Day 1: Monday, January 9. We studied the initial value problem

$$\begin{cases} f''(t) - 4f(t) = 0 \\ f(0) = 1 \\ f'(0) = -1 \end{cases}$$

and looked for a solution of the form $f(t) = c_1e^{2t} + c_2e^{-2t}$, since $f_1(t) = e^{2t}$ and $f_2(t) = e^{-2t}$ each solve the differential equation but don't meet the initial conditions. We chose this form because the derivative respects linearity:

$$f'(t) = c_1f_1'(t) + c_2f_2'(t).$$

To meet the initial conditions, we found that the numbers c_1 and c_2 needed to solve

$$\begin{cases} c_1 + c_2 = 1 \\ 2c_1 - 2c_2 = -1 \end{cases} \tag{1}$$

We solved this system (there are many approaches) to find $c_1 = 1/4$ and $c_2 = 3/4$. We then rewrote the system with x_1 and x_2 in place of c_1 and c_2 and turned that into an equality of column vectors:

$$\begin{bmatrix} c_1 + c_2 \\ 2c_1 - 2c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

This is the content of the first paragraph on p. 2 of the textbook.

Day 2: Wednesday, January 11. Almost everything we did today boiled down to understanding and manipulating the three equalities

$$\begin{bmatrix} x_1 + x_2 \\ 2x_1 - 2x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 2x_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ -2x_2 \end{bmatrix} = x_1\begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_2\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

The first equality is vector addition, the second is scalar multiplication, and the third is matrix-vector multiplication. The expression after the second equals sign is a linear combination. See pp. 2 and 32 for vector addition and scalar multiplication, pp. 3 and 33 for some examples of linear combinations, and p. 33 for matrix-vector multiplication. See p. 3 for a geometric interpretation of vector addition that works in two dimensions, but not really in higher dimensions. We ended by talking very briefly about the dot product, defined on p. 11, and how matrix-vector multiplication can also be done with dot products (p. 34). We will do this in much more detail next time. Worked Example 1.1 C compresses everything from today into about a third of a page of work. Read it!

Here is a claim for you to ponder. If b_1 and b_2 are real numbers, and if we define

$$\mathbf{v} := \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad \mathbf{w} := \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{b} := \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

then the vector equation

$$x_1\mathbf{v} + x_2\mathbf{w} = \mathbf{b}$$

has only one solution:

$$x_1 = \frac{b_1}{2} + \frac{b_2}{4} \quad \text{and} \quad x_2 = \frac{b_1}{2} - \frac{b_2}{4}.$$

Use algebraic techniques like we did on Monday to figure this out. Check that

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b_1 \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} + b_2 \begin{bmatrix} 1/2 \\ -1/4 \end{bmatrix}.$$

Everything that we did today allowed us to reinterpret Monday's problem (1) in new language, but we learned nothing new today about actually solving that problem! That will come later. First we need to be comfortable with the right language for problem solving.

Also, everything that we did today was for vectors with 2 entries and square matrices with 2 columns and 2 rows. This can be generalized tremendously! Section 1.3 does arithmetic with 3 entries, 3 columns, 3 rows, and so does Section 2.1. I want to think about a systematic way of solving problems like (1) first. Then we'll add more equations and unknowns and appreciate matrix-vector language more.

Day 3: Friday, January 13. We defined the dot product

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

calculated

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 11,$$

and then interpreted matrix-vector multiplication in terms of dot products. (Note that $\cdot$ can mean dot product of vectors but also ordinary multiplication of numbers, e.g., $2 \cdot 3 = 6$.) We have things like

$$\begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ 2x_1 - 2x_2 \end{bmatrix} \quad \text{with} \quad x_1 + x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad 2x_1 - 2x_2 = \begin{bmatrix} 2 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

It looks like matrix-vector multiplication can be expressed as the *dot product of the rows of the matrix, treated as columns, with the given vector*. Huh? We need a new operation: the transpose. This flips column vectors to rows and rows to columns:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \Longrightarrow \mathbf{v}^{\mathsf{T}} := \begin{bmatrix} v_1 & v_2 \end{bmatrix} \quad \text{and} \quad \mathbf{w} := \begin{bmatrix} w_1 & w_2 \end{bmatrix} \Longrightarrow \mathbf{w}^{\mathsf{T}} := \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

(I like writing transpose with the sans-serif T; some people just uppercase T.)

So here are two slogans. (1) The matrix-vector product $A\mathbf{x}$ is a linear combination of the columns of A with "weights" from the entries of $\mathbf{x}$. (2) The matrix-vector product $A\mathbf{x}$ is the vector whose entries are the dot products of the transposes of the rows of A with the vector $\mathbf{x}$. Yikes! I think the first slogan is better than the second. (I think Dr. Strang thinks so, too.) Do you understand all the words in the slogans? Can you calculate

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

using both slogans? The answer should be $\begin{bmatrix}17 & 39\end{bmatrix}^\mathsf{T}$ either way. (Transposes do save space while typing.)

Then we talked about dimensionality. Things like

$$\begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}3\\4\\5\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}1\\2\end{bmatrix} \cdot \begin{bmatrix}2\\1\\0\end{bmatrix}$$

are undefined because the vectors don't have the same number of entries. (A brief, lively discussion ensued and suggested ways of "embedding" a smaller vector inside a larger vector to make this work out.) In general, we denote the set of all column vectors with n entries by $\mathbb{R}^n$.

One particular vector shows up in all versions of $\mathbb{R}^n$: the vector whose entries are all 0. This is, of course, the zero vector. We write this as $\mathbf{0}$. (Note $0 \neq \mathbf{0}$.) Exercise for you: if A is a 2×2 matrix, what is $A\mathbf{0}$? More exercises for you (which we did not discuss in class but which come to mind now): let

$$I = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} \quad \text{and} \quad J = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}.$$

For $\mathbf{x} = \begin{bmatrix}x_1 & x_2\end{bmatrix}^\mathsf{T}$, calculate $I\mathbf{x}$ and $J\mathbf{x}$ and explain why we should call I an "identity" matrix and J a "flip" matrix.

We have significantly digressed from our original goal of solving a linear system of equations into a deep consideration of matrix vector notation, language, and arithmetic. Let's go back to that original problem, which was

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

Here is what we did. Multiply both sides of the first equation by 2 and subtract that from the second equation:

$$x_1 + x_2 = 1 \longrightarrow 2x_1 + 2x_2 = 2 \longrightarrow 2x_1 - 2x_2 - (2x_1 + 2x_2) = -1 - 2 \longrightarrow -4x_2 = -3.$$

Provided that $x_1 + x_2 = 1$, the equations $2x_1 - 2x_2 = -1$ and $-4x_2 = -3$ are the same. Here's why. Let $a = 2x_1 + 2x_2$. Then it is also the case that $a = 2$. And so

$$2x_1 - 2x_2 = -1 \iff 2x_1 - 2x_2 - a = -1 - a \iff 2x_1 - 2x_2 - (2x_1 + 2x_2) = -1 - 2.$$

This means that our original problem is *equivalent* to

$$\begin{cases} x_1 + x_2 = 1 \\ - 4x_2 = -3. \end{cases}$$

This problem is much nicer: solve for $x_2 = 3/4$ and then $x_1 + 3/4 = 1$, thus $x_1 = 1/4$. This is exactly what we saw on Monday.

Let me compress this work euphemistically as follows:

$$\begin{cases} x_1 + x_2 = 1 & (1) \\ 2x_1 - 2x_2 = -1 & (2) \end{cases} \iff \begin{cases} x_1 + x_2 = 1 & (1) \\ - 4x_2 = -3 & (2) - 2 \cdot (1) \end{cases}$$

Here's our job for next week and beyond: how do we generalize this procedure to any system of two equations in two unknowns? To any system of n equations in n unknowns?? To any system of m equations in n unknowns, maybe with $m \neq n$??? Specifically, how did we get the idea to "multiply by 2" and then "subtract that multiple from the second equation"?

The really new stuff today was our systematic approach to the original problem. See pp. 46–47 for the exact same arithmetic on a completely different system.

Day 4: Wednesday, January 18. We started out with one more 2×2 problem:

$$\begin{cases} 4x_1 + x_2 = 10 & (1) \\ 6x_1 + 2x_2 = 5 & (2). \end{cases}$$

The fundamental idea was to subtract a multiple of the first equation from the second equation to remove the $6x_1$ term in the second equation. The right multiplier was $6/4$. The result is the new system

$$\begin{cases} 4x_1 + x_2 = 10 & (1) \\ \dfrac{1}{2}x_2 = -10 & (2) - \dfrac{6}{4}(1) \end{cases}$$

This is upper-triangular: see p. 46. We called the coefficient 4 the pivot in the first equation; the pivot is the first nonzero coefficient in an equation that you use to eliminate the corresponding variable in the subsequent equations. See p. 47. We also called $1/2$ the pivot in the second equation of the new system; there is no more elimination to be done, but $1/2$ is still the first nonzero coefficient in that equation.

This is a major goal of linear algebra. Start with a problem $A\mathbf{x} = \mathbf{b}$. Convert this problem to an upper-triangular problem $U\mathbf{x} = \mathbf{c}$. The matrix $\mathbf{c}$ probably won't be $\mathbf{b}$! Then back-solve to find $\mathbf{x}$.

Then we worked on the larger problem

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

The strategy was the same as above just with more numbers: subtract multiples of the first equation from the second and third equations to eliminate the x_1-terms. The pivot in the first equation was 1. This brought us to

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 & (1) \\ - 7x_2 - 6x_3 = -10 & (2') = (2) - 3 \cdot (1) \\ - x_2 - x_3 = -2 & (3') = (3) - 2 \cdot (1). \end{cases}$$

The goal was an upper-triangular structure, so we wanted to eliminate x_2 from the third equation. The pivot in the second equation was -7. (A thoughtful discussion ensued about multiplying (2') and (3') by -1 and then exchanging them so that the pivot would be 1; this is mathematically valid but too much work for us. We want to restrict ourselves to solving this system to the "subtract a multiple of one row from another row" technique—these are simply the rules of the game right now.) After subtracting a multiple of the second equation from the third, we ended up with

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 & (1) \\ - 7x_2 - 6x_3 = -10 & (2') \\ - \dfrac{1}{7}x_3 = -\dfrac{4}{7} & (3'') = (3') - [(-1)/(-7)] \cdot (2'). \end{cases}$$

This is an upper-triangular problem, and the pivot in the third equation is $-1/7$.

See pp. 49–50 for another example very much like this one.

Day 5: Friday, January 20. We discussed the general method of elimination in a more abstract way than we have previously done. See the steps at the top of p. 46 and make sure you are comfortable with applying them to all of the systems that we have seen so far. See p. 51 for another discussion of the general elimination method.

We then talked about two "failures" of elimination. In general, a linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution or it doesn't. If it doesn't, then either it has no solution at all or infinitely many solutions. See the examples on pp. 49–50.

There is a third "temporary" failure in which the current row/equation does not have a pivot in the right place but another row does. In that case we interchange rows. See Example 3 on p. 49. I think it takes a large system to illustrate reasonably the need for row-swapping; you might argue in Example 3 that the system starts out as "lower-triangular" and you could just back-solve from the top down. Here is a problem in which you definitely want to flip some rows (what are they?):

$$\begin{cases} x_1 + 2x_2 + 3x_3 + 4x_4 = 5 \\ 6x_3 + 7x_4 = 8 \\ 9x_2 + 10x_3 + 11x_4 = 12 \\ 13x_2 + 14x_3 + 15x_4 = 16. \end{cases}$$

It will be worthwhile to read through and fill in the details of each of the worked examples on pp. 52–53.

Day 6: Monday, January 23. *Stop. Go no further.* Reread p. 51 *right now* to make sure you understand the goal and utility of elimination and reduction to an upper-triangular system. Then read p. 61. Does every single word make sense?

You may proceed. The goal is now to represent the act of elimination via the action of matrices. Previously matrices (and vectors) just encoded the data of a system in a convenient, evocative way. That was the *static* view of matrices (and vectors). Now we are going to take the *dynamic* view: matrices do something. Specifically, *matrices multiply vectors.* This philosophy is p. 58. Read that page to get a sense of our ultimate goals for the next few days.

I diverged from the text today and worked strictly with 2×2 problems. This is morally the same as pp. 60–62, which gives the 3×3 set-up that we'll do next time. Here is a brief sketch.

Our very first problem in the course was

$$A\mathbf{x} = \mathbf{b}, \qquad A := \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}, \qquad \mathbf{b} := \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{1}$$

Elimination turned this into

$$U\mathbf{x} = \mathbf{c}, \qquad U := \begin{bmatrix} 1 & 1 \\ 0 & -4 \end{bmatrix}, \qquad \mathbf{c} := \begin{bmatrix} 1 \\ -3 \end{bmatrix}. \tag{2}$$

How can we get U from A without doing elimination (in particular without introducing the auxiliary, but ultimately unimportant, variables x_1 and x_2)?

Here is the right idea: view elimination as actions on vectors. In passing from A to U and from $\mathbf{b}$ to $\mathbf{c}$, we had the three "evolutions" of vectors:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 \\ -2 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \qquad \text{and} \qquad \begin{bmatrix} 1 \\ -1 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ -3 \end{bmatrix}.$$

The first two "mappings" send the columns of A into the columns of U, and the third "transforms" $\mathbf{b}$ into $\mathbf{c}$.

I claim that all three mappings are the same, and they all have the same structure as the elimination procedure: don't change the first row, and define the new second row by subtracting 2 times the old first row from the old second row. In symbols,

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \mapsto \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix}.$$

Here is the brilliance of our forebears in linear algebra: represent this transformation by a matrix-vector multiplication. That is, if

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix},$$

then find a matrix E such that

$$E\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix}.$$

Here's how we do it. Remember that $E\mathbf{v}$ is fundamentally a *linear combination of the columns of E "weighted" by the entries of* $\mathbf{v}$. So let's look for a linear combination and rewrite

$$\begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix} = \begin{bmatrix} v_1 \\ -2v_1 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \mathbf{v}.$$

Do you completely understand each calculation above? I hope so, and I hope you agree that setting

$$E := \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$$

was the right idea. Double-check right now that with E as defined above (":=" for me means "defined to be equal"), we have

$$E\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad E\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \qquad \text{and} \qquad E\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}. \tag{3}$$

So here is what we know. The original problem (1) tells us that if $A\mathbf{x} = \mathbf{b}$, then $E(A\mathbf{x}) = E\mathbf{b}$. We're allowed to compute $E(A\mathbf{x})$ because $A\mathbf{x}$ is a vector, so $E(A\mathbf{x})$ is matrix-vector multiplication. But we can be more efficient.

Look back at (2) and notice (check) that $\mathbf{c} = E\mathbf{b}$. So we have $E(A\mathbf{x}) = \mathbf{c}$. But (2) also says that $\mathbf{c} = U\mathbf{x}$. So $E(A\mathbf{x}) = U\mathbf{x}$. Let's cut out the variables, and the parentheses. I vote that we *define* a new symbol EA to be equal to U. That is, the matrix-vector product EA is U:

$$EA = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -4 \end{bmatrix} = U.$$

That's a nice piece of bookkeeping, but the whole point of today was to compute U without doing elimination. How does this definition help? *Look at the columns.* Say that the columns of A are $\mathbf{a}_1$ and $\mathbf{a}_2$, i.e., $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$. The first two equalities in (3) say that the columns of U are just $E\mathbf{a}_1$ and $E\mathbf{a}_2$. So we really should define the matrix product EA to be

$$EA = E\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} := \begin{bmatrix} E\mathbf{a}_1 & E\mathbf{a}_2 \end{bmatrix}.$$

The columns of EA are the matrix-vector products of E with the columns of A.

Compute right now

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix}.$$

What happens if you flip the order of these two matrix factors?

I think this is a remarkable development. We have compressed all of the elimination into the action of multiplying by E.

What happens with a more general problem? We probably want to subtract a multiple of the first row from the second row and leave the first row unchanged. Here's how we do it. Let ℓ be any real number and define

$$E := \begin{bmatrix} 1 & 0 \\ -\ell & 1 \end{bmatrix}.$$

Let

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Calculate $E\mathbf{v}$ (this is old stuff) and EA (this is new stuff). Do you see that ℓ times the first row is subtracted from the second and that the first row is unchanged?

Last, look back at our old problem

$$\begin{bmatrix} 4 & 1 \\ 6 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 10 \\ 5 \end{bmatrix}.$$

I claim that the right elimination matrix is

$$E := \begin{bmatrix} 1 & 0 \\ -3/2 & 1 \end{bmatrix}.$$

When you multiply by E, do you get the same results as you did on January 18?

Day 7: Wednesday, January 25. We talked about elimination matrices for 3×3 systems on pp. 60–61 and augmented matrices on pp. 63–64.

More verbosely, we want to encode the action of subtracting ℓ times row j from row i. We do this by multiplying by the matrix E_{ij}, which is the identity matrix with the (i, j)-entry replaced by $-\ell$. I called this a "claim" in class. Proving this claim for a problem of arbitrary size could be a challenging exercise in notation, but here's a very particular situation.

Say we want to subtract 1/7 times row 2 from row 3 in a 3×3 problem. Then $\ell = 1/7$, $j = 2$, and $i = 3$. That is, the $(3, 2)$-entry of the elimination matrix should be $-1/7$. So we should represent this action by the matrix

$$E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/7 & 1 \end{bmatrix}.$$

I didn't say the following in class, but you should think about it. Subtracting 1/7 times row 1 from row 3 means that we want to do the following "transformation" on a vector:

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \mapsto \begin{bmatrix} v_1 \\ v_2 \\ v_3 - v_2/7 \end{bmatrix}. \tag{4}$$

Note that we didn't change the first two rows. We can represent this action by matrix-vector multiplication if we do some clever algebra:

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 - v_2/7 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ -1/7 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/7 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}. \tag{5}$$

And there we see E_{32} again.

For actions on small systems, you should be able to represent the "transformation" effected by "Subtract ℓ times row j from row i" for given values of ℓ, j, and i via a "mapping" like that in (4). Then you should be comfortable doing vector arithmetic and recognizing matrix-vector multiplication as *a linear combination of the columns* to get the equalities in (5).

Day 8: Friday, January 27. Here is a statement in English: "Subtract ℓ times row 1 of a column vector with 3 entries from row 2 of that vector, and do not change any other entries." Here is that statement in vector notation:

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \mapsto \begin{bmatrix} v_1 \\ v_2 - \ell v_1 \\ v_3 \end{bmatrix}. \tag{1}$$

We can encode this action into an matrix using the claim from Wednesday or the second box on p. 60. In the notation of that claim/box, we have $j = 2$ and $i = 1$, so the matrix is

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -\ell & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that only row 2 in this matrix is different from the 3×3 identity matrix, which is

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

Likewise, only column 1 in E is not the same as column 1 in I.

How can we prove this claim? The proof would have to work for arbitrarily large vectors, not just ones with 2 or 3 entries. We could easily get bogged down in notation. Instead, we try to recognize matrix-vector multiplication in the transformation (1). This is fundamentally a linear combination of vectors. We can see the linear combination via algebra:

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

Then we worked on a whole system with elimination. Specifically, we studied a problem from Day 4, which was

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & -1 & -3 \\ 2 & 3 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix}.$$

Call this matrix A and the vector on the right $\mathbf{b}$, so the problem is $A\mathbf{x} = \mathbf{b}$. We want to use the pivot of 1 from the first row (the $(1,1)$-entry of A) to eliminate the 3 and the 2 below. So we want to subtract 3 times row 1 from row 2, which requires the elimination matrix

$$E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

And we want to subtract 2 times row 1 from row 3, which requires the elimination matrix

$$E_{31} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}.$$

You can calculate (but I doubt you'd want to)

$$E_{31}E_{21}A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -6 \\ 0 & -1 & -1 \end{bmatrix}.$$

Now we want to use the pivot of -7 to eliminate the -1 below that, and so we want to subtract $-1/-7 = 1/7$ times row 2 from row 3. Then we want to multiply by the elimination matrix

$$E_{32} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/7 & 1 \end{bmatrix}.$$

This leads to

$$E_{32}E_{31}E_{21}A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -6 \\ 0 & 0 & -1/7 \end{bmatrix}.$$

This is upper-triangular!

I hope you saw E_{31} and E_{21} directly from the structure of A. But you definitely did not need to see (and probably could not see) E_{32} from A. The structure of E_{32} was only apparent from the structure of $E_{31}E_{21}A$. Constructing elimination matrices is an *iterative* process.

Also, look again at the structure of $E_{31}E_{21}A$. The elimination process has really become a 2×2 situation: we want to make

$$\begin{bmatrix} -7 & -6 \\ -1 & -1 \end{bmatrix}$$

upper-triangular, but we want to do so via action in a 3×3 world.

Here are some questions. We really have four factors in the matrix product $E_{32}E_{31}E_{21}A$ that produces the upper-triangular structure. How important is the order? We could probably have subtracted 2 times row 1 from row 3 before we subtracted 3 times row 1 from row 2, and so we might expect $E_{31}E_{21}A = E_{21}E_{31}A$. That is, in a just world, E_{31} and E_{21} should commute (but remember that $AB \neq BA$ typically for matrices—make up some matrices and do the arithmetic!). Next, we have three elimination matrices in a row: $E_{32}E_{31}E_{21}$. They all have a very similar, but not identical, structure. Can we possibly combine them into one big elimination matrix and cut down on the matrix multiplication?

Finally, our work above really shows that if $\mathbf{x}$ solves $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x}$ also solves $E_{32}E_{31}E_{21}A\mathbf{x} = E_{32}E_{31}E_{21}\mathbf{b}$. But why should a solution to $E_{32}E_{31}E_{21}A\mathbf{x} = E_{32}E_{31}E_{21}\mathbf{b}$ also solve $A\mathbf{x} = \mathbf{b}$. In other words, we know

$$A\mathbf{x} = \mathbf{b} \Longrightarrow E_{32}E_{31}E_{21}A\mathbf{x} = E_{32}E_{31}E_{21}\mathbf{b},$$

but do we have

$$E_{32}E_{31}E_{21}A\mathbf{x} = E_{32}E_{31}E_{21}\mathbf{b} \Longrightarrow A\mathbf{x} = \mathbf{b}?$$

Just because we have an arrow going one way doesn't mean it should go the other way!

These questions will guide some of our upcoming treatments of matrices, which will step away from systems for a bit.

Day 9: Monday, January 30. We finished one last elimination question: what if you suddenly don't have a pivot? Say your elimination work produces a matrix like

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

The first column looks good, but then you go to the second row and want to use that row to get rid of the 5 in the third row, right? But you can't, because of the 0 in the $(2, 2)$-entry. What do you do? If we were working at the level of the system of equations with the x-variables, you'd just interchange the second and third equations. Remember, *the goal is now to encode the action on the system via the action of a matrix.* What matrix P satisfies

$$P \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_3 \\ v_2 \end{bmatrix}?$$

I claim it's

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

You can check this in two ways: (1) just do the matrix-vector multiplication and (2) rewrite (v_1, v_3, v_2) as a linear combination weighted by v_1, v_2, and v_3, and then recognize that linear combination as a certain matrix times (v_1, v_2, v_3).

This is all covered on pp. 62–63. Now is a good time to read Worked Examples 2.3A and 2.3 B on pp. 64–65 before calling it quits on elimination.

We then switched focus to matrix arithmetic. This is how math goes: we started with a concrete problem (solving linear systems of equations), we developed a tool for systematically approaching that problem (elimination encoded in matrices), and now we are going to study the tool in part for the sake of studying the tool and in part in the hopes of learning more about the original problem.

We said that a matrix A is $m \times n$ (read "m times n"—this is not the product mn!) if A has m rows and n columns. Rows always come before columns (except alphabetically). Here are three important matrices in which 1 shows up in the size:

- The column vector: it's an $n \times 1$ matrix. Here is a 3×1 matrix:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Thus all vectors are matrices, but not all matrices are vectors. (Except maybe in parts of Chapter 3 later on.)

- The row vector: it's a $1 \times n$ matrix. Here is a 1×3 matrix:

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

• The 1×1 matrix, like $\begin{bmatrix}1\end{bmatrix}$. This is not the same as the number 1! The array containing the number 1 is just not the same as the number 1. (This is a matter of philosophy and deeper math—what exactly is an "array"?! I can answer that for you, but it will involve set theory from 2345/2390.)

We can add matrices if they are the same size—just add the corresponding entries. We can multiply matrices by scalars—just multiply all the entries by the given scalar. We had an enthusiastic discussion about how matrices of different size are inherently different objects, even if they have all the same entries:

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

We wanted to define the matrix product AB by saying that the columns of AB are the matrix-vector product of A with the columns of B. This means that each column of AB is a linear combination of the columns of A weighted by the entries of a column of B. This means that the columns of B have to have as many entries as A has columns. So if A is $m \times n$, then B has to be $n \times p$. The numbers m and p don't have to have anything to do with each other. If all else fails, keep this in mind:

$$AB = A\begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_p\end{bmatrix} = \begin{bmatrix}A\mathbf{b}_1 & \cdots & A\mathbf{b}_p\end{bmatrix}, \quad \text{where} \quad A \text{ is } m \times n \text{ and each } \mathbf{b}_k \text{ is } n \times 1.$$

This material is discussed on pp. 70–71. You should understand everything on these two pages. Go back to Worked Example 2.3 C on p. 65 and do the arithmetic—it's the product of a 3×2 and a 2×2 matrix. If you have the energy, look at the matrix S in Worked Example 2.4 A on p. 76 and check the calculations of S^2 and S^3. You might find the application of this matrix interesting (or maybe not).

Day 10: Wednesday, February 1. We talked over several different interpretations of matrix multiplication. The fundamental definition of AB when A is $m \times n$ and B is $n \times p$ is that AB is formed by A multiplying the columns of B:

$$AB = A\begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_p\end{bmatrix} = \begin{bmatrix}A\mathbf{b}_1 & \cdots & A\mathbf{b}_p\end{bmatrix}.$$

Let

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

Then

$$AB = \begin{bmatrix}11\end{bmatrix}.$$

This is the product of a 1×2 matrix with a 2×1 matrix, so the result is a 1×1 matrix. The only column of B is the vector

$$\mathbf{b}_1 = \begin{bmatrix}3\\4\end{bmatrix},$$

so $B = \begin{bmatrix} \mathbf{b}_1 \end{bmatrix}$. Then $AB = \begin{bmatrix} A\mathbf{b}_1 \end{bmatrix}$. This is really, therefore, a matrix-vector product. The matrix A has the two columns $\mathbf{a}_1 = \begin{bmatrix} 1 \end{bmatrix}$ and $\mathbf{a}_2 = \begin{bmatrix} 2 \end{bmatrix}$. Then

$$A\mathbf{b}_1 = 3\mathbf{a}_1 + 4\mathbf{a}_2 = 3\begin{bmatrix} 1 \end{bmatrix} + 4\begin{bmatrix} 2 \end{bmatrix} = \begin{bmatrix} 3 \end{bmatrix} + \begin{bmatrix} 8 \end{bmatrix} = \begin{bmatrix} 11 \end{bmatrix}.$$

Look what happens if we flip things around. Say

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

Then A is 1×2 and B is 2×1, so AB is defined and is 2×2. Now the columns of B are the vectors $\mathbf{b}_1 = \begin{bmatrix} 3 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 4 \end{bmatrix}$, so $AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 \end{bmatrix}$. The matrix-vector product $A\mathbf{b}_1$ is the linear combination of the columns of A weighted by the entries of $\mathbf{b}_1$. But A has only one column and $\mathbf{b}_1$ has only one entry, so

$$A\mathbf{b}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

Similarly, and laboriously,

$$A\mathbf{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 4 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}.$$

Thus

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}.$$

This is by no means the only way to view matrix multiplication. We keep asking what the columns are doing in AB: the columns of AB are formed by A multiplying the columns of B. But the rows are also doing something. I claim that the rows of AB are formed by multiplying the rows of A with all of B. If A is $m \times n$ and B is $n \times p$, then any row of A is $1 \times n$, so that row times B is $(1 \times n)$ times $(n \times p)$, which is $1 \times p$. That should feel comforting, because there are p columns in B, so each row of B has p entries. Here is the formula: if

$$A = \begin{bmatrix} \overrightarrow{a_1} \\ \vdots \\ \overrightarrow{a_m} \end{bmatrix},$$

then

$$AB = \begin{bmatrix} \overrightarrow{a_1}B \\ \vdots \\ \overrightarrow{a_m}B \end{bmatrix}.$$

Here I am writing the rows of A as $\overrightarrow{a_k}$ for $k = 1, \ldots, m$.

Put

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

Then A is 2×3 and B is 3×2, so AB is defined and 2×2. Let's calculate AB by "rows of A times B." We have

$$\begin{bmatrix} 1 & 3 & 5 \end{bmatrix} B = \begin{bmatrix} 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}.$$

This is a 1×3 matrix times a 3×2 matrix, so the output needs to be a 1×2 matrix. If we think about the action of $\begin{bmatrix} 1 & 3 & 5 \end{bmatrix}$ on the columns of B, we get

$$\begin{bmatrix} 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1\begin{bmatrix} 1 \end{bmatrix} + 0\begin{bmatrix} 3 \end{bmatrix} + 0\begin{bmatrix} 5 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \quad \text{and (more succinctly)} \quad \begin{bmatrix} 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \end{bmatrix}.$$

I thought this was excessive and pedantic on the board, but it feels worse to type! I hope you are seeing that all this is just repeated applications of the original definition of matrix multiplication. The first row of AB is then

$$\begin{bmatrix} 1 & 6 \end{bmatrix}.$$

Much more briefly, the second row is

$$\begin{bmatrix} 2 & 4 & 6 \end{bmatrix} B = \begin{bmatrix} 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 8 \end{bmatrix}.$$

All together,

$$AB = \begin{bmatrix} 1 & 6 \\ 2 & 8 \end{bmatrix}.$$

If you do this product again via the original definition (A acting on columns of B), do you get the same thing?

Here is a third way to multiply AB, and maybe I should have started with this. We've seen how A acts on the columns of B to produce columns of AB and how the rows of A act on all of B to produce rows of AB. We can also ask how a *single entry* of AB arises. Here's the answer: the (i, j) entry of AB is the dot product of row i of A and column j of B. Here's a proof in a very special case: A will be a 3×2 matrix and B will be a 2×3 matrix, and I'll show you the (1,1) entry of AB. Write $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix}$. Then $AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix}$. The $(1, 1)$ entry of AB falls in the first column, so look at $A\mathbf{b}_1$. Let's write

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_1 = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix}.$$

I am giving the entries of $\mathbf{b}_1$ two subscripts to indicate where they fall in $\mathbf{b}_1$ and that they belong to $\mathbf{b}_1$. Also, b_{21} is the $(2, 1)$-entry of B, right?

Anyway,

$$A\mathbf{b}_1 = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = b_{11}\mathbf{a}_1 + b_{12}\mathbf{a}_2.$$

We want the first entry of $A\mathbf{b}_1$ (right?), so we should write out the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ as

$$\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} \quad \text{and} \quad \mathbf{a}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix}$$

Observe that, for example, a_{32} is the $(3,2)$-entry of A. So we get

$$A\mathbf{b}_1 = b_{11}\mathbf{a}_1 + b_{12}\mathbf{a}_2 = b_{11}\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + b_{12}\begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} = \begin{bmatrix} b_{11}a_{11} \\ b_{11}a_{21} \\ b_{11}a_{31} \end{bmatrix} + \begin{bmatrix} b_{12}a_{12} \\ b_{12}a_{22} \\ b_{12}a_{32} \end{bmatrix} = \begin{bmatrix} b_{11}a_{11} + b_{12}a_{12} \\ b_{11}a_{21} + b_{12}a_{22} \\ b_{11}a_{31} + b_{12}a_{32} \end{bmatrix}.$$

The first entry of $A\mathbf{b}_1$, which is the $(1,1)$-entry of AB, is therefore

$$b_{11}a_{11} + b_{12}a_{12} = \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} \cdot \begin{bmatrix} b_{11} \\ b_{12} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} \cdot \mathbf{b}_1.$$

and this is the dot product of the first row of A, which is $\begin{bmatrix} a_{11} & a_{12} \end{bmatrix}$, with the first column of B, which is $\mathbf{b}_1$. Somebody tell me why I thought doing this on the board with A and B as matrices of arbitrary (but compatible) sizes was a good idea.

Here is one last product:

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

This is the product of a 2×2 matrix and a 2×3 matrix, so the result is 2×3. You might be happier drawing lines in these matrices to separate the rows and columns:

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

Here are all the dot products for the (i,j)-entries of the final product:

$$(1,1)\text{-entry} = \text{row 1 of } A \cdot \text{column 1 of } B = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = (1 \cdot 1) + (3 \cdot 0) = 1,$$

$$(1,2)\text{-entry} = \text{row 1 of } A \cdot \text{column 2 of } B = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} = (1 \cdot 0) + (3 \cdot 2) = 6$$

$$(1,3)\text{-entry} = \text{row 1 of } A \cdot \text{column 3 of } B = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \end{bmatrix} = (1 \cdot 3) + (3 \cdot 0) = 3$$

$$(2,1)\text{-entry} = \text{row 2 of } A \cdot \text{column 1 of } B = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = (2 \cdot 1) + (4 \cdot 0) = 2$$

$$(2,2)\text{-entry} = \text{row 2 of } A \cdot \text{column 2 of } B = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} = (2 \cdot 0) + (4 \cdot 2) = 8$$

$$(2,3)\text{-entry} = \text{row 2 of } A \cdot \text{column 3 of } B = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \end{bmatrix} = (2 \cdot 3) + (4 \cdot 0) = 6$$

All together,

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 6 & 3 \\ 2 & 8 & 6 \end{bmatrix}.$$

Here are all the matrix products that we did today. Can you do each of them in the three different ways? Which way do you like the most? The least? Why?

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

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

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

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

This material is covered on pp. 71–72. Look at "ways" 1, 2, and 3.

Day 11: Friday, February 3. Today's discussion covered (and augmented) pp. 72–73.

For practice, we multiplied two elimination matrices that we met on Day 8:

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

It looks like we just condensed the multipliers -3 and -2 into the appropriate slots in the same matrix!

Then we talked about some algebra. When working with real numbers, multiplication is associative:

$$2 \cdot (3 \cdot 4) = 2 \cdot 12 = 24 \quad \text{and} \quad (2 \cdot 3) \cdot 4 = 6 \cdot 4 = 24.$$

The same is true for matrices: if A, B, and C are such that the products AB and BC are defined, then

$$A(BC) = (AB)C.$$

This is the associativity of matrix multiplication: the way you group the factors doesn't matter. Exercise for you: how do the sizes of A, B, and C relate so that the products above are defined?

Multiplication of real numbers is also commutative:

$$2 \cdot 3 = 6 \quad \text{and} \quad 3 \cdot 2 = 6.$$

Unfortunately, the same is not true for matrices: if AB and BA are both defined (exercise for you: show that A and B must be square if both products are defined), then, in general,

$AB \neq BA$. *The order in which you write the factors in a matrix product matters very much.* Check out Example 2.4 B!

Last, multiplication of real numbers distributes over addition:

$$2 \cdot (3+4) = 2 \cdot 7 = 14 \quad \text{and} \quad 2 \cdot (3+4) = (2 \cdot 3) + (2 \cdot 4) = 6 + 8 = 14.$$

This is also true of matrix multiplication and addition:

$$A(B+C) = AB + AC \quad \text{and} \quad (B+C)D = BD + CD.$$

Again, what are the sizes of A, B, C, and D so that all the products above are defined? Note that in the first equality I have written $AB + AC$, not $BA + BC$, and similarly in the second equality.

Here is a nice application of how matrix multiplication distributes over addition. Consider the product

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 5 & 0 & 7 \\ 0 & 6 & 0 \end{bmatrix}.$$

Let me rewrite these matrices in a clever way:

$$\left(\begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 3 \\ 0 & 4 \end{bmatrix} \right) \left(\begin{bmatrix} 5 & 0 & 7 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 6 & 0 \end{bmatrix} \right).$$

I am going to call this product $(A+B)(C+D)$. Then I will distribute $A+B$ over the addition $C+D$ to get

$$(A+B)(C+D) = (A+B)C + (A+B)D.$$

Then I will distribute C and D over the addition $A+B$ to get

$$(A+B)C + (A+B)D = AC + BC + AD + BD.$$

Let's look at the first term, AC. This is the product

$$\begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 5 & 0 & 7 \\ 0 & 0 & 0 \end{bmatrix}.$$

Morally, this should feel like the first column of our original first factor times the first row of our original second factor. And it is. I claim that

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

You should check this claim by doing the arithmetic. This is a great chance to use the dot product—in the product on the right, the first factor has only one entry in each row, and the second factor has only one entry in each column.

I also claim that BC and AD are both the 2×2 zero matrix. And so in the end you get

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 5 & 0 & 7 \\ 0 & 6 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 5 & 0 & 7 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} \begin{bmatrix} 0 & 6 & 0 \end{bmatrix} = (\text{column } 1 \times \text{row } 1) + (\text{column } 2 \times \text{row } 2).$$

This is Strang's "fourth way" of multiplying matrices.

Day 12: Monday, February 6. We started with block matrices and block multiplication. You should read all of pp. 74–75. In equation (6) on p. 75, just say $A = \begin{bmatrix}1\end{bmatrix}$, since we haven't discussed inverse matrices.

Here is the example of block matrices that we did. (It's morally the same as Example 4 on p. 75.) Put

$$A_{11} \begin{bmatrix}1\end{bmatrix}, \qquad A_{12} = \begin{bmatrix}2 & 1\end{bmatrix}, \qquad A_{21} = \begin{bmatrix}3\\2\end{bmatrix}, \quad \text{and} \quad A_{22} = \begin{bmatrix}-1 & -3\\ 3 & 1\end{bmatrix}$$

to recognize an old friend:

$$A = \left[\begin{array}{c|cc} 1 & 2 & 1 \\ \hline 3 & -1 & -3 \\ 2 & 3 & 1 \end{array}\right] = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}.$$

Hopefully this block structure makes clear what we want to do with elimination: use A_{11} to zero out A_{21} while retaining A_{12} as it is and not caring about what A_{22} becomes. (We'll care later, but not now.)

We can represent elimination in the first column of A as a block matrix. Put

$$E_{11} = \begin{bmatrix}1\end{bmatrix}, \qquad E_{12} = \begin{bmatrix}0 & 0\end{bmatrix}, \qquad E_{21} = \begin{bmatrix}-3\\-2\end{bmatrix}, \quad \text{and} \quad E_{22} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

Then

$$E = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline -3 & 1 & 0 \\ -2 & 0 & 1 \end{array}\right] = \begin{bmatrix}E_{11} & E_{12}\\ E_{21} & E_{22}\end{bmatrix}.$$

We followed our noses to multiply

$$EA = \begin{bmatrix}E_{11} & E_{12}\\ E_{21} & E_{22}\end{bmatrix}\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix} = \left[\begin{array}{c|c} E_{11}A_{11} + E_{12}A_{21} & E_{11}A_{12} + E_{12}A_{22} \\ \hline E_{21}A_{11} + E_{21}A_{21} & E_{21}A_{12} + E_{22}A_{22} \end{array}\right] = \left[\begin{array}{c|cc} 1 & 2 & 1 \\ \hline 0 & -7 & -6 \\ 0 & -1 & -1 \end{array}\right].$$

You should check that all of the matrix products above are defined and that they come out to what I say they do. This is good practice with dimension counting!

I think block matrices make clear the structure of a general elimination problem. Say that A is an $n \times n$ matrix and that the $(1,1)$-entry of A is nonzero, so we can use it as a pivot. That is, A has the form

$$A = \left[\begin{array}{c|c} a_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right].$$

Exercise for you: what are the dimensions of A_{12}, A_{21}, and A_{22}? Assume that $a_{11} \neq 0$ and put

$$E = \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline -\dfrac{1}{a_{11}}A_{21} & I \end{array}\right].$$

Here $\mathbf{0}$ is a row vector of zeros (what is its size?). I claim that

$$EA = \left[\begin{array}{c|c} a_{11} & A_{12} \\ \hline \mathbf{0} & \widetilde{A}_{22} \end{array}\right],$$

where $\mathbf{0}$ is a column vector of zeros (what is its size?) and $\widetilde{A}_{22}$ is...something. You should check this claim. It's abstract, but we did a concrete 3×3 version just above.

Then we started talking about matrix inverses. We even proved something: say that $A\mathbf{x} = \mathbf{b}$ and there is a matrix M such that $MA = I$, where I is the identity matrix. Then $MA\mathbf{x} = M\mathbf{b}$. And $MA\mathbf{x} = I\mathbf{x} = \mathbf{x}$, and so $\mathbf{x} = M\mathbf{b}$. This is a *uniqueness* result: the only possible solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = M\mathbf{b}$. But this does not show that $\mathbf{x} = M\mathbf{b}$ actually *is* a solution; we need to check if $AM\mathbf{b} = \mathbf{b}$. If we also had $AM = I$, then we'd be done, for then $AM\mathbf{b} = I\mathbf{b} = \mathbf{b}$. We shouldn't expect this automatically, since we don't know if A and M commute.

Over the next day or two, we will cover all of pp. 83–89, although maybe not in the exact same order as Strang does.

Day 13: Wednesday, February 8. Last time we said we'd be happy if, given an $n \times n$ matrix A, we could find an $n \times n$ matrix M such that $AM = MA = I$, where I is the $n \times n$ identity matrix. The reason we'd be happy was Note 3 on p. 83, which you should read and think about now.

Here is reason to be happier: there is only one such M. Suppose that M and $\widetilde{M}$ are $n \times n$ matrices such that

$$AM = MA = I \quad \text{and} \quad A\widetilde{M} = \widetilde{M}A = I.$$

I claim that, then, $M = \widetilde{M}$. Here's why. Let's show $M - \widetilde{M} = \mathsf{O}$. (I will write O to denote the $n \times n$ matrix whose entries are all 0.)

We have

$$M - \widetilde{M} = (M - \widetilde{M})I = (M - \widetilde{M})AM = (MA - \widetilde{M}A)M = (I - I)M = \mathsf{O}M = \mathsf{O}.$$

Do you agree with each equality? Can you explain why each equality is true?

We therefore defined the *inverse* of A to be the $n \times n$ matrix M such that $AM = MA = I$, if such a matrix M exists. We said that such an A was *invertible*. We say "the" inverse, not "an"' inverse, because by the calculation $M = \widetilde{M}$ above, there can only be one inverse of A: the inverse is unique. We write $M = A^{-1}$.

The following also turns out to be true. If M is a matrix such that $AM = I$, then automatically $MA = I$. And if M is a matrix such that $MA = I$, then $AM = I$. So, we can say "$AM = I$ if and only if $MA = I$." This is not as easy to prove as the uniqueness of the inverse above. Now is a good time to read Note 2 on p. 83 and make sure you know how to prove that. These are good exam proofs!

We temporarily skipped Notes 4, 5, and 6 on p. 84 and worked on the inverse of the product AB. We also talked about the inverse of an elimination matrix and said that if E is the elimination matrix whose (i, j)-entry is $-\ell$, then E^{-1} is the elimination matrix whose (i, j)-entry is ℓ. You should read and work through Examples 2 and 3 on p. 85.

Last, we started to think about how to calculate the inverse in general, when we can't assume that the matrix has the special structure of an elimination matrix. Consider the 2×2 situation—this is just hard enough. If A is an invertible 2×2 matrix, then its inverse A^{-1} satisfies $AA^{-1} = I$, where I is the 2×2 identity matrix. Write

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

so $I = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}$. Write $A^{-1} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}$. We are thinking of A^{-1} as "the unknown" right now, and usually we denote unknowns with x.

So we have

$$\begin{aligned} AA^{-1} = I &\iff A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \\ &\iff \begin{bmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \\ &\iff A\mathbf{x}_1 = \mathbf{e}_1 \quad \text{and} \quad A\mathbf{x}_2 = \mathbf{e}_2. \end{aligned}$$

Do you agree with each step in the $\iff$ chain above?

Here is the good news: we have reduced the problem of finding the inverse A^{-1} to solving the two linear equations $A\mathbf{x}_1 = \mathbf{e}_1$ and $A\mathbf{x}_2 = \mathbf{e}_2$. Remember, in principle you know A, and the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are defined above. We've spent the whole course thinking about these two problems, and we should be able to solve each of them with elimination and back substitution. (Assuming there are two pivots...) This is basically p. 86 up to the sentence "The Gauss-Jordan method computes..."

Day 14: Friday, February 10. Today we did one 2×2 problem that proceeded analogously to the 3×3 problem on pp. 86–87. A briefer treatment of a 2×2 inverse appears in Example 4 on p. 87. We will do a 3×3 beast next time.

Let

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

We want to find the inverse of A (if it exists). Last time we said that we had to solve the two systems

$$A\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} =: \mathbf{e}_1 \quad \text{and} \quad A\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} =: \mathbf{e}_2$$

and then $\mathbf{x}_1$ and $\mathbf{x}_2$ would be the columns of A^{-1}. That is, $A^{-1} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}$.

The right idea is not to solve these two linear systems separately (although we could totally do that) but to do them simultaneously. Instead of doing elimination on the two 2×3 augmented matrices $\begin{bmatrix} A & \mathbf{e}_1 \end{bmatrix}$ and $\begin{bmatrix} A & \mathbf{e}_2 \end{bmatrix}$, we'll do elimination on the *one* 2×4 augmented matrix $\begin{bmatrix} A & \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} A & I \end{bmatrix}$. Here I is the 2×2 identity matrix. Since

$$\begin{bmatrix} A & I \end{bmatrix} = \left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 5 & 2 & 0 & 1 \end{array}\right],$$

we should subtract 5/3 times row 1 from row 2. (Quick! What is the appropriate 2×2 elimination matrix that does this? And look out below for the identity $E\begin{bmatrix}A & I\end{bmatrix} = \begin{bmatrix}EA & E\end{bmatrix}$.)

So we have

$$\begin{bmatrix}A & I\end{bmatrix} = \left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 5 & 2 & 0 & 1 \end{array}\right] \xrightarrow{(2) - 5/3\cdot(1)} \left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1/3 & -5/3 & 1 \end{array}\right].$$

What does this matrix on the right *mean*? Remember, sometimes matrices *act*, and sometimes matrices *encode* data. What's encoded here? It's the *pair* of upper-triangular systems

$$\begin{bmatrix}3 & 1 \\ 0 & 1/3\end{bmatrix}\mathbf{x}_1 = \begin{bmatrix}1 \\ -5/3\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}3 & 1 \\ 0 & 1/3\end{bmatrix}\mathbf{x}_2 = \begin{bmatrix}0 \\ 1\end{bmatrix}.$$

We could solve these via back-substitution as we have done in the past, but here's a chance to do something new, and better. To get the upper-triangular form, we used the pivot 3 in the $(1,1)$-entry of A to clear out the rest of the first column *below* the $(1,1)$-entry. This revealed the pivot of 1/3 in the $(2,2)$-entry. Now we are going to use this pivot in the $(2,2)$-entry to clear out the rest of the second column *above* the $(2,2)$-entry. Before we do that, though, we will divide by 1/3 (a.k.a. multiply by 3) in the second row to make this pivot nicer:

$$\left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1/3 & -5/3 & 1 \end{array}\right] \xrightarrow{3\cdot(2)} \left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1 & -5 & 3 \end{array}\right].$$

Why was this legal? Remember, if we strip away *all* the matrices and put the variables back in, the 2×4 matrix on the left represents the two problems

$$\begin{cases} 3x_1 + \quad x_2 = 1 \\ \quad (1/3)x_2 = -(5/3) \end{cases} \quad \text{and} \quad \begin{cases} 3x_1 + \quad x_2 = 0 \\ \quad (1/3)x_2 = 1 \end{cases}.$$

By multiplying both sides of the second equation in each system by 3, we should see that the two systems are the same as

$$\begin{cases} 3x_1 + x_2 = 1 \\ \quad x_2 = -5 \end{cases} \quad \text{and} \quad \begin{cases} 3x_1 + x_2 = 0 \\ \quad x_2 = 3 \end{cases}.$$

And this is why the pairs of systems represented by the matrices

$$\left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1/3 & -5/3 & 1 \end{array}\right] \quad \text{and} \quad \left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1 & -5 & 3 \end{array}\right]$$

are the same.

Now, how can we use the new pivot of 1 in the $(2,2)$-entry to clear out the 1 above that (i.e., the 1 in the $(2,1)$-entry)? I say we subtract row 2 from row 1:

$$\left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1 & -5 & 3 \end{array}\right] \xrightarrow{(1) - (2)} \left[\begin{array}{cc|cc} 3 & 0 & 6 & -3 \\ 0 & 1 & -5 & 3 \end{array}\right].$$

Look at the matrix in the left block: it's diagonal! This 2×4 problem represents the pair of systems

$$\begin{cases} 3x_1 = 6 \\ x_2 = -5 \end{cases} \quad \text{and} \quad \begin{cases} 3x_1 = -3 \\ x_2 = 3 \end{cases}.$$

These systems are *decoupled*: the unknowns don't both appear in the same equation. And so they are very easy to solve: there is nothing to do in the second equation, and we just divide the first equation by 3.

If we had divided the first row in our last matrix by 3, this would have been easier. So let's do that:

$$\left[\begin{array}{cc|cc} 3 & 0 & 6 & -3 \\ 0 & 1 & -5 & 3 \end{array}\right] \xrightarrow{1/3 \cdot (1)} \left[\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & 5 & 3 \end{array}\right].$$

This represents the pair of systems

$$\begin{cases} x_1 = 2 \\ x_2 = 5 \end{cases} \quad \text{and} \quad \begin{cases} x_1 = -1 \\ x_2 = 3, \end{cases}$$

which is the kind of system you've been hoping to meet since you started the class. There's nothing to do!

Let me summarize our work in three levels. First, we found A^{-1} by solving the systems $A\mathbf{x}_1 = \mathbf{e}_1$ and $A\mathbf{x}_2 = \mathbf{e}_2$, where

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$$

That is,

$$A^{-1} = \begin{bmatrix} 2 & 5 \\ -1 & 3 \end{bmatrix}.$$

You should check that $AA^{-1} = I$; this is just brute-force matrix multiplication.

Second, let's summarize our work without numbers. We started with the augmented matrix $\begin{bmatrix} A & I \end{bmatrix}$ and we did "downwards" elimination to convert this matrix into $\begin{bmatrix} U & \mathbf{?} \end{bmatrix}$, where U is upper-triangular and $\mathbf{?}$ is *some* matrix. This "downwards" elimination is the same elimination we've always done, and it hinged on having pivots. The new step was "upwards" elimination, and we used the pivots again to eliminate $\begin{bmatrix} U & \mathbf{?} \end{bmatrix}$ into the matrix $\begin{bmatrix} D & \mathbf{??} \end{bmatrix}$, where D is diagonal with *nonzero* numbers and $\mathbf{??}$ is *some other* matrix. This was only possible because the pivots, by definition, were *nonzero* numbers. The last step was to divide each row by its entry from D, and this gave us the matrix $\begin{bmatrix} I & A^{-1} \end{bmatrix}$.

Third, and last, let me put all the arithmetic together into one sequence:

$$\left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 5 & 2 & 0 & 1 \end{array}\right] \xrightarrow{(2) - 3 \cdot (1)} \left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1/3 & -5/3 & 1 \end{array}\right]$$

$$\xrightarrow{3 \cdot (2)} \left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1 & -5 & 3 \end{array}\right]$$

$$\xrightarrow{(1)-(2)} \left[\begin{array}{cc|cc} 3 & 0 & 6 & -3 \\ 0 & 1 & -5 & 3 \end{array}\right]$$

$$\xrightarrow{1/3\cdot(1)} \left[\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & 5 & 3 \end{array}\right].$$

Day 15: Monday, February 13. Today we inverted a 3×3 matrix. Another 3×3 example appears on pp. 86–87, and the strategy is the same there as here. Let A be our old friend

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

The inverse of A is the matrix $\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3\end{bmatrix}$ whose columns satisfy the *three different* linear systems

$$A\mathbf{x}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \qquad A\mathbf{x}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \qquad \text{and} \qquad A\mathbf{x}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}. \tag{2}$$

We could solve these three systems separately, but why do the same work three different times? Let's work on the augmented matrix

$$\begin{bmatrix}A & I\end{bmatrix} = \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 3 & -1 & -3 & 0 & 1 & 0 \\ 2 & 3 & 1 & 0 & 0 & 1 \end{array}\right].$$

In these typed notes, I'm not separating A and I with a horizontal line in the matrix $\begin{bmatrix}A & I\end{bmatrix}$, but I will put the line in when we're working with actual numbers.

We are going to do the same thing to $\begin{bmatrix}A & I\end{bmatrix}$ that we did to the 2×2 version on Day 14, but now we are going to keep much more careful track of the elimination matrices involved. This is going to be a painful amount of data to manage, but the payoff will be tremendous.

First, we want to use elimination to reduce A to upper-triangular form, with I coming along for the ride. Back on Day 8, we saw that the necessary elimination matrices were

$$E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{31} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad E_{32} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/7 & 1 \end{bmatrix}.$$

Let's use the pivot of 1 in the $(1,1)$-entry to clear out the rest of the first column of A and in the process act on I:

$$E_{31}E_{21}\begin{bmatrix}A & I\end{bmatrix} = \begin{bmatrix}E_{31}E_{21}A & E_{31}E_{21}\end{bmatrix} = \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -7 & -6 & -3 & 1 & 0 \\ 0 & -1 & -1 & -2 & 0 & 1 \end{array}\right].$$

Now let's use the pivot of -7 in the $(2,2)$-entry to clear out the rest of the second column:

$$E_{32}\begin{bmatrix}E_{31}E_{21}A & E_{31}E_{21}\end{bmatrix} = \begin{bmatrix}E_{32}E_{31}E_{21}A & E_{32}E_{31}E_{21}\end{bmatrix} = \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -7 & -6 & -3 & 1 & 0 \\ 0 & 0 & -1/7 & -11/7 & -1/7 & 1 \end{array}\right].$$

This says that the three problems (2) are equivalent to the three upper-triangular problems

$$U\mathbf{x}_1 = \begin{bmatrix} 1 \\ -3 \\ -11/7 \end{bmatrix}, \qquad U\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ -1/7 \end{bmatrix}, \qquad \text{and} \qquad U\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

where

$$U = E_{32}E_{31}E_{21}A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -6 \\ 0 & 0 & -1/7 \end{bmatrix}.$$

Of course, we could solve each by back-substitution, thanks to the upper-triangular structure of U. But why do the same thing three times when we could do one different thing only once? And, in particular, why work on an upper-triangular system when we could really work on a *diagonal* system?

Look at U. The diagonal elements of U are the pivots of A; they are all nonzero, and so we were able to use these pivots to perform elimination and get the upper-triangular structure. This was "downwards" elimination: we subtracted a multiple of an upper row from a lower row. For example, we subtracted 3 times row 1 of $\begin{bmatrix} A & I \end{bmatrix}$ from row 3 and got an equivalent system. Now we will perform "upwards" elimination: we will subtract a matrix of a lower row from an upper row. First, though, we will rescale the rows to avoid fractions.

Let

$$D_{33} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -7 \end{bmatrix}.$$

You should check that multiplying by D_{33} encodes the action of multiplying the third row by -7:

$$D_{33} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ -7v_3 \end{bmatrix}.$$

Then we calculate

$$\begin{aligned} D_{33} \begin{bmatrix} E_{32}E_{31}E_{21}A & E_{32}E_{31}E_{21} \end{bmatrix} &= \begin{bmatrix} D_{33}E_{32}E_{31}E_{21}A & D_{33}E_{32}E_{31}E_{21} \end{bmatrix} \\ &= \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -7 & -6 & -3 & 1 & 0 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right]. \end{aligned}$$

Now we want to use the 1 in the $(3,3)$-entry to clear out the first and second rows of the third column. This amounts to subtracting -6 times row 3 from row 2 and 1 times row 3 from row 1. You should check that the elimination matrices

$$E_{23} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad E_{13} := \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

do the job. Note that these elimination matrices have the multiplier above the diagonal—we haven't seen that before, but we also haven't had to subtract a lower row from an upper row before, either. Then

$$E_{23}E_{13}\begin{bmatrix} D_{33}E_{32}E_{31}E_{21}A & D_{33}E_{32}E_{31}E_{21}\end{bmatrix} = \begin{bmatrix} E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}A & E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}\end{bmatrix}$$
$$= \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & -10 & -1 & 7 \\ 0 & -7 & 0 & 63 & 7 & -42 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right].$$

I think you can guess that we'll divide the second row by -7, or, better yet, multiply by

$$D_{22} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/7 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

to find

$$D_{22}\begin{bmatrix} E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}A & E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}\end{bmatrix}$$
$$= \begin{bmatrix} D_{22}E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}A & D_{22}E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}\end{bmatrix}$$
$$= \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & -10 & -1 & 7 \\ 0 & 1 & 0 & 9 & 1 & -6 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right].$$

Last, we'll subtract 2 times row 2 from row 1 via the elimination matrix

$$E_{12} := \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

to conclude

$$E_{12}\begin{bmatrix} D_{22}E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}A & D_{22}E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}\end{bmatrix}$$
$$= \begin{bmatrix} E_{12}D_{22}E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}A & E_{12}D_{22}E_{23}E_{13}D_{33}E_{32}E_{31}E_{21}\end{bmatrix}$$
$$= \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 8 & -1 & -5 \\ 0 & 1 & 0 & -9 & -1 & 6 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right].$$

This should make us feel exhausted and burdened, but happy, too, for we have found A^{-1}. I claim

$$A^{-1} = \begin{bmatrix} 8 & -1 & -5 \\ -9 & -1 & 6 \\ 11 & 1 & -7 \end{bmatrix}.$$

Of course, you can go right ahead and check this by showing that $AA^{-1} = I$. We will analyze this in more detail next time.

Below I'll summarize the row-by-row arithmetic in case you want to check your work that way. On top of the arrow is the row operation; below is the matrix that we used.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 3 & -1 & -3 & 0 & 1 & 0 \\ 2 & 3 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow[E_{21}]{(2)-3\cdot(1)} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -7 & -6 & -3 & 1 & 0 \\ 2 & 3 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[E_{31}]{(3)-2\cdot(1)} \left[\begin{array}{rrr|rrr} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -7 & -6 & -3 & 1 & 0 \\ 0 & -1 & -1 & -2 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[E_{32}]{(3)-(-1/-7)\cdot(2)} \left[\begin{array}{rrr|rrr} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -7 & -6 & -3 & 1 & 0 \\ 0 & 0 & -1/7 & -11/7 & -1/7 & 1 \end{array}\right]$$

$$\xrightarrow[D_{22}]{-7\cdot(3)} \left[\begin{array}{rrr|rrr} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -7 & -6 & -3 & 1 & 0 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right]$$

$$\xrightarrow[E_{23}]{(2)-(-6)\cdot(3)} \left[\begin{array}{rrr|rrr} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -7 & 0 & 63 & 7 & -42 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right]$$

$$\xrightarrow[E_{13}]{(1)-1\cdot(3)} \left[\begin{array}{rrr|rrr} 1 & 2 & 0 & -10 & -1 & 7 \\ 0 & -7 & 0 & 63 & 7 & -42 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right]$$

$$\xrightarrow[D_{22}]{(-1/7)\cdot(3)} \left[\begin{array}{rrr|rrr} 1 & 2 & 0 & -10 & -1 & 7 \\ 0 & 1 & 0 & -9 & -1 & 6 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right]$$

$$\xrightarrow[E_{12}]{(1)-2\cdot(2)} \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 8 & 1 & -5 \\ 0 & 1 & 0 & -9 & -1 & 6 \\ 0 & 0 & 1 & 11 & 1 & -7 \end{array}\right]$$

yikes.

Day 16: Wednesday, February 15. We briefly revisited the terrible, horrible, no good, very bad calculation that we did on Day 15. Using the notation of that day, put

$$M_1 = E_{32}E_{31}E_{21} \quad \text{and} \quad M_2 = E_{12}D_{22}E_{23}E_{13}D_{33}.$$

I claim that all of our calculations showed

$$M_2M_1A = I.$$

So if we abbreviate $B = M_2M_1$, then we have $BA = I$, and so A is invertible with $A^{-1} = B$. That is, $A^{-1} = M_2M_1$. In other words, all of the elimination and diagonal matrices that we constructed to convert A into I multiply together to give A^{-1}.

Here is the algorithm that we developed for computing A^{-1}. We first use elimination to convert A to upper-triangular form. This is the matrix M_1A. The elimination proceeded "downwards" and is often called "Gaussian elimination." Then we did "upwards" elimination to convert M_1A into $I = M_2M_1A$. This is called "Gauss–Jordan elimination."

We succeeded with the "upwards" elimination because A has three pivots, i.e., the diagonal entries of the upper-triangular matrix M_1A are the nonzero numbers 1, -7, and $-1/7$. Look back at our work and convince yourself that the upwards elimination would have failed if we had a 0 in place of one of these three numbers. Note that we did not just work with elimination matrices but also "scaling" matrices (D_{22} and D_{33}) to make the diagonal entries equal to 1.

If you haven't already done so, you should read the example with K on pp. 86–87. Then read Example 4 and compare that result to Note 5 on p. 84; you can prove Note 5 by doing Problem 2.2.4. You should check the inverse calculation at the top of p. 88 by hand and then see if the steps in Example 5 make sense in the context of that calculation. Then read Example 6. Last, work through Worked Examples 2.5 A, B, and C. Computing a matrix inverse by hand via Gauss–Jordan elimination can be an ordeal, but you have lots of examples to consult!

We then reviewed for Friday's exam. There are several points that I want to make.

1. The dot product is only defined for column vectors of the same length:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + \cdots + x_ny_n.$$

While the dot product is a very useful tool for calculating the product of matrices, we do not define a dot product of matrices. In particular, please do not write $A \cdot B$ to denote the matrix product; rather, use "juxtaposition" AB.

2. If A is an $m \times n$ matrix and $\mathbf{x}$ is a $p \times 1$ column vector, then the matrix-vector product $A\mathbf{x}$ is defined only if $n = p$. In that case, if we write

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

then

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n.$$

Note that we have three different objects in play here: the scalars x_k for $k = 1, \ldots, n$, the $m \times 1$ column vectors $\mathbf{a}_k$ for $k = 1, \ldots, n$, and the matrix A. Note also that I am writing these three objects using different fonts in this typed document, and I handwrite them differently on the board—ordinary text for scalars, bold/underlined text for vectors, and uppercase text for matrices. Your writing should make it clear what kind of object is what. Good math needs good communication!

3. If A is an $m \times n$ matrix and B is a $p \times q$ matrix, then the matrix product AB is defined only if $n = p$. In that case, if we write

$$B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_q \end{bmatrix},$$

then

$$AB = A \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_q \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_q \end{bmatrix}.$$

Note that for $k = 1, \ldots, q$, each $\mathbf{b}_k$ is an $n \times 1$ column vector, so the matrix-vector product $A\mathbf{b}_k$ is defined, as A is an $m \times n$ matrix. Note also that AB is an $m \times q$ matrix.

4. Here is why we chose this particular definition of matrix-vector multiplication. Let A be $m \times n$, B be $n \times q$, and $\mathbf{x}$ be $q \times 1$. Then the matrix product AB and the matrix-vector products $(AB)\mathbf{x}$, $B\mathbf{x}$, and $A(B\mathbf{x})$ are all defined. (Quick: check that.) Way back on Day 6, we wanted AB to be the matrix such that $(AB)\mathbf{x} = A(B\mathbf{x})$. Does our definition above of AB allow us to juggle parentheses in this way? Let's check:

$$(AB)\mathbf{x} = \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_q \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_q \end{bmatrix} = x_1 A\mathbf{b}_1 + \cdots + x_q A\mathbf{b}_q,$$

$$B\mathbf{x} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_q \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_q \end{bmatrix} = x_1 \mathbf{b}_1 + \cdots + x_q \mathbf{b}_q,$$

and

$$A(B\mathbf{x}) = A(x_1\mathbf{b}_1 + \cdots + x_q\mathbf{b}_q) = A(x_1\mathbf{b}_1) + \cdots + A(x_q\mathbf{b}_q) = x_1 A\mathbf{b}_1 + \cdots + x_q A\mathbf{b}_q.$$

So we get, indeed,

$$(AB)\mathbf{x} = x_1 A\mathbf{b}_1 + \cdots + x_q A\mathbf{b}_q = A(B\mathbf{x}).$$

5. I love this calculation. It involves *algebra*: manipulating two seemingly different things and seeing they're really the same thing. And it involves *linearity*: we have used the identities

$$A(\mathbf{v}_1 + \mathbf{v}_2) = A\mathbf{v}_1 + A\mathbf{v}_2 \quad \text{and} \quad A(c\mathbf{v}) = cA\mathbf{v},$$

valid for any scalar c and matrix A and vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}$ for which the above matrix-vector multiplications are defined. This is *linear algebra*!

Day 17: Friday, February 17. You took an exam.

Day 18: Monday, February 20. We proved an if-and-only-if statement about matrix inverses. (To prove $P \iff Q$, you have to show $P \implies Q$ and $Q \implies P$.) Here is that statement: A is invertible if and only if for each $\mathbf{b}$, there is a unique $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$.

We've done the $\implies$ part a lot ($\mathbf{x} = A^{-1}\mathbf{b}$). How does the $\impliedby$ part go? Why is it the case that if we can always solve $A\mathbf{x} = \mathbf{b}$ uniquely, then A has an inverse? The answer comes from Problem 2.4.32. We choose the vector $\mathbf{b}$ to be something special: assuming A is $n \times n$, put $\mathbf{e}_k$ to be the vector with 1 in the kth row and 0 elsewhere. For example, if $n = 3$, then

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

Then for each k, we find a (unique) $\mathbf{x}_k$ such that $A\mathbf{x}_k = \mathbf{e}_k$. Put the $\mathbf{x}_k$ together into a matrix: $X = \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_k\end{bmatrix}$. Then $AX = \begin{bmatrix}\mathbf{e}_1 & \cdots & \mathbf{e}_n\end{bmatrix} = I$.

We now have a new test for invertibility. If you can find two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $A\mathbf{x}_1 = A\mathbf{x}_2$ but $\mathbf{x}_1 \neq \mathbf{x}_2$, then A is not invertible. Or if you can find a single vector $\mathbf{b}$ such that $A\mathbf{x} \neq \mathbf{b}$ for all $\mathbf{x}$, then A is not invertible. I like these tests because (1) they're useful and (2) they remind us that a matrix *acts* on vectors.

Then we deployed this test to create a new test involving pivots. All of our work has taught us that we can always transform a given matrix A into an upper-triangular matrix U by multiplying by a suitable matrix M. That is, $U = MA$ is upper-triangular. The matrix M will be the product of a bunch of elimination matrices and maybe permutation matrices, and therefore M is invertible. I claim that A is invertible if and only if MA is invertible. Try Problem 2.5.12.

So, when is MA invertible? Since MA is upper-triangular, we may as well ask when any upper-triangular matrix U is invertible. I claim that U is invertible if and only if all of the diagonal entries of U are nonzero. Look at the 3×3 case and consider the linear system $U\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix}\star & * & *\\0 & \star & *\\0 & 0 & \star\end{bmatrix}\mathbf{x} = \mathbf{b}.$$

If each $\star$ is nonzero (no one cares what $*$ is), then we can back-solve to figure out $\mathbf{x}$, and this $\mathbf{x}$ will be unique. Thus U is invertible by the work at the start of class. (I didn't say the following in class, but it's important for completeness: if a $\star$ is 0, then U is not invertible. This might be a little tricky to show in general. Here is Strang's argument from p. 88: if a $\star$ is 0, then we can do more elimination/permutation on U via a matrix $\widetilde{M}$ so that $\widetilde{MU}$ has a row whose entries are only 0. This matrix $\widetilde{MU}$ can't be invertible by the arguments that I give below in the specific example.)

So we have

$$A \text{ is invertible} \iff MA \text{ is invertible} \iff \text{Diagonal of } MA \text{ is all nonzero.}$$

But if all of the diagonal entries of MA are nonzero, then A has n pivots. (Recall that a pivot of A is a nonzero diagonal entry of the upper-triangular form of A.) We therefore concluded that A is invertible if and only if A has n pivots. You should now read the material under "Singular versus Invertible" on pp. 88–89.

Then we looked at an example from the exam. Let

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

(Finally, a new A.) If you do the elimination correctly, you find an invertible matrix M (what is M?) such that

$$MA = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 5 & -5 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then the pivots of A are 1 and 5, so A only has two, not three, pivots and therefore can't be invertible.

We can see that MA is not invertible due to that row of zeros. Indeed, for any $\mathbf{x}$, we have

$$MA \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 + 2x_3 \\ 5x_2 - 5x_3 \\ 0 \end{bmatrix}.$$

So if you are going to solve $MA\mathbf{x} = \mathbf{b}$, the third entry of $\mathbf{b}$ better be 0. In particular, the problem

$$MA\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

has no solution. Our work from the start of class then says that MA is not invertible.

Here is a good exercise for you: generalize the work above to show that if a matrix A has a row consisting entirely of 0's, then A is not invertible. If A has a column consisting of 0's, try to find a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. (Remember, $A\mathbf{x}$ is a linear combination of the columns of A weighted by entries of $\mathbf{x}$; make most of those entries 0.)

We can see the noninvertibility of A from the point of view of Gauss–Jordan elimination, too. If you try to do Gauss–Jordan elimination on A to find A^{-1}, you'll transform $\begin{bmatrix} A & I \end{bmatrix}$ into

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

There is no way to do "upwards" elimination starting from that $(3,3)$-entry of 0, and so Gauss–Jordan fails.

Day 19: Wednesday, February 22. We took a new perspective on our perennial problem $A\mathbf{x} = \mathbf{b}$: factoring. Recall that factoring in general reveals useful information (what is "useful" depends on what your problem is). Maybe you want prime factors: $24 = 2^3 \cdot 3$. Or maybe you want roots of a polynomial: $x^2 - 3x + 2 = (x-1)(x-3)$. Now we will factor matrices.

Our discussion today followed pp. 97–99, stopping at (and not including) "Better balance from LDU." Then we picked up with "One Square System = Two Triangular Systems" on pp. 100–101. There are many examples on these pages that substantially augment our work from class. Read also Worked Examples 2.6 A and B.

Recall from Days 8 and 15 that if

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

and

$$E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{31} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad E_{32} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/7 & 1 \end{bmatrix},$$

then

$$E_{32}E_{31}E_{21}A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -6 \\ 0 & 0 & -1/7 \end{bmatrix} =: U,$$

so U is upper-triangular. Put

$$M = E_{32}E_{31}E_{21},$$

so

$$MA = U.$$

Then M is the product of elimination matrices, which are invertible, and so M is invertible with

$$M^{-1} = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}.$$

Since $MA = U$, we have $A = M^{-1}U$. This is a factorization of A!

Here is why this is a *useful* factorization. On Day 13, we learned how to invert elimination matrices. We have

$$E_{21}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{31}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/7 & 1 \end{bmatrix}.$$

Side exercise for you: let E be the 3×3 elimination matrix that subtracts ℓ times row 1 from row 2. So

$$E \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 - \ell v_1 \\ v_3 \end{bmatrix}.$$

Morally, the inverse of E should be the 3×3 elimination matrix F that adds ℓ times row 1 to row 2, so we want

$$F \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 + \ell v_1 \\ v_3 \end{bmatrix}.$$

Check that $F(E\mathbf{v}) = \mathbf{v}$ for all vectors $\mathbf{v}$. Conclude that $FE = I$ using the more general principle that $A = B$ if and only if $A\mathbf{v} = B\mathbf{v}$ for all $\mathbf{v}$. Further exercise: prove this more general principle. More basic exercise: do you understand all the notation? Talk to me!

Anyway, if you grind it out,

$$M^{-1} = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/7 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1/7 & 1 \end{bmatrix} =: L.$$

And so we have factored

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & -1 & -3 \\ 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1/7 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -6 \\ 0 & 0 & -1/7 \end{bmatrix} = LU.$$

Look at the data that this factorization contains. The entries of L below its diagonal are the *negatives* of the multipliers that we used in elimination, and they end up in the correct entries of L. For example, the $(2,1)$-entry of L is 3, and the $(2,1)$-entry of E_{21} was -3. And the diagonal entries of U are the pivots of A. All along we've said that it's hard to see this data just from glancing at A, and yet there it is buried within this "LU-factorization."

Here is how this LU-factorization is useful. Our goal in life is to solve $A\mathbf{x} = \mathbf{b}$. Since $A = LU$, this is the same as solving $LU\mathbf{x} = \mathbf{b}$. Suppose that we have a solution $\mathbf{x}$. Then if we define $\mathbf{c} := U\mathbf{x}$, this vector $\mathbf{c}$ satisfies $L\mathbf{c} = \mathbf{b}$. So we see that if we can solve $LU\mathbf{x} = \mathbf{b}$, then we can also solve the pair of equations $U\mathbf{x} = \mathbf{c}$ and $L\mathbf{c} = \mathbf{b}$. Conversely, suppose that we have a solution $\mathbf{c}$ to $L\mathbf{c} = \mathbf{b}$ and a solution $\mathbf{x}$ to $U\mathbf{x} = \mathbf{c}$. Then $A\mathbf{x} = LU\mathbf{x} = L(U\mathbf{x}) = L\mathbf{c} = \mathbf{b}$. You should stare at this argument until it makes sense.

In summary, if $A = LU$, then

$$A\mathbf{x} = \mathbf{b} \iff LU\mathbf{x} = \mathbf{b} \iff \begin{cases} L\mathbf{c} = \mathbf{b} \\ U\mathbf{x} = \mathbf{c}. \end{cases}$$

This is true; why is it useful? We have turned one problem ($A\mathbf{x} = \mathbf{b}$) into two ($L\mathbf{c} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{c}$), and usually we'd prefer just one. ($1 < 2$.) But these two problems are triangular, and therefore "easy." We can solve both with back substitution.

Consider the problem

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & -1 & -3 \\ 2 & 3 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 10 \\ 4 \end{bmatrix}.$$

Using the LU-factorization of A above, this problem is equivalent to

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1/7 & 1 \end{bmatrix} \mathbf{c} = \begin{bmatrix} 1 \\ 10 \\ 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -6 \\ 0 & 0 & -1/7 \end{bmatrix} \mathbf{x} = \mathbf{c}.$$

We need $\mathbf{c}$ to solve the upper-triangular problem, so we do the lower-triangular one first. It's the system

$$\begin{cases} c_1 = 1 \\ 3c_1 + c_2 = 10 \\ 2c_1 + \dfrac{c_2}{7} + c_3 = 4. \end{cases}$$

Back-solving, we get $3 + c_2 = 10$, so $c_2 = 7$, and then $2 + 1 + c_3 = 4$, so $c_3 = 1$. Then the upper-triangular problem is

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -6 \\ 0 & 0 & -1/7 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 7 \\ 1 \end{bmatrix},$$

and as a system this reads

$$\begin{cases} x_1 + 2x_2 + x_3 = 1 \\ -7x_2 - 6x_3 = 7 \\ -\dfrac{x_3}{7} = 1. \end{cases}$$

Going up, we get $x_3 = -7$, then $-7x_2 + 42 = 7$, so $x_2 = 5$, and last $x_1 + 10 - 7 = -1$, thus $x_1 = -4$. So

$$\mathbf{x} = \begin{bmatrix} -2 \\ 5 \\ -7 \end{bmatrix}.$$

Day 20: Friday, February 24. We now have three methods for solving $A\mathbf{x} = \mathbf{b}$.

1. Form the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ and use elimination to convert this to $\begin{bmatrix} U & \mathbf{d} \end{bmatrix}$, where U is upper-triangular. Solve the system $U\mathbf{x} = \mathbf{d}$ via back-substitution (this is possible because U is upper-triangular). Be careful in that whatever you do to A to get U, you have to do to $\mathbf{b}$, which creates $\mathbf{d}$.

2. Calculate A^{-1} if A is invertible. Then $\mathbf{x} = A^{-1}\mathbf{b}$.

3. Obtain an LU-factorization of A and solve the two triangular problems $L\mathbf{c} = \mathbf{b}$ and then $U\mathbf{x} = \mathbf{c}$.

You should write out all of these methods on the same piece(s) of paper for the problem

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & -1 & -3 \\ 2 & 3 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 10 \\ 4 \end{bmatrix}.$$

We just did the LU-factorization on Day 19, but we never did elimination or an inverse for this particular $\mathbf{b}$.

The LU-factorization of A is always possible if we can reduce A to upper-triangular form using elimination matrices but not permutation matrices. Thus

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

does not have an LU-factorization since we need to use a permutation matrix to convert A into upper-triangular. (Exercise for you: what is that permutation matrix, and what is the resulting upper-triangular form?) The upshot is that when A has an LU-factorization, the diagonal of U contains the pivots of A (in order of use), and the entries of L below the diagonal are the multipliers from elimination: if you need to subtract ℓ times row j from row i in the elimination process, then $L(i,j) = \ell$. See p. 99.

So which method do we use, assuming that (1) A is invertible and (2) A has an LU-factorization, so that all three methods are available? By hand, everything is probably equally easy or difficult. In particular, elimination on the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ is practically the same work that will produce the LU-factorization. And Gauss–Jordan elimination is elimination.

Maybe the better question is what do we do on a *computer*. Think about all the arithmetic that we've done: it all boils down to addition, subtraction, multiplication, and division. If A is an invertible $n \times n$ matrix, it turns out that you need about n^3 arithmetic operations to compute A^{-1}. Then you need about n^2 operations to compute $\mathbf{x} = A^{-1}\mathbf{b}$. So solving $A\mathbf{x} = \mathbf{b}$ via the matrix inverse requires about

$$n^3 + n^2$$

operations. If you do elimination on A, it turns out that you need about $(n^3 - n)/3$ operations to convert A to upper-triangular U, and then about n^2 operations to solve $U\mathbf{x} = \mathbf{d}$. The point is that elimination requires about

$$\frac{n^3}{3} - \frac{n}{3} + n^2,$$

and

$$\frac{n^3}{3} - \frac{n}{3} + n^2 < n^3 + n^2.$$

So elimination could be more efficient.

Now imagine you are solving a system of systems of equations, like $A\mathbf{x} = \mathbf{b}_1, \ldots, A\mathbf{x} = \mathbf{b}_r$. You would need to do r multiplications by A^{-1}, or r rounds of elimination. But say you do the LU-factorization and solve the two systems $L\mathbf{c}_k = \mathbf{b}_k$ and $U\mathbf{x} = \mathbf{b}_k$ for $k = 1, \ldots, r$. You don't have to recalculate L and U each time; they don't change with k. So if you do the LU-factorization only once, you can use it as many times as you want!

See pp. 101 and 510. I don't expect you to memorize these operation counts, but I hope they give you some perspective as to why we are trying multiple approaches to the same problem.

Day 21: Monday, February 27. We added structure to our lives. The goal remains the same: to solve $A\mathbf{x} = \mathbf{b}$ or to understand our failure to do so. This discussion corresponds roughly to pp. 122–124, but with a little more set theory.

We need some basic terminology from sets to express new concepts clearly. A **set** is a collection of objects; we call these objects the **elements** of the set. You can, and should, argue that the word "collection" is not precisely defined—see how far you get with that before you circle back to the word "set."

If A is a set and x is an element of A, we write $x \in A$. If a set has only a few elements, we often list them between curly braces. So, we'll denote the set consisting of the integers 2, 4, 6, and 8 by $\{2,4,6,8\}$, and we'd say $2 \in \{2,4,6,8\}$. If y is not an element of A, we write $y \notin A$, so $5 \notin \{2,4,6,8\}$.

Often we form a set by selecting elements of another set with a certain property. For example, $\{2,4,6,8\}$ consists of all the even integers in the open interval $(0,10)$. We write

$$\{2,4,6,8\} = \{x \in (0,10) \mid x \text{ is an even integer}\}.$$

This is **set-builder notation**: we have a label for the elements (above they are x) and an indication of the "larger" set to which they belong (the open interval $(0,10)$, which is the set of all numbers between and not including 0 and 10) and then a description of the further property that those elements satisfy (the sentence "x is an even integer"). Sometimes we omit that larger set on the left side. Here's another example:

$$\{x \in (1,3) \mid x \text{ is an integer}\} = \{2\}.$$

In linear algebra, the most important sets will be sets of vectors. Remember that $\mathbb{R}$ is the set of all real numbers and $\mathbb{R}^n$ is the set of all column vectors with n entries. That is,

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \;\middle|\; x_1, \ldots, x_n \in \mathbb{R} \right\}.$$

We'd say

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \mathbb{R}^2 \quad \text{but} \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} \notin \mathbb{R}^3.$$

Likewise, we write $\mathbb{R}^{m\times n}$ for the set of all $m \times n$ matrices. Saying m and n allows for the possibility that $m \neq n$, and maybe our matrices are not square. Then

$$\mathbb{R}^{m\times n} = \left\{ \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \;\middle|\; a_{ij} \in \mathbb{R} \text{ for } i = 1, \ldots, m, j = 1, \ldots, n \right\}.$$

This is a lot of data to remember—-a matrix $A \in \mathbb{R}^{m\times n}$ has mn entries—so maybe the column perspective is more concise. (Eventually we might take the row perspective, too.) That is,

$$\mathbb{R}^{m\times n} = \left\{ \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \mid \mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m \right\}.$$

Make sure you agree that the columns of an $m \times n$ matrix are vectors in $\mathbb{R}^m$. Then, for example,

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}, \qquad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 3}, \quad \text{but} \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \notin \mathbb{R}^{2\times 2}.$$

While it is important and useful to be able to work concisely and clearly with set notation, always try to interpret your sets in words. For example, the set

$$\mathcal{V} := \left\{ \begin{bmatrix} c \\ 2c \end{bmatrix} \;\middle|\; c \in \mathbb{R} \right\}$$

is really the set of all scalar multiples of the vector $\begin{bmatrix}1\\2\end{bmatrix}$. Do you agree that $\begin{bmatrix}1\\5\end{bmatrix} \notin \mathcal{V}$?

We can use this notation to describe our main problem concisely. Let $A \in \mathbb{R}^{m \times n}$. We want to solve an equation of the form $A\mathbf{x} = \mathbf{b}$. Before we almost always worked with $m = n$; now we will not require that. For this equation to make sense, we need the matrix-vector product $A\mathbf{x}$ to be defined. Since A is $m \times n$, we need $\mathbf{x}$ to be $n \times 1$. Then the product $A\mathbf{x}$ will be $m \times 1$, so we need $\mathbf{b}$ to be $m \times 1$.

All together, here is our goal. Given $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, we want to find $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$. (We could have said all this in the first week of class, but I hope experience makes it easier for you to parse this sentence.)

The set $\mathbb{R}^{n \times 1}$ of $n \times 1$ matrices and the set $\mathbb{R}^n$ of column vectors with n entries are really the same, and so we will (rarely) bother to distinguish them. Technically the sets $\mathbb{R}^1 = \mathbb{R}^{1 \times 1}$ and $\mathbb{R}$ are different; after all, $0 \in \mathbb{R}$, but $\begin{bmatrix}0\end{bmatrix} \notin \mathbb{R}$. I think making such a distinction was useful earlier in the course, when we were thinking hard about how scalars, vectors, and matrices interacted. Now we will not make such a big deal about that distinction, but it will also arise only rarely (if ever) in practice.

Day 22: Wednesday, March 1. We defined the column space of $A \in \mathbb{R}^{m \times n}$. Strang denotes this by $\mathbf{C}(A)$, but I will usually write $\operatorname{col}(A)$. I say

$$\operatorname{col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\},$$

and every vector in $\operatorname{col}(A)$ is a vector in $\mathbb{R}^m$. The big question in solving $A\mathbf{x} = \mathbf{b}$ is if $\mathbf{b} \in \operatorname{col}(A)$. Maybe $\operatorname{col}(A) \neq \mathbb{R}^m$, in which case we'll fail to solve $A\mathbf{x} = \mathbf{b}$ sometimes. See pp. 126–127, up to but not including "Important." (That will be important next time.)

Motivated by some properties of the column space, which we proved, we defined the notion of subspace of $\mathbb{R}^n$. See pp. 124. We showed that the set

$$\mathcal{V} := \left\{ c \begin{bmatrix}1\\2\end{bmatrix} \;\middle|\; c \in \mathbb{R} \right\}$$

is a subspace of $\mathbb{R}^2$. Here's a generalization of that: let $\mathbf{v} \in \mathbb{R}^n$ be a particular vector. Then the set of scalar multiples of $\mathbf{v}$, which is

$$\{c\mathbf{v} \mid c \in \mathbb{R}\},$$

is a subspace of $\mathbb{R}^n$.

Now read pp. 124–125 on subspaces and stop at (do not read yet) Example 3. Subspaces of $\mathbb{R}^n$ are the most natural place to start, and we will spend most of our time there, but there is a further level of useful abstraction to consider: the vector space. Coming soon.

Day 23: Friday, March 3. We showed that the set

$$\mathcal{V} := \left\{ \begin{bmatrix}c\\2\end{bmatrix} \;\middle|\; c \in \mathbb{R} \right\}$$

is not a subspace of $\mathbb{R}^2$. All you have to do is break one of the subspace axioms, but this one breaks all three. (Caution: if $\mathbf{v} \in \mathcal{V}$, then $\alpha\mathbf{v} \in \mathcal{V}$ if $\alpha = 1$, so not quite all scalar multiples fail to belong to $\mathcal{V}$.)

Then we showed that

$$\mathcal{W} := \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is not a subspace of $\mathbb{R}^3$. Among other things, it lacks the zero vector.

More broadly, I claimed that if $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a *finite* set of vectors in $\mathbb{R}^n$, then $\mathcal{V}$ is never a subspace of $\mathbb{R}^n$, unless $\mathcal{V} = \{\mathbf{0}\}$. You should think about why this is true. Try checking scalar multiples and think about how many scalar multiples $c\mathbf{v}_1$ there are if $\mathbf{v}_1 \neq \mathbf{0}$.

This led us to ask how a finite subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ can be related to a subspace. Here is the new tool: the span. The span of $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ is the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$. Recall that a linear combination of $\mathbf{v}_1, \ldots \mathbf{v}_p$ is a vector of the form

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p,$$

where $c_1, \ldots, c_p \in \mathbb{R}$. I'll denote this set by $\operatorname{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_p\})$.

I claim that $\operatorname{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_p\})$ is always a subspace of $\mathbb{R}^n$. This is a little annoying to check for p in general, but think about it for $p = 2$ or $p = 3$, where you probably don't need to write "$\cdots$" all that much.

We checked that

$$\operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} \right) = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \;\middle|\; x_1, x_2 \in \mathbb{R} \right\} \neq \mathbb{R}^3.$$

Last, we rephrased the column space in terms of spans: for $A \in \mathbb{R}^{m \times n}$,

$$\begin{aligned} \operatorname{col}(A) &= \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} = \left\{ \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \;\middle|\; x_1, \ldots, x_n \in \mathbb{R} \right\} \\ &= \{x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \mid x_1, \ldots, x_n \in \mathbb{R}\} = \operatorname{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}). \end{aligned}$$

By now you can and should read all of Section 3.1.

Monday, March 13. We introduced the null space of $A \in \mathbb{R}^{m \times n}$, which we'll denote by $N(A)$ and maybe sometimes $\ker(A)$, as it's also called the kernel of A. We showed that

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

and I claimed more generally that if $A \in \mathbb{R}^{n \times n}$ is invertible, then $N(A) = \{\mathbf{0}\}$. You should prove this.

After some discussion and icky experience, we agreed to denote the zero vector in $\mathbb{R}^n$ by $\mathbf{0}_n$. For example,

$$\mathbf{0}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This notation is not universally accepted, so watch out. This allows us to write

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

Then we checked that $N(A)$ is a subspace of $\mathbb{R}^n$. This is an important result, but what's really important is our thought process in proving this. First, think about what you want. There are three things to check. With each thing, ask yourself what you already know and what you want to know. Also, ask yourself if you know what all the words mean.

1. Do we have $\mathbf{0}_n \in N(A)$? We want $A\mathbf{0}_n = \mathbf{0}_m$. Is this true? Grind it out:

$$A\mathbf{0}_n = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = 0\mathbf{a}_1 + \cdots + 0\mathbf{a}_n = \mathbf{0}_m + \cdots + \mathbf{0}_m = \mathbf{0}_m.$$

So, yes, $\mathbf{0}_n \in N(A)$.

2. If $\mathbf{x}_1$, $\mathbf{x}_2 \in N(A)$, do we have $\mathbf{x}_1 + \mathbf{x}_2 \in N(A)$? We know $A\mathbf{x}_1 = \mathbf{0}_m$ and $A\mathbf{x}_2 = \mathbf{0}_m$, and we want $A(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{0}_m$. Is this true? Let's compute

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{0}_m + \mathbf{0}_m = \mathbf{0}_m.$$

So, yes, $\mathbf{x}_1 + \mathbf{x}_2 \in N(A)$.

3. If $\mathbf{x} \in N(A)$ and $c \in \mathbb{R}$, do we have $c\mathbf{x} \in N(A)$? We know $A\mathbf{x} = \mathbf{0}_m$ and we want $A(c\mathbf{x}) = \mathbf{0}_m$. So we compute

$$A(c\mathbf{x}) = cA\mathbf{x} = c\mathbf{0}_m = \mathbf{0}_m.$$

So, yes, $N(A)$ is a subspace of $\mathbb{R}^n$.

We now have two "fundamental" subspaces associated with a matrix $A \in \mathbb{R}^{m\times n}$. There is the column space, $\text{col}(A)$, which is a subspace of $\mathbb{R}^m$. And now there is the null space, $N(A)$, which is a subspace of $\mathbb{R}^n$. Both of these spaces tell us things about our favorite problem $A\mathbf{x} = \mathbf{b}$. To be able to solve $A\mathbf{x} = \mathbf{b}$, we need $\mathbf{b} \in \text{col}(A)$. (You should be able to explain this by now.) So the column space tells us about *existence* of solutions: a solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ exists if $\mathbf{b} \in \text{col}(A)$.

It turns out that the null space tells us about *uniqueness* of solutions, and that is why we care about it. Say that $A\mathbf{x}_1 = \mathbf{b}$ and $A\mathbf{x}_2 = \mathbf{b}$ for some vectors $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathbb{R}^n$ with $\mathbf{x}_1 \neq \mathbf{x}_2$. (By the way, this means $\mathbf{b} \in \text{col}(A)$, right?) Calculate with me now:

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

So $\mathbf{x}_1 - \mathbf{x}_2 \in N(A)$. Moreover, since $\mathbf{x}_1 \neq \mathbf{x}_2$, we have $\mathbf{x}_1 - \mathbf{x}_2 \neq \mathbf{0}_n$, and so $N(A) \neq \{\mathbf{0}_n\}$.

Here is the slogan: if there are two different solutions to $A\mathbf{x} = \mathbf{b}$, then the null space of A contains more than just the zero vector. And here is why this matters: if the null space of A contains more than just the zero vector, then the problem $A\mathbf{x} = \mathbf{b}$ never has a unique solution. (Watch the tricky language: we're not assuming that this problem has any solution. Rather, we're saying that *if* the problem has one solution, then it doesn't *just* have one solution.)

Here's why. Say that $\mathbf{x} \in \mathbb{R}^n$ with $A\mathbf{x} = \mathbf{b}$ and $\mathbf{v} \in N(A)$ with $\mathbf{v} \neq \mathbf{0}_n$. Let $c \in \mathbb{R}$. Then compute

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

So $\mathbf{x} + c\mathbf{v}$ is another solution to the problem! Moreover, since $\mathbf{v} \neq \mathbf{0}_n$, each different value of $c \in \mathbb{R}$ gives a new solution $\mathbf{x} + c\mathbf{v}$. Not only is the solution not unique, there are infinitely many solutions!

Summing up, we can use the column space to tell us if we can solve a problem, and we can use the null space to tell us if the solution to that problem is unique. You should read pp. 134–135 up to, but not including, the box beginning "The two key steps of this section…"

Day 25: Wednesday, March 15. If you haven't read pp. 134–135, check them out now. We computed several concrete null spaces today, and these were more variations on Examples 1, 2, and 3 in Section 3.2. The general strategy that I hinted at appears on pp. 136–137, and we will do that next time.

Here were the concrete examples. We first showed

$$N\left(\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}\right) = \operatorname{span}\left(\left\{\begin{bmatrix} -3/2 \\ 1 \end{bmatrix}\right\}\right)$$

and then

$$N\left(\begin{bmatrix} 2 & 3 & 4 \\ 4 & 6 & 8 \end{bmatrix}\right) \operatorname{span}\left(\left\{\begin{bmatrix} -3/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}\right\}\right)$$

and finally

$$N\left(\begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix}\right) = \operatorname{span}\left(\left\{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\right\}\right).$$

Each of these calculations had some elements in common. To determine $N(A)$, we need to find all solutions to $A\mathbf{x} = \mathbf{0}$. One way to do this is to perform elimination on the augmented matrix $\begin{bmatrix} A & \mathbf{0} \end{bmatrix}$ and convert that augmented matrix to $\begin{bmatrix} U & \mathbf{0} \end{bmatrix}$, where U is upper-triangular. Here's an exercise for you: say that M is the product of all the elimination and maybe permutation matrices that you need to make A upper-triangular, i.e., $MA = U$. Check that $M\begin{bmatrix} A & \mathbf{0} \end{bmatrix} = \begin{bmatrix} U & \mathbf{0} \end{bmatrix}$.

The point is that we can just reduce A to upper-triangular form and then solve $U\mathbf{x} = \mathbf{0}$. For example,

$$\begin{bmatrix} 2 & 3 & 4 \\ 4 & 6 & 8 \end{bmatrix} \xrightarrow[E_{21}]{(2)-2\cdot(1)} \begin{bmatrix} 2 & 3 & 4 \\ 0 & 0 & 0 \end{bmatrix}.$$

Once in upper-triangular form, we studied the resulting system of equations:

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

We don't have nearly enough information to determine exactly the values of x_1, x_2, and x_3, but we can say

$$x_1 = -\frac{3}{2}x_2 - 2x_3.$$

For definiteness in our algorithmic process, we'll solve for the "earlier" variable in terms of the "later" variables. Thus any solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{0}$ has the form

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

Look at the upper-triangular form

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

There is one pivot: 2. This pivot appears in the first column. The second and third columns do not have pivots. It is not a coincidence that we were able to solve for x_1 (which, in the matrix-vector multiplication $U\mathbf{x}$, hits the column with a pivot) in terms of the variables x_2 and x_3 (which do not hit the pivots).

For the third null space, we took our sweet time and did a little more work that was basically Gauss–Jordan elimination. We didn't stop with an upper-triangular matrix; we "eliminated upwards," too. Here was the sequence of operations:

$$\begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix} \xrightarrow[E_{21}]{(2)-2\cdot(1)} \begin{bmatrix} 2 & 3 & 4 \\ 0 & 3 & 6 \end{bmatrix} \xrightarrow[D_2]{(1/3)\cdot(2)} \begin{bmatrix} 2 & 3 & 4 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow[E_{12}]{(1)-3\cdot(2)} \begin{bmatrix} 2 & 0 & -2 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow[D_1]{(1/2)\cdot(1)} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}.$$

If we put

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix}, \qquad R = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}, \qquad \text{and} \qquad M = D_1E_{12}D_2E_{21},$$

then $M \in \mathbb{R}^{2\times 2}$ (check that), M is invertible, and $R = MA$. I claim that since M is invertible, $N(A) = N(MA)$; you should check that. Thus $N(A) = N(R)$, and that's how we got the last null space above.

Each of the four matrices above represents a different stage in the evolution of A to a more useful, tractable form. We'll discuss those stages next time. The final stage, R, is the reduced row echelon form of A, which appears on pp. 136–138.

Day 26: Friday, March 17. We revisited the five related matrices from the end of Day 25. Start with

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix}.$$

We want to understand $N(A)$. Solving $A\mathbf{x} = \mathbf{0}_2$ is the same as solving $U\mathbf{x} = \mathbf{0}_2$, where

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

This is because we obtained U from A via elimination.

Note that U is upper-triangular. I don't think we ever gave a rigorous definition of upper-triangular, so let's do so now. A matrix $U \in \mathbb{R}^{m \times n}$ is upper-triangular if $U(i,j) = 0$ when $i > j$. Informally, if you're further down the matrix than across, you're 0. In U above, we have $U(2,1) = 0$, which we need because $2 > 1$, but we have $U(2,2) \neq 0$ and $U(2,3) \neq 0$, and both are okay because $2 = 2$ and $2 < 3$.

Solving $U\mathbf{x} = \mathbf{0}_2$ is a bit easier than solving $A\mathbf{x} = \mathbf{0}_2$, but we can do better. Going slightly out of order compared to Day 25, let's rescale and define

$$\widetilde{R} = \begin{bmatrix} 1 & 3/2 & 2 \\ 0 & 1 & 2 \end{bmatrix}.$$

Then $U\mathbf{x} = \mathbf{0}_2$ if and only if $\widetilde{R}\mathbf{x} = \mathbf{0}_2$. The matrix $\widetilde{R}$ is a bit nicer than U because not only is $\widetilde{R}$ upper-triangular, both pivots of $\widetilde{R}$ are 1.

A matrix like $\widetilde{R}$ has a name. A matrix in $\mathbb{R}^{m \times n}$ is in **row echelon form** if

(1) Rows with all zero entries are below any row that has at least one nonzero entry. (The matrix doesn't have to have a row of all zeros, but if it does, that row is at the bottom.)

(2) The first nonzero entry in each row is 1. (And so all pivots are 1.)

(3) If $i < j$, the first nonzero entry in row i is in a column strictly to the left of the column containing the first nonzero entry in row j. (If one row is below another row, the row below has to "start" its nonzero entries to the right of the row above.)

A glance at $\widetilde{R}$ above shows that it satisfies these three properties. Note that the first doesn't really count here, because $\widetilde{R}$ has no row of zeros.

We can always use elimination (and maybe permutation) and scaling to put a matrix in row echelon form. Row echelon form helps us count pivots and compute null spaces, because the data in row echelon form is arguably going to be simpler than the data of our original matrix.

We can go one step further from row echelon form. A matrix in $\mathbb{R}^{m \times n}$ is in **reduced row echelon form** if it is already in row echelon form (i.e., if (1), (2), and (3) above are satisfied) and if also

(4) If a column contains a pivot (which by (2) necessarily has the value of 1), then all of the other entries in that column are 0.

Look at the second column in $\widetilde{R}$. That entry of $3/2$ prevents $\widetilde{R}$ from being in reduced row echelon form. But we can "eliminate upwards" (i.e., perform Gauss–Jordan elimination) to convert $\widetilde{R}$ to

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

This is definitely in reduced row echelon form.

Reduced row echelon form cleanly and clearly reveals how many pivots a matrix has and makes it very easy to compute the null space of that matrix. Once we are in row echelon form (which we get via elimination/permutation and scaling), use Gauss–Jordan elimination ("upward" elimination) to arrive at reduced row echelon form.

Here's another example. Start with

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

Do elimination on the first column of A and convert A to the matrix

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

This matrix is not upper-triangular because its $(3, 2)$-entry is 1, not 0. So do elimination in the second column to get

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

Check that this matrix is in row echelon form; note that we do have a row of all zero entries and (rather naturally) it fell out at the bottom. Last, do Gauss-Jordan elimination to get rid of that 2; then the reduced row echelon form of A is

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

Exercise for you: find a matrix M such that $R = MA$. This M needs to be invertible and square—so $M \in \mathbb{R}^{n\times n}$ with $n =$ what?

Now we have $A\mathbf{x} = \mathbf{0}_3$ if and only if $R\mathbf{x} = \mathbf{0}_3$. (Note that $\mathbf{x} \in \mathbb{R}^2$, right?) And

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

So the only solution is $\mathbf{x} = \mathbf{0}_2$; that is, $N(A) = \{\mathbf{0}_2\}$. Look how easy it was to compute the null space once we had the reduced row echelon form.

Contrast the matrices

$$R_1 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \quad \text{and} \quad R_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Both matrices have two pivots, but one has the trivial null space $N(R_2) = \{\mathbf{0}_2\}$, while R_1 has a more complicated null space (which we previously found on Day 25).

Counting pivots relative to the dimensions of our matrix is going to be key to controlling null spaces (and column spaces)—and these spaces control our ability to solve $A\mathbf{x} = \mathbf{b}$, which is the point of life, more or less. You should now read about the reduced row echelon form on pp. 136–137.

Day 27: Monday, March 20. A good way to see if you understand the definition of the RREF is to build some matrices in RREF. Say that we want a matrix $R \in \mathbb{R}^{3\times4}$ with pivots in columns 1 and 3 only. So the matrix looks like

$$R = \left[\begin{array}{c|ccc} ? & * & * & * \\ ? & * & * & * \\ ? & * & * & * \end{array}\right].$$

Let's focus on that first column, which is why I singled it out. We want a pivot in that column, so one of the entries needs to be 1, and the rest are 0. Here are the three possibilities:

$$\left[\begin{array}{c|ccc} 1 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{array}\right] \quad \text{or} \quad \left[\begin{array}{c|ccc} 0 & * & * & * \\ 1 & * & * & * \\ 0 & * & * & * \end{array}\right] \quad \text{or} \quad \left[\begin{array}{c|ccc} 0 & * & * & * \\ 0 & * & * & * \\ 1 & * & * & * \end{array}\right]. \tag{1}$$

The first one looks fine to me, but I have problems with the second and third.

We know the first row of the second matrix in (1) can't be all 0, so maybe the second matrix really is

$$\left[\begin{array}{c|ccc} 0 & a & * & * \\ 1 & * & * & * \\ 0 & * & * & * \end{array}\right]$$

with $a \neq 0$. But then the first nonzero entry in row 1 is not strictly to the left of the first nonzero entry in row 2. The same problem happens if the second matrix in (1) has the form

$$\left[\begin{array}{c|ccc} 0 & 0 & a & * \\ 1 & * & * & * \\ 0 & * & * & * \end{array}\right] \quad \text{or} \quad \left[\begin{array}{c|ccc} 0 & 0 & 0 & a \\ 1 & * & * & * \\ 0 & * & * & * \end{array}\right]$$

with $a \neq 0$. And I claim the same problem happens with the third matrix in (1). So only the first matrix in (1) works as an RREF so far.

Now let's turn our attention to the second column in

$$\left[\begin{array}{c|c|cc} 1 & ? & * & * \\ 0 & ? & * & * \\ 0 & ? & * & * \end{array}\right].$$

The $(2,2)$- and $(3,2)$-entries have to be 0; if they're nonzero, then they're leading nonzero entries in their rows, and thus pivots. But this matrix has no pivots in column 2. So the matrix has the form

$$\left[\begin{array}{c|c|cc} 1 & ? & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{array}\right].$$

If we look carefully at the rules for the RREF, we see no restrictions on the $(1,2)$-entry. Making it nonzero will not create a pivot in the second column; making it zero will not create a row of zeros in row 1. So we can say that the matrix has the form

$$\left[\begin{array}{c|c|cc} 1 & a & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{array}\right]$$

for an arbitrary $a \in \mathbb{R}$.

Now we focus on the third column in

$$\left[\begin{array}{cc|c|c} 1 & a & ? & * \\ 0 & 0 & ? & * \\ 0 & 0 & ? & * \end{array}\right].$$

It has a pivot, so precisely one entry is 1 and the other two are 0. Here are the three possibilities:

$$\left[\begin{array}{cc|c|c} 1 & a & 1 & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \end{array}\right] \quad \text{or} \quad \left[\begin{array}{cc|c|c} 1 & a & 0 & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & * \end{array}\right] \quad \text{or} \quad \left[\begin{array}{cc|c|c} 1 & a & 0 & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 1 & * \end{array}\right].$$

The first matrix won't work as there is no leading 1 in any row of column 3; thus column 3 does not contain a pivot. The second matrix definitely works, as it has a leading 1 in row 2, thus a pivot there. The third matrix does have a leading 1 in row 3, thus a pivot there, but since column 4 can't contain a pivot, this third matrix would have to be

$$\left[\begin{array}{cc|c|c} 1 & a & 0 & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & * \end{array}\right].$$

Otherwise, if that $(2,4)$-entry is nonzero, then there is a leading 1 and thus a pivot in column 4.

So, our matrix now has the form

$$\left[\begin{array}{ccc|c} 1 & a & 0 & ? \\ 0 & 0 & 1 & ? \\ 0 & 0 & 0 & ? \end{array}\right].$$

The RREF rules don't restrict the $(1,4)$- and $(2,4)$-entries, as they will never be pivots in a matrix of this form. We do have to set the $(3,4)$-entry to be 0, as if it's nonzero, then it's a pivot. Thus *any* matrix $R \in \mathbb{R}^{3\times 4}$ that (1) is in RREF and (2) has pivots in columns 1 and 3 only has the form

$$R = \begin{bmatrix} 1 & a & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

for some a, b, $c \in \mathbb{R}$.

Here's a game to play with yourself or a friend. (It's not a fun game, but there are much worse games to play.) Pick dimensions for your matrix—say you want $R \in \mathbb{R}^{m\times n}$ with m and n given. Pick some columns between 1 and n to have pivots. How many matrices of this form can you make? What if you pick some rows between 1 and m to have pivots?

Remember that the point of RREF is that it's "easy" to get the null space from RREF. Let's do another example of that. Say that $A \in \mathbb{R}^{3\times 4}$ has the RREF above with the a, b, and c entries arbitrary. Then $A\mathbf{x} = \mathbf{0}_3$ if and only if $R\mathbf{x} = \mathbf{0}_3$, and

$$\begin{bmatrix} 1 & a & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \iff \begin{cases} x_1 + ax_2 + bx_4 = 0 \\ x_3 + cx_4 = 0 \\ 0 = 0. \end{cases}$$

We can therefore solve for x_1 and x_3 in terms of x_2 and x_4 (...dare I say that x_1 and x_3 are *functions* of x_2 and x_4?):

$$x_1 = -ax_2 - bx_4 \quad \text{and} \quad x_3 = -cx_4.$$

Then the solution $\mathbf{x}$ to $R\mathbf{x} = \mathbf{0}_3$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -ax_2 - bx_4 \\ x_2 \\ -cx_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} -ax_2 \\ x_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -bx_4 \\ 0 \\ -cx_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -a \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -b \\ 0 \\ -c \\ 1 \end{bmatrix}.$$

Thus

$$N(A) = N(R) = \text{span}\left(\left\{ \begin{bmatrix} -a \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -b \\ 0 \\ -c \\ 1 \end{bmatrix} \right\}\right).$$

Strang is calling the vectors

$$\begin{bmatrix} -a \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -b \\ 0 \\ -c \\ 1 \end{bmatrix}$$

the "special solutions" to $R\mathbf{x} = \mathbf{0}_3$; they are "special" because any other solution to $R\mathbf{x} = \mathbf{0}_3$ is a linear combination of them.

Then we introduced the terminology of free variables and pivot variables, and free columns and pivot columns. If $A \in \mathbb{R}^{3\times 4}$ has the RREF given by R, then the free variables are x_2 and x_4, the pivot variables are x_1 and x_3, the free columns are columns 2 and 4, and the pivot columns are columns 1 and 3. By the way, if $R \in \mathbb{R}^{m\times n}$ is the RREf of $A \in \mathbb{R}^{m\times n}$, then I'll sometimes write $R = \text{rref}(A)$. This presumes that A has only one RREF, which probably needs to be checked.

Counting pivots tells us how many free variables a matrix has, and that tells us something about the "size" of the matrix's null space. (We will be more precise with "size" later.) If there are no free variables, then the null space is trivial, because then the only solution to $R\mathbf{x} = \mathbf{0}_m$ is $\mathbf{x} = \mathbf{0}_n$. Otherwise, if there are free variables, then the null space contains nonzero vectors, and that destroys uniqueness of solutions to our favorite problem $A\mathbf{x} = \mathbf{b}$.

There is one situation in which a matrix *always* has free variables. If $A \in \mathbb{R}^{m\times n}$, then A has m rows; since there can be at most one pivot per row, A has at most m pivots. Solving any problem $A\mathbf{x} = \mathbf{b}$ requires n variables in $\mathbf{x}$. If $m < n$, then A cannot have as many pivots as variables in the problem, and so there will be $n - m > 0$ free variables. This shows that if $A \in \mathbb{R}^{m\times n}$ with $m < n$, then there are always nonzero vectors in $N(A)$, and so $A\mathbf{x} = \mathbf{b}$ cannot have a unique solution. (Maybe it has no solution!) Think of such a matrix as "short and wide"; think of this situation as having more unknowns than it does equations.

You can now read up to p. 138, stopping at "The rank of a matrix." Look at Worked Examples 3.2 A and B.

Day 28: Wednesday, March 22. We defined the **rank** of A to be the number of pivots in $\text{rref}(A)$; we write it as $\text{rank}(A)$. We argued that if $A \in \mathbb{R}^{m\times n}$, then $0 \le \text{rank}(A) \le m$. You should work through Worked Example 3.2 C right now.

We explored the peculiar case of a rank 1 matrix: its RREF has precisely one pivot. Say that $A \in \mathbb{R}^{3\times 4}$ has rank 1. Then

$$\text{rref}(A) = \begin{bmatrix} 1 & r_2 & r_3 & r_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where $r_2, r_3, r_4 \in \mathbb{R}$. A good exercise for you is to explain why $\text{rref}(A)$ has exactly this form. (What goes wrong if the 1 goes anywhere else?) We also know that there is an invertible matrix $E \in \mathbb{R}^{3\times 3}$ such that $EA = R$; the matrix E encodes all the elimination, permutation, and scaling needed to transform A into R. See Day 24 for an example of how such an E arises in practice. Then we have $A = E^{-1}R$; put $F = E^{-1}$, so this reads $A = FR$, and write $A = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 & \mathbf{f}_3 \end{bmatrix}$.

I claim that

$$A = FR = \begin{bmatrix} \mathbf{f}_1 & r_2\mathbf{f}_1 & r_3\mathbf{f}_1 & r_4\mathbf{f}_1 \end{bmatrix}.$$

Another good exercise for you is to do this matrix multiplication (multiply F against the columns of R). So, all the columns of A are just scalar multiples of the first column! In fact, check further that

$$A = \mathbf{f}_1 \begin{bmatrix} 1 & r_2 & r_3 & r_4 \end{bmatrix}.$$

That is, A is the product of a 3×1 matrix and a 1×4 matrix. See p. 139 for more examples. Can you generalize this line of argument to show that if $A \in \mathbb{R}^{m\times n}$ has rank 1, then $A = UV$ for some $U \in \mathbb{R}^{m\times 1}$ and $V \in \mathbb{R}^{1\times n}$. And if you don't have enough work to do, check that

$$A := \begin{bmatrix} 2 & 3 & 4 \\ 4 & 6 & 8 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & 3/2 & 2 \end{bmatrix}.$$

In the context of the work above, how does the vector $\begin{bmatrix}2\\4\end{bmatrix}$ relate to an elimination matrix that operates on A? (Yes, I know this vector is the first column of A, but it also shows up in elimination, or the inverse thereof.)

We finally stopped talking about null spaces exclusively and worked on a concrete version of the major problem $A\mathbf{x} = \mathbf{b}$. Let

$$A = \begin{bmatrix}2 & 3 & 4\\4 & 6 & 8\end{bmatrix} \qquad \text{and} \qquad \mathbf{b} = \begin{bmatrix}1\\2\end{bmatrix}.$$

A good way to approach $A\mathbf{x} = \mathbf{b}$ is to consider the augmented matrix $\begin{bmatrix}A & \mathbf{b}\end{bmatrix}$ and convert A to $R = \operatorname{rref}(A)$, doing the same operations to $\mathbf{b}$ that you do to A. Then you get $\begin{bmatrix}R & \mathbf{d}\end{bmatrix} = E\begin{bmatrix}A & \mathbf{b}\end{bmatrix}$ for some invertible matrix E.

Specifically, here's what happens:

$$\left[\begin{array}{ccc|c}2 & 3 & 4 & 1\\4 & 6 & 8 & 2\end{array}\right] \xrightarrow[E_{21}]{(2)-2\cdot(1)} \left[\begin{array}{ccc|c}2 & 3 & 4 & 1\\0 & 0 & 0 & 0\end{array}\right] \xrightarrow[D_1]{(1/2)\cdot(1)} \left[\begin{array}{ccc|c}1 & 3/2 & 2 & 1/2\\0 & 0 & 0 & 0\end{array}\right].$$

(So what's E?)

The matrix

$$\left[\begin{array}{ccc|c}1 & 3/2 & 2 & 1/2\\0 & 0 & 0 & 0\end{array}\right]$$

represents the system

$$\begin{cases}x_1 + (3/2)x_2 + 2x_3 = 1/2\\0 = 0.\end{cases}$$

We solve for x_1 as

$$x_1 = \frac{1}{2} - \frac{3x_2}{2} - \frac{2x_3}{3},$$

and so the solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ has the form

$$\begin{aligned}\mathbf{x} = \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}1/2 - (3/2)x_2 - (2/3)x_3\\x_2\\x_3\end{bmatrix} &= \begin{bmatrix}1/2\\0\\0\end{bmatrix} + \begin{bmatrix}(-3/2)x_2\\x_2\\0\end{bmatrix} + \begin{bmatrix}-(2/3)x_3\\0\\x_3\end{bmatrix}\\ &= \begin{bmatrix}1/2\\0\\0\end{bmatrix} + x_2\begin{bmatrix}(-3/2)\\1\\0\end{bmatrix} + x_3\begin{bmatrix}-(2/3)\\0\\1\end{bmatrix}.\end{aligned}$$

This problem has infinitely many solutions because we can take x_2 and x_3 to be any real numbers that we like. We can condense our notation by setting

$$\mathbf{x}_{\mathsf{p}} = \begin{bmatrix}1/2\\0\\0\end{bmatrix}, \qquad \mathbf{s}_1 = \begin{bmatrix}(-3/2)\\1\\0\end{bmatrix}, \qquad \text{and} \qquad \mathbf{s}_2 = \begin{bmatrix}-(2/3)\\0\\1\end{bmatrix}.$$

All our work above then says that any solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ has the form

$$\mathbf{x} = \mathbf{x}_{\mathsf{p}} + x_2\mathbf{s}_1 + x_3\mathbf{s}_2 \tag{2}$$

for some x_2, $x_3 \in \mathbb{R}$. Note that if we take $x_2 = x_3 = 0$, then we just get $\mathbf{x} = \mathbf{x}_\mathsf{p}$, and so $\mathbf{x}_\mathsf{p}$ is a "particular" solution to the problem.

Here is another aspect of our solution's structure. Back on Day 25 we showed that

$$N(A) = \text{span}(\{\mathbf{s}_1, \mathbf{s}_2\}).$$

Now we see those vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ show up in the solution to the problem $A\mathbf{x} = \mathbf{b}$, and we see precisely how $\mathbf{s}_1$ and $\mathbf{s}_2$ create infinitely many solutions to this problem. This is what we have said all along: the null space destroys uniqueness of solutions.

Here is a third aspect of our solution's structure. The formula (2) tells us how any two solutions to $A\mathbf{x} = \mathbf{b}$ are related: just add some vector from the null space. Specifically, and generally, let $A \in \mathbb{R}^{m \times n}$ and let $\mathbf{b} \in \mathbb{R}^m$. Suppose that $A\mathbf{x}_\mathsf{p} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{b}$. I'll let you check that $A(\mathbf{x} - \mathbf{x}_\mathsf{p}) = \mathbf{0}_m$. Thus $\mathbf{x} - \mathbf{x}_\mathsf{p} \in N(A)$. If we put $\mathbf{x}_0 := \mathbf{x} - \mathbf{x}_\mathsf{p}$, then you can also check that $\mathbf{x} = \mathbf{x}_\mathsf{p} + \mathbf{x}_0$. And so if we know one particular solution $\mathbf{x}_\mathsf{p}$, we just have to add an appropriate vector $\mathbf{x}_0$ in $N(A)$ to get any other solution $\mathbf{x}$.

This is a big deal. Our work above says that if just get one solution to the problem $A\mathbf{x} = \mathbf{b}$ and if we know all about the null space of A, then we know what every solution to $A\mathbf{x} = \mathbf{b}$ is.

You can now read Worked Example 3.2 C and then pp. 149–151 up to (but not including) Example 1. Skip p. 148 for now, but we will come back to it.

Day 29: Friday, March 24. We did two more illustrative examples of solving, or failing to solve, $A\mathbf{x} = \mathbf{b}$ via transformation to the RREF. Consider

$$\begin{bmatrix} 2 & 3 & 4 \\ 4 & 6 & 8 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Here is the conversion:

$$\left[\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 4 & 6 & 8 & 1 \end{array}\right] \xrightarrow[E_{21}]{\mathsf{(2)-2\cdot(1)}} \left[\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 0 & 0 & 0 & -1 \end{array}\right] \xrightarrow[D_1]{\mathsf{(1/2)\cdot(1)}} \left[\begin{array}{ccc|c} 1 & 3/2 & 2 & 1/2 \\ 0 & 0 & 0 & -1 \end{array}\right].$$

The last matrix above encodes the system

$$\begin{cases} x_1 + \dfrac{3x_2}{2} + 2x_3 = \dfrac{1}{2} \\ 0 = -1. \end{cases}$$

The equation $0 = -1$ is false, and so this system has no solutions.

Here is the broader lesson. Let $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and convert the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to $\begin{bmatrix} R & \mathbf{d} \end{bmatrix}$, where $R = \text{rref}(A)$. Then $R = EA$ for some invertible $E \in \mathbb{R}^{m \times m}$, and so $\mathbf{d} = E\mathbf{b}$. Then $\mathbf{x} \in \mathbb{R}^n$ satisfies $A\mathbf{x} = \mathbf{b}$ if and only if $R\mathbf{x} = \mathbf{d}$. If row i of R is all 0, then the ith entry of $R\mathbf{x}$ will be 0, and so the ith entry of $\mathbf{d}$ will be 0, too. This is a solvability condition for our problem: if we can solve $A\mathbf{x} = \mathbf{b}$ and if the ith row of $R = EA$ is all 0, then the ith entry of $\mathbf{d} = E\mathbf{b}$ is 0, too. And so if the ith row of R is all 0 and the ith entry of $\mathbf{d}$ is not 0, then we cannot solve the problem $A\mathbf{x} = \mathbf{b}$!

Let's do another example: let

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix}.$$

We work on the augmented matrix:

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 1 & 3 & 4 \\ 2 & 7 & 9 \end{array}\right] \xrightarrow[E_{21}]{(2)-(1)} \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 2 & 7 & 9 \end{array}\right] \xrightarrow[E_{31}]{(3)-2\cdot(1)} \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 3 & 3 \end{array}\right] \xrightarrow[E_{32}]{(3)-3\cdot(2)} \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right] \xrightarrow[E_{12}]{(1)-2\cdot(2)} \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right].$$

Here we have converted $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to $\begin{bmatrix} R & \mathbf{d} \end{bmatrix}$, where

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{d} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Again, $R = \operatorname{rref}(A)$ has a row of all 0 entries, but the corresponding row in $\mathbf{d}$ is 0. The problem $R\mathbf{x} = \mathbf{d}$ simply reads

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

Not only do we have a solution $\mathbf{x}$, we have only one solution. Observe that R has two pivots, but A only has two columns; thus both x_1 and x_2 are pivot variables of A, both columns 1 and 2 are pivot columns of A, and there are no free variables or free columns. In particular, $N(A) = \{\mathbf{0}_2\}$, so if $A\mathbf{x} = \mathbf{b}$ has a solution, then necessarily that solution is unique. Note that we got information about $N(A)$ even though we didn't set out to solve $A\mathbf{x} = \mathbf{0}_2$ here; we just read this information off from the RREF.

This is the situation of **full column rank**. Say that $A \in \mathbb{R}^{m\times n}$ and $\operatorname{rank}(A) = n$. (This presumes $m \geq n$, right? For we know $\operatorname{rank}(A) \leq m$ in general, so here in particular $n = \operatorname{rank}(A) \leq m$.) Then A has no free variables, since there are only n variables in play anyway, and all of them are pivot variables. Thus $N(A) = \{\mathbf{0}_n\}$, and so if $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$, then the only solution is that $\mathbf{x}$. No guarantees that a solution exists—in the last example above, try finding $\mathbf{b} \in \mathbb{R}^3$ such that

$$E_{12}E_{32}E_{31}E_{21}\mathbf{b} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

with $d_3 \neq 0$.

Please read Example 1 on p. 151 now and the very important box on p. 152.

Day 30: Monday, March 27. If $A \in \mathbb{R}^{m\times n}$ has full column rank, then $\operatorname{rank}(A) = n$. You might expect, then, that if A has **full row rank**, then $\operatorname{rank}(A) = m$. Full column rank guarantees that the null space is trivial. What good stuff does full column rank give us?

Let's look at an example: we have calculated that the RREF of

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix}$$

is

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

We see then that $\operatorname{rank}(A) = 2$, x_1 and x_2 are pivot variables, and x_3 is the only free variable.

I claim this means that we can always solve $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^2$. Let's do the particular example of $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. I'll show the elimination steps in full detail just to help us review:

$$\left[\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 4 & 9 & 14 & 1 \end{array}\right] \xrightarrow[E_{21}]{(2)-2\cdot(1)} \left[\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 0 & 3 & 6 & -1 \end{array}\right] \xrightarrow[D_2]{(1/3)\cdot(2)} \left[\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 0 & 1 & 2 & -1/3 \end{array}\right]$$
$$\xrightarrow[E_{12}]{(1)-3\cdot(2)} \left[\begin{array}{ccc|c} 2 & 0 & -2 & 2 \\ 0 & 1 & 2 & -1/3 \end{array}\right] \xrightarrow[D_1]{(1/2)\cdot(1)} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & -1/3 \end{array}\right].$$

Thus $A\mathbf{x} = \mathbf{b}$ if and only if

$$\begin{cases} x_1 - x_3 = 1 \\ x_2 + 2x_3 = -\dfrac{1}{3} \end{cases},$$

and so

$$x_1 = 1 + x_3 \quad \text{and} \quad x_2 = -\frac{1}{3} - 2x_3.$$

Thus all solutions $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ have the form

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

This example illustrates a general phenomenon: if $A \in \mathbb{R}^{m\times n}$ has rank m (full row rank), then we can always solve the problem $A\mathbf{x} = \mathbf{b}$, but, as we already know, the solution will not be unique if $m < n$, as then there will be free variables. Here is how you can argue that $A\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b} \in \mathbb{R}^m$; we'll keep working with the previous example. Do your elimination and convert the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to $\begin{bmatrix} R & \mathbf{d} \end{bmatrix}$. In the context of the work above, we get

$$\begin{bmatrix} R & \mathbf{d} \end{bmatrix} = \left[\begin{array}{ccc|c} 1 & 0 & -1 & d_1 \\ 0 & 1 & 2 & d_2 \end{array}\right].$$

Do you see the identity matrix there? We really have

$$\begin{bmatrix} R & \mathbf{d} \end{bmatrix} = \left[\begin{array}{cc|c} I & \mathbf{r}_3 & \mathbf{d} \end{array}\right],$$

where I is the 2×2 identity matrix. Then $R\mathbf{x} = \mathbf{d}$ if and only if

$$I \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + x_3\mathbf{r}_3 = \mathbf{d}.$$

We can force this equality to be true by taking

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{d} \quad \text{and} \quad x_3 = 0.$$

You should now read pp. 152–154. Pay careful attention to the four cases on p. 154—do you understand them all? Can you give concrete examples of A and $\mathbf{b}$ in each case? We've probably done all of them in class. Worked Examples 3.3 A, B, and C contain tons of useful details!

Day 31: Wednesday March 29. Everything that we do in this class is in service to understanding the problem $A\mathbf{x} = \mathbf{b}$, where $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. We know that to solve this problem we need $\mathbf{b} \in \text{col}(A)$ = column space = the set of all linear combinations of the columns of A. We know that to guarantee that the solution is unique (if it exists in the first place), we need $N(A) = \{\mathbf{0}_n\}$, where $N(A)$ = null space = $\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}_m\}$ and $\mathbf{0}_n$ is the zero vector in $\mathbb{R}^n$. Life teaches us that we should want solutions (1) to exist and (2) to be unique. So it's probably reasonable to say that we want the column space of A to be as large as possible and the null space of A to be as small as possible. That is, we want col(A) to contain as much of $\mathbb{R}^m$ as possible and $N(A)$ to contain as little of $\mathbb{R}^n$ as possible.

Our next task will be to quantify "as much" and "as little" in meaningful ways. A natural but fruitless idea is to ask how many vectors col(A) and $N(A)$ contain. This certainly would give a notion of "size" for these two key subspaces. But here's the problem: any subspace that is not $\{\mathbf{0}_n\}$ in fact contains infinitely many vectors. That is, if $\mathcal{V}$ is a subspace of $\mathbb{R}^n$ and there is $\mathbf{v} \in \mathcal{V}$ such that $\mathbf{v} \neq \mathbf{0}_n$, then $\mathcal{V}$ contains infinitely many elements.

Here's why. Let $\mathbf{v} \in \mathcal{V}$ be nonzero. Then $c_1\mathbf{v} \neq c_2\mathbf{v}$ for any $c_1, c_2 \in \mathbb{R}$. (You should figure this out on your own. Think about the vector $(c_1 - c_2)\mathbf{v}$: is it the zero vector?) Moreover, $c\mathbf{v} \in \mathcal{V}$ for each $c \in \mathbb{R}$ by the definition of subspace.

So, as soon as a subspace contains at least one vector different from the zero vector, that subset contains infinitely many vectors (all the scalar multiples of that nonzero vector), and therefore asking "how many vectors are in the column space or the null space?" won't give us really meaningful data about our favorite problem $A\mathbf{x} = \mathbf{b}$. Instead, we need to do more work and get some new ideas.

Let's think about the column space. We've spent a large amount of time computing null spaces from reduced row echelon forms, but we never really computed a column space. But we didn't really need to—a "formula" for the column space is built into its definition as the "span of all the columns of the matrix."

Here's an example. Let

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

We've worked with this matrix several times before. Its column space is

$$\text{col}(A) = \text{span}\left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \begin{bmatrix} 4 \\ 8 \end{bmatrix}\right) = \left\{ x_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ 8 \end{bmatrix} \;\middle|\; x_1,\ x_2,\ x_3 \in \mathbb{R} \right\}.$$

It looks like the column space is defined by three vectors (there are, after all, only three columns in A), but some of these vectors are redundant. Let's do a little algebra:

$$x_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ 8 \end{bmatrix} = 2x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 3x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 4x_3 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = (2x_1 + 3x_2 + 4x_3) \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Every vector in $\operatorname{col}(A)$ is therefore just a scalar multiple of the one vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. And so

$$\operatorname{col}(A) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right).$$

Consequently, there is a certain "redundancy" in writing the column space as the span of three vectors when really it is the span of one vector. Our real job going forward will be to eliminate that redundancy in describing subspaces—why say more when you could say less, and do less work in the process?

Let's look at the null space of this A. Here it is useful to appeal to the reduced row echelon form of A, which in the past we've calculated to be

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

This told us, after some work, that

$$N(A) = \operatorname{span}\left(\begin{bmatrix} -3/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}\right).$$

Cautioned by our experience with $\operatorname{col}(A)$, we might ask if $N(A)$ can be written as the span of fewer vectors. The only thing fewer than two is one, so could we have $N(A) = \operatorname{span}(\mathbf{v})$ for a single $\mathbf{v} \in \mathbb{R}^3$?

If so, then

$$\begin{bmatrix} -3/2 \\ 1 \\ 0 \end{bmatrix} = c_1\mathbf{v} \quad \text{and} \quad \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} = c_2\mathbf{v}$$

for some c_1, $c_2 \in \mathbb{R}$. Both c_1 and c_2 are nonzero (think about what happens otherwise), and so we can do a very clever rewriting:

$$\begin{bmatrix} -3/2 \\ 1 \\ 0 \end{bmatrix} = c_1\mathbf{v} = \left(\frac{c_2}{c_2}\right) c_1\mathbf{v} = \frac{c_1}{c_2}(c_2\mathbf{v}) = \frac{c_1}{c_2} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}.$$

Look at the second components of the vectors on the far left and the far right. The second component on the left is 1 and the right is 0. So $1 = 0$, which is impossible. We have therefore shown that $N(A)$ cannot be the span of a single vector. Thus, in the description of $N(A)$ above as a span, there is no redundancy, and we really need to have those two vectors "generating" the span.

By the way, you can and should check that another way of writing col(A) is

$$\text{col}(A) = \text{span}\left(\begin{bmatrix}2\\4\end{bmatrix}\right).$$

When we look at the reduced row echelon form, we see that $\begin{bmatrix}2\\4\end{bmatrix}$ is the only pivot column of A. We will see this pattern again...right now!

Here's another example. Let

$$A = \begin{bmatrix}2 & 3 & 4\\4 & 9 & 14\end{bmatrix}.$$

We've worked with this matrix a lot, too; you know this beastie already. Its RREF is

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

Previously we calculated

$$N(A) = \text{span}\left(\begin{bmatrix}1\\-2\\1\end{bmatrix}\right),$$

but now let's focus on the column space. It doesn't look like any two columns are scalar multiples of a third (unlike our previous example), so let's dig a little deeper.

The key idea is to relate the RREF back to A. Remember that $R = EA$ for some invertible $E \in \mathbb{R}^{2\times 2}$, where E "encodes" all the work that we did to convert A to R (elimination and scaling). Then $A = E^{-1}R$; abbreviate $F = E^{-1}$ to get $A = FR$. Say that $F = \begin{bmatrix}\mathbf{f}_1 & \mathbf{f}_2\end{bmatrix}$. It's not important right now exactly what the columns of F are.

Let's do the multiplication FR by multiplying F against the columns of R. Going column by column on R, we have

$$\begin{bmatrix}\mathbf{f}_1 & \mathbf{f}_2\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = \mathbf{f}_1, \qquad \begin{bmatrix}\mathbf{f}_1 & \mathbf{f}_2\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \mathbf{f}_2, \qquad \text{and} \qquad \begin{bmatrix}\mathbf{f}_1 & \mathbf{f}_2\end{bmatrix}\begin{bmatrix}-1\\2\end{bmatrix} = -\mathbf{f}_1 + 2\mathbf{f}_2.$$

Remember that $A = FR$, and so

$$\begin{bmatrix}2\\4\end{bmatrix} = \mathbf{f}_1, \qquad \begin{bmatrix}3\\9\end{bmatrix} = \mathbf{f}_2, \qquad \text{and} \qquad \begin{bmatrix}4\\14\end{bmatrix} = -\mathbf{f}_1 + 2\mathbf{f}_2.$$

The third equality is the key. It says that the third column of A is a linear combination of the first two columns. You can check in your spare time that

$$-\begin{bmatrix}2\\4\end{bmatrix} + 2\begin{bmatrix}3\\9\end{bmatrix} = \begin{bmatrix}4\\14\end{bmatrix}. \tag{3}$$

I claim this means that the column space is really

$$\text{span}\left(\begin{bmatrix}2\\4\end{bmatrix}, \begin{bmatrix}3\\9\end{bmatrix}\right).$$

Actually, I'm going to leave that for you to check: use the identity (3) to rewrite any linear combination

$$x_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 9 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ 14 \end{bmatrix}$$

in the form

$$y_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + y_2 \begin{bmatrix} 3 \\ 9 \end{bmatrix}.$$

And so the column space of A is not just a linear combinations of the columns of A—which includes some redundant data—but a linear combination of its *pivot* columns. You do see that

$$\begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 \\ 9 \end{bmatrix}$$

are the pivot columns of A, right?

We have calculated two column spaces and two null spaces today. All four spaces were the spans of certain sets of vectors—some redundant, some not. By removing the redundant vectors from consideration, we obtained simpler descriptions of the subspaces and maybe a better sense of their "size"—all nonzero subspaces are infinite, but maybe we can describe them as spans of as few vectors as possible. Our next task will be to make mathematically meaningful this intuitive feeling of "redundancy," and that will hinge on understanding the situation with the columns of our last matrix: when some vectors in the span are already linear combinations of the others.

Day 32: Friday March 31. You took an exam. What fun!

Day 33: Monday, April 3. Today's class was brought to you not by the Department of Mathematics but by the Department of Redundancy Department. On Day 31, you saw that the matrix

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$$

had the relationship

$$-\mathbf{a}_1 + 2\mathbf{a}_2 = \mathbf{a}_3$$

among its columns. (You can check that right now if you need to.) This relationship wasn't obvious, but it was there. It means that although $\text{col}(A) = \text{span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ by definition of the column space, really we just need two vectors in the span: $\text{col}(A) = \text{span}(\mathbf{a}_1, \mathbf{a}_2)$.

We are working so much with subspaces given by spans that we could use some jargon: a set $\mathcal{S}$ in $\mathbb{R}^m$ spans a subspace $\mathcal{V}$ of $\mathbb{R}^m$ if $\mathcal{V} = \text{span}(\mathcal{S})$, where $\text{span}(\mathcal{S})$ is, as always, the set of all linear combinations of vectors in $\mathcal{S}$. We might also say that $\mathcal{V}$ is spanned by $\mathcal{S}$, or that $\mathcal{S}$ is a spanning set for $\mathcal{V}$. For example, the columns of $A \in \mathbb{R}^{m \times n}$ form a spanning set for the subspace $\text{col}(A)$ of $\mathbb{R}^m$.

Back to the A with the redundant columns. You're going to have to trust me on this, but it turns out that the most useful way to view the redundancy is not that $\mathbf{a}_3$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$ (although it is) but rather that the following identity holds:

$$-\mathbf{a}_1 + 2\mathbf{a}_2 - \mathbf{a}_3 = \mathbf{0}_2.$$

I'll rewrite this a little to make the coefficients in the linear combination more apparent:

$$(-1)\mathbf{a}_1 + 2\mathbf{a}_2 + (-1)\mathbf{a}_3 = \mathbf{0}_2.$$

We are going to make this phenomenon both more abstract and more precise. I promise that, in the end, our new tools will help us answer the question "What is the best spanning set for a subspace?" or maybe "What is the least redundant spanning set?", as "best" might be subjective.

The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ are **linearly independent** if the only linear combination $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ of these vectors that adds up to $\mathbf{0}_m$ is the combination with $x_1 = 0, \ldots, x_n = 0$. Here is the "test for linear independence": assume that $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0}_m$ and show that $x_1 = 0, \ldots, x_n = 0$. We also have **linear dependence**: when the list of vectors is not linearly independent (big help). Here is the "test for linear dependence": find numbers $x_1, \ldots, x_n$ not all zero such that $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0}_m$.

We did several examples and had a spirited discussion. I will just list the examples but omit the work.

1. The vectors $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ are linearly independent.

2. The vectors $\begin{bmatrix}1\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\end{bmatrix}$, and $\begin{bmatrix}2\\3\end{bmatrix}$ are linearly dependent.

3. The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ are linearly independent if and only if the null space of the matrix $\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$ is the trivial null space $\{\mathbf{0}_n\}$. In the context of the example above, how does looking at the RREF and the null space help?

4. Any list of vectors containing the zero vector (in $\mathbb{R}^m$) is linearly dependent.

5. A single vector (in the definition above, take $n = 1$ and replace "are" with "is") is linearly independent if and only if that vector is not the zero vector.

You can now read pp. 163–166. Look at Worked Example 3.4 B.

Day 34: Wednesday, April 5. We did two more examples of linear (in)dependence.

6. A list of vectors in $\mathbb{R}^m$ that contains some repeated vectors is linearly dependent. For simplicity, suppose the list is $\mathbf{v}$, $\mathbf{v}$, $\mathbf{v}_3$, $\mathbf{v}_4$. The last two vectors can be anything, but the first two are definitely the same.

7. A list of vectors is linearly dependent if and only if one vector in the list is a linear combination of the others. This crystallizes how linear independence prevents redundancy in a spanning set: the set is LI if and only if no vector in the set is a linear combination of the others.

We then went further into how properties of matrices reveal linear (in)dependence of columns. We know that $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ is linearly independent if and only if $N(\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}) =$

$\{\mathbf{0}_n\}$. This happens if and only if the matrix $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$ has no free variables, which happens if and only if A has n pivot variables, which happens if and only if $\operatorname{rank}(A) = n$. Since $\operatorname{rank}(A) \le m$, if $\operatorname{rank}(A) = n$ as well, then $n \le m$. So there cannot be more columns than rows.

Here's the translation: any list of n vectors in $\mathbb{R}^m$ is linearly dependent if $n > m$. The columns of a matrix form a linearly dependent list if there are more columns than rows. Just by counting rows vs. columns, then, you can see that the columns of

$$\begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix}$$

are linearly dependent. But you have to do more work to figure out which column is a linear combination of the others.

The last thing we discussed, briefly, was the concept of basis. The basis is the resolution of the "Goldilocks" problem for spanning sets. A set may be too small to span a subspace—then that set is not a basis. The vectors in a set may not be linearly independent and may therefore be linear combinations of each other—that set is not a basis. The concept of basis prevents redundancies. See p. 167 for the definition. We proved that the vectors

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

form a basis for $\mathbb{R}^2$. (Note that $\mathbb{R}^m$ is a subspace of itself.)

You can now read pp. 166–168 (stop at and don't read Example 7 for now). Ignore row spaces for now; we'll do them later.

Day 34: Friday, April 7. Generalizing our last example, the columns of the $m \times m$ identity matrix form a basis for $\mathbb{R}^m$ for any m. We call this basis the standard basis for $\mathbb{R}^m$.

But a subspace can have plenty of bases. Read Worked Example 3.4 C right now. It's pretty amazing.

Here's another basis for $\mathbb{R}^2$:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

You can check that the rank of the matrix

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

is 2. So, this matrix has full column rank, and its null space is trivial. Therefore its columns are linearly independent. And this matrix has full column rank, and so we can always solve $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^2$. Thus $\operatorname{col}(A) = \mathbb{R}^2$, which is to say that every vector in $\mathbb{R}^2$ is in the span of the columns of A. So the two vectors above span $\mathbb{R}^2$.

There are *lots* of ways to check that a given list of vectors forms a basis. Unless told otherwise, you should probably do what feels right and easiest to you.

Without doing any real work, we then said that

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

can't form a basis for $\mathbb{R}^2$: there are too many vectors. Two rows, three columns, no way.

Then we did a proof by "$\cdots$." A basis is a spanning set—this gives us control over the subspace in question. And a basis is linearly independent—this prevents redundant vectors and gives us efficiency. We can get something even better.

Say that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis for the subspace $\mathcal{V}$ of $\mathbb{R}^m$. Then $\mathcal{V} = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, so if $\mathbf{v} \in \mathcal{V}$, then there are scalars $x_1, \ldots, x_n \in \mathbb{R}$ such that $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$. We can think of these scalars as "coordinates" of $\mathbf{v}$ with respect to the basis: they tell us how to assemble the vectors in the basis to form $\mathbf{v}$. What's great is that these coordinates are unique: there's only one way to pick them. Say that

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n.$$

Subtract to get

$$(x_1\mathbf{v}_1 - y_1\mathbf{v}_1) + \cdots + (x_n\mathbf{v}_n - y_n\mathbf{v}_n) = \mathbf{0}_m.$$

Factor to get

$$(x_1 - y_1)\mathbf{v}_1 + \cdots + (x_n - y_n)\mathbf{v}_n = \mathbf{0}_m.$$

Use linear independence to get

$$x_1 - y_1 = 0, \ldots, x_n - y_n = 0.$$

Add to get

$$x_1 = y_1, \ldots, y_1 = y_n.$$

Call it a day.

One last little thing: all this time we've been saying things like "Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis for $\mathcal{V}$." How do we know that a subspace even has a basis? In particular, how do we construct bases for really important subspaces like column spaces and null spaces?

We'll do the construction next week; as you can imagine, it boils down to the RREF. Here's an existence proof. I say that every subspace of $\mathbb{R}^m$ other than $\{\mathbf{0}_m\}$ has a basis. (Exercise for you: why can't $\{\mathbf{0}_m\}$ have a basis?)

For simplicity, let's put $m = 4$. Let $\mathcal{V}$ be a subspace of $\mathbb{R}^4$. Suppose $\mathcal{V} \neq \{\mathbf{0}_4\}$. Then there is a vector $\mathbf{v}_1 \in \mathcal{V}$ such that $\mathbf{v}_1 \neq \mathbf{0}_m$. Either $\mathcal{V} = \text{span}(\mathbf{v}_1)$ or $\mathcal{V} \neq \text{span}(\mathbf{v}_1)$. If $\mathcal{V} = \text{span}(\mathbf{v}_1)$, then the vector $\mathbf{v}_1$ by itself is a basis for $\mathcal{V}$: certainly $\mathbf{v}_1$ spans $\mathcal{V}$, and $\mathbf{v}_1$ is linearly independent because it's nonzero.

Now suppose $\mathcal{V} \neq \text{span}(\mathbf{v}_1)$. Then there is a vector $\mathbf{v}_2 \in \mathcal{V}$ such that $\mathbf{v}_2 \notin \text{span}(\mathbf{v}_1)$. I claim that $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. To prove this, suppose $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 = \mathbf{0}_m$. We need to show $x_1 = x_2 = 0$. Suppose $x_2 \neq 0$. Then we can get $\mathbf{v}_2 = -(x_1/x_2)\mathbf{v}_1$, and so $\mathbf{v}_2 \in \text{span}(\mathbf{v}_1)$. So $x_2 = 0$. Then $x_1\mathbf{v}_1 = 0$, so $x_1 = 0$ because $\mathbf{v}_1 \neq \mathbf{0}_4$. So we have $x_1 = x_2 = 0$.

Now we either have $\mathcal{V} = \text{span}(\mathbf{v}_1, \mathbf{v}_2)$ or $\mathcal{V} \neq \text{span}(\mathbf{v}_1, \mathbf{v}_2)$. If $\mathcal{V} = \text{span}(\mathbf{v}_1, \mathbf{v}_2)$, then $\mathbf{v}_1$ and $\mathbf{v}_2$ form a basis for $\mathcal{V}$ because they're linearly independent. If $\mathcal{V} \neq \text{span}(\mathbf{v}_1, \mathbf{v}_2)$, then there is $\mathbf{v}_3 \in \mathcal{V}$ such that $\mathbf{v}_3 \notin \text{span}(\mathbf{v}_1, \mathbf{v}_2)$. I claim that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent—can you figure that out as we did in the previous paragraph? So either $\mathcal{V} = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, in which case $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ form a basis for $\mathcal{V}$, or $\mathcal{V} \neq \text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. What do you think happens then?

Day 36: Monday, April 10. I hope you said that there is $\mathbf{v}_4 \in \mathcal{V}$ such that $\mathbf{v}_4 \notin \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. You can then argue that $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{v}_4$ are linearly independent. Now something special happens: we're working in $\mathbb{R}^4$, and we have four linearly independent vectors. I claim then that $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{v}_4$ therefore form a basis for $\mathbb{R}^4$. That is, $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4) = \mathbb{R}^4$. Since this span is already contained in $\mathcal{V}$, and since $\mathcal{V}$ is contained in $\mathbb{R}^4$, we set-theoretically "squeeze" $\mathcal{V}$ so that it equals $\mathbb{R}^4$. That is, the only, and last, possibility for $\mathcal{V}$ is that $\mathcal{V} = \mathbb{R}^4$.

Here's the justification of my claim. All we have to do is show that $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{v}_4$ span $\mathbb{R}^4$, since they are already linearly independent. Put the four vectors into a matrix: $A = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4\end{bmatrix}$. We want to show $\operatorname{col}(A) = \mathbb{R}^4$. We know $N(A) = \{\mathbf{0}_4\}$, so $\operatorname{rank}(A) = 4$, and so $\operatorname{rref}(A)$ has four pivots. But $\operatorname{rref}(A)$ is a 4×4 matrix, so $R = \operatorname{rref}(A) = I$, where I is the 4×4 identity matrix. So $A = ER = E$, where E is an invertible matrix, and therefore A is invertible, and therefore $\operatorname{col}(A) = \mathbb{R}^4$, right? (You want to be able to solve $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^4 \ldots$)

Can you generalize this argument to show that any list of m linearly independent vectors in $\mathbb{R}^m$ is a basis for $\mathbb{R}^m$?

We now know (1) what a basis is, (2) some methods of checking that a given list is (or is not) a basis for a given subspace, and (3) that every subspace of $\mathbb{R}^m$ has a basis. We really should figure out how to actually *find* a basis for a given subspace from scratch, but there's one more general concept that I want to share with you first. In the argument above, when we showed that any subspace of $\mathbb{R}^m$ (okay, $\mathbb{R}^4$), has a basis, there was an aspect of *counting*—either the basis had one vector, or it had two, or it had... m (well, 4). We might wonder if two bases for the same subspace can have different numbers of vectors. This would be bad! We are striving for efficiency with bases, and if my basis has more vectors than yours, I'd say that mine is less efficient than yours.

Here is the fact: every basis for a subspace of $\mathbb{R}^m$ has the same number of vectors. There is a great proof on pp. 169–170 of the book, which I will adapt here. Say that $\mathcal{V}$ is a subspace of $\mathbb{R}^m$ and we have two bases for $\mathcal{V}$: the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_p$. All of the vectors in each basis are distinct (because repetition destroys linear independence), so the first basis has n vectors and the second has p. We want to show that $n = p$.

What goes wrong if $n \neq p$? Then one has to be bigger than the other. To make things really concrete in class, let's say $n = 2$ and $p = 3$, and let's keep m arbitrary, so we're still working in $\mathbb{R}^m$. (You can always read the full argument in the book later.) Since $\mathbf{v}_1$ and $\mathbf{v}_2$ form a basis for $\mathcal{V}$, and since $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3 \in \mathcal{V}$, we can write each $\mathbf{w}_k$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$\mathbf{w}_1 = x_1\mathbf{v}_1 + x_2\mathbf{v}_2, \qquad \mathbf{w}_2 = y_1\mathbf{v}_1 + y_2\mathbf{v}_2, \qquad \text{and} \qquad \mathbf{w}_3 = z_1\mathbf{v}_1 + z_2\mathbf{v}_2.$$

This looks like matrix-vector multiplication to me:

$$\mathbf{w}_1 = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}x_1 \\ x_2\end{bmatrix} \qquad \mathbf{w}_2 = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}y_1 \\ y_2\end{bmatrix} \qquad \text{and} \qquad \mathbf{w}_3 = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}z_1 \\ z_2\end{bmatrix}.$$

And this looks like matrix-matrix multiplication now:

$$\begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2\end{bmatrix}.$$

Call the product above $C = AB$. Let's count sizes. So $C \in \mathbb{R}^{m\times 3}$, $A \in \mathbb{R}^{m\times 2}$, and $B \in \mathbb{R}^{2\times 3}$. There are more columns than rows in B, so $N(B) \neq \{\mathbf{0}_2\}$. That is, there is $\mathbf{x} \in \mathbb{R}^2$ such that $\mathbf{x} \neq \mathbf{0}_2$ and $B\mathbf{x} = \mathbf{0}_3$. Then $AB\mathbf{x} = \mathbf{0}_3$, too. And so $C\mathbf{x} = \mathbf{0}_3$, so $N(C) \neq \mathbf{0}_3$. But then the vectors $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$ are linearly dependent. Boom, contradiction.

So, now every subspace of $\mathbb{R}^m$ has a basis, and every basis has the same number of vectors—this number is the **dimension** of the subspace. Since the m columns of the $m \times m$ identity matrix are a basis for $\mathbb{R}^m$, the dimension of $\mathbb{R}^m$ is m—I mean, what else could it be? Let's start finding bases!

I say we begin with the column space, as this controls our ability to solve $A\mathbf{x} = \mathbf{b}$. Look at our old friend

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix}.$$

By definition, the column space is

$$\text{col}(A) = \text{span}\left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 9 \end{bmatrix}, \begin{bmatrix} 4 \\ 14 \end{bmatrix}\right),$$

but you also know that

$$\begin{bmatrix} 4 \\ 14 \end{bmatrix} = -\begin{bmatrix} 2 \\ 4 \end{bmatrix} + 2\begin{bmatrix} 3 \\ 9 \end{bmatrix},$$

and so really

$$\text{col}(A) = \text{span}\left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 9 \end{bmatrix}\right).$$

I'll leave it to you to check that these two vectors are linearly independent, and so they form a basis for $\text{col}(A)$, and the dimension of $\text{col}(A)$ is 2.

You figured this out via brute force (the best force), but now we can be more elegant. As usual, the answer lies in the RREF. Here it is

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

The pivot columns are columns 1 and 2, and these are the columns of A that formed the basis for $\text{col}(A)$. And the rank of A is 2, which is the dimension of the column space.

This sounds to me like a conjecture: for $A \in \mathbb{R}^{m\times n}$, the pivot columns of A form a basis for the column space of A, and the dimension of the column space is the rank of A. If this is true, then we have a nice algorithm for controlling the column space: get the RREF, figure out the pivot columns of A (careful—A and the RREF have different columns), and count pivots.

You should now read pp. 168–170 and Worked Examples 3.4 A and 3.4 C.

Day 37: Wednesday, April 12. So, tell me about the column space of

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

Let's get the RREF:

$$\begin{bmatrix}2 & 4\\4 & 8\end{bmatrix} \xrightarrow[E_{21}]{(2)-2\cdot(1)} \begin{bmatrix}2 & 4\\0 & 0\end{bmatrix} \xrightarrow[D_1]{(1/2)\cdot(1)} \begin{bmatrix}1 & 2\\0 & 0\end{bmatrix}.$$

So column 1 is the pivot column, and the rank is 1, so the dimension of the column space is 1. In particular, a basis for the column space is the single vector $\begin{bmatrix}2\\4\end{bmatrix}$, and

$$\text{col}(A) = \text{span}\left(\begin{bmatrix}2\\4\end{bmatrix}\right) = \left\{x\begin{bmatrix}2\\4\end{bmatrix} \;\middle|\; x \in \mathbb{R}\right\}.$$

Note that we cannot use the pivot column of R here! If you believe the conjecture, then the column space of R is spanned by its only pivot column:

$$\text{col}(R) = \text{span}\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \left\{x\begin{bmatrix}1\\0\end{bmatrix} \;\middle|\; x \in \mathbb{R}\right\}.$$

Then it's definitely the case that $\text{col}(A) \neq \text{col}(R)$. For example,

$$\begin{bmatrix}6\\12\end{bmatrix} \in \text{col}(A), \qquad \begin{bmatrix}6\\12\end{bmatrix} \notin \text{col}(R), \qquad \begin{bmatrix}6\\0\end{bmatrix} \in \text{col}(R), \qquad \text{and} \qquad \begin{bmatrix}6\\0\end{bmatrix} \notin \text{col}(A).$$

We use the RREF to understand the column space of A, but the column space of the RREF is (usually) not the column space of A.

Back to the claim: why do the pivot columns of A form a basis for its column space? Start with $A \in \mathbb{R}^{m\times n}$, let $R = \text{rref}(A)$, and let $E \in \mathbb{R}^{m\times m}$ be an invertible matrix such that $R = EA$. Put $F = E^{-1}$, so $A = FR$.

The pivot columns of R are some (maybe all?) of the columns in the $m \times m$ identity matrix. They have m rows, a 1 in precisely one entry, and a 0 in every other entry. The pivot columns of R are therefore linearly independent.

Moreover, the nonpivot columns of R have nonzero entries only in rows in which there is a pivot to the left. Otherwise, if an entry in a nonpivot column is the only nonzero entry in its row in R, then that column would be a pivot column. So a nonpivot columns of R is a linear combinations of the pivot columns—specifically, of the pivot columns to the left of that nonpivot column. This is a "proof in words" (all proofs are in words, but this is wordier than usual); you could do a proof in more precise notation, but I worry translating things precisely could obscure their meaning.

Here's a picture:

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

Columns 1 and 3 are pivot columns, column 2 is 2 times column 1, and column 4 is 3 times column 1 plus 4 times column 3. The nonpivot columns are linear combinations of the pivot columns. Note that the nonpivot column 2 is only a linear combination of the pivot column 1, i.e., the pivot column to the left of column 2.

Let's label the pivot columns as $\mathbf{p}_1, \ldots, \mathbf{p}_r$, where $r = \text{rank}(A)$. Then the pivot columns of A are $F\mathbf{p}_1, \ldots, F\mathbf{p}_r$. Worked Example 3.4 C in the book says that $F\mathbf{p}_1, \ldots, F\mathbf{p}_r$ are linearly independent. That is, the pivot columns of A are linearly independent. Now let $\mathbf{a}$ be any nonpivot column of A. Then $\mathbf{a} = F\mathbf{q}$, where $\mathbf{q}$ is a nonpivot column of R. So $\mathbf{q} = x_1\mathbf{p}_1 + \cdots + x_r\mathbf{p}_r$ for some $x_1, \ldots, x_r \in \mathbb{R}$. Then $\mathbf{a} = x_1F\mathbf{p}_1 + \cdots + x_rF\mathbf{p}_r$. So every nonpivot column of A is a linear combination of the pivot columns of A.

By definition, the columns of A form a spanning set for $\operatorname{col}(A)$. Let's remove the nonpivot columns from this spanning set. Then we are just left with the pivot columns. Since those nonpivot columns are linear combinations of the pivot columns, the span of the pivot columns equals the span of all the columns. So, the span of the pivot columns equals the column space of A. And the pivot columns are linearly independent. So, the pivot columns form a basis for the column space. And there are as many pivot columns as the rank of A. So, the dimension of the column space is the rank of A.

If we look back at the same matrix that we studied on Monday to motivate our conjecture, we had

$$\operatorname{rref}\left(\begin{bmatrix} 2 & 3 & 4 \\ 4 & 9 & 14 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix},$$

so the first and second columns of the original matrix are its pivot columns and therefore form a basis for its column space. That is,

$$\begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 \\ 9 \end{bmatrix}$$

form a basis for the column space. (Incidentally, the column space is $\mathbb{R}^2$, right?)

What about the null space of $A \in \mathbb{R}^{m\times n}$? We have seen that each free variable kicks another vector into a spanning set for $N(A)$, so maybe we should conjecture that the dimension of the null space is the number of free variables, which is $n - \operatorname{rank}(A)$. Say that $A \in \mathbb{R}^{3\times 4}$ has the RREF

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

We have

$$A\mathbf{x} = \mathbf{0}_3 \iff R\mathbf{x} = \mathbf{0}_3 \iff \begin{cases} x_1 + 2x_2 + 3x_4 = 0 \\ x_3 + 4x_4 = 0 \\ 0 = 0 \end{cases} \iff \begin{cases} x_1 = -2x_2 - 3x_4 \\ x_3 = -4x_4 \end{cases}$$

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

We definitely have, then

$$N(A) = \operatorname{span}\left(\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}\right).$$

Are these vectors actually a basis for $N(A)$, and is the dimension of $N(A)$ equal to 2?

The key observation is that these two vectors have a special arrangement of the numbers 0 and 1. Each vector has a 0 in a row where the other has a 1 and a 1 in a row where the other has a 0. I've illustrated this in blue for you:

$$\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}$$

So suppose

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

(I'm using c_1 and c_2 so as not to overwork x from above.) Then, just focusing on the blue rows, we must have

$$\begin{cases} (c_1 \cdot 1) + (c_2 \cdot 0) = 0 \\ (c_1 \cdot 0) + (c_2 \cdot 1) = 0 \end{cases} \Longrightarrow c_1 = 0 \text{ and } c_2 = 0.$$

The vectors are therefore linearly independent.

Now look again at the arrangement of 0 and 1 in these vectors. The vector

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

was weighted by the x_2 free variable, and this vector has a 1 in row 2 and a 0 in row 4. The vector

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

was weighted by the x_4 free variable, and this vector has a 1 in row 4 and a 0 in row 2. *This pattern is not an accident.*

If you haven't done so, now is a good time to read Example 9 on p. 169 (keep omitting the row space—we'll come back to that). Then read #2 on p. 181 and #2 on pp. 181–182.

Day 38: Friday, April 14. We are going to justify the conjecture that the dimension of $N(A)$, for $A \in \mathbb{R}^{m \times n}$, is $n - \text{rank}(A)$ = the number of free variables. Start with $A \in \mathbb{R}^{m \times n}$ and convert A to $R = \text{rref}(A)$. Then $A\mathbf{x} = \mathbf{0}_m$ if and only if $R\mathbf{x} = \mathbf{0}_m$. We solve $R\mathbf{x} = \mathbf{0}_m$ by working row-by-row and writing each pivot variable as a function of the free variables. However, we can't solve for a free variable in terms of any other variables, so we just leave the free variable as a function of itself. For example, from last time,

$$\begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \iff \begin{cases} x_1 = -2x_2 - 3x_3 \\ x_2 = x_2 \\ x_3 = -4x_4 \\ x_4 = x_4. \end{cases}$$

I emphasized the "free variables as functions of themselves" in blue here, and I hope you agree that those equations are silly and redundant, which is why we never wrote them before. But now we can rewrite each row of the solution $\mathbf{x}$ as a linear combination of the free variables—and if x_i is a free variable, then row i of $\mathbf{x}$ just contains x_i:

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

We can then rewrite $\mathbf{x}$ as a linear combination of as many vectors (in $\mathbb{R}^n$, by the way) as there are free variables; each vector is multiplied by one free variable. The vector multiplied by the free variable x_i will have a 1 in row i and a 0 in any other row corresponding to a different free variable.

Now we need a bit more precise notation (I'm sorry—I just can't say this clearly using "only words"!). Say that the free variables are $x_{k_1}, \ldots, x_{k_d}$ with $d = n - \text{rank}(A)$. In the example above, $d = 2$, $k_1 = 2$, and $k_2 = 4$. Say that the vectors are $\mathbf{v}_1, \ldots, \mathbf{v}_d$. That is, if $\mathbf{x} \in N(A)$, then $\mathbf{x} = x_{k_1}\mathbf{v}_1 + \cdots + x_{k_d}\mathbf{v}_d$. The entries in rows $k_1, \ldots, k_d$ of $\mathbf{v}_i$ are all 0 except for the entry in row k_i, which is 1. Then the entry in row k_i of any linear combination $c_1\mathbf{v}_1 + \cdots + c_d\mathbf{v}_d$ is c_i. So if $c_1\mathbf{v}_1 + \cdots + c_d\mathbf{v}_d = \mathbf{0}_n$, we look at rows $k_1, \ldots, k_d$ on both sides of the equality. On the left, row k_i is the coefficient c_i. On the right, row k_i is 0. So $c_i = 0$ for each i.

You should now read #3 on p. 182 and #3 on p. 183.

We have spent the course trying to solve $A\mathbf{x} = \mathbf{b}$ for $A \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^m$. Our ideal of success is probably that $\text{col}(A) = \mathbb{R}^m$ and $N(A) = \{\mathbf{0}_n\}$. Having $\text{col}(A) = \mathbb{R}^m$ means that we can always solve the problem for any $\mathbf{b} \in \mathbb{R}^m$. Having $N(A) = \{\mathbf{0}_n\}$ means that the solution is unique. Here is the bad news: if both $\text{col}(A) = \mathbb{R}^m$ and $N(A) = \{\mathbf{0}_n\}$, then $m = n$. In other words, if we can always solve our problem uniquely, then the underlying matrix must be square.

Here's why. Since $\text{col}(A) = \mathbb{R}^m$, and the dimension of $\mathbb{R}^m$ is m, the dimension of $\text{col}(A)$ is also m. That is, $\text{rank}(A) = m$. And $\text{rank}(A)$ is the number of pivot variables, so A has m pivot variables. But $N(A) = \{\mathbf{0}_n\}$, so A has no free variables and every variable is a pivot variable. Since there are n variables total in play, we must have $m = n$.

So, as soon as we study a nonsquare problem, we will either fail at existence or uniqueness. If $m < n$, then there are more columns than rows, and so more variables than equations. This kind of problem is "underdetermined," and it can't have a unique solution. If $n < m$, then there are more rows than columns, and so more equations than variables. This kind of problem is "overdetermined," and it need not have a solution at all.

What can we learn, then, in case of failure? Let's go back to a familiar RREF—actually, let's just suppose that our matrix A is already in RREF:

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

Columns 1 and 3 are the pivot columns, so

$$\operatorname{col}(A) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \left\{\begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \;\middle|\; x_1,\ x_2 \in \mathbb{R}\right\}.$$

Pretty quickly we see that any vector $\mathbf{b} \in \mathbb{R}^3$ whose third component is nonzero isn't in $\operatorname{col}(A)$. That is,

$$\mathcal{S} = \left\{\mathbf{b} \in \mathbb{R}^3 \;\middle|\; \mathbf{b} \notin \operatorname{col}(A)\right\} = \left\{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \;\middle|\; b_1,\ b_2,\ b_3 \in \mathbb{R},\ b_3 \neq 0\right\}.$$

This set $\mathcal{S}$ certainly captures all the $\mathbf{b} \in \mathbb{R}^3$ for which we can't solve $A\mathbf{x} = \mathbf{b}$, but it's not the best way of describing our failure. For example, $\mathcal{S}$ is not a subspace. (Check that: what's up with $\mathbf{0}_3$? Is the sum of two vectors in $\mathcal{S}$ still in $\mathcal{S}$? I grant you that $c\mathbf{v} \in \mathcal{S}$ for all $c \in \mathbb{R}$ and $\mathbf{v} \in \mathcal{S}$, as long as $c \neq 0$.) That's bad! This is *linear* algebra, and so we want all of our structures to respect linearity as much as possible.

Instead, look at the two vectors in the basis for $\operatorname{col}(A)$ above. They're just the first two columns of the 3×3 identity matrix. All we're missing is the third column, and then we'd have a basis for $\mathbb{R}^3$. So, put $\mathcal{V}_A = \operatorname{col}(A)$ and

$$\mathcal{W}_A = \operatorname{span}\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \left\{\begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} \;\middle|\; x_3 \in \mathbb{R}\right\}.$$

I'm decorating these with the subscript A because later I'll want to use $\mathcal{V}$ and $\mathcal{W}$ for more general subspaces, but those subspaces will always have some things in common with these very specific $\mathcal{V}_A$ and $\mathcal{W}_A$.

Then given $\mathbf{b} \in \mathbb{R}^3$, we can write

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ b_3 \end{bmatrix}.$$

The first vector in the sum above is in $\mathcal{V}_A$ and the second vector is in $\mathcal{W}_A$. Moreover, this decomposition is the *only* way to write a vector in $\mathbb{R}^3$ as the sum of a vector in $\mathcal{V}_A$ and a vector in $\mathcal{W}_A$. I hope that's obvious, but maybe you should check it yourself?

Summing up, we have shown that for each $\mathbf{b} \in \mathbb{R}^3$, there exist unique $\mathbf{v} \in \mathcal{V}_A$ and $\mathbf{w} \in \mathcal{W}_A$ such that $\mathbf{b} = \mathbf{v} + \mathbf{w}$. In other words, every vector in $\mathbb{R}^3$ is the sum of a unique vector $\mathbf{v}$ in col(A) and a "corrector" term $\mathbf{w}$ in a different subspace. Studying that "corrector" term can tell us just how "far off" a vector in $\mathbb{R}^3$ is from being in col(A).

That's half of the story. The other half is that there is a very useful *geometric* relationship between $\mathcal{V}_A$ and $\mathcal{W}_A$. Recall the dot product of $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^m$:

$$\mathbf{x} \cdot \mathbf{y} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = x_1 y_1 + \cdots + x_m y_m = \sum_{k=1}^{m} x_k y_k.$$

Careful—we never write $\mathbf{xy}$, always $\mathbf{x} \cdot \mathbf{y}$. And we don't take the dot product of matrices—always AB, never $A \cdot B$. If you feel a bit rusty on dot products, reread pp. 11–12 and Worked Example 1.2 C.

We used the dot product before as a slick tool for multiplying matrices: the (i, j)-entry of AB was the dot product of row i of A (viewed as a column vector) with column j of B. The dot product is also a handy tool for extracting data about vectors: I claim that if $\mathbf{e}_k$ is the kth column of the $m \times m$ identity matrix, then $\mathbf{v} \cdot \mathbf{e}_k = v_k$, where v_k is the entry in the kth row of $\mathbf{v}$. Can you show why?

Back to $\mathcal{V}_A$ and $\mathcal{W}_A$ from above. If $\mathbf{v} \in \mathcal{V}_A$ and $\mathbf{w} \in \mathcal{W}_A$, then we have

$$\mathbf{v} \cdot \mathbf{w} = 0.$$

Please check that, just to be sure.

Big picture—we are going to use material from Section 3.5 to understand the geometry of Section 4.1. But I'm going to do this out of order from the book. Bear with me, and I'll point you to specific parts of those sections soon!

Day 39: Monday, April 17. If $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ with $\mathbf{v} \cdot \mathbf{w} = 0$, then we say that $\mathbf{v}$ and $\mathbf{w}$ are orthogonal, which is a fancy word for perpendicular. We sometimes denote this as $\mathbf{v} \perp \mathbf{w}$. (I read this as "$\mathbf{v}$ perp $\mathbf{w}$.")

So, here is what we've built out of our failure to solve $A\mathbf{x} = \mathbf{b}$: a deeper understanding of the structure of $\mathbb{R}^3$. Every vector in $\mathbb{R}^3$ is the sum of a vector in col(A) and a term orthogonal to that first vector. If we pause for a bit to generalize some of these ideas, it will pay off down the line.

First, why do we let the dot product capture perpendicularity? Look at

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0 \quad \text{and} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

If you plot those vectors as arrows in $\mathbb{R}^2$, do you see the right angles? That's why we say that the dot product measures perpendicularity.

More generally, we say that two subspaces $\mathcal{V}$ and $\mathcal{W}$ are orthogonal if $\mathbf{v} \perp \mathbf{w}$ for each $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{W}$. We abbreviate this with the sentence $\mathcal{V} \perp \mathcal{W}$. (Now you see why I used $\mathcal{V}_A$ and $\mathcal{W}_A$ before for our specific matrix.)

And so in the case above, we have $\mathcal{V}_A \perp \mathcal{W}_A$, where

$$\mathcal{V}_A = \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) \quad \text{and} \quad \mathcal{W}_A = \text{span}\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right).$$

But there's more: it's really the case that

$$\mathcal{W}_A = \left\{ \mathbf{w} \in \mathbb{R}^3 \;\middle|\; \mathbf{v} \perp \mathbf{w} \text{ for all } \mathbf{v} \in \mathcal{V}_A \right\}.$$

Here's why. Certainly if $\mathbf{w} \in \mathcal{W}_A$, then $\mathbf{v} \cdot \mathbf{w} = 0$ for any $\mathbf{v} \in \mathcal{V}_A$. I mean, you just figured that out. But now suppose that $\mathbf{w}$ is *any* vector in $\mathbb{R}^3$ such that $\mathbf{v} \cdot \mathbf{w}$ for all $\mathbf{v} \in \mathcal{V}_A$. Let's choose our $\mathbf{v}$ judiciously: take $\mathbf{v} = \mathbf{e}_1$ to see that $w_1 = 0$ and $\mathbf{v} = \mathbf{e}_2$ to see that $w_2 = 0$. So the only nonzero entry of $\mathbf{w}$ is its third entry, and therefore $\mathbf{w} \in \mathcal{W}_A$.

This is an example of a more general phenomenon. Let $\mathcal{V}$ be a subspace of $\mathbb{R}^m$. Define

$$\mathcal{V}^\perp = \{ \mathbf{w} \in \mathbb{R}^m \mid \mathbf{v} \perp \mathbf{w} = 0 \text{ for all } \mathbf{v} \in \mathcal{V} \}.$$

We call $\mathcal{V}^\perp$ the orthogonal complement of $\mathcal{V}$ in $\mathbb{R}^m$, but I usually just pronounce it "$\mathcal{V}$ perp." So, in the work above, with

$$\mathcal{V}_A = \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right),$$

we have

$$\mathcal{V}_A^\perp = \text{span}\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) = \mathcal{W}_A.$$

I claim that if $\mathcal{V}$ is a subspace of $\mathbb{R}^m$, then so is $\mathcal{V}^\perp$ is a subspace of $\mathbb{R}^m$—you should check that! The rules for dot products in Problem 19 of Section 1.2 (p. 20) will be very helpful. You should also convince yourself that $\mathcal{V} \perp \mathcal{V}^\perp$ ...the notation would be silly if this didn't happen. However, I claim that just because $\mathcal{V} \perp \mathcal{W}$ for some subspace $\mathcal{W}$, that doesn't imply $\mathcal{W} = \mathcal{V}^\perp$. Take $\mathcal{V}$ and $\mathcal{W}$ each to be the span of a single, different column of the $m \times m$ identity matrix.

We got very lucky in the work above with A in finding $\mathcal{V}_A^\perp = \text{col}(A)^\perp = \mathcal{W}_A$, because A had an incredibly transparent structure. How do we generalize this? Given $A \in \mathbb{R}^{m \times n}$, can we always write a vector $\mathbf{b} \in \mathbb{R}^m$ as the sum of unique, orthogonal vectors in $\text{col}(A)$ and $\text{col}(A)^\perp$? And can we give a simple "formula" for $\text{col}(A)^\perp$, one that we could read off from the structure of A, the way we do for $\text{col}(A)$ or, more or less, $N(A)$? It's taken us a while to get to this question, and we have a bit more work to do, but the good news is that the answer, when we finally arrive there, is really simple. (Think "a basis for the column space is the pivot columns" simple.)

Before proceeding, we need a few more tools about orthogonal subspaces. First, I claim that an orthogonal decomposition, if it exists, always involves unique orthogonal vectors. Let $\mathcal{V}$ be a subspace of $\mathbb{R}^m$ and suppose that for some $\mathbf{x} \in \mathbb{R}^m$, we have $\mathbf{x} = \mathbf{v} + \mathbf{w}$ for some $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{V}^\perp$. Then those terms $\mathbf{v}$ and $\mathbf{w}$ are unique; there is only one way to choose them. Just believe me for now.

Day 40: Wednesday, April 19. So, if it is possible to write each vector $\mathbf{x} \in \mathbb{R}^m$ as the sum of a vector in $\text{col}(A)$ and a vector in $\text{col}(A)^\perp$, then there is only one way to choose those vectors. That's good—no one likes ambiguity or redundancy. But why would it even be possible to make this decomposition? We need more control over $\text{col}(A)^\perp$. Let's think about what it means to be in this subspace. All we have is the definition

$$\text{col}(A)^\perp = \{\mathbf{w} \in \mathbb{R}^m \mid \mathbf{v} \perp \mathbf{w} \text{ for all } \mathbf{v} \in \text{col}(A)\},$$

so let's use that. We have

$$\begin{aligned} \mathbf{w} \in \text{col}(A)^\perp &\iff \mathbf{v} \perp \mathbf{w} \text{ for all } \mathbf{v} \in \text{col}(A) \\ &\iff \mathbf{v} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{v} \in \text{col}(A) \\ &\iff (A\mathbf{x}) \cdot \mathbf{w} = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n. \end{aligned}$$

Here is the big idea. We know that if A were just a scalar $a \in \mathbb{R}$, then $(a\mathbf{x}) \cdot \mathbf{w} = \mathbf{x} \cdot (a\mathbf{w})$. At least, I hope we know that—you probably should check that this is true. Can we "jump" A across the dot product and write

$$(A\mathbf{x}) \cdot \mathbf{w} = \mathbf{x} \cdot (B\mathbf{w})$$

for some matrix B?

First, this matrix B would have to be in $\mathbb{R}^{n \times m}$, since $\mathbf{w} \in \mathbb{R}^m$, and we want $B\mathbf{w} \in \mathbb{R}^n$ so that we can take the dot product of $\mathbf{x}$ and $B\mathbf{w}$. If such a matrix exists (it does!), then we have

$$\mathbf{w} \in \text{col}(A)^\perp \iff \mathbf{x} \cdot (B\mathbf{w}) = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

But here's the great thing: if for some $\mathbf{y} \in \mathbb{R}^n$, it is the case that $\mathbf{x} \cdot \mathbf{y} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$, then really $\mathbf{y} = \mathbf{0}_n$. Why? Just take $\mathbf{x}$ to be the columns of the $n \times n$ identity matrix. Then you'll see that each of the components of $\mathbf{y}$ has to be 0. And so if $\mathbf{w} \in \text{col}(A)^\perp$, then $\mathbf{w} \in N(B)$. Conversely, if $\mathbf{w} \in N(B)$, then $\mathbf{w} \in \text{col}(A)^\perp$ by reversing all the work above. (Please do check that yourself.)

So what is B? I'm going to spoil the surprise and tell you the answer now so we can do some concrete calculations. Then we'll revisit this question and see why the answer is the right one in general. Here it is: $B = A^\mathsf{T}$, the transpose of A.

The transpose of any $A \in \mathbb{R}^{m \times n}$ is the matrix $A^\mathsf{T} \in \mathbb{R}^{m \times n}$ whose rows are the columns of A, or whose columns are the rows of A, or whose (i,j)-entry is the (j,i)-entry of A. For that last one, $A^\mathsf{T}(i,j) = A(j,i)$. For example, if

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

then

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

This is not as nice as before! Previously A was in RREF, but now A^T is not. So here's a warning: taking the transpose does not necessarily preserve the RREF. Look at p. 108 in the book, but don't read the last paragraph yet.

Returning to the problem of describing $\mathrm{col}(A)^\perp$ in a transparent way, all our work leads us to conjecture that $\mathrm{col}(A)^\perp = N(A^\mathsf{T})$. Just put $B = A^\mathsf{T}$ above.

Strang calls $N(A^\mathsf{T})$ the "left nullspace of A." (Gotta call it something.) This is defined on p. 180 in #4 of the "four fundamental subspaces"; see the second paragraph below that for why it's "left." Now you can read pp. 193–196, if you believe that $A\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A^\mathsf{T}\mathbf{y}$ for all $A \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{y} \in \mathbb{R}^m$.

Let's confirm that $\mathrm{col}(A)^\perp = N(A^\mathsf{T})$ in the context of

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

We've known

$$\mathrm{col}(A) = \mathrm{span}\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right) \quad \text{and} \quad \mathrm{col}(A)^\perp = \mathrm{span}\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right)$$

for some time, and now we can calculate $N(A^\mathsf{T})$. We need a bit of elimination and, believe it or not, permutation, to get $\mathrm{rref}(A^\mathsf{T})$:

$$A^\mathsf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 4 & 0 \end{bmatrix} \xrightarrow[E_{21}]{(2)-2\cdot(1)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 4 & 0 \end{bmatrix} \xrightarrow[E_{41}]{(4)-3\cdot(1)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 0 \end{bmatrix} \xrightarrow[E_{42}]{(4)-4\cdot(3)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow[P_{23}]{(2)\leftrightarrow(3)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Here P_{23} is the permutation matrix that interchanges rows 2 and 3—what is that matrix, exactly? Also, what are the dimensions of all the elimination matrices? (For that matter, what are they explicitly?) Last, note that $\mathrm{rank}(A^\mathsf{T}) = 2 = \mathrm{rank}(A)$. This is not a coincidence, even though $\mathrm{rref}(A)^\mathsf{T} \neq \mathrm{rref}(A^\mathsf{T})$, sadly.

We then have

$$A^\mathsf{T}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \mathbf{0}_4 \iff \begin{cases} x_1 = 0 \\ x_2 = 0 \\ x_3 = x_3, \end{cases}$$

and so all $\mathbf{x} \in N(A^\mathsf{T})$ have the form

$$\mathbf{x} = x_3\begin{bmatrix}0\\0\\1\end{bmatrix}$$

for some $x_3 \in \mathbb{R}$. That is,

$$N(A^\mathsf{T}) = \mathrm{span}\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right) = \mathrm{col}(A)^\perp,$$

exactly as we figured out earlier.

Day 41: Friday, April 21. Here's a recap of our work from Wednesday. Let $A \in \mathbb{R}^{m \times n}$. The key action of the transpose is that $A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^{\mathsf{T}}\mathbf{y}$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Then

$$\begin{aligned}
\mathbf{w} \in \operatorname{col}(A)^{\perp} &\iff \mathbf{v} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{v} \in \operatorname{col}(A) \text{ by definition of } \operatorname{col}(A)^{\perp} \\
&\iff A\mathbf{x} \cdot \mathbf{w} = 0 \text{ for all} \mathbf{x} \in \mathbb{R}^n, \text{ as } \operatorname{col}(A) = \{\mathbf{w} \in \mathbb{R}^m \mid \mathbf{w} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^n\} \\
&\iff \mathbf{x} \cdot A^{\mathsf{T}}\mathbf{w} = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n, \text{ by properties of the transpose} \\
&\iff A^{\mathsf{T}}\mathbf{w} = 0 \text{ by taking } \mathbf{x} = \text{ columns of identity} \\
&\iff \mathbf{w} \in N(A^{\mathsf{T}}) \text{ by definition of } N(A^{\mathsf{T}}).
\end{aligned}$$

This is our proof that $\operatorname{col}(A)^{\perp} = N(A^{\mathsf{T}})$.

We are going to keep studying this equality quite a bit further, but let's retread and ask why we'd use the transpose here in the first place. Who would have guessed that $A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^{\mathsf{T}}\mathbf{y}$ if they didn't know it already? Well, we started out with $\mathbf{w} \in \operatorname{col}(A)^{\perp}$, and we saw that this was equivalent to $A\mathbf{x} \cdot \mathbf{w} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Maybe someone thought that this equality would be easier if they could move the matrix over to $\mathbf{w}$. After all, if we just had scalar multiplication, that would work: $(a\mathbf{x}) \cdot \mathbf{w} = \mathbf{x} \cdot (a\mathbf{w})$ for any $a \in \mathbb{R}$. Is there a matrix B such that $A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot B\mathbf{y}$ all the time?

Fool around with a really simple case:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2 \times 2}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2, \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathbb{R}^2.$$

We compute

$$A\mathbf{x} = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix}$$

and then

$$(A\mathbf{x}) \cdot \mathbf{y} = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = (ax_1 + bx_2)y_1 + (cx_1 + dx_2)y_2.$$

We want to recognize this as a dot product of the form $\mathbf{x} \cdot (B\mathbf{y})$. We may not know what B is, but we should expect there to be two terms, one with a factor of x_1 and the other with a factor of x_2. So let's rewrite

$$\begin{aligned}
(ax_1 + bx_2)y_1 + (cx_1 + dx_2)y_2 &= ax_1y_1 + bx_2y_1 + cx_1y_2 + dx_2y_2 \\
&= (ax_1y_1 + cx_1y_2) + (bx_2y_1 + dx_2y_2) \\
&= (ay_1 + cy_2)x_1 + (by_1 + dy_2)x_2 \\
&= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} ay_1 + cy_2 \\ by_1 + dy_2 \end{bmatrix}.
\end{aligned}$$

Now we want to recognize the vector

$$\begin{bmatrix} ay_1 + cy_2 \\ by_1 + dy_2 \end{bmatrix}$$

as the matrix-vector product $B\mathbf{y}$ for some matrix B. After a moment of thinking, hopefully we see

$$\begin{bmatrix} ay_1 + cy_2 \\ by_1 + dy_2 \end{bmatrix} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \mathbf{y}.$$

That's our B! The rows of B are the columns of A; equivalently, the columns of B are the rows of A. The matrix B is the transpose of A: $B = A^\mathsf{T}$.

Now back to the equality $\text{col}(A)^\perp = N(A^\mathsf{T})$. This tells us stuff about dimension. Let me write $\dim[\mathcal{V}]$ for the dimension of a subspace $\mathcal{V}$. For example, $\dim[\text{col}(A)] = \text{rank}(A)$. Since $\text{col}(A)^\perp = N(A^\mathsf{T})$, we have

$$\dim[\text{col}(A)^\perp] = \dim[N(A^\mathsf{T})].$$

But we know how to find the dimension of a null space. For $A \in \mathbb{R}^{m \times n}$, it's $\dim[N(A)] = n - \text{rank}(A)$, so for $B \in \mathbb{R}^{n \times m}$, it's $\dim[N(B)] = m - \text{rank}(B)$. Thus

$$\dim[N(A^\mathsf{T})] = m - \text{rank}(A^\mathsf{T}),$$

and therefore

$$\dim[\text{col}(A)^\perp] = m - \text{rank}(A^\mathsf{T}).$$

So, what is $\text{rank}(A^\mathsf{T}) = \dim[\text{col}(A^\mathsf{T})]$? That's our new task. Let's look at our faithful friend:

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad A^\mathsf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 4 & 0 \end{bmatrix}.$$

We know $\text{rank}(A) = 2$ here, and we did figure out $\text{rref}(A^\mathsf{T})$ before, but we don't actually need that. Because the third column of A^T is $\mathbf{0}_4$, $\text{col}(A^\mathsf{T})$ is spanned just by the first two columns of A^T. And those first two columns of A^T are linearly independent. Look at the 1's that I've highlighted in blue:

$$c_1 \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Longrightarrow c_1 = 0 \text{ and } c_2 = 0.$$

So $\dim[\text{col}(A^\mathsf{T})] = 2$, and therefore $\text{rank}(A^\mathsf{T}) = 2 = \text{rank}(A)$.

This turns out to be true in general. In fact, we can say something explicit about not just $\text{rank}(A^\mathsf{T})$ but also a basis for $\text{col}(A^\mathsf{T})$. Say that a row in $\text{rref}(A)$ is a **pivot row** if that row contains a pivot, just like a pivot column is a column that contains a pivot. Here we have

$$\text{pivot columns are 1 and 3: } \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \text{pivot rows are 1 and 2: } \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The pivot rows of $\text{rref}(A)$ are the pivot columns of A^T:

$$\text{pivot columns are 1 and 2: } \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 4 & 0 \end{bmatrix}.$$

The only unfortunate thing with this example is that $A = \text{rref}(A)$ here. We'll need to do some more work to study A when it's not already in RREF. But the principles are the same:

$$\text{basis for } \text{col}(A) = \text{pivot columns of } A,$$

$$\text{rank}(A) = \text{rank}(A^\mathsf{T}),$$

and

$$\text{basis for } \text{col}(A^\mathsf{T}) = \text{pivot rows of } \text{rref}(A).$$

Day 42: Monday, April 24. Let $A \in \mathbb{R}^{m\times n}$. We know

$$\text{col}(A)^\perp = \{\mathbf{w} \in \mathbb{R}^m \mid \mathbf{v}\cdot\mathbf{w} = 0\} = N(A^\mathsf{T}).$$

And so

$$\dim[\text{col}(A)^\perp] = \dim[N(A^\mathsf{T})] = m - \text{rank}(A^\mathsf{T}).$$

So what is $\text{rank}(A^\mathsf{T})$? By definition it's $\dim[\text{col}(A^\mathsf{T})]$, but can we express that solely in terms of data about A? Our conjecture is that $\text{rank}(A^\mathsf{T}) = \text{rank}(A)$.

Before we always worked with a matrix A already in RREF. Suppose now that we just know $A = FR$, where $R = \text{rref}(A)$, and $F \in \mathbb{R}^{m\times m}$ is invertible. For example, what if

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

for some invertible $F \in \mathbb{R}^{3\times 3}$? There's a lot of good data stored in that R, but we still have work to do.

We want to study $A^\mathsf{T} = (FR)^\mathsf{T}$, so we should figure out what the transpose of a product is. This is not the first time that we've applied an operation to a product that is valid on each of the factors. Remember that if G and H (I'm running out of letters!) are square invertible matrices, then $(GH)^{-1} = H^{-1}G^{-1}$. We might guess, then, that $(FR)^\mathsf{T} = R^\mathsf{T}F^\mathsf{T}$.

This is true in general. First, please check that if G and H are matrices such that the product GH is defined, then the product $H^\mathsf{T}G^\mathsf{T}$ is also defined. There are (at least) a couple of proofs that $(GH)^\mathsf{T} = H^\mathsf{T}G^\mathsf{T}$. One argument appears on p. 109 in equation (4); this equation uses the "third way" of matrix multiplication back on p. 72.

Here's another argument, which I encourage you to chase through. First, convince yourself that if $J \in \mathbb{R}^{m\times n}$ satisfies $J\mathbf{x}\cdot\mathbf{y} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, then J is the $m \times n$ zero matrix. (Here's why: pick $\mathbf{x}$ to be one of the standard basis vectors for $\mathbb{R}^n$, so that $J\mathbf{x}$ is one of the columns of J. Which column? Then pick $\mathbf{y}$ to be one of the standard basis vectors for $\mathbb{R}^m$, so $J\mathbf{x} \cdot \mathbf{y}$ is an entry of one of the columns of J. Which entry? Conclude that all of the entries of J are 0.) Now look at the dot product

$$(H^\mathsf{T} G^\mathsf{T} - (GH)^\mathsf{T})\mathbf{x} \cdot \mathbf{y}.$$

Use the fundamental property of the transpose relative to dot products to show that this dot product is always 0.

So, we see that if $A = FR$, then $A^\mathsf{T} = R^\mathsf{T} F^\mathsf{T}$. Remember that F is invertible. I claim this means F^T is invertible. All we have to do is find some matrix H such that $F^\mathsf{T} H = I$, right? Let's see if we can squeeze F^{-1} in there somewhere. We have

$$F^\mathsf{T}(F^{-1})^\mathsf{T} = (F^{-1}F)^\mathsf{T} = I^\mathsf{T} = I.$$

The transpose of the identity matrix is the identity matrix, right? And so we have shown that F^T is invertible with

$$(F^\mathsf{T})^{-1} = (F^{-1})^\mathsf{T}.$$

Remember that the goal is to study $\text{col}(A^\mathsf{T}) = \text{col}(R^\mathsf{T} F^\mathsf{T})$. I claim that because F^T is invertible and multiplying R^T on the right, we really have $\text{col}(R^\mathsf{T} F^\mathsf{T}) = \text{col}(R^\mathsf{T})$. Hold on, you say! One of the (few) disappointments of the RREF is that a matrix and its RREF can have different column spaces, i.e., even though $A = FR$, we should expect $\text{col}(A) \neq \text{col}(FR)$. What's different here? It's the multiplication by F^T on the right in the product $R^\mathsf{T} F^\mathsf{T}$, in contrast to multiplication by F on the left in FR.

Here's the more general truth. Suppose that G and H are matrices such that the product GH is defined. (How big are they again?) Suppose also that H is invertible. Then $\text{col}(G) = \text{col}(GH)$. Here's why. If $\mathbf{b} \in \text{col}(G)$, then $\mathbf{b} = G\mathbf{x}$ for some $\mathbf{x}$. Here's the great trick: multiply by 1, or, really, by I. We have $G\mathbf{x} = (GH)(H^{-1}\mathbf{x}) \in \text{col}(GH)$. And if $\mathbf{b} \in \text{col}(GH)$, then $\mathbf{b} = G(H\mathbf{x}) \in \text{col}(G)$. Sound good?

And so it all comes down to this:

$$\text{col}(A^\mathsf{T}) = \text{col}((FR)^\mathsf{T}) = \text{col}(R^\mathsf{T} F^\mathsf{T}) = \text{col}(R^\mathsf{T}).$$

In particular, a basis for $\text{col}(R^\mathsf{T})$ will be a basis for $\text{col}(A^\mathsf{T})$, as they're the same spaces. *But it probably won't be the case that a basis for* $\text{col}(R)$ *will be a basis for* $\text{col}(A)$*!* And $\text{rank}(A^\mathsf{T}) = \text{rank}(R^\mathsf{T})$.

So, what is a basis for $\text{col}(R^\mathsf{T})$? Remember what R^T is. The columns of R^T are the rows of R. And the rows of R are either pivot rows (that is, rows with pivots) or rows of all 0. Remember our favorite R:

$$\text{pivot rows in } R = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{become} \quad \text{pivot columns in } R^\mathsf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 4 & 0 \end{bmatrix}.$$

So, every column of R^T is either a pivot row of R (written as a column) or a column of all 0. A column of all 0 adds nothing to a span. Thus $\text{col}(R^\mathsf{T})$ is the span of the pivot rows of R.

I claim that those pivot rows are linearly independent because, if $\mathrm{rank}(R) = r$, those rows contain the $r \times r$ identity matrix. Here is that identity matrix highlighted in R and R^T:

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad R^\mathsf{T} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 4 & 0 \end{bmatrix}.$$

There is no way for a linear combination of vectors whose rows contain an identity matrix to add up to the zero vector unless all the coefficients in that combination are 0. Consequently, the pivot rows are linearly independent. Since $\mathrm{col}(R^\mathsf{T})$ is already the span of the pivot rows of R, the pivot rows of R form a basis for $\mathrm{col}(R^\mathsf{T})$. And there are $r = \mathrm{rank}(A) = \mathrm{rank}(R)$ pivot rows, because there are r pivots.

We conclude that $\mathrm{col}(A^\mathsf{T}) = \mathrm{col}(R^\mathsf{T})$, that the pivot rows of R form a basis for $\mathrm{col}(R^\mathsf{T})$ and thus for $\mathrm{col}(A^\mathsf{T})$, and that $\mathrm{rank}(A^\mathsf{T}) = \mathrm{rank}(A)$. In particular, $\dim[\mathrm{col}(A)^\perp] = m - \mathrm{rank}(A^\mathsf{T}) = m - \mathrm{rank}(A)$. Wow.

You should now read #1 on p. 181 and then #1 on p. 183. In fact, you could put all of our work together and read pp. 180–184. Watch out, though, as there are serious typos in the "Four fundamental subspaces" box on p. 180. The row space should be $\mathbf{C}(A^\mathsf{T})$ and the column space should be $\mathbf{C}(A)$.

Day 43: Wednesday, April 26. Here is a summary of some of the highlights of our class. Remember that the dimension of any subspace $\mathcal{V}$ of $\mathbb{R}^n$ is the number of vectors in a basis for $\mathcal{V}$; we write this number as $\dim[\mathcal{V}]$. Every subspace has a basis, and every basis for a subspace has the same number of vectors.

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

1. The pivot rows of A form a basis for $\mathrm{col}(A)$.

2. $\dim[\mathrm{col}(A)] = \mathrm{rank}(A) =$ the number of pivot variables.

3. $\dim[N(A)] = n - \mathrm{rank}(A) =$ the number of free variables.

4. $\mathrm{col}(A)^\perp = N(A^\mathsf{T})$.

5. $\mathrm{rank}(A^\mathsf{T}) = \mathrm{rank}(A)$.

6. The pivot rows of $\mathrm{rref}(A)$ form a basis for $\mathrm{col}(A^\mathsf{T})$, which is sometimes called the row space of A.

7. $\dim[\mathrm{col}(A)^\perp] = \dim[N(A^\mathsf{T})] = m - \mathrm{rank}(A)$.

Let's do an example with actual numbers. Let

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

I claim that

$$R = \text{rref}(A) = \begin{bmatrix} 1 & 0 & 0 & 10 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 10 \end{bmatrix}.$$

At this point in the course, you should be able to figure that out without too much grief. The pivot columns of A are columns 1, 2, and 3, and the only free variable is x_4. Then a basis for the column space of A is

$$\begin{bmatrix} 1 \\ -3 \\ -3 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ -4 \\ 8 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix}.$$

This is a three-dimensional subspace of $\mathbb{R}^3$, so $\text{col}(A) = \mathbb{R}^3$. Great—we can always solve $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^3$.

Why? I claim we did something like that on Day 36. Here's a more general proof—you should fill in some of the justification gaps. Let $\mathcal{V}$ be a subspace of $\mathbb{R}^m$ such that $\dim[\mathcal{V}] = m$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be a basis for $\mathcal{V}$. Put $B = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m \end{bmatrix}$. Then $B \in \mathbb{R}^{m \times nm}$ and the columns of B are linearly independent, so B is invertible. Now let $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{b} = B(B^{-1}\mathbf{b}) \in \text{col}(B)$, and $\text{col}(B) = \mathcal{V}$. So $\mathbb{R}^m = \mathcal{V}$.

Back to A. Do you see the linear dependence relation in the columns of A:

$$10 \begin{bmatrix} 1 \\ -3 \\ -3 \end{bmatrix} - \begin{bmatrix} 2 \\ -4 \\ 8 \end{bmatrix} + 10 \begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ 2 \end{bmatrix}?$$

Next, we determine $N(A)$ by solving $R\mathbf{x} = \mathbf{0}_3$: we get

$$R\mathbf{x} = \mathbf{0}_3 \iff \begin{cases} x_1 + 10x_4 = 0 \\ x_2 - x_4 = 0 \\ x_3 + 10x_4 = 0 \end{cases} \iff \mathbf{x} = \begin{bmatrix} -10x_4 \\ x_4 \\ -10x_4 \\ x_4 \end{bmatrix} = x_4 \begin{bmatrix} -10 \\ 1 \\ -10 \\ 1 \end{bmatrix}.$$

Hence a basis for $N(A)$ consists of the single vector

$$\begin{bmatrix} -10 \\ 1 \\ -10 \\ 1 \end{bmatrix}.$$

Consequently, we can always solve $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^3$, but we can never do so uniquely. Indeed, if $A\mathbf{x}_\star = \mathbf{b}$ for some $\mathbf{x}_\star \in \mathbb{R}^4$, then every solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ has the form

$$\mathbf{x} = \mathbf{x}_\star + c \begin{bmatrix} -10 \\ 1 \\ -10 \\ 1 \end{bmatrix}$$

for some $c \in \mathbb{R}$. I claim we talked about this on Day 28.

Third, a basis for $\mathrm{col}(A^{\mathsf{T}})$ consists of the pivot rows of R written as column vectors in $\mathbb{R}^4$. Each row of R is a pivot row, so a basis for $\mathrm{col}(A^{\mathsf{T}})$ is

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 10 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \\ 10 \end{bmatrix}.$$

Finally, to get a basis for $N(A^{\mathsf{T}})$, we would need to solve $A^{\mathsf{T}}\mathbf{x} = \mathbf{0}_4$. Unfortunately, we can't use the RREF of A here, since $\mathrm{rref}(A)^{\mathsf{T}} \neq \mathrm{rref}(A^{\mathsf{T}})$. I mean, $\mathrm{rref}(A)^{\mathsf{T}}$ isn't even in RREF.

For more examples, see Examples 1, 2, and 3 on pp. 184–186. In Example 3, you can ignore the discussion of graphs and just focus on the four fundamental subspaces of the given matrix.

There is one last thing for us to figure out. A long time ago we looked at a particular matrix $A \in \mathbb{R}^{3\times 4}$ and saw explicitly that we could write each $\mathbf{b} \in \mathbb{R}^3$ as the sum of unique vectors $\mathbf{v} \in \mathrm{col}(A)$ and $\mathbf{w} \in \mathrm{col}(A)^{\perp}$. That is, for each $\mathbf{b} \in \mathbb{R}^3$, there are vectors $\mathbf{v} \in \mathrm{col}(A)$ and $\mathbf{w} \in \mathrm{col}(A)^{\perp}$ such that $\mathbf{b} = \mathbf{v} + \mathbf{w}$, and there is only one way to choose these $\mathbf{v}$ and $\mathbf{w}$.

Given $A \in \mathbb{R}^{m\times n}$, why can we write each $\mathbf{b} \in \mathbb{R}^m$ as a sum $\mathbf{b} = \mathbf{v} + \mathbf{w}$ for $\mathbf{v} \in \mathrm{col}(A)$ and $\mathbf{w} \in \mathrm{col}(A)^{\perp} = N(A^{\mathsf{T}})$? And why are the vectors $\mathbf{v}$ and $\mathbf{w}$ unique? First we handle existence: why are there $\mathbf{v}$ and $\mathbf{w}$ that do this in the first place? Abbreviate $r = \mathrm{rank}(A)$, so $\dim[\mathrm{col}(A)] = r$ and, thanks to our hard work above, $\dim[N(A^{\mathsf{T}})] = m - r$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be a basis for $\mathrm{col}(A)$ and let $\mathbf{w}_1, \ldots, \mathbf{w}_{m-r}$ be a basis for $N(A^{\mathsf{T}})$. Then there are m vectors in the list $\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}_1, \ldots, \mathbf{w}_{m-r}$, so if we just show that this list is linearly independent, then it will be a basis for $\mathbb{R}^m$. (Why? Suppose they're linearly independent. Let $\mathcal{V} = \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}_1, \ldots, \mathbf{w}_{m-r})$, so $\dim[\mathcal{V}] = m$. We showed above that if $\mathcal{V}$ is a subspace of $\mathbb{R}^m$ with $\dim[\mathcal{V}] = m$, then $\mathcal{V} = \mathbb{R}^m$.)

Assume

$$c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r + c_{r+1}\mathbf{w}_1 + \cdots + c_m\mathbf{w}_{m-r} = \mathbf{0}_m.$$

Abbreviate

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r \quad \text{and} \quad \mathbf{w} = c_{r+1}\mathbf{w}_1 + \cdots + c_m\mathbf{w}_{m-r}.$$

Then we know three things: $\mathbf{v} \in \mathrm{col}(A)$, $\mathbf{w} \in N(A^{\mathsf{T}})$, and $\mathbf{v} + \mathbf{w} = \mathbf{0}_m$. These things tell us $\mathbf{v} = -\mathbf{w} \in N(A^{\mathsf{T}}) = \mathrm{col}(A)^{\perp}$. Thus $\mathbf{v} \in \mathrm{col}(A)$ and $\mathbf{v} \in \mathrm{col}(A)^{\perp}$. I claim that if $\mathcal{V}$ is a subspace of $\mathbb{R}^m$, then the only vector in both $\mathcal{V}$ and $\mathcal{V}^{\perp}$ is the zero vector. (Check this yourself: if $\mathbf{v} \in \mathcal{V}$ and $\mathcal{V}^{\perp}$, then $\mathbf{v} \cdot \mathbf{v} = 0$, right? Show that $0 = \mathbf{v} \cdot \mathbf{v} = v_1^2 + \cdots + v_m^2 \geq 0$. Since $v_k^2 \geq 0$ always, if $v_k^2 > 0$, then $\mathbf{v} \cdot \mathbf{v} > 0$, which can't be true. So $v_k^2 = 0$ for all k.)

And so $\mathbf{v} = \mathbf{0}_m$. Can you also show that $\mathbf{w} = \mathbf{0}_m$ using the same reasoning? And so $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = \mathbf{0}_m$. Since $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent, we have $c_1 = \cdots = c_r = 0$. Can you also show that $c_{r+1} = \cdots = c_m = 0$ using the same reasoning?

For uniqueness, I claim that if $\mathcal{V}$ is a subspace of $\mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^m$ can be written as $\mathbf{b} = \mathbf{v} + \mathbf{w}$ for some $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{V}^{\perp}$, then there is only one way to choose these $\mathbf{v}$ and $\mathbf{w}$. Otherwise, if $\mathbf{b} = \widetilde{\mathbf{v}} + \widetilde{\mathbf{w}}$ for some $\widetilde{\mathbf{v}} \in \mathcal{V}$ and $\widetilde{\mathbf{w}} \in \mathcal{V}^{\perp}$, then we have $\mathbf{v} + \mathbf{w} = \widetilde{\mathbf{v}} + \widetilde{\mathbf{w}}$, thus $\mathbf{v} - \widetilde{\mathbf{v}} = \widetilde{\mathbf{w}} - \mathbf{w}$. Since $\mathbf{w}$, $\widetilde{\mathbf{w}} \in \mathcal{V}^{\perp}$, and $\mathcal{V}^{\perp}$ is a subspace, we also have $\widetilde{\mathbf{w}} - \mathbf{w} \in \mathcal{V}^{\perp}$. But then $\mathbf{v} - \widetilde{\mathbf{v}} \in \mathcal{V}^{\perp}$, and we already know $\mathbf{v} - \widetilde{\mathbf{v}} \in \mathcal{V}$, right? Doesn't this mean $\mathbf{v} - \widetilde{\mathbf{v}} = \mathbf{0}_m$? Is that useful? And can you also show that $\mathbf{w} - \widetilde{\mathbf{w}} = \mathbf{0}_m$?

So there we go. Every vector in $\mathbb{R}^m$ is the sum of orthogonal vectors in $\text{col}(A)$ and $N(A^\mathsf{T})$, and the vectors in this sum are necessarily unique. See p. 199.

Day 45: Friday, April 28. Look at the pictures on the front cover (and the blurb on p. iii under "The front cover captures a central idea of linear algebra"), in Figure 3.5 on p. 183, and in Figure 4.3 on p. 198. This is their story, and ours.

Let $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = r$. We have shown that every vector $\mathbf{b} \in \mathbb{R}^m$ is the sum of a unique vector $\mathbf{v} \in \text{col}(A)$ and a unique vector $\mathbf{w} \in \text{col}(A)^\perp = N(A^\mathsf{T})$. Moreover, $\mathbf{v}\cdot\mathbf{w} = 0$. We might euphemistically write $\mathbb{R}^m = \text{col}(A) \oplus N(A^\mathsf{T})$. Also, $\dim[\text{col}(A)] = r$, and $\dim[N(A^\mathsf{T})] = m - r$.

Now let $B = A^\mathsf{T}$. Then $B \in \mathbb{R}^{n\times m}$, and we also know $\text{rank}(B) = \text{rank}(A^\mathsf{T}) = \text{rank}(A) = r$. The results in the previous paragraph say that every vector $\mathbf{x} \in \mathbb{R}^n$ is the sum of a unique vector $\mathbf{c} \in \text{col}(B)$ and a unique vector $\mathbf{d} \in N(B^\mathsf{T})$ such that $\mathbf{c}\cdot\mathbf{d} = 0$. But $B^\mathsf{T} = (A^\mathsf{T})^\mathsf{T} = A$, so every vector in $\mathbb{R}^n$ is the sum of unique orthogonal vectors in $\text{col}(A^\mathsf{T})$ and $N(A)$. So $\mathbb{R}^n = N(A) \oplus \text{col}(A^\mathsf{T})$. Also, $\dim[\text{col}(A^\mathsf{T})] = r$ and $\dim[N(A)] = n - r$.

This is what Figure 3.5 says and one of the things that A does: A induces meaningful orthogonal decompositions of $\mathbb{R}^n$ and $\mathbb{R}^m$. Now look at Figure 4.3. This is one of the other things that A does: A acts on vectors in $\mathbb{R}^n$ and sends them to the column space. *This is what the front cover says; this is what linear algebra is.*

While we started out using matrices as a convenient organization tool for storing and representing systems of equations, they do much more than that. For me, these are the things to remember from the course long after you've forgotten how to compute a RREF or check if something is a subspace.

1. Yes, a matrix fundamentally encodes data about a linear system $A\mathbf{x} = \mathbf{b}$.

2. But a matrix also *acts* on data: if $A \in \mathbb{R}^{m\times n}$, then A transforms $\mathbf{x} \in \mathbb{R}^n$ into $A\mathbf{x} \in \mathbb{R}^m$.

3. And so a matrix links two kinds of data: things in $\mathbb{R}^n$ with things in $\mathbb{R}^m$. But more than linking, a matrix *decomposes*: if $A \in \mathbb{R}^{m\times n}$, then we can recognize and orient everything in $\mathbb{R}^n$ and $\mathbb{R}^m$ relative to A via the "orthogonal direct sum" representations $\mathbb{R}^n = \text{col}(A^\mathsf{T}) \oplus N(A)$ and $\mathbb{R}^m = \text{col}(A) \oplus N(A^\mathsf{T})$.