA = LU$ with P a product of permutation matrices, L lower-triangular, and U upper-triangular. Figuring out how to get that P factor out front is a little tricky, and a deeper discussion of this situation is probably more appropriate in a subsequent numerical linear algebra course. But once we know $PA = LU$, to solve $A\mathbf{x} = \mathbf{b}$, first permute $PA\mathbf{x} = P\mathbf{b}$, and then solve $LU\mathbf{x} = P\mathbf{b}$ as we did above.

Content from Strang's *ILA* 6E. See p. 65. This is wholly optional reading and requires a little more knowledge of permutation matrices than I expect or desire right now.

2.11.4 Problem ($\star$). One challenge in teaching this course is writing good problems! Even the most innocent-looking matrix can involve nasty arithmetic when it comes to Gaussian elimination because of *all those fractions*. Good problems involve matrices with integer entries for which Gaussian elimination does not require any division when doing the elimination steps to get to the upper-triangular form. Here is a way to get good problems and generate your own practice.

(i) Write down a square (3×3 or 4×4) matrix L whose diagonal entries are all 1 and whose entries below the diagonal are nonnegative integers between 0 and 9. (At least one of those subdiagonal entries should be nonzero, to keep things interesting.)

(ii) Let U be a square upper-triangular matrix of the same size as L whose entries on and above the diagonal are also nonnegative integers between 0 and 9. The diagonal entries of U do not need to be 1. (Again, at least one of those superdiagonal entries should be nonzero.)

(iii) Multiply LU and call that product A.

(iv) Do Gaussian elimination on A to reduce it to upper-triangular form. How do L and U show up?

2.12. Upper-triangular matrices control existence and uniqueness.

Gauss–Jordan elimination says that if the diagonal entries of an upper-triangular matrix are nonzero, then that matrix is invertible. The arrow of our logic goes the other way, too.

2.12.1 Lemma. *Let $U \in \mathbb{R}^{m \times m}$ be upper-triangular. If U has a zero diagonal entry, then $\mathbf{N}(U) \neq \{\mathbf{0}_m\}$, and so U is not invertible.*

Proof. We consider two possible structures of U: one where the first diagonal entry is zero and one where it is nonzero and a zero diagonal entry occurs further down along the diagonal.

1. *The first diagonal entry is zero.* We do a specific case first to see the strategy.

(i) *The case $m = 4$.* Here U has the form

$$U = \begin{bmatrix} 0 & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix}.$$

Then $U\mathbf{e}_1 = \mathbf{0}_4$, and since $\mathbf{e}_1 \neq \mathbf{0}_4$, so $\mathbf{e}_1 \in \mathbf{N}(U)$, and therefore $\mathbf{N}(U) \neq \{\mathbf{0}_4\}$, thus U is not invertible.

(ii) *The general case.* Here U has the form

$$U = \begin{bmatrix} \mathbf{0}_m & \widetilde{U} \end{bmatrix},$$

where $\widetilde{U}$ is "the rest" of U (columns 2 through m). Again, since $U\mathbf{e}_1$ is the first column of U, we have $U\mathbf{e}_1 = \mathbf{0}_m$ and $\mathbf{e}_1 \neq \mathbf{0}_m$, so $\mathbf{e}_1 \in \mathbf{N}(U)$, and therefore $\mathbf{N}(U) \neq \{\mathbf{0}_m\}$, thus U is not invertible.

2. *An entry on the diagonal in rows 2 or below is zero.* We look at the *first* zero entry on the diagonal (first starting from the top row), and so now this zero entry has to occur in row 2 or below. That is, $u_{jj} = 0$ for some $j \geq 2$ but $u_{ii} \neq 0$ for $1 \leq i \leq j-1$.

(i) *An example in the case $m = 4$.* One such possibility when $m = 4$ is

$$U = \begin{bmatrix} ☺ & * & ☹ & * \\ 0 & ☺ & ☹ & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \end{bmatrix}.$$

Here ☺ denotes nonzero entries, and in the notation above $j = 3$. Now look at the upper-triangular matrix

$$\widehat{U} := \begin{bmatrix} ☺ & * \\ 0 & ☺ \end{bmatrix}.$$

This has nonzero diagonal entries and so we can find x_1, $x_2 \in \mathbb{R}$ such that

$$\widehat{U}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ☹ \\ ☹ \end{bmatrix}.$$

But then

$$x_1\begin{bmatrix} ☺ \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} * \\ ☺ \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} ☹ \\ ☹ \\ 0 \\ 0 \end{bmatrix},$$

and so the third column of U is in the span of the first two. Thus the columns of U are dependent, so $\mathbf{N}(U) \neq \{\mathbf{0}_m\}$. In particular, $(x_1, x_2, -1, 0) \in \mathbf{N}(U)$, and $(x_1, x_2, -1, 0) \neq \mathbf{0}_4$

(ii) *The general case.* As before, assume that there is $j \geq 2$ such that $u_{jj} = 0$ but $u_{ii} \neq 0$ for $1 \leq i \leq j-1$. Write

$$U = \left[\begin{array}{cc|c} \widehat{U} & \widehat{\mathbf{u}}_j & \widetilde{U} \\ 0 & \mathbf{0}_{m-(j-1)} & \end{array}\right].$$

This block expression for U contains a lot of condensed notation: $\widehat{U}$ is a $(j-1)\times(j-1)$ upper-triangular matrix with nonzero diagonal entries, $\widehat{\mathbf{u}}_{j-1} \in \mathbb{R}^{j-1}$, $\widetilde{U}$ contains the remaining columns of U, and, irritatingly, 0 means a matrix whose entries are all 0. If $j = m$, then $\widetilde{U}$ isn't there. Since the diagonal entries of $\widehat{U}$ are nonzero, we can find $\widehat{\mathbf{x}}_{j-1} \in \mathbb{R}^{j-1}$ such that $\widehat{U}\widehat{\mathbf{x}}_{j-1} = \widehat{\mathbf{u}}_{j-1}$. From this, we can show that the jth column of U is in the span of the first $j-1$ columns, so the columns of U are dependent, and therefore $\mathbf{N}(U) \neq \{\mathbf{0}_m\}$. ■

We can combine Gauss–Jordan elimination (Theorem 2.10.4) and this lemma to conclude the following.

2.12.2 Theorem. *An upper-triangular matrix is invertible if and only if all of its diagonal entries are nonzero.*

Proof. ($\Longrightarrow$) Suppose that $U \in \mathbb{R}^{m\times m}$ is invertible and upper-triangular but that it has a nonzero diagonal entry. Since U is invertible, Problem 2.9.13 implies that $\mathbf{N}(U) = \{\mathbf{0}_m\}$. But since U is upper-triangular with a nonzero diagonal entry, Lemma 2.12.1 implies that $\mathbf{N}(U) \neq \{\mathbf{0}_m\}$. This is a contradiction, so U cannot be invertible and upper-triangular and have a nonzero diagonal entry; thus all diagonal entries must be zero.

($\Longleftarrow$) This is Gauss–Jordan elimination (Theorem 2.10.4). ■

2.12.3 Problem (!). By considering the matrices

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

explain why knowing whether or not the diagonal entries of an *arbitrary* matrix are zero does not say anything about the invertibility of that matrix.

Another nasty consequence for an upper-triangular matrix U with a zero diagonal entry is that we cannot always solve $U\mathbf{x} = \mathbf{c}$ for any $\mathbf{c} \in \mathbb{R}^m$. This happened in Example 2.7.5, which is worth rereading right now.

2.12.4 Lemma. *Let $U \in \mathbb{R}^{m \times m}$ be upper-triangular. If U has a zero diagonal entry, then $\mathbf{C}(U) \neq \mathbb{R}^m$, and so U is not invertible.*

Proof. First, why do we need to prove this? We know from Lemma 2.12.1 that since U has a zero diagonal entry, U is not invertible. *But did we ever show that if a square matrix is not invertible, then its column space isn't all of $\mathbb{R}^m$?* **No**. This lemma turns out to be the key step in figuring that out.

As in the proof of Lemma 2.12.1, we consider two different locations for the zero diagonal entry—just "flipped" from that proof.

1. *The (m, m)-entry of U is zero.* That is, the "last" entry on the diagonal is zero.

(i) *The case $m = 4$.* Here U has the form

$$U = \begin{bmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The last row is the problem. If $\mathbf{b} = (b_1, b_2, b_3, b_4) \in \mathbf{C}(U)$, then $b_4 = 0$. For example, $\mathbf{e}_4 \notin \mathbf{C}(U)$, and so $\mathbf{C}(U) \neq \mathbb{R}^4$.

(ii) *The general case.* If the (m, m)-entry of U is zero, then the mth (last) row of U has all zero entries. This is because U is upper-triangular and the other entries in that row are all below the diagonal. Suppose that $\mathbf{b} \in \mathbf{C}(U)$, so $\mathbf{b} = U\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^m$. Then the mth entry of $\mathbf{b}$ is $b_m = 0$ because this entry is the dot product of the mth row of U with $\mathbf{x}$. That dot product is zero since the mth row of U is the zero vector. And so not every $\mathbf{b} \in \mathbb{R}^m$ can be in $\mathbf{C}(U)$; for example, $\mathbf{e}_m \notin \mathbf{C}(U)$.

2. *The (j, j)-entry of U is zero for some $j < m$.* That is, an entry "further up" the diagonal is zero. Additionally, we assume that the (j, j)-entry is the first zero entry on the diagonal from the bottom, so $u_{ii} \neq 0$ for $j + 1 \leq i \leq m$.

(i) *The case $m = 4$.* One possibility here is that U has the form

$$U = \begin{bmatrix} * & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & ☺ & * \\ 0 & 0 & 0 & ☺ \end{bmatrix}.$$

Here ☺ $\neq 0$, and so the $(2, 2)$-entry is the first nonzero entry on the diagonal when we go up from the bottom.

Use those nonzero diagonal entries to do Gauss–Jordan elimination in rows 1 and 2 of columns 3 and 4. Then there is an invertible matrix E such that

$$EU = \begin{bmatrix} * & * & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

So, if $\mathbf{b} \in \mathbf{C}(U)$, then there is $\mathbf{v} \in \mathbb{R}^4$ such that $U\mathbf{v} = \mathbf{b}$, and then $EU\mathbf{v} = E\mathbf{b}$. Consequently, the second entry of $E\mathbf{b}$ is 0. But that certainly is not true for all $\mathbf{b}$: take $\mathbf{b} = E^{-1}\mathbf{e}_2$. That is, $E^{-1}\mathbf{e}_2 \notin \mathbf{C}(U)$. Something like this happened in Example 2.7.5.

(ii) *The general case.* Here U has the form

$$U = \begin{bmatrix} * & * \\ \overrightarrow{\mathbf{0}} & \overrightarrow{\mathbf{u}} \\ 0 & \widehat{U} \end{bmatrix}.$$

Here $*$ is just some matrix, the symbol $\overrightarrow{\mathbf{0}}$ is a *row vector* whose entries are all 0, the symbol $\overrightarrow{\mathbf{u}}$ is a row vector, the symbol 0 is (irritatingly!) a matrix whose entries are all 0, and, critically, $\widehat{U}$ is upper-triangular with nonzero entries on the diagonal. (A good exercise for you is to figure out the dimensions of those blocks. To start, $\widehat{U}$ has $m-(j+1)$ rows and columns.) Use the nonzero entries of $\widehat{U}$ to do elimination in the rows above $\widehat{U}$. Then there is an invertible matrix E such that

$$EU = \begin{bmatrix} * & 0 \\ \overrightarrow{\mathbf{0}} & \overrightarrow{\mathbf{0}} \\ 0 & I \end{bmatrix}.$$

The important thing is that row j of EU is all zero. Then if $\mathbf{b} \in \mathbf{C}(U)$, the jth entry of $E\mathbf{b}$ is 0, and, as before, this cannot happen for every $\mathbf{b} \in \mathbb{R}^m$. ■

2.13. The invertible matrix theorem summarizes everything.

Here is the payoff for all of our work on linear systems and inverses. The following is only valid for square systems, because it talks about inverses and hinges on those technical, demanding results for upper-triangular matrices, but what a payoff for this case!

2.13.1 Theorem (Invertible matrix theorem). *Let $A \in \mathbb{R}^{m \times m}$. The following statements are equivalent in the sense that if any one of them is true, then all of the others are true.*

(i) *A is invertible.*

(ii) *For each $\mathbf{b} \in \mathbb{R}^m$, there is exactly one $\mathbf{x} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{b}$. (The problem $A\mathbf{x} = \mathbf{b}$ always has a unique solution.)*

(iii) $\mathbf{C}(A) = \mathbb{R}^m$. *(A solution to the problem* $A\mathbf{x} = \mathbf{b}$ *always exists.)*

(iv) $\mathbf{N}(A) = \{\mathbf{0}_m\}$. *(If the problem* $A\mathbf{x} = \mathbf{b}$ *has a solution, it's unique.)*

(v) *The columns of A are independent. (The problem* $A\mathbf{x} = \mathbf{b}$ *has no redundant data.)*

Proof. We first collect some facts that we already know.

• If $A \in \mathbb{R}^{m \times m}$, then there is an invertible matrix $E \in \mathbb{R}^{m \times m}$ such that EA is upper-triangular. (This is Gaussian elimination: Theorem 2.8.1.)

• If $U \in \mathbb{R}^{mtimesm}$ is upper-triangular with nonzero diagonal entries, then U is invertible. (This is Gauss–Jordan elimination: Theorem 2.10.4.)

• If $U \in \mathbb{R}^{m \times m}$ is upper-triangular with $\mathbf{N}(U) = \{\mathbf{0}_m\}$ or $\mathbf{C}(U) = \mathbb{R}^m$, then U is invertible. (These conditions ensure that no diagonal entry of U is zero, so U must be invertible.)

Next, we have already some of the proof.

• That part (i) implies (ii) is Theorem 2.9.12.

• That part (i) implies parts (iii), (iv), and (v) is Problem 2.9.13.

If we can prove that either of parts (iii) or (iv) implies part (i), then we will have established that part (ii) implies part (i), since parts (iii) and (iv) together imply part (ii). So, we focus on showing that the last three parts imply the very first. Throughout, we will use Gauss–Jordan elimination to write $EA = U$ for some invertible E and upper-triangular U. Equivalently, $A = E^{-1}U$, and so for A to be invertible, we just need U to be invertible. And since U is upper-triangular, Theorem 2.12.2 says that we just need the diagonal entries of U to be nonzero. That is what we establish.

(iii) $\Longrightarrow$ (i) We claim that $\mathbf{C}(U) = \mathbb{R}^m$, too. If $\mathbf{b} \in \mathbb{R}^m$, then since $\mathbf{C}(A) = \mathbb{R}^m$, there is $\mathbf{v} \in \mathbb{R}^m$ such that $A\mathbf{x} = E^{-1}\mathbf{b}$. Then $U\mathbf{x} = EA\mathbf{x} = \mathbf{b}$. (This might feel like Problem 2.9.14. What is different here?) So, $\mathbf{C}(U) = \mathbb{R}^m$. Lemma 2.12.4 then says that all of the diagonal entries of U are nonzero.

(iv) $\Longrightarrow$ (i) We claim that $\mathbf{N}(U) = \{\mathbf{0}_m\}$, too. If $U\mathbf{x} = \mathbf{0}_m$, then $A\mathbf{x} = E^{-1}U\mathbf{x} = \mathbf{0}_m$, too. So $\mathbf{x} = \mathbf{0}_m$, as desired, and $\mathbf{N}(U) = \{\mathbf{0}_m\}$. Lemma 2.12.1 then says that all of the diagonal entries of U are nonzero.

(v) $\Longrightarrow$ (i) We have known for a long time that independent columns imply that the null space is as small as possible. Specifically, from Corollary 1.11.1, if the columns of A are independent, then $\mathbf{N}(A) = \{\mathbf{0}_m\}$. (This, by the way, is true even when A is not square.) So part (iv) is true, which implies the invertibility of A. (But when A is not square, it does not make sense to talk about an inverse.) ■

2.13.2 Problem (!). Reread the proof of the invertible matrix theorem and convince yourself that any one part does imply the other four. For example, if you assume that (iii) is true, why is part (v) true?

The invertible matrix theorem proves our longstanding Conjectures 1.9.8 and 1.11.6. We still do not have the tools to prove Conjecture 1.14.11.

It should be surprising that, *for square systems*, existence and uniqueness by themselves, separately, are enough to imply existence and uniqueness together! This is a special property of *square* systems that nonsquare problems need not share, as we will see. We will give examples of *nonsquare* problems for which existence is always true but uniqueness fails, and for which existence sometimes fails but uniqueness is always true.

2.13.3 Problem (!). Uniqueness can never fail "only sometimes." Let $A \in \mathbb{R}^{m \times n}$ and let $\mathbf{b}_1$, $\mathbf{b}_2 \in \mathbb{R}^m$. Suppose that there are two different $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathbb{R}^n$ such that $A\mathbf{x}_1 = \mathbf{b}_1$ and $A\mathbf{x}_2 = \mathbf{b}_1$. Explain why if the problem $A\mathbf{x} = \mathbf{b}_2$ has a solution, it really has infinitely many solutions. [Hint: $\mathbf{x}_1 - \mathbf{x}_2 \in \mathbf{N}(A)$. *Adapt the proof of Theorem 2.5.2.*]

2.13.4 Problem ($\star$). Recall that a permutation matrix $\mathbb{R}^{m \times m}$ is a matrix in which each standard basis vector for $\mathbb{R}^m$ appears once, and only once, as a column. Use the invertible matrix theorem to prove that any permutation matrix is invertible. (Previously we talked about the invertibility of the special permutation matrices P_{ij}, for which $P_{ij}^{-1} = P_{ij}$. Subsequently we will develop a slick formula for the inverse of an arbitrary permutation matrix.)

2.13.5 Problem ($\star$). Let $A \in \mathbb{R}^{m \times m}$. Recall that $E \in \mathbb{R}^{m \times m}$ is a **LEFT INVERSE** of A if $EA = I_m$ and $F \in \mathbb{R}^{m \times m}$ is a **RIGHT INVERSE** of A if $AF = I_m$. (Important: we are not assuming that E or F is invertible here. Also, reread Corollary 2.9.8 right now.)

(i) Prove that if A has a left inverse, then A is invertible. [Hint: *if* $A\mathbf{x} = \mathbf{0}_m$, *what is* $EA\mathbf{x}$*?*]

(ii) Prove that if A has a right inverse, then A is invertible. [Hint: *if* $\mathbf{b} \in \mathbb{R}^m$, *what is* $AF\mathbf{b}$*?*]

2.13.6 Remark. *The nonzero diagonal entries of an upper-triangular matrix are sometimes called its* ***PIVOTS****. The pivots of a general* $A \in \mathbb{R}^{m \times m}$ *are the nonzero diagonal entries of the upper-triangular matrix to which* A *can always be transformed by elimination and row interchanges, i.e., by Theorem 2.8.1. This language is a little perilous, as we never proved that the matrix* U *from Theorem 2.8.1 was unique—could we write* $E_1A = U_1$ *and* $E_2A = U_2$ *with* U_1 *and* U_2 *both upper-triangular,* $U_1 \neq U_2$*, and* E_1 *and* E_2 *as the product of elimination and/or permutation matrices? What is important from the point of view of invertibility is not the exact value of these "pivots" but rather whether they are all nonzero or not.*

2.13.7 Problem (+). Let $A \in \mathbb{R}^{m\times m}$ and suppose that E_1, $E_2 \in \mathbb{R}^{m\times m}$ are invertible with E_1A and E_2A both upper-triangular. Prove that if E_1A has no nonzero diagonal entries, then E_2A also has no nonzero diagonal entries. [Hint: *show that A is invertible.*] We will eventually prove that E_1A and E_2A must have the same number of nonzero diagonal entries, although we need more technology for that.

Content from Strang's *ILA* 6E. Page 41 introduces the terminology "pivot." I personally feel that the phrase "nonzero pivot" is redundant. Informally, you should think of the pivots as "the nonzero things that you multiply by when doing elimination." Because we can permute rows even when we don't need to avoid zero diagonal entries, we can select an "ideal" pivot at any state of elimination—see "'Partial Pivoting' to Reduce Roundoff Errors" on p. 66 and think once more about taking a numerical linear algebra class after this one.

2.13.8 Problem ($\star$). Give an example of an upper-triangular matrix U such that for some j, column j is a pivot column, but the (j, j)-entry of U is 0. The point is that a pivot column may not contain a diagonal pivot! Language is slippery.

3. The Subspace Perspective

3.1. Introduction: what goes wrong with nonsquare systems?

Our best successes in this course arguably come from square systems: $A\mathbf{x} = \mathbf{b}$ with $A \in \mathbb{R}^{m\times m}$ and $\mathbf{b} \in \mathbb{R}^m$, same number of equations as unknowns. We will see that it is with square systems alone that we have a chance (not a guarantee) for both existence and uniqueness of solutions—it is possible both to be able to solve the problem and have only one solution for it. For nonsquare systems, we will show that either existence or uniqueness always fails (maybe both). Understanding how to quantify and qualify our failures, and how to move on from them, will be the central part of our forthcoming story. We can see this happen with relatively small systems using relatively few numbers.

3.1.1 Example. We consider the problem $A\mathbf{x} = \mathbf{b}$ for the variety of A below.

(i) For

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

the problem $A\mathbf{x} = \mathbf{b}$ always has the unique solution $\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^2$.

(ii) For

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

existence fails for $A\mathbf{x} = \mathbf{b}$ when $\mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$ with $b_2 \neq 0$. Uniqueness also fails because $A\mathbf{e}_2 = \mathbf{0}_2$.

(iii) For

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

existence fails for $A\mathbf{x} = \mathbf{b}$ when $\mathbf{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$ with $b_3 \neq 0$. However, $\mathsf{N}(A) = \{\mathbf{0}_2\}$, so if a solution to $A\mathbf{x} = \mathbf{b}$ exists, then it is unique.

(iv) For

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

both existence and uniqueness fail. We cannot solve $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$ with $b_2 \neq 0$ and $b_3 \neq 0$, and $A\mathbf{0}_2 = \mathbf{0}_3$, so uniqueness fails.

(v) For

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

we have existence but not uniqueness: take $\mathbf{x} = (b_1, b_2, 0)$ to solve $A\mathbf{x} = (b_1, b_2)$. However, uniqueness fails because $A\mathbf{0}_3 = \mathbf{0}_2$.

(vi) For

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

both existence and uniqueness fail. We cannot solve $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$ with $b_2 \neq 0$, and $A\mathbf{e}_3 = \mathbf{0}_2$.

Here is what the previous example suggests as we move beyond square systems.

3.1.2 Conjecture. *Let $A \in \mathbb{R}^{m \times n}$.*

(i) *If $m > n$ (more equations than unknowns, more rows than columns, A is taller than it is wide), then we will always fail to solve $A\mathbf{x} = \mathbf{b}$ for some $\mathbf{b} \in \mathbb{R}^m$. That is, $\mathbf{C}(A) \neq \mathbb{R}^m$. It may or may not be possible to get unique solutions.*

(ii) *If $m < n$ (more unknowns than equations, more columns than rows, A is wider than it is tall), then we will never be able to solve $A\mathbf{x} = \mathbf{b}$ uniquely. Solutions may or may not exist in the first place.*

Additionally, unlike with the invertible matrix theorem for square matrices, neither existence nor uniqueness by itself is enough to guarantee the other condition.

Content from Strang's *ILA* 6E. Now is a good time to reread p. 38.

Our goal is to prove Conjecture 3.1.2 and do a bit more: to qualify and quantify our inherent failure to obtain existence and uniqueness for nonsquare problems. Previously, we focused on existence first (the column space) and then uniqueness (the null space), the idea being that it is more natural to have *some* solutions to work with first and then to ask if they are the *only* solutions. Now that we are a little more experienced, we are going to do the reverse. First we will study the specific problem $A\mathbf{x} = \mathbf{0}_m$ for $A \in \mathbb{R}^{m \times n}$ (usually with $m \neq n$), and from that we will learn techniques for approaching the more general problem $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} \in \mathbb{R}^m$ arbitrary.

3.2. The CR-factorization reveals the null space.

The CR-factorization makes null space calculations very easy. We first treat two "easy" extreme cases and then focus on the interesting intermediate case. Let $A \in \mathbb{R}^{m \times n}$.

If $\text{rank}(A) = 0$, then A is the zero matrix, and $A\mathbf{x} = \mathbf{0}_m$ for all $\mathbf{x} \in \mathbb{R}^n$, thus $\mathbf{N}(A) = \mathbb{R}^n$. This is straightforward and boring.

If $\text{rank}(A) = n$ and all of the columns of A are pivot columns, then the columns of A are independent, so $\mathbf{N}(A) = \{\mathbf{0}_n\}$. This is nice (solutions to $A\mathbf{x} = \mathbf{b}$, if they exist, are unique), but also boring.

So, from now on we assume $1 \le \operatorname{rank}(A) < n$. Then $\mathbf{N}(A)$ is both nontrivial and not everything, and therefore interesting.

Put $r := \operatorname{rank}(A) \ge 1$, and write A in its CR-factorization: $A = CR$, where the columns of $C \in \mathbb{R}^{m\times r}$ are independent, and $R \in \mathbb{R}^{r\times n}$. Suppose that $A\mathbf{x} = \mathbf{0}_m$. Then $CR\mathbf{x} = \mathbf{0}_m$. Insert parentheses: $C(R\mathbf{x}) = \mathbf{0}_m$. Since the columns of C are independent, if $C\mathbf{v} = \mathbf{0}_m$, then $\mathbf{v} = \mathbf{0}_r$. Thus $R\mathbf{x} = \mathbf{0}_r$, and so $\mathbf{x} \in \mathbf{N}(R)$. Conversely, suppose that $R\mathbf{x} = \mathbf{0}_r$. Then

$$A\mathbf{x} = (CR)\mathbf{x} = C(R\mathbf{x}) = C\mathbf{0}_r = \mathbf{0}_m.$$

We have therefore proved that $\mathbf{N}(A) = \mathbf{N}(R)$. This is good, because the structure of R is probably simpler than the structure of A. In particular, at least r columns of R are the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r \in \mathbb{R}^r$, so R has many zero entries, which is always helpful. Now we will look further at the structure of $\mathbf{N}(R)$.

3.2.1 Example. Let

$$A = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 0 & 4 \end{bmatrix}.$$

From the point of view of the CR-factorization, we have $C = I_2$ and $R = A$.

We study $\mathbf{N}(A)$. For $\mathbf{x} \in \mathbb{R}^4$, we have $A\mathbf{x} = \mathbf{0}_2$ if and only if

$$\begin{cases} x_1 \phantom{{}+x_2} + 2x_3 + 3x_4 = 0 \\ \phantom{x_1+{}} x_2 \phantom{{}+2x_3} + 4x_4 = 0. \end{cases}$$

This is not as nice as the square upper-triangular systems that we have previously studied. Every equation here has at least two variables in it.

The right, if not immediately obvious, strategy is to solve for what we can easily solve for. The unknowns x_1 and x_2 have coefficients of 1 on them, so solving for those two variables in terms of x_3 and x_4 is easier, relatively speaking, than solving for x_3 or x_4. We get

$$\begin{cases} x_1 = -2x_3 - 3x_4 \\ x_2 = -4x_4, \end{cases}$$

and if we put $\mathbf{x} = (x_1, x_2, x_3, x_4)$, then

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_3 - 3x_4 \\ -4x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_3 \\ 0 \\ x_3 \\ 0 \end{bmatrix} + \begin{bmatrix} -3x_4 \\ -4x_4 \\ 0 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ -4 \\ 0 \\ 1 \end{bmatrix}.$$

We have shown that every $\mathbf{x} \in \mathbf{N}(A)$ is a linear combination of those two vectors on the right. More compactly,

$$\mathbf{N}\left(\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 0 & 4 \end{bmatrix}\right) = \mathbf{C}\left(\begin{bmatrix} -2 & -3 \\ 0 & -4 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

(Strictly speaking, we have shown that *if* $\mathbf{x} \in \mathbf{N}(A)$, *then* $\mathbf{x}$ is in the column space of that 4×2 matrix. You should check your work and show that each column in that 4×2 matrix is in $\mathbf{N}(A)$.)

It should look like there are two degrees of freedom in describing the null space, which come from those two variables x_3 and x_4 whose values we did not (could not) specify. We might call those "free" variables, because we were "free" to choose them, and once we did, the values of x_1 and x_2 were specified. It cannot be an accident than our matrix has rank 2, as well.

Content from Strang's *ILA* 6E. This example is basically the same as Example 1 on p. 93. Strang calls the columns of

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

the "special solutions" for $A\mathbf{x} = \mathbf{0}_2$. What is "special" about these solutions is that they are linearly independent, and every solution to $A\mathbf{x} = \mathbf{0}_2$ is in the span of these solutions.

Here is the pattern from Example 3.2.1. The matrix A had the block structure

$$A = \begin{bmatrix} I_2 & F \end{bmatrix}, \qquad F := \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix},$$

and its null space is

$$\mathbf{N}(A) = \mathbf{C}\left(\begin{bmatrix} -F \\ I_2 \end{bmatrix} \right).$$

Here is how this generalizes. (Before proceeding, doing Problems 2.4.5 and 2.4.6 on block matrices would be a very, very good idea.) Let $A \in \mathbb{R}^{m \times n}$ with $r := \text{rank}(A)$ and suppose $1 \le r < n$. Suppose further that the CR-factorization of A is $A = CR$ with $R = \begin{bmatrix} I_r & F \end{bmatrix}$. This is a special case without the permutation matrix that usually appears in R; we will treat that shortly after another example. We know that $\mathbf{N}(A) = \mathbf{N}(R)$.

If $R\mathbf{x} = \mathbf{0}_r$, split up $\mathbf{x}$ into the block vector

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(r)} \\ \mathbf{x}_{(n-r)} \end{bmatrix}.$$

For example, if $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ and $r = 2$, then $\mathbf{x}^{(2)} = (x_1, x_2)$ and $\mathbf{x}_{(3)} = (x_3, x_4, x_5)$. In the general case, we have

$$\mathbf{0}_r = R\mathbf{x} = \begin{bmatrix} I_r & F \end{bmatrix} \begin{bmatrix} \mathbf{x}^{(r)} \\ \mathbf{x}_{(n-r)} \end{bmatrix} = I_r\mathbf{x}^{(r)} + F\mathbf{x}_{(n-r)} = \mathbf{x}^{(r)} + F\mathbf{x}_{(n-r)},$$

and so

$$\mathbf{x}^{(r)} = -F\mathbf{x}_{(n-r)}.$$

Now put it back together:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(r)} \\ \mathbf{x}_{(n-r)} \end{bmatrix} = \begin{bmatrix} -F\mathbf{x}_{(n-r)} \\ \mathbf{x}_{(n-r)} \end{bmatrix} = \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{x}_{(n-r)} \in \mathbf{C}\left(\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right).$$

3.2.2 Problem ($\star$). Let $1 \le r < n$ and $F \in \mathbb{R}^{r\times(n-r)}$. Suppose that

$$\mathbf{y} \in \mathbf{C}\left(\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right).$$

Prove that $\mathbf{y} \in \mathbf{N}(\begin{bmatrix} I_r & F \end{bmatrix})$. [Hint: *split* $\mathbf{y}$ *up as*

$$\mathbf{y} = \begin{bmatrix} -F\mathbf{v} \\ \mathbf{v} \end{bmatrix}$$

for some $\mathbf{v} \in \mathbb{R}^{n-r}$ *and note that* $F\mathbf{v} \in \mathbb{R}^r$.]

We conclude (using Definition 1.8.4 of set equality) that

$$\mathbf{N}(\begin{bmatrix} I_r & F \end{bmatrix}) = \mathbf{C}\left(\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right). \tag{3.2.1}$$

Thus in the special case that $A \in \mathbb{R}^{m\times n}$ has the CR-factorization $A = C\begin{bmatrix} I_r & F \end{bmatrix}$, we can write $\mathbf{N}(A)$ as the column space above.

More generally, however, the R-factor in a CR-factorization need not have the identity block first.

3.2.3 Example. Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}.$$

Again $\operatorname{rank}(A) = 2$ and the CR-factorization of A uses $C = I_2$ and $R = A$. We proceed as in Example 3.2.1: assume $A\mathbf{x} = \mathbf{0}_2$ and write this as the linear system

$$\begin{cases} x_2 + 2x_3 + 3x_5 = 0 \\ x_4 + 4x_5 = 0. \end{cases}$$

We solve for the variables with the simplest coefficients of 1; these are now x_2 and x_4:

$$\begin{cases} x_2 = -2x_3 - 3x_5 \\ x_4 = -4x_5. \end{cases}$$

Vectorizing, we have

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_1 \\ -2x_3 - 3x_5 \\ x_3 \\ -4x_5 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -2x_2 \\ x_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -3x_5 \\ 0 \\ -4x_5 \\ x_5 \end{bmatrix} = x_1\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3\begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5\begin{bmatrix} 0 \\ -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

Thus

$$\mathbf{N}\left(\begin{bmatrix}0 & 1 & 2 & 0 & 3\\ 0 & 0 & 0 & 1 & 4\end{bmatrix}\right) = \mathbf{C}\left(\begin{bmatrix}1 & 0 & 0\\ 0 & -2 & -3\\ 0 & 1 & 0\\ 0 & 0 & -4\\ 0 & 0 & 1\end{bmatrix}\right). \tag{3.2.2}$$

Hopefully it looks like the columns of the 2×2 identity matrix I_2 are jumbled up in the columns of A and the rows of the 3×3 identity matrix I_3 are jumbled up in the column space that now controls the null space.

Content from Strang's *ILA* 6E. This was basically Example 2 on p. 94, except the matrix in that example has a row of all zero entries, which cannot happen in the R-factor from the CR-factorization (Problem 2.3.7).

Recall that, in general, if $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$, and if $1 \le r < n$, then the CR-factorization of A has R in the form

$$R = \begin{bmatrix} I_r & F \end{bmatrix} P,$$

where $F \in \mathbb{R}^{r \times (n-r)}$ and $P \in \mathbb{R}^{n \times n}$ is a permutation matrix. The point of P is that multiplying on the right by P jumbles up the columns of the $r \times r$ identity matrix I_r among the "junky" columns F. We still know that $\mathbf{N}(A) = \mathbf{N}(R) = \mathbf{N}(\begin{bmatrix} I_r & F \end{bmatrix} P)$. What is this last null space?

The good news is that things are not much more complicated than the case of $P = I_n$ from before. What matters right now is much less than P is a *permutation* matrix than that P is *invertible* by Problem 2.13.4. So, suppose that $R\mathbf{x} = \mathbf{0}_m$. Then $\begin{bmatrix} I_r & F \end{bmatrix} P\mathbf{x} = \mathbf{0}_m$. The trick is to abbreviate $\mathbf{y} = P\mathbf{x}$. We then have

$$\begin{bmatrix} I_r & F \end{bmatrix} P\mathbf{x} = \mathbf{0}_r \iff \begin{cases} \begin{bmatrix} I_r & F \end{bmatrix} \mathbf{y} = \mathbf{0}_r \\ P\mathbf{x} = \mathbf{y}. \end{cases} \tag{3.2.3}$$

If $\begin{bmatrix} I_r & F \end{bmatrix} \mathbf{y} = \mathbf{0}_r$, then $\mathbf{y} \in \mathbf{N}(\begin{bmatrix} I_r & F \end{bmatrix})$, and so from (3.2.1) we know that

$$\mathbf{y} = \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{v}$$

for some $\mathbf{v} \in \mathbb{R}^{n-r}$. Then

$$P\mathbf{x} = \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{v},$$

and so

$$\mathbf{x} = P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{v}.$$

We have shown that if $\mathbf{x} \in \mathbf{N}(\begin{bmatrix} I_r & F \end{bmatrix} P)$, then

$$\mathbf{x} \in \mathbf{C}\left(P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right).$$

The same idea as in Problem 3.2.2 then gives

$$\mathbf{N}(\begin{bmatrix} I_r & F \end{bmatrix} P) = \mathbf{C}\left(P^{-1}\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right). \tag{3.2.4}$$

3.2.4 Problem (!). Use (3.2.4) and some related ideas in (2.4.2) to find a permutation matrix $P \in \mathbb{R}^{5\times 5}$ and a matrix $F \in \mathbb{R}^{2\times 3}$ such that

$$\mathbf{N}\left(\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}\right) = \mathbf{C}\left(P^{-1}\begin{bmatrix} -F \\ I_3 \end{bmatrix}\right).$$

Compare this to (3.2.2).

Content from Strang's *ILA* 6E. The expressions (3.2.1) and (3.2.4) for the null space appear in the box on p. 97 and the subsequent "Review" paragraph. Interpret P^{T} there as P^{-1} for now.

Here is a summary of everything that we have learned about the form of the null space.

3.2.5 Theorem. *Let $A \in \mathbb{R}^{m\times n}$ with $r := \operatorname{rank}(A)$.*

(i) *If $r = 0$, then $\mathbf{N}(A) = \mathbb{R}^n$. In particular, $\mathbf{N}(A) = \mathbf{C}(I_n)$.*

(ii) *If $r = n$, then $\mathbf{N}(A) = \{\mathbf{0}_n\}$. In particular, $\mathbf{N}(A) = \mathbf{C}(Z)$, where $Z \in \mathbb{R}^{m\times n}$ is the zero matrix.*

(iii) *If $1 \le r < n$ and A has the CR-factorization $A = CR$ with $R = \begin{bmatrix} I_r & F \end{bmatrix} P$ for some $F \in \mathbb{R}^{r\times(n-r)}$ and some permutation matrix $P \in \mathbb{R}^{n\times n}$, then*

$$\mathbf{N}(A) = \mathbf{N}(R) = \mathbf{C}\left(P^{-1}\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right).$$

3.2.6 Problem (!). Adding more rows of all zero entries to a matrix does not change its null space.

(i) Convince yourself that this is true by finding the null spaces of

$$\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and comparing your results to Examples 3.2.1 and 3.2.3.

(ii) Now let $A \in \mathbb{R}^{m\times n}$ and denote by 0 the zero matrix in $\mathbb{R}^{p\times n}$ for any $p \ge 1$. Prove that

$$\mathbf{N}\left(\begin{bmatrix} A \\ 0 \end{bmatrix}\right) = \mathbf{N}(A).$$

By now we should be convinced that the CR-factorization is not only a natural and meaningful way to represent important data about a matrix. It is a useful tool for solving $A\mathbf{x} = \mathbf{0}$. However, we still do not have an efficient or “easy” method for computing this factorization other than a column-by-column examination of the matrix to identify its pivot columns. The time has come to develop such a method.

3.3. CR and RREF are friends.

We have given an “existential” construction of the CR-factorization of a matrix that proceeded column by column. Now we develop the CR-factorization in a row-by-row approach. We return to the great algorithm of the course—Gaussian and Gauss–Jordan elimination—and extend it to nonsquare matrices. Previously, we used Gaussian elimination (“downwards elimination”) to reduce a square matrix to an upper-triangular matrix, and then we used Gauss–Jordan elimination (“upwards elimination”) to convert an upper-triangular matrix with nonzero diagonal entries to the identity matrix. Here we apply elimination to arbitrary matrices; we will first obtain the CR-factorization more transparently, and later we will extract the complete solution formula for a linear system. Gloriously, there is very little new in the actual arithmetic.

3.3.1 Example. Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$

This matrix A appeared in Example 1.14.7. Previously we only performed elementary row operations on square matrices, but they certainly work on nonsquare matrices, too. We compute

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix} \xrightarrow[E_{21}]{\text{R2} \mapsto \text{R2}-2\times\text{R1}} \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}, \qquad E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[D_{33}]{\text{R3} \mapsto (1/2)\times\text{R3}} \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}, \qquad D_{33} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$$

$$\xrightarrow[E_{13}]{\text{R1} \mapsto \text{R1}-\text{R3}} \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}, \qquad E_{12} := \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[P_{23}]{\text{R2} \mapsto \text{R3},\ \text{R3} \mapsto \text{R2}} \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad P_{23} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

That is,

$$EA = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad E := P_{23}E_{12}D_{33}E_{21}.$$

This matrix EA appeared in Examples 1.14.1 and 2.1.1.

Abbreviate $R_0 := EA$. This notation is meant to emphasize the block structure

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} R \\ 0 \end{bmatrix}, \qquad R := \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix},$$

where 0 denotes the row of all zero entries. Here R is the R-factor in the CR-factorization of both R_0 (Example 2.1.1) and A (part (ii) of Example 2.3.1). That is, by doing Gaussian and Gauss–Jordan elimination to A, we found an invertible matrix E such that

$$EA = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

reveals the R-factor in the CR-factorization of A.

The matrix $EA = R_0$ in the preceding example has four defining characteristics.

3.3.2 Definition. *A matrix is in* ***REDUCED ROW ECHELON FORM (RREF)*** *if it has the following four properties.*

Row Property 1. *Any row whose entries are all zero is below any row with some nonzero entries.*

Row Property 2. *If a row contains nonzero entries, the first nonzero entry of that row is* 1*, called the* ***LEADING*** 1 *or the* ***PIVOT*** *for that row.*

Column Property 1. *The other entries of any column containing a leading* 1 *are* 0*. That is, a column containing a leading* 1 *is a column of the* $m \times m$ *identity matrix* I_m*, equivalently, a standard basis vector for* $\mathbb{R}^m$*.*

Column Property 2. *If* $1 \le i < j \le m$ *and rows* i *and* j *both contain nonzero entries, then the leading* 1 *of row* i *appears in a column before the column containing the leading* 1 *of row* j*. That is, the leading* 1 *of a given row is "to the left" of the leading* 1*'s in the rows below. (Here is a more precise, and possibly more annoying, way of saying this. Suppose that* $1 \le i_1 < i_2 \le m$ *and rows* i_1 *and* i_2 *both contain nonzero entries. Let the leading* 1 *of row* i_1 *be an entry in column* j_{i_1} *and the leading* 1 *of row* i_2 *be an entry in column* j_{i_2}*. Then* $j_{i_1} < j_{i_2}$*.)*

3.3.3 Example. We take another look at the matrix $EA = R_0$ from Example 3.3.1:

$$EA = R_0 = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in RREF.

Row Property 1. There is only one row of all 0 entries, and at the bottom, below the rows with nonzero entries.

Row Property 2. The rows with nonzero entries are rows 1 and 2, and their first nonzero entries are 1.

Column Property 1. Columns 1 and 3 contain these leading 1's, and the other entries of columns 1 and 3 are all 0. (That is, columns 1 and 3 are columns of I_3.)

Column Property 2. Last, row 2 contains nonzero entries, and the leading 1 of row 2 is in column 3, which is after the leading 1 in row 1 (which is in column 1).

3.3.4 Problem (!). Explain *all* of the reasons why

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

is not in RREF.

3.3.5 Example. For practice with the axioms of the RREF from Definition 3.3.2, we construct all matrices in $\mathbb{R}^{3\times 4}$ that are in RREF and that have leading 1's in columns 2 and 4 only. We proceed via the following steps.

1. Start with the first column (a very good place to start). If any entry is nonzero, that entry is the leading nonzero entry in its row (nothing comes before the first column), and so column 1 has a leading 1. This is not allowed under the rules of our current game, so the first column is $\mathbf{0}_3$, and therefore

$$R = \begin{bmatrix} 0 & ? & ? & ? \\ 0 & ? & ? & ? \\ 0 & ? & ? & ? \end{bmatrix}.$$

2. The second column has a leading 1, so it is either $\mathbf{e}_1$, $\mathbf{e}_2$, or $\mathbf{e}_3$. If column 2 is $\mathbf{e}_2$, then

$$R = \begin{bmatrix} 0 & 0 & ? & ? \\ 0 & 1 & ? & ? \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The other two entries in row 1 (in columns 3 and 4) can't both be 0, as that would violate Row Property 1. So, at least one of them is nonzero, thus a leading nonzero entry. But then $\mathbf{e}_1$ appears in column 3 or 4, contradicting Column Property 2.

3. We now know

$$R = \begin{bmatrix} 0 & 1 & ? & ? \\ 0 & 0 & ? & ? \\ 0 & 0 & ? & ? \end{bmatrix}.$$

Look at the third column. If it has a nonzero entry in rows 2 or 3, that is the leading nonzero entry in that row, and so column 3 has a leading 1. This is not allowed in our game. However, there does not appear to be any restrictions on the $(1,3)$-entry, since that would not be a leading nonzero entry in row 1. Write

$$R = \begin{bmatrix} 0 & 1 & * & ? \\ 0 & 0 & 0 & ? \\ 0 & 0 & 0 & ? \end{bmatrix}.$$

We have upgraded the $(1,3)$-entry from ? to $*$ to emphasize that it can be any number right now, zero or not.

4. The fourth column is a pivot column, so it is $\mathbf{e}_1$, $\mathbf{e}_2$, or $\mathbf{e}_3$. If it's $\mathbf{e}_1$, then

$$\begin{bmatrix} 0 & 1 & * & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

but then the 1 in the $(1,4)$-entry is not the leading nonzero entry in row 1, so column 4 doesn't have a leading 1. If column 4 is $\mathbf{e}_3$, then

$$R = \begin{bmatrix} 0 & 1 & * & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and that contradicts Row Property 1. The only choice left is that column 4 is $\mathbf{e}_2$.

We conclude that all matrices in $\mathbb{R}^{3\times 4}$ that are in RREF with leading 1's in columns 2 and 4 only have the form

$$R = \begin{bmatrix} 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where the $(1,3)$-entry is arbitrary. This is a pretty restricted family of matrices.

3.3.6 Problem (!). Write down all possible 2×2 matrices that are in RREF. How do you know that you have found them all? (Some of the entries of your matrices will have to be more or less arbitrary real numbers, but be sure to specify if certain values are excluded.)

Here is the first fruit of Gauss–Jordan elimination: we can reduce any matrix to RREF.

3.3.7 Theorem. *Let $A \in \mathbb{R}^{m \times n}$. There exist an invertible matrix $E \in \mathbb{R}^{m \times m}$ and a matrix $R_0 \in \mathbb{R}^{m \times n}$ such that $EA = R_0$ and R_0 is in RREF.*

Proof. This is Gauss–Jordan elimination. Start with the first nonzero column *from the left* (at least one column is nonzero since A is nonzero). If needed, multiply by a permutation matrix so the first entry in this column is nonzero. If there are more zero entries in this column, multiply by elimination matrices to make the entries in rows 2 and below all zero. Then go to the first row below row 1 that has a nonzero entry; call that row i (so $i \geq 2$). Go to the column that has the first nonzero entry in that row; call that column j (so $j \geq 2$). Repeat the permutation (if needed) and elimination operations to make the entries in rows $i+1$ and below of column j all zero. Continue until you have reached the last row of the matrix or until all rows below have only zero entries.

Now start with the first nonzero row from the bottom; call this row i. Multiply by a scaling matrix so that the first nonzero entry in this row (starting *from the left*) is one. Say that this first nonzero entry appears in column j. If $i \geq 2$, multiply by elimination matrices to make the entries in rows $i-1$ to 1 of column j all zero. Then go to the first nonzero row above row i (if there is any such row) and the first nonzero element of that row, which will be in column $j-1$ or before. Go to the column that has the first nonzero entry in that row and multiply by a scaling matrix so that the first nonzero entry is one. Use elimination to create zeros in the entries of that column that fall in the rows above. Continue until you have reached the first row of the matrix or until all rows above have only zero entries.

If the $(1,1)$-entry of the matrix is nonzero, multiply by a scaling matrix to make it one. Finally, multiply the matrix by a permutation matrix so that all rows whose entries are only zero are below all rows with nonzero entries. ∎

Content from Strang's *ILA* 6E. A reduction to RREF is given at the top of p. 95 and another is done in Example 2 at the bottom of the page. A third is Example 3 on pp. 97–98, and this also includes a null space calculation and remarks on the CR-factorization (which we will revisit shortly). Page 96 gives the algorithm for computing the RREF column by column. Read p. 142 up to but not including the "Factorization" box.

3.3.8 Problem ($\star$). Find a matrix $A \in \mathbb{R}^{3 \times 4}$ whose entries are all nonzero such that

$$\mathsf{rref}(A) = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Provide a matrix $E \in \mathbb{R}^{3 \times 3}$ such that $EA = \mathsf{rref}(A)$; you may express E as a product of elementary matrices, and you do not have to multiply that product out.

Theorem 3.3.7 does not say that we can reduce a matrix to only *one* kind of RREF. It would be terribly awkward if Gauss–Jordan elimination could transform a matrix into two different kinds of RREF. We do have some freedom in how we carry out the elimination, so perhaps going in different orders could yield different RREFs. Of course, that cannot happen, but it needs to be checked. Our approach is founded on the belief that the RREF and the CR-factorization are best understood when they are put in dialogue with each other.

Matrices in RREF should look familiar because we have been seeing them since we met the CR-factorization.

3.3.9 Lemma. *Let $A \in \mathbb{R}^{m \times n}$ with $r := \operatorname{rank}(A) \geq 1$, and let $A = CR$ be the CR-factorization of A. Then R is in RREF.*

Proof. Let $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$ be the pivot columns of A. By Theorem 2.3.5, we know that $\mathbf{r}_{j_i} = \mathbf{e}_i \in \mathbb{R}^r$ for $i = 1, \ldots, r$.

1. No row of R has all zero entries by Problem 2.3.7, so R trivially satisfies Row Property 1.

2. For the other properties, we treat the first row as a special case.

Case 1: $j_1 = 1$. Then $\mathbf{r}_1 = \mathbf{e}_1 \in \mathbb{R}^r$, so the $(1,1)$-entry of R is 1. This proves Row Property 2 for row 1 and Column Property 1 for the leading 1 in row 1 in the case that $j_1 = 1$.

Case 2: $j_1 \geq 2$. Then the first $j_1 - 1$ columns of A are $\mathbf{0}_m$, so $C\mathbf{r}_j = \mathbf{0}_m$ for $j = 1, \ldots, j_1 - 1$. By independence, $\mathbf{r}_j = \mathbf{0}_r$ for $j = 1, \ldots, j_1 - 1$. The entries of row 1 in columns 1 through $j_1 - 1$ are therefore 0, but row 1 now has the entry of 1 in column j_1 since $\mathbf{r}_{j_1} = \mathbf{e}_1$. This proves Row Property 2 for row 1 and Column Property 1 for the first leading 1 in the case that $j_1 \geq 2$.

Finally, this also proves Column Property 2 for row 1 (i.e., for $i_1 = 1$ in Column Property 2), since the first nonzero column of R is $\mathbf{e}_1$ in column $j_1 \geq 1$; thus all of the entries in rows 2 and below are 0 for columns 1 through j_1. That is, the first nonzero entry in rows 2 and below must appear in columns $j_1 + 1$ or later.

3. Now let $i \geq 2$. Since $\mathbf{r}_{j_i} = \mathbf{e}_i$, we know that row i has a nonzero entry in column j_i. If we can show that the entries in row i of columns 1 through $j_i - 1$ are zero, then the first nonzero entry in row i will be the 1 from $\mathbf{e}_i$. Additionally, the other entries of the column in which that 1 is located (column j_i) will be zero, since that column is $\mathbf{e}_i$.

So, suppose that the first nonzero entry in row i occurs in some column j with $j < j_i$. We know that $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$ and also $\mathbf{a}_j = C\mathbf{r}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$. By independence, rows i through r of $\mathbf{r}_j$ must be 0.

We have therefore established that the first nonzero entry in row i is 1, that this entry occurs in column j_i, and that column j_i is $\mathbf{e}_i$. This proves Row Property 2 and Column Property 1.

4. Last, if $1 \leq i_1 < i_2 \leq m$, we now know that the leading 1 in row i_1 appears in column j_{i_1}, and the leading 1 in column i_2 appears in column j_{i_2}. By construction, $j_{i_1} < j_{i_2}$ since $i_1 < i_2$. This proves Column Property 2. ■

Not every matrix in RREF shows up as the R-factor in a CR-factorization. A matrix in RREF can have a row of all zero entries (provided that they are below rows with nonzero entries), but Problem 2.3.7 forbids the R-factor in a CR-factorization from having a row of all zero entries. However, up to the rows of zeros, a matrix in RREF has effectively the same behavior as the R-factor in a CR-factorization. In particular, the pivot columns of such an R-factor are standard basis vectors, and the same is true for a matrix in RREF.

3.3.10 Example. Let

$$R_0 = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The leading 1's appear in columns 2 and 4, and, by long experience, these are the pivot columns of R_0.

3.3.11 Lemma. *Let $R_0 \in \mathbb{R}^{m \times n}$ be a nonzero matrix in RREF. The pivot columns of R_0 are exactly the columns of R_0 with leading 1's.*

Proof. As stated above, since R_0 is a nonzero matrix, it has at least one nonzero entry, and so at least one row has a nonzero entry. That row therefore has a leading nonzero entry, thus a leading 1, and so R_0 does have some columns with leading 1's. Call those columns $\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r}$ with $j_i < j_{i+1}$ for $i = 1, \ldots, r-1$. We show that this list of columns satisfies the properties of pivot columns given precisely in Remark 1.14.3. We break the proof into the following punishingly intricate steps.

1. *For each $i = 1, \ldots, r$, there is k with $1 \le k \le r$ such that $\mathbf{r}_{j_i} = \mathbf{e}_k$.* Any column with a leading 1 is a standard basis vector $\mathbf{e}_k \in \mathbb{R}^m$ by Column Property 1. We need to show $1 \le k \le r$. Suppose that for some i, there is $k \ge r+1$ such that $\mathbf{r}_{j_i} = \mathbf{e}_k$. In particular, row $k \ge r+1$ has a nonzero entry.

However, rows 1 through r now contain at most $r-1$ leading 1's, and so at least one row between rows 1 and r has no leading 1. Then that row must have all zero entries (as otherwise it has a nonzero entry, thus a leading nonzero entry, thus a leading 1). But then a row between rows 1 and r has all zero entries, whereas a row between rows $r+1$ and m has a nonzero entry. This contradicts Row Property 1.

So, for each $i = 1, \ldots, r$, there exists k_i with $1 \le k_i \le r$ such that $\mathbf{r}_{j_i} = \mathbf{e}_{k_i}$. We want to show $k_i = i$. To be clear, saying $\mathbf{r}_{j_1} = \mathbf{e}_{k_i}$ means that the leading 1 in row k_i occurs in column j_i. Moreover, each row from row 1 to row r has a leading 1; otherwise, as demonstrated in the paragraph above, that row would have all zero entries, whereas a row below it would have nonzero entries.

2. *If $j < j_1$, then $\mathbf{r}_j = \mathbf{0}_m$.* Suppose instead that $\mathbf{r}_j \neq \mathbf{0}_m$ for some $j < j_1$. Then row i of $\mathbf{r}_j$ is nonzero for some i, and so row i has a leading nonzero entry, thus a leading 1, in some column $k \le j$. But column j_1 is the first column with a leading 1.

3. $\mathbf{r}_{j_1} = \mathbf{e}_1$. Suppose instead that $\mathbf{r}_{j_1} = \mathbf{e}_k$ with $k \ge 2$. Since R_0 is a nonzero matrix, row 1 cannot have all zero entries, for then there would be a nonzero entry in row 2 or below, and that would contradict Row Property 1. So, row 1 has a nonzero entry, thus a leading nonzero entry, thus a leading 1. This entry cannot occur in column j with $j < j_1$, as column j_1 is the first column with a leading 1, and it cannot occur in column j_1, since row 1 of column j is zero. The leading 1 in row 1 therefore occurs in column $j > j_1$.

Now take $i_1 = 1$, $j_{i_1} = j$, $i_2 = k$, and $j_{i_2} = j_1$. Then $i_1 < i_2$ but $j_{i_2} < j_{i_1}$. The leading 1 of row i_1 is an entry in column j_{i_1}; the leading 1 of row i_2 is an entry in column j_{i_2}. This contradicts Column Property 2. We must therefore have $\mathbf{r}_{j_1} = \mathbf{e}_1$.

4. $\mathbf{r}_{j_i} = \mathbf{e}_i$ *for* $2 \le i \le r$. Suppose that this is not true and that i is the first column for which it fails. That is, $i \ge 2$, $\mathbf{r}_{j_i} \neq \mathbf{e}_i$, but

$$\mathbf{r}_{j_\ell} = \mathbf{e}_\ell, \ 1 \le \ell \le i-1. \tag{3.3.1}$$

We must have $\mathbf{r}_{j_i} = \mathbf{e}_k$ for some $1 \le k \le r$, so now $k \neq i$. Then the leading 1 of row k is in column j_i. Let the leading 1 of row i be in column j.

Case 1: $i < k$. Take $i_1 = i$, $i_2 = k$, $j_{i_1} = j$, and $j_{i_2} = j_i$ to find, from Column Property 2, that $j < j_i$. Since j is the index of a column with a leading 1, and $j < j_i$, it must be the case that $j = j_\ell$ for some $1 \le \ell \le i-1$. But then $\mathbf{r}_j = \mathbf{r}_{j_\ell} = \mathbf{e}_\ell$ by (3.3.1), and so the leading 1 in column j is in row $\ell < i$, not in row i.

Case 2: $k < i$. Then $\mathbf{r}_{j_k} = \mathbf{e}_k$ by (3.3.1), but also $\mathbf{r}_{j_i} = \mathbf{e}_k$ by the assumption above. This says that columns j_k and j_i both contain leading 1's for row k, which is impossible, because a leading 1 for a row can only appear in one column.

5. *If* $j_i < j < j_{i+1}$, *then* $\mathbf{r}_j \in \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_i})$. Since $\operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_i}) = \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_i)$, this is true if $\mathbf{r}_j$ has zero entries in rows $i+1$ through m. Suppose instead that $\mathbf{r}_j$ has a nonzero entry in one of those rows $i+1$ through m. Suppose that this happens in row $\ell \ge i+1$, so row ℓ has a nonzero entry in column j and so it has a leading nonzero entry in column j or before, thus a leading 1 in column j or before. As column j is not a pivot column, it does not have a leading nonzero entry in any of its rows, and so the leading 1 in row $\ell \ge i+1$ occurs in a column before column j. If this column is column $j_k < j$, then $\mathbf{r}_{j_k} = \mathbf{e}_\ell$. We now know that $\mathbf{r}_{j_k} = \mathbf{e}_k$, so $k = \ell \ge i+1$. Then $j_{i+1} = j_k < j < j_{i+1}$, a contradiction.

6. *If* $j > j_r$, *then* $\mathbf{r}_j \in \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r})$. First, the entries in rows $r+1$ through m of R are all zero. Otherwise, a row in rows $r+1$ through m would have a nonzero entry, thus a leading nonzero entry, thus a leading 1. But R_0 already has leading 1's in rows 1 through r, and so R_0 would have at least $r+1$ leading 1's, a contradiction. So, every column in R_0 has zero entries in rows $r+1$ through m, and so every column of R_0 is in $\operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_r) = \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r})$. ■

We now have two notions of pivot columns. First, there is the original concept from Definition 1.14.2 and Remark 1.14.3; this is a "dynamic" view of pivot columns, as it has them talking to each other via spans. Second, there is the new idea that the pivot columns of a matrix in RREF are those columns with leading 1's; this is a "static" view, as the columns are just sitting there, and we check which ones have leading 1's. But this second perspective is *helpful* because it reveals all of the pivot columns all at once if the matrix is in RREF. We will soon generalize this to matrices not in RREF.

3.3.12 Example. **(i)** Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}$$

We know from part (ii) of Example 2.3.1 that the CR-factorization of A is

$$A = CR = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

and that the RREF of A is

$$R_0 = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Hopefully the connections between C, R, and R_0 are obvious. First, the pivot columns of A occur in the same locations as the pivot columns of the RREF, and the latter are easy to find, since they are the columns with leading 1's. This gives us the factor C in the CR-factorization. Second, we obtain the factor R in the CR-factorization by chopping off the zero row(s) in the RREF.

(ii) Suppose that we only know the RREF of the matrix A above and the matrix E that collects the elementary matrices whose action converts A to RREF. That is, we have

$$EA = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} =: R_0,$$

where $E \in \mathbb{R}^{3\times 3}$ is invertible. Without even knowing the exact entries of E (which, in Gaussian/Gauss–Jordan elimination, we typically would not want to compute), we can recover the CR-factorization of A. Let $E^{-1} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{v}_1 \end{bmatrix}$. Then

$$A = E^{-1}R_0 = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{c}_1 & 2\mathbf{c}_1 & \mathbf{c}_2 & (3\mathbf{c}_1 + 4\mathbf{c}_2) \end{bmatrix}.$$

The pivot columns of A are therefore columns 2 and 4. Here is why. Column 2 equals $\mathbf{c}_1$, which is nonzero since it is a column of the invertible matrix E^{-1}, and so column 2 is the first nonzero column of A. And column 4 is the first column of A not in the span of column 2. Now factor

$$A = \begin{bmatrix} \mathbf{0}_3 & \mathbf{c}_1 & 2\mathbf{c}_1 & \mathbf{c}_2 & (3\mathbf{c}_1 + 4\mathbf{c}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}.$$

This is the CR-factorization of A.

The perspectives of this example give us a strategy for accomplishing two things simultaneously. First, we will prove the uniqueness of the RREF of a matrix, which Theorem 3.3.7 did not provide. Second, in the process we will show how to extract the CR-factorization from the RREF.

We begin with an auxiliary result that gives a slightly different perspective on the CR-factorization.

3.3.13 Lemma. *Let $A \in \mathbb{R}^{m \times n}$. Suppose that $A = CR$ for some $C \in \mathbb{R}^{m \times r}$ and $R \in \mathbb{R}^{r \times n}$, where C has independent columns and R is in RREF with no nonzero rows. Then the columns of C are the pivot columns of A, the pivot columns of A occur in the same locations as the pivot columns of R, and $A = CR$ is the CR-factorization of A.*

Proof. Since R has no nonzero rows, every row of R has a nonzero entry, thus a leading nonzero entry, thus a leading 1. And since R has r rows, R has r pivot columns, which we list as $(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r})$. And because R is in RREF, we know that $\mathbf{r}_{j_i} = \mathbf{e}_i \in \mathbb{R}^r$. Finally, because $\mathbf{a}_{j_i} = C\mathbf{r}_{j_i} = C\mathbf{e}_i$, the columns of C are columns of A. We show that $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$ is the list of pivot columns of A by checking the four conditions of Remark 1.14.3.

(i) Let $j < j_1$. Then $\mathbf{r}_j = \mathbf{0}_r$, so $\mathbf{a}_j = C\mathbf{r}_j = \mathbf{0}_m$. Since the columns of C are independent and $\mathbf{r}_{j_1} \neq \mathbf{0}_r$, likewise $\mathbf{a}_{j_1} = C\mathbf{r}_{j_1} \neq \mathbf{0}_m$.

(ii) Now let $r \geq 2$ and $i \geq 2$. If $\mathbf{a}_{j_i} \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$, then $C\mathbf{r}_{j_i} \in \operatorname{span}(C\mathbf{r}_{j_1}, \ldots, C\mathbf{r}_{j_{i-1}})$. Since the columns of C are independent, it follows that $\mathbf{r}_{j_i} \in \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_{i-1}})$, a contradiction.

(iii) Continue to assume $r \geq 2$ and $i \geq 2$ and now suppose $j < j_i$. Then $\mathbf{a}_j = C\mathbf{r}_j \in \operatorname{span}(C\mathbf{r}_{j_1}, \ldots, C\mathbf{r}_{j_{i-1}}) = \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$.

(iv) Last, let $j > j_r$. Then $\mathbf{a}_j = C\mathbf{r}_j \in \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r}) = \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$.

Since the columns of C are therefore the pivot columns of A, the factorization $A = CR$ is the CR-factorization of A, by uniqueness of that factorization. ■

Here is the improvement of Theorem 3.3.7.

3.3.14 Theorem. *Let $A \in \mathbb{R}^{m \times n}$.*

(i) *There exist an invertible matrix $E \in \mathbb{R}^{m \times m}$ and a unique matrix $R_0 \in \mathbb{R}^{m \times n}$ such that R_0 is in RREF. We call R_0 the* ***RREF of*** *A and write $R_0 = \mathsf{rref}(A)$.*

(ii) *Let $A = CR$ be the CR-factorization of A. If R_0 has no zero rows, then $R = R_0$. If R_0 has one or more zero rows, then R_0 has the block structure*

$$R_0 = \begin{bmatrix} R \\ 0 \end{bmatrix}.$$

(iii) *The pivot columns of A, R, and R_0 all occur in the same locations. That is, if* *columns $j_1, \ldots, j_r$ of A are the pivot columns of A, then columns $j_1, \ldots, j_r$ of R are the pivot columns of R, and columns $j_1, \ldots, j_r$ of R_0 are the pivot columns of R_0.*

Proof. Theorem 3.3.7 gave the existence of the RREF: we can write $EA = R_0$, where $E \in \mathbb{R}^{m\times m}$ is invertible and $R_0 \in \mathbb{R}^{m\times n}$ is in RREF. Since $E \in \mathbb{R}^{m\times m}$ is invertible, $\operatorname{rank}(E) = m$, and so Problem 2.4.10 tells us that the pivot columns of A and of $R_0 = EA$ occur in the same locations. We show below the uniqueness of R_0 and the relation of R_0 to the factor R from the CR-factorization of A; we already know from Theorem 2.3.5 that the pivot columns of A and of R occur in the same locations.

If A is the zero matrix, then $R_0 = EA$ is the zero matrix, and this is the only choice for R_0. So, assume that A is not the zero matrix and write $A = E^{-1}R_0$. Then R_0 is not the zero matrix and so has at least one row with nonzero entries. We show that there is only one possible choice for R_0 in this case, too.

1. *R_0 has no zero rows.* Put $C := E^{-1}$ and $R := R_0$. Then $A = CR$, where the columns of C are independent and $R = R_0$ is in RREF with no nonzero rows. Lemma 3.3.13 shows that $A = CR$ is the CR-factorization of A, and so there is only one choice for $R = R_0$.

2. *R_0 has one or more zero rows.* Write R_0 in the block form

$$R_0 = \begin{bmatrix} R \\ 0 \end{bmatrix},$$

where R has at least one nonzero entry in each row, and where the symbol 0 denotes one or more zero rows. Since R_0 is in RREF, so is R.

Specifically, suppose that $R \in \mathbb{R}^{r\times n}$. Now write $E^{-1} = \begin{bmatrix} C & V \end{bmatrix}$, where $C \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{r\times(m-r)}$. Then

$$A = E^{-1}R_0 = CR.$$

Again, we have written A in the form $A = CR$, where the columns of C are independent and R is in RREF with no nonzero rows. Lemma 3.3.13 again applies to show that $A = CR$ is the CR-factorization of A, and so there is only one choice for R. This choice of R completely determines R_0. ■

3.4. The RREF reveals the complete structure of solutions to linear systems.

Once we know the RREF of a matrix, it is easy to find its CR-factorization. Since the null space of a matrix and the R-factor in its CR-factorization are the same, and since the data in the R-factor is arguably less complicated than the data in the original matrix, having the R-factor allows us to find the null space of a matrix more easily. However, the algorithmic process of Gaussian and Gauss–Jordan elimination that reduces a matrix to RREF does not merely give us the RREF and the null space. Rather, this process enables us to find *all* solutions to the fundamental problem $A\mathbf{x} = \mathbf{b}$.

3.4.1 Example. Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$

We study $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$ arbitrary.

For a small problem like this, the most efficient approach is to put the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ into RREF in the form $\begin{bmatrix} R_0 & E\mathbf{b} \end{bmatrix}$, where $R_0 = \mathsf{rref}(A)$ and $EA = R_0$ with E invertible. We basically repeat the steps of Example 3.3.1, where we were taking $\mathbf{b} = \mathbf{0}_3$ throughout. This time, however, we skip writing out the elementary matrices that do all the elimination; refer back to that example as needed. While there are many ways that we could do the arithmetic, we follow here the pseudocode of the proof of Theorem 3.3.7.

$$\left[\begin{array}{ccccc|c} 0 & 1 & 2 & 1 & 7 & b_1 \\ 0 & 2 & 4 & 2 & 14 & b_2 \\ 0 & 0 & 0 & 2 & 8 & b_3 \end{array}\right] \xrightarrow{\mathsf{R2} \mapsto \mathsf{R2}-2\times\mathsf{R1}} \left[\begin{array}{ccccc|c} 0 & 1 & 2 & 1 & 7 & b_1 \\ 0 & 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 0 & 2 & 8 & b_3 \end{array}\right]$$

$$\xrightarrow{\mathsf{R3} \mapsto (1/2)\times\mathsf{R3}} \left[\begin{array}{ccccc|c} 0 & 1 & 2 & 1 & 7 & b_1 \\ 0 & 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 0 & 1 & 4 & b_3/2 \end{array}\right]$$

$$\xrightarrow{\mathsf{R1} \mapsto \mathsf{R1}-\mathsf{R3}} \left[\begin{array}{ccccc|c} 0 & 1 & 2 & 0 & 3 & b_1 - b_3/2 \\ 0 & 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 0 & 1 & 4 & b_3/2 \end{array}\right]$$

$$\xrightarrow{\mathsf{R2} \mapsto \mathsf{R3},\ \mathsf{R3}\mapsto\mathsf{R2}} \left[\begin{array}{ccccc|c} 0 & 1 & 2 & 0 & 3 & b_1 - b_3/2 \\ 0 & 0 & 0 & 1 & 4 & b_3/2 \\ 0 & 0 & 0 & 0 & 0 & b_2 - 2b_1 \end{array}\right].$$

Then $A\mathbf{x} = \mathbf{b}$ is equivalent to

$$\left\{ \begin{array}{rcl} x_2 + 2x_3 + 3x_5 & = & b_1 - b_3/2 \\ x_4 + 4x_5 & = & b_3/2 \\ 0 & = & b_2 - 2b_1. \end{array} \right. \tag{3.4.1}$$

The third equation is a "solvability condition": if $A\mathbf{x} = \mathbf{b}$, then we must have $b_2 = 2b_1$. This is not the first time that we have seen this condition, and it should be apparent from the row structure of A (the second row is twice the first row). But here we emphasize that we can solve $A\mathbf{x} = \mathbf{b}$ if and only if we can solve the (simpler) problem (3.4.1); this is just because all of the row operations are reversible (elementary matrices are invertible!). And if $b_2 = 2b_1$, then we can solve (3.4.1). Thus $\mathbf{C}(A) = \{\mathbf{b} \in \mathbb{R}^3 \mid b_2 = 2b_1\}$.

When the solvability condition is met, the solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_1(b_1 - b_3/2) - 2x_3 - 3x_5 \\ x_3 \\ b_3/2 - 4x_5 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ b_1 - b_3/2 \\ 0 \\ b_3/2 \\ 0 \end{bmatrix} + x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

This is a wonderfully transparent solution formula. The "free" variables x_3 and x_5 quantify the nonuniqueness: each choice of x_3 and x_5 gives a different solution. And the solvability condition $b_2 = 2b_1$ makes precise the lack of existence: if $b_2 \neq b_1$, then there is no solution. (By the way, only b_1 and b_3 show up in the solution, since $b_2 = 2b_1$.)

Assuming $b_2 = 2b_1$ and taking $x_3 = x_5 = 0$, we conclude that one "particular" solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x}_\star := (0, b_1 - b_3/2, 0, b_3/2, 0)$, while all other solutions are $\mathbf{x} = \mathbf{x}_\star + c_1\mathbf{z}_1 + c_2\mathbf{z}_2$, where $\mathbf{z}_1 = (1,0,0,0,0)$, $\mathbf{z}_2 = (0,-2,1,0,0)$, and $\mathbf{z}_3 = (0,-3,0,-4,1)$. By the way, taking $\mathbf{b} = \mathbf{0}_3$ (i.e., $b_1 = b_2 = b_3 = 0$), we obtain

$$\mathbf{N}(A) = \mathbf{C}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & -4 \\ 0 & 0 & 1 \end{bmatrix}\right),$$

which is the same column space that appeared in Example 3.2.3. This is because the R-factor in the CR-factorization of this A is the same matrix from that example.

3.4.2 Example. Let 3.4.2

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & 4 & 0 \\ 1 & 2 & 2 \\ 7 & 14 & 8 \end{bmatrix}.$$

This is, of course, the transpose of the matrix from Examples 3.3.1 and 3.4.1. We study $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} = (b_1, b_2, b_3, b_4, b_5)$.

We have

$$[A \quad \mathbf{b}] = \left[\begin{array}{ccc|c} 0 & 0 & 0 & b_1 \\ 1 & 2 & 1 & b_2 \\ 2 & 4 & 0 & b_3 \\ 1 & 2 & 2 & b_4 \\ 7 & 14 & 8 & b_5 \end{array}\right] \xrightarrow{\mathsf{R3} \mapsto \mathsf{R3} - 2\times\mathsf{R2}} \left[\begin{array}{ccc|c} 0 & 0 & 0 & b_1 \\ 1 & 2 & 0 & b_2 \\ 0 & 0 & 0 & b_3 - 2b_2 \\ 1 & 2 & 2 & b_4 \\ 7 & 14 & 8 & b_5 \end{array}\right]$$

$$\xrightarrow{\text{R4} \mapsto \text{R4}-\text{R2}} \left[\begin{array}{ccc|c} 0 & 0 & 0 & b_1 \\ 1 & 2 & 0 & b_2 \\ 0 & 0 & 0 & b_3 - 2b_2 \\ 0 & 0 & 2 & b_4 - b_2 \\ 7 & 14 & 8 & b_5 \end{array}\right]$$

$$\xrightarrow{\text{R5} \mapsto \text{R5}-7\times\text{R2}} \left[\begin{array}{ccc|c} 0 & 0 & 0 & b_1 \\ 1 & 2 & 0 & b_2 \\ 0 & 0 & 0 & b_3 - 2b_2 \\ 0 & 0 & 2 & b_4 - b_2 \\ 0 & 0 & 8 & b_5 - 7b_2 \end{array}\right]$$

$$\xrightarrow{\text{R5} \mapsto \text{R5}-4\times\text{R4}} \left[\begin{array}{ccc|c} 0 & 0 & 0 & b_1 \\ 1 & 2 & 0 & b_2 \\ 0 & 0 & 0 & b_3 - 2b_2 \\ 0 & 0 & 2 & b_4 - b_2 \\ 0 & 0 & 0 & (b_5 - 7b_2) - 4(b_4 - b_2) \end{array}\right]$$

$$\xrightarrow{\text{R4} \mapsto (1/2)\times\text{R4}} \left[\begin{array}{ccc|c} 0 & 0 & 0 & b_1 \\ 1 & 2 & 0 & b_2 \\ 0 & 0 & 0 & b_3 - 2b_2 \\ 0 & 0 & 1 & (b_4 - b_2)/2 \\ 0 & 0 & 0 & b_5 - 4b_4 - 3b_2 \end{array}\right]$$

$$\xrightarrow{\text{R2} \mapsto \text{R1},\ \text{R1} \mapsto \text{R2}} \left[\begin{array}{ccc|c} 1 & 2 & 0 & b_2 \\ 0 & 0 & 0 & b_1 \\ 0 & 0 & 1 & (b_4 - b_2)/2 \\ 0 & 0 & 0 & b_3 - 2b_2 \\ 0 & 0 & 0 & b_5 - 4b_4 - 3b_2 \end{array}\right]$$

$$\xrightarrow{\text{R3} \mapsto \text{R2},\ \text{R2} \mapsto \text{R3}} \left[\begin{array}{ccc|c} 1 & 2 & 0 & b_2 \\ 0 & 0 & 1 & (b_4 - b_2)/2 \\ 0 & 0 & 0 & b_1 \\ 0 & 0 & 0 & b_3 - 2b_2 \\ 0 & 0 & 0 & b_5 - 4b_4 - 3b_2 \end{array}\right]$$

Along the way, we simplified

$$(b_5 - 7b_2) - 4(b_4 - b_2) = b_5 - 4b_4 - 3b_2.$$

The problem $A\mathbf{x} = \mathbf{b}$ is then equivalent to

$$\begin{cases} x_1 + 2x_2 & = b_2 \\ x_3 & = (b_4 - b_2)/2 \\ 0 & = b_1 \\ 0 & = b_3 - 2b_2 \\ 0 & = b_5 - 4b_4 - 3b_2 \end{cases}$$

We now have *three* solvability conditions:

$$b_1 = 0, \qquad b_3 - 2b_2 = 0, \qquad \text{and} \qquad b_5 - 4b_4 - 3b_2 = 0.$$

If these are met, then the solution $\mathbf{x}$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_2 - 2x_2 \\ x_2 \\ (b_4 - b_2)/2 \end{bmatrix} = \begin{bmatrix} b_2 \\ 0 \\ (b_4 - b_2)/2 \end{bmatrix} + \begin{bmatrix} -2x_2 \\ x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} b_2 \\ 0 \\ (b_4 - b_2)/2 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

We should see the same solution structure as in Example 3.4.1: taking $\mathbf{x}_\star = (b_2, 0, (b_4 - b_2)/2)$ gives a solution $A\mathbf{x}_\star = \mathbf{b}$ when the solvability conditions are met, and then every other solution is $\mathbf{x} = \mathbf{x}_\star + x_2\mathbf{z}_1$ for some $x_2 \in \mathbb{R}$, where $\mathbf{z}_1 = (-2, 1, 0)$. A byproduct of this calculation is that

$$\mathbf{N}(A) = \mathbf{C}\left(\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}\right) = \text{span}\left(\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}\right). \tag{3.4.2}$$

By the way, we also figured out that

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

3.4.3 Problem (!). Find elementary matrices that perform all of the row operations in the previous example. Since $A \in \mathbb{R}^{4\times 3}$ in that example, all of the elementary matrices will be 4×4.

Here is the structural pattern of solutions that we are seeing from Examples 3.4.1 and 3.4.2.

3.4.4 Theorem. *Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Suppose that $\mathbf{x}_\star \in \mathbb{R}^n$ satisfies $A\mathbf{x}_\star = \mathbf{b}$. Then any other solution $\mathbf{x} \in \mathbb{R}^n$ to $A\mathbf{x} = \mathbf{b}$ has the form $\mathbf{x} = \mathbf{x}_\star + \mathbf{z}$ for some $\mathbf{z} \in \mathbf{N}(A)$.*

3.4.5 Problem (!). Prove it. [Hint: *what does $\mathbf{x} - \mathbf{x}_\star$ do?*]

It looks like we have a "decomposition" from Theorem 3.4.4 for solutions to $A\mathbf{x} = \mathbf{b}$. Any solution $\mathbf{x}$ is the sum of one "particular" solution and a vector in the null space. As in so many other places in the course, we need to build some more tools, but eventually we will be able to say a bit more about what that "particular" solution is doing, and maybe how to choose it best when we have many options.

Content from Strang's *ILA* 6E. Read all of p. 104 again and now read p. 105. Then look at Figure 3.1 on p. 107 and think about how we are representing solutions to $A\mathbf{x} = \mathbf{b}$ "parametrically" with parameters involving solutions to $A\mathbf{x} = \mathbf{0}$.

3.4.6 Remark. *A column of a matrix that is not a pivot column is often called a* ***FREE COLUMN****. In the problem $A\mathbf{x} = \mathbf{b}$, we call the unknown x_j a* ***PIVOT VARIABLE*** *of A if column j of A is a pivot column, and call x_j a* ***FREE VARIABLE*** *if column j of A is a free column. To solve $A\mathbf{x} = \mathbf{b}$, convert the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to RREF $\begin{bmatrix} R_0 & \mathbf{c} \end{bmatrix}$, and then solve the equation $R\mathbf{x} = \mathbf{c}$ for the pivot variables in terms of the free variables.*

3.4.7 Problem (!). Explain why $\mathbf{C}(A) \neq \mathbf{C}(\mathsf{rref}(A))$ in general. But explain why $\mathbf{N}(A) = \mathbf{N}(\mathsf{rref}(A))$ always.

3.5. Extreme rank matrices are special.

First, we can use the RREF to get a sharper bound on rank. Let $A \in \mathbb{R}^{m \times n}$ be nonzero. Since $\text{rank}(A)$ is the number of pivot columns of A, and A has n columns, there can be at most n pivot columns, so $\text{rank}(A) \leq n$. We have known this since we first introduced the concept of rank.

Here is what is new. Theorem 3.3.14 tells us that the pivot columns of A and $\mathsf{rref}(A)$ occur in the same locations, so in particular $\text{rank}(A) = \text{rank}(\mathsf{rref}(A))$. By Lemma 3.3.11, the pivot columns of $\mathsf{rref}(A)$ are the columns with leading 1's. Each leading 1 needs to occur in a new row, and there are only m rows in $\mathsf{rref}(A)$. So, there can be at most m leading 1's in $\mathsf{rref}(A)$, thus at most m pivot columns in $\mathsf{rref}(A)$. Then $\text{rank}(A) = \text{rank}(\mathsf{rref}(A)) \leq m$.

3.5.1 Lemma. *Let $A \in \mathbb{R}^{m \times n}$. Then* $\text{rank}(A) \leq \min\{m, n\}$.

Much of our work in this course focuses on the columns of a matrix. This estimate is one of those times where knowledge of the rows is valuable. Of course we have seen this computationally when using the dot product to obtain matrix-vector and matrix-matrix products. The theoretical value of a "row perspective" will become more apparent when we take up orthogonality and "fill in" what is "missing" from $\mathbb{R}^n$ beyond the null space and from $\mathbb{R}^m$ beyond the column space.

Examples 3.4.1 and 3.4.2 involved matrices $A \in \mathbb{R}^{m \times n}$ with $1 \leq \text{rank}(A) < \min\{m, n\}$. Some interesting things happen in the "extreme" case of $\text{rank}(A) = \min\{m, n\}$.

3.5.2 Example. **(i)** Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

so $\text{rank}(A) = 2$, which is the number of rows of A. We can always solve $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} = (b_1, b_2)$ by taking $x_1 = b_1$, $x_2 = b_2$, and x_3 to be anything, although that freedom in x_3 means $\mathbf{N}(A) \neq \{\mathbf{0}_3\}$, so solutions to $A\mathbf{x} = \mathbf{b}$ are not unique. This is also not surprising, as $\text{rank}(A) \neq 3$, so the columns of A must be dependent.

(ii) Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Then $\text{rank}(A) = 2$, which is the number of columns of A, so every column is a pivot column, and therefore $\mathbf{N}(A) = \{\mathbf{0}_2\}$. (We have done this before.) Consequently, solutions to $A\mathbf{x} = \mathbf{b}$, if they exist, are unique; however, if $\mathbf{b} = (b_1, b_2, b_3) \in \mathbf{C}(A)$, then $b_3 = 0$, so $\mathbf{C}(A) \neq \mathbb{R}^3$.

Content from Strang's *ILA* 6E. Read Example 1 on pp. 105–106 and Example 2 on p. 107.

Here is what these examples teach us.

3.5.3 Theorem. *Let* $A \in \mathbb{R}^{m \times n}$.

(i) *Suppose that* $m \leq n$ *and* $\text{rank}(A) = m$. *Then* $\mathbf{C}(A) = \mathbb{R}^m$. *That is, we can always solve* $A\mathbf{x} = \mathbf{b}$ *for any* $\mathbf{b} \in \mathbb{R}^m$. *In this case, we say that* A *has* ***FULL ROW RANK***.

(ii) *Suppose that* $n \leq m$ *and* $\text{rank}(A) = n$. *Then* $\mathbf{N}(A) = \{\mathbf{0}_n\}$. *That is, if we can solve* $A\mathbf{x} = \mathbf{b}$ *for some* $\mathbf{b} \in \mathbb{R}^m$, *then the solution* $\mathbf{x}$ *is unique. In this case, we say that* A *has* ***FULL COLUMN RANK***.

Proof. **(i)** Since $\text{rank}(A) = m$ and $\text{rank}(A) = \min\{m, n\}$, we have $m \leq n$ here. If $m = n$, then $A \in \mathbb{R}^{m \times m}$ is square with m independent columns and therefore invertible, so $\mathbf{C}(A) = \mathbb{R}^m$ by the invertible matrix theorem.

Otherwise, suppose $m < n$. (By the way, since $m \geq 1$, here $n \geq 2$.) Let $A = CR$ be the CR-factorization of A. Since $\text{rank}(A) = m$, the factor $C \in \mathbb{R}^{m \times m}$ is square with independent columns and therefore is invertible. To have $A\mathbf{x} = \mathbf{b}$ for some $\mathbf{b} \in \mathbb{R}^m$, it therefore suffices to be able to solve $R\mathbf{x} = C^{-1}\mathbf{b} =: \mathbf{c} \in \mathbb{R}^m$. And since $m < n$, we can write $R = \begin{bmatrix} I_m & F \end{bmatrix} P$ for some $F \in \mathbb{R}^{m \times (n-m)}$ and a permutation matrix $P \in \mathbb{R}^{n \times n}$.

We therefore have $R\mathbf{x} = \mathbf{c}$ if and only if $\begin{bmatrix} I_m & F \end{bmatrix} P\mathbf{x} = \mathbf{c}$. Inspired by (3.2.3), take $\mathbf{y} = P\mathbf{x}$. Then

$$R\mathbf{x} = \mathbf{c} \iff \begin{cases} \begin{bmatrix} I_m & F \end{bmatrix} \mathbf{y} = \mathbf{c} \\ P\mathbf{x} = \mathbf{y}. \end{cases}$$

To solve $\begin{bmatrix} I_m & F \end{bmatrix} \mathbf{y} = \mathbf{c}$, one option is to take

$$\mathbf{y} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0}_{n-m} \end{bmatrix}$$

and then put $\mathbf{x} = P^{-1}\mathbf{y}$.

(ii) If $\text{rank}(A) = n$, then every column of A is a pivot column, so all of the columns of A are independent, and therefore $\mathbf{N}(A) = \{\mathbf{0}_n\}$. ■

3.5.4 Problem (!). Let $1 \leq m < n$, $F \in \mathbb{R}^{m \times (n-m)}$, and $\mathbf{c} \in \mathbb{R}^m$. Check that taking

$$\mathbf{y} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0}_{n-m} \end{bmatrix}$$

does solve $\begin{bmatrix} I_m & F \end{bmatrix} \mathbf{y} = \mathbf{c}$.

Content from Strang's *ILA* 6E. Read the rest of p. 106 starting from "This example is typical...." Then read all of p. 108.

Content from Strang's *ILA* 6E. Read all of the "Worked Examples" on pp. 109–110.

We finally resolve Conjecture 3.1.2 using the RREF.

3.5.5 Theorem. *Let $A \in \mathbb{R}^{m \times n}$.*

(i) *If $n < m$, then $\mathbf{C}(A) \neq \mathbb{R}^m$, and so we cannot always solve $A\mathbf{x} = \mathbf{b}$.*

(ii) *If $m < n$, then $\mathbf{N}(A) \neq \{\mathbf{0}_n\}$, and so solutions to $A\mathbf{x} = \mathbf{b}$, if they exist (and maybe they do not), are never unique.*

Proof. **(i)** If $n < m$, then $\text{rank}(A) \leq \min\{m, n\} = n < m$. Let $R_0 = \mathsf{rref}(A)$. Then R_0 has fewer than m pivot columns, so R_0 has fewer than m leading 1's. That is, at least one row of R_0 cannot have a leading 1, and so that row has all zero entries. To be specific, say that row i of R_0 has all zero entries. Then there is no $\mathbf{x} \in \mathbb{R}^n$ such that $R_0\mathbf{x} = \mathbf{e}_i$.

Now let $E \in \mathbb{R}^{m \times m}$ be invertible with $EA = R_0$. We claim that there is no $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = E^{-1}\mathbf{e}_i$. Otherwise, we would have $R_0\mathbf{x} = EA\mathbf{x} = \mathbf{e}_i$. So, $E^{-1}\mathbf{e}_i \notin \mathbf{C}(A)$, and therefore $\mathbf{C}(A) \neq \mathbb{R}^m$. This proves part (i) of the conjecture.

(ii) If $m < n$, then $\text{rank}(A) \leq \min\{m, n\} = m < n$, and so not every column of A is a pivot column. Then the columns of A are dependent, so $\mathbf{N}(A) \neq \{\mathbf{0}_n\}$, and therefore solutions to $A\mathbf{x} = \mathbf{b}$, if they exist, cannot be unique. This proves part (ii) of the conjecture. ■

Content from Strang's *ILA* 6E. This is discussed in the "Important" box on p. 98 and the two paragraphs preceding that. Now read Example 4 on p. 98.

3.5.6 Problem (!). Prove that any list in $\mathbb{R}^n$ of length greater than n is dependent.

3.5.7 Problem (!). Let $A \in \mathbb{R}^{m\times n}$ and let $r := \text{rank}(A)$. What do you know about existence and/or uniqueness of solutions to the problem $A\mathbf{x} = \mathbf{b}$ if $r < n$? If $r < m$?

3.6. The RREF resolves the rank conjecture

We now, at last, have the technology (namely, Theorem 3.5.5, which we proved with the RREF) to prove Conjecture 1.14.11. This conjecture posited that the rank of a matrix is a critical threshold for efficiently describing the essential data in the matrix. We begin with the last part of the conjecture.

3.6.1 Theorem. *Let $A \in \mathbb{R}^{m\times n}$ with $r := \text{rank}(A) \geq 1$. Any list in $\mathbf{C}(A)$ of length greater than r is dependent.*

Proof. We consider a special illustrative case and then give the general proof.

1. Suppose that $A = \begin{bmatrix}\mathbf{0}_m & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5\end{bmatrix} \in \mathbb{R}^{m\times 5}$ and that columns 2 and 4 are the only pivot columns, so $r = 2$ here. (The exact value of m is irrelevant.) Then the list $\mathbf{a}_2$, $\mathbf{a}_4$ is independent, $\mathbf{a}_3 \in \text{span}(\mathbf{a}_2)$, and $\mathbf{a}_5 \in \text{span}(\mathbf{a}_2, \mathbf{a}_4)$. It suffices to show that any list of length 3 in $\mathbf{C}(A)$ is dependent, as then any longer list contains a dependent sublist (Problem 1.13.5).

Here is the trick. Let $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ be a list in $\mathbf{C}(A)$. Write $\mathbf{v}_1 = w_1\mathbf{a}_2 + w_2\mathbf{a}_4$, $\mathbf{v}_2 = x_1\mathbf{a}_2 + x_w\mathbf{a}_4$, $\mathbf{v}_3 = y_1\mathbf{a}_2 + y_w\mathbf{a}_4$. Put

$$V := \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_4\end{bmatrix} \in \mathbb{R}^{m\times 3} \quad \text{and} \quad C := \begin{bmatrix}\mathbf{a}_2 & \mathbf{a}_4\end{bmatrix} \in \mathbb{R}^{m\times 2}.$$

This matrix C is, of course, the C-factor from the CR-factorization of A, although (shockingly) that factorization does not play an explicit role here.

Then

$$V = \begin{bmatrix}(x_1\mathbf{a}_2 + x_3\mathbf{a}_4) & (y_1\mathbf{a}_2 + y_3\mathbf{a}_4) & (u_1\mathbf{a}_2 + u_3\mathbf{a}_4)\end{bmatrix} = \begin{bmatrix}\mathbf{a}_2 & \mathbf{a}_4\end{bmatrix}\begin{bmatrix}x_1 & y_1 & u_1 \\ x_3 & y_3 & u_3\end{bmatrix} =: CB.$$

The killer is that $B \in \mathbb{R}^{2\times 3}$, so B has more columns than rows. By Theorem 3.5.5, we know $\mathbf{N}(B) \neq \{\mathbf{0}_3\}$. That is, there is $\mathbf{z} \in \mathbb{R}^3$ such that $B\mathbf{z} = \mathbf{0}_2$ and $\mathbf{z} \neq \mathbf{0}_3$. But then $V\mathbf{z} = CB\mathbf{z} = \mathbf{0}_3$, so $\mathbf{z} \in \mathbf{N}(V)$ with $\mathbf{z} \neq \mathbf{0}_3$. Thus $\mathbf{N}(V) \neq \{\mathbf{0}_3\}$, and therefore the columns of V are dependent.

2. Here is the general proof. Let $\mathbf{v}_1, \ldots, \mathbf{v}_{r+1}$ be a list in $\mathbf{C}(A)$. We want to show that this list is dependent.

Say that $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}$ is the list of pivot columns of A and put $C = \begin{bmatrix}\mathbf{a}_{j_1} & \cdots & \mathbf{a}_{j_r}\end{bmatrix}$. The pivot columns span $\mathbf{C}(A)$, so we can write each $\mathbf{v}_k$ as a linear combination of the pivot columns: $\mathbf{v}_k = C\mathbf{x}_k$ for some $\mathbf{x}_k \in \mathbb{R}^r$. Then $V = CX$, where

$$V = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_{r+1}\end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_{r+1}\end{bmatrix} \in \mathbb{R}^{r\times(r+1)}.$$

The matrix X has more columns than rows, so $\mathbf{N}(X) \neq \{\mathbf{0}_{r+1}\}$ by Theorem 3.5.5. Then there is $\mathbf{z} \in \mathbb{R}^{r+1}$ such that $X\mathbf{z} = \mathbf{0}_r$ and $\mathbf{z} \neq \mathbf{0}_{r+1}$. We compute $V\mathbf{z} = CX\mathbf{z} = \mathbf{0}_r$, so $\mathbf{z} \in \mathbf{N}(V)$ with $\mathbf{z} \neq \mathbf{0}_{r+1}$. Then $\mathbf{N}(V) \neq \{\mathbf{0}_{r+1}\}$, and therefore the columns of V are dependent. ■

We prove the next part of the conjecture with a similar argument.

3.6.2 Theorem. *Let $A \in \mathbb{R}^{m\times n}$ with $r := \operatorname{rank}(A) \geq 1$. Then no list of fewer than r vectors in $\mathbf{C}(A)$ can span $\mathbf{C}(A)$.*

Proof. Let $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}$ be the pivot columns of A and put $C := \begin{bmatrix} \mathbf{a}_{j_1} & \cdots & \mathbf{a}_{j_r} \end{bmatrix}$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_s$ be a list in $\mathbf{C}(A)$ that spans $\mathbf{C}(A)$, and suppose $s < r$. Put $V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_s \end{bmatrix} \in \mathbb{R}^{m\times s}$.

Each $\mathbf{a}_{j_i}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_s$, so for $i = 1, \ldots, r$, there is $\mathbf{x}_i \in \mathbb{R}^s$ such that $\mathbf{a}_{j_i} = V\mathbf{x}_i$. Put $X = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_r \end{bmatrix} \in \mathbb{R}^{r\times s}$. Thus $C = VX$.

Since $s < r$, Theorem 3.5.5 implies $\mathbf{N}(X) \neq \{\mathbf{0}_s\}$. Then there is $\mathbf{z} \in \mathbb{R}^r$ such that $\mathbf{z} \neq \mathbf{0}_r$ and $X\mathbf{z} = \mathbf{0}_s$. Thus $C\mathbf{z} = VX\mathbf{z} = V\mathbf{0}_s = \mathbf{0}_m$, so $\mathbf{z} \in \mathbf{N}(C)$ and $\mathbf{z} \neq \mathbf{0}_r$. So, $\mathbf{N}(C) \neq \{\mathbf{0}_r\}$, which means that the columns of C are dependent. This is a contradiction since the columns of C are the pivot columns of A and therefore independent. ■

3.6.3 Problem (!). Why do you think people sometimes call the list of pivot columns of a matrix both a "minimal spanning list" for the column space and a "maximal linearly independent list" in the column space?

3.6.4 Problem ($\star$). The proof of Theorem 3.6.2 contained a technique worth extracting. Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times p}$ with $\mathbf{C}(A) = \mathbf{C}(B)$. Prove that there exists a matrix $X \in \mathbb{R}^{p\times n}$ such that $A = BX$.

Finally, we prove the first part of the conjecture.

3.6.5 Theorem. *Let $A \in \mathbb{R}^{m\times n}$ with $r := \operatorname{rank}(A) \geq 1$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be an independent list in $\mathbf{C}(A)$. Then $\mathbf{C}(A) = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$.*

Proof. Problem 2.4.8 tells us that if $\mathbf{v} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$, then $\mathbf{v} \in \mathbf{C}(A)$. We need to show that if $\mathbf{w} \in \mathbf{C}(A)$, then $\mathbf{w} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$. What goes wrong if there is $\mathbf{w} \in \mathbf{C}(A)$ such that $\mathbf{w} \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$?

Put $\mathbf{v}_{r+1} := \mathbf{w}$. We show that the list $\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}$ is independent. Here we use the linear independence lemma. We know $\mathbf{v}_1 \neq \mathbf{0}_m$, since the list $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is independent, and we also know $\mathbf{v}_j \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1})$ for $j = 2, \ldots, r$, again by the independence of this list $\mathbf{v}_1, \ldots, \mathbf{v}_r$. And the defining property of $\mathbf{v}_{r+1}$ is that $\mathbf{v}_{r+1} = \mathbf{w} \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$.

So, the list $\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}$ satisfies $\mathbf{v}_1 \neq \mathbf{0}_m$ and $\mathbf{v}_j \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1})$ for $j = 2, \ldots, r+1$. The linear independence lemma implies that the list $\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}$ is independent. But this is a list of length $r+1 > r$ in $\mathbf{C}(A)$, and so it must be dependent by Theorem 3.6.1. ■

3.7. We should have more questions.

It looks like we have accomplished the major goal of the course: solve $A\mathbf{x} = \mathbf{b}$ and understand when we cannot. We apply Gaussian and Gauss–Jordan elimination to the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to convert it to $\begin{bmatrix} R_0 & \mathbf{c} \end{bmatrix}$ with $R_0 = \mathsf{rref}(A)$, and then we study $R_0\mathbf{x} = \mathbf{c}$. The upshot is that $A\mathbf{x} = \mathbf{b}$ and $R_0\mathbf{x} = \mathbf{c}$ have the same solutions (if any at all), but the structure of R_0 is probably much simpler than the structure of A (specifically, there are zero entries in strategic places in R_0).

For the problem to have a solution, if the ith row of R_0 is all zero, then the ith entry of $\mathbf{c}$ must be 0; this translates into a "solvability condition" on $\mathbf{b}$. Assuming those "solvability conditions" to be true, we then rewrite the system $R_0\mathbf{x} = \mathbf{c}$ as a system of equations and solve for the "free variables" in terms of the "pivot variables" (recall Remark 3.4.6). In the special case that A is square, we could just do Gaussian elimination to convert A to its upper-triangular form U; if all of the diagonal entry of U are nonzero, then we can back-substitute. If a diagonal entry is 0, then we probably should go all the way to RREF to have some control over the null space.

Considering all of this good work, we are now pretty adept at solving $A\mathbf{x} = \mathbf{b}$ (especially when A is square and invertible), but we could still be better at *understanding* $A\mathbf{x} = \mathbf{b}$, particularly at understanding *failure*. This good work motivates some more questions.

Question 1. If uniqueness fails and $\mathbf{N}(A) \neq \{\mathbf{0}_n\}$, can we quantify and qualify how many "degrees of freedom" the null space gives to the problem, and in particular how many "different" solutions there are beyond "infinitely many?" Can we do this in terms of m, n, and $\text{rank}(A)$? Is it possible to pick a "best" solution out of the infinitely many that exist?

Question 2. If existence fails and $\mathbf{C}(A) \neq \mathbb{R}^m$, can we quantify and qualify how much of $\mathbb{R}^m$ the column space "hits" versus "misses"? Can we do this in terms of m, n, and $\text{rank}(A)$? And can we solve a "related" problem that somehow "approximates" our unsolvable problem in a meaningful way?

To answer these questions—and to ask them precisely—we need some new tools. The overarching tool is *abstraction*: what are the *really* essential properties of the things that we're studying?

3.8. Null and column spaces are dynamic.

Our successes going forward will hinge in no small part on a new perspective: how vectors within a given set interact with each other. This may sound weird at first, but trust me that it will feel completely natural soon. We pause from concrete numbers and focus on the *dynamic* aspects of the null space.

Let $A \in \mathbb{R}^{m\times n}$. The null space

$$\mathbf{N}(A) = \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \mathbf{0}_m\}$$

behaves well with respect to the fundamental objects of vector arithmetic.

1. Suppose $\mathbf{v}_1,\ \mathbf{v}_2 \in \mathbf{N}(A)$. Then $A\mathbf{v}_1 = \mathbf{0}_m$ and $A\mathbf{v}_2 = \mathbf{0}_m$, so

$$A(\mathbf{v}_1 + \mathbf{v}_2) = A\mathbf{v}_1 + A\mathbf{v}_2 = \mathbf{0}_m + \mathbf{0}_m = \mathbf{0}_m.$$

Thus $\mathbf{v}_1 + \mathbf{v}_2 \in \mathbf{N}(A)$. Like the column space, the null space is "closed under addition": adding two vectors in $\mathbf{N}(A)$ yields another vector in $\mathbf{N}(A)$.

2. Similarly, if $\mathbf{v} \in \mathbf{N}(A)$ and $c \in \mathbb{R}$, then since $A\mathbf{v} = \mathbf{0}_m$, we have

$$A(c\mathbf{v}) = c(A\mathbf{v}) = c\mathbf{0}_m = \mathbf{0}_m,$$

so $c\mathbf{v} \in \mathbf{N}(A)$. That is, the null space is "closed under scalar multiplication": multiplying a vector in $\mathbf{N}(A)$ by a real number yields another vector in $\mathbf{N}(A)$.

3. Finally, since $A\mathbf{0}_n = \mathbf{0}_m$, we have $\mathbf{0}_n \in \mathbf{N}(A)$. Thus the null space is never empty, and it contains one of the most important vectors for vector and matrix arithmetic.

Content from Strang's *ILA* 6E. These properties of the null space appear in the very last paragraph of p. 88.

The column space

$$\mathbf{C}(A) = \{A\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^n\}.$$

also behaves well with respect to vector arithmetic.

1. Let $\mathbf{w}_1,\ \mathbf{w}_2 \in \mathbf{C}(A)$. Then there are $\mathbf{v}_1,\ \mathbf{v}_2 \in \mathbb{R}^n$ such that $\mathbf{w}_1 = A\mathbf{v}_1$ and $\mathbf{w}_2 = A\mathbf{v}_2$. So,

$$\mathbf{w}_1 + \mathbf{w}_2 = A\mathbf{v}_1 + A\mathbf{v}_2 = A(\mathbf{v}_1 + \mathbf{v}_2) \in \mathbf{C}(A)$$

since $\mathbf{v}_1 + \mathbf{v}_2 \in \mathbb{R}^n$. That is, $\mathbf{C}(A)$ is "closed under addition": adding two vectors in $\mathbf{C}(A)$ yields another vector in $\mathbf{C}(A)$.

2. Similarly, if $\mathbf{w} \in \mathbf{C}(A)$ with $\mathbf{w} = A\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^n$, and if $c \in \mathbb{R}$, then

$$c\mathbf{w} = c(A\mathbf{v}) = A(c\mathbf{v}) \in \mathbf{C}(A),$$

since $c\mathbf{v} \in \mathbb{R}^n$. That is, $\mathbf{C}(A)$ is "closed under scalar multiplication": multiplying a vector in $\mathbf{C}(A)$ by a real number yields another vector in $\mathbf{C}(A)$.

3. Finally, since $A\mathbf{0}_n = \mathbf{0}_m$, we have $\mathbf{0}_m \in \mathbf{C}(A)$. Thus the column space is never empty, and in particular it contains one of the most important vectors for vector and matrix arithmetic alike.

Sets of vectors that have these properties—closure under vector addition and scalar multiplication and containing the zero vector—are among the most special and useful kinds of sets. They do not just exist and contain things; they are *dynamic* with respect to vector operations. We will see just how special these sets—these *spaces*—are in the context of understanding, and maybe even solving, $A\mathbf{x} = \mathbf{b}$ for A nonsquare.

3.9. Subspaces encode dynamic properties of null and column spaces.

Subsets of $\mathbb{R}^n$ (or $\mathbb{R}^m$) that have these three properties—closure under vector addition, closure under scalar multiplication, presence of the zero vector—are just particularly "nice" for linear algebra. They respect the fundamental arithmetic and algebra that we do, and they arise often in connection with our fundamental problem of solving and understanding and approximating $A\mathbf{x} = \mathbf{b}$. So, they deserve a special name that reflects their dynamism—they are not merely sets but *spaces* of vectors that interact well together.

3.9.1 Definition. *A subset $\mathcal{V}$ of $\mathbb{R}^p$ is a* ***SUBSPACE*** *of $\mathbb{R}^p$ if the following are true.*

(i) [Closure under vector addition] *If $\mathbf{v}$, $\mathbf{w} \in \mathcal{V}$, then $\mathbf{v} + \mathbf{w} \in \mathcal{V}$.*

(ii) [Closure under scalar multiplication] *If $\mathbf{v} \in \mathcal{V}$ and $c \in \mathbb{R}$, then $c\mathbf{v} \in \mathcal{V}$.*

(iii) [Presence of the zero vector] $\mathbf{0}_p \in \mathcal{V}$.

3.9.2 Remark. *We read $\mathbf{0}_p \in \mathcal{V}$ as "The zero vector belongs to $\mathcal{V}$" or "The zero vector is in $\mathcal{V}$." Students often like to write "The zero vector exists in $\mathcal{V}$"; this is awkward, as the zero vector always exists in the ambient space $\mathbb{R}^p$.*

Content from Strang's *ILA* 6E. Page 86 discusses the axioms for a subspace. Examples 1 and 2 on p. 87 present concrete (non)examples of subspaces of $\mathbb{R}^p$.

3.9.3 Example. Let $A \in \mathbb{R}^{m \times n}$.

(i) The column space $\mathbf{C}(A) = \{A\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^n\}$ is a subspace of $\mathbb{R}^m$. We previously saw that $\mathbf{C}(A)$ is a subspace; here we emphasize that $\mathbf{C}(A)$ is a subspace of $\mathbb{R}^m$ because every vector in $\mathbf{C}(A)$ has the form $A\mathbf{v} \in \mathbb{R}^m$.

(ii) The null space $\mathbf{N}(A) = \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \mathbf{0}_m\}$ is a subspace of $\mathbb{R}^n$. We previously saw that $\mathbf{N}(A)$ is a subspace; here we emphasize that $\mathbf{N}(A)$ is a subspace of $\mathbb{R}^n$ because, from its definition, every vector in $\mathbf{N}(A)$ is a vector in $\mathbb{R}^n$. Theorem 3.2.5 also tells us how to write any null space as a column space, and eventually we will see that any vector space can be written as a column space.

3.9.4 Problem (!). Let $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathbb{R}^p$. Prove that $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ is a subspace of $\mathbb{R}^p$. [Hint: *every span is a column space.*]

3.9.5 Example. Let

$$\mathcal{V} = \left\{ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; v_1 + v_2 = 0 \right\}.$$

We can show that $\mathcal{V}$ is a subspace of $\mathbb{R}^3$ in two ways.

(i) The first way is to practice the definition.

• Let $\mathbf{v} = (v_1, v_2, v_3)$, $\mathbf{w} = (w_1, w_2, w_3) \in \mathcal{V}$, so $v_1 + v_2 = w_1 + w_2 = 0$. Then $\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, v_3 + w_3)$. We want $(v_1 + w_1) + (v_2 + w_2) = 0$, and we rearrange $(v_1 + w_1) + (v_2 + w_2) = (v_1 + v_2) + (w_1 + w_2) = 0$. So $\mathbf{v} + \mathbf{w} \in \mathcal{V}$.

• Next, let $\mathbf{v} = (v_1, v_2, v_3) \in \mathcal{V}$, so $v_1 + v_2 = 0$, and $c \in \mathbb{R}$. Then $c\mathbf{v} = (cv_1, cv_2, cv_3)$, and we want $cv_1 + cv_2 = 0$. We factor $cv_1 + cv_2 = c(v_1 + v_2) = 0$, so $c\mathbf{v} \in \mathcal{V}$.

• Last, we want $\mathbf{0}_3 = (0, 0, 0) \in \mathcal{V}$, and we check $0 + 0 = 0$, so, yes, $\mathbf{0}_3 \in \mathcal{V}$.

(ii) Here is the other, sneakier way. If $\mathbf{v} = (v_1, v_2, v_3) \in \mathcal{V}$, then $v_1 + v_2 = 0$, so $v_1 = -v_2$. Then

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} -v_2 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_2 \\ v_3 \end{bmatrix} \in \mathbf{C}\left(\begin{bmatrix} -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

Conversely, if $\mathbf{v}$ is in this column space, then $v_1 + v_2 = 0$, and so $\mathbf{v} \in \mathcal{V}$. (Check that, please.) So $\mathcal{V}$ is a column space and therefore a subspace by the first part of this example.

We will eventually show that every subspace is both a column space and a null space (probably for different matrices). This is a miracle of definitions and algebra: the abstract conditions of the definition of subspace realize themselves concretely in matrices. For the purposes of this course, the only important subspaces that we will study will eventually be column and null spaces. However, there will be times when working with the three axioms for a subspace will be more convenient than representing the subspace as a particular column or null space.

3.9.6 Problem (!). Check that both $\{\mathbf{0}_p\}$ and $\mathbb{R}^p$ are subspaces of $\mathbb{R}^p$. These are "extreme" subspaces: $\{\mathbf{0}_p\}$ is contained in every subspace of $\mathbb{R}^p$, and every subspace of $\mathbb{R}^p$ is contained in $\mathbb{R}^p$ itself.

3.9.7 Problem (!). Let

$$\mathcal{V} = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; x_1,\ x_2 \in \mathbb{R} \right\}.$$

Explain how each of the three conditions for a subspace fails for $\mathcal{V}$.

3.9.8 Problem (!).

(i) Let $A \in \mathbb{R}^{m \times n}$ with $m \neq n$. Do $\mathbf{N}(A)$ and $\mathbf{C}(A)$ have any vectors in common?

(ii) Let $A \in \mathbb{R}^{m \times m}$. What one vector do $\mathbf{N}(A)$ and $\mathbf{C}(A)$ definitely have in common?

(iii) By considering the matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

show that perhaps $\mathbf{N}(A)$ and $\mathbf{C}(A)$ have only that one vector in common.

Content from Strang's *ILA* 6E. Section 3.1 discusses the much more general, and hugely important, concept of a VECTOR SPACE. This is a set of elements called VECTORS that we can add together and multiply by scalars (real or complex numbers), and for which these operations of VECTOR ADDITION and SCALAR MULTIPLICATION basically behave the way that we expect arithmetic to behave. See the eight axioms on p. 89.

Maybe the two most important vector spaces are the column vectors with n entries, which, of course, is $\mathbb{R}^n$, and, from calculus, the space of continuous functions on an interval $I \subseteq \mathbb{R}$, which we denote by $\mathcal{C}(I)$. You know from calculus that if f and g are continuous on I, then so are $f+g$ and cf for any real c. (The space $\mathcal{C}(I)$ has the additional algebraic operation of function multiplication, fg, whereas we cannot multiply vectors in $\mathbb{R}^n$ in any "natural" way to get another vector in $\mathbb{R}^n$.) The r-times continuously differentiable functions (functions whose first r derivatives exist and are continuous) form the subspace $\mathcal{C}^r(I)$ of $\mathcal{C}(I)$, which is a natural player in differential equations.

The structure of vector spaces transcends matrix problems and provide the "right" framework for understanding the linear structure that pervades calculus. See pp. 84–85 for just a little on this. We will focus mostly on subspaces of $\mathbb{R}^n$, not general vector spaces, in this course.

3.10. Bases are the ideal coordinate systems for subspaces.

Content from Strang's *ILA* 6E. You should be very comfortable with the notion of independence by now. Read pp. 115–117 thoroughly. Think carefully about the "guilty" remark at the end of p. 117. Which way of saying that the columns of $A \in \mathbb{R}^{m \times n}$ are independent feels easier to you—that $A\mathbf{x} = \mathbf{0}_m$ forces $\mathbf{x} = \mathbf{0}_n$ (the "democratic" way) or that one column of A is a combination of other columns (the "guilty" way)?

You should also be very comfortable with the notion of span by now. Read "Vectors that Span a Subspace" at the start of p. 118.

3.10.1 Example. Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$

We know from long experience going back to Example 1.14.7 that columns 2 and 4 are pivot columns. Consequently, the list $\mathbf{a}_2$, $\mathbf{a}_4$ is independent and $\mathbf{C}(A) = \operatorname{span}(\mathbf{a}_2, \mathbf{a}_4)$. We also saw in Example 3.4.1 that

$$\mathbf{N}(A) = \mathbf{C}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & -4 \\ 0 & 0 & 1 \end{bmatrix}\right).$$

The columns of this matrix giving $\mathbf{N}(A)$ are also independent, for if

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ -3 \\ 0 \\ -4 \\ 1 \end{bmatrix} = \mathbf{0}_5,$$

then

$$\begin{bmatrix} c_1 \\ * \\ c_2 \\ * \\ c_3 \end{bmatrix} = \mathbf{0}_5,$$

thus $c_1 = c_2 = c_3 = 0$. What matters here are balance of 0's and 1's in the three vectors above, not the other entries.

While $\mathbf{C}(A)$ and $\mathbf{N}(A)$ are very different spaces, we have described them in the same way: as column spaces of matrices with independent columns, as spans of lists of independent vectors.

3.10.2 Problem (!). Let A be as in the previous example. Example 3.4.1 showed that $\mathbf{b} = (b_1, b_2, b_3) \in \mathbf{C}(A)$ if and only if $b_2 = 2b_1$. Use this fact to show that

$$\mathbf{C}(A) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right).$$

Check that the vectors in this span are also independent but only one is a column of A. The new result here is that we have written $\mathbf{C}(A)$ as a span of independent vectors, not all of which were columns of A. But the old result is that we still only needed two vectors to do it. This is probably not surprising, since $\operatorname{rank}(A) = 2$.

Here is the pattern that we have seen throughout this course: writing column spaces and null spaces as spans of independent vectors is an efficient way of describing them. It turns out that we can always do this, which should not surprise us. An independent list of vectors that spans a subspace has a special name.

3.10.3 Definition. *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. A list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathcal{V}$ is a* ***BASIS*** *for $\mathcal{V}$ if the following hold.*

(i) *The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are independent.*

(ii) $\mathcal{V} = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$.

3.10.4 Example.

(i) The standard basis vectors for $\mathbb{R}^n$ are a basis for $\mathbb{R}^n$. (If they were not, it would be a pretty awful use of the word "basis.") Here is the proof for $n = 3$, which should feel familiar. If $\mathbf{v} \in \mathbb{R}^3$, then

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3.$$

Thus $\mathbb{R}^3 = \operatorname{span}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$. If $x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 = \mathbf{0}_3$, then with $\mathbf{x} = (x_1, x_2, x_3)$, we have $\mathbf{x} = \mathbf{0}_3$, thus $x_1 = x_2 = x_3 = 0$. This is the linear independence of $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$.

(ii) The pivot columns of a matrix are a basis for that matrix's column space. This is unsurprising: the pivot columns are independent and they span that matrix's column space.

(iii) The columns of an invertible matrix $A \in \mathbb{R}^{m \times m}$ are a basis for $\mathbb{R}^m$. This is also unsurprising: these columns are independent and $\mathbf{C}(A) = \mathbb{R}^m$.

(iv) Let $A \in \mathbb{R}^{m \times n}$ with $r := \operatorname{rank}(A)$ satisfying $1 \le r < n$. If the CR-factorization of A is $A = C \begin{bmatrix} I_r & F \end{bmatrix} P$, then Theorem 3.2.5 gives

$$\mathbf{N}(A) = \mathbf{C}\left(P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right).$$

If we can show that the columns of this matrix

$$P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}$$

are independent, then they will be a basis for $\mathbf{N}(A)$, since they already span $\mathbf{N}(A)$.

So, suppose that

$$P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{v} = \mathbf{0}_n$$

for some $\mathbf{v} \in \mathbb{R}^{n-r}$. The goal is to show that $\mathbf{v} = \mathbf{0}_{n-r}$. Since P^{-1} is invertible, we immediately have

$$\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{v} = \mathbf{0}_n,$$

Doing the block matrix multiplication, we obtain

$$\begin{bmatrix} -F\mathbf{v} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_r \\ \mathbf{0}_{n-r} \end{bmatrix},$$

thus $\mathbf{v} = \mathbf{0}_{n-r}$.

3.10.5 Problem (!). Explain why the subspace $\{\mathbf{0}_p\}$ of $\mathbb{R}^p$ cannot have a basis according to our definition of basis. This is not as serious an issue as it might sound; some people define the "empty list" to be a basis for $\{\mathbf{0}_p\}$, with the idea that a linear combination of no vectors is defined to be $\mathbf{0}_p$.

Content from Strang's *ILA* 6E. Read all of "A Basis for a Vector Space" on pp. 118–119. Every single thing here is important. Then read Worked Example 3.4 A on p. 122. This is a very important example that you should know how to prove.

3.10.6 Problem (!). Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. Prove that $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathcal{V}$ are a basis for $\mathcal{V}$ if and only if the following are both true.

(i) The matrix $\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{bmatrix} \in \mathbb{R}^{p \times r}$ has full column rank.

(ii) $\mathcal{V} = \mathbf{C}\left(\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{bmatrix} \right)$.

We made the definition of a basis as a consequence of our observations that the most efficient way to describe subspaces was as spans of lists of linearly independent vectors. We can be a little more dynamic. A basis is fundamentally a "coordinate system" for a subspace: we can "reach" every vector in the subspace in a unique way "via" the basis. We know this in our hearts with the standard basis $\mathbf{e}_1$, $\mathbf{e}_2$ for $\mathbb{R}^2$: any $\mathbf{x} = (x, y) \in \mathbb{R}^2$ has the unique form $\mathbf{x} = x\mathbf{e}_1 + y\mathbf{e}_2$.

3.10.7 Problem (!). Which of the two conditions in Definition 3.10.3 do you think encodes this "reaching" property and which encodes the uniqueness?

Here is the rigorous proof.

3.10.8 Theorem. *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. A list $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathcal{V}$ is a basis for $\mathcal{V}$ if and only if for any $\mathbf{v} \in \mathcal{V}$, there are unique $c_1, \ldots, c_r \in \mathbb{R}$ such that*

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_d\mathbf{v}_r.$$

Proof. ($\Longrightarrow$) Let $\mathbf{v} \in \mathcal{V}$. Then $\mathbf{v} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$, so there are $c_1, \ldots, c_r \in \mathbb{R}$ such that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r$. What if there is another way to write $\mathbf{v}$ as the span of $\mathbf{v}_1, \ldots, \mathbf{v}_r$?

Then $\mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_r\mathbf{v}_r$, too. Do a little arithmetic and algebra:

$$\mathbf{0}_p = \mathbf{v} - \mathbf{v} = (c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r) - (d_1\mathbf{v}_1 + \cdots + d_r\mathbf{v}_r) = (c_1 - d_1)\mathbf{v}_1 + \cdots + (c_r - d_r)\mathbf{v}_r.$$

The independence of $\mathbf{v}_1, \ldots, \mathbf{v}_r$ means $c_1 - d_1 = 0, \ldots, c_r - d_r = 0$, so $c_j = d_j$ for all j. So, there is only *one* way to write $\mathbf{v}$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_r$.

($\Longleftarrow$) We already know $\mathcal{V} = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$. To check independence, suppose $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = \mathbf{0}_p$. We already know that we can get $\mathbf{0}_p$ in the span by writing $\mathbf{0}_p = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_r$. So we have

$$c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_r,$$

and since we are assuming that coefficients are unique, this gives $c_j = 0$ for all j. ■

A basis for a subspace cannot be much use as a coordinate system or an ideally efficient representation mechanism if the basis does not exist in the first place! We know this to be true for the most important subspaces, column spaces and null spaces, but it turns out that the three subspace axioms alone guarantee the existence of a basis. To prove that, we need an unsurprising auxiliary result.

3.10.9 Lemma. *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. If $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathcal{V}$ and $c_1, \ldots, c_n \in \mathbb{R}$, then $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \in \mathcal{V}$. That is, $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is contained in $\mathcal{V}$.*

Proof. This is really an induction argument on n. Here is the proof for $n = 3$. Since $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3 \in \mathcal{V}$, the subspace axioms guarantee $c_1\mathbf{v}_1$, $c_2\mathbf{v}_2$, $c_3\mathbf{v}_3 \in \mathcal{V}$. Then the axioms guarantee $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in \mathcal{V}$, and so $(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) + c_3\mathbf{v}_3 \in \mathcal{V}$. ■

3.10.10 Theorem. *Let $\mathcal{V} \neq \{\mathbf{0}_p\}$ be a subspace of $\mathbb{R}^p$. Then $\mathcal{V}$ has a basis.*

Proof. This is a proof by exhaustion, which means that we exhaust all possible cases and also ourselves. Since $\mathcal{V} \neq \{\mathbf{0}_p\}$, there is $\mathbf{v}_1 \in \mathcal{V}$ such that $\mathbf{v}_1 \neq \mathbf{0}_p$. One of two things is true: either $\mathcal{V} = \operatorname{span}(\mathbf{v}_1)$ or $\mathcal{V} \neq \operatorname{span}(\mathbf{v}_1)$. In the first case, $\mathbf{v}_1$ by itself is a basis for $\mathcal{V}$ since it's nonzero and therefore independent and also spans $\mathcal{V}$.

In the second case, there is $\mathbf{v}_2 \in \mathcal{V}$ such that $\mathbf{v}_2 \notin \operatorname{span}(\mathbf{v}_1)$. Since $\mathbf{v}_1 \neq \mathbf{0}_p$, the list $\mathbf{v}_1$, $\mathbf{v}_2$ is independent. Once again, one of two things is true: either $\mathcal{V} = \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2)$ or $\mathcal{V} \neq \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2)$. In the first case, $\mathbf{v}_1$, $\mathbf{v}_2$ is a basis for $\mathcal{V}$. In the second case, there is $\mathbf{v}_3 \in \mathcal{V}$ such that $\mathbf{v}_3 \notin \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2)$. Turn the crank and keep going.

Assuming $\mathcal{V} \neq \operatorname{span}(\mathbf{v}_1)$, one of two things has to happen in the end. First, we could have a list $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathcal{V}$ such that $\mathcal{V} = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$, $\mathbf{v}_1 \neq \mathbf{0}_p$ and $\mathbf{v}_j \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1})$ for $j = 2, \ldots, r$. This list is therefore independent and so is a basis for $\mathcal{V}$.

The other possibility is that we have turned the crank far enough to have an independent list $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathcal{V}$ of p (necessarily distinct!) vectors. Put $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_p \end{bmatrix}$, so $A \in \mathbb{R}^{p \times p}$ has independent columns and therefore is invertible. Thus $\mathbf{C}(A) = \mathbb{R}^p$. This means $\mathcal{V} = \mathbb{R}^p$: if $\mathbf{b} \in \mathbb{R}^p$, then $\mathbf{b} \in \mathbf{C}(A)$ since $\mathbf{C}(A) = \mathbb{R}^p$. And Lemma 3.10.9 says that if $\mathbf{b} \in \mathbf{C}(A)$, then $\mathbf{b} \in \mathcal{V}$. ■

3.10.11 Problem (!). This problem revisits the proof of Theorem 3.10.10.

(i) Rewrite the proof in the specific case of $p = 4$. In place of the vague sentence "Turn the crank and keep going," continue going explicitly until you reach "The other possibility" with $p = 4$.

(ii) Reread the proof and identify exactly where the assumption that $\mathcal{V}$ was a subspace was used.

3.11. Dimension quantifies the true size of a subspace.

The point of a basis is efficient representation. Theorem 3.10.8 gives us part of that efficiency: there is only one way to represent vectors with respect to a basis. And now we know there is always a basis. One more big thing remains: the amount of data in a basis is effectively always the same in that any basis contains the same number of vectors. (Deeper question: is there a "best" basis for a subspace? What more could we want? Think about it...)

3.11.1 Theorem. *Let $\mathcal{V} \neq \{\mathbf{0}_p\}$ be a subspace of $\mathbb{R}^p$. All bases for $\mathcal{V}$ have the same length.*

Proof. We know that at least one basis for $\mathcal{V}$ exists by Theorem 3.10.10. Call that basis $\mathbf{v}_1, \ldots, \mathbf{v}_r$ for some $r \geq 1$. Put $A := \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{bmatrix} \in \mathbb{R}^{p \times r}$. So $\mathcal{V} = \mathbf{C}(A)$.

The columns of A are independent, so $\operatorname{rank}(A) = r$. Now let the list $\mathbf{w}_1, \ldots, \mathbf{w}_s$ also be a basis for $\mathcal{V}$. We want to show $s = r$. This list $\mathbf{w}_1, \ldots, \mathbf{w}_s$ is therefore independent in $\mathbf{C}(A)$, so it cannot be longer than r by Theorem 3.6.1. Thus $s \leq r$. This list $\mathbf{w}_1, \ldots, \mathbf{w}_s$ also spans $\mathbf{C}(A)$, so it has to be at least as long as $\operatorname{rank}(A)$ by Theorem 3.6.2. Thus $s \geq r$. Since both $s \leq r$ and $s \geq r$, we have $s = r$. ■

Content from Strang's *ILA* 6E. Page 120 proves Theorem 3.11.1. Read the paragraphs after the boxed "Definition." Then read Worked Example 3.4 B on p. 122.

Since any (nontrivial) subspace has a basis, and any basis for a (nontrivial) subspace has the same length, it's fair to give a name to that length.

3.11.2 Definition. *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. If $\mathcal{V} \neq \{\mathbf{0}_p\}$, then the* **DIMENSION** *of $\mathcal{V}$, denoted $\dim(\mathcal{V})$, is the length of any basis for that subspace. We define $\dim(\{\mathbf{0}_p\}) := 0$.*

3.11.3 Example. **(i)** Since the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ for $\mathbb{R}^n$ contains n vectors, $\dim(\mathbb{R}^n) = n$.

(ii) Let $A \in \mathbb{R}^{m \times n}$ have rank $r \geq 1$. Then $\dim(\mathbf{C}(A)) = r$, since A has r pivot columns, the pivot columns are independent, and the pivot columns span $\mathbf{C}(A)$. *That is, rank is the dimension of the column space.*

There are a lot of useful properties of bases and dimension that are basically analogues of things that we figured out for column spaces relative to rank.

3.11.4 Problem ($\star$). Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$ with dimension $\dim(\mathcal{V}) = r \geq 1$.

(i) Show that $r \leq p$. [Hint: *if $r > p$, what contradicts Problem 3.5.6?*]

(ii) Explain why any list of more than r vectors in $\mathcal{V}$ is dependent. [Hint: *write $\mathcal{V}$ as the column space of a matrix of rank r. Then use Theorem 3.6.1.*]

(iii) Explain why any list of r independent vectors in $\mathcal{V}$ is a basis for $\mathcal{V}$. [Hint: *write $\mathcal{V}$ as the column space of a matrix of rank r. Then use Theorem 3.6.2.*]

(iv) Explain why any list of r vectors in $\mathcal{V}$ that spans $\mathcal{V}$ is a basis for $\mathcal{V}$. [Hint: *if the list is dependent, apply Lemma 1.13.6. What contradiction results?*]

3.11.5 Problem ($\star$). It is useful to be able to compare dimensions of related subspaces.

(i) Suppose that $\mathcal{V}$ and $\mathcal{W}$ are subspaces of $\mathbb{R}^p$ such that every vector in $\mathcal{V}$ is also in $\mathcal{W}$. (We might think that $\mathcal{V}$ is a subspace of the subspace $\mathcal{W}$!) Prove that $\dim(\mathcal{V}) \leq \dim(\mathcal{W})$. [Hint: *if $\dim(\mathcal{W}) = r \leq p$, what goes wrong, per part (ii) of Problem 3.11.4 if there is a list of more than r independent vectors in $\mathcal{V}$?*]

(ii) In the previous part, if $\dim(\mathcal{V}) = \dim(\mathcal{W})$, explain why $\mathcal{V} = \mathcal{W}$. [Hint: *use part (iii) of Problem 3.11.4 to explain why any basis for $\mathcal{V}$ is then a basis for $\mathcal{W}$.*]

3.11.6 Problem ($\star$). It is also useful to see how a subspace does, or does not, "change" under the "action" of a matrix. Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$ with $\dim(\mathcal{V}) \geq 1$. Let $A \in \mathbb{R}^{p \times p}$ be invertible and let

$$\mathcal{W} = \{A\mathbf{v} \mid \mathbf{v} \in \mathcal{V}\}.$$

We can think of $\mathcal{W}$ as the part of $\mathbf{C}(A)$ where the inputs (the $\mathbf{v}$) are "coming from" $\mathcal{V}$.

(i) Prove that $\mathcal{W}$ is a subspace of $\mathbb{R}^p$. [Hint: *modify the proof that $\mathbf{C}(A)$ is a subspace of $\mathbb{R}^p$.*]

(ii) Prove that $\dim(\mathcal{W}) = \dim(\mathcal{V})$. [Hint: *let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be a basis for $\mathcal{V}$. Why is $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ a basis for $\mathcal{W}$?*]

3.11.7 Problem ($\star$). And it is useful to be able to compare ranks of related matrices. Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$.

(i) Prove that $\text{rank}(AB) \leq \text{rank}(A)$. [Hint: *Problems 2.2.7 and 3.11.5.*]

(ii) Prove that if B is invertible (so here we're assuming $n = p$), then $\mathbf{C}(A) = \mathbf{C}(AB)$ and so $\text{rank}(AB) = \text{rank}(A)$. [Hint: *this is one of the rare times when we need to work at both parts of Definition 1.8.4. For one part, use Problem 2.9.14; for the other, consider* *the trick of rewriting* $A\mathbf{v} = (AB)(B^{-1}\mathbf{v})$.]

(iii) Give an example to show that if B is not invertible, then we may have $\operatorname{rank}(AB) < \operatorname{rank}(A)$.

(iv) Prove that $\operatorname{rank}(AB) \le \operatorname{rank}(B)$. [Hint: *let* $r := \operatorname{rank}(B)$. *If* $r = 0$, *then* B *is the zero matrix. If* $r \ge 1$, *suppose that* $\operatorname{rank}(AB) > \operatorname{rank}(B)$ *and find vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_{r+1}$ *that are columns of* B *with* $A\mathbf{v}_1, \ldots, A\mathbf{v}_{r+1}$ *independent. Find something that goes wrong if* $\mathbf{v}_1, \ldots, \mathbf{v}_{r+1}$ *is dependent, which is what you expect since this is a list of columns of* B *longer than* $\operatorname{rank}(B)$.]

(v) Prove that if A is invertible (so here $m = n$), then $\operatorname{rank}(AB) = \operatorname{rank}(B)$. [Hint: *apply Problem 3.11.6.*]

(vi) Give an example to show that if A is not invertible, then we may have $\operatorname{rank}(AB) < \operatorname{rank}(B)$.

Content from Strang's *ILA* 6E. Look at the matrix R_0 on p. 130. Then read #2 and #3 on pp. 130–131 on the dimensions of its column and null spaces. Next, look at the matrix A at the bottom of p. 131 and read about its column and null spaces in #2 and #3 on pp. 132–133. We'll come back to the row space and left null space shortly. Read Worked Example 3.5 B on p. 137.

3.11.8 Problem (!). You might wonder why, in considering the "size" of a subspace, we never just try to count its elements. Explain why if $\mathcal{V} \ne \{\mathbf{0}_p\}$ is a subspace of $\mathbb{R}^p$, then $\mathcal{V}$ contains infinitely many vectors. [Hint: *explain why there is* $\mathbf{v} \in \mathcal{V}$ *with* $\mathbf{v} \ne \mathbf{0}_p$ *and then* $c\mathbf{v} \in \mathcal{V}$ *for all* $c \in \mathbb{R}$.]

3.12. Rank–nullity quantifies existence and uniqueness.

Now we can answer our lingering questions on how to quantify and qualify failure for the problem $A\mathbf{x} = \mathbf{b}$.

3.12.1 Theorem (Rank–nullity). *Let* $A \in \mathbb{R}^{m \times n}$. *Then*

$$\operatorname{rank}(A) + \dim(\mathbf{N}(A)) = n.$$

Proof. Let $r := \operatorname{rank}(A)$.

1. $r = 0$. Then A is the zero matrix, so $A\mathbf{x} = \mathbf{0}_m$ for all $\mathbf{x} \in \mathbb{R}^n$. Thus $\mathbf{N}(A) = \mathbb{R}^n$, so $\dim(\mathbf{N}(A)) = n$, and we have

$$\operatorname{rank}(A) + \dim(\mathbf{N}(A)) = 0 + n = n.$$

2. $r = n$. Then the columns of A are independent, so $\mathbf{N}(A) = \{\mathbf{0}_n\}$, and therefore

$\dim(\mathbf{N}(A)) = 0$. Thus

$$\operatorname{rank}(A) + \dim(\mathbf{N}(A)) = n + 0 = n.$$

3. $1 \le r < n$. From the CR-factorization of A, we can write

$$\mathbf{N}(A) = \mathbf{C}\left(P^{-1}\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right).$$

Part (iv) of Example 3.10.4 showed that the columns of

$$P^{-1}\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}$$

are a basis for $\mathbf{N}(A)$, and there are $n - r$ such columns. Hence $\dim(\mathbf{N}(A)) = n - r$, and so

$$\operatorname{rank}(A) + \dim(\mathbf{N}(A)) = r + (n - r) = n. \qquad \blacksquare$$

The point is that if you know one of the dimensions $\operatorname{rank}(A) = \dim(\mathbf{C}(A))$ or $\dim(\mathbf{N}(A))$, then we know the other. It should be interesting, and maybe a bit weird, to realize that even though $\mathbf{C}(A)$ is not a subspace of $\mathbb{R}^n$ (it is a subspace of $\mathbb{R}^m$), its dimension still talks to the dimension of $\mathbf{N}(A)$ (which is a subspace of $\mathbb{R}^n$) and the dimension of $\mathbb{R}^n$ itself.

3.12.2 Example. Let

$$A = \begin{bmatrix} 1 & 2 & 1 & 7 & 0 \\ 2 & 4 & 2 & 14 & 0 \\ 0 & 0 & 2 & 8 & 0 \end{bmatrix}.$$

We know that A has two pivot columns, so $\operatorname{rank}(A) = 2$, and Example 3.4.1 gave

$$\mathbf{N}(A) = \mathbf{C}\left(\begin{bmatrix} -2 & -3 & 0 \\ 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right).$$

We showed in Example 3.10.1 that the columns of the matrix on the right are independent, so $\dim(\mathbf{N}(A)) = 3$. As rank–nullity promises,

$$\operatorname{rank}(A) + \dim(\mathbf{N}(A)) = 2 + 3 = 5.$$

Here are some ways to interpret rank–nullity and the quantitative (and morally qualitative) extents to which solutions to $A\mathbf{x} = \mathbf{b}$ exist and are unique.

• If $A \in \mathbb{R}^{m \times n}$ with $m \neq n$ (so A is not square), then either existence or uniqueness of solutions fails. Of course, either or both could fail even when A is square! But there is just a little more hope to avoid failure in the square case. Rank–nullity allows us to quantify this failure.

• First, if existence fails and $\mathbf{C}(A) \neq \mathbb{R}^m$, then $\dim(\mathbf{C}(A))$ is not as "large" as it could be. If $\dim(\mathbf{C}(A))$ is "too small," then $\dim(\mathbf{N}(A))$ will have to be "large enough' to make $\dim(\mathbf{N}(A)) + \dim(\mathbf{C}(A)) = n$ true. This means that if existence fails "sufficiently much," then uniqueness also fails with a "certain large degree of freedom."

• Second, if uniqueness fails and $\mathbf{N}(A) \neq \{\mathbf{0}_n\}$, then $\dim(\mathbf{N}(A))$ is not as "small" as it could be. If $\dim(\mathbf{N}(A))$ is "too big," then $\dim(\mathbf{C}(A))$ will have to be "small enough" to make, again, $\dim(\mathbf{N}(A)) + \dim(\mathbf{C}(A)) = n$ true. This means that if uniqueness fails with a "certain large degree of freedom," then uniqueness also fails "sufficiently much."

3.12.3 Problem ($\star$). Let $A \in \mathbb{R}^{m\times m}$. In this problem we will assume that the rank–nullity theorem is true (it is, by the work above), but that we do not know that some of our prior results are true.

(i) Use rank–nullity to prove that $\mathbf{N}(A) = \{\mathbf{0}_m\}$ if and only if $\mathbf{C}(A) = \mathbb{R}^m$. This gives another proof of some parts of the invertible matrix theorem.

(ii) Use rank–nullity to give another proof of the result from Theorem 3.5.3 that if A has full row rank, then $\mathbf{C}(A) = \mathbb{R}^m$. [Hint: *if* $\operatorname{rank}(A) = m$, *what do you know about* $\mathbf{N}(A^\mathsf{T})$ *from rank–nullity applied to* A^T*? Use this in conjunction with* $\mathbb{R}^m = \mathbf{C}(A) \oplus \mathbf{N}(A^\mathsf{T})$.]

4. Orthogonality

4.1. Introduction: how do subspaces fit together?

The notion of dimension allows us to quantify how much of $\mathbb{R}^m$ the column space of $A \in \mathbb{R}^{m \times n}$ misses: it misses $m - \dim[\mathbf{C}(A)]$ "dimensions" of $\mathbb{R}^m$. But what is going on elsewhere in $\mathbb{R}^m$ beyond $\mathbf{C}(A)$? How does $\mathbf{C}(A)$ "fit into" or "subsist within" the whole of $\mathbb{R}^m$? Often $\mathbf{C}(A) \neq \mathbb{R}^m$, and so we cannot always solve the fundamental problem $A\mathbf{x} = \mathbf{b}$; here the deeper question is not just *how much* of $\mathbb{R}^m$ does $\mathbf{C}(A)$ miss but rather *what exactly* in $\mathbb{R}^m$ does $\mathbf{C}(A)$ miss. Are there other meaningful, useful ways to characterize and describe $\mathbf{C}(A)$ beyond just the definition?

We could also ask about $\mathbb{R}^n$ and $\mathbf{N}(A)$. When $\mathbf{v} \in \mathbf{N}(A)$, we have $A\mathbf{v} = \mathbf{0}_m$; the action of A on such $\mathbf{v}$ is therefore pretty simple. For what $\mathbf{w} \in \mathbb{R}^n$ does A "act nontrivially" with $A\mathbf{w} \neq \mathbf{0}_m$? On what parts of $\mathbb{R}^n$ is A "more interesting"? How does $\mathbf{N}(A)$ "fit into" the whole of $\mathbb{R}^n$?

It turns out that these questions are "dual" to each other in that if we know how to handle one of them, we can understand the other pretty quickly.

4.2. A toy problem will show how they fit.

At the start of the course, our focus was on the column space. We prioritized existence of solutions to $A\mathbf{x} = \mathbf{b}$ first, and then we thought about uniqueness. This is natural—talking about the uniqueness of something that may not exist in the first place is unnatural. Conversely, we eased our way into the RREF by thinking about the null space first; solving $A\mathbf{x} = \mathbf{0}$ was (probably) easier than tackling the full problem $A\mathbf{x} = \mathbf{b}$. We will take a morally similar approach here and ask first how the null space of $A \in \mathbb{R}^{m \times n}$ fits into $\mathbb{R}^n$; many results for the column space will then follow with pleasantly little effort.

4.2.1 Example. Let

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then $\mathbf{N}(A) = \operatorname{span}(\mathbf{e}_1, \mathbf{e}_2)$. Hopefully what is missing from $\mathbf{N}(A)$ is obvious: $\mathbf{e}_3 \in \mathbb{R}^3$.

More precisely, any $\mathbf{x} \in \mathbb{R}^3$ can be written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix}.$$

That is, any $\mathbf{x} \in \mathbb{R}^3$ can be written (or, more evocatively, "decomposed") in the form $\mathbf{x} = \mathbf{v} + \mathbf{w}$ for $\mathbf{v} \in \mathbf{N}(A)$ and $\mathbf{w} \in \operatorname{span}(\mathbf{e}_3)$. And there is only one way to choose these $\mathbf{v}$ and $\mathbf{w}$.

Put $\mathcal{W} := \operatorname{span}(\mathbf{e}_3)$. Can we see $\mathcal{W}$ directly from A itself, without passing to the null space? Certainly: $\mathbf{e}_3$ appears in the first row of A. We are more used to thinking about columns than rows, so we flip every row of A to a column and every column of A to a row by taking the **TRANSPOSE**:

$$A^{\mathsf{T}} := \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then we see

$$\mathcal{W} = \mathbf{C}(A^{\mathsf{T}}).$$

And so we have written—"decomposed"—any $\mathbf{x} \in \mathbb{R}^3$ uniquely as a sum of the form $\mathbf{x} = \mathbf{v} + \mathbf{w}$ with $\mathbf{v} \in \mathbf{N}(A)$ and $\mathbf{w} \in \mathbf{C}(A^{\mathsf{T}})$.

The vectors $\mathbf{v}$ and $\mathbf{w}$ "talk" to each in a special way. Here, $\mathbf{v} = (x_1, x_2, 0)$ and $\mathbf{w} = (0, 0, x_3)$ for some x_1, x_2, $x_3 \in \mathbb{R}$. We compute

$$\mathbf{v} \cdot \mathbf{w} = x_1(0) + x_2(0) + 0(x_3) = 0.$$

We will view this dot product relation as encoding our physical intuition that $\mathbf{N}(A)$, which is effectively the xy-plane in three-dimensional space, is perpendicular to $\mathbf{C}(A^{\mathsf{T}})$, which is effectively the z-axis in three-dimensional space.

Did we just get really lucky, since the bases for $\mathbf{N}(A)$ and $\mathbf{C}(A^{\mathsf{T}})$ in the previous example just involved the coordinate axes for three-dimensional space? Here is what might be possible.

4.2.2 Conjecture. *Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$. There exist unique $\mathbf{v} \in \mathbf{N}(A)$ and $\mathbf{w} \in \mathbf{C}(A^{\mathsf{T}})$ such that $\mathbf{x} = \mathbf{v} + \mathbf{w}$ and $\mathbf{v} \cdot \mathbf{w} = 0$. (More precisely, by "unique" we mean that if we can write $\mathbf{x} \in \mathbb{R}^n$ as both $\mathbf{x} = \mathbf{v} + \mathbf{w}$ and $\mathbf{x} = \widetilde{\mathbf{v}} + \widetilde{\mathbf{w}}$ for some $\mathbf{v}$, $\widetilde{\mathbf{v}} \in \mathbf{N}(A)$ and $\mathbf{w}$, $\widetilde{\mathbf{w}} \in \mathbf{C}(A^{\mathsf{T}})$, then $\mathbf{v} = \widetilde{\mathbf{v}}$ and $\mathbf{w} = \widetilde{\mathbf{w}}$.)*

4.3. Orthogonality and transposes are friends.

Here is the fundamental interaction between vectors that dot products encode.

4.3.1 Definition. *The vectors $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ are* ***ORTHOGONAL*** *if $\mathbf{v} \cdot \mathbf{w} = 0$.*

4.3.2 Problem (!). Let x, $y \in \mathbb{R}$ be nonzero and let $\mathbf{v} = (x, y)$ and $\mathbf{w} = (-y, x)$. Check that $\mathbf{v} \cdot \mathbf{w} = 0$. Then draw $\mathbf{v}$ and $\mathbf{w}$ in the xy-plane (here you will have to pick some specific nonzero values for x and y) and convince yourself that $\mathbf{v}$ and $\mathbf{w}$ are perpendicular.

We will use the dot product in more calculations than we have before. It has many algebraic properties that resemble the familiar algebraic properties of multiplication of real numbers. One moral difference: the product of real numbers is a real number, while the dot product of vectors is a real number, not a vector.

4.3.3 Problem ($\star$). Prove the following. All vectors below are in the same space, e.g., $\mathbb{R}^p$. (If it makes things more concrete for you, do it for $p = 3$.)

(i) [Commutativity] $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$.

(ii) [Distributivity] $\mathbf{v} \cdot (\mathbf{w}_1 + \mathbf{w}_2) = (\mathbf{v} \cdot \mathbf{w}_1) + (\mathbf{v} \cdot \mathbf{w}_2)$.

(iii) [Factoring] $\mathbf{v} \cdot (c\mathbf{w}) = c(\mathbf{v} \cdot \mathbf{w})$.

The properties above are equalities. Here is an inequality: if $\mathbf{v} = (v_1, \ldots, v_p) \in \mathbb{R}^p$, then

$$\mathbf{v} \cdot \mathbf{v} = v_1^2 + \cdots + v_p^2 \geq 0.$$

Suppose that $\mathbf{v} \cdot \mathbf{v} = 0$. Then

$$0 \leq v_1^2 \leq v_1^2 + \cdots + v_p^2 = \mathbf{v} \cdot \mathbf{v} = 0,$$

and so $v_1^2 = 0$. Then $v_1 = 0$. Replace v_1 with v_j for $j = 1, \ldots, n$ to conclude $v_j = 0$ for all j, thus $\mathbf{v} = \mathbf{0}_p$. This is important enough to record by itself.

4.3.4 Theorem. *Let $\mathbf{v} \in \mathbb{R}^p$. Then $\mathbf{v} \cdot \mathbf{v} \geq 0$ and $\mathbf{v} \cdot \mathbf{v} = 0$ only when $\mathbf{v} = \mathbf{0}_p$.*

We have learned several major operations involving matrices, chiefly matrix multiplication (when matrix-matrix products are defined, which is not always) and matrix inversion (when the matrix under consideration is invertible, which is not always). The key matrix operation that interacts best with the dot product and orthogonality is the transpose.

4.3.5 Definition (What the transpose is). *The* **TRANSPOSE** *of $A \in \mathbb{R}^{m \times n}$ is the matrix $A^{\mathsf{T}} \in \mathbb{R}^{n \times m}$ such that the (i, j)-entry of A^{T} is the (j, i)-entry of A. We write*

$$A_{ij}^{\mathsf{T}} = A_{ji}.$$

4.3.6 Remark. *Recall that we write $\mathbf{v} \in \mathbb{R}^n$ as*

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = (v_1, \ldots, v_n).$$

If we think of $\mathbf{v}$ as an $n \times 1$ matrix, which is sometimes convenient, then

$$\mathbf{v}^{\mathsf{T}} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix},$$

but we never write $\mathbf{v}^{\mathsf{T}} = (v_1, \ldots, v_n)$. Parentheses and square brackets are different kinds of notation!

4.3.7 Example. If

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$

then

$$A^{\mathsf{T}} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

4.3.8 Problem (⋆). Suppose that $P \in \mathbb{R}^{m \times m}$ is a permutation matrix. Prove that P^{T} is also a permutation matrix. [Hint: *what are the columns of* P^{T}*?*]

Now we can generalize how $\mathbf{N}(A)$ and $\mathbf{C}(A^{\mathsf{T}})$ interact in general.

4.3.9 Theorem. *Let* $A \in \mathbb{R}^{m \times n}$*,* $\mathbf{v} \in \mathbf{N}(A)$*, and* $\mathbf{w} \in \mathbf{C}(A^{\mathsf{T}})$*. Then* $\mathbf{v} \cdot \mathbf{w} = 0$*.*

Proof. If $\mathbf{v} \in \mathbf{N}(A)$, then $A\mathbf{v} = \mathbf{0}_m$. One way to compute $A\mathbf{v}$ is by taking dot products of $\mathbf{v}$ with the rows of A viewed as columns in $\mathbb{R}^n$. That $A\mathbf{v} = \mathbf{0}_m$ says that each such dot product is 0. Say that the ith row of A, viewed as a column in $\mathbb{R}^n$, is $\mathbf{b}_i \in \mathbb{R}^n$. Then $\mathbf{v} \cdot \mathbf{b}_i = 0$ for all i. Any vector in $\mathbf{C}(A^{\mathsf{T}})$ is a linear combination of the rows of A viewed as columns in $\mathbb{R}^n$. Say that $\mathbf{w} \in \mathbf{C}(A^{\mathsf{T}})$ has the form

$$\mathbf{w} = c_1\mathbf{b}_1 + \cdots + c_m\mathbf{b}_m.$$

Then

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot (c_1\mathbf{b}_1 + \cdots + c_m\mathbf{b}_m) = c_1(\mathbf{v} \cdot \mathbf{b}_1) + \cdots + c_m(\mathbf{v} \cdot \mathbf{b}_m) = 0.$$ ■

This theorem allows us to prove the uniqueness result in Conjecture 4.2.2.

4.3.10 Corollary. *Let* $A \in \mathbb{R}^{m \times n}$ *and suppose that* $\mathbf{z} \in \mathbb{R}^n$ *satisfies both* $\mathbf{z} \in \mathbf{N}(A)$ *and* $\mathbf{z} \in \mathbf{C}(A^{\mathsf{T}})$*. Then* $\mathbf{z} = \mathbf{0}_n$*.*

4.3.11 Problem (!). Prove it: what is $\mathbf{z} \cdot \mathbf{z}$?

Suppose that we have written $\mathbf{x} \in \mathbb{R}^n$ as both

$$\mathbf{x} = \mathbf{v} + \mathbf{w} \quad \text{and} \quad \mathbf{x} = \widetilde{\mathbf{v}} + \widetilde{\mathbf{w}}$$

for some $\mathbf{v}$, $\widetilde{\mathbf{v}} \in \mathbf{N}(A)$ and $\mathbf{w}$, $\widetilde{\mathbf{w}} \in \mathbf{C}(A^{\mathsf{T}})$. We are going to show that

$$\mathbf{v} = \widetilde{\mathbf{v}} \quad \text{and} \quad \mathbf{w} = \widetilde{\mathbf{w}}.$$

First, we rearrange

$$\mathbf{v} + \mathbf{w} = \mathbf{x} = \widetilde{\mathbf{v}} + \widetilde{\mathbf{w}}$$

into

$$\mathbf{v} - \widetilde{\mathbf{v}} = \widetilde{\mathbf{w}} - \mathbf{w}.$$

Then $\mathbf{v} - \widetilde{\mathbf{v}} \in \mathbf{N}(A)$, since $\mathbf{v}$, $\widetilde{\mathbf{v}} \in \mathbf{N}(A)$ and $\mathbf{N}(A)$ is a subspace of $\mathbb{R}^n$. And, likewise, $\widetilde{\mathbf{w}} - \mathbf{w} \in \mathbf{C}(A^\mathsf{T})$. So, $\mathbf{v} - \widetilde{\mathbf{v}} = \widetilde{\mathbf{w}} - \mathbf{w} \in \mathbf{C}(A^\mathsf{T})$.

But then both $\mathbf{v} - \widetilde{\mathbf{v}} \in \mathbf{N}(A)$ and $\mathbf{v} - \widetilde{\mathbf{v}} \in \mathbf{C}(A^\mathsf{T})$, so by Lemma 4.3.10, we have $\mathbf{v} - \widetilde{\mathbf{v}} = \mathbf{0}_n$. This gives the desired equality $\mathbf{v} = \widetilde{\mathbf{v}}$. And this also gives $\widetilde{\mathbf{w}} - \mathbf{w} = \mathbf{v} - \widetilde{\mathbf{v}} = \mathbf{0}_n$, thus $\widetilde{\mathbf{w}} = \mathbf{w}$. And there we are: we have the uniqueness part of the decomposition in Conjecture 4.2.2.

4.3.12 Lemma. *Let $A \in \mathbb{R}^{m \times n}$. Suppose that*

$$\mathbf{v} + \mathbf{w} = \widetilde{\mathbf{v}} + \widetilde{\mathbf{w}}$$

for some $\mathbf{v}$, $\widetilde{\mathbf{v}} \in \mathbf{N}(A)$ and $\mathbf{w}$, $\widetilde{\mathbf{w}} \in \mathbf{C}(A^\mathsf{T})$. Then $\mathbf{v} = \widetilde{\mathbf{v}}$ and $\mathbf{w} = \widetilde{\mathbf{w}}$.

Onwards to the existence part of the decomposition part of Conjecture 4.2.2, which will take somewhat more effort than uniqueness. Our first step is to think more about A^T. Our proof of Theorem 4.3.9 was a perfectly adequate proof based on what A^T *is*: the matrix formed by swapping the rows and columns of A. This is a "static" way to think about A^T: it *is* an array of data, and there is nothing wrong with that.

But we can be dynamic: *what things do defines what things are.* Here is what A^T *does.* Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis vectors for $\mathbb{R}^n$ and $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_m$ be the standard basis vectors for $\mathbb{R}^m$. For example, if $n = 4$ and $m = 3$, then

$$\mathbf{e}_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \quad \text{and} \quad \mathbf{e}_4 = \begin{bmatrix}0\\0\\0\\1\end{bmatrix},$$

while

$$\widetilde{\mathbf{e}}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad \widetilde{\mathbf{e}}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \quad \text{and} \quad \widetilde{\mathbf{e}}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Now recall that multiplying a matrix by standard basis vectors extracts its columns, while taking the dot product of a vector with standard basis vectors extracts its entries. Since $A \in \mathbb{R}^{m \times n}$, the jth column of A is $A\mathbf{e}_j \in \mathbb{R}^m$, and then $A\mathbf{e}_j \cdot \widetilde{\mathbf{e}}_i$ is the ith entry in that column. That is,

$$A_{ij} = A\mathbf{e}_j \cdot \widetilde{\mathbf{e}}_i.$$

4.3.13 Problem (!). Here is a test for matrix equality. Suppose that A, $B \in \mathbb{R}^{m \times n}$. Prove that if

$$A\mathbf{v} \cdot \mathbf{w} = B\mathbf{v} \cdot \mathbf{w} \tag{4.3.1}$$

for all $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^m$, then $A = B$. [Hint: *the generous "for all" quantifier allows you to take* $\mathbf{v}$ *and* $\mathbf{w}$ *to be standard basis vectors.*]

Likewise, since $A^\mathsf{T} \in \mathbb{R}^{n \times m}$, the jth column of A^T is $A^\mathsf{T}\widetilde{\mathbf{e}}_j \in \mathbb{R}^n$, and then $A^\mathsf{T}\widetilde{\mathbf{e}}_j \cdot \mathbf{e}_i$ is the ith entry in that column. That is,

$$A^\mathsf{T}_{ij} = A^\mathsf{T}\widetilde{\mathbf{e}}_j \cdot \mathbf{e}_i.$$

Thus
$$A^{\mathsf{T}}\widetilde{\mathbf{e}}_j \cdot \mathbf{e}_i = A^{\mathsf{T}}_{ij} = A_{ji} = A\mathbf{e}_i \cdot \widetilde{\mathbf{e}}_j,$$
and so
$$A\mathbf{e}_i \cdot \widetilde{\mathbf{e}}_j = A^{\mathsf{T}}\widetilde{\mathbf{e}}_j \cdot \mathbf{e}_i.$$
The commutativity of the dot product ($\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$) gives
$$A\mathbf{e}_i \cdot \widetilde{\mathbf{e}}_j = \mathbf{e}_i \cdot A^{\mathsf{T}}\widetilde{\mathbf{e}}_j. \tag{4.3.2}$$
This is what A^{T} *does*: it pops across the dot product.

4.3.14 Remark. *There are actually two dot products in* (4.3.2). *The one on the left is in* $\mathbb{R}^m$, *since* $A\mathbf{e}_i$, $\widetilde{\mathbf{e}}_j \in \mathbb{R}^m$. *The one on the right is in* $\mathbb{R}^n$, *since* $\mathbf{e}_i$, $A^{\mathsf{T}}\widetilde{\mathbf{e}}_j \in \mathbb{R}^n$.

This "popping" behavior of the transpose is not limited to standard basis vectors.

4.3.15 Theorem (What the transpose does). *Let* $A \in \mathbb{R}^{m\times n}$, $\mathbf{v} \in \mathbb{R}^n$, *and* $\mathbf{w} \in \mathbb{R}^m$. *Then*
$$A\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot A^{\mathsf{T}}\mathbf{w}.$$
Moreover, the transpose is the only matrix in $\mathbb{R}^{n\times m}$ *to do this: if* $B \in \mathbb{R}^{n\times m}$ *satisfies*
$$A\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot B\mathbf{w}$$
for all $\mathbf{v} \in \mathbb{R}^n$ *and* $\mathbf{w} \in \mathbb{R}^m$, *then* $B = A^{\mathsf{T}}$.

4.3.16 Problem (+). Prove it. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis vectors for $\mathbb{R}^n$ and $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_m$ be the standard basis vectors for $\mathbb{R}^m$.

(i) Fix $\mathbf{v} \in \mathbb{R}^n$ and expand $\mathbf{v} = v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n$. Use (4.3.2), the linearity of matrix-vector multiplication, and dot product arithmetic from Problem 4.3.3 to show
$$A\mathbf{v} \cdot \widetilde{\mathbf{e}}_i = \mathbf{v} \cdot A^{\mathsf{T}}\widetilde{\mathbf{e}}_i, \ i = 1, \ldots, m. \tag{4.3.3}$$

(ii) Fix $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^m$ and now expand $\mathbf{w} = w_1\widetilde{\mathbf{e}}_1 + \cdots + w_m\widetilde{\mathbf{e}}_m$. Use again the linearity of matrix-vector multiplication and dot product arithmetic from Problem 4.3.3, now in conjunction with (4.3.3), to show
$$A\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot A^{\mathsf{T}}\mathbf{w}. \tag{4.3.4}$$

(iii) Suppose that $B \in \mathbb{R}^{n\times m}$ satisfies
$$A\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot B\mathbf{w} \tag{4.3.5}$$
for all $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^m$. Combine (4.3.4) and (4.3.5) (and use the commutativity of the dot product) to obtain $B\mathbf{w} \cdot \mathbf{v} = A^{\mathsf{T}}\mathbf{w} \cdot \mathbf{v}$ for all $\mathbf{w} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$. Then apply Problem 4.3.13 to obtain $B = A^{\mathsf{T}}$.

4.3.17 Problem ($\star$). Here are some nice properties of the transpose that can easily be deduced from what it does, rather than what it is.

(i) Let $A \in \mathbb{R}^{m \times n}$. First explain why $(A^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{m \times n}$, too. Then use Problem 4.3.13 to prove that $(A^{\mathsf{T}})^{\mathsf{T}} = A$ by showing that

$$(A^{\mathsf{T}})^{\mathsf{T}}\mathbf{v} \cdot \mathbf{w} = A\mathbf{v} \cdot \mathbf{w}$$

for all $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^m$.

(ii) Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. First explain why $(AB)^{\mathsf{T}}$, $B^{\mathsf{T}}A^{\mathsf{T}} \in \mathbb{R}^{p \times m}$. Then show that

$$(B^{\mathsf{T}}A^{\mathsf{T}})\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot AB\mathbf{w}$$

for all $\mathbf{v} \in \mathbb{R}^m$ and $\mathbf{w} \in \mathbb{R}^p$. Use the uniqueness result of Theorem 4.3.15 to conclude $(AB)^{\mathsf{T}} = B^{\mathsf{T}}A^{\mathsf{T}}$.

(iii) Let $A \in \mathbb{R}^{m \times m}$ be invertible. Prove that A^{T} is also invertible with inverse $(A^{\mathsf{T}})^{-1} = (A^{-1})^{\mathsf{T}}$ by computing $(AA^{-1})^{\mathsf{T}} = I_m^{\mathsf{T}}$ and $(AA^{-1})^{\mathsf{T}} = A^{\mathsf{T}}(A^{-1})^{\mathsf{T}}$. What does this tell you?

(iv) Let $P \in \mathbb{R}^{m \times m}$ be a permutation matrix, so P contains all of the columns of the identity matrix I_m (each column appearing once, and only once) in some order. Argue that

$$P\mathbf{v} \cdot P\mathbf{w} = \mathbf{v} \cdot \mathbf{w}, \ \mathbf{v}, \ \mathbf{w} \in \mathbb{R}^m. \tag{4.3.6}$$

Conclude $P^{\mathsf{T}}P\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w}$ for all $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^m$. Why does this imply that P is invertible with $P^{-1} = P^{\mathsf{T}}$?

Content from Strang's *ILA* 6E. Pages 67–68 discuss fundamental properties of the transpose. Pages 68–69 show how the transpose interacts with dot products (I wholly disagree that $\cdot$ is "unprofessional"—I like that it emphasizes how the dot product takes in two inputs and how it's "linear in each input." I like dot products.). If you have seen integration by parts in calculus, read Example 2 on p. 69.

Now we can give another proof of Theorem 4.3.9 that relies on what the transpose does, rather than is. Let $A \in \mathbb{R}^{m \times n}$, $\mathbf{v} \in \mathbf{N}(A)$, and $\mathbf{w} \in \mathbf{C}(A^{\mathsf{T}})$. Then $A\mathbf{v} = \mathbf{0}_m$ and there is $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{w} = A^{\mathsf{T}}\mathbf{y}$. We compute

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot A^{\mathsf{T}}\mathbf{y} = A\mathbf{v} \cdot \mathbf{y} = \mathbf{0}_m \cdot \mathbf{y} = 0. \tag{4.3.7}$$

So slick! By the way, the first two dot products in (4.5.1) were dot products in $\mathbb{R}^n$, but the second two were in $\mathbb{R}^m$.

Content from Strang's *ILA* 6E. Read p. 144 starting with the box "The **nullspace of** A...." Then read the "Important" paragraph on p. 145 about the orthogonality of $\mathbf{C}(A)$ and $\mathbf{N}(A^{\mathsf{T}})$, which we will discuss in greater detail later, and Example 1. Note that Strang typically likes to write the dot product as $\mathbf{x}^{\mathsf{T}}\mathbf{y}$, not $\mathbf{x} \cdot \mathbf{y}$.

4.3.18 Problem (!). Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$

Use results from Examples 3.3.1 and 3.4.2 to give bases for $\mathbf{N}(A)$ and $\mathbf{C}(A^\mathsf{T})$, and check directly that the vectors in the basis for $\mathbf{N}(A)$ are orthogonal to the vectors in the basis for $\mathbf{C}(A^\mathsf{T})$.

4.3.19 Problem (+). Here is a much less slick way to show the orthogonality of vectors in $\mathbf{N}(A)$ and $\mathbf{C}(A^\mathsf{T})$ that reinforces properties of the CR-factorization. Let $A \in \mathbb{R}^{m\times n}$ with $r = \operatorname{rank}(A)$. We show again that if $\mathbf{v} \in \mathbf{N}(A)$ and $\mathbf{w} \in \mathbf{C}(A^\mathsf{T})$, then $\mathbf{v}\cdot\mathbf{w} = 0$.

(i) If $r = n$, why do you have very little work to do? Do it.

(ii) If $r = 0$, again, why do you have very little work to do? Go forth and do it.

(iii) Suppose for the remainder of the problem that $1 \le r < n$. Write $A = CR$ in its CR-factorization with $C \in \mathbb{R}^{m\times r}$, $R = \begin{bmatrix} I_r & F \end{bmatrix} P$ for some $F \in \mathbb{R}^{r\times(n-r)}$ and some permutation matrix $P \in \mathbb{R}^{n\times n}$. First show that

$$A^\mathsf{T} = P^\mathsf{T} \begin{bmatrix} I_r \\ F^\mathsf{T} \end{bmatrix} C^\mathsf{T}.$$

For $\mathbf{w} \in \mathbf{C}(A^\mathsf{T})$, there is $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{w} = A^\mathsf{T}\mathbf{y}$; put $\mathbf{z} = C^\mathsf{T}\mathbf{y} \in \mathbb{R}^r$ to have

$$\mathbf{w} = P^\mathsf{T} \begin{bmatrix} I_r \\ F^\mathsf{T} \end{bmatrix} \mathbf{z}.$$

(iv) Use part (iii) of Theorem 3.2.5 (now with $P^{-1} = P^\mathsf{T}$ by part (iv) of Problem 4.3.17) to write any $\mathbf{v} \in \mathbf{N}(A)$ as

$$\mathbf{v} = P^\mathsf{T} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{x}.$$

(v) Use (4.3.6), with P^T in place of P (this is allowed, since P^T is a permutation matrix when P is a permutation matrix, by Problem 4.3.8) to compute

$$\mathbf{v}\cdot\mathbf{w} = \left(\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{x}\right) \cdot \left(\begin{bmatrix} I_r \\ F^\mathsf{T} \end{bmatrix} \mathbf{z}\right).$$

Break $\mathbf{x}$ and $\mathbf{z}$ into appropriately sized blocks to do the matrix-vector multiplications; then compute the dot products.

4.4. The row space is the missing piece.

We are working enough with $\mathbf{C}(A^{\mathsf{T}})$ that we should give it a special name to reflect the fact that the columns of A^{T} come from the rows of A.

4.4.1 Definition. *The* ***ROW SPACE*** *of* $A \in \mathbb{R}^{m\times n}$ *is* $\mathbf{C}(A^{\mathsf{T}})$.

We are very close to proving the existence part of the decomposition in Conjecture 4.2.2. We need one more fact: $\operatorname{rank}(A) = \operatorname{rank}(A^{\mathsf{T}})$.

Here is one way to develop this. Let $A \in \mathbb{R}^{m\times n}$ have the RREF R_0, so there is $E \in \mathbb{R}^{m\times m}$ invertible such that $EA = R_0$. Then $A = E^{-1}R_0$ and so

$$A^{\mathsf{T}} = (E^{-1}R_0)^{\mathsf{T}} = R_0^{\mathsf{T}}(E^{-1})^{\mathsf{T}}.$$

Since E^{-1} is invertible, so is $(E^{-1})^{\mathsf{T}}$ by part (ii) of Problem 4.3.17, and then

$$\mathbf{C}(A^{\mathsf{T}}) = \mathbf{C}(R_0^{\mathsf{T}}(E^{-1})^{\mathsf{T}}) = \mathbf{C}(R_0^{\mathsf{T}})$$

by Problem 3.11.7. Thus

$$\operatorname{rank}(A^{\mathsf{T}}) = \dim(\mathbf{C}(A^{\mathsf{T}})) = \dim(\mathbf{C}(R_0^{\mathsf{T}})) = \operatorname{rank}(R_0^{\mathsf{T}}).$$

If $\operatorname{rank}(R_0^{\mathsf{T}}) = \operatorname{rank}(A)$, then we are done.

4.4.2 Remark. *A nice consequence of the identity* $\mathbf{C}(A^{\mathsf{T}}) = \mathbf{C}(R_0^{\mathsf{T}})$ *is that if we know a basis for* $\mathbf{C}(R_0^{\mathsf{T}})$*, then that is also a basis for* $\mathbf{C}(A^{\mathsf{T}})$*. Annoyingly, a basis for* $\mathbf{C}(R_0)$ *still does not have to be a basis for* $\mathbf{C}(A)$*!*

4.4.3 Example. Here is a way to "see" why $\operatorname{rank}(R_0^{\mathsf{T}}) = \operatorname{rank}(A)$ should be true from a specific situation. Suppose that $\operatorname{rank}(A) = 2$ and the RREF has the form

$$R_0 = \begin{bmatrix} 0 & 1 & * & 0 & * \\ 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.4.1}$$

Then

$$R_0^{\mathsf{T}} = \begin{bmatrix} 1 & 0 & 0 \\ * & 0 & 0 \\ 0 & 1 & 0 \\ * & * & 0 \\ * & * & 0 \end{bmatrix}. \tag{4.4.2}$$

The first two columns of R_0^{T} are then the pivot columns of R_0^{T}, and so $\operatorname{rank}(R_0^{\mathsf{T}}) = 2$.

4.4.4 Problem (!). By considering (4.4.1) and (4.4.2), explain why $\mathsf{rref}(A)^\mathsf{T} \neq \mathsf{rref}(A^\mathsf{T})$ in general.

It is somewhat more annoying to prove that $\text{rank}(A) = \text{rank}(R_0^\mathsf{T})$ in general, but the CR-factorization gives a related proof, albeit with less visualization of rows turning into columns.

4.4.5 Theorem. *Let $A \in \mathbb{R}^{m\times n}$. Then $\text{rank}(A) = \text{rank}(A^\mathsf{T})$.*

Proof. Let $\text{rank}(A) = r$. If $r = 0$, then A is the zero matrix in $\mathbb{R}^{m\times n}$, so A^T is the zero matrix in $\mathbb{R}^{n\times m}$. Then $\mathbf{C}(A^\mathsf{T}) = \{\mathbf{0}_n\}$, so $\text{rank}(A^\mathsf{T}) = 0$.

Otherwise, suppose $r \geq 1$, and let $A = CR$ be the CR-factorization of A. We have $C \in \mathbb{R}^{m\times r}$ and $R \in \mathbb{R}^{r\times n}$, so $R^\mathsf{T} \in \mathbb{R}^{n\times r}$. Then $A^\mathsf{T} = (CR)^\mathsf{T} = R^\mathsf{T}C^\mathsf{T}$, so $\text{rank}(A^\mathsf{T}) \leq \text{rank}(R^\mathsf{T})$ by Problem 3.11.7. And since $R^\mathsf{T} \in \mathbb{R}^{n\times r}$, we can estimate $\text{rank}(R^\mathsf{T}) \leq \min\{n, r\} \leq r$.

All of this is to say that $\text{rank}(A^\mathsf{T}) \leq \text{rank}(A)$. This is true for any matrix A, so replace A with A^T:

$$\text{rank}(A) = \text{rank}((A^\mathsf{T})^\mathsf{T}) \leq \text{rank}(A^\mathsf{T}).$$

Put the inequalities $\text{rank}(A^\mathsf{T}) \leq \text{rank}(A)$ and $\text{rank}(A) \leq \text{rank}(A^\mathsf{T})$ together to get $\text{rank}(A) = \text{rank}(A^\mathsf{T})$. ■

Now we can make precise the observations about how the rows of the RREF become a basis for the row space.

4.4.6 Corollary. *Let $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) \geq 1$ have the CR-factorization $A = CR$.*

(i) $\mathbf{C}(R^\mathsf{T}) = \mathbf{C}(A^\mathsf{T})$. *In particular, the columns of R^T are a basis for $\mathbf{C}(A^\mathsf{T})$; equivalently, the nonzero rows of $\mathsf{rref}(A)$, viewed as vectors in $\mathbb{R}^n$, are a basis for $\mathbf{C}(A^\mathsf{T})$.*

(ii) $\mathbf{N}(A^\mathsf{T}) = \mathbf{N}(C^\mathsf{T})$.

Proof. Suppose $r := \text{rank}(A)$ with $1 \leq r \leq n$, so $R \in \mathbb{R}^{r\times n}$.

(i) Since $A^\mathsf{T} = R^\mathsf{T}C^\mathsf{T}$, every vector in $\mathbf{C}(A^\mathsf{T}) = \mathbf{C}(R^\mathsf{T}C^\mathsf{T})$ is in $\mathbf{C}(R^\mathsf{T})$ by Problem 2.2.7. We know that $\text{rank}(A^\mathsf{T}) = \text{rank}(A) = \text{rank}(R) = \text{rank}(R^\mathsf{T})$; the first and third equalities are the new result of Theorem 4.4.5, and the second is the long-familiar result of Problem 2.3.7. That is, every vector in $\mathbf{C}(A^\mathsf{T})$ is in $\mathbf{C}(R^\mathsf{T})$, and $\dim(\mathbf{C}(A^\mathsf{T})) = \dim(\mathbf{C}(R^\mathsf{T}))$, so $\mathbf{C}(A^\mathsf{T}) = \mathbf{C}(R^\mathsf{T})$ by Problem 3.11.5. Any basis for $\mathbf{C}(R^\mathsf{T})$ is therefore a basis for $\mathbf{C}(A^\mathsf{T})$. Now, $R^\mathsf{T} \in \mathbb{R}^{n\times r}$, and $\text{rank}(R^\mathsf{T}) = r$, so every column of R^T is a pivot column of R^T, and therefore the columns of R^T are a basis for $\mathbf{C}(R^\mathsf{T})$, thus for $\mathbf{C}(A^\mathsf{T})$.

(ii) We have $A^\mathsf{T}\mathbf{y} = \mathbf{0}_n$ if and only if $R^\mathsf{T}C^\mathsf{T}\mathbf{y} = \mathbf{0}_n$. Since the columns of $R^\mathsf{T} \in \mathbb{R}^{n\times r}$ are independent, $\mathbf{N}(R^\mathsf{T}) = \{\mathbf{0}_n\}$, and so $R^\mathsf{T}C^\mathsf{T}\mathbf{y} = \mathbf{0}_n$ if and only if $C^\mathsf{T}\mathbf{y} = \mathbf{0}_n$. Thus $\mathbf{y} \in \mathbf{N}(A^\mathsf{T})$ precisely when $\mathbf{y} \in \mathbf{N}(C^\mathsf{T})$, so $\mathbf{N}(A^\mathsf{T}) = \mathbf{N}(C^\mathsf{T})$. ■

Here is a summary of this corollary, Problem 2.3.7, and the work preceding Example 3.2.1:

$$A = CR \Longrightarrow \begin{cases} \mathbf{C}(A) = \mathbf{C}(C) \\ \mathbf{N}(A) = \mathbf{N}(R) \\ \\ \mathbf{C}(A^{\mathsf{T}}) = \mathbf{C}(R^{\mathsf{T}}) \\ \mathbf{N}(A^{\mathsf{T}}) = \mathbf{N}(C^{\mathsf{T}}) \end{cases}$$

Content from Strang's *ILA* 6E. Read #1 on p. 130, #4 on p. 131, #1 on p. 132, and #4 on p. 133. Actually, probably best to reread all of pp. 130–133 and see all four subspaces talk to each other.

Now we are ready to prove the existence part of the decomposition in Conjecture 4.2.2. We do so by building a special basis for $\mathbb{R}^n$, and here is where the abstract tools of subspace, basis, and dimension really shine. Suppose that $A \in \mathbb{R}^{m\times n}$ has rank r. Then $\dim(\mathbf{N}(A)) = n - r$ by rank–nullity and $\dim(\mathbf{C}(A^{\mathsf{T}})) = r$ by the result above. Since $(n-r)+r = n$, this should make us feel optimistic.

4.4.7 Problem (!). We first handle two "easy" and "extreme" cases.

(i) Prove that if A has full column rank ($r = n$), then every $\mathbf{x} \in \mathbb{R}^n$ can be written uniquely in the form $\mathbf{x} = \mathbf{v} + \mathbf{w}$ for $\mathbf{v} \in \mathbf{N}(A)$ and $\mathbf{w} \in \mathbf{C}(A^{\mathsf{T}})$. [Hint: *there are not many choices for* $\mathbf{v}$.]

(ii) What happens if $r = 0$?

Going forward, suppose $1 \le r < n$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_{n-r}$ be a basis for $\mathbf{N}(A)$ and let $\mathbf{w}_1, \ldots, \mathbf{w}_r$ be a basis for $\mathbf{C}(A^{\mathsf{T}})$. The list $\mathbf{v}_1, \ldots, \mathbf{v}_{n-r}$ is independent with length $n-r$, and the list $\mathbf{w}_1, \ldots, \mathbf{w}_r$ is independent with length r. Our optimism should be growing to the point that we consider the list

$$\mathbf{v}_1, \ldots, \mathbf{v}_{n-r}, \mathbf{w}_1, \ldots, \mathbf{w}_r$$

as a possible basis for $\mathbb{R}^n$. Since this list has length n, we just have to check that it is independent. So, suppose that

$$\underbrace{y_1\mathbf{v}_1 + \cdots + y_{n-r}\mathbf{v}_{n-r}}_{\mathbf{v}} + \underbrace{z_1\mathbf{w}_1 + \cdots + z_r\mathbf{w}_r}_{\mathbf{w}} = \mathbf{0}_n$$

for some scalars $y_1, \ldots, y_{n-r}$, $z_1, \ldots, z_r \in \mathbb{R}$. We want to show $y_j = z_j = 0$ for all j.

Since $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathbf{N}(A)$ and $\mathbf{N}(A)$ is a subspace of $\mathbb{R}^n$, we have $\mathbf{v} \in \mathbf{N}(A)$. The same reasoning gives $\mathbf{w} \in \mathbf{C}(A^{\mathsf{T}})$. And so $\mathbf{v}$ and $\mathbf{w}$ relate to each other in two ways:

$$\mathbf{v} + \mathbf{w} = \mathbf{0}_n \qquad \text{and} \qquad \mathbf{v} \cdot \mathbf{w} = 0.$$

Now here is the great trick: take the dot product. We have

$$(\mathbf{v}+\mathbf{w})\cdot\mathbf{v} = \mathbf{0}_n\cdot\mathbf{v},$$

and so

$$(\mathbf{v}\cdot\mathbf{v}) + (\mathbf{w}\cdot\mathbf{v}) = 0,$$

and therefore

$$\mathbf{v}\cdot\mathbf{v} = 0,$$

and so finally

$$\mathbf{v} = \mathbf{0}_n.$$

Then

$$\mathbf{0}_n = \mathbf{v}+\mathbf{w} = \mathbf{0}_n + \mathbf{w} = \mathbf{w}$$

as well.

This is both true and useful, for now we know

$$y_1\mathbf{v}_1 + \cdots + y_{n-r}\mathbf{v}_{n-r} = \mathbf{v} = \mathbf{0}_n \quad \text{and} \quad z_1\mathbf{w}_1 + \cdots + z_r\mathbf{w}_r = \mathbf{w} = \mathbf{0}_n.$$

Independence of the lists $\mathbf{v}_1, \ldots, \mathbf{v}_{n-r}$ and $\mathbf{w}_1, \ldots, \mathbf{w}_r$ forces all of the coefficients y_j and z_j to be 0, as desired.

Thus the list $\mathbf{v}_1, \ldots, \mathbf{v}_{n-r}, \mathbf{w}_1, \ldots, \mathbf{w}_r$ is a basis for $\mathbb{R}^n$, so for each $\mathbf{x} \in \mathbb{R}^n$, there are (unique) $y_1, \ldots, y_{n-r}, z_1, \ldots, z_r \in \mathbb{R}$ such that

$$\mathbf{x} = \underbrace{y_1\mathbf{v}_1 + \cdots + y_{n-r}\mathbf{v}_{n-r}}_{\mathbf{v}} + \underbrace{z_1\mathbf{w}_1 + \cdots + z_r\mathbf{w}_r}_{\mathbf{w}}.$$

With $\mathbf{v} \in \mathbf{N}(A)$ and $\mathbf{w} \in \mathbf{C}(A^\mathsf{T})$ as just defined, this gives the existence of the decomposition in Conjecture 4.2.2.

4.4.8 Remark. *This is one of those times when having a basis for a subspace in the abstract was very useful. Without knowing precisely the forms of the bases for $\mathbf{N}(A)$ (which we could extract from the RREF of A) or $\mathbf{C}(A^\mathsf{T})$ (which we could extract from the pivot rows of the RREF of A), we built a basis for $\mathbb{R}^n$ and used that to get our desired decomposition.*

Here is the confirmation of Conjecture 4.2.2.

4.4.9 Theorem. *Let $A \in \mathbb{R}^{m\times n}$. For each $\mathbf{x} \in \mathbb{R}^n$ there exist unique $\mathbf{v} \in \mathbf{N}(A)$ and $\mathbf{w} \in \mathbf{C}(A^\mathsf{T})$ such that $\mathbf{x} = \mathbf{v}+\mathbf{w}$ and $\mathbf{v}\cdot\mathbf{w} = 0$. We summarize this symbolically by writing*

$$\mathbb{R}^n = \mathbf{N}(A) \oplus \mathbf{C}(A^\mathsf{T}),$$

and we call $\mathbb{R}^n$ the ***ORTHOGONAL DIRECT SUM*** *of $\mathbf{N}(A)$ and $\mathbf{C}(A^\mathsf{T})$.*

Content from Strang's *ILA* 6E. "Combining Bases from Subspaces" on p. 147 contains these "counting" arguments that lead to a basis for all of $\mathbb{R}^n$ out of bases for $\mathbf{N}(A)$ and

$\mathbf{C}(A^{\mathsf{T}})$. Read Examples 3 and 4. Then go back to the box on p. 145 with the inequality $\dim(\mathcal{V})+\dim(\mathcal{W}) \le n$. Can you prove this? [Hint: *start with bases for $\mathcal{V}$ and $\mathcal{W}$, and show that together, the vectors in both bases are still independent.*] Can you give an example of orthogonal subspaces for which the inequality is strict? [Hint: *look at some, but not all, of the standard basis vectors.*]

4.4.10 Problem (!). Let $A \in \mathbb{R}^{m\times n}$. Here is how $\mathbf{C}(A)$ "fits into" $\mathbb{R}^m$: prove that we can write any $\mathbf{b} \in \mathbb{R}^m$ uniquely in the form $\mathbf{b} = \mathbf{v} + \mathbf{w}$ for some $\mathbf{v} \in \mathbf{C}(A)$ and $\mathbf{w} \in \mathbf{N}(A^{\mathsf{T}})$, and that $\mathbf{v}\cdot\mathbf{w} = 0$. We summarize this symbolically by writing

$$\mathbb{R}^m = \mathbf{C}(A) \oplus \mathbf{N}(A^{\mathsf{T}}).$$

[Hint: *replace A in Theorem 4.4.9 with A^{T}.*]

We should not interpret the dual results $\mathbb{R}^n = \mathbf{N}(A) \oplus \mathbf{C}(A^{\mathsf{T}})$ and $\mathbb{R}^m = \mathbf{C}(A) \oplus \mathbf{N}(A^{\mathsf{T}})$ as saying that any vector in $\mathbb{R}^n$ or $\mathbb{R}^m$ is in one or another of these FOUR FUNDAMENTAL SUBSPACES $\mathbf{N}(A)$, $\mathbf{C}(A^{\mathsf{T}})$, $\mathbf{C}(A)$, and $\mathbf{N}(A^{\mathsf{T}})$ associated with A. Rather, we can build $\mathbb{R}^n$ and $\mathbb{R}^m$ out of the four fundamental subspaces.

4.4.11 Example. Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$

Over the course of many examples, we have already found bases for the four fundamental subspaces of A.

(i) Example 3.10.1 showed that a basis for $\mathbf{N}(A)$ is the list

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

We found this by computing

$$R_0 = \mathsf{rref}(A) = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

in Examples 3.3.1 and 3.4.1 and then (in the latter example) solving $R_0\mathbf{x} = \mathbf{0}_3$.

(ii) Example 1.14.7 showed that a basis for $\mathbf{C}(A)$ is

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

because these are the pivot columns of A. Of course, we can also get this from the RREF.

(iii) Corollary 4.4.6 tells us that a basis for $\mathbf{C}(A^{\mathsf{T}})$ consists of the "pivot rows" of R_0. That is, a basis for $\mathbf{C}(A^{\mathsf{T}})$ is

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

However, in Example 3.4.2 we computed $\mathsf{rref}(A^{\mathsf{T}})$ and read from it that the pivot columns of A^{T} are columns 1 and 3, so a basis for $\mathbf{C}(A^{\mathsf{T}})$ is also

$$\begin{bmatrix}0\\1\\2\\1\\7\end{bmatrix}, \begin{bmatrix}0\\0\\0\\2\\8\end{bmatrix}.$$

(iv) Finally, Example 3.4.2 also showed that a basis for $\mathbf{N}(A^{\mathsf{T}})$ is the list of length 1

$$\begin{bmatrix}-2\\1\\0\end{bmatrix}.$$

We found this by computing $\widetilde{R}_0 := \mathsf{rref}(A^{\mathsf{T}})$ and solving $\widetilde{R}_0\mathbf{x} = \mathbf{0}_5$.

We combine the bases for $\mathbf{N}(A)$ and $\mathbf{C}(A^{\mathsf{T}})$ to obtain the basis

$$\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\-2\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\-3\\0\\-4\\1\end{bmatrix}, \begin{bmatrix}0\\1\\2\\1\\7\end{bmatrix}, \begin{bmatrix}0\\0\\0\\2\\8\end{bmatrix}$$

for $\mathbb{R}^5$, and we combine the bases for $\mathbf{C}(A)$ and $\mathbf{N}(A^{\mathsf{T}})$ to obtain the basis

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

for $\mathbb{R}^3$. We could check, with some effort, that these are bases from the definition of basis—or we could use the procedure that proved Theorem 4.4.9.

4.4.12 Problem ($\star$). Let A be as in the previous example (as A almost always is). Give an example of a vector $\mathbf{v} \in \mathbb{R}^5$ such that $\mathbf{v} \notin \mathbf{N}(A)$ and $\mathbf{v} \notin \mathbf{C}(A^\mathsf{T})$. Then give an example of $\mathbf{w} \in \mathbb{R}^3$ such that $\mathbf{w} \notin \mathbf{C}(A)$ and $\mathbf{w} \notin \mathbf{N}(A^\mathsf{T})$. [Hint: *add some nonzero vectors chosen from the right places.*]

4.5. Orthogonality encodes solvability conditions.

We started our study of orthogonality with how the null space of $A \in \mathbb{R}^{m \times n}$ fits into $\mathbb{R}^n$. After a lot of work on the null space, we quickly extracted information about how the column space fits into $\mathbb{R}^n$. Now we will focus on the column space and obtain an alternate characterization of this valuable object that goes beyond its original definition.

We need two auxiliary results. First we demonstrate another way that the dot product extracts information about vectors. If we test or measure a given vector against all vectors under the lens of the dot product and we always get 0, then that given vector is the zero vector.

4.5.1 Problem (!). Suppose that $\mathbf{v} \in \mathbb{R}^n$ satisfies $\mathbf{v} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in \mathbb{R}^n$. Prove that $\mathbf{v} = \mathbf{0}_n$. [Hint: *take advantage of that generous quantifier "for all" and let* $\mathbf{w}$ *be one of the standard basis vectors.*]

Next, we revisit and reinterpret a familiar result.

4.5.2 Lemma. *Let* $A \in \mathbb{R}^{m \times n}$*.*

$$\mathbf{N}(A^\mathsf{T}) = \{\mathbf{v} \in \mathbb{R}^m \mid \mathbf{v} \cdot \mathbf{w} = 0 \textit{ for all } \mathbf{w} \in \mathbf{C}(A)\}. \tag{4.5.1}$$

Proof. Let

$$\mathcal{W} := \{\mathbf{v} \in \mathbb{R}^m \mid \mathbf{v} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in \mathbf{C}(A)\}.$$

1. *The proof of* $\mathbf{v} \in \mathbf{N}(A^\mathsf{T}) \implies \mathbf{v} \in \mathcal{W}$. Suppose $A^\mathsf{T}\mathbf{v} = \mathbf{0}_n$ and let $\mathbf{w} \in \mathbf{C}(A)$. Then $\mathbf{w} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, and we compute

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot A\mathbf{x} = A^\mathsf{T}\mathbf{v} \cdot \mathbf{x} = \mathbf{0}_n \cdot \mathbf{x} = 0.$$

2. *The proof of* $\mathbf{v} \in \mathcal{W} \implies \mathbf{v} \in \mathbf{N}(A^\mathsf{T})$. Now suppose $\mathbf{v} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in \mathbf{C}(A)$. We want to show $A^\mathsf{T}\mathbf{v} = \mathbf{0}_n$. Any $\mathbf{w} \in \mathbf{C}(A)$ has the form $\mathbf{w} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, and so

$$0 = \mathbf{v} \cdot A\mathbf{x} = A^\mathsf{T}\mathbf{v} \cdot \mathbf{x}$$

for all $\mathbf{x} \in \mathbb{R}^n$. Problem 4.5.1 tells us that $A^\mathsf{T}\mathbf{v} = \mathbf{0}_n$. ■

The equality (4.5.1) motivates a new kind of structure.

4.5.3 Definition. *Let* $\mathcal{V}$ *be a subset of* $\mathbb{R}^p$ *(not necessarily a subspace). The* ***ORTHOGONAL***

***COMPLEMENT** of $\mathcal{V}$ in $\mathbb{R}^p$ is*

$$\mathcal{V}^\perp := \{\mathbf{w} \in \mathbb{R}^p \mid \mathbf{v} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{v} \in \mathcal{V}\}.$$

We pronounce the symbol $\mathcal{V}^\perp$ as "vee perp."

Content from Strang's *ILA* 6E. The last paragraph on p. 145 defines orthogonal complements.

4.5.4 Problem ($\star$). Let $\mathcal{V}$ be a subset of $\mathbb{R}^p$. Prove that $\mathcal{V}^\perp$ is a subspace of $\mathbb{R}^p$. Did you need $\mathcal{V}$ to be a subspace?

4.5.5 Example.

(i) The equality (4.5.1) says that

$$\mathbf{C}(A)^\perp = \mathbf{N}(A^\mathsf{T}) \tag{4.5.2}$$

for any $A \in \mathbb{R}^{m \times n}$.

(ii) Let $\mathcal{V} = \mathbb{R}^p$ and suppose that $\mathbf{w} \in \mathbb{R}^p$ with $\mathbf{v} \cdot \mathbf{w} = 0$ for all $\mathbf{v} \in \mathbb{R}^p$. Problem 4.5.1 says that $\mathbf{w} = \mathbf{0}_p$, so $(\mathbb{R}^p)^\perp = \{\mathbf{0}_p\}$.

(iii) Let $\mathcal{V} = \{\mathbf{0}_p\}$. Then $\mathbf{0}_p \cdot \mathbf{w} = 0$ for any $\mathbf{w} \in \mathbb{R}^p$, so $\{\mathbf{0}_n\}^\perp = \mathbb{R}^p$.

4.5.6 Problem (!). For $A \in \mathbb{R}^{m \times n}$, prove that $\mathbf{N}(A) = \mathbf{C}(A^\mathsf{T})^\perp$.

We just saw the extreme cases of $(\mathbb{R}^p)^\perp = \{\mathbf{0}_p\}$ and $\{\mathbf{0}_p\}^\perp = \mathbb{R}^p$. Thus

$$\big((\mathbb{R}^p)^\perp\big)^\perp = \{\mathbf{0}_p\}^\perp = \mathbb{R}^p \quad \text{and} \quad \big(\{\mathbf{0}_p\}^\perp\big)^\perp = (\mathbb{R}^p)^\perp = \{\mathbf{0}_p\}.$$

4.5.7 Problem (!). Here is a less extreme case in $\mathbb{R}^2$. Let $\mathcal{V} = \operatorname{span}(\mathbf{e}_1)$. Draw pictures to convince yourself that $(\mathcal{V}^\perp)^\perp = \mathcal{V}$ and then prove it.

These examples might make us wonder the following.

4.5.8 Conjecture. *$(\mathcal{V}^\perp)^\perp = \mathcal{V}$ for any subspace $\mathcal{V}$ of $\mathbb{R}^p$.*

In particular, if true Conjecture 4.5.8 would imply

$$\mathbf{C}(A) = \big(\mathbf{C}(A)^\perp\big)^\perp = \mathbf{N}(A^\mathsf{T})^\perp. \tag{4.5.3}$$

That is,

$$\mathbf{C}(A) = \left\{\mathbf{b} \in \mathbb{R}^m \;\middle|\; A^\mathsf{T}\mathbf{z} = \mathbf{0}_n \implies \mathbf{b} \cdot \mathbf{z} = 0\right\}.$$

This would give us a new way of deciding solvability of $A\mathbf{x} = \mathbf{b}$: check that $\mathbf{b}$ is orthogonal to everything in $\mathbf{N}(A^\mathsf{T})$. Or, and this might be faster, just check that $\mathbf{b}$ is orthogonal to a basis for $\mathbf{N}(A^\mathsf{T})$.

4.5.9 Problem (!). Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be a basis for $\mathcal{V}$. Suppose that $\mathbf{v} \in \mathbb{R}^p$ satisfies $\mathbf{v} \cdot \mathbf{v}_j = 0$ for $j = 1, \ldots, r$. Prove that $\mathbf{v} \in \mathcal{V}^\perp$.

The point is that if (4.5.3) is true, then it would give us a different way of describing the column space. In particular, we might get an easier way of checking that a vector is *not* in the column space than doing elementary row operations and going to the RREF. We would still have to solve $A^\mathsf{T}\mathbf{z} = \mathbf{0}_n$, but that is probably easier than solving $A\mathbf{x} = \mathbf{b}$, since $\mathbf{0}_n$ is a much simpler vector than an arbitrary $\mathbf{b}$.

4.5.10 Example. We check that (4.5.3) is true for the familiar matrix

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$

Example 3.4.1 gave us the "solvability condition" $b_2 = 2b_1$ for the problem $A\mathbf{x} = \mathbf{b}$. That is, $\mathbf{b} = (b_1, b_2, b_3) \in \mathbf{C}(A)$ if and only if $b_2 = 2b_1$. Actually, since the second row of A is twice the first, we have known that *if* $\mathbf{b} \in \mathbf{C}(A)$, *then* $b_2 = 2b_1$ for as long as we have known how to multiply matrices and vectors. The other direction, that *if* $b_2 = 2b_1$, *then* $\mathbf{b} \in \mathbf{C}(A)$, was a consequence of elimination.

Now we develop this solvability condition from (4.5.3). Example 4.4.11 showed that a basis for $\mathbf{N}(A^\mathsf{T})$ is the very short list consisting of the single vector $\mathbf{z} := (-2, 1, 0)$. Consequently, if $\mathbf{b} \in \mathbf{C}(A) = \mathbf{N}(A^\mathsf{T})^\perp$, then $\mathbf{b} \cdot \mathbf{z} = 0$, since $\mathbf{z} \in \mathbf{N}(A^\mathsf{T})$. We have $\mathbf{b} \cdot \mathbf{z} = -2b_1 + b_2$, so $\mathbf{b} \cdot \mathbf{z} = 0$ means $b_2 = 2b_1$.

Conversely, if $b_2 = 2b_1$, then

$$0 = b_2 - 2b_1 = (-2)b_1 + (1 \cdot b_2) + (0 \cdot b_3) = \mathbf{z} \cdot \mathbf{b}.$$

Then for any $\mathbf{y} \in \mathbf{N}(A^\mathsf{T})$, we have $\mathbf{y} = c\mathbf{z}$ for some $c \in \mathbb{R}$, and then

$$\mathbf{b} \cdot \mathbf{y} = \mathbf{b} \cdot (c\mathbf{z}) = c(\mathbf{b} \cdot \mathbf{z}) = 0.$$

That is, if $b_2 = 2b_1$, then $\mathbf{b} \in \mathbf{N}(A^\mathsf{T})^\perp = \mathbf{C}(A^\mathsf{T})$.

4.5.11 Problem ($\star$). Assuming Conjecture 4.5.8 to be true, prove the **Fredholm alternative**: if $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, then one, and only one, of the following is true.

(i) $\mathbf{b} \in \mathbf{C}(A)$ and so the problem $A\mathbf{x} = \mathbf{b}$ has a solution.

(ii) There is $\mathbf{v} \in \mathbb{R}^m$ such that $A^\mathsf{T}\mathbf{v} = \mathbf{0}_n$ and $\mathbf{b} \cdot \mathbf{v} \neq 0$.

The second condition gives us a test for $\mathbf{b} \notin \mathbf{C}(A)$ that does not actually require us to fail explicitly at solving $A\mathbf{x} = \mathbf{b}$.

Now we prove Conjecture 4.5.8. For a subspace $\mathcal{V}$ of $\mathbb{R}^p$, we want to show that if $\mathbf{v} \in \mathcal{V}$, then $\mathbf{v} \in (\mathcal{V}^\perp)^\perp$, and if $\mathbf{x} \in (\mathcal{V}^\perp)^\perp$, then $\mathbf{x} \in \mathcal{V}$. The first of these is fairly straightforward.

4.5.12 Lemma. *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. If $\mathbf{v} \in \mathcal{V}$, then $\mathbf{v} \in (\mathcal{V}^\perp)^\perp$.*

Proof. Let $\mathbf{v} \in \mathcal{V}$. We want to show $\mathbf{v} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in \mathcal{V}^\perp$. And that is exactly what it means for $\mathbf{w}$ to be a vector in $\mathcal{V}^\perp$! ■

Showing that any vector in $(\mathcal{V}^\perp)^\perp$ is also in $\mathcal{V}$ is a bit harder, and we need a trick. (Actually, two tricks.) Any subspace $\mathcal{V}$ of $\mathbb{R}^p$, other than $\{\mathbf{0}_p\}$, has the form $\mathcal{V} = \mathbf{C}(B)$ for some matrix $B \in \mathbb{R}^{p \times r}$, where $r = \dim(\mathcal{V})$. Let $A = B^\mathsf{T}$, so $A \in \mathbb{R}^{r \times p}$ and $\mathcal{V} = \mathbf{C}(A^\mathsf{T})$. Then

$$\mathcal{V}^\perp = \mathbf{C}(A^\mathsf{T})^\perp = \mathbf{N}(A) \tag{4.5.4}$$

by Problem 4.5.6. And by Theorem 4.4.9, we can write any $\mathbf{x} \in \mathbb{R}^p$ as $\mathbf{x} = \mathbf{v} + \mathbf{w}$ for unique $\mathbf{v} \in \mathbf{C}(A^\mathsf{T}) = \mathcal{V}$ and $\mathbf{w} \in \mathbf{N}(A) = \mathcal{V}^\perp$. Here is what we have proved.

4.5.13 Lemma. *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. For each $\mathbf{x} \in \mathbb{R}^p$, there exist unique $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{V}^\perp$ such that $\mathbf{x} = \mathbf{v} + \mathbf{w}$. We might go so far as to write $\mathbb{R}^p = \mathcal{V} \oplus \mathcal{V}^\perp$ in the style of the orthogonal decompositions with fundamental subspaces.*

4.5.14 Problem (!). If $\mathcal{V}$ is a subspace of $\mathbb{R}^p$, prove that $\dim(\mathcal{V}^\perp) = p - \dim(\mathcal{V})$. [Hint: *write $\mathcal{V} = \mathbf{C}(A)$ where A has full column rank; use rank–nullity.*]

Here is why all of this jumping around between subspaces and matrices matters. Start with a subspace $\mathcal{V}$ of $\mathbb{R}^p$ and $\mathbf{x} \in (\mathcal{V}^\perp)^\perp$. We want to show $\mathbf{x} \in \mathcal{V}$. Lemma 4.5.13 tells us that $\mathbf{x} = \mathbf{v} + \mathbf{w}$ for some $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{V}^\perp$. If $\mathbf{w} = \mathbf{0}_p$, then $\mathbf{x} = \mathbf{v} \in \mathcal{V}$, as desired. Since $\mathbf{w} = \mathbf{x} - \mathbf{v}$, we have

$$\mathbf{w} \cdot \mathbf{w} = (\mathbf{x} - \mathbf{v}) \cdot \mathbf{w} = (\mathbf{x} \cdot \mathbf{w}) - (\mathbf{v} \cdot \mathbf{w}). \tag{4.5.5}$$

Since $\mathbf{x} \in (\mathcal{V}^\perp)^\perp$ and $\mathbf{w} \in \mathcal{V}^\perp$, we have $\mathbf{x} \cdot \mathbf{w} = 0$. And since $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{V}^\perp$, we have $\mathbf{v} \cdot \mathbf{w} = 0$. Thus from (4.5.5) we find $\mathbf{w} \cdot \mathbf{w} = 0$, so $\mathbf{w} = \mathbf{0}_p$.

4.5.15 Problem ($\star$). Prove that if $\mathcal{V}$ is a subspace of $\mathbb{R}^p$ and $\mathbf{v} \in \mathbb{R}^p$ with both $\mathbf{v} \in \mathcal{V}$ and $\mathbf{v} \in \mathcal{V}^\perp$, then $\mathbf{v} = \mathbf{0}_p$. Draw a picture illustrating this phenomenon in $\mathbb{R}^2$. How does this result imply Corollary 4.3.10?

This completes the second proof of Conjecture 4.5.8, which we now upgrade to a theorem.

4.5.16 Theorem. *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. Then $(\mathcal{V}^\perp)^\perp = \mathcal{V}$.*

It is also possible to prove the conjecture by focusing more on matrices first and counting dimensions. The value of the following proof is that it gives the essential characterization $\mathbf{C}(A) = \mathbf{N}(A^{\mathsf{T}})^{\perp}$ of the column space *directly* by working at the level of matrices, without passing to abstract subspaces.

4.5.17 Theorem. *Let* $A \in \mathbb{R}^{m\times n}$. *Then* $\mathbf{C}(A) = \mathbf{N}(A^{\mathsf{T}})^{\perp}$.

Proof. **1.** *The proof of* $\mathbf{w} \in \mathbf{C}(A) \Longrightarrow \mathbf{w} \in \mathbf{N}(A^{\mathsf{T}})^{\perp}$. Let $\mathbf{w} \in \mathbf{C}(A)$, so $\mathbf{w} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. We want to show $\mathbf{w} \in \mathbf{N}(A^{\mathsf{T}})^{\perp}$, so we need to prove that $\mathbf{w}\cdot\mathbf{v} = 0$ for all $\mathbf{v} \in \mathbf{N}(A^{\mathsf{T}})$.

That is, we want to show that if $A^{\mathsf{T}}\mathbf{v} = \mathbf{0}_n$, then $A\mathbf{x} \cdot \mathbf{v} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Assuming this, we have

$$A\mathbf{x} \cdot \mathbf{v} = \mathbf{x} \cdot A^{\mathsf{T}}\mathbf{v} = \mathbf{x} \cdot \mathbf{0}_n = 0.$$

2. *The proof of* $\mathbf{w} \in \mathbf{N}(A^{\mathsf{T}})^{\perp} \Longrightarrow \mathbf{w} \in \mathbf{C}(A)$. Now let $\mathbf{w} \in \mathbf{N}(A^{\mathsf{T}})^{\perp}$. We want to show $\mathbf{w} \in \mathbf{C}(A)$. This is harder: we know that if $A^{\mathsf{T}}\mathbf{v} = \mathbf{0}_n$, then $\mathbf{w}\cdot\mathbf{v} = 0$, and somehow we want to use this to summon up $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{w}$. And this is the fundamental problem of the course.

The trick here is contradiction. What goes wrong if $\mathbf{w} \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in \mathbb{R}^m$ with $A^{\mathsf{T}}\mathbf{v} = \mathbf{0}_n$ but $\mathbf{w} \notin \mathbf{C}(A)$? Consider the matrix $B = \begin{bmatrix} A & \mathbf{w} \end{bmatrix} \in \mathbb{R}^{m\times(n+1)}$. Since $\mathbf{w} \notin \mathbf{C}(A)$, the independent columns of B are the independent columns of A along with $\mathbf{w}$. So, if $\operatorname{rank}(A) = r$, then $\operatorname{rank}(B) = r + 1$.

Now we consider the transpose:

$$B^{\mathsf{T}} = \begin{bmatrix} A^{\mathsf{T}} \\ \mathbf{w}^{\mathsf{T}} \end{bmatrix} \in \mathbb{R}^{(n+1)\times m}.$$

Here we are thinking of $\mathbf{w} \in \mathbb{R}^m$ as an $m \times 1$ block in B, so $\mathbf{w}^{\mathsf{T}}$ is a $1 \times m$ block in B^{T}. We know $\operatorname{rank}(B^{\mathsf{T}}) = \operatorname{rank}(B) = r + 1$, too. And we know

$$\operatorname{rank}(B^{\mathsf{T}}) + \dim[\mathbf{N}(B^{\mathsf{T}})] = m,$$

so

$$\dim[\mathbf{N}(B^{\mathsf{T}})] = m - (r + 1).$$

We will show $\mathbf{N}(B^{\mathsf{T}}) = \mathbf{N}(A^{\mathsf{T}})$. (Recall that $A^{\mathsf{T}} \in \mathbb{R}^{n\times m}$, so both $\mathbf{N}(B^{\mathsf{T}})$ and $\mathbf{N}(A^{\mathsf{T}})$ are subspaces of $\mathbb{R}^m$, so it is reasonable to ask if they are equal.) Assuming this to be true, we have

$$m = \operatorname{rank}(A^{\mathsf{T}}) + \dim[\mathbf{N}(A^{\mathsf{T}})] = r + \dim[\mathbf{N}(B^{\mathsf{T}})],$$

and so

$$\dim[\mathbf{N}(B^{\mathsf{T}})] = m - r.$$

This is the contradiction and another illustration of the power of abstract dimension counting.

We conclude by proving $\mathbf{N}(B^{\mathsf{T}}) = \mathbf{N}(A^{\mathsf{T}})$.

(i) *The proof that* $\mathbf{v} \in \mathbf{N}(B^{\mathsf{T}}) \Longrightarrow \mathbf{v} \in \mathbf{N}(A^{\mathsf{T}})$. If $\mathbf{v} \in \mathbf{N}(B^{\mathsf{T}})$, then $B^{\mathsf{T}}\mathbf{v} = \mathbf{0}_{n+1}$ and so

$$\begin{bmatrix} \mathbf{0}_n \\ 0 \end{bmatrix} = \mathbf{0}_{n+1} = B^{\mathsf{T}}\mathbf{v} = \begin{bmatrix} A^{\mathsf{T}}\mathbf{v} \\ \mathbf{w}^{\mathsf{T}}\mathbf{v} \end{bmatrix} = \begin{bmatrix} A^{\mathsf{T}}\mathbf{v} \\ \mathbf{v}\cdot\mathbf{w} \end{bmatrix}.$$

Comparing components, we get $A^{\mathsf{T}}\mathbf{v} = \mathbf{0}_n$, so $\mathbf{v} \in \mathbf{N}(A^{\mathsf{T}})$.

(ii) *The proof that* $\mathbf{v} \in \mathbf{N}(A^{\mathsf{T}}) \Longrightarrow \mathbf{v} \in \mathbf{N}(B^{\mathsf{T}})$. Conversely, suppose $\mathbf{v} \in \mathbf{N}(A^{\mathsf{T}})$, so both $A^{\mathsf{T}}\mathbf{v} = \mathbf{0}_n$ and $\mathbf{w} \cdot \mathbf{v} = 0$. Then

$$B^{\mathsf{T}}\mathbf{v} = \begin{bmatrix} A^{\mathsf{T}}\mathbf{v} \\ \mathbf{v} \cdot \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_n \\ 0 \end{bmatrix} = \mathbf{0}_{n+1},$$

so $\mathbf{v} \in \mathbf{N}(B^{\mathsf{T}})$. ■

4.5.18 Problem ($\star$). Here is a proof of Conjecture 4.5.8 starting from Theorem 4.5.17. We already know that the conjecture is true if $\mathcal{V} = \{\mathbf{0}_p\}$, so assume that $\mathcal{V}$ is any subspace of $\mathbb{R}^p$ with $r := \dim(\mathcal{V}) \geq 1$. Then $\mathcal{V} = \mathbf{C}(A)$ for some $A \in \mathbb{R}^{p \times r}$. Combine (4.5.1) and Theorem 4.5.17 to get $(\mathbf{C}(A)^{\perp})^{\perp} = \mathbf{C}(A)$.

4.6. The fundamental theorem of linear algebra encodes everything.

Here is a summary of all of our work. This answers the question "What is missing beyond the null space or the column space?" and provides a complete overview of how a matrix in $\mathbb{R}^{m \times n}$ determines the structure of $\mathbb{R}^n$ and $\mathbb{R}^m$.

4.6.1 Theorem (Fundamental theorem of linear algebra). *Let* $A \in \mathbb{R}^{m \times n}$.

(i) $\mathbb{R}^n = \mathbf{N}(A) \oplus \mathbf{C}(A^{\mathsf{T}})$.

(ii) $\mathbb{R}^m = \mathbf{C}(A) \oplus \mathbf{N}(A^{\mathsf{T}})$.

(iii) $\mathbf{N}(A) = \mathbf{C}(A^{\mathsf{T}})^{\perp}$.

(iv) $\mathbf{C}(A) = \mathbf{N}(A^{\mathsf{T}})^{\perp}$.

(v) $\dim[\mathbf{N}(A)] = n - \operatorname{rank}(A)$.

(vi) $\operatorname{rank}(A) = \operatorname{rank}(A^{\mathsf{T}})$.

Content from Strang's *ILA* 6E. Figure 4.1 on p. 146 says all of this. Study the figure carefully and read the paragraph following its caption. I like to start reading the figure by beginning with $\mathbf{b}$, then tracking it back to $\mathbf{x}_r$ and $\mathbf{x}_n$. (The subscript n there is for "null space," not the n in $\mathbb{R}^n$.)

4.7. Projections make orthogonal decompositions explicit.

There is just one major problem with our fundamental theorem: all of these results are highly existential. We developed those existential results by starting with the null space and asking "What else is missing from $\mathbb{R}^n$?" Then we proceeded to the column space and gave a new characterization of it.

But the results are still existential. Specifically, while we can say that $\mathbb{R}^m = \mathbf{C}(A) \oplus \mathbf{N}(A^\mathsf{T})$ in the sense that each $\mathbf{b} \in \mathbb{R}^m$ can be written uniquely as $\mathbf{b} = \mathbf{v} + \mathbf{w}$ for some $\mathbf{v} \in \mathbf{C}(A)$ and $\mathbf{w} \in \mathbf{N}(A^\mathsf{T})$ and that $\mathbf{v} \cdot \mathbf{w} = 0$, how do we find those $\mathbf{v}$ and $\mathbf{w}$ explicitly and easily? First, we only need one of them. For if we know $\mathbf{b} = \mathbf{v} + \mathbf{w}$, then $\mathbf{w} = \mathbf{b} - \mathbf{v}$. So how do we get $\mathbf{v}$?

How do we do anything? *We multiply by a matrix.* Can we find $P \in \mathbb{R}^{m \times m}$ such that if $\mathbf{b} \in \mathbb{R}^m$, then $P\mathbf{b} \in \mathbf{C}(A)$ and $\mathbf{b} - P\mathbf{b} \in \mathbf{N}(A^\mathsf{T})$? Then we have $\mathbf{b} = P\mathbf{b} + (\mathbf{b} - P\mathbf{b})$ as our decomposition.

Content from Strang's *ILA* 6E. Read all of pp. 151–152 up to, but not including, "Projection Onto a Line." This is the mission statement of Section 4.2, and it's a very helpful overview of where we're going.

It turns out to be very helpful to assume that A has full column rank. This is not as huge a restriction as we might initially think. After all, $\mathbf{C}(A) = \mathbf{C}(C)$, where C is the matrix containing just the pivot columns of A, and C has full column rank. So, if we are going to understand the decomposition $\mathbb{R}^m = \mathbf{C}(A) \oplus \mathbf{N}(A^\mathsf{T})$, we may as well do it when A has full column rank.

Assume $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(A) = n$ and let $\mathbf{b} \in \mathbb{R}^m$. We want to find $\mathbf{v} \in \mathbf{C}(A)$ and $\mathbf{w} \in \mathbf{N}(A^\mathsf{T})$ such that

$$\mathbf{b} = \mathbf{v} + \mathbf{w} \quad \text{and} \quad \mathbf{v} \cdot \mathbf{w} = 0.$$

Since $\mathbf{v} \in \mathbf{C}(A)$, we have $\mathbf{v} = A\widehat{\mathbf{x}}$ for some $\widehat{\mathbf{x}} \in \mathbb{R}^n$. Then $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbf{N}(A^\mathsf{T})$ must satisfy

$$\begin{cases} \mathbf{b} = A\widehat{\mathbf{x}} + \mathbf{w} \\ \mathbf{a}_j \cdot \mathbf{w} = 0, \ j = 1, \ldots, n. \end{cases} \tag{4.7.1}$$

That second (set of) equation(s) is the orthogonality of $\mathbf{w}$ to everything in $\mathbf{C}(A)$, equivalently, to the columns of A.

This reduces to n equations:

$$0 = \mathbf{a}_j \cdot \mathbf{w} = \mathbf{a}_j \cdot (\mathbf{b} - A\widehat{\mathbf{x}}), \ j = 1, \ldots, n.$$

And now for the trick: rewrite $\mathbf{a}_j = A\mathbf{e}_j$, so

$$0 = (A\mathbf{e}_j) \cdot (\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{e}_j \cdot A^\mathsf{T}(\mathbf{b} - A\widehat{\mathbf{x}}), \ j = 1, \ldots, n.$$

The powerful Problem 4.5.1 implies that

$$A^\mathsf{T}(\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{0}_n,$$

which rearranges to

$$A^\mathsf{T}A\widehat{\mathbf{x}} = A^\mathsf{T}\mathbf{b}.$$

If only $A^\mathsf{T}A$ were invertible, we could peel it off to solve for $\widehat{\mathbf{x}}$:

$$\widehat{\mathbf{x}} = (A^\mathsf{T}A)^{-1}A^\mathsf{T}\mathbf{b}.$$

If only. Then we would have

$$\mathbf{v} = A\widehat{\mathbf{x}} = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\mathbf{b},$$

and so putting

$$P := A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$$

would give $P\mathbf{b} \in \mathbf{C}(A)$, $\mathbf{b} - P\mathbf{b} \in \mathbf{N}(A^{\mathsf{T}})$, and $\mathbf{b} = P\mathbf{b} + (\mathbf{b} - P\mathbf{b})$.

Good news: $A^{\mathsf{T}}A$ is invertible here, and this is exactly why we assumed that A has full column rank.

4.7.1 Lemma. *Let $A \in \mathbb{R}^{m\times n}$ with* $\operatorname{rank}(A) = n$. *Then $A^{\mathsf{T}}A$ is invertible.*

Proof. We might initially think $A^{\mathsf{T}}A$ = independent rows in A^{T} dotted with independent columns in A has to give us something good. It does, but the trick is to show $\mathbf{N}(A^{\mathsf{T}}A) = \{\mathbf{0}_n\}$. For if $A^{\mathsf{T}}A\mathbf{x} = \mathbf{0}_n$, then

$$0 = \mathbf{x}\cdot\mathbf{0}_n = \mathbf{x}\cdot(A^{\mathsf{T}}A\mathbf{x}) = (A\mathbf{x})\cdot(A\mathbf{x}),$$

and so $A\mathbf{x} = \mathbf{0}_n$, thus $\mathbf{x} \in \mathbf{N}(A)$. Since A has full column rank, $\mathbf{N}(A) = \{\mathbf{0}_n\}$, so $\mathbf{x} = \mathbf{0}_n$. ■

Content from Strang's *ILA* 6E. This lemma is proved on p. 157. Read the warning at the top of the page and then the calculations at the bottom of the page of how this breaks when A has dependent columns.

4.7.2 Problem ($\star$). Let $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = m$. Show that AA^{T} is invertible. [Hint: *either mimic the strategy of Lemma 4.7.1 or apply that lemma to $B^{\mathsf{T}}B$ with $B = A^{\mathsf{T}}$.*]

4.7.3 Problem ($\star$). Let $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) \ge 1$ have the CR-factorization $A = CR$. Prove that $C^{\mathsf{T}}C$, RR^{T}, and $C^{\mathsf{T}}AR^{\mathsf{T}}$ are invertible.

Bad news: getting to the application of Lemma 4.7.1 was a lot of working backward, and some gaps need filling.

4.7.4 Problem ($\star$). Let $A \in \mathbb{R}^{m\times n}$ have full column rank and set

$$P_A := A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}.$$

(i) Show that if $\mathbf{v} \in \mathbf{C}(P_A)$, then $\mathbf{v} \in \mathbf{C}(A)$. [Hint: *use the formula for P_A.*]

(ii) Show that if $\mathbf{b} \in \mathbf{C}(A)$, then $P_A\mathbf{b} = \mathbf{b}$, and so $\mathbf{b} \in \mathbf{C}(P_A)$. [Hint: *again, use the formula for P_A.*] This, combined with the previous result, gives $\mathbf{C}(A) = \mathbf{C}(P_A)$.

(iii) Show that $P_A^2 = P_A$.

(iv) Show that $P_A^{\mathsf{T}} = P_A$.

(v) Justify each of the following equalities:

$$\mathbf{N}(P_A) = \mathbf{N}(P_A^{\mathsf{T}}) = \mathbf{C}(P_A)^{\perp} = \mathbf{C}(A)^{\perp} = \mathbf{N}(A^{\mathsf{T}}). \tag{4.7.2}$$

(vi) Show that $\mathbf{b} - P_A\mathbf{b} \in \mathbf{N}(P_A)$ for each $\mathbf{b} \in \mathbb{R}^m$. [Hint: *compute* $P_A(\mathbf{b} - P_A\mathbf{b})$. *What is* P_A^2*?*] Conclude that $\mathbf{b} - P_A\mathbf{b} \in \mathbf{N}(A^{\mathsf{T}})$.

The fruit of this problem is that we can write any $\mathbf{b} \in \mathbb{R}^m$ as

$$\mathbf{b} = \mathbf{v} + \mathbf{w}, \qquad \mathbf{v} = P_A\mathbf{b}, \qquad \mathbf{w} = \mathbf{b} - P_A\mathbf{b}, \tag{4.7.3}$$

and we will have $\mathbf{v} \in \mathbf{C}(A)$ and $\mathbf{w} \in \mathbf{N}(A^{\mathsf{T}})$. Lemma 4.3.12 assures us that this decomposition is unique.

Content from Strang's *ILA* 6E. Pages 155–156 develop all of this. I don't think memorizing equations (5), (6), and (7) is a good idea, or even memorizing the structure of our P_A above. I think it's more important to be able to *replicate* the derivation of P_A on your own. Check Worked Example 4.2 A on p. 158.

4.7.5 Example. Let

$$B = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix},$$

so B has full column rank. Then we can write any $\mathbf{b} \in \mathbb{R}^3$ uniquely as $\mathbf{b} = P_B\mathbf{b} + \mathbf{w}$, $\mathbf{w} = \mathbf{b} - P_B\mathbf{b} \in \mathbf{N}(B^{\mathsf{T}})$, where $P_B = B(B^{\mathsf{T}}B)^{-1}B^{\mathsf{T}}$.

Now we figure out P_B. We compute

$$B^{\mathsf{T}}B = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix},$$

$$(B^{\mathsf{T}}B)^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1 \end{bmatrix},$$

and

$$P_B = B(B^{\mathsf{T}}B)^{-1}B^{\mathsf{T}}$$

$$= \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/5 & 2/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/5 & 2/5 & 0 \\ 2/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It was nice that $B^{\mathsf{T}}B$ turned out to be diagonal, as that made computing its inverse very easy.

4.7.6 Problem (!). With B from the previous example, show that

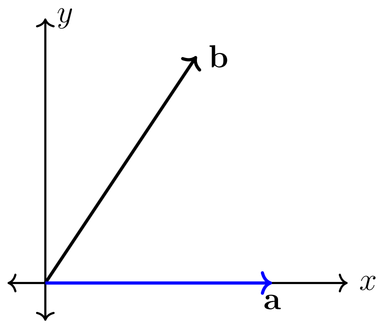
$$P_B\mathbf{b} \cdot \mathbf{e}_2 = 2(P_B\mathbf{b} \cdot \mathbf{e}_1).$$

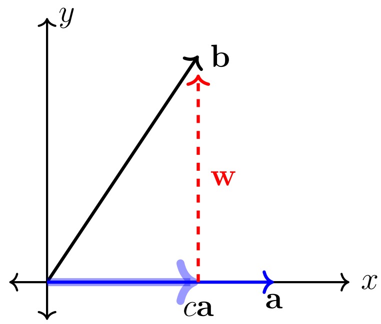
How does this relate to Problem 3.10.2

The situation in which $A \in \mathbb{R}^{m \times n}$ has only one column, so $n = 1$, is worth considering. Suppose that $A = \begin{bmatrix}\mathbf{a}\end{bmatrix}$, where $\mathbf{a} \in \mathbb{R}^m$ is nonzero. Given $\mathbf{b} \in \mathbb{R}^m$, there are (necessarily unique) $\mathbf{v} \in \mathbf{C}(A) = \text{span}(\mathbf{a})$ and $\mathbf{w} \in \mathbf{N}(A^{\mathsf{T}})$ such that $\mathbf{b} = \mathbf{v} + \mathbf{w}$ with $\mathbf{v} \cdot \mathbf{w} = 0$. Since $\mathbf{v} \in \text{span}(\mathbf{a})$, we can write $\mathbf{v} = c\mathbf{a}$ for some $c \in \mathbb{R}$; we will have $\mathbf{v} \cdot \mathbf{w} = 0$ if $\mathbf{a} \cdot \mathbf{w} = 0$. Then

$$\mathbf{b} = c\mathbf{a} + \mathbf{w}.$$

Here is a picture when $m = 2$ and $\mathbf{a}$ is a multiple of $\mathbf{e}_1 = (1, 0)$.





Now we slightly reinterpret the formula for the projection $P_A = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$ onto $\mathbf{C}(A) = \text{span}(\mathbf{a})$. If

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \in \mathbb{R}^m = \mathbb{R}^{m \times 1},$$

then

$$A^{\mathsf{T}}A = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots a_m \end{bmatrix} = \begin{bmatrix} a_1^2 + \cdots + a_m^2 \end{bmatrix} = \begin{bmatrix}\mathbf{a} \cdot \mathbf{a}\end{bmatrix} \in \mathbb{R}^{1 \times 1}.$$

We follow the convention established long ago in Undefinition 1.4.1 that 1×1 matrices are just numbers, and so we conclude here that

$$A^{\mathsf{T}}A = \mathbf{a} \cdot \mathbf{a}.$$

In particular, since $\mathbf{a} \neq \mathbf{0}_m$, we have $\mathbf{a} \cdot \mathbf{a} \neq 0$, and so

$$(A^\mathsf{T} A)^{-1} = \frac{1}{\mathbf{a} \cdot \mathbf{a}}.$$

We simplify notation further by putting

$$\mathbf{a}^\mathsf{T} = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix} \in \mathbb{R}^{1 \times m}.$$

We only defined the transpose for a matrix, not a vector, earlier, so we do need to single out this special case. We also write $\mathbf{a}$ in lieu of $\begin{bmatrix} \mathbf{a} \end{bmatrix} \in \mathbb{R}^{m \times 1}$. Then

$$P_A = A(A^\mathsf{T} A)^{-1} A^\mathsf{T} = \frac{1}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a}\mathbf{a}^\mathsf{T} \in \mathbb{R}^{m \times m}. \tag{4.7.4}$$

4.7.7 Problem (!). Continuing to identify $m \times 1$ matrices with vectors in $\mathbb{R}^m$ and 1×1 matrices with real numbers, check that if $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^m$, then $\mathbf{x}^\mathsf{T}\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$, but describe the *entries* of the $m \times m$ matrix $\mathbf{x}\mathbf{y}^\mathsf{T}$.

What is the action of P_A as defined in (4.7.4)? We have

$$P_A \mathbf{b} = \left(\frac{1}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a}\mathbf{a}^\mathsf{T} \right) \mathbf{b} = \left(\frac{1}{\mathbf{a} \cdot \mathbf{a}} \right) \mathbf{a}(\mathbf{a}^\mathsf{T}\mathbf{b}) = \frac{1}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a}(\mathbf{a} \cdot \mathbf{b}) = \left(\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}} \right) \mathbf{a}. \tag{4.7.5}$$

The last equality presented us with the uncomfortable situation of scalar multiplication with the scalar on the right of the vector. (Remember that we always write $c\mathbf{a}$, never $\mathbf{a}c$.) From (4.7.5), the action of P_A here is really "take a dot product and do scalar multiplication," which should sound like matrix-vector multiplication, but a little easier.

We could also obtain (4.7.5) without using the prior formula for P_A. The situation is that we have $\mathbf{a} \in \mathbb{R}^m$ with $\mathbf{a} \neq \mathbf{0}_n$, and we want to write $\mathbf{b} \in \mathbb{R}^m$ as $\mathbf{b} = c\mathbf{a} + \mathbf{w}$ with $\mathbf{a} \cdot \mathbf{w} = 0$. Now we have two equations and two unknowns:

$$\begin{cases} \mathbf{b} = c\mathbf{a} + \mathbf{w} \\ \mathbf{a} \cdot \mathbf{w} = 0. \end{cases} \tag{4.7.6}$$

This is (4.7.1) with $n = 1$, $A = \begin{bmatrix} \mathbf{a} \end{bmatrix} = \mathbf{a}$, and $\widehat{\mathbf{x}} = c$.

As before, a little algebraic trickery will reduce (4.7.6) to one equation: rewrite

$$\mathbf{w} = \mathbf{b} - c\mathbf{a}$$

and plug in to get

$$0 = \mathbf{a} \cdot (\mathbf{b} - c\mathbf{a}).$$

Rearrange a little:

$$0 = (\mathbf{a} \cdot \mathbf{b}) - c(\mathbf{a} \cdot \mathbf{a}),$$

and a little more:

$$c(\mathbf{a} \cdot \mathbf{a}) = \mathbf{a} \cdot \mathbf{b},$$

and divide:

$$c = \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}}.$$

This division is perfectly legal since $\mathbf{a} \neq \mathbf{0}_m$, and therefore $\mathbf{a} \cdot \mathbf{a} \neq 0$.

We worked backwards, so we should check our work. Certainly

$$\left(\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}}\right) \mathbf{a} \in \operatorname{span}(\mathbf{a}).$$

4.7.8 Problem (!). Let

$$\mathbf{w} = \mathbf{b} - \left(\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}}\right) \mathbf{a}.$$

Check that $\mathbf{a} \cdot \mathbf{w} = 0$, so $\mathbf{w} \in \mathbf{C}(A)^{\perp} = \mathbf{N}(A^{\mathsf{T}})$.

4.7.9 Example. We check how this one-dimensional result respects our intuition. Say $\mathbf{a} = \mathbf{e}_1$ in $\mathbb{R}^2$. Of course, we expect

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = b_1\mathbf{e}_1 + \begin{bmatrix} 0 \\ b_2 \end{bmatrix}.$$

We compute

$$\left(\frac{\mathbf{e}_1 \cdot \mathbf{b}}{\mathbf{e}_1 \cdot \mathbf{e}_1}\right) \mathbf{e}_1 = \frac{b_1}{1}\mathbf{e}_1 = b_1\mathbf{e}_1.$$

How nice it was that $\mathbf{e}_1 \cdot \mathbf{e}_1 = 1$.

Content from Strang's *ILA* 6E. Read "Projection Onto a Line" from pp. 152–154. Check Examples 1 and 2.

The matrix P_A that we cooked up to achieve the explicit decomposition (4.7.3) deserves a special name.

4.7.10 Definition. *Let $P \in \mathbb{R}^{m \times m}$.*

(i) *P is a* ***PROJECTION*** *if $P^2 = P$.*

(ii) *P is an* ***ORTHOGONAL PROJECTION*** *if $P^2 = P$ and $P^{\mathsf{T}} = P$.*

(iii) *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^m$. The matrix P is an* ***ORTHOGONAL PROJECTION ONTO*** *$\mathcal{V}$ if P is an orthogonal projection with $\mathbf{C}(P) = \mathcal{V}$.*

4.7.11 Example. **(i)** The zero matrix in $\mathbb{R}^{m \times m}$ and I_m are both orthogonal projections.

(ii) Problem 4.7.4 shows that $P_A = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$ is an orthogonal projection onto $\mathbf{C}(A)$ when $A \in \mathbb{R}^{m \times n}$ has full column rank.

(iii) Let $\mathcal{V}$ be a subspace of $\mathbb{R}^m$. If $\mathcal{V} = \{\mathbf{0}_m\}$, then the matrix $P \in \mathbb{R}^{m\times m}$ whose entries are all 0 is an orthogonal projection onto $\mathcal{V}$. Otherwise, let $r = \dim(\mathcal{V})$ and let $A \in \mathbb{R}^{m\times r}$ be a matrix whose columns are a basis for $\mathcal{V}$. Then $P_A = A(A^\mathsf{T}A)^{-1}A^\mathsf{T}$ is an orthogonal projection onto $\mathcal{V}$.

4.7.12 Problem (!). Without doing any matrix calculations, in $\mathbb{R}^3$, what do you expect an orthogonal projection onto $\mathrm{span}(\mathbf{e}_1, \mathbf{e}_2)$ to be? Now do those calculations.

4.7.13 Problem ($\star$). Let $\mathcal{V}$ be a subspace of $\mathbb{R}^{m\times m}$ and let $P \in \mathbb{R}^{m\times m}$ be an orthogonal projection onto $\mathcal{V}$.

(i) Show that if $\mathbf{v} \in \mathcal{V}$, then $P\mathbf{v} = \mathbf{v}$.

(ii) Explain why $I_m - P$ is an orthogonal projection onto $\mathcal{V}^\perp$.

(iii) Show that if $\mathbf{w} \in \mathcal{V}^\perp$, then $P\mathbf{w} = \mathbf{0}_m$.

After all of our work, we probably want to call P_A *the* orthogonal projection *onto* $\mathbf{C}(A)$ when A has full column rank. Is it unique? What an insult it would be if it were not. This is asking if there is only one $P \in \mathbb{R}^{m\times m}$ such that $P^2 = P$, $P^\mathsf{T} = P$, and $\mathbf{C}(P) = \mathbf{C}(A)$. And this is true.

4.7.14 Problem (!). Let $P \in \mathbb{R}^{m\times m}$ be an orthogonal projection. Show that for each $\mathbf{x} \in \mathbb{R}^m$, there is $\mathbf{w} \in \mathbf{N}(P)$ such that

$$\mathbf{x} = P\mathbf{x} + \mathbf{w}.$$

[Hint: *the equality suggests taking* $\mathbf{w} = \mathbf{x} - P\mathbf{x}$*; show that with this definition of* $\mathbf{w}$*, we have* $\mathbf{w} \in \mathbf{N}(P)$.]

4.7.15 Problem (+). Let P_1, $P_2 \in \mathbb{R}^{m\times m}$ be orthogonal projections with $\mathbf{C}(P_1) = \mathbf{C}(P_2)$. We prove $P_1 = P_2$ by showing $P_1\mathbf{x} = P_2\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^m$. (*What things do defines what things are.*)

(i) Show that $\mathbf{N}(P_1) = \mathbf{N}(P_2)$. [Hint: $\mathbf{N}(P) = \mathbf{C}(P^\mathsf{T})^\perp$ *for any* $P \in \mathbb{R}^{m\times m}$*. What does the equality* $\mathbf{C}(P_1) = \mathbf{C}(P_2)$ *and the assumption that* P_1 *and* P_2 *are orthogonal projections say about* $\mathbf{C}(P_1^\mathsf{T})$ *and* $\mathbf{C}(P_2^\mathsf{T})$ *and thus about their orthogonal complements?*]

(ii) Use Problem 4.7.14 to explain why we can write any $\mathbf{x} \in \mathbb{R}^m$ as both

$$\mathbf{x} = P_1\mathbf{x} + \mathbf{w}_1 \quad \text{and} \quad \mathbf{x} = P_2\mathbf{x} + \mathbf{w}_2$$

for some $\mathbf{w}_1 \in \mathbf{N}(P_1)$ and $\mathbf{w}_2 \in \mathbf{N}(P_2)$. Conclude that

$$P_1\mathbf{x} - P_2\mathbf{x} \in \mathbf{C}(P_1) \quad \text{and} \quad P_1\mathbf{x} - P_2\mathbf{x} \in \mathbf{N}(P_1).$$

Then invoke Problem 4.5.15.

We have so far found orthogonal projections onto column spaces of matrices with full column rank only, but we can practically accomplish the orthogonal decomposition $\mathbb{R}^m = \mathbf{C}(A) \oplus \mathbf{N}(A^\mathsf{T})$ for $A \in \mathbb{R}^{m \times n}$ whether or not A has full column rank. The decomposition for the zero matrix is easy, since then $\mathbf{C}(A) = \{\mathbf{0}_m\}$ and $\mathbf{N}(A^\mathsf{T}) = \mathbb{R}^m$. So, assume $\operatorname{rank}(A) = r \geq 1$. Let $B \in \mathbb{R}^{m \times r}$ be a matrix whose columns form a basis for $\mathbf{C}(A)$, so $\mathbf{C}(A) = \mathbf{C}(B)$ and B has full column rank. (For example, we could choose $B = C$ from the CR-factorization, but this may not be the ideal choice.) The orthogonal projection $P_B = B(B^\mathsf{T}B)^{-1}B^\mathsf{T}$ is defined since B has full column rank. Every $\mathbf{b} \in \mathbb{R}^m$ can therefore be written in the form $\mathbf{b} = P_B\mathbf{b} + \mathbf{w}$, where $P_B\mathbf{b} \in \mathbf{C}(B) = \mathbf{C}(A)$ and $\mathbf{w} \in \mathbf{N}(B^\mathsf{T})$. And

$$\mathbf{N}(B^\mathsf{T}) = \mathbf{C}(B)^\perp = \mathbf{C}(A)^\perp = \mathbf{N}(A^\mathsf{T}).$$

Thus every $\mathbf{b} \in \mathbb{R}^m$ has the form $\mathbf{b} = P_B\mathbf{b} + \mathbf{w}$ with $P_B\mathbf{b} \in \mathbf{C}(A)$ and $\mathbf{w} \in \mathbf{N}(A^\mathsf{T})$.

4.7.16 Problem (!). Let

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$

Use Problem 3.10.2 and Example 4.7.5 to find an orthogonal projection $P \in \mathbb{R}^{3 \times 3}$ such that every $\mathbf{b} \in \mathbb{R}^3$ has the form $\mathbf{b} = P\mathbf{b} + \mathbf{w}$ for some $\mathbf{w} \in \mathbf{N}(A^\mathsf{T})$.

We now possess a much deeper understanding of how a matrix induces *structure* from the decompositions $\mathbb{R}^n = \mathbf{N}(A) \oplus \mathbf{C}(A^\mathsf{T})$ and $\mathbb{R}^m = \mathbf{C}(A) \oplus \mathbf{N}(A^\mathsf{T})$ for $A \in \mathbb{R}^{m \times n}$, and how to perform those decompositions via matrix multiplication. We also have the characterization $\mathbf{C}(A) = \mathbf{N}(A^\mathsf{T})^\perp$ and the resulting "solvability conditions" from the Fredholm alternative (Problem 4.5.11).

4.8. Least squares approximates solutions to unsolvable problems.

What else do we gain from these results? There has been something of a dichotomy in our approach to linear systems. Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Either we can solve $A\mathbf{x} = \mathbf{b}$ (uniquely or not) or we cannot We have focused on the solving part. Otherwise, if $A\mathbf{x} = \mathbf{b}$ has no solution, what is the point in talking about it?

Very often in life, mathematically or otherwise, we cannot solve the problems that we face. The next best thing is to solve an easier problem. (If the question is too hard, give up and ask a different question.) If we cannot solve $A\mathbf{x} = \mathbf{b}$, could we find $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that $A\widehat{\mathbf{x}}$ is "as close" to $\mathbf{b}$ as possible? This hinges in part on having a suitable definition of "close" for vectors.

Any vector of the form $A\widehat{\mathbf{x}}$ will be in $\mathbf{C}(A)$. Desiring $A\widehat{\mathbf{x}}$ to be as close to $\mathbf{b}$ as possible is then equivalent to picking $\mathbf{p} \in \mathbf{C}(A)$ such that $\mathbf{p}$ is as close to $\mathbf{b}$ as possible (for if $\mathbf{p} \in \mathbf{C}(A)$, then $\mathbf{p}$ has the form $\mathbf{p} = A\widehat{\mathbf{x}}$). That is, we approximate the unsolvable problem $A\mathbf{x} = \mathbf{b}$ by $A\widehat{\mathbf{x}} = \mathbf{p}$. Once we know how to find $\mathbf{p}$, then we can use prior techniques to solve $A\widehat{\mathbf{x}} = \mathbf{p}$. Finding $\mathbf{p}$ is something of a miracle, and we have already done most of the hard work with projections.

We first make precise these notions of "close" and "approximation." Both require a new concept: the notion of *size*, which is really a notion of *length*. The following definition generalizes the notion that the length of the line segment in two dimensions from the origin $(0,0)$ to a point (x,y) is $\sqrt{x^2+y^2}$.

4.8.1 Definition. *The* ***NORM*** *of* $\mathbf{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m$ *is*

$$\|\mathbf{v}\| := \left(v_1^2 + \cdots + v_m^2\right)^{1/2} = \sqrt{\mathbf{v}\cdot\mathbf{v}}.$$

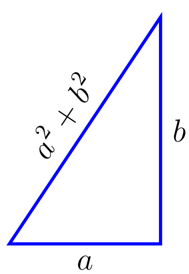
4.8.2 Example. If $\mathbf{v} = (1,2,3)$, then

$$\|\mathbf{v}\| = \sqrt{1^2+2^2+3^2} = \sqrt{1+4+9} = \sqrt{14}.$$

4.8.3 Problem (!). Let $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^m$. Prove that $\|c\mathbf{v}\| = |c|\,\|\mathbf{v}\|$ and interpret this as a statement about "stretching" vectors.

Content from Strang's *ILA* 6E. Reread all of p. 9 right now. There are plenty of other meaningful ways of measuring the length of a vector in $\mathbb{R}^m$ that we won't need. You might enjoy reading pp. 355–356 up to and including Figure 9.8.

Length and orthogonality interact in a helpful way. We know this already because we believe the Pythagorean theorem, which the definition of $\|\cdot\|$ is basically designed to respect.



4.8.4 Theorem (Pythagorean theorem). *Let* $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^m$ *with* $\mathbf{v}\cdot\mathbf{w} = 0$. *Then*

$$\|\mathbf{v}+\mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2.$$

4.8.5 Problem (!). Prove it. [Hint: *use the definition* $\|\mathbf{x}\|^2 = \mathbf{x}\cdot\mathbf{x}$ *to compute* $\|\mathbf{v}+\mathbf{w}\|^2$ *and get* $\mathbf{v}\cdot\mathbf{w} = 0$ *to show up somewhere.*]

4.8.6 Problem ($\star$). Let $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^m$ with $\mathbf{v}\cdot\mathbf{w} = 0$. Prove that

$$\|\mathbf{v}\| \le \|\mathbf{v}+\mathbf{w}\|.$$

Draw a picture illustrating this in $\mathbb{R}^2$. [Hint: *square both sides and use the Pythagorean theorem to get an expression for* $\|\mathbf{v}+\mathbf{w}\|^2$.]

Here is how we use this new tool of the norm. We think that two vectors $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^m$ are "close" if the difference $\|\mathbf{v} - \mathbf{w}\|$ is "small."

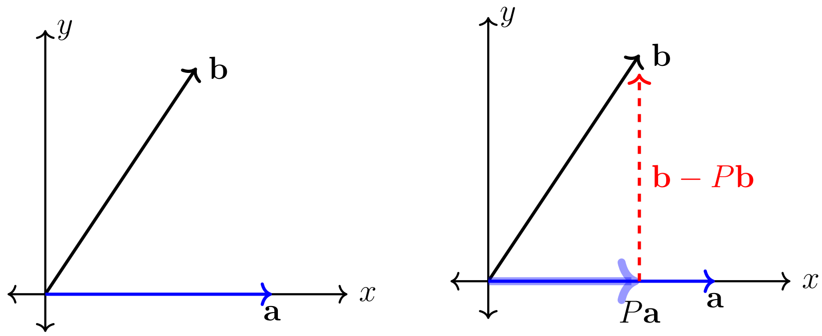
4.8.7 Remark. *And what exactly does "small" mean? Say $\|\mathbf{v}\| < \epsilon$ for some $\epsilon > 0$. Then since the square root is increasing,*

$$|v_j| = \sqrt{v_j^2} \le \sqrt{v_1^2 + \cdots + v_m^2} = \|\mathbf{v}\| < \epsilon.$$

So if $\|\mathbf{v}\|$ is "small" in the sense that it is less than some threshold $\epsilon > 0$, then each component v_j is "small" in the same way: $|v_j| < \epsilon$ for all j. A vector with small components is probably a small vector.

Given $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ with $\mathbf{b} \notin \mathsf{C}(A)$, can we find $\mathbf{p} \in \mathsf{C}(A)$ such that $\mathbf{b}$ and $\mathbf{b}$ are "close"? Then maybe solving $A\widehat{\mathbf{x}} = \mathbf{p}$ will be an adequate substitute for failing to solve $A\mathbf{x} = \mathbf{b}$.

Gloriously, finding this $\mathbf{p}$ is quite easy when we have the orthogonal projection $P_A\mathbf{b}$. Here is a picture that we drew before, more or less, when finding the projection onto the column space of a matrix with a single column.



Here

$$A = \begin{bmatrix}\mathbf{a}\end{bmatrix} \in \mathbb{R}^{m \times 1}, \qquad \mathbf{a} \neq \mathbf{0}_m, \qquad \text{and} \qquad P = P_A = P_{\begin{bmatrix}\mathbf{a}\end{bmatrix}}.$$

Hopefully the picture makes it clear that the closest vector in $\mathsf{C}(\begin{bmatrix}\mathbf{a}\end{bmatrix}) = \operatorname{span}(\mathbf{a})$ to $\mathbf{b}$ is $P\mathbf{b}$.

Here is the general result. Going forward, as before, we assume that $A \in \mathbb{R}^{m \times n}$ has full column rank. This is what makes the projection P_A exist. And, as before, we will handle the case $\operatorname{rank}(A) < n$ later.

Take any $\mathbf{v} \in \mathsf{C}(A)$. The following inequality encodes the idea that $P_A\mathbf{b}$ is the closest vector in $\mathsf{C}(A)$ to $\mathbf{b}$:

$$\|\mathbf{b} - P_A\mathbf{b}\| \le \|\mathbf{b} - \mathbf{v}\| \text{ for all } \mathbf{v} \in \mathsf{C}(A). \tag{4.8.1}$$

Our goal is to prove (4.8.1).

This inequality is equivalent to

$$\|\mathbf{b} - \mathbf{v}\|^2 \ge \|\mathbf{b} - P_A\mathbf{b}\|^2, \tag{4.8.2}$$

and so this is what we will really prove. The upshot to (4.8.2) is that the square roots from the norm are gone.

We can make $P_A\mathbf{b}$ show up on the left side of (4.8.2) by adding and subtracting:

$$\|\mathbf{b}-\mathbf{v}\|^2 = \|\mathbf{b}-P_A\mathbf{b}+P_A\mathbf{b}-\mathbf{v}\|^2. \tag{4.8.3}$$

Now we group terms cleverly:

$$\|\mathbf{b}-P_A\mathbf{b}+P_A\mathbf{b}-\mathbf{v}\|^2 = \|(\mathbf{b}-P_A\mathbf{b})+(P_A\mathbf{b}-\mathbf{v})\|^2. \tag{4.8.4}$$

We know $\mathbf{b}-P_A\mathbf{b} \in \mathbf{N}(A^\mathsf{T})$ by (4.7.2), $P_A\mathbf{b} \in \mathbf{C}(A)$, and $\mathbf{v} \in \mathbf{C}(A)$. So, $P_A\mathbf{b}-\mathbf{v} \in \mathbf{C}(A) = \mathbf{N}(A^\mathsf{T})^\perp$. That is, $\mathbf{b}-P_A\mathbf{b}$ and $P_A\mathbf{b}-\mathbf{v}$ are orthogonal. The Pythagorean theorem implies

$$\|(\mathbf{b}-P_A\mathbf{b})+(P_A\mathbf{b}-\mathbf{v})\|^2 = \|\mathbf{b}-P_A\mathbf{b}\|^2 + \|P_A\mathbf{b}-\mathbf{v}\|^2 \ge \|\mathbf{b}-P_A\mathbf{b}\|^2. \tag{4.8.5}$$

Combining (4.8.3), (4.8.4), and (4.8.5) gives (4.8.2).

4.8.8 Theorem (Least squares). *Let $A \in \mathbb{R}^{m\times n}$ have full column rank. Then for any $\mathbf{b} \in \mathbb{R}^m$, the closest vector in $\mathbf{C}(A)$ to $\mathbf{b}$ is $P_A\mathbf{b}$:*

$$\|\mathbf{b}-P_A\mathbf{b}\| \le \|\mathbf{b}-\mathbf{v}\| \ \textit{ for all } \mathbf{v} \in \mathbf{C}(A). \tag{4.8.6}$$

No other vector in $\mathbf{C}(A)$ is "as close" to $\mathbf{b}$ as $P_A\mathbf{b}$ in the sense that the inequality in (4.8.6) is strict for $\mathbf{v} \neq P_A\mathbf{b}$. Moreover, with $\widehat{\mathbf{x}} := (A^\mathsf{T}A)^{-1}A^\mathsf{T}\mathbf{b}$, we have

$$\|A\widehat{\mathbf{x}}-\mathbf{b}\| \le \|A\mathbf{x}-\mathbf{b}\| \ \textit{ for all } \mathbf{x} \in \mathbb{R}^n. \tag{4.8.7}$$

*The **LEAST SQUARES SOLUTION** $\widehat{\mathbf{x}}$ is the best "approximate solution" to the (possibly unsolvable) problem $A\mathbf{x} = \mathbf{b}$, and $\widehat{\mathbf{x}}$ solves*

$$A\widehat{\mathbf{x}} = P_A\mathbf{b}, \tag{4.8.8}$$

which is the "best approximation to" the (possibly unsolvable) problem $A\mathbf{x} = \mathbf{b}$.

Proof. We proved this in the discussion above, but here is a terser recap. Let $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbf{C}(A)$. Then

$$\begin{aligned}\|\mathbf{b}-\mathbf{v}\|^2 &= \|(\mathbf{b}-P_A\mathbf{b})+(P_A\mathbf{b}-\mathbf{v})\|^2 \\ &= \|\mathbf{b}-P_A\mathbf{b}\|^2 + \|P_A\mathbf{b}-\mathbf{v}\|^2 \\ &\ge \|\mathbf{b}-P_A\mathbf{b}\|^2.\end{aligned}$$

If $\mathbf{v} \neq P_A\mathbf{b}$, then $\|P_A\mathbf{b}-\mathbf{v}\|^2 > 0$, and so the inequality above is strict, which gives the strict inequality in (4.8.6).

The second inequality (4.8.7) is just (4.8.6) with

$$P_A\mathbf{b} = A\widehat{\mathbf{x}}, \qquad \widehat{\mathbf{x}} := (A^\mathsf{T}A)^{-1}A^\mathsf{T}\mathbf{b},$$

and $\mathbf{v} \in \mathbf{C}(A)$ replaced by $\mathbf{v} = A\mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^n$. ■

We use the phrase "least squares solution" because the sum of the squares in $\|A\widehat{\mathbf{x}}-\mathbf{b}\|$ is the smallest of all sums of squares of the form $\|A\mathbf{x}-\mathbf{b}\|$. More euphemistically, we might write

$$\|A\widehat{\mathbf{x}}-\mathbf{b}\| = \min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x}-\mathbf{b}\|.$$

4.8.9 Problem (!). If $A \in \mathbb{R}^{m\times m}$ is invertible, what are P_A and $\widehat{\mathbf{x}}$? Are you surprised?

Content from Strang's *ILA* 6E. Read p. 163 up to and including the box before Example 1. Then read "Minimizing the Error" on pp. 164–165. Skip the "By calculus" section on pp. 165–166 if you haven't taken multivariable calculus. Then read "The Big Picture for Least Squares" on pp. 166–167. Spend some time contrasting Figure 4.7 on p. 166 with Figure 4.1 back on p. 146. How is $\mathbf{b}$ behaving differently between the two figures?

4.8.10 Remark. *Let $A \in \mathbb{R}^{m\times n}$ have full column rank. The closest vector in $\mathbf{C}(A)$ to $\mathbf{b}$ is $P_A\mathbf{b}$, but can another vector $\mathbf{v}_\star \in \mathbf{C}(A)$ be "equally closest" to $\mathbf{b}$ with $\mathbf{v}_\star \neq P_A\mathbf{b}$? That is, we know that*

$$\|\mathbf{b}-P_A\mathbf{b}\| < \|\mathbf{b}-\mathbf{v}\| \text{ for all } \mathbf{v}\in\mathbf{C}(A) \text{ with } \mathbf{v}\neq P_A\mathbf{b}, \tag{4.8.9}$$

but can $\mathbf{v}_\star \in \mathbf{C}(A)$ satisfy

$$\|\mathbf{b}-\mathbf{v}_\star\| < \|\mathbf{b}-\mathbf{w}\| \text{ for all } \mathbf{w}\in\mathbf{C}(A) \text{ with } \mathbf{w}\neq \mathbf{v}_\star? \tag{4.8.10}$$

No. For then we could take $\mathbf{v}=\mathbf{v}_\star$ in (4.8.9) and $\mathbf{w}=P_A\mathbf{b}$ in (4.8.10) to find

$$\|\mathbf{b}-P_A\mathbf{b}\| < \|\mathbf{b}-\mathbf{v}_\star\| < \|\mathbf{b}-P_A\mathbf{b}\|,$$

and that is impossible.

Here is the lesson of least squares: when we cannot solve $A\mathbf{x}=\mathbf{b}$ because $\mathbf{b}\notin\mathbf{C}(A)$, we first find the best approximation to $\mathbf{b}\in\mathbf{C}(A)$, which we call $\mathbf{p}$, and then we solve $A\widehat{\mathbf{x}}=\mathbf{p}$. So far, this approach requires A to have full column rank.

4.8.11 Example. Let

$$A = \begin{bmatrix}1 & 0\\ 2 & 0\\ 0 & 1\end{bmatrix} \quad\text{and}\quad \mathbf{b} = \begin{bmatrix}1\\0\\0\end{bmatrix}.$$

If $\mathbf{y}\in\mathbf{C}(A)$, then $y_2 = 2y_1$; here $\mathbf{b}\notin\mathbf{C}(A)$. We could just use the formula from Theorem 4.8.8 to find the least squares solution $\widehat{\mathbf{x}}\in\mathbb{R}^2$ that makes $\|A\widehat{\mathbf{x}}-\mathbf{b}\|$ as small as possible, but here it might be enlightening to see how the structure of $P_A\mathbf{b}$ allows us to solve $A\widehat{\mathbf{x}} = P_A\mathbf{b}$ directly.

From Example 4.7.5, the orthogonal projection onto $\mathbf{C}(A)$ is

$$P_A = \begin{bmatrix} 1/5 & 2/5 & 0 \\ 2/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We compute

$$P_A\mathbf{b} = \begin{bmatrix} 1/5 & 2/5 & 0 \\ 2/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 2/5 \\ 0 \end{bmatrix}.$$

Then the problem $A\widehat{\mathbf{x}} = P_A\mathbf{b}$ becomes

$$\begin{cases} \widehat{x}_1 & & = & 1/5 \\ 2\widehat{x}_1 & & = & 2/5 \\ & \widehat{x}_2 & = & 0, \end{cases}$$

which gives $\widehat{x}_1 = 1/5$ and $\widehat{x}_2 = 0$, so the least squares solution is

$$\widehat{\mathbf{x}} = \begin{bmatrix} 1/5 \\ 0 \end{bmatrix}.$$

We can overthink this. Saying that this is the least squares solution means $\|A\widehat{\mathbf{x}} - \mathbf{b}\| \le \|A\mathbf{x} - \mathbf{b}\|$ for all $\mathbf{x} \in \mathbb{R}^2$. We are never going to make $\|A\mathbf{x} - \mathbf{b}\|$ smaller than when we choose $\mathbf{x} = \widehat{\mathbf{x}}$. We compute (squaring for simplicity)

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \left\| \begin{bmatrix} x_1 \\ 2x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} x_1 - 1 \\ 2x_1 \\ x_2 \end{bmatrix} \right\|^2 = (x_1-1)^2 + (2x_1)^2 + x_2^2 \ge (x_1-1)^2 + 4x_1^2.$$

That last inequality holds because $x_2^2 \ge 0$. What this says is that $\|A\mathbf{x} - \mathbf{b}\|^2$ is always at least as large as $(x_1-1)^2 + 4x_1^2$. And what is that? A function of x_1 alone! A little calculus, or graphing the parabola, will show that the minimum of $f(x_1) = (x_1-1)^2 + 4x_1^2$ occurs at $x_1 = 1/5$. Thus

$$\|A\mathbf{x} - \mathbf{b}\|^2 \ge (1/5-1)^2 + 4(1/5)^2 + 0^2 = \|A\widehat{\mathbf{x}} - \mathbf{b}\|^2,$$

exactly as least squares predicts.

4.8.12 Problem (!). Use the formula for $\widehat{\mathbf{x}}$ from Theorem 4.8.8 to compute $\widehat{\mathbf{x}}$ in the previous example directly, and check that the result is, indeed, $\widehat{\mathbf{x}} = (1/5, 0)$.

The following is probably the most legitimate "application" of linear algebra that we will ever do, beyond the overarching application of solving and understanding the fundamental problem $A\mathbf{x} = \mathbf{b}$.

4.8.13 Example. We can find a line between any two points in the plane, but no line is guaranteed to exist between any three or more points. Suppose that we have p sample points of data: $(x_1, y_1), \ldots, (x_p, y_p)$. If we try to find a single line $y = mx + b$ on which all of the points lie, we will need m and b to satisfy $y_k = mx_k + b$ for $k = 1, \ldots, p$. This will probably fail if $p \geq 3$. For example, when $p = 4$, we want m and b to meet

$$\begin{cases} mx_1 & + & b & = & y_1 \\ mx_2 & + & b & = & y_2 \\ mx_3 & + & b & = & y_3 \\ mx_4 & + & b & = & y_4. \end{cases} \tag{4.8.11}$$

Four equations, two unknowns (note that here x_k and y_k are given, and in particular x_k is *known*)—not a recipe for success.

But think about this from the point of view of linear algebra: the problem (4.8.11) is the matrix-vector equation

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}.$$

Again, remember that in the notation of this problem, x_k and y_k are given, while m and b are unknown.

Let

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{bmatrix}.$$

That is, $A = \begin{bmatrix} \mathbf{x} & \mathbf{1} \end{bmatrix}$, where $\mathbf{x}$ is the vector with the x-coordinates of our data points, and $\mathbf{1}$ is the vector whose entries are all 1. We want to do least squares, so A better have independent columns. For that, the first column should not be a multiple of the second. The first column *is* a multiple of the second precisely when all of the x_k's are the same number. (Check that.) But in that case, all of the data points have the same x-coordinate, in which case they all lie on the same vertical line, and that is boring.

So, assume that at least one of the x_k's is not equal to the other. Then we can do least squares and say that the best choice of slope and y-intercept is

$$\begin{bmatrix} \widehat{m} \\ \widehat{b} \end{bmatrix} = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}.$$

We could compute what $\widehat{m}$ and $\widehat{b}$ are explicitly, or we could go to a computer and replace thinking with typing.

This has been a very algebraic treatment of the problem. (The course is called linear algebra.) How does the line $y = \widehat{m}x + \widehat{b}$ interact geometrically with our original data points

(x_k, y_k)? Let $\mathbf{y} = (y_1, \ldots, y_p)$, and recall that $\mathbf{x} = (x_1, \ldots, x_p)$ and $\mathbf{1} = (1, \ldots, 1)$. Least squares tells us that the vector $(\widehat{m}, \widehat{b}) \in \mathbb{R}^2$ satisfies

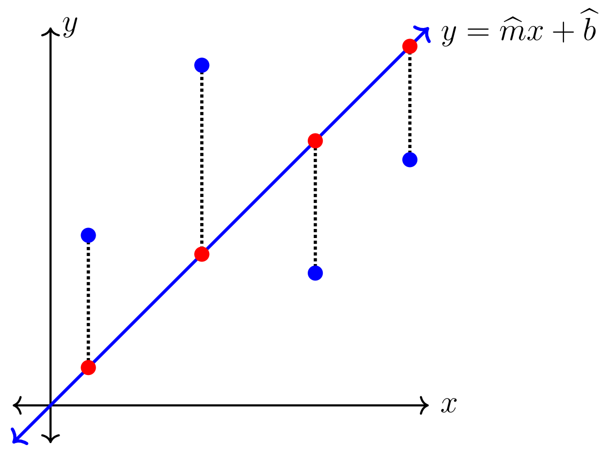
$$\left\| \mathbf{y} - \begin{bmatrix} \mathbf{x} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \widehat{m} \\ \widehat{b} \end{bmatrix} \right\| \leq \left\| \mathbf{y} - \begin{bmatrix} \mathbf{x} & \mathbf{1} \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} \right\|$$

for any $(m, b) \in \mathbb{R}^2$. If we square (for notational convenience) and expand this, we get

$$(y_1 - (\widehat{m}x_1 + \widehat{b}))^2 + \cdots + (y_p - (\widehat{m}x_p + \widehat{b}))^2 \leq (y_1 - (mx_1 + b))^2 + \cdots + (y_p - (mx_p + b))^2$$

for all m, $b \in \mathbb{R}$.

Each difference $|y_k - (\widehat{m}x_k + \widehat{b})|$ is the vertical distance between the data point (x_k, y_k) and the line $y = \widehat{m}x + \widehat{b}$, since the point on this line with x-coordinate equal to x_k is $(x_k, \widehat{m}x_k + \widehat{b})$. Likewise, each difference $|y_k - (mx_k + b)|$ is the vertical distance between the data point (x_k, y_k) and the line $y = mx + b$, since the point on this line with x-coordinate equal to x_k is $(x_k, mx_k + b)$.



So, the least squares line $y = \widehat{m}x + \widehat{b}$ is the line such that the sum of the squares of the *vertical* distances between the data points (x_k, y_k) and the points on the line is as small as possible. (Since we usually think of distance in the plane as *perpendicular* distance, we might wonder how to minimize *that*. . .)

This example suggests that an extremely natural, and important, application results in a matrix A that transparently has full column rank but not full row rank. This justifies our emphasis on A having full column rank in the least squares developments so far.

Content from Strang's *ILA* 6E. Least squares for data fitting to lines appears in Example 1 on pp. 163–164, Figure 4.6, and pp. 167–168. Pay careful attention to the utility of orthogonal columns in A in Example 2 on p. 168. There's no reason to stop with lines. What if you wanted to find the "best" parabola approximating a set of data? Add one more column to A to account for the extra coefficient in the parabola and read p. 170.

Nonetheless, a careful review of the work leading to Theorem 4.8.8 will convince you that we did not need A to have full column rank to find a best approximation to $\mathbf{b}$.

4.8.14 Problem ($\star$). Let $\mathcal{V}$ be a subspace of $\mathbb{R}^m$ and let $\mathbf{b} \in \mathbb{R}^m$. Find $\mathbf{v}_\star \in \mathcal{V}$ such that

$$\|\mathbf{b} - \mathbf{v}_\star\| \leq \|\mathbf{b} - \mathbf{v}\|$$

for any $\mathbf{v} \in \mathcal{V}$. [Hint: *if* $\mathcal{V} = \{\mathbf{0}_m\}$, *there are not many options for* $\mathbf{v}_\star$. *Otherwise, start by writing* $\mathcal{V} = \mathbf{C}(A)$ *for some* $A \in \mathbb{R}^{m \times r}$ *with* $\operatorname{rank}(A) = r \geq 1$.]

So, if A does not have full column rank, we could still find the closest point $\mathbf{p} \in \mathbf{C}(A)$ to $\mathbf{b}$ and then try to solve $A\widehat{\mathbf{x}} = \mathbf{p}$. We will definitely succeed in solving this because $\mathbf{p} \in \mathbf{C}(A)$! The challenge is that because A does not have full column rank, we will succeed with too many degrees of freedom: $\mathbf{N}(A) \neq \{\mathbf{0}_n\}$, and so we have many choices for $\widehat{\mathbf{x}}$. Which is best?

4.8.15 Remark. *Here is one way of proceeding, motivated by the notion that less complicated data is probably better than complicated data.*

Let $A \in \mathbb{R}^{m \times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$, *and let* $\mathbf{p} \in \mathbf{C}(A)$ *be the closest point in* $\mathbf{C}(A)$ *to* $\mathbf{b}$. *Let* $\widehat{\mathbf{x}} \in \mathbb{R}^n$ *satisfy* $A\widehat{\mathbf{x}} = \mathbf{p}$. *By Theorem 4.4.9, write* $\widehat{\mathbf{x}} = \widehat{\mathbf{v}} + \widehat{\mathbf{w}}$, *where* $\widehat{\mathbf{v}} \in \mathbf{C}(A^\mathsf{T})$ *and* $\widehat{\mathbf{w}} \in \mathbf{N}(A)$. *Then* $A\widehat{\mathbf{v}} = \mathbf{p}$, *and so by Theorem 3.4.4 any other solution* $\widehat{\mathbf{y}}$ *to* $A\widehat{\mathbf{y}} = \mathbf{p}$ *also has the form* $\widehat{\mathbf{y}} = \widehat{\mathbf{v}} + \widehat{\mathbf{z}}$ *for some* $\widehat{\mathbf{z}} \in \mathbf{N}(A)$.

By Problem 4.8.6, $\|\widehat{\mathbf{v}}\| \leq \|\widehat{\mathbf{y}}\|$. *That is,* $\widehat{\mathbf{v}}$ *has the smallest norm of any solution* $\widehat{\mathbf{y}}$ *to* $A\widehat{\mathbf{y}} = \mathbf{p}$. *We might call* $\widehat{\mathbf{v}}$ *the* ***MINIMUM-NORM LEAST SQUARES SOLUTION****.*

But how do we find $\widehat{\mathbf{v}}$? *This requires first finding that* $\widehat{\mathbf{x}}$ *that solves* $A\widehat{\mathbf{x}} = \mathbf{p}$, *which requires knowing* $\mathbf{p}$; *this requires the orthogonal projection onto* $\mathbf{C}(A)$. *Then to get* $\widehat{\mathbf{v}}$ *from* $\widehat{\mathbf{x}}$, *we need the orthogonal projection onto* $\mathbf{C}(A^\mathsf{T})$ *This seems like a lot of work.*

It would be nice if there were a simpler formula for $\widehat{\mathbf{v}}$ *in terms of* A *and* $\mathbf{b}$, *and experience teaches us that such a formula probably involves multiplying* $\mathbf{b}$ *by a special matrix. This turns out to be true: there is a matrix* $A^+ \in \mathbb{R}^{n \times m}$ *such that* $\widehat{\mathbf{v}} = A^+\mathbf{b}$, *and this* A^+ *is the* ***PSEUDOINVERSE*** *of* A.

Content from Strang's *ILA* 6E. Page 169 gives a concrete example of what to do when A doesn't have full column rank. The construction of the pseudoinverse is best resolved via the glorious tool of the singular value decomposition. Read the comment at the bottom of p. 169 for a nice review of the three possibilities for solutions to linear systems.

Optionally (this is wholly, totally optional), read Section 4.5, which details what the pseudoinverse does. You can skip the example from the "Incidence Matrix of a Graph" on p. 194. Ideally we will develop the SVD, so the formula for A^+ on p. 195 will eventually make sense.

4.8.16 Problem ($\star$). Here is the opposite question: what is the best solution when we have too many solutions? Suppose that $A \in \mathbb{R}^{m \times n}$ has full row rank, so we can always solve $A\mathbf{x} = \mathbf{b}$. However, perhaps A is not square, in which case A won't have full column rank as well, and so solutions are not unique. This resembles the situation in Remark 4.8.15, which floated the idea of choosing the "minimum norm solution."

First reread that remark carefully. Since $\mathbf{b} \in \mathbf{C}(A)$ here, we can assume $\mathbf{p} = \mathbf{b}$ throughout, and we may as well dispense with the hats since there is actually a solution to $A\mathbf{x} = \mathbf{b}$ now; call this solution $\mathbf{x} = \mathbf{x}_0$, so $A\mathbf{x}_0 = \mathbf{b}$. Use Problem 4.7.4 to get the orthogonal projection P_{A^T} onto $\mathbf{C}(A^\mathsf{T})$. Write $\mathbf{x}_\star = P_{A^\mathsf{T}}\mathbf{x}_0$. Check the following.

(i) $\mathbf{x}_\star = A^\mathsf{T}(AA^\mathsf{T})^{-1}\mathbf{b} \in \mathbf{C}(A^\mathsf{T})$. [Hint: *what is $A\mathbf{x}_0$?*]

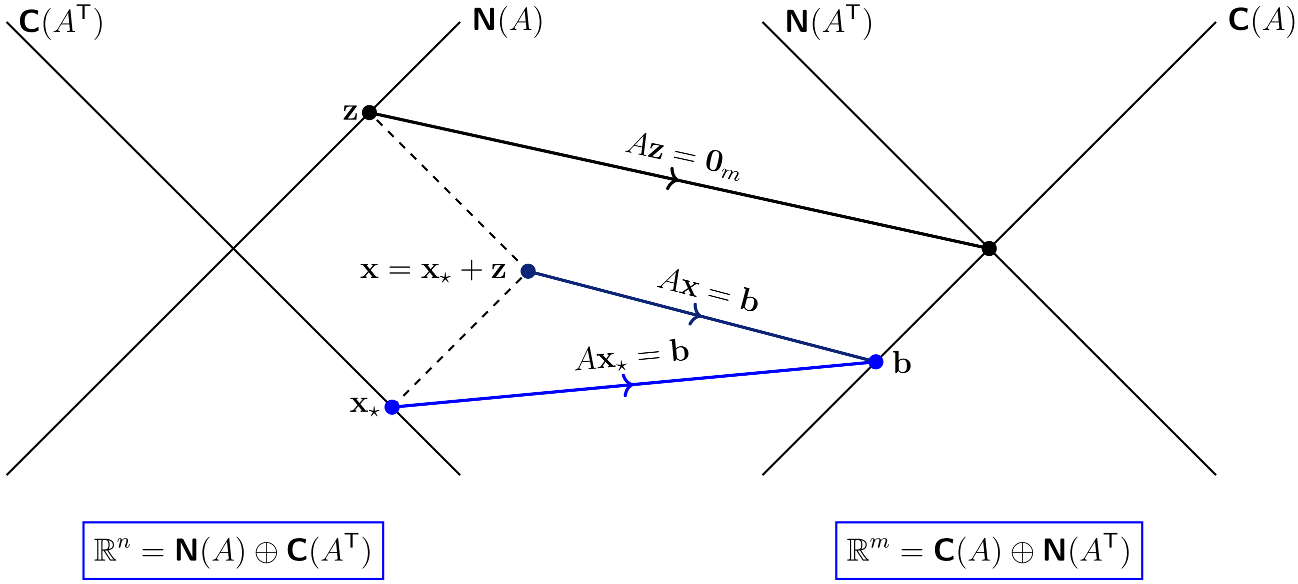
(ii) $A\mathbf{x}_\star = \mathbf{b}$.

(iii) If $A\mathbf{x} = \mathbf{b}$, then $\|\mathbf{x}_\star\| \le \|\mathbf{x}\|$. [Hint: *use Theorem 3.4.4 to write $\mathbf{x} = P_{A^\mathsf{T}}\mathbf{x} + \mathbf{z}$ for some $\mathbf{z} \in \mathbf{N}(A)$; show $P_{A^\mathsf{T}}\mathbf{x} = \mathbf{x}_\star$ and then use Problem 4.8.6 to estimate $\|\mathbf{x}\| \ge \|\mathbf{x}_\star\|$.*]

4.9. Three pictures summarize orthogonality and least squares.

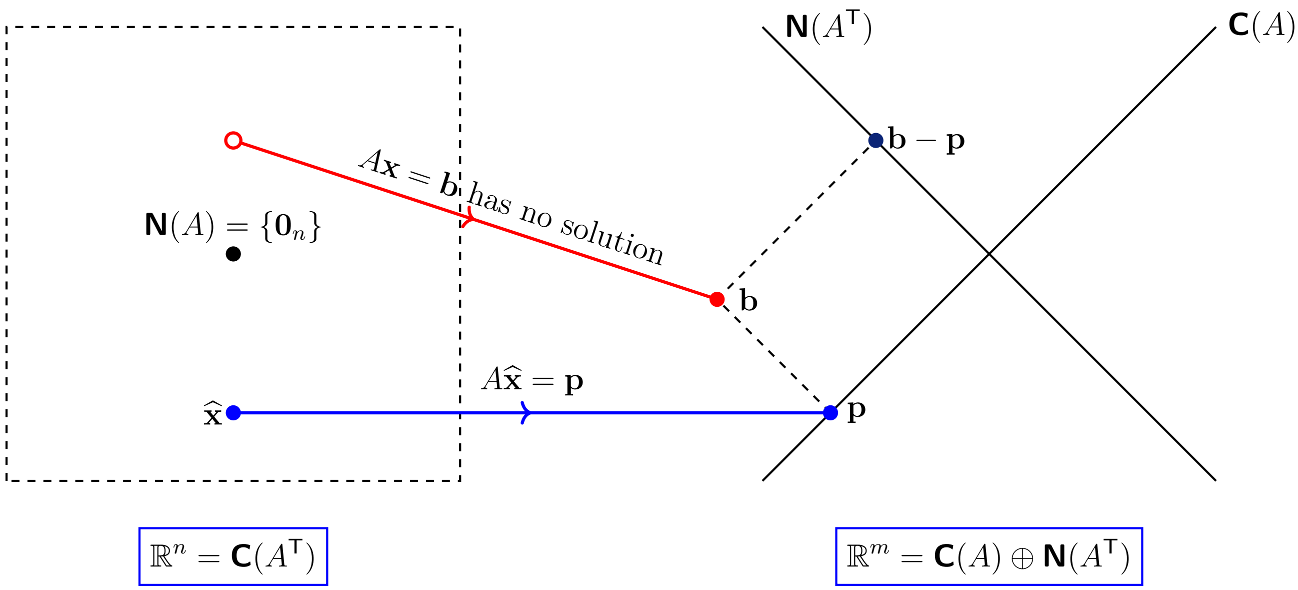
Here is a summary, in pictures, of everything that we have done. Literally: these three pictures encapsulate most of the ideas of the course. In these pictures, which are really fake cartoons, we should imagine that all of the four fundamental subspaces are one-dimensional (except in the second, where the null space of A is trivial), and so we can picture them as coordinate axes in a two-dimensional plane.

• The best case is that we can solve $A\mathbf{x} = \mathbf{b}$, although maybe $\mathbf{N}(A) \ne \{\mathbf{0}_n\}$ and we have infinitely many solutions. In that case perhaps the "best" solution is the "minimum norm solution," i.e., the $\mathbf{x}_\star \in \mathbb{R}^n$ such that $A\mathbf{x}_\star = \mathbf{b}$ and if $\mathbf{x} \in \mathbb{R}^n$ with $A\mathbf{x} = \mathbf{b}$, then $\|\mathbf{x}\star\| \le \|\mathbf{x}\|$. And in this case, we would need $\mathbf{x}_\star \in \mathbf{C}(A^\mathsf{T})$. Here is why: since $\mathbb{R}^n = \mathbf{N}(A) \oplus \mathbf{C}(A^\mathsf{T})$, we could write $\mathbf{x}_\star = \mathbf{v}+\mathbf{w}$ for $\mathbf{v} \in \mathbf{N}(A)$ and $\mathbf{w} \in \mathbf{C}(A^\mathsf{T})$. If $\mathbf{x}_\star \notin \mathbf{C}(A^\mathsf{T})$, then $\mathbf{v} \ne \mathbf{0}_n$ (otherwise $\mathbf{x}_\star = \mathbf{0}_n + \mathbf{w} = \mathbf{w} \in \mathbf{C}(A^\mathsf{T})$, and then $\|\mathbf{x}_\star\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 > \|\mathbf{w}\|^2$, thus $\|\mathbf{x}_\star\| > \|\mathbf{w}\|$. But $A\mathbf{w} = A(\mathbf{v} + \mathbf{w}) = A\mathbf{x}_\star = \mathbf{b}$ since $A\mathbf{v} = \mathbf{0}_n$. Such a minimum norm solution can be constructed with the pseudoinverse; Problem 4.8.16 outlines an approach when A has full row rank.



Content from Strang's *ILA* 6E. This picture is basically Figure 4.1 on p. 146, which you should revisit right now.

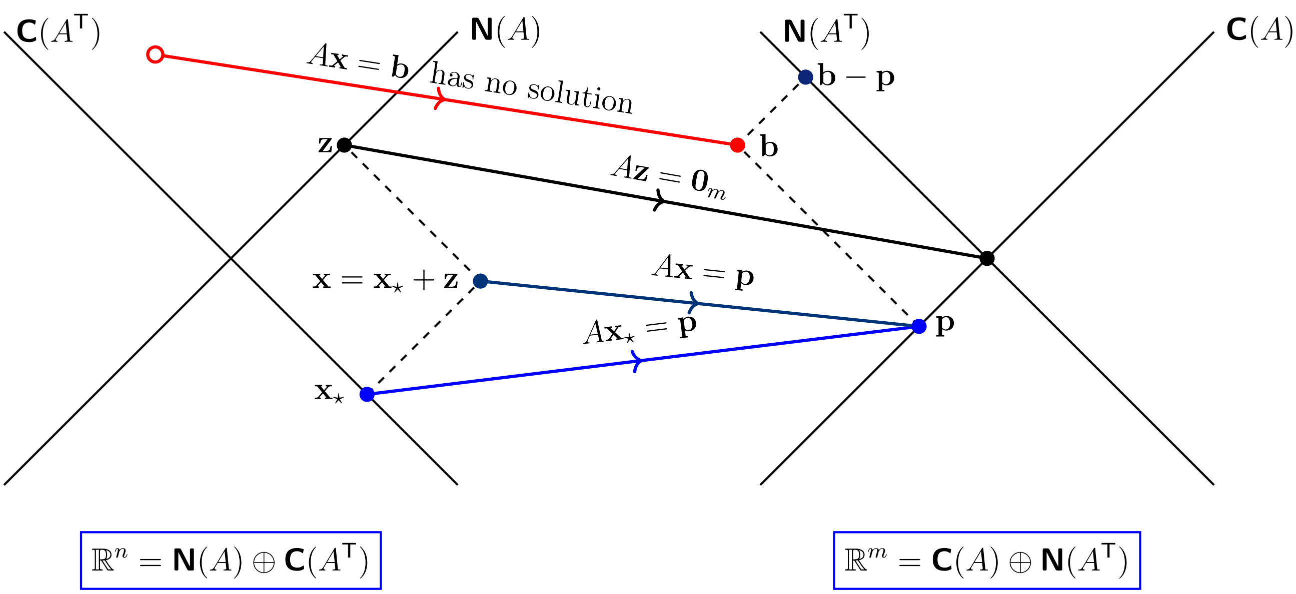
• The next best case is that while we cannot solve $A\mathbf{x} = \mathbf{b}$, since $\mathbf{b} \notin \mathbf{C}(A)$, A does have full column rank, so we can do a least squares approximation.



Content from Strang's *ILA* 6E. This picture is basically Figure 4.7 on p. 166, which you should revisit right now.

• The worst case is that we cannot solve $A\mathbf{x} = \mathbf{b}$ and A does not have full column rank. Then while we can approximate $A\mathbf{x} = \mathbf{b}$ with the problem $A\widehat{\mathbf{x}} = \mathbf{p}$, where $\mathbf{p}$ is the projection of $\mathbf{b}$ onto $\mathbf{C}(A)$, this new approximate problem will not have a unique solution, since $\mathbf{N}(A) \neq \{\mathbf{0}_n\}$. So, can we find $\mathbf{x}_\star \in \mathbb{R}^n$ such that $A\mathbf{x}_\star$ is the closest vector in $\mathbf{C}(A)$ to $\mathbf{b}$ and that $\mathbf{x}_\star$ is the smallest vector $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x}$ is the closest vector in $\mathbf{C}(A)$ to $\mathbf{b}$? Saying that aloud sounds awful, and a slicker way of putting it is that $\mathbf{x}_\star$ should be the MINIMUM NORM LEAST SQUARES SOLUTION. Here is what $\mathbf{x}_\star$ should do:

$$\begin{cases} \mathbf{x}_\star \in \mathbf{C}(A^\mathsf{T}) \\ \mathbf{v} \in \mathbf{C}(A) \Longrightarrow \|\mathbf{b} - A\mathbf{x}_\star\| \leq \|\mathbf{b} - \mathbf{v}\| \\ \begin{pmatrix} \mathbf{x} \in \mathbb{R}^n \\ \mathbf{v} \in \mathbf{C}(A^\mathsf{T}) \Longrightarrow \|\mathbf{b} - A\mathbf{x}\| \leq \|\mathbf{b} - \mathbf{v}\| \end{pmatrix} \Longrightarrow \|\mathbf{x}_\star\| \leq \|\mathbf{x}\|. \end{cases} \tag{4.9.1}$$



As in the first, possibly best, case, we would want $\mathbf{x}_\star \in \mathbf{C}(A^\mathsf{T})$ here. Again, this can be accomplished with the pseudoinverse.

4.9.1 Problem (+). We construct this mythical pseudoinverse. Let $A \in \mathbb{R}^{m\times n}$ with $r := \operatorname{rank}(A) \geq 1$. Suppose that A has the CR-factorization $A = CR$. Let $\mathbf{b} \in \mathbb{R}^m$.

(i) Explain why the closest vector in $\mathbf{C}(A)$ to $\mathbf{b}$ is $P_C\mathbf{b}$. Then the least squares problem is $A\widehat{\mathbf{x}} = P_C\mathbf{b}$.

(ii) If $r < n$, explain why there exist infinitely many solutions to $A\widehat{\mathbf{x}} = P_C\mathbf{b}$. The strategy is now to pick the MINIMUM NORM LEAST SQUARES SOLUTION: choose $\widehat{\mathbf{x}}_\star$ to have the smallest norm possible and still meet $A\widehat{\mathbf{x}}_\star = P_C\mathbf{b}$.

(iii) Let $\widehat{\mathbf{x}} \in \mathbb{R}^n$ solve $A\widehat{\mathbf{x}} = P_C\mathbf{b}$, and conclude that $R\widehat{\mathbf{x}} = (C^\mathsf{T}C)^{-1}C^\mathsf{T}\mathbf{b}$.

(iv) Write $\widehat{\mathbf{x}} = P_{R^\mathsf{T}}\widehat{\mathbf{x}} + \widehat{\mathbf{w}}$ for some $\widehat{\mathbf{w}} \in \mathbf{N}(R)$. Set $\widehat{\mathbf{x}}_\star := P_{R^\mathsf{T}}\widehat{\mathbf{x}}$. Show that

$$\widehat{\mathbf{x}}_\star = (R^\mathsf{T}(RR^\mathsf{T})^{-1})((C^\mathsf{T}C)^{-1}C^\mathsf{T})\mathbf{b}.$$

In particular, regardless of the choice of $\widehat{\mathbf{x}}$, the vector $P_{R^\mathsf{T}}\widehat{\mathbf{x}}$ is the same. Explain why $A\widehat{\mathbf{x}}_\star = P_C\mathbf{b}$.

(v) Simplify the large product above to

$$(R^\mathsf{T}(RR^\mathsf{T})^{-1})((C^\mathsf{T}C)^{-1}C^\mathsf{T}) = R^\mathsf{T}(C^\mathsf{T}AR^\mathsf{T})^{-1}C^\mathsf{T} =: A^+$$

(vi) This matrix A^+ is the PSEUDOINVERSE of A. Explain why $A(A^+\mathbf{b}) = P_C\mathbf{b}$ and $\|A^+\mathbf{b}\| \le \|\widehat{\mathbf{x}}\|$ for any $\widehat{\mathbf{x}} \in \mathbb{R}^n$ with $A\widehat{\mathbf{x}} = P_C\mathbf{b}$. [Hint: *write $A = CR$ and use the definitions of A^+ and P_C to prove $A(A^+\mathbf{b}) = P_C\mathbf{b}$. The estimate is mostly a retread of the work above: if $A\widehat{\mathbf{x}} = P_C\mathbf{b}$, obtain $R\widehat{\mathbf{x}} = (C^\mathsf{T}C)^{-1}C\mathbf{b}$, and then write $\widehat{\mathbf{x}} = P_{\mathbb{R}^\mathsf{T}}\widehat{\mathbf{x}} + \widehat{\mathbf{w}}$ for some $\widehat{\mathbf{w}} \in \mathbf{N}(R)$. Since $P_{\mathbb{R}^\mathsf{T}}\widehat{\mathbf{x}} = A^+\mathbf{b}$, apply Problem 4.8.6 to conclude $\|A^+\mathbf{b}\| \le \|\widehat{\mathbf{x}}\|$.*]

(vii) If $\operatorname{rank}(A) = n$, show that $A^+ = (A^\mathsf{T}A)^{-1}A^\mathsf{T}$. (This is "ordinary" least squares.) If $\operatorname{rank}(A) = m$, show that $A^+ = A^\mathsf{T}(AA^\mathsf{T})^{-1}$. (This is Problem 4.8.16.) And if A is square and invertible, show that $A^+ = A^{-1}$. [Hint: *for each case, what do you know about C and R?*]

Content from Strang's *ILA* 6E. This is wholly optional reading. Page 190 starts out discussing left and right inverses for arbitrary matrices, which generalizes Problem 2.13.5 for square matrices. The rest of the page floats the idea of the "minimum norm least squares solution." Figure 11 on p. 191 views the pseudoinverse dynamically as a way of moving (mapping?) back to the row space. The first half o p. 193 discusses that dynamism, and the second half discusses computing the pseudoinverse from the CR-factorization. Page 192 gives some small examples of pseudoinverses and discusses the special cases of full row and column rank.

4.10. Orthonormal bases are the best bases.

We continue to accumulate victories. We know how the fundamental subspaces associated with a matrix fit into the surrounding Euclidean spaces, and we know how to compute explicitly orthogonal decompositions of vectors. Unfortunately, "explicitly" does not mean "easily." Actually calculating the orthogonal projection onto the column space of $A \in \mathbb{R}^{m\times n}$ can be annoying, because we have to invert a matrix of the form $B^\mathsf{T}B$ (where $B = A$ if A has full column rank). This could make solving a least squares problem hard, and we know anyway that numerically computing inverses is rarely a good idea.

There is actually a way to avoid inverses if $A \in \mathbb{R}^{m\times n}$ has full column rank. Per Theorem 4.8.8, if we cannot solve $A\mathbf{x} = \mathbf{b}$, we consider the approximate problem $A\widehat{\mathbf{x}} = P_A\mathbf{b}$ with $P_A = A(A^\mathsf{T}A)^{-1}A^\mathsf{T}\mathbf{b}$, and the least squares solution is $\widehat{\mathbf{x}} = (A^\mathsf{T}A)^{-1}A^\mathsf{T}\mathbf{b}$. This is equivalent to the so-called NORMAL EQUATION

$$A^{\mathsf{T}}A\widehat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}. \tag{4.10.1}$$

And the normal equation is just a matrix-vector equation, which we could solve with Gaussian elimination and back-substitution, or perhaps an LU-factorization; the point is that we do not really have to compute $(A^{\mathsf{T}}A)^{-1}$ to get $\widehat{\mathbf{x}}$ if we are open to exploring other avenues of solution.

Content from Strang's *ILA* 6E. Reread the first three paragraphs on p. 163.

It turns out that if we ask a little more of our matrix, the projection onto its column space becomes much nicer. The right thing to do is exploit geometry further. Long ago (in Problem 1.6.7) we saw why the standard basis vectors in $\mathbb{R}^m$ were so nice. Since

$$\mathbf{e}_j \cdot \mathbf{e}_k = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

and $\mathbb{R}^m = \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_m) = \mathbf{C}(I_m)$, we have the representation

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{e}_1)\mathbf{e}_1 + \cdots + (\mathbf{v} \cdot \mathbf{e}_m)\mathbf{e}_m$$

for any $\mathbf{v} \in \mathbb{R}^m$. What is special here is not the formulas for the standard basis vectors but how they interact under the dot product. And what is most important is their mutual orthogonality.

4.10.1 Definition. *A list* $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^m$ *is* ***ORTHOGONAL*** *if*

$$\mathbf{u}_j \cdot \mathbf{u}_k = 0$$

for $j \neq k$.

To keep things simple, look at an orthogonal list $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3 \in \mathbb{R}^m$ and say $\mathbf{v} \in \operatorname{span}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$. Then $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3$ for some c_1, c_2, $c_3 \in \mathbb{R}$. The trick is to compute

$$\begin{aligned}
\mathbf{v} \cdot \mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3) \cdot \mathbf{u}_1 \\
&= \big((c_1\mathbf{u}_1) \cdot \mathbf{u}_1\big) + \big((c_2\mathbf{u}_2) \cdot \mathbf{u}_1\big) + \big((c_3\mathbf{u}_3) \cdot \mathbf{u}_1\big) \\
&= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + c_3(\mathbf{u}_3 \cdot \mathbf{u}_1) \\
&= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1).
\end{aligned}$$

If $\mathbf{u}_1 \neq \mathbf{0}_m$, then $\mathbf{u}_1 \cdot \mathbf{u}_1 = \|\mathbf{u}_1\|^2 \neq 0$, and so we have

$$c_1 = \frac{\mathbf{v} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2}.$$

Assume that none of the $\mathbf{u}_j$ are $\mathbf{0}_m$; otherwise, they contribute nothing worthwhile to the span. Taking dot products of $\mathbf{v}$ against the other $\mathbf{u}_j$ then yields

$$c_j = \frac{\mathbf{v} \cdot \mathbf{u}_j}{\|\mathbf{u}_j\|^2}.$$

This generalizes to an arbitrary orthogonal list.

4.10.2 Theorem. *Let $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^m$ be orthogonal and let $\mathbf{v} \in \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$. Then*

$$\mathbf{v} = \left(\frac{\mathbf{v} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2}\right)\mathbf{u}_1 + \cdots + \left(\frac{\mathbf{v} \cdot \mathbf{u}_n}{\|\mathbf{u}_n\|^2}\right)\mathbf{u}_n.$$

A nice consequence is that any orthogonal list of *nonzero* vectors is independent. For if $c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n = \mathbf{0}_m$, then each c_j must be 0. (In the theorem above, take $\mathbf{v} = \mathbf{0}_m$, so the dot products collapse to 0.)

4.10.3 Problem (!). What is the maximum length of any list of orthogonal vectors in $\mathbb{R}^m$?

All that division, however, gets annoying, and it is much more efficient to assume $\|\mathbf{u}_j\| = 1$ for all j.

4.10.4 Definition. *A list $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^m$ is **ORTHONORMAL** if*

$$\mathbf{q}_j \cdot \mathbf{q}_k = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

The $j \neq k$ condition means that an orthonormal list is orthogonal, while the $j = k$ condition gives $\|\mathbf{q}_j\| = \sqrt{\mathbf{q}_j \cdot \mathbf{q}_j} = 1$.

4.10.5 Problem ($\star$). Let $\mathbf{q}_1$, $\mathbf{q}_2 \in \mathbb{R}^m$ be orthonormal and let $\mathbf{v} \in \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2)$. Use the Pythagorean theorem (Theorem 4.8.4) to show that

$$\|\mathbf{v}\|^2 = |\mathbf{v} \cdot \mathbf{q}_1|^2 + |\mathbf{v} \cdot \mathbf{q}_2|^2.$$

This generalizes: if $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^m$ are orthonormal and $\mathbf{v} \in \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_n)$, then

$$\|\mathbf{v}\|^2 = |\mathbf{v} \cdot \mathbf{q}_1|^2 + \cdots + |\mathbf{v} \cdot \mathbf{q}_n|^2.$$

We work with matrices and column spaces as much as we do with lists of vectors and spans, so we put those orthonormal vectors into a matrix and get an unfortunate definition.

4.10.6 Definition. *A matrix $Q \in \mathbb{R}^{m \times n}$ is **ORTHOGONAL** if the columns of Q are orthonormal.*

4.10.7 Example. **(i)** The identity matrix is always orthogonal.

(ii) Let $\theta \in \mathbb{R}$ and

$$Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Then trig and the Pythagorean identity $\sin^2(\theta) + \cos^2(\theta) = 1$ say that Q is orthogonal.

Content from Strang's *ILA* 6E. Everything on pp. 176–178 is important. I am being a little more general and calling any matrix, square or not, with orthogonal columns an "orthogonal matrix."

Here is a nice consequence of definitions. Let $Q = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$ be orthogonal. Then

$$\mathbf{q}_i \cdot \mathbf{q}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Remember that row i of Q^T is just column i of Q, and that the (i,j)-entry of $Q^\mathsf{T}Q$ is the dot product of row i of Q^T and column j of Q. That is, the (i,j)-entry of $Q^\mathsf{T}Q$ is $\mathbf{q}_i \cdot \mathbf{q}_j$, and so this (i,j)-entry is 1 when $i = j$ (= on the diagonal) and 0 otherwise (= off the diagonal). This sounds a lot like an identity matrix, and it is! Since $Q \in \mathbb{R}^{m \times n}$, we have $Q^\mathsf{T} \in \mathbb{R}^{n \times m}$, and so $Q^\mathsf{T}Q \in \mathbb{R}^{n \times n}$.

4.10.8 Theorem. *Let $Q \in \mathbb{R}^{m \times n}$ be orthogonal. Then $Q^\mathsf{T}Q = I_n$.*

4.10.9 Problem (!). Is every orthogonal matrix invertible?

4.10.10 Problem ($\star$). State and prove an analogue of Theorem 4.10.8 for the case when the columns of Q are only orthogonal, not orthonormal.

4.10.11 Problem (!). Is every orthogonal projection (Definition 4.7.10) an orthogonal matrix?

We can use Theorem 4.10.8 to get a slick representation of vectors in $\mathbf{C}(Q)$ for orthogonal Q. Let $Q \in \mathbb{R}^{m \times n}$ be orthogonal and $\mathbf{b} \in \mathbf{C}(Q)$. Then $\mathbf{b} = Q\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, and so $Q^\mathsf{T}\mathbf{b} = Q^\mathsf{T}Q\mathbf{x} = \mathbf{x}$. Thus

$$\mathbf{b} = Q\mathbf{x} = QQ^\mathsf{T}\mathbf{b}.$$

Now we think about matrix multiplication. One way to compute the entries of $Q^\mathsf{T}\mathbf{b}$ is to take the dot product of the rows of Q^T with $\mathbf{b}$. (This is probably how we usually compute matrix-vector products by hand.) And the rows of Q^T are the columns of Q, so

$$Q^\mathsf{T}\mathbf{b} = \begin{bmatrix} \mathbf{q}_1 \cdot \mathbf{b} \\ \vdots \\ \mathbf{q}_n \cdot \mathbf{b} \end{bmatrix}.$$

Next, one way to compute $QQ^\mathsf{T}\mathbf{b}$ is to take the linear combination of the columns of Q weighted by the entries of $Q^\mathsf{T}\mathbf{b}$:

$$\mathbf{b} = QQ^\mathsf{T}\mathbf{b} = \begin{bmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_n\end{bmatrix}\begin{bmatrix}\mathbf{q}_1 \cdot \mathbf{b} \\ \vdots \\ \mathbf{q}_n \cdot \mathbf{b}\end{bmatrix} = (\mathbf{q}_1 \cdot \mathbf{b})\mathbf{q}_1 + \mathbf{q}_1 \cdots + (\mathbf{q}_n \cdot \mathbf{b})\mathbf{q}_n.$$

This brings together two ways of looking at matrix-vector multiplication: the dot product way for quick and dirty calculations by hand, the linear combination of columns way to actually understand what happens.

These are old ideas reinterpreted in a new framework. Here is what is really new, and useful: orthonormality and orthogonal matrices make least squares so much easier. Suppose that $Q \in \mathbb{R}^{m\times n}$ is orthogonal and we want to solve $Q\mathbf{x} = \mathbf{b}$, but $\mathbf{b} \notin \mathbf{C}(Q)$. Then we would solve the least squares problem

$$Q\widehat{\mathbf{x}} = P_Q\mathbf{b},$$

where

$$P_Q = Q(Q^\mathsf{T}Q)^{-1}Q^\mathsf{T} = QI_n^{-1}Q^\mathsf{T} = QQ^\mathsf{T}. \tag{4.10.2}$$

Look at that: the orthogonal projection onto $\mathbf{C}(Q)$ collapses to QQ^T. No inverses needed.

4.10.12 Theorem. *Let $Q = \begin{bmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_n\end{bmatrix} \in \mathbb{R}^{m\times n}$ be orthogonal. Then the orthogonal projection onto $\mathbf{C}(Q)$ is QQ^T, and every $\mathbf{v} \in \mathbf{C}(Q)$ has the form*

$$\mathbf{v} = QQ^\mathsf{T}\mathbf{v} = (\mathbf{v}\cdot\mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{v}\cdot\mathbf{q}_n)\mathbf{q}_n.$$

Content from Strang's *ILA* 6E. You should hold the answer to Worked Example 4.4 B on p. 185 deep within your heart.

Then the least squares problem is just

$$Q\widehat{\mathbf{x}} = QQ^\mathsf{T}\mathbf{b}.$$

Since Q has full column rank, $\mathbf{N}(Q) = \{\mathbf{0}_n\}$, and so the only solution $\widehat{\mathbf{x}}$ to $Q\widehat{\mathbf{x}} = Q(Q^\mathsf{T}\mathbf{b})$ is, basically by design, and that says

$$\widehat{\mathbf{x}} = Q^\mathsf{T}\mathbf{b}.$$

Computing the least squares solution requires no inverses, only transposing.

4.10.13 Problem (!). Reread Example 4.7.5 and explain how orthonormality made calculating the projection operator easier. How would things have been more complicated there if we used the pivot columns of A as the basis for the column space, not the columns of B?

Content from Strang's *ILA* 6E. Page 179 through the top of p. 180 discuss least squares with orthogonal matrices.

We conclude with a special, and hugely important, case. Every orthonormal list is an orthogonal list with no zero vectors in it. And every orthogonal list with no zero vectors is independent. So an orthonormal list of length m in $\mathbb{R}^m$ is an independent list of length m: thus a basis for $\mathbb{R}^m$. (No need to check spanning because we can count.) This is the best basis.

4.10.14 Definition. *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^p$. A list $\mathbf{q}_1, \ldots, \mathbf{q}_r$ in $\mathcal{V}$ is an **ORTHONORMAL BASIS** for $\mathcal{V}$ if it is both a basis for $\mathcal{V}$ and an orthonormal list.*

Since orthonormality implies orthogonality, and orthogonality implies independence, to check that a list is an orthonormal basis for a subspace, we just have to check that the list is orthonormal and spans the subspace; we get independence for free from orthonormality. Theorem 4.10.2 gives us a really slick way to represent a vector $\mathbf{v} \in \mathbb{R}^m$ in terms of an orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_m$ for $\mathbb{R}^m$: since $\mathbf{v} \in \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_m)$, we get

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{v} \cdot \mathbf{q}_m)\mathbf{q}_m.$$

This is so great because we *explicitly* have the coefficients for $\mathbf{v}$ with respect to this basis. Just take the dot product.

4.10.15 Problem ($\star$). Let $\mathbf{q}_1, \ldots, \mathbf{q}_m$ be an orthonormal basis for $\mathbb{R}^m$ and let $1 \le r < m$. Put $\mathcal{V} = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_r)$ and explain why $\mathcal{V}^\perp = \operatorname{span}(\mathbf{q}_{r+1}, \ldots, \mathbf{q}_m)$. [Hint: *we need Definition 1.8.4 here. If $\mathbf{v} \in \operatorname{span}(\mathbf{q}_{r+1}, \ldots, \mathbf{q}_m)$, use Problem 4.5.9 to show that $\mathbf{v} \in \mathcal{V}_r^\perp$. Next, if $\mathbf{w} \in \mathcal{V}_r^\perp$, we know $\mathbf{w} = (\mathbf{w} \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{w} \cdot \mathbf{q}_m)\mathbf{q}_m$ and $\mathbf{w} \cdot \mathbf{q}_j = 0$ for $j = 1, \ldots, r$. What goes away in that expansion of $\mathbf{w}$?*]

4.11. Gram–Schmidt turns independence into orthonormality.

Life is best when the vectors under consideration are orthonormal. Often they are not. How do we turn a problem governed by “ordinary” vectors into a problem controlled by orthonormal vectors? Of course this depends on the exact problem, and the problem that we have most recently studied is least squares.

Let $A \in \mathbb{R}^{m \times n}$. How can we find an orthonormal basis for $\mathbf{C}(A)$? This will make computing the orthogonal projection P_A onto $\mathbf{C}(A)$ easy. The hard way to find P_A is to start with a matrix $B \in \mathbb{R}^{m \times r}$ such that the columns of B are independent and $\mathbf{C}(A) = \mathbf{C}(B)$ and then compute $P_A = B(B^\mathsf{T} B)^{-1} B^\mathsf{T}$. But if we have an orthogonal matrix $Q \in \mathbb{R}^{m \times r}$ such that $\mathbf{C}(A) = \mathbf{C}(Q)$, then $P_A = QQ^\mathsf{T}$, per (4.10.2).

We therefore may as well just start with an independent list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{R}^m$. Here is our goal: given an independent list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{R}^m$, find an orthonormal list $\mathbf{q}_1, \ldots, \mathbf{q}_r$ in $\mathbb{R}^m$ such that

$$\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_n).$$

We will figure out a very transparent, iterative algorithm for doing this. It is less important to be able to perform this algorithm by hand than it is to understand how to figure it out in the first place. We look at three increasingly more complicated cases on r to get some ideas.

1. $n = 1$. We want the very small list $\mathbf{q}_1$ to be orthonormal with $\operatorname{span}(\mathbf{v}_1) = \operatorname{span}(\mathbf{q}_1)$. We therefore want $\|\mathbf{q}_1\| = 1$, and we will get the spanning property if $\mathbf{q}_1 = c_1\mathbf{v}_1$ for some $c_1 \in \mathbb{R}$. How should we choose c_1? All that we know is

$$1 = \|\mathbf{q}_1\| = \|c_1\mathbf{v}_1\| = |c_1|\,\|\mathbf{v}_1\|\,.$$

Since the list $\mathbf{v}_1$ is independent, $\mathbf{v}_1 \neq \mathbf{0}_m$, so $\|\mathbf{v}_1\| \neq 0$. Then we may solve for

$$|c_1| = \frac{1}{\|\mathbf{v}_1\|}.$$

This suggests taking

$$c_1 = \frac{1}{\|\mathbf{v}_1\|}.$$

2. $n = 2$. We want the list $\mathbf{q}_1, \mathbf{q}_2$ to be orthonormal and $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2)$. Since declaring $\mathbf{q}_1 := \mathbf{v}_1/\|\mathbf{v}_1\|$ was enough in the $r = 1$ case, we might try doing that here, too. This gives $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2) = \operatorname{span}(\mathbf{q}_1, \mathbf{v}_2)$. (Think about it…) If we then want $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2)$, we need $\mathbf{q}_2 \in \operatorname{span}(\mathbf{q}_1, \mathbf{v}_2)$. So, we must be able to write $\mathbf{q}_2 = c_1\mathbf{q}_1 + c_2\mathbf{v}_2$, and then to have $\mathbf{q}_2 \cdot \mathbf{q}_1 = 0$, we need

$$0 = \mathbf{q}_2 \cdot \mathbf{q}_1 = (c_1\mathbf{q}_1 + c_2\mathbf{v}_2) \cdot \mathbf{q}_1 = c_1(\mathbf{q}_1 \cdot \mathbf{q}_1) + c_2(\mathbf{v}_2 \cdot \mathbf{q}_1) = c_1 + c_2(\mathbf{v}_2 \cdot \mathbf{q}_1) = c_1 + c_2(\mathbf{v}_2 \cdot \mathbf{q}_1)$$

since $\mathbf{q}_1 \cdot \mathbf{q}_1 = \|\mathbf{q}_1\|^2 = 1$. Then

$$c_1 = -c_2(\mathbf{v}_2 \cdot \mathbf{q}_1)$$

and so

$$\mathbf{q}_2 = -c_2(\mathbf{v}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 + c_2\mathbf{v}_2 = c_2(\mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{q}_1)\mathbf{q}_1) \tag{4.11.1}$$

Put

$$\mathbf{w}_2 := \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{q}_1)\mathbf{q}_1,$$

so we want $\mathbf{q}_2 = c_2\mathbf{w}_2$. Since the list $\mathbf{v}_1, \mathbf{v}_2$ is independent and $\mathbf{q}_1$ is a nonzero scalar multiple of $\mathbf{v}_1$, the list $\mathbf{q}_1, \mathbf{v}_2$ is also independent (check that), and therefore $\mathbf{w}_2 \neq \mathbf{0}_m$. To have $\|\mathbf{q}_2\| = 1$, we need $|c_2|\,\|\mathbf{w}_2\| = 1$; this suggests taking $c_2 = 1/\|\mathbf{w}_2\|$, which is permissible.

To summarize, we put

$$\begin{cases} \mathbf{q}_1 := \mathbf{v}_1/\|\mathbf{v}_1\| \\ \mathbf{w}_2 := \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 \\ \mathbf{q}_2 := \mathbf{w}_2/\|\mathbf{w}_2\| \end{cases} \tag{4.11.2}$$

to get an orthonormal list $\mathbf{q}_1$, $\mathbf{q}_2$ with the desired property that $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2)$, and the added bonus that $\operatorname{span}(\mathbf{v}_1) = \operatorname{span}(\mathbf{q}_1)$.

3. $n = 3$. We want the list $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ to be orthonormal with

$$\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3). \tag{4.11.3}$$

Based on our prior success, we might use (4.11.2) to define $\mathbf{q}_1$ and $\mathbf{q}_2$. Then we would have (4.11.3). We therefore want $\mathbf{q}_3 \in \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{v}_3)$, and so we need $\mathbf{q}_3 = c_1\mathbf{q}_1 + c_2\mathbf{q}_2 + c_3\mathbf{v}_3$. To have $\mathbf{q}_3 \cdot \mathbf{q}_1 = 0$, we need

$$0 = \mathbf{q}_3 \cdot \mathbf{q}_1 = (c_1\mathbf{q}_1 + c_2\mathbf{q}_2 + c_3\mathbf{v}_3) \cdot \mathbf{q}_1 = c_1(\mathbf{q}_1 \cdot \mathbf{q}_1) + c_2(\mathbf{q}_2 \cdot \mathbf{q}_1) + c_3(\mathbf{v}_3 \cdot \mathbf{q}_1) = c_1 + c_3(\mathbf{v}_3 \cdot \mathbf{q}_1)$$

since $\mathbf{q}_1 \cdot \mathbf{q}_1 = \|\mathbf{q}_1\|^2 = 1$ and $\mathbf{q}_2 \cdot \mathbf{q}_1 = 0$. Similarly, we want

$$0 = \mathbf{q}_3 \cdot \mathbf{q}_2 = c_1(\mathbf{q}_1 \cdot \mathbf{q}_2) + c_2(\mathbf{q}_2 \cdot \mathbf{q}_2) + c_3(\mathbf{v}_3 \cdot \mathbf{q}_2) = c_2 + c_3(\mathbf{v}_3 \cdot \mathbf{q}_2)$$

since $\mathbf{q}_1 \cdot \mathbf{q}_2 = 0$ and $\mathbf{q}_2 \cdot \mathbf{q}_2 = \|\mathbf{q}_2\|^2 = 1$. Then we need

$$c_1 = -c_3(\mathbf{v}_3 \cdot \mathbf{q}_1) \quad \text{and} \quad c_2 = -c_3(\mathbf{v}_3 \cdot \mathbf{q}_2).$$

So, $\mathbf{q}_3$ must have the form

$$\mathbf{q}_3 = -c_3(\mathbf{v}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 - c_3(\mathbf{v}_3 \cdot \mathbf{q}_2)\mathbf{q}_2 + c_3\mathbf{v}_3 = c_3(\mathbf{v}_3 - (\mathbf{v}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{v}_3 \cdot \mathbf{q}_2)\mathbf{q}_2).$$

Put $\mathbf{w}_3 = \mathbf{v}_3 - (\mathbf{v}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{v}_3 \cdot \mathbf{q}_2)\mathbf{q}_2$. If $\mathbf{w}_3 = \mathbf{0}_m$, then $\mathbf{v}_3 \in \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2) = \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2)$, which contradicts the independence of the list $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. So, $\mathbf{w}_3 \neq \mathbf{0}_m$, and therefore to have $\|\mathbf{q}_3\| = 1$, we could take $c_3 = 1/\|\mathbf{w}_3\|$.

To summarize, we put

$$\begin{cases} \mathbf{q}_1 := \mathbf{v}_1/\|\mathbf{v}_1\| \\ \mathbf{w}_2 := \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 \\ \mathbf{q}_2 := \mathbf{w}_2/\|\mathbf{w}_2\| \\ \mathbf{w}_3 := \mathbf{v}_3 - (\mathbf{v}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{v}_3 \cdot \mathbf{q}_2)\mathbf{q}_2 \\ \mathbf{q}_3 := \mathbf{w}_3/\|\mathbf{w}_3\| \end{cases}$$

to get an orthonormal list $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ with the desired property that

$$\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3),$$

and the added bonus that

$$\operatorname{span}(\mathbf{v}_1) = \operatorname{span}(\mathbf{q}_1) \quad \text{and} \quad \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2).$$

Content from Strang's *ILA* 6E. Pages 180–181 do Gram–Schmidt for three vectors. See in particular the 3D drawings in Figure 4.10 on p. 181.

We are almost ready for the general result and need one auxiliary tool first.

4.11.1 Problem ($\star$). Let A, $B \in \mathbb{R}^{m \times n}$ with $\mathbf{C}(A) = \mathbf{C}(B)$ and $\operatorname{rank}(B) = n$. Let $\mathbf{v} \in \mathbb{R}^m$. Show that

$$\mathbf{C}(\begin{bmatrix} A & \mathbf{v} \end{bmatrix}) = \mathbf{C}(\begin{bmatrix} B & (\mathbf{v} - P_B\mathbf{v}) \end{bmatrix}).$$

[Hint: *let* $c \in \mathbb{R}$ *and* $\mathbf{x} \in \mathbb{R}^n$ *and write* $A\mathbf{x} + c\mathbf{v} = A\mathbf{x} + c(\mathbf{v} - P_B\mathbf{v}) + cP_B\mathbf{v}$. *Explain* *why* $A\mathbf{x} + cP_B\mathbf{v} \in \mathbf{C}(B)$. *Next, let* $c \in \mathbb{R}$ *and* $\mathbf{x} \in \mathbb{R}^n$ *and write* $B\mathbf{x} + c(\mathbf{v} - P_B\mathbf{v}) = (B\mathbf{x} - cP_B\mathbf{v}) + c\mathbf{v}$. *Explain why* $B\mathbf{x} - cP_B\mathbf{v} \in \mathbf{C}(A)$.]

4.11.2 Theorem (Gram–Schmidt procedure). *Suppose that the list* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *in* $\mathbb{R}^m$ *is independent. There exists an orthonormal list* $\mathbf{q}_1, \ldots, \mathbf{q}_n$ *in* $\mathbb{R}^m$ *such that* $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j)$ *for* $j = 1, \ldots, n$. *Specifically,*

$$\begin{cases} \mathbf{q}_1 := \mathbf{v}_1 / \|\mathbf{v}_1\| \\ \mathbf{w}_j := \mathbf{v}_j - \big((\mathbf{v}_j \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{v}_j \cdot \mathbf{q}_{j-1})\mathbf{q}_{j-1}\big), \ j \geq 2 \\ \mathbf{q}_j := \mathbf{w}_j / \|\mathbf{w}_j\|, \ j \geq 2. \end{cases} \tag{4.11.4}$$

With

$$Q_{j-1} := \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_{j-1} \end{bmatrix}$$

for $j \geq 2$, *we can also write*

$$\begin{cases} \mathbf{q}_1 := \mathbf{v}_1 / \|\mathbf{v}_1\| \\ \mathbf{w}_j := (I_m - Q_{j-1}Q_{j-1}^{\mathsf{T}})\mathbf{v}_j, \ j \geq 2. \mathbf{q}_j := \mathbf{w}_j / \|\mathbf{w}_j\|, \ j \geq 2. \end{cases}$$

Proof. This is really a proof by induction, but the key ideas are outlined in the $n = 3$ case above. The point is that we know how to construct $\mathbf{q}_1$, and then we assume that we have constructed through $\mathbf{q}_j$ with $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j)$ and $\mathbf{q}_1, \ldots, \mathbf{q}_j$ orthonormal. Then we check that the algorithm above defines $\mathbf{q}_{j+1}$ correctly.

1. We put $\mathbf{w}_{j+1} := (I_m - Q_{j+1}Q_{j+1}^{\mathsf{T}})\mathbf{v}_{j+1}$.

2. By Theorem 4.10.12, the matrix $Q_{j+1}Q_{j+1}^{\mathsf{T}}$ is the orthogonal projection onto

$$\mathbf{C}(Q_{j+1}) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j) = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j).$$

And by Problem 4.7.13, $I_m - Q_{j+1}Q_{j+1}^{\mathsf{T}}$ is the orthogonal projection onto $\mathbf{C}(Q_{j+1})^{\perp}$. Thus $\mathbf{w}_{j+1}$ is orthogonal to every vector in $\mathbf{C}(Q_{j+1})$, in particular the vectors $\mathbf{q}_1, \ldots, \mathbf{q}_j$.

3. To show that $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j, \mathbf{v}_{j+1}) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j, \mathbf{w}_{j+1})$, invoke Problem 4.11.1 with $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_j \end{bmatrix}$, $B = Q_j$, and $\mathbf{v} = \mathbf{w}_{j+1}$.

4. We check that $\mathbf{w}_{j+1} \neq \mathbf{0}_m$: if $\mathbf{w}_{j+1} = \mathbf{0}_m$, then $\mathbf{v}_{j+1} = Q_jQ_j^{\mathsf{T}}\mathbf{v}_{j+1} \in \mathbf{C}(Q_j) = \mathbf{C}(A)$. This contradicts the independence of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

5. And so we can renormalize $\mathbf{q}_{j+1} := \mathbf{w}_{j+1} / \|\mathbf{w}_{j+1}\|$ and obtain

$$\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j, \mathbf{v}_{j+1}) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j, \mathbf{w}_{j+1}) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j, \mathbf{q}_{j+1}).$$

Since this was true for an arbitrary $j \geq 1$, we can "turn the crank" and keep going to get the full list $\mathbf{q}_1, \ldots, \mathbf{q}_n$, regardless of the value of n. ■

Content from Strang's *ILA* 6E. Page 183 presents some pseudocode for computing Gram–Schmidt. See also the confession on p. 184. Read it and take a numerical linear algebra class.

To write things out concisely, the projection $Q_jQ_j^\mathsf{T}$ condenses notation, and the formula $\mathbf{w}_j = (I_m - Q_{j-1}Q_{j-1}^\mathsf{T})\mathbf{v}_j$ is an instance of our course's theme that we do things by multiplying by matrices—in this case, the matrix $I_m - Q_{j-1}Q_{j-1}^\mathsf{T}$. To do calculations by hand, the formula for $\mathbf{w}_j$ and $\mathbf{q}_j$ in (4.11.4) is probably more transparent.

4.11.3 Problem ($\star$). Using the hypotheses and notation of the Gram–Schmidt procedure, prove that $\mathbf{v}_j \cdot \mathbf{q}_j > 0$ as follows.

(i) Show that $\mathbf{v}_1 \cdot \mathbf{q}_1 > 0$.

(ii) To show $\mathbf{v}_j \cdot \mathbf{q}_j > 0$ for $j \geq 2$, explain why we just need to show that $\mathbf{v}_j \cdot \mathbf{w}_j > 0$.

(iii) For $j \geq 2$, rewrite

$$\mathbf{v}_j = \mathbf{w}_j + Q_{j-1}Q_{j-1}^\mathsf{T}\mathbf{v}_j.$$

Explain why $Q_{j-1}Q_{j-1}^\mathsf{T}\mathbf{v}_j \cdot \mathbf{w}_j = 0$. [Hint: $Q_{j-1}Q_{j-1}^\mathsf{T}\mathbf{v}_j = Q_{j-1}(Q_{j-1}^\mathsf{T}\mathbf{v}_j)$ *and* $\mathbf{w}_j \in \mathbf{C}(Q_{j-1})^\perp$.]

(iv) Conclude that $\mathbf{v}_j \cdot \mathbf{w}_j = \|\mathbf{w}_j\|^2$.

4.11.4 Example. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

The list $\mathbf{v}_1$, $\mathbf{v}_2$ is the list of pivot columns of the long-suffering matrix

$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix},$$

so $\mathbf{v}_1$, $\mathbf{v}_2$ is a basis for $\mathbf{C}(A)$. From Example 4.4.11, the (short) list $\mathbf{v}_3$ is a basis for $\mathbf{N}(A^\mathsf{T})$. When we combine a basis for $\mathbf{C}(A)$ with a basis for $\mathbf{N}(A^\mathsf{T})$, orthogonality gives us a basis for the whole space. That is, $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ is a basis for $\mathbb{R}^3$. In particular, this list is independent, so we can apply Gram–Schmidt to it.

1. Start by computing

$$\|\mathbf{v}_1\| = \sqrt{1^2 + 2^2 + 0^2} = \sqrt{5}$$

and put

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}.$$

2. Then we want to set

$$\mathbf{w}_2 = \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{q}_1)\mathbf{q}_1.$$

We compute

$$\mathbf{v}_2 \cdot \mathbf{q}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix} = \frac{1}{\sqrt{5}} + \frac{4}{\sqrt{5}} = \frac{5}{\sqrt{5}} = \sqrt{5},$$

so

$$\mathbf{w}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \sqrt{5} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

Then

$$\|\mathbf{w}_2\| = \sqrt{0^2 + 0^2 + 2^2} = 2,$$

so we put

$$\mathbf{q}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \frac{1}{2} \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

3. Last, we want to set

$$\mathbf{w}_3 = \mathbf{v}_3 - (\mathbf{v}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{v}_3 \cdot \mathbf{q}_2)\mathbf{q}_2.$$

We have

$$\mathbf{v}_3 \cdot \mathbf{q}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix} = -\frac{2}{\sqrt{5}} + \frac{2}{\sqrt{5}} = 0$$

and

$$\mathbf{v}_3 \cdot \mathbf{q}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0,$$

so there is not much work to do here: $\mathbf{w}_3 = \mathbf{v}_3$. Compute

$$\|\mathbf{w}_3\| = \|\mathbf{v}_3\| = \sqrt{(-2)^2 + 1^2 + 0^2} = \sqrt{5},$$

and set

$$\mathbf{q}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|} = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}.$$

The result is that the list $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ is orthonormal and preserves spans in the sense that

$$\operatorname{span}(\mathbf{v}_1) = \operatorname{span}(\mathbf{q}_1), \qquad \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2),$$
$$\text{and} \qquad \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3).$$

In particular, since there are three vectors in the list $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, it is an orthonormal basis for $\mathbb{R}^3$. The best basis. Also, since $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2) = \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2)$, the list $\mathbf{q}_1$, $\mathbf{q}_2$ is an orthonormal basis for $\mathbf{C}(A)$ with A from above.. Again, the best basis.

4.11.5 Problem ($\star$). Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$.

(i) Suppose that for some integer j with $1 \le j \le n-1$, the list $\mathbf{v}_1, \ldots, \mathbf{v}_j$ is independent but $\mathbf{v}_{j+1} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j)$, so the list $\mathbf{v}_1, \ldots, \mathbf{v}_{j+1}$ is dependent. What happens at the $(j+1)$st step in the Gram–Schmidt process? [Hint: *reread the proof of Theorem 4.11.2.*]

(ii) Let j be an integer with $1 \le j \le n-1$, and now suppose that the list $\mathbf{v}_1, \ldots, \mathbf{v}_{j+1}$ is independent, so we can do Gram–Schmidt through the $(j+1)$st step. Suppose that $\mathbf{v}_{j+1} \cdot \mathbf{v}_k = 0$ for $k = 1, \ldots, j$. What now happens at this $(j+1)$st step in Gram–Schmidt? [Hint: *think about the third step in Example 4.11.4.*]

4.12. The QR-factorization makes least squares easy.

That the Gram–Schmidt procedure "preserves spans" is probably not a consequence that we expected when we originally started out with an independent list and wanted to get an orthonormal list with the same span as the whole list. Sometimes accidental consequences are nice.

Look at the $n = 3$ situation. We have an independent list $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ in $\mathbb{R}^m$ and orthonormal vectors $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3 \in \mathbb{R}^m$ such that the spans are preserved:

$$\begin{cases} \mathbf{v}_1 \in \operatorname{span}(\mathbf{q}_1) \\ \mathbf{v}_2 \in \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2) \\ \mathbf{v}_3 \in \operatorname{span}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3). \end{cases}$$

Since the $\mathbf{q}_k$ are orthonormal, we have the expansions

$$\begin{cases} \mathbf{v}_1 = (\mathbf{v}_1 \cdot \mathbf{q}_1)\mathbf{q}_1 \\ \mathbf{v}_2 = (\mathbf{v}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{v}_2 \cdot \mathbf{q}_2)\mathbf{q}_2 \\ \mathbf{v}_3 = (\mathbf{v}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{v}_3 \cdot \mathbf{q}_2)\mathbf{q}_2 + (\mathbf{v}_3 \cdot \mathbf{q}_3)\mathbf{q}_3. \end{cases}$$

The very intentional typesetting should reveal a "triangular" structure.

Work backwards:

$$\mathbf{v}_1 = (\mathbf{v}_1 \cdot \mathbf{q}_1)\mathbf{q}_1 = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{q}_1 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{v}_2 = (\mathbf{v}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{v}_2 \cdot \mathbf{q}_2)\mathbf{q}_2 = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} \begin{bmatrix} \mathbf{v}_2 \cdot \mathbf{q}_1 \\ \mathbf{v}_2 \cdot \mathbf{q}_2 \end{bmatrix},$$

and

$$\mathbf{v}_3 = (\mathbf{v}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{v}_3 \cdot \mathbf{q}_2)\mathbf{q}_2 + (\mathbf{v}_3 \cdot \mathbf{q}_3)\mathbf{q}_3 = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} \begin{bmatrix} \mathbf{v}_3 \cdot \mathbf{q}_1 \\ \mathbf{v}_3 \cdot \mathbf{q}_2 \\ \mathbf{v}_3 \cdot \mathbf{q}_3 \end{bmatrix}.$$

Then

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} \begin{bmatrix} (\mathbf{v}_1 \cdot \mathbf{q}_1) & (\mathbf{v}_2 \cdot \mathbf{q}_1) & (\mathbf{v}_3 \cdot \mathbf{q}_1) \\ 0 & (\mathbf{v}_2 \cdot \mathbf{q}_2) & (\mathbf{v}_3 \cdot \mathbf{q}_2) \\ 0 & 0 & (\mathbf{v}_3 \cdot \mathbf{q}_3) \end{bmatrix}.$$

Put

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}, \qquad Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix}, \qquad \text{and} \qquad R = \begin{bmatrix} (\mathbf{v}_1 \cdot \mathbf{q}_1) & (\mathbf{v}_2 \cdot \mathbf{q}_1) & (\mathbf{v}_3 \cdot \mathbf{q}_1) \\ 0 & (\mathbf{v}_2 \cdot \mathbf{q}_2) & (\mathbf{v}_3 \cdot \mathbf{q}_2) \\ 0 & 0 & (\mathbf{v}_3 \cdot \mathbf{q}_3) \end{bmatrix}$$

to see that we have factored the matrix A (which has independent columns) into the product $A = QR$, with Q orthogonal and R upper-triangular. In fact, the diagonal entries of R are positive (not just nonzero) by Problem 4.11.3.

We have a lot of recent knowledge about why orthogonal matrices are nice, and we have a lot of past knowledge about why upper-triangular matrices with nonzero diagonal entries are nice. We put all of that together with this "QR-factorization" to obtain the ultimate form of least squares.

Content from Strang's *ILA* 6E. Page 182 develops the QR-factorization for a matrix with three independent columns.

4.12.1 Theorem (QR-factorization). *Let $A \in \mathbb{R}^{m\times n}$ have independent columns (so A has full column rank: $\operatorname{rank}(A) = n$). There exist an orthogonal matrix $Q \in \mathbb{R}^{m\times n}$ and an upper-triangular matrix $R \in \mathbb{R}^{n\times n}$ such that $A = QR$. Specifically, the columns of Q are the vectors constructed from the columns of A by the Gram–Schmidt procedure, and the (i, j)-entry of R is $\mathbf{a}_j \cdot \mathbf{q}_i$, where $\mathbf{a}_j$ is the jth column of A, and $\mathbf{q}_i$ is the ith column of Q.*

We proved the $n = 3$ case of this. The general proof just hinges on (1) the orthonormality of the vectors produced by Gram–Schmidt, (2) the "span preservation property" that $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j)$ for each j, not just $j = n$, and (3) Problem 4.11.3 to get the positive diagonal entries in R.

4.12.2 Example. Let

$$A = \begin{bmatrix} 1 & 1 & -2 \\ 2 & 2 & 1 \\ 0 & 2 & 0 \end{bmatrix}.$$

We performed Gram–Schmidt on the columns of A in Example 4.11.4. Collect the Gram–Schmidt output in

$$Q = \begin{bmatrix} 1/\sqrt{5} & 0 & -2/\sqrt{5} \\ 2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \end{bmatrix}.$$

If we have forgotten the coefficients from the Gram–Schmidt work, we can compute them quickly (and we only do this for R_{ij} with $j \geq i$):

$$R_{11} = \mathbf{a}_1 \cdot \mathbf{q}_1 = \sqrt{5},$$

$$R_{12} = \mathbf{a}_2 \cdot \mathbf{q}_1 = \sqrt{5},$$

$$R_{13} = \mathbf{a}_3 \cdot \mathbf{q}_1 = 0,$$

$$R_{22} = \mathbf{a}_2 \cdot \mathbf{q}_2 = 2,$$
$$R_{23} = \mathbf{a}_3 \cdot \mathbf{q}_2 = 0,$$
$$R_{33} = \mathbf{a}_3 \cdot \mathbf{q}_3 = \sqrt{5}.$$

Then

$$R = \begin{bmatrix} \sqrt{5} & \sqrt{5} & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \sqrt{5} \end{bmatrix},$$

and we have

$$\begin{bmatrix} 1 & 1 & -2 \\ 2 & 2 & 1 \\ 0 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} & 0 & -2/\sqrt{5} \\ 2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{5} & \sqrt{5} & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \sqrt{5} \end{bmatrix}$$

4.12.3 Problem ($\star$). The QR-factorization is our third major matrix factorization, after CR and LU.

(i) Let $A \in \mathbb{R}^{m \times n}$ have independent columns. What is the CR-factorization of A? When is the R in that factorization the same as the R in the QR-factorization?

(ii) Trefethen & Bau's beautiful *Numerical Linear Algebra* describes Gaussian elimination and the LU-factorization as "triangular triangularization" and the QR-factorization as "triangular orthogonalization" (p. 148). Why do you think Trefethen & Bau chose those phrases to describe LU and QR?

(iii) Given $A \in \mathbb{R}^{m \times n}$, briefly summarize what guarantees the existence of the CR-, LU-, and QR-factorizations. Does a factorization always exist, or do A, m, and n need to satisfy some extra hypotheses?

Here is how the QR-factorization is useful for least squares. Start with $A \in \mathbb{R}^{m \times n}$ with independent columns and factor $A = QR$ with $Q \in \mathbb{R}^{m \times n}$ orthogonal and $R \in \mathbb{R}^{n \times n}$ upper-triangular with positive diagonal entries.

If $\mathbf{b} \notin \mathbf{C}(A)$, then solving $A\widehat{\mathbf{x}} = P_A\mathbf{b}$ is the next best thing to solving the unsolvable problem $A\mathbf{x} = \mathbf{b}$. While we do have a formula for $\widehat{\mathbf{x}}$, the annoying thing is that it requires computing the inverse $(A^\mathsf{T} A)^{-1}$. Better to solve a linear system than compute an inverse, and, from (4.10.1), solving $A\widehat{\mathbf{x}} = P_A\mathbf{b}$ is equivalent to solving the normal equation

$$(A^\mathsf{T} A)\widehat{\mathbf{x}} = A^\mathsf{T}\mathbf{b}. \tag{4.12.1}$$

We could always use Gaussian elimination to solve (4.12.1). Again, better to solve a linear system than compute an inverse.

But this system (4.12.1) is pretty nice after the QR-factorization. Since $A = QR$, we have $A^\mathsf{T} = (QR)^\mathsf{T} = R^\mathsf{T} Q^\mathsf{T}$, and since Q is orthogonal, we have $Q^\mathsf{T} Q = I_n$. Then

$$A^\mathsf{T} A = (R^\mathsf{T} Q^\mathsf{T})(QR) = R^\mathsf{T}(Q^\mathsf{T} Q)R = R^\mathsf{T} I_n R = R^\mathsf{T} R.$$

The problem (4.12.1) now reads

$$R^{\mathsf{T}}R\widehat{\mathbf{x}} = R^{\mathsf{T}}Q^{\mathsf{T}}\mathbf{b}. \tag{4.12.2}$$

This is great! Since R is upper-triangular with positive diagonal entries, R^{T} is lower-triangular with positive diagonal entries.

4.12.4 Problem ($\star$). Prove that. [Hint: *recall that $R^{\mathsf{T}}_{ij} = R_{ji}$. Since R has positive diagonal entries, $R_{ii} > 0$. Since R is upper-triangular, $R_{ij} = 0$ for $i > j$. To show that R^{T} is lower-triangular, you want $R^{\mathsf{T}}_{ij} = 0$ for $j > i$. Is this true?*]

Any triangular matrix with nonzero diagonal entries is invertible, so R^{T} is invertible. Then (4.12.2) is just

$$R\widehat{\mathbf{x}} = Q^{\mathsf{T}}\mathbf{b}.$$

Again, R is upper-triangular with positive diagonal entries, so we can solve this system by back-substitution (no need even for Gaussian elimination!) No inverses anywhere in the actual calculations, just in the theory.

Content from Strang's *ILA* 6E. Pages 182–183 discuss how to use the QR-factorization in least squares. Read the second half of p. 185, which summarizes everything. You don't need to read about the pseudoinverses.

Then read about the "victory of orthogonality" on p. 197. Stop with #5 for now, and there just be able to explain why if $Q \in \mathbb{R}^{n\times n}$ is orthogonal, so is any power Q^k for $k \geq 1$.

In the paragraph after that, note the "sum of squares definition of length." There are many valid, meaningful ways of defining the length of a vector (pp. 355–356), but the way that interacts best with the dot product is saying length is $\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}}$. As you think about least squares, keep in mind how length and dot product interact so nicely.

4.12.5 Problem (!). Let $A \in \mathbb{R}^{m\times n}$ have full column rank, and let $A = QR$ be its QR-factorization.

(i) Suppose that $A\mathbf{x} = \mathbf{b}$. Show then that $R\mathbf{x} = Q^{\mathsf{T}}\mathbf{b}$, as expected.

(ii) Conversely, we can always solve $R\mathbf{x} = Q^{\mathsf{T}}\mathbf{b}$. (Why?) Why does multiplying by Q not necessarily get us back to $A\mathbf{x} = \mathbf{b}$?