MATH 3260: LINEAR ALGEBRA I

Daily Log for Lectures and Readings

Timothy E. Faver

February 19, 2025

CONTENTS

How to Use This Daily Log . . . 3
Day 1: Monday, January 6 . . . 4
Day 2: Wednesday, January 8 . . . 8
Day 3: Friday, January 10 . . . 14
Day 4: Monday, January 13 . . . 15
Day 5: Wednesday, January 15 . . . 19
Day 6: Friday, January 17 . . . 25
Day 7: Wednesday, January 22 . . . 29
Day 8: Friday, January 24 . . . 34
Day 9: Monday, January 27 . . . 37
Day 10: Wednesday, January 29 . . . 43
Day 11: Friday, January 31 . . . 50
Day 12: Monday, February 3 . . . 53
Day 13: Wednesday, February 5 . . . 59
Day 14: Friday, February 7 . . . 64
Day 15: Monday, February 10 . . . 67
Day 16: Wednesday, February 12 . . . 71
Day 17: Friday, February 14 . . . 74
Day 18: Monday, February 17 . . . 74
Day 19: Wednesday, February 19 . . . 79

How to Use This Daily Log

This document is our primary reference for the course. It contains all of the material that we discuss in class along with some supplementary remarks that may not be mentioned in a class meeting. Each individual day has references, when applicable, to relevant material from the text *Introduction to Linear Algebra (Sixth Edition)* by Gilbert Strang. These references are spread throughout a day's notes, and you should be consulting both the daily log and Strang's text more or less simultaneously.

This log contains several classes of problems.

(!) Problems marked (!) are meant to be attempted *immediately*. They will directly address or reinforce something that we covered (or perhaps omitted) in class. It will be to your great benefit to pause and work (!)-problems as you encounter them.

($\star$) Problems marked ($\star$) are intentionally more challenging and deeper than (!)-problems. The ($\star$)-problems will summarize and generalize ideas that we have discussed in class and give you broader, possibly more abstract perspectives. You should attempt the ($\star$)-problems on a second rereading of the lecture notes, after you have completed the (!)-problems. Completing all of the ($\star$)-problems constitutes the *minimal* preparation for exams.

(+) Problems marked (+) are meant to be more challenging than the (!)- and ($\star$)-problems and will take you deeper into calculations and proofs and make connections to concepts across and beyond the course. It will not be necessary to do any (+)-problems to master the essential material of the course, but your experience may be richer (and more meaningful, and more fun) by considering them. If you have done all of the (!)- and ($\star$)-problems, and the required and recommended problems from the textbook, and if you're still feeling bored or wondering if something is "missing," check out the (+)-problems.

Day 1: Monday, January 6.

Linearity pervades mathematics and science. An "operator" is LINEAR if (1) we can "add" its inputs and outputs in ways that "respect" all of the "usual" properties of addition of real numbers, (2) we can "multiply" its inputs and outputs by numbers in ways that, again, "respect" all of the "usual" properties of multiplication of real numbers, and (3) if the operator itself "respects" addition and "scalar" multiplication. Lots of quotes, lots of words, here are some symbols.

Let $\mathcal{A}$ be that operator and let x and y be inputs. Then there is a notion of adding x and y so that $x + y$ is another input and behaves the way that we expect $+$ to behave; for example, $x+y = y+x$. We can also add the outputs $\mathcal{A}x$ and $\mathcal{A}y$, and this addition behaves as we expect, e.g., $\mathcal{A}x + \mathcal{A}y = \mathcal{A}y + \mathcal{A}x$. Here we are using the same symbol $+$ for addition on both the input side and the output side, even though the inputs and outputs could "live" in totally different "universes."

And there is a notion of multiplying inputs and outputs by real numbers, which we denote by "juxtaposition." That is, if c is a real number and x is an input, then cx is another input, and we have properties like $c(x + y) = cx + cy$ and $(cd)x = c(dx)$.

Now here is how $\mathcal{A}$ respects addition and scalar multiplication:

$$\mathcal{A}(x + y) = \mathcal{A}x + \mathcal{A}y \quad \text{and} \quad \mathcal{A}(cx) = c(\mathcal{A}x). \tag{1.1}$$

These two identities are what we mean by the linearity of $\mathcal{A}$.

You already know some linear operators because you can do arithmetic and calculus. Say that we define $\mathcal{A}x := 2x$ for real numbers x. Or that we define $\mathcal{A}f := f'$, with f' as the derivative of a differentiable function f.

1.1 Problem (!). Prove that if $\mathcal{A}$ is defined in either of these ways, then the linear identities (1.1) hold.

Many, many problems possess a linear structure. The inputs and outputs obey natural rules of addition and scalar multiplication, and the problem is either governed by or well-approximated by a linear operator. You've already met such problems in calculus. Many differential equations are linear; a fundamental problem of physics asks us to find functions f such that $f'' + f = g$ for a given function g. This compresses as $\mathcal{A}f = g$ with $\mathcal{A}f := f'' + f$ as a linear operator satisfying (1.1). Or maybe you want to approximate a hard problem in a nice way; you know how to do this with a local linear approximation. If you want to study a function f around a point x, and f is complicated, study instead $f(x+h) \approx f(x) + f'(x)h$. Here $\mathcal{A}h := f'(x)h$ is linear.

Much of the point of calculus is to *be wise and linearize*; the point of linear algebra is to understand the linear structure of that approximation. However, this is about as much calculus as we'll do in this course. We will focus on a particular kind of linear operator that arises from *linear systems of equations*. These are hugely worthwhile problems in and of themselves, and many other problems that don't look like linear systems of equations either hinge obliquely or are well-approximated by such systems. (Unfortunately, or fortunately, we will not see many, if any, "concrete" examples of *applications* of such systems—every subdiscipline of math, and every scientific discipline allied with math, has its own favorite examples, and you probably won't be convinced of the worth of any one of those examples if you aren't already convinced.)

We can tease out a tremendous amount of structure and theory from very simple motivating examples, and here will be our favorite for the foreseeable future. Let's try to solve the **LINEAR SYSTEM**

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11. \end{cases} \tag{1.2}$$

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

By the way, you didn't need to get out of bed today and come to class to figure out how to solve it, but imagine if the system had 50 variables and 50 equations. You'd probably want a precise and systematic way of approaching it.

1.2 Problem (!). Try to solve (1.2). What does your gut instinct say to do? (If you're reading these notes for the first time and haven't been in class, *don't* read below this problem for our approach just yet—try it by yourself.)

Before we do anything to (1.2), 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) true?

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

3. 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) true? That is, is the solution **UNIQUE**?

4. If not, why not? Can we quantify or qualify why a linear system might have more than one solution?

We will solve (1.2) by transforming it into an "equivalent" system of equations that is much easier to solve—actually, several "equivalent" systems. We'll say that two systems are **EQUIVALENT** if they have precisely the same solutions. And we'll do this via algebra.

Recall that if a and b are real numbers, then $a = b$ if and only if $ac = bc$ for all nonzero c. That is, if you know $a = b$, then you also know $ac = bc$ for all nonzero c (actually for $c = 0$, too, although that's boring). And if you know $ac = ab$ for all nonzero c (actually, for just *one* nonzero c), then you can divide to get $a = b$. In the context of linear systems, scaling *both* sides of the *same* equation by the *same* **nonzero** number doesn't change things. Let's multiply the first equation by the very convenient number -3:

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff \begin{cases} -3x + 6y = -3 \\ 3x + 2y = 11. \end{cases}$$

Now we'll use another property of algebra. Recall that if a and b are real numbers, then $a = b$ if and only if $a + c = b + c$ for all real numbers c. That is, if you know $a = b$, then you can add c to both sides to get $a + c = b + c$. And if you know $a + c = b + c$ for some c, then just subtract c from both sides (or add $-c$ to both sides) to get $a = b$. (Actually, if you're watching your language, we said "for all" c, so if you know $a + c = b + c$ for *all* c, just take $c = 0$.)

Thus

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff \begin{cases} -3x + 6y = -3 \\ 3x + 2y + c = 11 + c \end{cases}$$

for any c that we like. What if we take $c = -3x + 6y$ on the left and $c = -3$ on the right? The first equation says that these two versions of c have to be equal. Thus

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff \begin{cases} -3x + 6y = -3 \\ 3x + 2y + (-3x + 6y) = 11 + (-3) \end{cases}$$

$$\iff \begin{cases} -3x + 6y = -3 \\ 8y = 8 \end{cases}$$

$$\iff \begin{cases} -3x + 6y = -3 \\ y = 1 \end{cases}$$

The second equation is extremely transparent, but the first looks worse than it originally did because of that extra factor of -3. But our work above was redundant; there was no need to keep that -3 multiplying both sides of the first equation, and we could have divided by -3 at any time that we wanted. That is,

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

What this is saying is that x and y satisfy the first system precisely when they satisfy the second—and we know what y is from the second system. With this value of y, the first equation in the second system becomes

$$x - 2 = 1 \iff x = 3.$$

All of our work boils down to

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff \begin{cases} x = 3 \\ y = 1. \end{cases} \tag{1.3}$$

This is an existence and uniqueness result for (1.2): there exists a solution ($x = 3$ and $y = 1$), and it is the only solution. Way to go.

1.3 Problem (!). What would you do if I asked you to *check* that $x = 3$ and $y = 1$ solves (1.2)? Would you repeat all of the work above, or would you just plug in these values and do arithmetic?

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's a third operation—interchanging two equations, which sounds silly but actually is worthwhile—that we'll 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 you are consistent, it doesn't matter if you 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'll just call "lists of numbers" right now. Here are the three important vectors in (1.2), and we'll also write them as ordered pairs to make typesetting easier:

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

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

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

1.4 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) 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) that we did not already know. Let's 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:

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

When multiplying a vector by a number, we always write the number first:

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

1.5 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. There is more componentwise arithmetic on pp. 1–2.

We rewrite (1.2) 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)—yet.

The expression

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

is something that we'll see often: it's a **LINEAR COMBINATION** of the vectors $(1, 3)$ and $(-2, 2)$. By the way, this is an example of typesetting a column vector as an ordered pair to save space. 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 p3 that is written in vector form and then solved with elimination, as we did (1.2).

Day 2: Wednesday, January 8.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Column vector of length n, linear combination of vectors, $m \times n$ matrix, matrix-vector product

Here are some more precise (well, mostly 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 $x \in S$. For example, $1 \in \{1, 2, 3\}$. We denote by $\mathbb{R}$ the set of all real numbers, so $1 \in \mathbb{R}$.

2.1 Undefinition. *Let $n \geq 1$ be an integer. 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 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 distinguish $\mathbb{R}^1$ and $\mathbb{R}$, so $\mathbb{R}^1 = \mathbb{R}$.

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

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

Why is this an "undefinition," not a definition? Because we didn't give a rigorous definition of "ordered list." I like to 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, $f = \{(j, v_j)\}_{j=1}^n$. This is probably a *useless* way to think about column vectors for day-to-day purposes, but it comforts me to know that there is deeper math behind that undefinition. If it doesn't comfort you, it's okay to move on.

We continue to 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 $\mathbf{v}+\mathbf{w} \in \mathbb{R}^n$ and $c\mathbf{v} \in \mathbb{R}^n$. However, we only add vectors that have the same number of components, so something like

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

is not defined.

2.2 Example. We compute

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

I hope it's obvious why we want to call the vector on the right the "zero vector in $\mathbb{R}^3$."

2.3 Definition. *The* **ZERO VECTOR** *in $\mathbb{R}^n$ is the vector $\mathbf{0}$ whose entries are all 0. Sometimes we will write $\mathbf{0}_n$ to emphasize that this is the zero vector with n entries.*

For example,

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

2.4 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?

In studying our motivating toy problem, we encountered a "linear combination" of vectors. Here is that object in general.

2.5 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*

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

We may also express this in sigma notation:

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

2.6 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?

2.7 Problem (!). Let

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

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.

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*.

Here is a major step toward that right language. Recall that our original problem (1.2) can be written as a system of linear equations or as a vector equation involving a linear combination:

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

Let's put the coefficient vectors together into a matrix:

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

I hope you'll agree that this is a "square" matrix: it has 2 rows and 2 columns. We most often think of matrices in terms of columns (though rows are also 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'll say $A \in \mathbb{R}^{2\times 2}$ for the matrix A above, and I hope you believe that

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

More generally, we say the following.

2.8 Definition. *Let m, $n \geq 1$ be integers. An $m \times n$ **MATRIX** is a rectangular array of numbers with m rows and n columns. We denote the set of all $m \times n$ matrices by $\mathbb{R}^{m\times n}$.*

Since a matrix with m rows and 1 column is really just an ordered list of m numbers, we will not distinguish $\mathbb{R}^{m\times 1}$ and $\mathbb{R}^m$, so $\mathbb{R}^{m\times 1} = \mathbb{R}^m$. Also, $\mathbb{R}^{1\times 1} = \mathbb{R}$. But we do not equate $\mathbb{R}^{1\times n}$ and $\mathbb{R}^n$.

*The **(i,j)-ENTRY** 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 $A_{i,j}$. Two matrices are **EQUAL** if and only if they have the same number of rows and columns and if all of their corresponding entries are equal.*

Regarding that last caveat, we have things like

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

2.9 Problem (!). Reread that until it makes sense.

2.10 Example. Let

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

The $(1,2)$-entry of A is 2, and the $(2,1)$-entry of A is 3.

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

As with column vectors, our attempt at defining a matrix is really an undefinition because we did not rigorously define "rectangular array" of numbers. If you really want to, you can 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., $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$. 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.

2.11 Problem (!). Compute

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

We are finally ready to think about linear systems. With

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

how should we define the symbol $A\mathbf{x}$ so that

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

The answer is pretty much staring us in the face:

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

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

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

Let's 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.

2.12 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.

2.13 Example. **(i)** $\begin{bmatrix}1 & 3 & 5 \\ 2 & 4 & 6\end{bmatrix}\begin{bmatrix}1 \\ 0 \\ 1\end{bmatrix} = 1\begin{bmatrix}1 \\ 2\end{bmatrix} + 0\begin{bmatrix}3 \\ 4\end{bmatrix} + 1\begin{bmatrix}5 \\ 6\end{bmatrix} = \begin{bmatrix}1 \\ 2\end{bmatrix} + \begin{bmatrix}0 \\ 0\end{bmatrix} + \begin{bmatrix}5 \\ 6\end{bmatrix} = \begin{bmatrix}1+0+5 \\ 2+0+6\end{bmatrix} = \begin{bmatrix}6 \\ 8\end{bmatrix}$

(ii) $\begin{bmatrix}1 & 0 \\ 0 & 2 \\ 3 & 0\end{bmatrix}\begin{bmatrix}2 \\ 4\end{bmatrix} = 2\begin{bmatrix}1 \\ 0 \\ 3\end{bmatrix} + 4\begin{bmatrix}0 \\ 2 \\ 0\end{bmatrix} = \begin{bmatrix}2 \\ 0 \\ 6\end{bmatrix} + \begin{bmatrix}0 \\ 8 \\ 0\end{bmatrix} = \begin{bmatrix}2+0 \\ 0+8 \\ 6+0\end{bmatrix} = \begin{bmatrix}2 \\ 8 \\ 6\end{bmatrix}$

And now for the definition in symbols.

2.14 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*

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

Content from Strang's *ILA* 6E. Examples of matrix-vector multiplication appear on p. 1.

2.15 Problem (!). Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{v} \in \mathbb{R}^n$. How many entries does $A\mathbf{v}$ have? *Use the definition of $A\mathbf{v}$ from Definition 2.14.*

2.16 Problem ($\star$). Let $A \in \mathbb{R}^{m \times n}$. Prove that $A\mathbf{0}_n = \mathbf{0}_m$. *Use the definition of $A\mathbf{0}_n$ from Definition 2.14.*

2.17 Problem ($\star$). Prove that 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

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

This could involve a lot of $\cdots$ that might obscure the actual arithmetic going on; if it makes things more transparent, do it for $n = 2$ or $n = 3$ first. *However you do it, use the definition of matrix-vector multiplication from Definition 2.14.*

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

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

Our original questions remain the same—how to solve it, how to understand failure to solve it. The new question is probably *Why is writing it as $A\mathbf{x} = \mathbf{b}$ any better than the original way?*

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

Day 3: Friday, January 10.

No class due to weather. The following problems will reinforce our work on matrices and matrix-vector multiplication.

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

3.2 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, $x_2 \in \mathbb{R}$

Day 4: Monday, January 13.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Dot product of vectors in $\mathbb{R}^n$

No class due to conflict with department chair interviews. Please read and work through the following material—word by word, line by line. Check all calculations and details.

The goal of the class is the same as always: solve $A\mathbf{x} = \mathbf{b}$, and when we can't solve it, understand why. 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 2.14. We'll tease it out in an example.

4.1 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

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

right?

Here is another way of looking at this arithmetic:

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

Do you see how the vectors $(3, 1)$ and $(1, -2)$ appear in the first component on the right? And how $(3, 1)$ and $(3, 2)$ appear in the second component? It's almost as though the vector by which we're multiplying the matrix, and the rows of the matrix *viewed as column vectors*, are doing all of the arithmetic.

Let's introduce a new structure: the **DOT PRODUCT** of vectors in $\mathbb{R}^2$. (Just $\mathbb{R}^2$ now for starters.) Put

$$\begin{bmatrix}v_1\\v_2\end{bmatrix} \cdot \begin{bmatrix}w_1\\w_2\end{bmatrix} := v_1w_1 + v_2w_2.$$

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.

Let's generalize this example.

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

$$\mathbf{v} \cdot \mathbf{w} = \begin{bmatrix}v_1\\\vdots\\v_n\end{bmatrix} \cdot \begin{bmatrix}w_1\\\vdots\\w_n\end{bmatrix} = v_1w_1 + \cdots + v_nw_n = \sum_{j=1}^n v_jw_j.$$

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.

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

I will do my best to reserve the symbol $\cdot$ for the dot product and use "juxtaposition" to denote multiplication of real numbers, e.g., $3(1)$, not $3 \cdot 1$. But I guess the dot product in $\mathbb{R}^1 = \mathbb{R}$ is just ordinary multiplication, so no big deal.

4.4 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 numbers x and y, right?

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

4.5 Example. Here is how this works in $\mathbb{R}^3$. (I like $\mathbb{R}^3$: it's big enough to be interesting but not so big that it's intimidating.) 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. I 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 4.3, so here is the second:

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

I'll let you check the third.

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

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ v_3 \end{bmatrix}$$

$$= v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$$

$$= (\mathbf{v} \cdot \mathbf{e}_1)\mathbf{e}_1 + (\mathbf{v} \cdot \mathbf{e}_2)\mathbf{e}_2 + (\mathbf{v} \cdot \mathbf{e}_3)\mathbf{e}_3 = \sum_{j=1}^{3} (\mathbf{v} \cdot \mathbf{e}_j)\mathbf{e}_j.$$

This is a really clean representation of a vector in terms of its components and some other, simpler vectors. We'll return to such representations many times in the future.

4.6 Problem (+). 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.

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

(ii) Prove that

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

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

$$\mathbf{v} = \sum_{j=1}^{n} (\mathbf{v} \cdot \mathbf{e}_j)\mathbf{e}_j.$$

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 4.1. We'll work with a matrix with three columns to see this a little more abstractly. Let $A \in \mathbb{R}^{m \times 3}$ and write

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

I just want to focus on the first row of A, so I've listed that out explicitly. The symbols $*$ below denote the remaining $m - 1$ rows of A. The exact values of the entries in those rows are wholly unimportant right now. (If it makes you feel better, take $m = 2$ and replace each $*$ with 0.)

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

$$A\mathbf{v} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ * & * & * \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = v_1 \begin{bmatrix} a_{11} \\ * \end{bmatrix} + v_2 \begin{bmatrix} a_{12} \\ * \end{bmatrix} + v_3 \begin{bmatrix} a_{13} \\ * \end{bmatrix} = \begin{bmatrix} v_1a_{11} + v_2a_{12} + v_3a_{13} \\ * \end{bmatrix}.$$

At the risk of being annoying, I am using the same symbol $*$ to denote rows 2 through m of the columns of A and then the vector $A\mathbf{v}$; I sincerely don't care what's going on there right now. Here is what we have shown: the first component of $A\mathbf{v}$ is

$$v_1a_{11}+v_2a_{12}+v_3a_{13}=\begin{bmatrix}a_{11}\\a_{12}\\a_{13}\end{bmatrix}\cdot\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}=\begin{bmatrix}a_{11}\\a_{12}\\a_{13}\end{bmatrix}\cdot\mathbf{v},$$

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.

4.7 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 $\mathbf{v}$.*

4.8 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 2.14?

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. 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.

4.9 Problem ($\star$). Go back and redo each of the matrix-vector products in Example 2.13 and Problem 3.2 with dot products. What do you find easier for work by hand: Definition 2.14 or Theorem 4.7?

Day 5: Wednesday, January 15.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Column space of a matrix

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" (dare I say "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. I first claim that matrix-vector multiplication can "extract" the columns of a matrix. Let's start small. As before, we'll write

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

Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \in \mathbb{R}^{m \times 3}$. It's important that A has only three columns, but here the number of rows doesn't matter. 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* $\mathbf{e}_1$ *extracted the first column of* A*.*

5.1 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.

5.2 Theorem. *Let* $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ *be the standard basis vectors for* $\mathbb{R}^n$*: the components of* $\mathbf{e}_j$ *are all* 0*, except for the component in row* j*, which is* 1*. Let* $A \in \mathbb{R}^{m \times n}$*. Then* $A\mathbf{e}_j$ *is the* j*th column of* A*.*

5.3 Problem (⋆). Prove it!

5.4 Problem (⋆). Let $I_n \in \mathbb{R}^{n \times n}$ be the matrix whose jth column is $\mathbf{e}_j$. We might write $I_n = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix}$. Prove that $I_n\mathbf{v} = \mathbf{v}$ for any $\mathbf{v} \in \mathbb{R}^n$. We therefore call I_n the **IDENTITY MATRIX**: multiplying $\mathbf{v}$ by I_n just tells you what $\mathbf{v}$ is. [Hint: *prove it for* $n = 3$ *first to see the pattern of the arithmetic before doing it for* n *arbitrary.*]

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 can go further than data extraction via matrix-vector multiplication. I like to say that *what things do defines what things are.* And what matrices do is multiply vectors! Recall that two matrices A, $B \in \mathbb{R}^{m \times n}$ are equal if their corresponding entries are all equal: $A_{ij} = B_{ij}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. That is a "static" way of viewing matrix equality (and not a bad way at all). Here is the "dynamic" way: A and B are equal if they always do the same thing to the same vector.

5.5 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. This is an "if and only if" statement, so we need to prove two things. First we want to assume that $A = B$ and then show that $A\mathbf{v} = B\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. This feels pretty silly, right? We should just be able to "substitute" A in for B. If we want to be pickier and more precise about what $=$ means here, $A = B$ means that A and B have equal entries, so also equal columns. That is, $A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}$ and $B = \begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_n\end{bmatrix}$ with $\mathbf{a}_j = \mathbf{b}_j$ for all j. (And what does $\mathbf{a}_j = \mathbf{b}_j$ mean? Componentwise equality.) So, if $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$, then

$$A\mathbf{v} = \sum_{j=1}^n v_j \mathbf{a}_j = \sum_{j=1}^n v_j \mathbf{b}_j = B\mathbf{v}.$$

Now we want to show that if $A\mathbf{v} = B\mathbf{v}$ 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 $\mathbf{v} = \mathbf{e}_j$. 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. That means $A = B$. ■

5.6 Problem (!). Are you sure about that? If $A, B \in \mathbb{R}^{m \times n}$, and the jth column of A equals the jth column of B for $j = 1, \ldots, n$, why do we have $A = B$? [Hint: *there are several versions of equality here: equality of vectors in $\mathbb{R}^m$, equality of matrices in $\mathbb{R}^{m\times n}$, equality of numbers in $\mathbb{R}$. What role does each version play in answering the question?*]

We now have as good an understanding of matrix-vector multiplication as we're going to get without doing anything new. 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. We give this a special name.

5.7 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,*

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

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

5.8 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 $\mathbf{C}(A) = \mathbb{R}^m$. Then every $\mathbf{b} \in \mathbb{R}^m$ would be in $\mathbf{C}(A)$, so 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 $\mathbf{b} \in \mathbb{R}^m$. Whether or not that is true, $\mathbf{C}(A)$ consists of all the vectors in $\mathbb{R}^m$ for which we can solve $A\mathbf{x} = \mathbf{b}$. Right now, understanding the column space does not replace solving $A\mathbf{x} = \mathbf{b}$, and understanding the column space gives us no new tools or algorithms for solving $A\mathbf{x} = \mathbf{b}$ efficiently. That's coming.

5.9 Example. With A as in Example 5.8, 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 5.8, 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.

5.10 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 brought us from* (1.2) *to* (1.3) *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: *don't reinvent the wheel. 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*

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

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?*]

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.

5.11 Example. **(i)** Let

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

If $\mathbf{v} = (v_1, v_2)$, then $A\mathbf{v} = (2v_1, 0)$. That is, the second component of $A\mathbf{v}$ is always 0, so if $\mathbf{b} = (b_1, b_2) \in \mathbf{C}(A)$, then $b_2 = 0$. Surely not all vectors in $\mathbb{R}^2$ have 0 as their second component; for example, $(1, 1) \notin \mathbf{C}(A)$.

(ii) Let

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

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

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

(You believe that multiplication distributes over addition, right? That $c_1\mathbf{v} + c_2\mathbf{v} = (c_1 + c_2)\mathbf{v}$?) 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

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

I think you'll agree that 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. If not, let's use dot products for a change:

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

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*.

5.12 Example. Let's take another look at those matrices. I think it's easier to start with

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

and recall our arithmetic to see that the second column is -2 times the first column. Then maybe we'll 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!

Is there redundancy in

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

I claim that no column is a multiple of another—this is annoying to check, but it builds character, so you should do it. (Here's how to get started: if the first column is c times the second column, won't some of the zero and nonzero entries interact badly?) But maybe, if we're lucky, we'll notice patterns relating the third column to the first and second. Because life is short, I'll tell you those patterns:

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

The third column is a linear combination of the first two. This is redundancy again: at a "linear" level, the third column can be recovered from the first two.

In fact, the third column just disappears when looking at the column space:

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

Content from Strang's *ILA* 6E. Look at Examples 1, 2, and 3 on p. 20. Check that we can always solve $A_1\mathbf{x} = \mathbf{b}$ for any $\mathbf{b}$, and that there is only one choice of $\mathbf{x}$ that works. Then check all of the arithmetic that appears in the statements about A_2 and A_3.

So why is this bad? Why do "redundant" columns make the column space smaller than we'd like Do they always do that? And can we be more precise than "redundant"?

5.13 Problem (+). Here is a generalization of these issues.

(i) Let $\mathbf{a} \in \mathbb{R}^2$, $c \in \mathbb{R}$, and $A = \begin{bmatrix} \mathbf{a} & c\mathbf{a} \end{bmatrix}$. First explain why every vector in $\mathbf{C}(A)$ is a constant multiple of $\mathbf{a}$. Then find $\mathbf{b} \in \mathbb{R}^2$ such that $\mathbf{b} \notin \mathbf{C}(A)$. (By the way, $\notin$ just means "is not an element of.") [Hint: *what happens if both* $\mathbf{e}_1$, $\mathbf{e}_2 \in \mathbf{C}(A)$*?*]

(ii) Let $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^3$ and $c_1, c_2 \in \mathbb{R}$, and

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & (c_1\mathbf{a}_1 + c_2\mathbf{a}_2) \end{bmatrix}.$$

First explain why every vector in $\mathbf{C}(A)$ can be written as $v_1\mathbf{a}_1 + v_2\mathbf{a}_2$ for some $v_1, v_2 \in \mathbb{R}$. If $\mathbf{C}(A) = \mathbb{R}^3$, does it feel weird that the "three-dimensional" space $\mathbb{R}^3$ can be described by varying only two parameters v_1 and v_2? Try to find $\mathbf{b} \in \mathbb{R}^3$ such that $\mathbf{b} \notin \mathbf{C}(A)$. I expect this to be annoying, since I'm not telling you what the entries of $\mathbf{a}_1$ and $\mathbf{a}_2$ are, but go as far as you can and see if you get stuck. It's a +-problem, after all.

Day 6: Friday, January 17.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Span of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$, matrix with dependent columns (N), matrix with independent columns (N)

Here was the problem with the matrices in Examples 5.11 and 5.12: 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. Somehow this prevented the column space from being as large as possible. Our job is to understand why.

First, now is a good time to review and augment our vocabulary. We want to understand $A\mathbf{x} = \mathbf{b}$ for $A \in \mathbb{R}^{m \times n}$, and to be able to solve this we want $\mathbf{b} \in \mathbf{C}(A)$, where

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

If $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ with $\mathbf{a}_j \in \mathbb{R}^m$, we can also write

$$\mathbf{C}(A) = \{c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n \mid c_1, \ldots, c_n \in \mathbb{R}\}.$$

That is, $\mathbf{C}(A)$ is the set of all linear combinations of the columns of A. We may want to consider such sets of linear combinations not strictly in the context of columns of a matrix.

6.1 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* $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

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

6.2 Example. **(i)** Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. Then $\mathbf{C}\big(\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}\big) = \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

(ii) Let $\mathbf{v} \in \mathbb{R}^m$. Then $\operatorname{span}(\mathbf{v}) = \{c\mathbf{v} \mid c \in \mathbb{R}\}$.

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

6.4 Problem ($\star$). Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$. Prove that $\mathbf{0}_m \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

6.5 Problem ($\star$). Let $\mathbf{v}_1$, $\mathbf{v}_2 \in \mathbb{R}^m$ and c, c_1, $c_2 \in \mathbb{R}$. Prove that

$$\operatorname{span}(\mathbf{v}_1, c\mathbf{v}_1) = \operatorname{span}(\mathbf{v}_1) \quad \text{and} \quad \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2).$$

Explain how these "small" spans illustrate the following general principle: the span of a list of vectors equals the span of those vectors in the list that are not linear combinations of other vectors in the list.

The problem with the matrices in Examples 5.11 and 5.12 was twofold: 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 $\mathbf{b}$), and 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.

6.6 Definition. *The columns of a matrix $A \in \mathbb{R}^{m \times n}$ are* **DEPENDENT** *if (at least) one column is in the span of the others, i.e., 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 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." And here is our conjecture.

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

This conjecture encapsulates the situation of the matrices in Examples 5.11 and 5.12. Unfortunately, it's only a conjecture right now, and we don't yet have the tools to prove it. And even when we know it's 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.

6.8 Problem (!). We can talk about a nonsquare matrix with dependent columns, but the conjecture was only for a square matrix. Here's 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.

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

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

are dependent. [Hint: $\mathbf{v} = 1\mathbf{v}$.] Conclude that if a matrix contains the same column two or more times, then its columns are dependent.

We probably think that the opposite of "dependent" is "independent," and so the columns of a matrix should be "independent" if no column is in the span of the others, i.e., if no column is a linear combination of the others. This could be hard to check! We'd have to fix our attention on one column at a time and the compare it to every other column. That could take forever. Here is a better definition of "independent," although it requires a little more technical notation.

6.10 Definition. *The columns of $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$ are* **INDEPENDENT** *if $\mathbf{a}_1 \neq \mathbf{0}_m$ and if $\mathbf{a}_j$ is not a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}$ for $j = 2, \ldots, n$ (if, indeed, $n \geq 2$). That is, $\mathbf{a}_1 \neq \mathbf{0}_m$ and $\mathbf{a}_j \notin \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1})$ for $j = 2, \ldots, n$.*

In the particular case that $n = 1$ and A has only one column, then we say that this column is independent if it is not the zero vector.

I think we better do a concrete example right away.

6.11 Example. Let

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

Certainly

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

Next, we want to check that $\mathbf{a}_2 \notin \operatorname{span}(\mathbf{a}_1)$. 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 $\mathbf{a}_3 \notin \text{span}(\mathbf{a}_1, \mathbf{a}_2)$. 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 $6 = 0$.

6.12 Problem (!). Where does the algorithm for independence in Definition 6.10 break down with the matrix

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

6.13 Remark. *Why do we exclude the case $\mathbf{a}_1 = \mathbf{0}_m$ so specifically in the definition of independent columns? Honestly, "because it's the right thing to do." This definition of independent columns is, in a moment, going to give us the correct analogue of Conjecture 6.7. (Well, "correct" once we prove it.)*

That's not very satisfying right now, I realize. Perhaps the better reason to exclude $\mathbf{a}_1 = \mathbf{0}_m$ is because if any column of a matrix is the zero vector, then the columns are dependent: that zero column is the linear combination of all the other columns with weights equal to 0. Think about part (i) of Example 5.11.

And why do we not say that subsequent columns of the matrix beyond the first can't be the zero vector? Because the condition $\mathbf{a}_j \notin \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1})$ excludes that: we know $\mathbf{0}_m \in \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1})$ by Problem 6.4, so if $\mathbf{a}_j \notin \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1})$, then it's definitely the case that $\mathbf{a}_j \neq \mathbf{0}_m$.

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 here is the analogue of Conjecture 6.7 for a matrix with independent columns.

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

6.15 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 $\mathbf{C}(A) \neq \mathbb{R}^3$.

6.16 Problem ($\star$). Here is another reason to prevent the first column from being the zero vector, beyond Remark 6.13. Suppose that at least one column of $A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_m\end{bmatrix} \in \mathbb{R}^{m\times m}$ is the zero vector $\mathbf{0}_m$. Explain why $\operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ can be "described" by at most $m-1$ different parameters and therefore $\mathbf{C}(A)$ is at most "$(m-1)$-dimensional."

As with Conjecture 6.7, we do not yet have the tools to prove Conjecture 6.14, nor do we have a systematic way of verifying that a matrix's columns are independent. 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.

6.17 Problem ($+$). Let

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

Of course, $\mathbf{C}(A) = \operatorname{span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. However, we can be more efficient. Show that we can write $\mathbf{C}(A)$ in the following ways:

$$\mathbf{C}(A) = \operatorname{span}(\mathbf{a}_1, \mathbf{a}_2), \qquad \mathbf{C}(A) = \operatorname{span}(\mathbf{a}_2, \mathbf{a}_3), \qquad \mathbf{C}(A) = \operatorname{span}(\mathbf{a}_1, \mathbf{a}_3).$$

But we can't beat that: explain why $\mathbf{C}(A)$ is not the span of any one column of A.

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.

Day 7: Wednesday, January 22.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Independent list of vectors (N), dependent list of vectors (N), rank of a matrix

To describe what columns really matter, we need a variation on our notions of dependence and independence that free us from thinking solely about vectors as columns of matrices. Before we state that, we clarify our expectation that a matrix with dependent columns can't have independent columns.

7.1 Theorem. *Let $A \in \mathbb{R}^{m \times n}$. The columns of A are dependent if and only if they are not independent.*

Proof. This is an "if and only if" statement, so we need our logic to go in two directions. First suppose that the columns of A are dependent; write, as usual, $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$. We need to show that either $\mathbf{a}_1 = \mathbf{0}_m$ or $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1})$ for some $j \geq 2$, if indeed $n \geq 2$. If $\mathbf{a}_1 = \mathbf{0}_m$, then we're done, so assume $\mathbf{a}_1 \neq \mathbf{0}_m$ (and, implicitly, $n \geq 2$). This could get messy in the abstract, so consider the very special case of $n = 4$ with $\mathbf{a}_3$ as a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_4$:

$$\mathbf{a}_3 = c_1\mathbf{a}_1 + c_4\mathbf{a}_4.$$

Our program is to look for the jth column as a linear combination of the previous $j - 1$ columns; it doesn't look like $\mathbf{a}_3$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$ here.

Or is it? Since $c_4 \in \mathbb{R}$, we have two options: $c_4 = 0$ or $c_4 \neq 0$. If $c_4 = 0$, then

$$\mathbf{a}_3 = c_1\mathbf{a}_1 = c_1\mathbf{a}_1 + 0\mathbf{a}_2 \in \operatorname{span}(\mathbf{a}_1, \mathbf{a}_2).$$

This violates the definition of independent columns with $j = 2$. If $c_4 \neq 0$, some algebra (which I will leave you to check, please) gives

$$\mathbf{a}_4 = \left(-\frac{c_1}{c_4}\right)\mathbf{a}_1 + 0\mathbf{a}_2 + \left(\frac{1}{c_4}\right)\mathbf{a}_3 \in \operatorname{span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3).$$

Now here is how this works in general (and I am not expecting you to read this unless you're curious). Say that the columns of $A \in \mathbb{R}^{m \times n}$ are dependent and that column ℓ is a linear combination of the other columns:

$$\mathbf{a}_\ell = \sum_{\substack{k=1 \\ k \neq \ell}}^{n} c_k\mathbf{a}_k.$$

Let j be the largest integer such that $c_j \neq 0$. If $j = \ell - 1$, then $\mathbf{a}_\ell \in \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{\ell-1})$. Otherwise, assume $j \geq \ell + 1$ (since c_ℓ does not exist in the notation above). Then $c_k = 0$ for $k \geq j + 1$, so

$$\mathbf{a}_\ell = \sum_{\substack{k=1 \\ k \neq \ell}}^{j-1} c_k\mathbf{a}_k + c_j\mathbf{a}_j.$$

Since $c_j \neq 0$, this rearranges to

$$\mathbf{a}_j = \left(\frac{1}{c_j}\right)\mathbf{a}_\ell + \sum_{\substack{k=1 \\ k \neq \ell}}^{j-1} \left(-\frac{c_k}{c_j}\right)\mathbf{a}_k \in \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}).$$

Here $\mathbf{a}_\ell$ is in the list $\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}$ since $\ell < \ell + 1 \leq j$. Glad I didn't do that in class?

Now for the other direction: if the columns of A are not independent, they are dependent. First suppose $\mathbf{a}_1 = \mathbf{0}_m$, so $\mathbf{a}_1 \in \text{span}(\mathbf{a}_2, \ldots, \mathbf{a}_n)$ by Problem 6.4. Next suppose $\mathbf{a}_j \in \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1})$ for some $j \geq 2$. Then

$$\mathbf{a}_j = c_1\mathbf{a}_1 + \cdots + c_{j-1}\mathbf{a}_{j-1} = c_1\mathbf{a}_1 + \cdots + c_{j-1}\mathbf{a}_{j-1} + 0\mathbf{a}_{j+1} + \cdots + 0\mathbf{a}_n,$$

and so $\mathbf{a}_j$ is a linear combination of the other columns. ■

This should be fundamentally comforting: "dependent" should mean "not independent." Now we free ourselves from talking about columns of matrices by saying exactly the same thing for lists of vectors.

7.2 Definition. *A list of vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ *is* **INDEPENDENT** *if* $\mathbf{v}_1 \neq \mathbf{0}_m$ *and if* $\mathbf{v}_j \notin \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1})$. *The list is* **DEPENDENT** *if it is not independent. In particular, a list consisting of one vector is dependent if and only if that vector is the zero vector.*

This vocabulary and the following procedure are the keys to talking *efficiently* about the column space of a matrix—to describing it with the minimal amount of data necessary. Say that we start with a list of vectors and consider its span (like we do with the columns of a matrix and the column space of that matrix). How can we remove all of the "unnecessary" vectors from that list so that we arrive at a "sublist" of just enough vectors whose span equals that of the original list? I hope an example helps. We are going to need the general principle suggested by Problem 6.5, so you should (re)read, and maybe (re)do, that problem now.

7.3 Example. Consider the following list of vectors in $\mathbb{R}^4$:

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_5 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{v}_6 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix}.$$

The zero vector contributes nothing to the span:

$$\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_6) = \text{span}(\mathbf{v}_2, \ldots, \mathbf{v}_6).$$

Since $\mathbf{v}_2 \neq \mathbf{0}_4$, we have $\text{span}(\mathbf{v}_1, \mathbf{v}_2) = \text{span}(\mathbf{v}_2)$, and also the list consisting of the single vector $\mathbf{v}_2$ is independent.

Next, $\mathbf{v}_3 = 2\mathbf{v}_2$, so $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \text{span}(\mathbf{v}_2)$, and the list consisting of the single vector $\mathbf{v}_2$ is still independent. Nothing really new yet.

Onward: we have $\mathbf{v}_4 \notin \text{span}(\mathbf{v}_2)$. Why? So, the list $\mathbf{v}_2$, $\mathbf{v}_4$ is independent and $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4) = \text{span}(\mathbf{v}_2, \mathbf{v}_4)$.

I claim that $\mathbf{v}_5 \in \text{span}(\mathbf{v}_2, \mathbf{v}_4)$ and I will leave that for you to check. Thus $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5) = \text{span}(\mathbf{v}_2, \mathbf{v}_4)$. Last, I claim that $\mathbf{v}_6 \notin \text{span}(\mathbf{v}_2, \mathbf{v}_4)$, and I'll also ask you to check that. Thus the list $\mathbf{v}_2$, $\mathbf{v}_4$, $\mathbf{v}_6$ is independent and $\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_6) = \text{span}(\mathbf{v}_2, \mathbf{v}_4, \mathbf{v}_6)$.

In summary, the "sublist" $\mathbf{v}_2$, $\mathbf{v}_4$, $\mathbf{v}_6$ is independent and has the same span as the original list. This is efficient: we have cut the number of vectors needed to describe the span in half.

7.4 Problem (!). Check the things that I asked you to check in the previous example.

Here is the general result.

7.5 Lemma. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a list in $\mathbb{R}^m$ with at least one nonzero vector in the list. There exists an independent sublist $\mathbf{v}_{j_1}, \ldots, \mathbf{v}_{j_r}$ of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that* $\operatorname{span}(\mathbf{v}_{j_1}, \ldots, \mathbf{v}_{j_r}) = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

In Example 7.3 we had $r = 3$, $j_1 = 2$, $j_2 = 4$, and $j_3 = 6$.

Proof (of Lemma 7.5). 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 $\mathbf{v}_{j_1} \neq \mathbf{0}$.

Let $\mathbf{v}_{j_2}$ be 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 $\mathbf{v}_{j_2} \notin \operatorname{span}(\mathbf{v}_{j_1})$.

Let $\mathbf{v}_{j_3}$ be 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. ■

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 the example and lemma above to extract the linearly independent columns that span the column spaces.

Just because we found one sublist that preserves the span doesn't mean there isn't another: reread, and maybe attempt, Problem 6.17 right now. But that problem suggests the following conjecture.

7.6 Conjecture. *Let $A \in \mathbb{R}^{m \times n}$. If $\mathbf{C}(A)$ is the span of r independent columns, then any list of r independent columns of A also spans the column space. No list of fewer than r columns can span the column space, and any list of more than r columns is dependent.*

We don't have the tools to prove this conjecture yet, but it suggests that there is a "threshold" for independence and spans: a number of columns that is "just right" to span the column space efficiently without any redundancy. We give this number a name, even though we don't know how to compute it yet.

7.7 Definition. *The* **RANK** *of a matrix $A \in \mathbb{R}^{m \times n}$ is the length of the longest list of linearly independent columns of A.*

7.8 Example. The rank of the matrix whose columns are the vectors from Example 7.3 is 3.

Actually, we do know how to compute rank if we start out knowing what it should be. Here are some of the "worst" matrices from the point of view of redundancy: they have far more columns than necessary to describe their column spaces. Actually, we only need one column to span the column space.

7.9 Example. **(i)** $\mathbf{C}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = \{\mathbf{0}_2\}$

(ii) $\mathbf{C}\left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\right) = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$

(iii) $\mathbf{C}\left(\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}\right) = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$ as each column is a multiple of the first.

What a waste of storage space!

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.

Perhaps it would be nice if we had a more efficient way of representing such redundant matrices. Is there a way to extract only the columns that are absolutely necessary for representing the column space? (And will this help us solve, and understand, linear systems?)

The right approach is a new tool: matrix multiplication and matrix factorizations. Think about the factoring that you've already done in life before linear algebra. You've factored integers into products of powers of primes:

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

And you've 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.

For example, maybe we could have something like

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

where the first matrix (okay, column vector) is the only column that matters, and the second matrix (okay, row vector) contains the data needed for constructing all of the columns out of this first column. This hinges on defining matrix products in such a way that the definition yields the equality above. How do we do it?

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

Day 8: Friday, January 24.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Matrix-matrix product (a.k.a. matrix multiplication)

As we often say, the goal of this course is to study the problem $A\mathbf{x} = \mathbf{b}$ for given $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Ideally we could solve it and find $\mathbf{x} \in \mathbb{R}^n$ that makes this equation true. If we can't solve it, we should understand why—can we quantify our failure with further information about why $\mathbf{b}$ doesn't work, or can we approximate the problem somehow so that we could solve a related version?

Along the way, we've picked up notation and language to manipulate this problem (linear combinations, spans, matrix-vector multiplication, dot products) and to develop alternate ways of phrasing it. In particular, we have $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{b} \in \mathbf{C}(A)$, where $\mathbf{C}(A)$ is the column space of A: the set of all linear combinations of columns of A. Some of those columns may be redundant and contribute nothing new to the column space, so we are developing the language of dependent and independent vectors to ensure that we work with the minimal amount of data necessary to describe those $\mathbf{b}$ for which the problem $A\mathbf{x} = \mathbf{b}$ makes sense.

Our next great leap forward will be a notion of multiplying two matrices, not just a matrix and a vector. We will, in time, reverse-engineer that multiplication to *factor* matrices to reveal further useful and meaningful data about matrices, and thus about our fundamental problem. In fact, matrix multiplication will give us an algorithm for solving $A\mathbf{x} = \mathbf{b}$, something we haven't really done yet!

So, what is a "good" definition of matrix multiplication? Starting small might help: let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{v} \in \mathbb{R}^n$. I know we've said that $\mathbb{R}^n \neq \mathbb{R}^{n\times 1}$, and we've basically never thought about $n \times 1$ matrices anyway. But any $B \in \mathbb{R}^{n\times 1}$ has the form $B = \begin{bmatrix}\mathbf{b}\end{bmatrix}$ for some $\mathbf{b} \in \mathbb{R}^n$. Of course, usually we think of $\mathbf{b} \in \mathbb{R}^n$ and $\begin{bmatrix}\mathbf{b}\end{bmatrix}$ as being the same object.

Let's break that pattern. For $A \in \mathbb{R}^{m\times n}$ and $\mathbf{v} \in \mathbb{R}^n$, we have $A\mathbf{v} \in \mathbb{R}^m$. Any $C \in \mathbb{R}^{m\times 1}$ has the form $C = \begin{bmatrix}\mathbf{c}\end{bmatrix}$ for some $\mathbf{c} \in \mathbb{R}^m$. So, I think that we can think of matrix-vector multiplication $A\mathbf{v}$ as matrix-matrix multiplication $A\begin{bmatrix}\mathbf{v}\end{bmatrix}$, and through that lens we have

$$A\begin{bmatrix}\mathbf{v}\end{bmatrix} = \begin{bmatrix}A\mathbf{v}\end{bmatrix}.$$

What did we do? The matrix-matrix product $A\begin{bmatrix}\mathbf{v}\end{bmatrix}$ is just the matrix whose only column is the vector $A\mathbf{v}$.

What if the second factor has more columns? Let $B = \begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_p\end{bmatrix}$. If we want to compute AB and continue the pattern above, then we want to multiply each column of B by A. But if $A \in \mathbb{R}^{m\times n}$, then each column of B needs to be in $\mathbb{R}^n$ so that we can do that multiplication. And the matrix-vector product A times column of B yields a vector in $\mathbb{R}^m$. We don't care how many columns of B there are, so p can be arbitrary. Thus $AB \in \mathbb{R}^{m\times p}$.

8.1 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*

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

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.

Here is the first reason for this definition and the restriction on the sizes of A and B: we want this definition to return the usual definition of matrix-vector multiplication when B is a column vector. There are other good reasons (possibly better reasons). They're coming. For now, let's practice.

8.2 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

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

Feel free to do this with the original definition of matrix-vector multiplication or dot products. 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 $A \in \mathbb{R}^{3\times3}$ and $B \in \mathbb{R}^{3\times2}$, the product AB is defined and $AB \in \mathbb{R}^{3\times2}$.

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

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

8.3 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 3.2.

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

8.4 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. ■

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

(i) If a whole row of A is all 0, what do you know about AB?

(ii) If a whole column of B is all 0, what do you know about AB?

Here is something less nice. We expect that the order in which we multiply real numbers doesn't matter: if x, $y \in \mathbb{R}$, then $xy = yx$. Not so for matrices.

8.6 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 8.2 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? Frankly, no. Example 8.2 and Problem 8.3 suggest that matrices fundamentally act not just on vectors but on other matrices: they are dynamic. We can live without commutativity of matrix multiplication but not without dynamic matrix action. Also, most actions aren't commutative—try putting your shoes on before your socks. Order matters.

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.

Day 9: Monday, January 27.

Here is a matrix product that may look a little weird at first glance.

9.1 Example. Let

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

The columns of B are vectors in $\mathbb{R}^1$, and of course we usually think of these as real numbers: $\mathbb{R} = \mathbb{R}^1 (= \mathbb{R}^{1 \times 1})$. But here it is helpful, if silly, to keep the vector point of view. That is, we think that

$$B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix}, \quad \text{where} \quad \mathbf{b}_1 = \begin{bmatrix} 1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 0 \end{bmatrix}, \quad \mathbf{b}_3 = \begin{bmatrix} 2 \end{bmatrix}, \quad \mathbf{b}_4 = \begin{bmatrix} 3 \end{bmatrix}.$$

Now we compute AB by multiplying A against the columns of B. Of course we are going to have

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 & A\mathbf{b}_4 \end{bmatrix},$$

so what are these matrix-vector products?

We start with

$$A\mathbf{b}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

The second equality is the definition of matrix-vector multiplication $A\mathbf{v}$ for $A \in \mathbb{R}^{n\times 1}$ and $\mathbf{v} \in \mathbb{R}^1$: it's a linear combination of one vector. Pretty silly, I know.

Let's do it again in gory detail:

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

I think you'll agree that

$$A\mathbf{b}_3 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \quad \text{and} \quad A\mathbf{b}_4 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}.$$

All together,

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

Just look at that matrix: it's so redundant! Every column is a multiple of the first, and the entries of B tell us how to do that multiplication. Far better to keep the matrix factored as AB so that we can see the important data: the one column in A and the multipliers in B.

This pattern generalizes nicely: for any $\mathbf{a} \in \mathbb{R}^m$ and $c_1, \ldots, c_{n-1} \in \mathbb{R}$

$$\begin{bmatrix} \mathbf{a} & c_1\mathbf{a} & \cdots & c_{n-1}\mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{a} \end{bmatrix} \begin{bmatrix} 1 & c_1 & \cdots & c_{n-1} \end{bmatrix}. \tag{9.1}$$

9.2 Problem (!). Stare at this equality until you believe it. Maybe write something, too.

Content from Strang's *ILA* 6E. Read and work through all of the calculations on pp. 29–30 under "Rank One Matrices and $A = CR$."

The factorization (9.1) is the sort of factorization that we desire. It takes a matrix with a lot of redundant data and breaks it up into the important chunks. The first factor in (9.1) contains the only independent column of the product, and the second factor tells you how to build the other columns out of that one and only one necessary column. Can we get something like this for a matrix with more than one independent column and for which the "building" might be more complicated?

This is one of those times where it's helpful to know where you're going before you get there.

9.3 Example. We compute

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

and we do so by thinking very intentionally about columns this time (not dot products, please, even though that's faster):

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

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

and

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

Thus

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

We've seen this matrix before. We know that, going from left to right, its first two columns are linearly independent, and its third column is a linear combination of the two. The product on the left makes that explicit: the first factor in the product contains the linearly independent columns, and the second factor tells you how to put those columns together.

The structure of the second factor is interesting. The first two columns of the identity matrix $I_2 \in \mathbb{R}^{2\times 2}$ (Problem 5.4) are there, and I will block them off as follows:

$$\begin{bmatrix}1&0&3\\0&1&2\end{bmatrix}=\left[\begin{array}{cc|c}1&0&3\\0&1&2\end{array}\right].$$

The matrix on the right is an example of a **PARTITIONED MATRIX**. This is totally euphemistic and just a nice way of breaking matrices into "submatrices" so that you see more meaningful patterns in the data.

Let me put

$$I_2=\begin{bmatrix}1&0\\0&1\end{bmatrix}\quad\text{and}\quad F=\begin{bmatrix}3\\2\end{bmatrix}.$$

Then

$$\begin{bmatrix}1&0&3\\0&1&2\end{bmatrix}=\begin{bmatrix}I_2&F\end{bmatrix}.$$

This is a **BLOCK MATRIX**: a matrix whose entries are other matrices. Again, totally euphemistic, just a convenient way of seeing the most important parts from a bird's-eye perspective.

The factorization above is the dream. We want to take a matrix $A \in \mathbb{R}^{m \times n}$ with r independent columns and no more than r independent columns—so the rank of A is r—and write it as a product $A = CR$. The matrix C should contain those r independent columns, so $C \in \mathbb{R}^{m \times r}$. We want the matrix $R \in \mathbb{R}^{r \times n}$ to tell us how to build the columns of A out of linear combinations of the columns of C. So C times the jth column of R should give us the jth column of A.

Content from Strang's *ILA* 6E. 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 $\times$ 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.

Unfortunately, we still do not have the tools to do prove that such a factorization exists or to develop an algorithm for computing it "reasonably" and effectively for all but the most trivial and obvious matrices. This is just like we don't have efficient tools for checking dependence or independence of vectors. Let me at least state this factorization as a dream.

9.4 Conjecture. *Let $A \in \mathbb{R}^{m \times n}$ have rank r. Then there exist matrices $C \in \mathbb{R}^{m \times r}$ and $R \in \mathbb{R}^{r \times n}$ such that $A = CR$. In particular, the columns of C are r independent columns of A.*

Content from Strang's *ILA* 6E. 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.

Ideally, we could write

$$R = \begin{bmatrix} I_r & F \end{bmatrix} \tag{9.2}$$

with r as the $r \times r$ identity matrix. If $r = n$ and all of the columns of A are independent, then we just have $R = I_n$, and there is no F block present, since no column is a linear combination of the others in this case. You should think of that F block being possibly fictitious.

9.5 Problem (!). If $A \in \mathbb{R}^{m \times n}$ can be written as $A = CR$ with $C \in \mathbb{R}^{m \times r}$ and R as in (9.2), and if $r < n$, what are the dimensions of that F?

This is actually not quite the structure of R all the time. The snag can be that the first r columns of A may not be the independent ones, and maybe those r independent columns are "interspersed" throughout A.

9.6 Problem (!). Let $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^m$ be independent and $c_1, c_2 \in \mathbb{R}$. Compute

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} 1 & c_1 & 0 \\ 0 & c_2 & 1 \end{bmatrix}$$

and comment on the structure of the matrix that results (which columns are independent and why?).

Content from Strang's *ILA* 6E. This is what Strang means by the parenthetical remark "in correct order" on p. in the displayed equations after "$A = CR$ becomes."

To get around this, the reality is that we'll have to write

$$R = \begin{bmatrix} I_r & F \end{bmatrix} P$$

for a "permutation" matrix P that will reshuffle the columns appropriately. We'll get around to talking about permutation matrices later, but this will mean that A would have the factorization

$$A = C \begin{bmatrix} I_r & F \end{bmatrix} P.$$

Is that allowed? Can we multiply three matrices at once? Will it matter which two matrices we multiply first?

Nope!

9.7 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 you *group* matrices during multiplication doesn't 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.

Content from Strang's *ILA* 6E. Read "AB times $C = A$ times BC" on p. 29.

The proof of Theorem 9.7 is largely a thankless exercise in juggling parentheses, so I will leave that for you to suss out.

9.8 Problem (+). **(i)** Let $A \in \mathbb{R}^{m \times n}$, $B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_p \end{bmatrix} \in \mathbb{R}^{n \times p}$, and $\mathbf{v} = (v_1, \ldots, v_p) \in \mathbb{R}^p$. Explain why each of the following four equalities is true:

$$\begin{aligned} A(B\mathbf{v}) &= A(v_1\mathbf{b}_1 + \cdots + v_p\mathbf{b}_p) \\ &= v_1 A\mathbf{b}_1 + \cdots + v_p A\mathbf{b}_p \text{ [Hint: \textit{Problem 2.17}]} \\ &= \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_p \end{bmatrix} \\ &= (AB)\mathbf{v}. \end{aligned}$$

(ii) Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, and $C \in \mathbb{R}^{p \times q}$. Let $\mathbf{e}_1, \ldots, \mathbf{e}_q$ be the standard basis vectors for $\mathbb{R}^q$ (Problem 4.6). Explain why to prove that $(AB)C = A(BC)$, it suffices to show that

$((AB)C)\mathbf{e}_j = (A(BC))\mathbf{e}_j$.

(iii) Use only the fact that $(DE)\mathbf{v} = D(E\mathbf{v})$ for matrices D, E and vectors $\mathbf{v}$ for which both sides of that equality are defined (as proved above), justify each equality below:

$$((AB)C)\mathbf{e}_j = (AB)(C\mathbf{e}_j) = A(B(C\mathbf{e}_j)) = A((BC)\mathbf{e}_j) = (A(BC))\mathbf{e}_j.$$

All this being said, we still have no idea of how to compute that "CR-factorization" of a matrix unless we are really lucky and see the dependence relations among the columns from the get-go. There is quite a systematic way of doing that, and it is related to proving Conjectures 6.7 and 6.14, and to developing an explicit algorithm for solving $A\mathbf{x} = \mathbf{b}$ when we can actually solve it. That is, all of our dreams will come true through very related techniques.

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.

Let's finally start solving linear systems. We're going to take a break from matrix manipulations—very briefly—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 $A\mathbf{x} = \mathbf{b}$.

9.9 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

$$\begin{cases} x_1 & - & 2x_2 & = & 1 \\ & & 8x_2 & = & 8 \end{cases}$$

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, and don't laugh:

$$\begin{cases} x_1 - 2x_2 = 1 \\ 0 = 8. \end{cases}$$

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, keep laughing:

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

There is really not much to do, since the second equation is both true and doesn't involve unknowns. There's not much more we can do with the first equation, since we don't know the 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. You can pick any x_2 that you want, 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.

Day 10: Wednesday, January 29.

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 9.9.

10.1 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. One way for this to work out is to assume that the first two are false and then show why the third must be true. So, 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). That is, 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 (Problem 2.17).

Now let $c \in \mathbb{R}$ be arbitrary and $\mathbf{x} = \mathbf{x}_1 + c\mathbf{z}$. Then

$$A\mathbf{x} = A(\mathbf{x}_1 + c\mathbf{z}) = A\mathbf{x}_1 + A(c\mathbf{z}) = A\mathbf{x}_1 + c(A\mathbf{z}) = \mathbf{b} + c\mathbf{0}_m = \mathbf{b} + \mathbf{0}_m = \mathbf{b}.$$

The second and third equalities are, again, the linearity of matrix-vector multiplication. Make sure you understand why all of the other equalities are true.

So why does this give infinitely many solutions? Maybe we should have put $\mathbf{x}_c := \mathbf{x}_1 + c\mathbf{z}$ instead to emphasize the dependence of $\mathbf{x}_c$ on the parameter c. (I guess there might be conflicts of notation with $c = 1$ and $c = 2$?) The point is that each different $c \in \mathbb{R}$ generates a different $\mathbf{x}_1 + c\mathbf{z} \in \mathbb{R}^n$: you can, and should, check that if $c_1 \neq c_2$, then $\mathbf{x}_1 + c_1\mathbf{z} \neq \mathbf{x}_1 + c_2\mathbf{z}$. ■

10.2 Problem ($\star$). **(i)** Let $A \in \mathbb{R}^{m \times n}$. Suppose that there is $\mathbf{z} \in \mathbb{R}^n$ such that $\mathbf{z} \neq \mathbf{0}_n$ and $A\mathbf{z} = \mathbf{0}_m$. Let $\mathbf{b} \in \mathbb{R}^m$. Prove that if the problem $A\mathbf{x} = \mathbf{b}$ has a solution, it is not unique.

(ii) Consider the other side of this: if the only solution to $A\mathbf{x} = \mathbf{0}_m$ is $\mathbf{x} = \mathbf{0}_n$, then solutions to $A\mathbf{x} = \mathbf{b}$ (if they exist) are unique. Here's why: if $A\mathbf{x}_1 = \mathbf{b}$ and $A\mathbf{x}_2 = \mathbf{b}$, what does $\mathbf{z} := \mathbf{x}_1 - \mathbf{x}_2$ solve? Why does that imply $\mathbf{x}_1 = \mathbf{x}_2$ and thus uniqueness?

(iii) By considering the vector $\mathbf{z} = (2, -1)$, explain how the previous part generalizes the situation in part (iii) of Example 9.9.

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 $\mathbf{z}$.

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

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 11 \end{bmatrix} \iff \begin{bmatrix} 1 & -2 \\ 0 & 8 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}.$$

The latter system was easy to solve with "back-substitution."

What's up with the $\iff$? Why are these two problems equivalent (in the sense that $\mathbf{x}$ solves one of them precisely when it solves the other)? More abstractly, we started with $A \in \mathbb{R}^{m \times m}$ (for now A will be square), we wanted to solve $A\mathbf{x} = \mathbf{b}$, and we somehow converted or "reduced" the problem to $U\mathbf{x} = \mathbf{c}$, where U was upper-triangular. Then we back-substituted.

10.3 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 $i > j$.*

10.4 Example. Each matrix below is upper-triangular:

$$\begin{bmatrix} 1 & -2 \\ 0 & 8 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \text{and} \qquad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}.$$

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 you do this? How do you "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: $A\mathbf{x} = \mathbf{b} \Longrightarrow U\mathbf{x} = \mathbf{c}$ and $U\mathbf{x} = \mathbf{c} \Longrightarrow A\mathbf{x} = \mathbf{b}$. Having an arrow go one way in math doesn't always mean it goes the other way.

The good news is that we already know how to do this. It's 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.

Specifically, 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 revolution of linear algebra is that we can encode this via matrix multiplication. Whenever we want to "do something" in this class, you should ask yourself how we can accomplish this by multiplying by a suitable matrix.

So, 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

$$E\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix}. \tag{10.1}$$

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 (10.1).

10.5 Problem (!). Check that. Then compute $EA = U$ with A and U as above.

Here is how we're 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. We can, and that's called inverting a matrix, and we'll do that nice and abstractly soon.

10.6 Problem ($\star$). 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 feels like we're 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

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 3 & 2 & 11 \end{array}\right].$$

I like to draw a line separating the $\mathbf{b}$ column when I'm working with actual numbers. 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.

I'm going to tell you the path forward, even if it isn't obvious right now. 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

$$\begin{bmatrix} \circledast & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \to \begin{bmatrix} * & * & * \\ \color{blue}{0} & \color{blue}{*} & \color{blue}{*} \\ \color{blue}{0} & \color{blue}{*} & \color{blue}{*} \end{bmatrix}.$$

I've written the changed entries 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? I am going to tell you the answer, which generalizes all our work with E above. 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.

10.7 Problem (!). 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.

We do an example in glacially slow detail.

10.8 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:

$$E_{21}A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 8 & 7 & 9 \end{bmatrix}.$$

I'll 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 $\mathsf{R2} \mapsto \mathsf{R2} - 2 \times \mathsf{R1}$ 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

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 8 & 7 & 9 \end{bmatrix} \xrightarrow[E_{31}]{\mathsf{R3} \mapsto \mathsf{R3}-4\times\mathsf{R1}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 3 & 5 \end{bmatrix}, \qquad E_{31} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}.$$

Finally, we want to clear out the 3 in the $(3,2)$-entry:

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 3 & 5 \end{bmatrix} \xrightarrow[E_{32}]{\mathsf{R3} \mapsto \mathsf{R3}-3\times\mathsf{R2}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \qquad E_{32} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}.$$

We're 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).

We are going to focus on "reducing" A to an upper-triangular form, and I am going to leave practicing with back-substitution to you. It's mostly just a longer version of part (i) of Example 9.9.

10.9 Problem (!). Use the results (and the notation) of Example 10.8 to solve $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (0, 1, 5)$.

10.10 Problem ($\star$). We prefer upper-triangular matrices, in part for consistency, but "lower-triangular" matrices can be equally nice. Solve

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

10.11 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} and E_{32} *not* to commute. Then do the multiplication to check $E_{31}E_{32} \neq E_{32}E_{31}$.

Day 11: Friday, January 31.

For larger matrices, the pattern of elimination is the same. Use the $(1,1)$-entry to "create zeros" in rows 2 and below of column 1 by subtracting appropriate multiples of row 1 from those lower rows. Then use the *new* $(2,2)$-entry to "create zeros" in the *new* rows 3 and below of the *new* column 2 by subtracting appropriate multiples of the *new* row 2 from those *new* lower rows. Keep going until you've reached the last row and the matrix has been "reduced" to an upper-triangular structure.

This approach to 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.

11.1 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

$$\begin{bmatrix} 2 & 2 & 1 \\ 4 & 4 & 3 \\ 8 & 9 & 9 \end{bmatrix} \xrightarrow[E_{21}]{\text{R2} \mapsto \text{R2}-2\times\text{R1}} \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 8 & 9 & 9 \end{bmatrix} \xrightarrow[E_{31}]{\text{R3} \mapsto \text{R3}-4\times\text{R1}} \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 5 \end{bmatrix}.$$

The matrices E_{21} and E_{31} are the same as before in Example 10.8, so I didn't write them out again.

The problem is that the $(2,2)$-entry is now 0. We can't use that to eliminate the 3 in the $(3,2)$-entry. But if we could "flip" rows 2 and 3, we'd be done. (This is totally legitimate: you can interchange the order of equations in a system of equations and not change the solution structure at all.) 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:

$$\begin{bmatrix} v_1 \\ v_3 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

Are you okay with how I got the last equality? Maybe it would help to rearrange the sum so that v_1, v_2, and v_3 come in order:

$$v_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + v_3 \begin{bmatrix}0\\1\\0\end{bmatrix} + v_2 \begin{bmatrix}0\\0\\1\end{bmatrix} = v_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + v_2 \begin{bmatrix}0\\0\\1\end{bmatrix} + v_3 \begin{bmatrix}0\\1\\0\end{bmatrix}.$$

Here is the result:

$$\begin{bmatrix}2&2&1\\0&0&1\\0&1&5\end{bmatrix} \xrightarrow[P_{23}]{\text{R3} \mapsto \text{R2, R2} \mapsto \text{R3}} \begin{bmatrix}2&2&1\\0&1&5\\0&0&1\end{bmatrix}, \qquad P_{23} := \begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}.$$

I am calling this P_{23} to emphasize that we get it by interchanging columns 2 and 3 of the identity matrix. We'll call such a matrix formed by swapping columns of the identity a **PERMUTATION MATRIX**.

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 10.8, as we have to include a factor of a permutation matrix, not just an elimination mtrix.

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**. So, if at some stage of elimination, the diagonal entry that you want to use to eliminate entries below is 0, but other entries in that column are nonzero, just "permute" the rows to bring that nonzero entry up to the row that you want. Then eliminate as usual in the remaining rows.

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.

11.2 Problem (!). Explain in words (no need for any calculations) why $P_{ij}A = P_{ji}A$.

11.3 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.)

11.4 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 you want to use to eliminate entries below is 0 and *all* other entries in that column are 0, too? Good news is that you don't have to do any more elimination on entries in that column, as they're already 0. Bad news is that you won't be able to solve $A\mathbf{x} = \mathbf{b}$ for all $\mathbf{b}$. Here's a particular example of why.

11.5 Example. Here is a problematic matrix:

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

We eliminate:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 0 & 0 & 5 \end{bmatrix} \xrightarrow[E_{21}]{\text{R2} \mapsto \text{R2}-2\times\text{R1}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \qquad E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Maybe it doesn't 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?

Let's 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 can't always solve $A\mathbf{x} = \mathbf{b}$.

It's worth interpreting this 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.

11.6 Problem ($\star$). Does the "arrow go the other way"? We have shown

$$\mathbf{b} \in \mathbf{C}(A) \Longrightarrow b_2 = 2b_1.$$

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!"

Day 12: Monday, February 3.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Invertible matrix (N), inverse of a matrix

These results will follow and support us for the rest of the course and beyond. Here is an abstraction of our elimination procedure.

12.1 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.*

12.2 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's 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 if you want to solve $U\mathbf{x} = \mathbf{c}$ with $\mathbf{c} = (c_1, c_2, c_3)$, first you'd look at

$$u_{33}x_3 = c_3.$$

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?

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's an algorithm that will let you do that, and we'll see it briefly, but we'll 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 $F(E\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^m$? 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 $F(E\mathbf{v}) = \mathbf{v}$. This just says $(FE)\mathbf{v} = \mathbf{v}$. What does that tell us about the matrix product FE? If $(FE)\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^m$, then we could take $\mathbf{v} = \mathbf{e}_j$ 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. Much later, we'll see that this extra bit is delightfully redundant—this is a surprise, since matrix multiplication usually is not commutative.

12.3 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*

$$FE = I_m \quad \text{and} \quad EF = I_m. \tag{12.1}$$

12.4 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 $F = P$. I suggest that you check that $P^2 = I_2$. By the way, this is the first time we're 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 12.4 should be comforting in that it suggests that elimination and permutation matrices are invertible. We'd probably like 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's why. Suppose that E has "two" inverses F_1 and F_2, so

$$F_1E = EF_1 = F_2E = EF_2 = I_m. \tag{12.2}$$

We need to show that $F_1 = F_2$. Here is a great trick: multiply by 1. You know that $1x = x$ for any $x \in \mathbb{R}$, and the same is true for matrices.

12.5 Problem (!). Check that $AI_m = I_mA = A$ for any $A \in \mathbb{R}^{m\times m}$.

So,

$$F_1 = F_1I_m = F_1(EF_2) = (F_1E)F_2 = I_mF_2 = F_2. \tag{12.3}$$

Here is the formal result.

12.6 Theorem. *Let $E \in \mathbb{R}^{m\times m}$. There exists at most one $F \in \mathbb{R}^{m\times m}$ satisfying* (12.1).

Content from Strang's *ILA* 6E. This is Note 2 on p. 50.

We can now talk about "the" inverse of a matrix.

12.7 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 $F = E^{-1}$.

Let's generalize Example 12.4.

12.8 Theorem. **(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 $P_{ij}^{-1} = P_{ij}$.*

Now go back and look very carefully at the calculation in (12.3). We did not use all of the equalities in (12.2). 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.

12.9 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. Okay, maybe this needs a teensy bit of proof. First, the calculation in (12.3) 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 12.7. ■

12.10 Remark. *We will eventually prove that the existence of a left or right inverse alone is enough to guarantee the invertibility of a matrix! That is, if E, $F \in \mathbb{R}^{m \times m}$ with $EF = I_m$, then both E and F are invertible. We will need some more technology to do that, however.*

12.11 Problem (!). We probably expect that undoing the undoing of an action does that action. Totally makes sense, right? More precisely, if $E \in \mathbb{R}^{m \times m}$ is invertible, we should expect that E^{-1} is also invertible and $(E^{-1})^{-1} = E$. (That's how exponents work, right?) Prove this by showing that E satisfies the definition of inverse for E^{-1}. *What things do defines what things are.*

12.12 Problem ($\star$). 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 you do B to a vector $\mathbf{v}$ by multiplying $B\mathbf{v}$, and then you do A by multiplying $A(B\mathbf{v}) = (AB)\mathbf{v}$. To undo AB, you 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:

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_mB = B^{-1}B = I_m.$$

12.13 Problem (!). Check that $(AB)(B^{-1}A^{-1}) = I_m$ as well.

Here is the formal result.

12.14 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 12.1 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. I 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: $A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_m\mathbf{b} = \mathbf{b}$.

12.15 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.

12.16 Problem (!). Hugely important: convince yourself of the following for $A \in \mathbb{R}^{m\times m}$.

(i) If there is $\mathbf{z} \in \mathbb{R}^m$ with $\mathbf{z} \neq \mathbf{0}_m$ and $A\mathbf{z} = \mathbf{0}_m$, then A is not invertible.

(ii) If A is invertible, then $\mathbf{C}(A) = \mathbb{R}^m$.

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

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

Day 13: Wednesday, February 5.

Using the solution formula $\mathbf{x} = A^{-1}\mathbf{b}$ from Theorem 12.15 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:

$$\texttt{https}: //\texttt{math.mit.edu}/\ \texttt{gs/linearalgebra/ila5/linearalgebra5\_11} - \texttt{1.pdf}.$$

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.

Let's 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'll start with the nicest case: upper-triangular. I claim that 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.

13.1 Example. Let

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

We met this matrix in Example 10.8. I want to 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

$$D \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3/2 \end{bmatrix}.$$

I think you know what to do by now: expand the vector on the right as a linear combination weighted by v_1, v_2, and v_3, and you'll see that D should be

$$D_{33} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}.$$

I'm calling 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'll 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 $U^{-1} = E$.

13.2 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}.

The arithmetic in Example 13.1 is called **Gauss–Jordan elimination**. I'll state how this works in the abstract.

13.3 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 $EU = I_m$.*

Proof. This should feel basically the same as the proof of Theorem 12.1. 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 so that it's 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. ■

13.4 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 doesn't even give any problems asking you to do it in this edition—that's how deprecated the method is. Far better to *understand* A^{-1} than have a formula for it.

13.5 Problem (!). Explain why the matrix A from Example 10.8 is invertible. What is A^{-1}? (Don't actually compute it—no one really cares—but express A^{-1} as the product of the inverses of a bunch of elimination, scaling, and/or permutation matrices.)

What we really care about is not a formula for matrix inverses but the existence and behavior of inverses. We've seen a bunch of behaviors already: how the inverse of a product works, how the inverse is itself invertible, and, admittedly, the special formulas for inverses of elimination and permutation matrices.

13.6 Theorem. *Let $U \in \mathbb{R}^{m\times m}$ be upper-triangular. Then U is invertible if and only if all of its diagonal entries are nonzero.*

Proof. ($\Longleftarrow$) This is easier, so I'll do it first. It's just Gauss–Jordan elimination: since the diagonal entries of U are nonzero, we can find $E \in \mathbb{R}^{m \times m}$ invertible such that $EU = I_m$, thus $U = E^{-1}$ is invertible.

($\Longrightarrow$) We are going to use contradiction. What if U is invertible and a diagonal entry is zero? Something has to go wrong, and I am going to spoil the surprise for you: we are going to find a nonzero vector $\mathbf{x} \in \mathbb{R}^m$ such that $U\mathbf{x} = \mathbf{0}_m$. This will contradict Theorem 12.15, which says that since U is invertible, the only solution is $\mathbf{x} = \mathbf{0}_m$.

I want to consider two possible structures of U: one where the first diagonal entry is zero and one where it isn't, but a zero diagonal entry occurs further down along the diagonal. Here is the first when $m = 4$:

$$U = \begin{bmatrix} 0 & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix}.$$

More generally, 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). Recall now how we "extract" columns from a matrix: multiply by the standard basis vectors. So $U\mathbf{e}_1$ is the first column of U, where $\mathbf{e}_1$ is 1 in row 1 and 0 everywhere else. That is, $U\mathbf{e}_1 = \mathbf{0}_m$ and $\mathbf{e}_1 \neq \mathbf{0}_m$. That's the contradiction.

Next case: a zero entry further down on the diagonal. That is, $u_{jj} = 0$ for some $j \geq 2$ but $u_{ii} \neq 0$ for $1 \leq i \leq j-1$. Here is one such possibility when $m = 4$:

$$U = \begin{bmatrix} \smiley & * & \frownie & * \\ 0 & \smiley & \frownie & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \end{bmatrix}.$$

By $\smiley$ I mean nonzero entries, and in the notation above $j = 3$. Now look at the upper-triangular matrix

$$\widehat{U} := \begin{bmatrix} \smiley & * \\ 0 & \smiley \end{bmatrix}.$$

This has nonzero diagonal entries and so we can find x_1, $x_2 \in \mathbb{R}$ such that

$$\widehat{U} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frownie \\ \frownie \end{bmatrix}.$$

But then

$$x_1 \begin{bmatrix} \smiley \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} * \\ \smiley \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frownie \\ \frownie \\ 0 \\ 0 \end{bmatrix},$$

and from that

$$U\mathbf{x} = \mathbf{0}_4, \qquad \mathbf{x} := \begin{bmatrix} x_1 \\ x_2 \\ -1 \\ 0 \end{bmatrix}.$$

So what is the problem? We have found $\mathbf{x} \in \mathbb{R}^4$ such that $\mathbf{x} \neq \mathbf{0}_4$ but $U\mathbf{x} = \mathbf{0}_4$. This contradicts Theorem 12.15; since U is invertible, the only $\mathbf{x}$ that should work there is $\mathbf{x} = \mathbf{0}_4$.

Here is the generalization of this, which I didn't do in class, and for which you are not responsible. 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

$$U = \left[\begin{array}{cc|c} \widehat{U} & \widehat{\mathbf{u}}_j & \widetilde{U} \\ 0 & \mathbf{0}_{m-j} & \end{array}\right].$$

Here $\widehat{U}$ is a $(j-1)\times(j-1)$ upper-triangular matrix with nonzero diagonal entries, $\widehat{\mathbf{u}}_j \in \mathbb{R}^{j-1}$, and $\widetilde{U}$ contains the remaining columns of U. I am irritatingly using 0 to mean a matrix whose entries are all 0.

Since the diagonal entries of $\widehat{U}$ are nonzero, we can find $\widehat{\mathbf{x}} \in \mathbb{R}^{j-1}$ such that $\widehat{U}\widehat{\mathbf{x}} = \widehat{\mathbf{u}}_j$. Then

$$U\mathbf{x} = \mathbf{0}_m, \qquad \mathbf{x} := \begin{bmatrix} \widehat{\mathbf{x}} \\ -1 \\ \mathbf{0}_{m-(j+1)} \end{bmatrix},$$

and so the equation $U\mathbf{x} = \mathbf{0}_m$ has a nonzero solution. Thus U cannot be invertible. ■

13.7 Problem (+). What goes wrong if a diagonal entry of $U \in \mathbb{R}^{m\times m}$ is zero? We saw this in Example 11.5, but it's worth revisiting in the abstract. There are two possibilities: either the last diagonal entry is 0 or an entry further up the diagonal is 0, but the entries below it are nonzero.

(i) In the first case, the bottom row of U is 0. What does that say about the $\mathbf{c} \in \mathbb{R}^{m\times m}$ for which we can solve $U\mathbf{x} = \mathbf{c}$? And what does that say in the context of Theorem 12.15?

(ii) In the second case, U might have a structure like the following:

$$U = \begin{bmatrix} * & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & ☺ & * \\ 0 & 0 & 0 & ☺ \end{bmatrix}.$$

As before ☺ $\neq 0$. Explain why for this particular U, there is an invertible $\widetilde{E} \in \mathbb{R}^{4\times 4}$ such that $\widetilde{E}U$ has a row of all 0 entries. [Hint: *eliminate upward.*] What does this say about the $\mathbf{c} \in \mathbb{R}^4$ for which we can have $U\mathbf{x} = \mathbf{c}$?

Day 14: Friday, February 7.

The tool that we used in the second part of the proof of Theorem 13.6 is incredibly nice, and it's going to resolve some of our long-standing conjectures. First we isolate that part of the proof for future reference.

14.1 Corollary. *Let $U \in \mathbb{R}^{m \times m}$ be upper-triangular. Suppose that a diagonal entry of U is nonzero. Then there exists a nonzero vector $\mathbf{x} \in \mathbb{R}^m$ such that $U\mathbf{x} = \mathbf{0}_m$.*

Here's a blast from the past. We've said that the columns of a matrix $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$ are independent if $\mathbf{a}_1 \neq \mathbf{0}_m$ and $\mathbf{a}_j \notin \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1})$. Here A does not have to be square. This is just a more precise way of saying that no column is a linear combination of the others, and so independence is a precise way of controlling redundancy in a matrix's data. It turns out that independence is intimately connected with the HOMOGENEOUS problem $A\mathbf{x} = \mathbf{0}_m$.

14.2 Theorem. *The columns of a matrix $A \in \mathbb{R}^{m \times n}$ are independent if and only if the only solution to $A\mathbf{x} = \mathbf{0}_m$ is $\mathbf{x} = \mathbf{0}_n$.*

Proof. It's actually easier to prove

$$\text{dependent columns} \iff A\mathbf{x} = \mathbf{0}_m \text{ has a nonzero solution,}$$

so we'll do that. (Being negative is much more fun than being positive.)

If the columns are dependent, then one column is a linear combination of the others. Let $\mathbf{x}$ be the vector consisting of the weights from that combination with -1 for the "bad vector." That's the proof in words. In symbols, this is basically the proof of Theorem 7.1. For simplicity, if $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & (c_2\mathbf{a}_2 + c_4\mathbf{a}_4) & \mathbf{a}_4 \end{bmatrix} \in \mathbb{R}^{m \times 4}$, then

$$A \begin{bmatrix} 0 \\ c_2 \\ -1 \\ c_4 \end{bmatrix} = c_2\mathbf{a}_2 - (c_2\mathbf{a}_2 + c_4\mathbf{a}_4) + c_4\mathbf{a}_4 = \mathbf{0}_m.$$

Now suppose that $A\mathbf{x} = \mathbf{0}_m$ has a nonzero solution. So, at least one entry of $\mathbf{x}$ is nonzero. Pick any such nonzero entry, divide $A\mathbf{x} = \mathbf{0}_m$ by it, and rewrite the corresponding column of A as a linear combination of the rest. For example, if $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix} \in \mathbb{R}^{m \times 4}$ with $A\mathbf{x} = \mathbf{0}_m$, $\mathbf{x} = (x_1, x_2, x_3, x_4)$, and $x_3 \neq 0$, then

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

so

$$\frac{x_1}{x_3}\mathbf{a}_1 + \frac{x_2}{x_3}\mathbf{a}_2 + \mathbf{a}_3 + \frac{x_4}{x_3}\mathbf{a}_4 = \mathbf{0}_m.$$

and thus

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

is a linear combination of the other columns. ■

We could have done this back when we were talking about independence and dependence. Maybe we should have?

Content from Strang's *ILA* 6E. Look at "Independent columns" toward the bottom of p. 30.

14.3 Problem ($\star$). We are talking a lot right now about square systems: $A\mathbf{x} = \mathbf{b}$ with $A \in \mathbb{R}^{m \times m}$. Number of equations = number of variables. Independence and dependence make sense for all systems, not just square. Prove that if the columns of $A \in \mathbb{R}^{m \times n}$ are independent, then if a solution to $A\mathbf{x} = \mathbf{b}$ exists (no guarantee that it does!), it is unique. [Hint: *reread the proof of Theorem 10.1 and then the statement of Problem 10.2.*]

Remarkably, for a square system, independent columns and invertibility are the same thing.

14.4 Theorem. *A matrix $A \in \mathbb{R}^{m \times m}$ is invertible if and only if its columns are independent.*

Proof. ($\Longrightarrow$) By Theorem 14.2, to prove that the columns are independent, we should assume that $A\mathbf{x} = \mathbf{0}_m$ for some $\mathbf{x} \in \mathbb{R}^m$ and show $\mathbf{x} = \mathbf{0}_m$. This is almost automatic by Theorem 12.15, as $\mathbf{x} = A^{-1}\mathbf{0}_m = \mathbf{0}_m$.

($\Longleftarrow$) Assume that the columns are independent, and watch Gaussian elimination rise and shine! Whether or not the columns of A are independent, we can find an invertible $E \in \mathbb{R}^{m \times m}$ such that $EA = U$ with U upper-triangular. Then $A = E^{-1}U$. If U is also invertible, Theorem 12.14 guarantees that A is invertible. Theorem 13.6 says that U will be invertible if its diagonal entries are nonzero.

So are they? What goes wrong if U has a nonzero diagonal entry? Corollary 14.1 goes wrong: there is $\mathbf{x} \neq \mathbf{0}_m$ such that $U\mathbf{x} = \mathbf{0}_m$. But then $A\mathbf{x} = E^{-1}U\mathbf{x} = E^{-1}\mathbf{0}_m = \mathbf{0}_m$. This contradicts the independence of the columns of A. That's wrong. ■

Let us celebrate this result with a proof of Conjecture 6.14 which conjectured that if $A \in \mathbb{R}^{m \times m}$ has independent columns, then $\mathbf{C}(A) = \mathbb{R}^m$. Theorem 14.4 says that independence implies invertibility. And if A is invertible, then $\mathbf{C}(A) = \mathbb{R}^m$ because we can always solve $A\mathbf{x} = \mathbf{b}$. (We get *unique* solvability from independence to boot.) This is the victory of independence: having just enough data in your matrix, and no more, lets you solve $A\mathbf{x} = \mathbf{b}$, and solve it uniquely.

14.5 Problem (!). How does this lead to a quick proof of Conjecture 6.7? [Hint: *"$P \iff Q$" is the same as "not $P \iff$ not Q."*]

We have proved a lot of things recently, and all of them are saying basically the same thing. It can get hard to keep track of all of this, so here's a summary.

14.6 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) *There is an invertible matrix E such that $U := EA$ is upper-triangular and all of the diagonal entries of U are nonzero.*

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

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

(v) *For any $\mathbf{b} \in \mathbb{R}^m$, the problem $A\mathbf{x} = \mathbf{b}$ always has a unique solution.*

14.7 Problem (!). We have essentially proved Theorem 14.6, but the work is spread out among a number of results. Now is a good opportunity for you to review the logical connections among those results. Here is a cartoon summarizing those connections, which you can tease out in the steps below.

$$\begin{array}{ccccc} \text{(v)} & \Longrightarrow & \text{(i)} & \Longleftrightarrow & \text{(ii)} \\ & & \Updownarrow & & \\ & & \text{(iii)} & \Longleftrightarrow & \text{(iv)} \end{array}$$

(i) To prove that parts (i) and (ii) are equivalent, use Theorems 12.1, 12.14, and 13.6.

(ii) Review the proof of Theorem 14.4 to remind yourself why parts (i) and (iii) are equivalent.

(iii) Review the proof of Theorem 14.2 to remind yourself why parts (iii) and (iv) are equivalent.

(iv) Review the argument leading up to Theorem 12.15 to remind yourself why part (i) implies part (v). Then argue that (v) implies part (iv), and we know that implies (iii) and thus part (i).

14.8 Problem (+). Corollary 12.9 told us that $A \in \mathbb{R}^{m \times m}$ is invertible if and only if A has left and right inverses, i.e., matrices L, $R \in \mathbb{R}^{m \times m}$ such that $LA = AR = I_m$. The upshot of this condition is that we don't need to verify $L = R$; all we need is $LA = I_m$ and $AR = I_m$, and ostensibly the matrices L and R don't have to talk to each other. Here we show that we only need a left inverse or a right inverse to guarantee invertibility.

(i) Suppose that $LA = I_m$. Prove that the columns of A are linearly independent and thus A is invertible. [Hint: *if $A\mathbf{x} = \mathbf{0}_m$, apply L to both sides and figure out $\mathbf{x}$.*]

(ii) Suppose that $AR = I_m$. Let $EA = U$ be upper-triangular for some invertible matrix

E, so $UR = E$. If U is not invertible, then U has a zero diagonal entry. Problem 13.7 then implies that for some invertible $\widetilde{E}$, the product $\widetilde{E}U$ has a row whose entries are all 0. Use Problem 8.5 to deduce something about $\widetilde{E}UR$, and thus about $\widetilde{E}E$. Since $\widetilde{E}$ and E are invertible, what contradiction results?

(iii) We now can weaken part (v) of Theorem 14.6 slightly, but crucially. It turns out that $A \in \mathbb{R}^{m \times m}$ is invertible if and only if the problem $A\mathbf{x} = \mathbf{b}$ always has *a* solution for each $\mathbf{b} \in \mathbb{R}^m$. We do not need to require the solution to be unique. Certainly the existence of a solution is a consequence of invertibility (uniqueness, too). Now suppose that we can always solve $A\mathbf{x} = \mathbf{b}$. Explain how choosing $\mathbf{b}$ to be the standard basis vectors produces a matrix R such that $AR = I_m$.

14.9 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 12.1. This language is a little perilous, as we never proved that the matrix U from Theorem 12.1 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's 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.*

14.10 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: *what goes wrong if E_2A has some nonzero diagonal entries?*] 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.

Day 15: Monday, February 10.

We've 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 shouldn't 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.

15.1 Example. Let

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{bmatrix}.$$

We saw in Example 10.8 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 13.1 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. Go ahead and try it and see how opaque the work is. (I mean, we haven't really computed a matrix inverse other than Remark 13.4.) 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":

$$E^{-1} = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = E_{21}^{-1}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 4 & 3 & 1\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 2 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 4 & 3 & 1\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 2 & 1 & 0\\ 4 & 3 & 1\end{bmatrix} =: L.$$

Just look at that matrix L. It's **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? Let's solve $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} = (0, 1, 5)$. Ideally you did this in Problem 10.9. This problem 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 $\mathbf{b}$.

Let's 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

$$\begin{cases} 2x_1 & + \; x_2 & + \; x_3 & = 0\\ & x_2 & + \; x_3 & = 1\\ & & 2x_3 & = 2\end{cases}$$

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*. (It pains me to say this, but brute force isn't always the best force.)

All of this turns out to be more generally true.

15.2 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*

$$\begin{cases} L\mathbf{c} = \mathbf{b} \\ U\mathbf{x} = \mathbf{c}. \end{cases} \tag{15.1}$$

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 15.1. (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. Or check out the readings in Strang mentioned below.)

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 (15.1) are true.

Now suppose that both equations in (15.1) 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 (15.1) 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. ■

15.3 Problem ($\star$). 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 15.1 probably felt no more efficient than a routine back-substitution approach (which you did in Problem 10.9, right?) Maybe it felt more inefficient! That's a valid feeling. All of our examples in this class are effectively toy problems designed so that the on-the-fly arithmetic is easy.

But what if you need to solve $A\mathbf{x} = \mathbf{b}_j$ for many $\mathbf{b}_j$? If you have only a finite number of $\mathbf{b}_j$, maybe you 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 you 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, you get a new $\mathbf{b}_j$, but A stays the same. If you 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's 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 I think this is better covered in a numerical linear algebra course. But once you 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.

Day 16: Wednesday, February 12.

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

* I found this StackExchange post really helpful:

https://math.stackexchange.com/questions/266355/necessity-advantage-of-lu-decomposition-over-gaussian-elimination.

for it. With nonsquare systems—$A\mathbf{x} = \mathbf{b}$, $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m \neq n$—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.

16.1 Example. We consider our favorite problem $A\mathbf{x} = \mathbf{b}$ for the variety of A below.

(i) It's hard to get nicer than

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

because then the unique solution to $A\mathbf{x} = \mathbf{b}$ is always just $\mathbf{x} = \mathbf{b}$ (for $\mathbf{b} \in \mathbb{R}^2$).

(ii) It's easy to get less nice, though. Take

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

Then existence fails for $\mathbf{b}$ with $b_2 \neq 0$, while uniqueness also fails. Inspired by Problem 10.2 and the fact that $A\mathbf{e}_2 = \mathbf{0}_2$, where $\mathbf{e}_2 = (0, 1)$, we can check that

$$A\left(\begin{bmatrix} b_1 \\ 0 \end{bmatrix} + c\mathbf{e}_2\right) = \begin{bmatrix} b_1 \\ 0 \end{bmatrix}$$

for any b_1, $c \in \mathbb{R}$. Thus solutions, when they exist, are never unique. The dependence of the columns affects both existence and uniqueness here, per the invertible matrix theorem.

(iii) With

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

existence fails for those $\mathbf{b} \in \mathbb{R}^3$ with $b_3 \neq 0$. But solutions, when they exist are unique, because the only solution to $A\mathbf{x} = \mathbf{0}_3$ is $\mathbf{x} = \mathbf{0}_2$. We saw this in Problem 10.2, and it's closely related to the independence of the columns. We'll revisit this soon.

(iv) With

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

both existence and uniqueness fail, since we can't solve $A\mathbf{x} = (0, 1, 0)$, while when we can solve $A\mathbf{x} = (b_1, 0, 0)$ with $\mathbf{x} = (b_1, 0)$, we can also solve it with $\mathbf{x} = (b_1, c)$ for any $c \in \mathbb{R}$. And the columns are dependent.

(v) With

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

we have existence but not uniqueness: take $\mathbf{x} = (b_1, b_2, c)$ to solve $A\mathbf{x} = (b_1, b_2)$. Again, dependent columns.

(vi) Last, existence and uniqueness fail for

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

Once again, we can only solve $A\mathbf{x} = \mathbf{b}$ for $b_2 = 0$, and when we can, $\mathbf{x} = (b_1, c_2, c_3)$ is a solution for any c_2, $c_3 \in \mathbb{R}$. Dependent columns are the worst.

Here is what the previous example suggests as we move beyond square systems.

16.2 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.*

Content from Strang's *ILA* 6E. Now is a good time to reread p. 38.

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. Think about the column space. For $A \in \mathbb{R}^{m \times n}$, we have

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

Every vector in $\mathbf{C}(A)$ is a vector in $\mathbb{R}^m$. We know very well that controlling the column space is the key (well, a key) to existence of solutions to $A\mathbf{x} = \mathbf{b}$, for this equation is true if and only if $\mathbf{b} \in \mathbf{C}(A)$.

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). A *space* is more dynamic.

Specifically, the column space behaves well with respect to the fundamental objects of vector arithmetic. 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 $\mathbf{C}(A)$.

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 $\mathbf{C}(A)$.

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 don't just exist and contain things; they are *dynamic* with respect to vector operations. We'll 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.

Day 17: Friday, February 14.

We took Exam 1.

Day 18: Monday, February 17.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Null space of a matrix, subspace (N)

The column space governs *existence*: we can solve $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{b} \in \mathbf{C}(A)$. But the column space says nothing about *uniqueness*: having $\mathbf{b} \in \mathbf{C}(A)$ does not guarantee that there is *only* one $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$, but rather that there is *at least* one such $\mathbf{x}$. To understand uniqueness, we need to study a new set—more precisely, because of its dynamism, a unique space.

We have discussed the following several times, and you proved it in Problem 10.2.

18.1 Theorem. *Let* $A \in \mathbb{R}^{m \times n}$.

(i) Suppose that the only $\mathbf{z} \in \mathbb{R}^n$ such that $A\mathbf{z} = \mathbf{0}_m$ is $\mathbf{z} = \mathbf{0}_n$. Then for any $\mathbf{b} \in \mathbb{R}^m$, the problem $A\mathbf{x} = \mathbf{b}$ has at most one solution. (Maybe it has none.)

(ii) Suppose that for all $\mathbf{b} \in \mathbb{R}^m$, the problem $A\mathbf{x} = \mathbf{b}$ has at most one solution. (Maybe it has none.) Then the only solution to $A\mathbf{x} = \mathbf{0}_m$ is $\mathbf{x} = \mathbf{0}_n$.

You saw another version of this in Theorem 14.2, right? Rereading that theorem right now is probably a good idea.

Here is the point: we can understand uniqueness of solutions to the problem $A\mathbf{x} = \mathbf{b}$ for *any* $\mathbf{b}$ by studying the problem for the special case of $\mathbf{b} = \mathbf{0}_m$. This motivates the study of a new dynamic set related to A.

18.2 Definition. *Let $A \in \mathbb{R}^m$. The* NULL SPACE *of A is*

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

18.3 Example. **(i)** The null space of I_2 is just $\{\mathbf{0}_2\}$, for if $I_2\mathbf{v} = \mathbf{0}_2$, then since $I_2\mathbf{v} = \mathbf{v}$, we just have $\mathbf{v} = \mathbf{0}_2$.

(ii) Let

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

The system $A\mathbf{x} = \mathbf{0}_3$ (for $\mathbf{x} \in \mathbb{R}^2$) is just

$$\begin{cases} x_1 = 0 \\ x_2 = 0 \\ 0 = 0, \end{cases}$$

so $\mathbf{N}(A) = \{\mathbf{0}_2\}$ once again.

18.4 Problem (!). Prove that

$$\mathbf{N}(I_n) = \{\mathbf{0}_n\} \quad \text{and} \quad \mathbf{N}\left(\begin{bmatrix} I_n \\ 0 \end{bmatrix}\right) = \{\mathbf{0}_n\}.$$

In the second, block matrix, you should interpret the symbol 0 as representing one or more rows of zeros. [Hint: *convince yourself that*

$$\begin{bmatrix} A \\ B \end{bmatrix} \mathbf{x} = \begin{bmatrix} A\mathbf{x} \\ B\mathbf{x} \end{bmatrix}$$

whenever $A \in \mathbb{R}^{m_1 \times n}$, $B \in \mathbb{R}^{m_2 \times n}$, and $\mathbf{x} \in \mathbb{R}^n$.]

18.5 Example. Let

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

For $\mathbf{x} \in \mathbb{R}^4$, we have $A\mathbf{x} = \mathbf{0}_2$ if and only if

$$\begin{cases} x_1 \quad + 2x_3 + 3x_4 = 0 \\ \quad x_2 \quad\quad\quad + 4x_4 = 0 \end{cases}$$

This is not as nice as the square upper-triangular systems that we have previously studied. There's no equation with just one variable 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, comparatively speaking, than solving for x_3 or x_4. Gotta solve for something, anyway. 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

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_3 - 3x_4 \\ -4x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_3 \\ 0 \\ x_3 \\ 0 \end{bmatrix} + \begin{bmatrix} -3x_4 \\ -4x_4 \\ 0 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ -4 \\ 0 \\ 1 \end{bmatrix}.$$

Think about that for a moment. 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* $\mathbf{x}$ is in the null space of 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 $\mathbf{N}(A)$.)

Do you see the pattern here? Our original 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

$$\mathbf{N}(A) = \mathbf{C}\left(\begin{bmatrix} -F \\ I_2 \end{bmatrix}\right).$$

This can't be an accident.

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.

18.6 Problem ($\star$). Let $n > m$ and $F \in \mathbb{R}^{m \times (n-m)}$. Prove that

$$\mathbf{N}\left(\begin{bmatrix} I_m & F \end{bmatrix}\right) = \mathbf{C}\left(\begin{bmatrix} -F \\ I_m \end{bmatrix}\right).$$

[Hint: *write any* $\mathbf{x} \in \mathbb{R}^n$ *as* $\mathbf{x} = (\mathbf{x}_m, \mathbf{x}_{n-m})$ *with* $\mathbf{x}_m \in \mathbb{R}^m$ *and* $\mathbf{x}_{n-m} \in \mathbb{R}^{n-m}$*, and argue that*

$$\begin{bmatrix} A & B \end{bmatrix} \mathbf{x} = \begin{bmatrix} A\mathbf{x}_m & B\mathbf{x}_{n-m} \end{bmatrix}$$

when $A \in \mathbb{R}^{m \times m}$ *and* $B \in \mathbb{R}^{m \times (n-m)}$.]

18.7 Problem (!). Putting more zero rows into the matrix doesn't change the null space. We saw this already in Example 18.3. Adapt the work of Example 18.5 to express the null space of

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

as a column space.

18.8 Problem (+). Let $1 \leq r < n$, $F \in \mathbb{R}^{r \times (n-r)}$, and

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

Here the two occurrences of the symbol 0 are meant to represent matrices whose entries are all the number 0. (What are the dimensions of those matrices?) Prove that

$$\mathbf{N}(A) = \mathbf{C}\left(\begin{bmatrix} -F \\ I_r \end{bmatrix}\right).$$

We are doing examples and problems in these special forms for a reason, and I'll tell you what that reason is soon. Let's pause from concrete numbers and focus on the *dynamic* aspects of the null space. Spoiler: it's the same dynamism as the column space.

Let $A \in \mathbb{R}^{m \times n}$. 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 $\mathbf{N}(A)$.

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 $\mathbf{N}(A)$.

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.

Subsets of $\mathbb{R}^n$ (or $\mathbb{R}^m$, or whatever) 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.

18.9 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] $\mathbf{0}_p \in \mathcal{V}$.

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 $\mathbb{R}^p$.

18.10 Example. Let $A \in \mathbb{R}^{m \times n}$.

(i) $\mathbf{C}(A) = \{A\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^n\}$ is a subspace of $\mathbb{R}^m$. We proved this some time ago; the important thing here is that every vector in $\mathbf{C}(A)$ has the form $A\mathbf{v} \in \mathbb{R}^m$.

(ii) $\mathbf{N}(A) = \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \mathbf{0}_m\}$ is a subspace of $\mathbb{R}^n$. It should be obvious from the definition of $\mathbf{N}(A)$ that every vector in the null space is a vector in $\mathbb{R}^n$.

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.

18.11 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 $\mathcal{V}$.

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.

Day 19: Wednesday, February 19.

Vocabulary from today

You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

reduced row echelon form (RREF) (N), pivot column (of a matrix in RREF), free column (of a matrix in RREF)

We return to the problem of concretely describing null spaces with a new wrinkle.

19.1 Example. Let

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

We proceed as in Example 18.5: assume $A\mathbf{x} = \mathbf{0}_2$ and write this as the linear system

$$\begin{cases} x_1 + 2x_2 + 3x_4 = 0 \\ x_3 + 4x_4 = 0. \end{cases}$$

We solve for the variables with the simples coefficients of 1; these are now x_1 and x_3:

$$\begin{cases} x_1 = -2x_2 - 3x_4 \\ x_3 = -4x_4. \end{cases}$$

Vectorizing, we have

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_2 - 3x_4 \\ x_2 \\ -4x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_2 \\ x_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -3x_4 \\ 0 \\ -4x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

Thus

$$\mathbf{N}\left(\begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{bmatrix}\right) = \mathbf{C}\left(\begin{bmatrix} -2 & -3 \\ 1 & 0 \\ 0 & -4 \\ 0 & 1 \end{bmatrix}\right). \tag{19.1}$$

All of our previous examples and problems about finding null spaces had the identity matrix show up in a pretty obvious way. It looks like the 2×2 identity is jumbled here. How can we sort it out?

We've handled "jumbled" matrices before. Recall that a permutation matrix $P \in \mathbb{R}^{m \times m}$ is a matrix formed by reordering the columns of the $m \times m$ identity matrix. If $B \in \mathbb{R}^{m \times n}$, then PB reorders the rows of B per the ordering of the columns in P. With our A in this example, however, it's a matter of reordering *columns*. We'd be happier if the columns of the 2×2 identity matrix appeared first in A.

How can we make this happen? If multiplying on the left by a permutation matrix reorders rows, multiplying on the right reorders columns. Here $A \in \mathbb{R}^{2 \times 4}$, so if we multiply on the right by a permutation matrix P, we better have $P \in \mathbb{R}^{4 \times 4}$. What we want is

$$AP = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{bmatrix} P = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_3 & \mathbf{a}_2 & \mathbf{a}_4 \end{bmatrix},$$

and we know that if $P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix}$, then

$$AP = \begin{bmatrix} A\mathbf{p}_1 & A\mathbf{p}_2 & A\mathbf{p}_3 & A\mathbf{p}_4 \end{bmatrix}.$$

So, we better have

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_3 & \mathbf{a}_2 & \mathbf{a}_4 \end{bmatrix} = \begin{bmatrix} A\mathbf{p}_1 & A\mathbf{p}_2 & A\mathbf{p}_3 & A\mathbf{p}_4 \end{bmatrix}.$$

We know that the columns of P are going to be columns of I_4, and we know that $A\mathbf{e}_j = \mathbf{a}_j$, where $\mathbf{e}_j$ is the jth column of I_4, i.e., the jth standard basis vectors. All together, this says that we should take

$$P = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_3 & \mathbf{e}_2 & \mathbf{e}_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

to find

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

Inverting P, we have

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

How does this compare to what we already know from (19.1)? Example 18.5 taught us that

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

and now we have shown that

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

So where is the permutation on the right?

In the rows! Multiplying on the left by P interchanges rows 2 and 3, and we have

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

So here is the conclusion:

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

Content from Strang's *ILA* 6E. This was basically Example 2 on p. 94.

19.2 Problem (+). Let $m < n$ and $F \in \mathbb{R}^{m \times (n-m)}$. Let $P \in \mathbb{R}^{n \times n}$ be invertible. Prove that

$$\mathbf{N}\left(\begin{bmatrix} I_m & F \end{bmatrix} P\right) = \mathbf{C}\left(P^{-1}\begin{bmatrix} -F \\ I_m \end{bmatrix}\right).$$

[Hint: *you want to solve* $\begin{bmatrix} I_m & F \end{bmatrix} P\mathbf{x} = \mathbf{0}_m$. *Put* $\mathbf{y} = P\mathbf{x}$. *Now you want to solve* $\begin{bmatrix} I_m & F \end{bmatrix} \mathbf{y} = \mathbf{0}_m$. *You know how to do this from Problem 18.6. Along the way, check that the matrix product in the column space above is actually defined.*]

19.3 Problem (!). Adapt the work of Example 19.1 to express the null space of

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

as a column space.

19.4 Problem (+). Let $1 \le r < n$ and $F \in \mathbb{R}^{r \times (n-r)}$. Let $P \in \mathbb{R}^{n \times n}$ be invertible. Prove that

$$\mathbf{N}\left(\begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P\right) = \mathbf{C}\left(P^{-1}\begin{bmatrix} -F \\ I_m \end{bmatrix}\right).$$

As before, the symbols 0 denote matrices whose entries are all 0.

The conclusion from all of the recent problems and examples should be that the null space is "easy" to describe when the matrix under consideration has one of the following special forms:

$$I_n, \qquad \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} I_m & F \end{bmatrix}, \qquad \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} I_m & F \end{bmatrix} P, \qquad \text{or} \qquad \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P. \tag{19.2}$$

Above, P is in practice a permutation matrix (and actually a rather specific kind of permutation matrix), although the only thing that we really required in the null space calculations was the invertibility of P.

Certainly not every matrix has one of these six forms—common to all of these forms is the appearance of an identity matrix within the columns of overall matrix. But every matrix can be *reduced* to one of these forms by the elementary row operations that we already know and love—by Gauss–Jordan elimination. Once again, the major technique in computational linear algebra is putting zeros in matrices.

19.5 Example. Let

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

Previously we only performed elementary row operations on square matrices, but they certainly work on nonsquare matrices, too. We compute

$$\begin{bmatrix}1&2&1&7\\2&4&2&14\\0&0&2&8\end{bmatrix} \xrightarrow[E_{21}]{\mathsf{R2} \mapsto \mathsf{R2}-2\times\mathsf{R1}} \begin{bmatrix}1&2&1&7\\0&0&0&0\\0&0&2&8\end{bmatrix}, \qquad E_{21} := \begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}$$

$$\xrightarrow[P_{23}]{\mathsf{R2} \mapsto \mathsf{R3},\ \mathsf{R3}\mapsto\mathsf{R2}} \begin{bmatrix}1&2&1&7\\0&0&2&8\\0&0&0&0\end{bmatrix}, \qquad P_{23} := \begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}$$

$$\xrightarrow[D_{22}]{\mathsf{R2} \mapsto (1/2)\times\mathsf{R2}} \begin{bmatrix}1&2&1&7\\0&0&1&4\\0&0&0&0\end{bmatrix}, \qquad D_{22} := \begin{bmatrix}1&0&0\\0&1/2&0\\0&0&1\end{bmatrix}$$

$$\xrightarrow[E_{12}]{\mathsf{R1} \mapsto \mathsf{R1}-\mathsf{R2}} \begin{bmatrix}1&2&0&3\\0&0&1&4\\0&0&0&0\end{bmatrix}, \qquad E_{12} := \begin{bmatrix}1&-1&0\\0&1&0\\0&0&1\end{bmatrix}.$$

That is,

$$EA = \begin{bmatrix}1&2&0&3\\0&0&1&4\\0&0&0&0\end{bmatrix}, \qquad E := E_{12}D_{22}P_{23}E_{21}.$$

Now put

$$P = \begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix}2&3\\0&4\end{bmatrix}.$$

Then

$$EA = \begin{bmatrix}I_2&F\\0&0\end{bmatrix} P =: R_0$$

where the symbol 0 denotes the matrix $\begin{bmatrix}0&0\end{bmatrix}$.

We saw in Problem 19.3 that

$$\mathbf{N}(R_0) = \mathbf{N}\left(\begin{bmatrix}I_2&F\\0&0\end{bmatrix} P\right) = \mathbf{C}\left(P^{-1}\begin{bmatrix}-F\\I_2\end{bmatrix}\right) = \mathbf{C}\left(\begin{bmatrix}-2&-3\\1&0\\0&-4\\0&1\end{bmatrix}\right).$$

This is helpful here because we have $EA = R_0$ with E invertible. If $\mathbf{v} \in \mathbf{N}(A)$, then $A\mathbf{v} = \mathbf{0}_2$, so $E(A\mathbf{v}) = \mathbf{0}_m$. And then

$$\mathbf{0}_2 = (EA)\mathbf{v} = R_0\mathbf{v},$$

so $\mathbf{v} \in \mathbf{N}(R_0)$. Conversely, if $\mathbf{v} \in \mathbf{N}(R_0)$, then $R_0\mathbf{v} = \mathbf{0}_2$, so

$$A\mathbf{v} = E^{-1}R_0\mathbf{v} = E^{-1}\mathbf{0}_2 = \mathbf{0}_2.$$

Thus $\mathbf{N}(A) = \mathbf{N}(R_0)$. This is a nice auxiliary fact: multiplying on the left by an invertible matrix does not change the kernel!

All together, we conclude

$$\mathbf{N}\left(\begin{bmatrix} 1 & 2 & 1 & 7 \\ 2 & 4 & 2 & 14 \\ 0 & 0 & 2 & 8 \end{bmatrix}\right) = \mathbf{C}\left(\begin{bmatrix} -2 & -3 \\ 0 & -4 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

The example above is prototypical: Gauss–Jordan elimination "reduces" any $A \in \mathbb{R}^{m \times n}$ to a matrix of the following structure.

19.6 Definition. *A matrix $R \in \mathbb{R}^{m \times n}$ is in* **REDUCED ROW ECHELON FORM (RREF)** *if it has the following four properties.*

Row Property 1. *Any nonzero row of R is below any row with 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 $\mathbb{R}^m$. Such a column is called a* **PIVOT COLUMN**. *A column that is not a pivot column is called a* **FREE COLUMN**.

Column Property 2. *If $\mathbf{e}_i$ and $\mathbf{e}_j$ are columns of R with $i < j$, then the first appearance of $\mathbf{e}_i$ must occur before any appearance of $\mathbf{e}_j$.*

19.7 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.

19.8 Problem (!). Explain why

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

is in RREF and comment on the role of the adjective "first" in Column Property 2 of Definition 19.6.

Here is the fruit of Gauss–Jordan elimination.

19.9 Theorem. *Let $A \in \mathbb{R}^{m \times n}$. There exists an invertible matrix $E \in \mathbb{R}^{m \times m}$ such that EA is in RREF with one of the following forms:*

(i) I_n, *in which case A is square and invertible;*

(ii) $\begin{bmatrix} I_n \\ 0 \end{bmatrix}$, *in which case $n < m$ (more rows than columns);*

(iii) $\begin{bmatrix} I_m & F \end{bmatrix}$, *in which case $m < n$ (more columns than rows) and $F \in \mathbb{R}^{m \times (n-m)}$;*

(iv) $\begin{bmatrix} I_m & F \end{bmatrix} P$, *with the same conditions as in form (iii) and P a permutation matrix;*

(v) $\begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix}$, *in which case $1 \le r \le \min\{m, n\}$ and $F \in \mathbb{R}^{r \times (n-r)}$;*

(vi) $\begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P$, *with the same conditions as in form (v) and P a permutation matrix.*

This form is unique in the sense that if $\widetilde{E} \in \mathbb{R}^{m \times m}$ is invertible with $\widetilde{E}A$ in RREF, then $EA = \widetilde{E}A$. We write $EA = \mathsf{rref}(A)$ and call $\mathsf{rref}(A)$ the **RREF of** *A.*

We are not going to prove this theorem in detail. Existence, again, is just Gauss–Jordan elimination. Uniqueness is surprisingly more annoying.

19.10 Problem ($\star$). Example 19.5 constructs $E \in \mathbb{R}^{3 \times 3}$ such that

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

(i) By revisiting the elementary row operations in that example, explain why E in Theorem 19.9 might not be unique. [Hint: *could P_{23} or D_{33} have appeared earlier or later?*]

(ii) With E from Example 19.5, find a permutation matrix $\widetilde{P}$ such that

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

Contrast this with the result of Example 19.5 and explain how this shows that P and F from Theorem 19.9 may not be unique.

(iii) Explain why there cannot exist a matrix $A \in \mathbb{R}^{3\times 4}$ such that

$$\mathsf{rref}(A)=\begin{bmatrix}1&0&3&2\\0&1&4&0\\0&0&0&0\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}$$

Conclude that not every permutation matrix P can appear in the forms of Theorem 19.9.

(iv) Give examples of two matrices $A \neq B$ that have the same RREF. [Hint: *look no further than the first form in Theorem 19.9.*]

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. Page 96 gives the algorithm column by column. Read p. 142 up to but not including the "Factorization" box.

19.11 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.

19.12 Problem (+). For each of the six RREF forms in Theorem 19.9, find a matrix whose RREF has that form. Construct your matrix so that it has at least two rows and at least two columns and that all of its entries are nonzero. For the forms with a permutation matrix, ensure that a permutation matrix is actually needed in your form (don't just let P be the identity, which is a permutation matrix, but a boring one).