MATH 4260: LINEAR ALGEBRA II

Daily Log for Lectures, Readings, and Vocabulary

Timothy E. Faver

October 14, 2024

CONTENTS

Day 1: Monday, August 12.

Material from *Linear Algebra* by Meckes & Meckes

I expect that you are very familiar with the notation and techniques of Sections 1.1, 1.2, and 1.3. We will revisit Gaussian elimination in a more rapid and abstract context later when we review matrix multiplication and the RREF.

Here are three examples, two totally made up and one not so made up, that all have some common linear algebraic features despite their heavy cosmetic differences.

1.1 Example. For what b_1, b_2, $b_3 \in \mathbb{R}$ can we find $x_1, \ldots, x_5 \in \mathbb{R}$ such that

$$\begin{cases} x_1 + 2x_2 + x_3 + 7x_4 + x_5 = b_1 \\ 2x_1 + 4x_2 + 2x_3 + 14x_4 + 2x_5 = b_2 \\ 2x_3 + 8x_4 = b_3? \end{cases}$$

If we can find such x_k, are they unique?

We proceed with "elementary row operations." Subtract 2 times the first equation from the second to find that the problem is equivalent to

$$\begin{cases} x_1 + 2x_2 + x_3 + 7x_4 + x_5 = b_1 \\ 0 = b_2 - 2b_1 \\ 2x_3 + 8x_4 = b_3. \end{cases}$$

Divide both sides of the third equation by 2 to get a second equivalent problem

$$\begin{cases} x_1 + 2x_2 + x_3 + 7x_4 + x_5 = b_1 \\ 0 = b_2 - 2b_1 \\ x_3 + 4x_4 = b_3/2. \end{cases}$$

For readability, interchange the second and third equations to get another equivalent problem

$$\begin{cases} x_1 + 2x_2 + x_3 + 7x_4 + x_5 = b_1 \\ x_3 + 4x_4 = b_3/2 \\ 0 = b_2 - 2b_1. \end{cases}$$

Last, for no good reason except hope and faith, subtract the (new) second equation from the (original) first equation to end with

$$\begin{cases} x_1 + 2x_2 + 3x_4 + x_5 = b_1 - b_3/2 \\ x_3 + 4x_4 = b_3/2 \\ 0 = b_2 - 2b_1. \end{cases}$$

The upshot of our final problem is that the "data" involving the variables is rather simpler, as there are fewer of them in play. We also see a necessary condition for solving the problem: b_1 and b_2 must satisfy $b_2 - 2b_1 = 0$. That is, $b_2 = 2b_1$, and so we will not be able to solve the problem for arbitrary right sides b_1, b_2, and b_3. For example, there will be no solutions when $b_1 = 1$ and $b_2 = 0$. Of course, the final system is the reduced row echelon form (or row reduced echelon form) of the original, and we achieved it via the three elementary row operations: subtracting a multiple of one equation (row) from another, multiplying both sides of an equation (every entry in a row) by the same nonzero number, and interchanging two equations (rows).

Now we solve the final problem: we must have

$$x_1 = (b_1 - b_3/2) - 2x_2 - 3x_4 - x_5 \quad \text{and} \quad x_3 = b_3/2 - 4x_4.$$

This immediately destroys uniqueness of the solution, even if the "solvability condition" $b_2 - 2b_1 = 0$ is met, as we are "free" (pun intended) to pick x_2, x_4, and x_5 to be any values we like, and each choice of these "free variables" will create a different solution. Consequently, the problem has infinitely many solutions when $b_2 - 2b_1 = 0$ and no solution when $b_2 - 2b_1 \neq 0$.

However, saying that a set has infinitely many members is not a very useful measurement in linear algebra; indeed, most of the interesting sets that we will study are infinite. We can get better control over the "size" of the solution set by focusing on the degrees of freedom in the solution: it looks like there are 3, coming from each of the free variables x_2, x_4, and x_5. The language of vectors will make this precise.

Here is another problem that we can phrase using functions and calculus but that really reduces to a linear system. Like the first, this is totally made up.

1.2 Example. For what quadratics q can we find a quadratic p such that

$$(x+1)p'(x) = q(x) \text{ for all } x?$$

Some further notation (really, some *coordinates*) will help: say $p(x) = a_0 + a_1x + a_2x^2$ and $q(x) = b_0 + b_1x + b_2x^2$. We should view b_0, b_1, and b_2 as given, and we want to find a_0, a_1, and a_2. Since $q'(x) = 2a_2x + a_1$, we have

$$\begin{aligned}(x+1)p'(x) = q(x) &\iff (x+1)(2a_2x + a_1) = b_0 + b_1x + b_2x^2 \\ &\iff 2a_2x^2 + (a_1 + 2a_2)x + a_1 = b_0 + b_1x + b_2x^2.\end{aligned}$$

We recall that two polynomials are equal for all x if and only if their corresponding coefficients are equal, so we need

$$2a_2 = b_2, \qquad a_1 + 2a_2 = b_1, \qquad \text{and} \qquad b_0 = a_1.$$

This immediately lets us solve for

$$a_2 = \frac{b_2}{2} \quad \text{and} \quad a_1 = b_0$$

in terms of the given coefficients, but it imposes no restrictions on a_0. So, we have infinitely many solutions

$$p(x) = a_0 + b_0x + \frac{b_2x^2}{2},$$

and there is one "degree of freedom" here, coming from a_0.

However, we also have a solvability condition on the coefficients of q:

$$b_1 = a_1 + 2a_2 = b_0 + b_2.$$

That is, we can only solve this problem if $b_1 = b_0 + b_2$, and we cannot solve the problem uniquely. This is just like our first example.

1.3 Problem. We can rephrase the solution to the previous example more cleanly using derivatives. Show that if p and q are quadratics with $(x+1)p'(x) = q(x)$, then

$$p(x) = a_0 + q(0)x + \frac{q''(0)}{4}x^2$$

for some $a_0 \in \mathbb{R}$. [Hint: *what are the Taylor coefficients of q?*]

Our last example comes from differential equations (no knowledge of which is presumed for this course).

1.4 Example. A popular second-order linear ordinary differential equation is

$$\epsilon^2 f'' + f = g.$$

Here $\epsilon > 0$ and a continuous function g defined on $\mathbb{R}$ are given, and we want to find a function f on $\mathbb{R}$ such that $\epsilon^2 f''(x) + f(x) = g(x)$ for all x. The dreaded method of variation of parameters furnishes us such a solution:

$$f(x) = c_1 \cos\left(\frac{x}{\epsilon}\right) + c_2\epsilon \sin\left(\frac{x}{\epsilon}\right) + \frac{1}{\epsilon}\int_0^x \sin\left(\frac{x-\tau}{\epsilon}\right) g(\tau)\ d\tau.$$

This is not something that I presume you know off the top of your head; just accept it that every solution to the ODE above has this form for some c_1, $c_2 \in \mathbb{R}$. In fact, it should not be hard to see that $c_1 = f(0)$; with a little more work (possibly expanding the sine inside the integral using a trig addition formula), you can show $c_2 = f'(0)$.

The difference between this problem and the previous two examples is that we can *always* solve the ODE; there are no apparent solvability conditions on g. We still lack uniqueness, but we always get a solution. However, as posed this problem is too general and vague for physical reasonableness. Often we want some extra conditions on the "forcing" function g and/or the solution f. Let's impose the following.

1. The problem is inherently symmetric: if f is even, then so is $\epsilon^2 f'' + f$. Symmetries often cut down on the amount of data that we need to manage and the amount of work that we need to do. So, assume that f and g are even: $f(-x) = f(x)$ and $g(-x) = g(x)$.

2. Assume that g is sufficiently well-behaved for large x that the improper integral

$$\int_0^\infty |g(x)|\ dx := \lim_{b\to\infty}\int_0^b |g(x)|\ dx$$

converges. This integral is one way of measuring the "size" of g as a function, and its convergence says that g is "not too large." Incidentally, since g is even, this also implies that the integral $\int_{-\infty}^{0} |g(x)| \, dx$ converges, and so the integral $\int_{-\infty}^{\infty} |g(x)| \, dx$ converges, too.

3. Assume that we want f to vanish at ∞ in the sense that

$$\lim_{x \to \infty} f(x) = 0.$$

If we think of f as the response to the driving force g (maybe with f measuring the displacement of a harmonic oscillator), this says that the response dies out over long scales. By the way, since f is even, this says $\lim_{x \to -\infty} f(x) = 0$, too.

I claim that with some algebra and calculus (nothing too fancy), these conditions specify the coefficients c_1 and c_2 for us. Namely, we are forced to take

$$c_1 = \frac{1}{\epsilon} \int_0^\infty \sin\left(\frac{x}{\epsilon}\right) g(x) \, dx \quad \text{and} \quad c_2 = \frac{1}{\epsilon} \int_0^\infty \cos\left(\frac{x}{\epsilon}\right) g(x) \, dx.$$

With these choices, some more algebra of integrals gives an explicit formula for f:

$$f(x) = \frac{1}{\epsilon} \int_x^\infty \sin\left(\frac{x - \tau}{\epsilon}\right) g(\tau) \, d\tau.$$

But despite the uniqueness of the solution assuming the extra conditions, there is a solvability condition lurking around. Since f is even, $f'(0) = 0$, and since $c_2 = f'(0)$, we really need

$$\int_0^\infty \cos\left(\frac{x}{\epsilon}\right) g(x) \, dx = 0.$$

Certainly not all g satisfy this condition.

While these examples look very different, they all have substantial features in common. Every example asks us to solve an equation of the form $\mathcal{T}v = w$, where $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ is a linear operator, $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, $w \in \mathcal{W}$ is given, and $v \in \mathcal{V}$ is unknown. Big picture, we can add elements of $\mathcal{V}$ and multiply them by real numbers (or maybe complex numbers) to get new elements of $\mathcal{V}$, and all the arithmetic works exactly as we think it should. Same for $\mathcal{W}$. And $\mathcal{T}$ respects linearity:

$$\mathcal{T}(v_1 + v_2) = \mathcal{T}v_1 + \mathcal{T}v_2 \quad \text{and} \quad \mathcal{T}(\alpha v) = \alpha(\mathcal{T}v)$$

for all v_1, v_2, $v \in \mathcal{V}$ and $\alpha \in \mathbb{R}$. (There's a sneaky point in here in that $v_1 + v_2$ is addition in $\mathcal{V}$ and $\mathcal{T}v_1 + \mathcal{T}v_2$ is addition in $\mathcal{W}$, but basically no one cases about that.)

The challenge common to these problems is twofold. First, $\mathcal{T}$ has a nontrivial kernel: there exists $v_0 \in \mathcal{V} \setminus \{0\}$ such that $\mathcal{T}v_0 = 0$. This destroys uniqueness: if $\mathcal{T}v = w$, then

$$\mathcal{T}(v + \alpha v_0) = \mathcal{T}v + \alpha \mathcal{T}v_0 = \mathcal{T}v + 0 = \mathcal{T}v = w,$$

too. Here $\alpha \in \mathbb{R}$ is arbitrary, and since $v_0 \neq 0$, we get infinitely many more solutions $v + \alpha v_0$.

Second, the range of $\mathcal{T}$ is not all of $\mathcal{W}$: there exists $w \in \mathcal{W}$ such that $\mathcal{T}v \neq w$ for all $v \in \mathcal{V}$. And so boom, we cannot solve $\mathcal{T}v = w$ for all $w \in \mathcal{W}$. This destroys existence of solutions.

However, it is possible to find subspaces $\mathcal{V}_0$ of $\mathcal{V}$ and $\mathcal{W}_0$ of $\mathcal{W}$ such that $\mathcal{T}v \neq 0$ for all $v \in \mathcal{V}_0 \setminus \{0\}$, and such that for all $w \in \mathcal{W}_0$, there exists $v \in \mathcal{V}_0$ such that $\mathcal{T}v = w$. That is, we can solve $\mathcal{T}v = w$ uniquely for $v \in \mathcal{V}_0$ given $w \in \mathcal{W}_0$.

That being said, the first two examples have more in common with each other than they do with the third. The first two are really finite-dimensional problems, while the third is infinite-dimensional. As much as possible in this course, we will develop ideas and results for arbitrary vector spaces, regardless of dimension—and we will fail quite often. We will see how the assumption of finite-dimensionality leads to quick proofs, how results might fail in infinite dimensions, and how more structure (specifically, the structure of functional analysis) is needed to get things to work in infinite dimensions. And as much as possible, we'll do examples following a "rule of three": see it in $\mathbb{R}^n$, see it in a finite-dimensional space that is not $\mathbb{R}^n$ (but, necessarily, isomorphic to $\mathbb{R}^n$), and see it in infinite-dimensional space (typically a vector space of functions).

If none of that made sense to you, you're probably in good company. The goal of this course is to get it to make sense: to see the common linear algebraic structure underlying these seemingly disparate problems, and to see how the tools of linear algebra make cosmetically complicated problems much simpler.

Day 2: Wednesday, August 14.

Material from *Linear Algebra* by Meckes & Meckes

Pages 378–382 of Appendix A.1 contain effectively all the information on functions that we will need in the course (and some things that we don't need right now). Note that the text does not use the ordered pair definition of function, and after this introductory material we will not, either, in practice (virtually no one does).

Vocabulary from today

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

Function (N), domain, codomain, range, image of a set under a function

Functions are foundational to all of mathematics. We will need functions to define vector spaces, the primary setting in which we will work, and linear operators, the primary connection between vector spaces. Moreover, essentially all vector spaces consist of functions; we will see that column vectors and matrices are functions of "discrete" variables, while some of the most interesting infinite-dimensional vector spaces consist of functions.

Here is a first stab at the definition of function.

2.1 Undefinition. *A* FUNCTION *from a set A to a set B is a rule or operation that pairs (or associates, or maps) every element of A with one and only one element of B.*

The problem with this definition (which is why it is an undefinition) is the use of weasel words: "rule," "operation," "pairs," "associates," "maps." What do these words mean? We will make this annoyingly precise, but first we consider some examples to see how broad functions can be.

2.2 Example. The following should all be functions.

(i) The pairing of real numbers x with their doubles $2x$ is a function: every real number is paired with another number, and only one number at that.

(ii) The pairing of people in a room with the date (1 through 31) on which they were born. Everyone has only one birthday.

(iii) The pairing of people in a room with the color of the chair in which they are seated (assuming everyone is sitting in a chair and every chair has a discernible color). This last function does not involve numbers at all!

The better definition of function involves more set-theoretic machinery, specifically, the ordered pair. The idea of an ordered pair (x, y) is that another ordered pair (a, b) equals (x, y) if and only if $x = a$ and $y = b$. That is, ordered pairs are equal if and only if their corresponding components are equal—that encodes the idea of "order." It is not necessary to memorize the following definition, but it is here for completeness.

2.3 Definition. *Let x and y be elements of a set. The* ORDERED PAIR *whose first component is x and whose second component is y is the set*

$$(x, y) := \{\{x\}, \{x, y\}\}.$$

2.4 Problem (Wholly optional). Use this definition of ordered pair to prove that $(x, y) = (a, b)$ if and only if $x = a$ and $y = b$.

Now we are ready to define functions.

2.5 Definition. *Let A and B be sets. A* FUNCTION f FROM A TO B *is a set of ordered pairs with the following properties.*

(i) *If $(x, y) \in f$, then $x \in A$ and $y \in B$.*

(ii) *For each $x \in A$, there is a unique $y \in B$ such that $(x, y) \in f$.*

We often use the notation $f : A \to B$ to mean that f is a function from A to B. If $(x, y) \in f$, then we write $y = f(x)$. The set A is the DOMAIN *of f, and the set B is the*

CODOMAIN *of f. The* IMAGE *or* RANGE *of f is the set*

$$f(A) := \{f(x) \mid x \in A\}.$$

More generally, if $E \subseteq A$, then the IMAGE OF E UNDER f *is*

$$f(E) := \{f(x) \mid x \in E\}.$$

The first condition in this definition encodes the act of pairing: elements of A are paired with elements of B as ordered pairs. The second condition encodes the idea that *every* element of A is pair with *one and only one* element of B. (There is a slicker way of phrasing this definition using Cartesian products, but we can avoid that extra bit of technology for now.)

2.6 Example. Let

$$f = \{(1, -1), (2, 1), (3, -1), (4, 1)\}.$$

Then f is clearly a set of ordered pairs. We study possible domains and codomains of f.

(i) Let $A = \{1, 2, 3, 4\}$ and $B = \{1, -1\}$. Then for each $x \in A$, there is one and only one $y \in B$ such that $(x, y) \in f$, and so f is a function from A to B. Moreover, $f(A) = B$. It happens that $f(1) = f(3)$, and also $f(2) = f(4)$, but that does not violate any part of the definition of function. (It does mean that f is not one-to-one or injective, a condition that we will discuss later.)

(ii) Let $A = \{1, 2, 3\}$ and $B = \{1, -1\}$. Since $(4, 1) \in f$ but $4 \notin A$, f cannot be a function from A to B; the first condition in the definition of function is violated.

(iii) Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, -1\}$. Since $5 \in A$ but $(5, y) \notin f$ for all $y \in B$, f cannot be a function from A to B; part of the second condition in the definition of function is violated.

(iv) Let $A = \{1, 2, 3, 4\}$ and $B = \{1, -1, 0\}$. Again, for each $x \in A$, there is one and only one $y \in B$ such that $(x, y) \in f$, and so f is a function from A to B. It happens that $f(A) \neq B$, since $0 \notin f(A)$, but that does not violate any part of the definition of function. (It does mean that f is not onto or surjective, a condition that we will discuss later.)

2.7 Problem. **(i)** Why is $\{(1, -1), (1, 1), (2, 1), (3, -1), (4, 1)\}$ not a function from $\{1, 2, 3, 4\}$ to $\{1, -1\}$?

(ii) Let $f = \{(x, x^2) \mid x \in \mathbb{R}\}$. Let $I = [0, \infty)$. Show that $f(I) = I$.

(iii) Why is $\{(x, y) \mid x,\ y \in \mathbb{R} \text{ and } y^2 = x\}$ not a function from $\mathbb{R}$ to $\mathbb{R}$?

2.8 Problem (Optional but worth at least reading). Let A, B, C, and D be sets and let $f\colon A \to B$ and $g\colon C \to D$ be functions. Prove that $f = g$ if and only if $A = C$ and $f(x) = g(x)$ for all $x \in A$ (equivalently, for all $x \in C$). [Hint: *remember that f and g are sets of ordered pairs. To prove the forward implication, if $f = g$, we want to show $x \in A \iff x \in C$ and $f(x) = g(x)$ for all $x \in A$. So, take some $x \in A$ and obtain $(x, f(x)) \in g$. Why does this force $x \in C$ and $g(x) = f(x)$? To prove the reverse implication and show $f = g$, we want to establish $(x, y) \in f \iff (x, y) \in g$. If $(x, y) \in f$, why do we have $x \in A$ and thus $x \in C$? Since $f(x) = g(x)$, why does this lead to $(x, y) \in g$?*]

Life starts with sets and then we connect them with functions (which are themselves sets). Naturally, we may also want to consider sets of functions. If A and B are sets, we denote by

$$B^A$$

the set of all functions from A to B.

2.9 Example. The set $\{1, 2\}^{\{1\}}$ is the set of all functions from $\{1\}$ to $\{1, 2\}$. Any function from $\{1\}$ to $\{1, 2\}$ must be a set consisting of a single ordered pair whose first coordinate is 1 and whose second coordinate is either 1 or 2. So,

$$\{1, 2\}^{\{1\}} = \big\{\{(1, 1)\}, \{(1, 2)\}\big\}.$$

2.10 Problem. What are all the elements of $\{1, -1\}^{\{1,2,3,4\}}$? [Hint: *there are eight.*]

Now we show how functions give a rigorous definition of column vectors and matrices, the building blocks of elementary linear algebra. Consider the vector

$$\begin{bmatrix} 2 \\ 4 \\ 8 \end{bmatrix} \in \mathbb{R}^3.$$

This vector must be different from

$$\begin{bmatrix} 4 \\ 2 \\ 8 \end{bmatrix},$$

even though the same numbers appear in both—the two vectors have different *entries*. The fundamental difference is one of *ordering*: 2, 4, and 8 appear in different entries, slots, or positions between the two vectors. We might say that the first vector should mean the same as the function f from $\{1, 2, 3\}$ to $\mathbb{R}$ such that $f(1) = 2$, $f(2) = 4$, and $f(3) = 8$. Then we might say that $\mathbb{R}^3$ should be the set of all functions from $\{1, 2, 3\}$ to $\mathbb{R}$.

Consider next the matrix

$$\begin{bmatrix} 2 & 6 & 10 \\ 4 & 8 & 12 \end{bmatrix} \in \mathbb{R}^{2\times 3}.$$

This is a "two-dimensional" array of data, and so while there are six numbers in play, it will be more meaningful to describe this matrix with two directions—rows and columns. If we think of the (i, j)-entry of this matrix as the number in row i and column j (rows before columns, always), then this matrix could be the function f with

$$f(1,1) = 2, \quad f(2,1) = 4, \quad f(1,2) = 6, \quad f(2,2) = 8, \quad f(1,3) = 10, \quad \text{and} \quad f(2,3) = 12.$$

Here f is a function from $\{(i, j) \mid i = 1,\ 2,\ j = 1,\ 2,\ 3\}$ to $\mathbb{R}$.

Day 3: Friday, August 16.

Material from *Linear Algebra* by Meckes & Meckes

We are dancing around the idea of a vector space by considering addition and scalar multiplication in the function spaces $\mathbb{R}^n$ (column vectors), $\mathbb{R}^{m\times n}$ ($m \times n$ matrices), $\mathbb{R}^\infty$ (sequences = functions from $\mathbb{N}$ = natural numbers to $\mathbb{R}$), and $\mathbb{R}^X$ (functions from any set X to $\mathbb{R}$). You may want to read the discussions of column vector arithmetic on pp.24–26. The essential arithmetical properties of real numbers that we tacitly and joyfully assume appear as properties of fields on pp.39–43. Throughout this course, the only fields that we will consider are $\mathbb{R}$ and the complex numbers $\mathbb{C}$, and whenever you see the symbol $\mathbb{F}$ you can think of it as meaning either the real or complex numbers. Matrix and sequence arithmetic appear on pp.55–56 in the context of vector spaces. The book uses the notation $M_{m,n}(\mathbb{R}) = \mathbb{R}^{m\times n}$ and $M_n(\mathbb{R}) = \mathbb{R}^{n\times n}$; I will not.

Vocabulary from today

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

Sequence in a set X

After all the bluster last time about a rigorous definition for functions, we will be content with the original undefinition: if A and B are sets, a function f from A to B (denoted by $f\colon A \to B$) is a rule that pairs each $x \in A$ with a unique $y \in B$, written $y = f(x)$. To do linear algebra, we need algebra, and to do algebra, we need arithmetic. We study arithmetic with functions, which is really arithmetic with vectors.

Recall that $\mathbb{R}^n$ is the set of all functions from the set $\{1, \ldots, n\}$ to $\mathbb{R}$. If $\mathbf{x} \in \mathbb{R}^n$, then $\mathbf{x}$ is a function on $\{1, \ldots, n\}$ that takes real values. We write $\mathbf{x}(k) = x_k$ and express

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

In other words, the column vector on the right is just a (very convenient!) notational device for the function $\mathbf{x}$ on $\{1, \ldots, n\}$ that takes the values $\mathbf{x}(k) = x_k$ for $k = 1, \ldots, n$ and $x_k \in \mathbb{R}$.

We work with $\mathbb{R}^3$ for some time just for convenience. There is really only one natural way to add functions (...column vectors...) in $\mathbb{R}^3$, and that is componentwise, or entrywise:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}.$$

Abbreviate

$$\mathbf{x} := \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \mathbf{y} := \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{z} := \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}.$$

Then $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are all elements of $\mathbb{R}^3$, which is to say, real-valued functions on the set $\{1, 2, 3\}$. We say that $\mathbf{z} = \mathbf{x} + \mathbf{y}$, and what this means is that $\mathbf{z}(k) = \mathbf{x}(k) + \mathbf{y}(k)$ for all k.

In this last sentence, and in the displayed calculation above, the symbol $+$ is doing double duty. We have the familiar addition of real numbers ($x_1 + y_1$, $x_2 + y_2$, $x_3 + y_3$), and we have the addition of elements of $\mathbb{R}^3$. Perhaps we should be pedantic and give a new symbol for the latter—say that $\mathbf{x} +_{\mathbb{R}^3} \mathbf{y}$ is the function that takes the value $\mathbf{x}(k) + \mathbf{y}(k)$ at k. That is,

$$\big(\mathbf{x} +_{\mathbb{R}^3} \mathbf{y}\big)(k) = \mathbf{x}(k) + \mathbf{y}(k),$$

where on the right $+$ is the familiar addition of the real numbers $\mathbf{x}(k)$ and $\mathbf{y}(k)$. What a horrible, burdensome way of living. We will not write like this. The point, to be repeated often, is that addition of column vectors respects our hopefully intuitive addition of functions: do it pointwise, entrywise, componentwise.

We move to matrix addition. Recall that $\mathbb{R}^{m \times n}$ is the set of all functions from

$$\{(i, j) \mid i = 1, \ldots, m, \ j = 1, \ldots, n\}$$

to $\mathbb{R}$. For

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \ B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \in \mathbb{R}^{2 \times 2},$$

we expect

$$A + B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}.$$

How does this respect our notion of A and B as functions from $\{(1,1), (1,2), (2,1), (2,2)\}$ to $\mathbb{R}$? We have $A(1,1) = a_{11}$, $B(1,1) = b_{11}$, and $(A + B)(11) = a_{11} + b_{11}$. That is, $(A + B)(1,1) = A(1,1) + B(1,1)$. This perfectly respects our idea that function addition is performed pointwise.

More generally, let X be any set, and let f, $g\colon X \to \mathbb{R}$ be functions. Probably the most natural way to assign meaning to the symbol $f + g$ is to make it the function defined by $(f + g)(x) := f(x) + g(x)$. Again, we add functions pointwise.

3.1 Problem. With this definition, explain why

$$f + g = \{(x, f(x) + g(x)) \mid x \in X\}.$$

And, again, we could be horribly pedantic and write something like $f +_{\mathbb{R}^X} g$ instead of $f + g$, to emphasize that $+$ is an operation on real numbers, while $+_{\mathbb{R}^X}$ is an operation on the set of all functions from X to $\mathbb{R}$, which we are calling $\mathbb{R}^X$. We are all probably happier not doing that.

Column vectors and matrices are inherently "finite-dimensional" objects. There is a nice kind of function that retains the discrete structure of column vectors and matrices while being infinite: the sequence.

3.2 Definition. *Let $\mathbb{N}$ denote the natural numbers ($\mathbb{N} = \{1, 2, 3, \ldots\}$). A* **SEQUENCE** *in a set X is a function $f\colon \mathbb{N} \to X$. If $f(k) = x_k$, then we often write $f = (x_k)$.*

3.3 Example. Define

$$f\colon \mathbb{N} \to \mathbb{R}\colon k \mapsto k^2.$$

Then f is a sequence in $\mathbb{R}$, and we might write $f = (k^2)$. Strictly speaking, as a set of ordered pairs,

$$(k^2) = \left\{(k, k^2) \;\middle|\; k \in \mathbb{N}\right\}.$$

Sequences are hugely useful in analysis because they allow us to express "continuous concepts" (like limits and continuity) in terms of "discrete" objects (the domain of a sequence is $\mathbb{N}$ which, while infinite, is still tamer—more discrete—than $\mathbb{R}$). For us in linear algebra, sequences will serve as a pleasant source of examples that introduce infinite-dimensional complexities while largely retaining the manipulability of column vectors and matrices. In particular, since all function addition is componentwise, we should have

$$(x_k) + (y_k) = (x_k + y_k).$$

Again, we are overworking the symbol $+$, once for addition of sequences, once for addition of real numbers.

Per our prior notation, we might say that $\mathbb{R}^{\mathbb{N}}$ is the set of all real-valued sequences. Another evocative notation for this set is $\mathbb{R}^\infty$, since a sequence is morally an infinitely long vector. We will use both notations, although eventually we may want to consider "doubly infinite" sequences indexed by negative numbers, too. (Such sequences arise with Fourier coefficients, among other uses.)

We have thus far defined addition in $\mathbb{R}^n$, $\mathbb{R}^{m \times n}$, $\mathbb{R}^\infty$, and $\mathbb{R}^X$ for any set X (this last set $\mathbb{R}^X$ subsuming all prior ones). In each case, we used the intuitive notion of addition in $\mathbb{R}$ to define pointwise addition of functions. Function addition inherits many useful properties of real addition, in particular commutativity and associativity:

$$f + g = g + f \quad \text{and} \quad (f + g) + h = f + (g + h),$$

where, above, f, g, and h are real-valued functions defined on a common set. That is, the order in which we add real-valued functions should not matter, nor should the order in which we group them. Hopefully all of this perfectly respects our intuition from calculus.

Among many other useful properties of real addition is the notion of an **IDENTITY ELEMENT** for addition: the number $0 \in \mathbb{R}$ satisfies $x + 0 = x$ for all $x \in \mathbb{R}$. We can get a "zero" for function addition by defining the zero function pointwise. In the case of $\mathbb{R}^n$, we put

$$\mathbf{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

and of course $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. (Sadly, there is no universal notation for the zero matrix in $\mathbb{R}^{m \times n}$. Maybe O? Z?)

For sequences, the sequence (0) takes the value 0 at each $k \in \mathbb{N}$, which may be hard to see because there is no k-dependence in (0). But, working componentwise,

$$(x_k) + (0) = (x_k + 0) = (x_k),$$

and that is certainly what a zero sequence should do.

Most generally, for any set X, we can define

$$z \colon X \to \mathbb{R} \colon x \mapsto 0,$$

and then for any function $f \colon X \to \mathbb{R}$, we have $f(x) + z(x) = 0$ for all $x \in \mathcal{X}$. That is, z serves as the additive identity for function addition in the set $\mathbb{R}^X$. (Should we write z_X?)

There is another essential algebraic operation for linear algebra: multiplication by real numbers. (It may be surprising that we do not really multiply vectors—dot products and cross products notwithstanding—after all the function multiplication that we do in calculus.) Again, this should work componentwise. For example, in the case of $\mathbb{R}^3$, if $\alpha \in \mathbb{R}$, then

$$\alpha \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{bmatrix}.$$

For sequences,

$$\alpha(x_k) = (\alpha x_k).$$

For functions $f \colon X \to \mathbb{R}$, the symbol αf should denote the function satisfying $(\alpha f)(x) = \alpha f(x)$.

3.4 Problem. Explain why $\alpha f = \{(x, \alpha f(x)) \mid x \in X\}$.

Arithmetic with multiplying functions by real numbers works *exactly as it should*, and we will not belabor the obvious. The point of today was to build a stock of examples for what we will call vector spaces in the future, and all of these vector spaces ($\mathbb{R}^n$, $\mathbb{R}^{m \times n}$, $\mathbb{R}^\infty$, $\mathbb{R}^X$) were built on functions and properties of real numbers. We conclude with some picky comments about column and row vectors vs.matrices.

3.5 Remark. *We do not think of $\mathbb{R}^{m \times 1}$ as the set of all functions from*

$$\{(i, 1) \mid i = 1, \ldots, m\}$$

to $\mathbb{R}$, but rather we define $\mathbb{R}^{m \times 1} := \mathbb{R}^m$. We do not think of $\mathbb{R}^1$ as the set of all functions *from* $\{1\}$ *to* $\mathbb{R}$*, but rather we define* $\mathbb{R}^1 := \mathbb{R}$*. Thus* $\mathbb{R}^{1\times 1} = \mathbb{R}$*. However, for* $n \geq 2$*, we do not say* $\mathbb{R}^{1\times n} = \mathbb{R}^n$ *but instead continue to think of* $\mathbb{R}^{1\times n}$ *as the set of all functions from*

$$\{(1, j) \mid j = 1, \ldots, n\}$$

to $\mathbb{R}$*. The point is that we want to distinguish column vectors in* $\mathbb{R}^n$ *from row vectors in* $\mathbb{R}^{1\times n}$*. Most broadly, we could say*

$$\mathbb{R}^n = \begin{cases} \mathbb{R}, & n = 1 \\ \mathbb{R}^{\{1,\ldots,n\}}, & n \geq 2 \end{cases}$$

and

$$\mathbb{R}^{m\times n} = \begin{cases} \mathbb{R}, & m = n = 1 \\ \mathbb{R}^m, & n = 1 \\ \mathbb{R}^{\{(i,j) \ \mid \ i=1,\ldots,m, \ j=1,\ldots,n\}}, & n \geq 2. \end{cases}$$

Of course, no one thinks like this on a daily basis.

Day 4: Monday, August 19.

Material from *Linear Algebra* by Meckes & Meckes

The axioms for a vector space appear on p.51. The main point is the "bottom line" box, repeated on p.58. See p.50 and Proposition 1.9 for examples of vector spaces. See pp.57–58 for essential arithmetic in vector spaces that follows from the axioms. Do Quick Exercise #21 on p.52.

Vocabulary from today

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

Zero vector, additive inverse. I will *not* ask you to define what a vector space is or memorize the axioms.

Be able to explain what the zero vector and the additive inverse *do*. Don't just say that the zero vector is 0 or $\mathbf{0}$ or $0_{\mathcal{V}}$; rather, $0_{\mathcal{V}} + v = v$ for all $v \in \mathcal{V}$. Likewise, the additive inverse isn't $-v$ but rather the unique vector w such that $v + w = 0_{\mathcal{V}}$.

The function sets $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$, $\mathbb{R}^\infty$, and $\mathbb{R}^X$ (with X an arbitrary set) that we considered are all, of course, examples of vector spaces—and in some sense they will be almost all of the vector spaces that we study. For functions f and g in any one of these sets, we define $f + g$ and αf (for $\alpha \in \mathbb{R}$) pointwise (or componentwise) and we get a new function in that set. (While we can naturally multiply functions pointwise, and we do so all the time in calculus, we will not typically do so in this course.) All of the arithmetic works exactly as it should, because all of the arithmetic is inherited from arithmetic on $\mathbb{R}$, and all arithmetic on $\mathbb{R}$ works as it should.

We now abstract from these situations the absolute essentials of the structure—the properties without which we cannot do anything worthwhile and from which we can prove everything worthwhile. First, we take the convention that the symbol $\mathbb{F}$ represents either $\mathbb{R}$ or $\mathbb{C}$, with $\mathbb{C} := \{a + ib \mid a,\ b \in \mathbb{R},\ i^2 = -1\}$.

4.1 Definition. *A* VECTOR SPACE OVER $\mathbb{F}$ *consists of four objects:* $\mathcal{V}$, $\mathbb{F}$, $+_{\mathcal{V}}$, *and* $\cdot$, *which are described below.*

- $\mathcal{V}$ *is a nonempty set.*
- $\mathbb{F}$ *is either* $\mathbb{R}$ *or* $\mathbb{C}$.
- *For each* $v,\ w \in \mathcal{V}$, *there exists* $v +_{\mathcal{V}} w \in \mathcal{V}$, *which satisfies the axioms below.*
- *For each* $\alpha \in \mathbb{F}$ *and* $v \in \mathcal{V}$, *there exists* $\alpha \cdot v \in \mathcal{V}$, *which satisfies the axioms below.*

That is, $+_{\mathcal{V}}$ *is a function from* $\{(v, w) \mid v,\ w \in \mathcal{V}\}$ *to* $\mathcal{V}$ *and* $\cdot$ *is a function from* $\{(\alpha, v) \mid \alpha \in \mathbb{R},\ v \in \mathcal{V}\}$ *to* $\mathcal{V}$. *We call* $+_{\mathcal{V}}$ VECTOR ADDITION *and* $\cdot$ SCALAR MULTIPLICATION. *Often we abuse terminology and call just* $\mathcal{V}$ *the vector space.*

Vector addition and scalar multiplication satisfy the following axioms.

Axioms for vector addition.

1. *Commutativity:* $v +_{\mathcal{V}} w = w +_{\mathcal{V}} v$ *for all* $v,\ w \in \mathcal{V}$.

2. *Associativity:* $v +_{\mathcal{V}} (w +_{\mathcal{V}} u) = (v +_{\mathcal{V}} w) +_{\mathcal{V}} u$ *for all* $v,\ w,\ u \in \mathcal{V}$.

3. *Identity: there exists* $0_{\mathcal{V}} \in \mathcal{V}$ *such that* $v + 0_{\mathcal{V}} = v$ *for all* $v \in \mathcal{V}$.

4. *Inverse: for each* $v \in \mathcal{V}$, *there exists* $-v \in \mathcal{V}$ *such that* $v +_{\mathcal{V}} (-v) = 0_{\mathcal{V}}$.

Axioms for scalar multiplication.

5. *Identity:* $1 \cdot v = v$ *for all* $v \in \mathcal{V}$.

6. *Associativity:* $\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v$ *for all* $\alpha,\ \beta \in \mathbb{F}$ *and* $v \in \mathcal{V}$.

Axioms relating vector addition and scalar multiplication.

7. *Distributivity:* $(\alpha + \beta) \cdot v = (\alpha \cdot v) +_{\mathcal{V}} (\beta \cdot v)$ *for all* $\alpha,\ \beta \in \mathbb{F}$ *and* $v \in \mathcal{V}$.

8. *Distributivity again:* $\alpha \cdot (v +_{\mathcal{V}} w) = (\alpha \cdot v) +_{\mathcal{V}} (\beta \cdot w)$ *for all* $\alpha \in \mathbb{F}$ *and* $v,\ w \in \mathcal{V}$.

We discuss at length these foundational, essential axioms.

4.2 Remark. *The grouping of the axioms is taken from Strang's Introduction to Linear Algebra. The phrase that a vector space "consists of four objects" is weasel words; really, we might think of a vector space as an "ordered 4-tuple"* $(\mathcal{V}, \mathbb{F}, +_{\mathcal{V}}, \cdot)$, *where* $\mathcal{V}$ *is a nonempty set,* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *and* $+_{\mathcal{V}}$ *and* $\cdot$ *are the maps above. (And here an ordered* 4*-tuple* (a, b, c, d) *is the function* $f: \{1, 2, 3, 4\} \to \{a, b, c, d\}$ *with* $f(1) = a$, $f(2) = b$, $f(3) = c$, *and* $f(4) = d$.*) Of course, no one ever does this in public.*

Here are some more focused comments on the axioms.

(i) *Commutativity of vector addition means that the order in which we add vectors is irrelevant. (Mathematicians are typically uncomfortable using the plus symbol for something that does not commute.)*

(ii) *Associativity of vector addition means that the way in which we group vectors is irrelevant for addition.*

(iii) *We will shortly show that the zero vector is unique and therefore merits the definite article "the."*

(iv) *The symbol* $-v$ *for the additive inverse is just that: a symbol. It is possible to prove that given* $v \in \mathcal{V}$, *there is only one vector* $w \in \mathcal{V}$ *such that* $v +_{\mathcal{V}} w = 0_{\mathcal{V}}$. *It is also possible to show that* $-v = (-1) \cdot v$; *that is, there is an intimate, and expected, connection between the additive inverse in* $\mathcal{V}$ *and scalar multiplication by the additive inverse of the multiplicative identity in* $\mathbb{F}$.

(v) *For associativity of scalar multiplication, given* α, $\beta \in \mathbb{F}$ *and* $v \in \mathcal{V}$, *we obtain* $\beta \cdot v \in \mathcal{V}$ *and thus* $\alpha \cdot (\beta \cdot v) \in \mathcal{V}$. *But we also have* $\alpha\beta \in \mathbb{F}$, *where juxtaposition of* α *and* β *here indicates their product according to arithmetic in* $\mathbb{F}$, *and so we have* $(\alpha\beta) \cdot v \in \mathcal{V}$. *Associativity of scalar multiplication asserts that these two instances of multiplication are really the same, as we would expect.*

(vi) *The first distributive axiom illustrates why we might want to decorate vector addition as* $+_{\mathcal{V}}$. *On the left,* $\alpha + \beta$ *is addition of numbers in* $\mathbb{F}$, *while on the right* $(\alpha \cdot v) +_{\mathcal{V}} (\beta \cdot v)$ *is vector addition of the vectors* $\alpha \cdot v$ *and* $\beta \cdot v$ *in* $\mathcal{V}$.

4.3 Example. We prove some consequences of the vector space axioms using *only* these axioms. Below, $\mathcal{V}$ is a vector space over $\mathbb{F}$.

(i) First we show that the zero vector is unique. The axioms tell us that there exists $0_{\mathcal{V}} \in \mathcal{V}$ such that $v +_{\mathcal{V}} 0_{\mathcal{V}} = v$ for all $v \in \mathcal{V}$. Suppose there is another vector that does this. ("What things do defines what things are.") That is, suppose that $w \in \mathcal{V}$ also satisfies $v +_{\mathcal{V}} w = v$ for all $v \in \mathcal{V}$. We can make $0_{\mathcal{V}}$ "talk" to w by taking $v = 0_{\mathcal{V}}$ in the previous equality; after all, it holds for all $v \in \mathcal{V}$. Then we get $0_{\mathcal{V}} +_{\mathcal{V}} w = 0_{\mathcal{V}}$. By the vector space axioms, $0_{\mathcal{V}} +_{\mathcal{V}} w = w$. Thus $w = 0_{\mathcal{V}}$.

(ii) Next we show the useful fact that if $\alpha \in \mathbb{F}$ and $v \in \mathcal{V}$ with $\alpha \cdot v = 0_{\mathcal{V}}$, then either $\alpha = 0$ or $v = 0_{\mathcal{V}}$. By $\alpha = 0$ we mean the number $0 \in \mathbb{F}$. This is a statement of the form

"$P \implies Q$ or R," and such statements are logically equivalent to both "P and not $Q \implies R$" and P and not $R \implies Q$." We work with the first version: assume that $\alpha \cdot v = 0_{\mathcal{V}}$ but $\alpha \neq 0$. Then α has a reciprocal $\alpha^{-1} \in \mathbb{F}$, which satisfies $\alpha^{-1}\alpha = 1$. Here the juxtaposition $\alpha^{-1}\alpha$ means multiplication in $\mathbb{F}$.

So, from $\alpha \cdot v = 0_{\mathcal{V}}$, we have $\alpha^{-1} \cdot (\alpha \cdot v) = \alpha^{-1} \cdot 0_{\mathcal{V}}$. On the left, we use vector space axioms to rewrite

$$\alpha^{-1} \cdot (\alpha \cdot v) = (\alpha^{-1}\alpha) \cdot v = 1 \cdot v = v.$$

On the right, we actually need to do a little more work and prove separately that $\beta \cdot 0_{\mathcal{V}} = 0_{\mathcal{V}}$ for all $\beta \in \mathbb{F}$. (The axioms do not explicitly tell us anything about how the additive identity for vector addition interacts with scalar multiplication.) Assuming this to be true, we get $\alpha^{-1} \cdot 0_{\mathcal{V}} = 0_{\mathcal{V}}$ and thus $v = 0_{\mathcal{V}}$.

4.4 Problem. In the example above, try proving that if $\alpha \cdot v = 0_{\mathcal{V}}$ and $v \neq 0_{\mathcal{V}}$, then $\alpha = 0$. How far do you get? Is this any harder than our approach?

Typically we will not use $+_{\mathcal{V}}$ or $\cdot$ anymore. That is, we write $v +_{\mathcal{V}} w = v + w$ and $\alpha \cdot v = \alpha v$. Also, we will write 0 instead of $0_{\mathcal{V}}$ and $\mathbf{0}$ for the zero vector in $\mathbb{R}^n$. As needed, we may include subscripts for clarity. For example, we have shown that if $\alpha v = 0$ for $\alpha \in \mathbb{F}$ and $v \in \mathcal{V}$, then either $\alpha = 0$ or $v = 0$; context tells us that the 0 appearing in $\alpha v = 0$ and $v = 0$ is the zero vector, while the 0 appearing in $\alpha = 0$ is the scalar 0.

We will adhere to the conventions that letters like u, v, and w denote elements of a vector space, while α, β, c, and d denote elements of $\mathbb{F}$ (with more letters in each case as needed from earlier or later in the alphabet). We will reserve boldface letters for column vectors in $\mathbb{R}^n$ (they deserve this special treatment because they have scalar components, and because they are just so special to linear algebra).

Day 5: Wednesday, August 21.

Material from *Linear Algebra* by Meckes & Meckes

Pages 55–57 have many examples of vector spaces. Our notation in this log and in class uses $\mathcal{C}^k([a,b])$ to mean what the book calls $D^k[a,b]$ on p.57. The book also adopts a more formal/algebraic view of polynomials (p.57 again) than our view of polynomials as functions on $\mathbb{R}$ or $\mathbb{C}$.

We study a number of examples, and nonexamples, of vector spaces. Most really arise as *subspaces* of other vector spaces, and we will discuss that presently. Here is the general situation. One starts with a vector space $\mathcal{W}$ that is often "too large to be interesting." For example, for an arbitrary set X, the set of all functions $\mathbb{F}^X$ is a vector space (as we review momentarily), but rarely in life do we consider *all* functions from X to $\mathbb{F}$; in calculus, for example, we really only care about continuous and differentiable functions. Instead, we find a subset $\mathcal{V} \subseteq \mathcal{W}$ that is also a vector space when vector addition and scalar multiplication are restricted to $\mathcal{V}$, i.e., $v + w \in \mathcal{V}$ and $\alpha v \in \mathcal{V}$ for all v, $w \in \mathcal{V}$ and $\alpha \in \mathbb{F}$, and $0 \in \mathcal{V}$. Often

$\mathcal{V}$ is chosen to be "small enough to be interesting."

5.1 Example. As we have said repeatedly, $\mathbb{F}^X$ is a vector space with vector addition and scalar multiplication defined pointwise (or componentwise, or entrywise). Axioms like commutativity, associativity, and distributivity follow from arithmetic in $\mathbb{F}$. We emphasize that the zero vector is the function

$$z \colon X \to \mathbb{F} \colon x \mapsto 0.$$

That is,

$$z = \{(x, 0) \mid x \in X\},$$

and sometimes we just write 0 instead of z (which could lead to unfortunate expressions like $0 = \{(x, 0) \mid x \in X\}$ if we overthink things—better than underthinking them?) The additive inverse of $f \in \mathbb{F}^X$ is the function $-f$ defined pointwise by

$$-f \colon X \to \mathbb{F} \colon x \mapsto -f(x).$$

For practice in picky reading comprehension, we emphasize that the symbol $-f$ denotes one function, while $-f(x)$ is the additive inverse in $\mathbb{F}$. Thus

$$-f = \{(x, -f(x)) \mid x \in X\}.$$

Of course, $-f(x) = -1 \cdot f(x)$ as multiplication in $\mathbb{F}$.

Here are the most important function spaces for calculus and differential equations.

5.2 Example. Let $I \subseteq \mathbb{R}$ be an interval, which may be open or closed, bounded or unbounded. (Most of calculus is done on intervals, after all.)

(i) Let

$$\mathcal{C}(I) := \left\{ f \in \mathbb{R}^I \;\middle|\; f \text{ is continuous on } I \right\}.$$

Much of calculus works because *limits are linear*: if f, $g \in \mathbb{R}^I$ and $x_0 \in I$ and

$$\lim_{x \to x_0} f(x) \quad \text{and} \quad \lim_{x \to x_0} g(x)$$

exist, then

$$\lim_{x \to x_0} \big(f(x) + g(x)\big)$$

exists with

$$\lim_{x \to x_0} \big(f(x) + g(x)\big) = \lim_{x \to x_0} f(x) + \lim_{x \to x_0} g(x).$$

Likewise, for any $\alpha \in \mathbb{R}$,

$$\lim_{x \to x_0} \alpha f(x) = \alpha \lim_{x \to x_0} f(x).$$

Since $f \in \mathcal{C}(I)$ if and only if

$$\lim_{x \to x_0} f(x) = f(x_0)$$

for all $x_0 \in I$ (ignoring the possible particular cases of left/right limits at the endpoints of I, if an endpoint even belongs to I), linearity of limits implies $f + g \in \mathcal{C}(I)$ and $\alpha f \in \mathcal{C}(I)$ for f, $g \in \mathcal{C}(I)$ and $\alpha \in \mathbb{R}$. Last, since all constant functions are continuous, the zero function 0 (defined by $0(x) = 0$ for all $x \in I$) is an element of $\mathcal{C}(I)$. Thus $\mathcal{C}(I)$ is a vector space (over $\mathbb{R}$).

(ii) Let $r \geq 0$ be an integer and let

$$\mathcal{C}^r(I) := \left\{ f \in \mathbb{R}^I \;\middle|\; f \text{ is } r\text{-times differentiable on } I \text{ and } f^{(r)} \text{ is continuous on } I \right\}.$$

Here $f^{(k)}$ is the kth derivative of f; for example, $f^{(3)} = f'''$. We put

$$\mathcal{C}^0(I) := \mathcal{C}(I)$$

since $f^{(0)} = f$. Each $\mathcal{C}^r(I)$ is a vector space because differentiation is linear: if f and g are differentiable, then $(f+g)' = f' + g'$ and $(\alpha f)' = \alpha f'$. We impose the requirement that $f^{(r)}$ be continuous mostly for "mathematical niceness," e.g., in differential equations, one often wants that the solution to an rth order differential equation has a continuous rth derivative.

5.3 Example. Let

$$\mathcal{C}^\infty(I) := \cap_{r=0}^{\infty} \mathcal{C}^r(I) = \left\{ f \in \mathbb{R}^I \;\middle|\; f \in \mathcal{C}^r(I) \text{ for all } r \geq 1 \right\}.$$

We call functions in $\mathcal{C}^\infty(I)$ **INFINITELY DIFFERENTIABLE**. Prove that $\mathcal{C}^\infty(I)$ is a vector space.

Here are some "finite-dimensional" (non)examples.

5.4 Example. **(i)** Let

$$\mathcal{V} := \left\{ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \;\middle|\; x_1 \in \mathbb{R} \right\}.$$

We claim that $\mathcal{V}$ is a vector space with the usual componentwise vector addition and scalar multiplication. We just check

$$\begin{bmatrix} x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} y_1 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ 0 + 0 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ 0 \end{bmatrix} \in \mathcal{V} \quad \text{and} \quad \alpha \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha \cdot 0 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ 0 \end{bmatrix} \in \mathcal{V}.$$

What is critical here is that vector addition and scalar multiplication keep 0 as the second component. Also,

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \mathcal{V},$$

again because, critically, the second component is 0.

(ii) The set

$$\mathcal{W} := \left\{ \begin{bmatrix} x_1 \\ 1 \end{bmatrix} \;\middle|\; x_1 \in \mathbb{R} \right\}$$

is not a vector space with, again, the usual componentwise vector addition and scalar multiplication. We only need to break one of the axioms, but we show that many fail.

We probably expect that $\mathcal{W}$ is not closed under addition because the second component will have us adding $1+1=2$, which destroys the 1 in the second component. To be explicit, we give a concrete example of how this fails:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \in \mathcal{W} \quad \text{but} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1+2 \\ 1+1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \notin \mathcal{W}.$$

Next, we probably expect that $\mathcal{W}$ is not closed under scalar multiplication because the second component will have us multiplying $\alpha \cdot 1 = \alpha \neq 1$ when $\alpha \neq 1$. To be explicit, we give a concrete example of how this fails:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathcal{W} \quad \text{but} \quad 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \notin \mathcal{W}.$$

Finally, $\mathcal{W}$ lacks an additive identity for vector addition. The only possible additive identity is the zero vector in $\mathbb{R}^2$, and

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \notin \mathcal{W},$$

because of that unpleasant second component.

We go "infinite-dimensional" again.

5.5 Example. Denote by ℓ^∞ the set of all **BOUNDED** sequences:

$$\ell^\infty := \{(a_k) \in \mathbb{R}^\infty \mid \exists M > 0 \; \forall k \in \mathbb{N} : |a_k| \le M\}.$$

For example, if $a_k = 1$ for all k, then $|a_k| \le 1$ for all k, and so $(a_k) \in \ell^\infty$. Likewise, if $b_k = 1/2^k$ for all k, then $|b_k| \le 1/2$ for all k, and so $(b_k) \in \ell^\infty$.

We show that ℓ^∞ is a vector space over $\mathbb{R}$. The zero sequence (0) is certainly an element of ℓ^∞, since $|0| < 1$. (Remember that (0) is the map $\mathbb{N} \to \mathbb{R} \colon k \mapsto 0$.)

Next we check scalar multiplication. Let $\alpha \in \mathbb{R}$ and $(a_k) \in \ell^\infty$. We want to show $\alpha(a_k) \in \ell^\infty$, and we know $\alpha(a_k) = (\alpha a_k)$. Our goal, therefore, is to find $M > 0$ such that $|\alpha a_k| \le M$ for all k. Since $(a_k) \in \ell^\infty$, we know there is $N > 0$ such that $|a_k| \le N$ for all k. Now we need a property of absolute value:

$$|xy| = |x||y|, \;\; x, \;\; y \in \mathbb{R}.$$

Then $|\alpha a_k| = |\alpha||a_k| \le |\alpha|N$. Taking $M = |\alpha|N$ is the bound we want.

Last, we check vector addition. Let (a_k), $(b_k) \in \ell^\infty$. We want to show $(a_k) + (b_k) \in \ell^\infty$, and we know $(a_k) + (b_k) = (a_k + b_k)$. Our goal, therefore, is to find $M > 0$ such that

$|a_k + b_k| \le M$ for all k. Since (a_k), $(b_k) \in \ell^\infty$, we know there are M_1, $M_2 > 0$ such that $|a_k| \le M_1$ and $|b_k| \le M_2$. Now we need another property of absolute value (the **TRIANGLE INEQUALITY**):

$$|x + y| \le |x| + |y|, \ x, \ y \in \mathbb{R}.$$

Then $|a_k + b_k| \le |a_k| + |b_k| \le M_1 + M_2$. Taking $M = M_1 + M_2$ is the bound we want.

By the way, everything above works if we replace $\mathbb{R}$ with $\mathbb{C}$, since the modulus on $\mathbb{C}$ is multiplicative and enjoys the triangle inequality.

5.6 Problem. Limits of sequences behave exactly as we expect. First, if $(a_k) \in \mathbb{R}^\infty$ and $L \in \mathbb{R}$, we say that $\lim_{k\to\infty} a_k = L$ if we can make a_k arbitrarily close to L by taking k sufficiently large. It follows that if $\lim_{k\to\infty} a_k = L_1$ and $\lim_{k\to\infty} b_k = L_2$, then $\lim_{k\to\infty}(a_k + b_k) = L_1 + L_2$ and $\lim_{k\to\infty} \alpha a_k = \alpha L_1$.

(i) Prove that

$$\mathcal{V} := \left\{(a_k) \in \mathbb{R}^\infty \ \middle| \ \lim_{k\to\infty} a_k \text{ exists}\right\}$$

is a vector space.

(ii) Prove that

$$c_0 := \left\{(a_k) \in \mathbb{R}^\infty \ \middle| \ \lim_{k\to\infty} a_k = 0\right\}$$

is a vector space (the notation c_0 is unfortunate, as it looks like a coefficient in some sum, but traditional).

(iii) Prove that

$$\mathcal{V}_\alpha := \left\{(a_k) \in \mathbb{R}^\infty \ \middle| \ \lim_{k\to\infty} a_k = \alpha\right\}$$

is not a vector space when $\alpha \ne 0$. Explain *all* of the ways in which $\mathcal{V}_\alpha$ fails to be a vector space.

5.7 Problem. So far, we have not paid too much attention to the field over which we are considering our vector spaces. Explain why $\mathbb{R}$ is a vector space over the field $\mathbb{R}$, $\mathbb{C}$ is a vector space over both $\mathbb{R}$ and $\mathbb{C}$, but $\mathbb{R}$ is not a vector space over $\mathbb{C}$.

Here is a finite-dimensional example in disguise. We begin with a slightly atypical definition (caution: your experience in other texts and algebra classes may radically differ).

5.8 Definition. *A* **POLYNOMIAL ON** $\mathbb{F}$ *is a function of the form*

$$p \colon \mathbb{F} \to \mathbb{F} \colon x \mapsto \sum_{k=0}^{n} a_k x^k$$

for some integer $n \ge 0$ *and some coefficients* $a_0, \ldots, a_n \in \mathbb{F}$. *If* $a_n \ne 0$, *the* **DEGREE** *of* p *is* $\deg(p) := n$.

Informally, a polynomial is just a sum of multiples of nonnegative integer powers of x.

5.9 Example. **(i)** Denote (again, atypically) by $\mathbb{P}(\mathbb{F})$ the set of all polynomial functions on $\mathbb{F}$. Here is the informal proof that $\mathbb{P}(\mathbb{F})$ is a vector space: adding polynomials results in polynomials, and multiplying polynomials by constants results in polynomials. To be a little more precise about the zero vector (function), note that $p(x) = 0$ for all x is a polynomial with $n = 0$ and $a_0 = 0$.

(ii) Let

$$\mathcal{V} := \{p \in \mathbb{P}(\mathbb{F}) \mid \deg(p) = 2\}.$$

Then $\mathcal{V}$ is not a vector space. First, it has no zero vector, since $\deg(0) = 0$. Next, $\mathcal{V}$ is not closed under vector addition, as we could subtract quadratics and get a linear or constant polynomial. To be concrete, with $p(x) = 3x^2$ and $q(x) = -3x^2 + x$, we have $p, q \in \mathcal{V}$ but $(p + q)(x) = x$, thus $p + q \notin \mathcal{V}$. Finally, $\mathcal{V}$ is not closed under scalar multiplication, as $0p = 0 \notin \mathcal{V}$ for all $p \in \mathcal{V}$. However, $\alpha p \in \mathcal{V}$ for all $\alpha \neq 0$ and $p \in \mathcal{V}$. This shows that sometimes we must be very precise in how we break the vector space axioms.

5.10 Problem. In contrast to the last example,

$$\mathcal{V} := \{p \in \mathbb{P}(\mathbb{F}) \mid \deg(p) \leq 2\},$$

is a vector space. Prove that.

5.11 Problem. We have now met two of the three kinds of vector spaces from our original three motivating problems. Certainly $\mathbb{R}^5$ and $\mathbb{R}^3$ appeared in Example 1.1, while the space of (at most) quadratics from Problem 5.10 above appeared in Example 1.2. Prove that

$$\mathcal{V} := \left\{ f \in \mathcal{C}^2(\mathbb{R}) \;\middle|\; f \text{ is even, } \lim_{x \to \infty} f(x) = 0 \right\}$$

and

$$\mathcal{W} := \left\{ f \in \mathcal{C}(\mathbb{R}) \;\middle|\; f \text{ is even, } \int_0^\infty |f(x)| \, dx < \infty \right\}$$

are vector spaces; these appeared in Example 1.4. [Hint: *for $\mathcal{W}$, use the triangle inequality and the comparison test for improper integrals (look up the comparison test as needed) to establish that if f and g are absolutely integrable on $[0, \infty)$, then so is $f + g$.*]

Day 6: Friday, August 23.

Material from *Linear Algebra* by Meckes & Meckes

Linear combinations are defined on p.53 and subspaces on p.55. See also the discussion of subspaces on pp.58–59. Read the philosophy on p.63 about "recognizing sameness" and pp.64–65 on linear operators. See Examples 1, 2, and 3 of linear operators on pp.86–87.

Do Quick Exercises #22, #23, #24 in Section 1.5 and #1 in Section 2.1.

Vocabulary from today

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

Subspace (N), linear combination, span, linear operator (N)

We have seen that vector spaces can arise as nonempty subsets of larger vector spaces that are closed under addition and scalar multiplication. Often the original vector space is "too large" to be interesting, like $\mathbb{R}^{\mathbb{R}}$, while the "smaller" space is more restrictive and has nicer features, like $\mathcal{C}(\mathbb{R})$. Depending on our point of view, we may still want to relate the smaller space to the original space via the notion of subspace.

6.1 Definition. *Let $\mathcal{W}$ be a vector space over $\mathbb{F}$. A set $\mathcal{V} \subseteq \mathcal{W}$ is a* SUBSPACE *of $\mathcal{W}$ if the following hold.*

(i) *$\mathcal{V}$ contains the zero vector: $0 \in \mathcal{V}$.*

(ii) *$\mathcal{V}$ is closed under addition: $v + w \in \mathcal{V}$ for all v, $w \in \mathcal{V}$.*

(iii) *$\mathcal{V}$ is closed under scalar multiplication: $\alpha v \in \mathcal{V}$ for all $\alpha \in \mathbb{F}$, $v \in \mathcal{V}$.*

All of our examples of vector spaces so far have been subspaces of vector spaces like $\mathbb{F}^n$, $\mathbb{R}^I$ for some interval $I \subseteq \mathbb{R}$, or $\mathbb{R}^\infty$. However, plenty of subsets of a vector space fail to be subspaces for violating one or more of the subspace axioms above.

6.2 Problem. Let $\mathcal{V}$ be a vector space and let $v_0 \in \mathcal{V} \setminus \{0\}$. Explain all of the ways in which $\{v_0\}$ fails to be a subspace of $\mathcal{V}$.

Nonetheless, every subset of a vector space "generates" a subspace via the following important interaction between vector addition and scalar multiplication.

6.3 Definition. *Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$.*

(i) *A* **LINEAR COMBINATION** *of the vectors* $v_1, \ldots, v_n \in \mathcal{V}$ *is a vector of the form*

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = \sum_{k=1}^{n} \alpha_k v_k$$

for some $\alpha_k \in \mathbb{F}$.

(ii) *Let* $\mathcal{B} \subseteq \mathcal{V}$ *be nonempty. (This set* $\mathcal{B}$ *need not be a subspace of* $\mathcal{V}$*.) The* **SPAN** *of* $\mathcal{B}$ *is the set of all linear combinations of vectors in* $\mathcal{B}$*. That is,*

$$\operatorname{span}(\mathcal{B}) := \left\{ \sum_{k=1}^{n} \alpha_k v_k \;\middle|\; \alpha_1, \ldots, \alpha_n \in \mathbb{F},\; v_1, \ldots, v_n \in \mathcal{V}, n \in \mathbb{N} \right\}.$$

If $\mathcal{B}$ *is a finite set, say,* $\mathcal{B} = \{w_j\}_{j=1}^{m}$*, then we write*

$$\operatorname{span}(\mathcal{B}) = \operatorname{span}(\{w_j\}_{j=1}^{m}) = \operatorname{span}(w_1, \ldots, w_m)$$

and omit the curly braces.

By the way, the sigma notation here reads as follows. If $w_1, \ldots, w_n \in \mathcal{V}$, then we define $\sum_{k=1}^{n} w_k$ recursively as

$$\sum_{k=1}^{n} w_k := \begin{cases} w_1, & n = 1 \\ w_n + \sum_{k=1}^{n-1} w_k, & n \geq 2. \end{cases}$$

Sometimes we may start the sum at an index other than 1. If $w_m, \ldots, w_n \in \mathcal{V}$ with $m \leq n$, put

$$\sum_{k=m}^{n} w_k := \sum_{k=1}^{n-(m-1)} w_{k+(m-1)}.$$

Last, we adopt the useful convention that the "empty sum" is the zero vector: if $m > n$, then

$$\sum_{k=m}^{n} w_k := 0.$$

6.4 Theorem. *Let* $\mathcal{V}$ *be a vector space over the field* $\mathbb{F}$ *and let* $\mathcal{B} \subseteq \mathcal{V}$ *be nonempty. Then* $\operatorname{span}(\mathcal{B})$ *is a subspace of* $\mathcal{V}$.

Proof. We prove this in the extremely special case when $\mathcal{B} = \{v_1, v_2\}$ to avoid too much sigma notation. We have

$$\operatorname{span}(\mathcal{B}) = \operatorname{span}(v_1, v_2) = \{\alpha_1 v_1 + \alpha_2 v_2 \mid \alpha_1,\ \alpha_2 \in \mathbb{F}\}.$$

First we check that the zero vector is in $\operatorname{span}(\mathcal{B})$. Since we are free to pick α_1 and α_2 to be any numbers in $\mathbb{F}$, we set them to be 0, so $0v_1 + 0v_2 \in \operatorname{span}(\mathcal{B})$. And $0v_1 + 0v_2 = 0 + 0 = 0$. (The first two instances of 0 in the preceding sentence were scalars; the last two were the zero vector in $\mathcal{V}$.)

Next we check closure under scalar multiplication. Given α_1, $\alpha_2 \in \mathbb{F}$, we want to show $\alpha(\alpha_1 v_1 + \alpha_2 v_2) \in \operatorname{span}(\mathcal{B})$. We just distribute the multiplication and rearrange parentheses:

$$\alpha(\alpha_1 v_1 + \alpha_2 v_2) = (\alpha\alpha_1)v_1 + (\alpha\alpha_2)v_2.$$

This is certainly a linear combination of v_1 and v_2, since $\alpha\alpha_1$, $\alpha\alpha_2 \in \mathbb{F}$. And so $\alpha(\alpha_1 v_1 + \alpha_2 v_2) \in \operatorname{span}(\mathcal{B})$.

Last, we check closure under vector addition. We want to show

$$(\alpha_1 v_1 + \alpha_2 v_2) + (\beta_1 v_1 + \beta_2 v_2) \in \mathbb{F}$$

for any α_1, α_2, β_1, $\beta_2 \in \mathbb{F}$. We do some arithmetic:

$$(\alpha_1 v_1+\alpha_2 v_2)+(\beta_1 v_1+\beta_2 v_2) = (\alpha_1 v_1+\beta_1 v_1)+(\alpha_2 v_2+\beta_2 v_2) = (\alpha_1+\beta_1)v_1+(\alpha_2+\beta_2)v_2 \in \operatorname{span}(\mathcal{B}).$$

Here we "factored" not the scalars but the vectors v_1 and v_2, using distributive properties of vector space arithmetic. ■

We "live" in vector spaces in linear algebra; all of our interesting and relevant questions can somehow be posed (and solved?) using the vector space structure, possibly gussied up with additional features (like norms and inner products). However, to state those problems exactly, we need to be able to "move between" vector spaces in a way that "respects" the vector space operations (of vector addition and scalar multiplication). We also need an instrument to determine when two similar-looking vector spaces are really the same. We achieve this via the following.

6.5 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces (over the same field $\mathbb{F}$). A function $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$ is a* **LINEAR OPERATOR** *(or* **LINEAR MAP** *or* **LINEAR TRANSFORMATION***) if $\mathcal{T}$ satisfies the following two properties.*

1. *Additivity.* $\mathcal{T}(v + w) = \mathcal{T}(v) + \mathcal{T}(w)$ *for all* v, $w \in \mathcal{V}$.

2. *Homogeneity.* $\mathcal{T}(\alpha v) = \alpha\mathcal{T}(v)$ *for all* $\alpha \in \mathbb{F}$ *and* $v \in \mathcal{V}$.

Note that the vector space operations on both sides of the two equals signs in the above are different. The addition in $\mathcal{T}(v + w)$ is addition in $\mathcal{V}$, while the addition in $\mathcal{T}(v) + \mathcal{T}(w)$ is addition in $\mathcal{W}$. The scalar multiplication in $\mathcal{T}(\alpha v)$ is scalar multiplication in $\mathcal{V}$, while the scalar multiplication in $\alpha\mathcal{T}(v)$ is scalar multiplication in $\mathcal{W}$.

We start by working through many examples of linear operators.

6.6 Example. The map

$$\mathcal{T}\colon \mathbb{R} \to \mathbb{R}\colon v \mapsto 2v$$

is linear. First we check homogeneity:

$$\mathcal{T}(\alpha v) = 2(\alpha v) = \alpha(2v) = \alpha\mathcal{T}(v)$$

Next we check additivity:

$$\mathcal{T}(v+w) = 2(v+w) = 2v + 2w = \mathcal{T}(v) + \mathcal{T}(w).$$

6.7 Problem. Generalize the preceding example vastly by showing that scalar multiplication is a linear operator. More precisely, assume that $\mathcal{V}$ is a vector space over $\mathbb{F}$, and do the following.

(i) Fix $\lambda \in \mathbb{F}$ and show that the map

$$\mathcal{T} \colon \mathcal{V} \to \mathcal{V} \colon v \mapsto \lambda v$$

is a linear operator.

(ii) By choosing λ above appropriately, show that the map

$$I \colon \mathcal{V} \to \mathcal{V} \colon v \mapsto v$$

is linear. This is the **IDENTITY OPERATOR**.

(iii) By choosing λ above appropriately, show that the map

$$\mathcal{T} \colon \mathcal{V} \to \mathcal{V} \colon v \mapsto 0$$

is linear. This is the **ZERO OPERATOR**. (Outside this problem, we will typically call it 0 and thus vastly overwork that one poor symbol by making it denote an element of $\mathbb{F}$, an element of $\mathcal{V}$, and a linear operator simultaneously.)

6.8 Problem. Let $\mathcal{V}$ be a vector space and let $v_0 \in \mathcal{V} \setminus \{0\}$. Prove that the "constant"

$$\mathcal{T} \colon \mathcal{V} \to \mathcal{V} \colon v \mapsto v_0$$

is not linear. Explain why $\mathcal{T}$ fails both additivity and homogeneity. This problem shows that every vector space has a "nonlinear" operator defined on it.

Most of the familiar operations from calculus yield linear operators, since *limits are linear*.

6.9 Example. If $f \in \mathcal{C}^1([0,1])$, then f is differentiable and $f' \in \mathcal{C}([0,1])$, so the map

$$\mathcal{T} \colon \mathcal{C}^1([0,1]) \to \mathcal{C}([0,1]) \colon f \mapsto f'$$

is defined. It is also linear, because the derivative is linear:

$$\mathcal{T}(f+g) = (f+g)' = f' + g' = \mathcal{T}(f) + \mathcal{T}(g) \quad \text{and} \quad \mathcal{T}(\alpha f) = (\alpha f)' = \alpha f' = \alpha \mathcal{T}(f).$$

6.10 Example. Define

$$\mathcal{T}\colon \mathbb{R}^2 \to \mathbb{R}^2\colon \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}.$$

We show that $\mathcal{T}$ is linear (and eventually we will connect it intimately with a matrix). First we check homogeneity:

$$\mathcal{T}\left(\alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \mathcal{T}\left(\begin{bmatrix} \alpha x_1 \\ \alpha x_2 \end{bmatrix}\right) = \begin{bmatrix} -\alpha x_2 \\ \alpha x_1 \end{bmatrix} = \alpha \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix} = \alpha\mathcal{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right).$$

Next we check additivity. This gets bulky quickly, so we do two calculations and hope they meet in the middle:

$$\mathcal{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \mathcal{T}\left(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}\right) = \begin{bmatrix} -\alpha(x_2 + y_2) \\ \alpha(x_1 + y_1) \end{bmatrix}$$

and

$$\mathcal{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) + \mathcal{T}\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{bmatrix} -\alpha x_2 \\ x_1 \end{bmatrix} + \begin{bmatrix} -\alpha y_2 \\ y_1 \end{bmatrix} = \begin{bmatrix} -\alpha x_2 - \alpha y_2 \\ x_1 + y_1 \end{bmatrix} = \begin{bmatrix} -\alpha(x_2 + y_2) \\ x_1 + y_1 \end{bmatrix}.$$

They do.

6.11 Example. The fundamental theorem of calculus tells us that if $f \in \mathcal{C}([0,1])$, then the map

$$F\colon [0,1] \to \mathbb{R}\colon x \mapsto \int_0^x f(s)\ ds$$

is differentiable on $[0,1]$ with $F' = f$. Then F' is continuous, so $F \in \mathcal{C}^1([0,1])$. We may therefore define a map

$$\mathcal{T}\colon \mathcal{C}([0,1]) \to \mathcal{C}^1([0,1])$$

by

$$(\mathcal{T}f)(x) := \int_0^x f(s)\ ds.$$

Observe carefully what we did: we started with a function f on $[0,1]$ and we wanted to define a new function $\mathcal{T}f$ on $[0,1]$, which required us to specify the values $(\mathcal{T}f)(x)$ for each $x \in [0,1]$.

We check that $\mathcal{T}$ is linear. First,

$$\big(\mathcal{T}(\alpha f)\big)(x) = \int_0^x \alpha f(s)\ ds = \alpha \int_0^x f(s)\ ds = \alpha(\mathcal{T}f)(x)$$

by the linearity of the definite integral. Next,

$$(\mathcal{T}(f+g))(x) = \int_0^x \big(f(s) + g(s)\big)\ ds = \int_0^x f(s)\ ds + \int_0^s g(s)\ ds = (\mathcal{T}f)(x) + (\mathcal{T}f)(x),$$

again by the linearity of the integral. Since we have these pointwise equalities for all $x \in [0,1]$, the functions $\mathcal{T}(\alpha f)$ and $\alpha\mathcal{T}f$ are equal, as are $\mathcal{T}(f+g)$ and $\mathcal{T}(f) + \mathcal{T}(g)$.

The preceding example shows that antidifferentiation (properly defined via definite integrals) is a linear operator. We expect that differentiation and integration undo each other, and we will see how, through the right lenses, the operators of Examples 6.9 and 6.11 are "inverses" of each other. For now, it is worthwhile to reflect on our mathematical progress: in precalculus, we studied functions probably "in isolation" from each other; in calculus, we studied functions together via the common features of continuity, differentiability, and integrability; in linear algebra, we are studying functions (linear operators) that act on other functions (which are now viewed as vectors).

6.12 Problem. Here is a chance to think about linear operators as functions and thus as sets of ordered pairs. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, both over $\mathbb{F}$. Prove that a function $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$ is linear if and only if the following both hold.

1. If (v_1, w_1), $(v_2, w_2) \in \mathcal{T}$, then $(v_1 + v_2, w_1 + w_2) \in \mathcal{T}$.

2. If $(v, w) \in \mathcal{T}$ and $\alpha \in \mathbb{F}$, then $(\alpha v, \alpha w) \in \mathcal{T}$.

As we saw in Examples 1.1, 1.2, and 1.4, many problems (realistic or artificial, interesting or boring) can be written in the form "Solve $\mathcal{T}(v) = w$ for v given w, with $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$ a linear operator, $v \in \mathcal{V}$, and $w \in \mathcal{W}$." To become competent at this, we will do a few more examples of linear operators in isolation and then consider how linear operators interact with subspaces of $\mathcal{V}$ and $\mathcal{W}$ and with each other.

6.13 Problem. To be fair, Example 1.1 involves matrix-vector multiplication as the linear operator, and we have not yet discussed that. What was the linear operator in Example 1.2?

Hereafter, we will often remove the parentheses when evaluating a linear operator $\mathcal{T}$. That is, we frequently write

$$\mathcal{T}v := \mathcal{T}(v).$$

Day 7: Monday, August 26.

Material from *Linear Algebra* by Meckes & Meckes

Pages 67–69 introduce matrix-vector multiplication "componentwise"; you will probably recognize this as a dot product formulation. Lemma 2.11 deduces as a consequence of this definition our in-class definition of matrix-vector multiplication. Pages 69–73 discuss eigenvalues. Pages 73–75 review linear systems as matrix-vector equations and give an application of eigenvalues. See Example 3 on p.87 for a connection between a "continuous" integral operator and the "discrete" operator of matrix-vector multiplication. Example 4 on pp.87–88 generalizes our sequence "shift" example.

Do Quick Exercises #3 and #5 in Section 2.1.

Vocabulary from today

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

The linear operator induced by a matrix. Eigenvalue of a linear operator (N), eigenvector of a linear operator (N), eigenpair of a linear operator. Also, eigenvalue, eigenvector, and eigenpair for a matrix.

Sequences in $\mathbb{R}^\infty$ are morally "infinitely long column vectors," and, being infinite, their entries can be "shifted" in ways that those of a finite column vector cannot.

7.1 Example. For $(a_k) \in \mathbb{R}^\infty$, let $\mathcal{T}(a_k) = (a_{k+1})$. That is, $\mathcal{T}$ "shifts" all of the entries in (a_k) ahead by 1. For example, if we cartoonishly write

$$(a_k) = (a_1, a_2, a_3, a_4, \ldots),$$

then

$$\mathcal{T}(a_k) = (a_2, a_3, a_4, a_5, \ldots).$$

We show that $\mathcal{T}$ is linear. First, for (a_k), $(b_k) \in \mathbb{R}^\infty$, we have

$$\mathcal{T}\big[(a_k) + (b_k)\big] = \mathcal{T}(a_k + b_k) = (a_{k+1} + b_{k+1}) = (a_{k+1}) + (b_{k+1}) = \mathcal{T}(a_k) + \mathcal{T}(b_k).$$

Next, for $\alpha \in \mathbb{R}$ and $(a_k) \in \mathbb{R}^\infty$, we have

$$\mathcal{T}\big[\alpha(a_k)\big] = \mathcal{T}(\alpha a_k) = (\alpha a_{k+1}) = \alpha(a_{k+1}) = \alpha\mathcal{T}(a_k).$$

Now we take up the study of an essential linear operator whose presence we have delayed for quite a while: matrix-vector multiplication. We motivate its definition by considering what equality in the following linear system means.

7.2 Example. We are probably trained to write something like

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

as a matrix-vector equation, with matrix-vector multiplication on the left. Here is how the matrix emerges naturally from vector addition and scalar multiplication. This system is equivalent to the vector equality

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

and on the left we expand

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

Of course this should also equal the matrix-vector product

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

That is, we want to define

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

Here, then, is the idea: a matrix-vector product should be a linear combination of the columns of the matrix weighted by the entries of the vector. This is perhaps not the most familiar way of defining the product (and indeed an equivalent, and probably easier, way of doing so involves the dot product of the rows of the matrix with the vector), but it is highly useful for *understanding* matrix-vector multiplication.

7.3 Definition. *Let*

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{F}^{m \times n} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n.$$

Then

$$A\mathbf{x} := x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \sum_{j=1}^{n} x_j\mathbf{a}_j \in \mathbb{F}^m.$$

Note carefully that $\mathbf{x} \in \mathbb{F}^n$, $A \in \mathbb{F}^{m \times n}$, and $A\mathbf{x} \in \mathbb{F}^m$. Since $A\mathbf{x}$ is a linear combination of the columns of A, all of which have m rows, this is what we should expect.

7.4 Example. We compute

$$\begin{aligned} \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} &= 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 2+0+3 \\ 4+0+4 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ 8 \end{bmatrix}. \end{aligned}$$

Matrix-vector multiplication is of course linear in the sense that

$$A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} \tag{7.1}$$

and

$$A(\alpha\mathbf{x}) = \alpha A\mathbf{x} \tag{7.2}$$

for all $A \in \mathbb{F}^{m\times n}$, $\mathbf{x}$, $\mathbf{y} \in \mathbb{F}^n$, and $\alpha \in \mathbb{F}$. For example,

$$A(\alpha\mathbf{x}) = \sum_{j=1}^{n} \alpha x_j \mathbf{a}_j = \alpha \sum_{j=1}^{n} x_j \mathbf{a}_j = \alpha A\mathbf{x}.$$

7.5 Problem. Show that $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}$.

The identities in (7.1) and (7.2) show that every $A \in \mathbb{F}^{m\times n}$ "induces" a linear operator from $\mathbb{F}^n$ to $\mathbb{F}^m$.

7.6 Theorem. *Let $A \in \mathbb{F}^{m\times n}$. The map*

$$\mathcal{T}_A \colon \mathbb{F}^n \to \mathbb{F}^m \colon \mathbf{v} \mapsto A\mathbf{v}$$

is linear. We call $\mathcal{T}_A$ the LINEAR OPERATOR INDUCED BY THE MATRIX A.

We emphasize that $\mathcal{T}_A$ and A are different objects! Both are functions, but A is a function from $\{(i,j) \mid i = 1,\ldots,m,\ j = 1,\ldots,n\}$, whereas $\mathcal{T}_A$ is a function from $\mathbb{F}^n$ to $\mathbb{F}^m$. Later we will develop the notion of isomorphism to explain how one might reasonably consider $\mathcal{T}_A$ and A to be "the same." Later we will also address the reverse question: if $\mathcal{T} \colon \mathbb{F}^n \to \mathbb{F}^m$ is linear, is there a matrix $A \in \mathbb{F}^{m\times n}$ such that $\mathcal{T}\mathbf{v} = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$? (Yes.)

For now we pursue a different question: what makes a linear operator "easy"? Arguably the "easiest" linear operator defined on any vector space is scalar multiplication, as proved in Problem 6.7. Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. A linear operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ acts as scalar multiplication on a vector $v \in \mathcal{V}$ if

$$\mathcal{T}v = \lambda v$$

for some $\lambda \in \mathbb{F}$. This happens for any scalar λ if $v = 0$.

7.7 Problem. Explain that. More precisely, using only Definition 6.5, show that $\mathcal{T}0 = 0$ for any linear operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$. (To be clear, the 0 in $\mathcal{T}0$ is the zero vector for $\mathcal{V}$, while on the right it is the zero vector for $\mathcal{W}$. Perhaps subscripts would be nice here: $\mathcal{T}0_{\mathcal{V}} = 0_{\mathcal{W}}$.

Consequently, in considering the problem $\mathcal{T}v = \lambda v$, we restrict to the more interesting case of nonzero v.

7.8 Definition. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and let $\mathcal{T}\colon \mathcal{V} \to \mathcal{V}$ be linear. A scalar $\lambda \in \mathbb{F}$ is an* EIGENVALUE *of $\mathcal{T}$ if there exists $v \in \mathcal{V} \setminus \{0\}$ such that*

$$\mathcal{T}v = \lambda v.$$

Such a vector $v \neq 0$ is an EIGENVECTOR *of $\mathcal{T}$ corresponding to λ (equivalently, λ is an eigenvalue of $\mathcal{T}$ "corresponding" to v), and the ordered pair (λ, v) is an* EIGENPAIR *of $\mathcal{T}$.*

It is important to be aware that we only define eigenvalues and eigenvectors for linear operators mapping a space $\mathcal{V}$ back into itself. Indeed, if we want $\mathcal{T}v = \lambda v$, then since $v \in \mathcal{V}$, we also have $\lambda v \in \mathcal{V}$, and thus $\mathcal{T}v \in \mathcal{V}$. That is, eigenvalues simply do not make sense for an operator $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$ with $\mathcal{V} \neq \mathcal{W}$ (or maybe with $\mathcal{V}$ not a *subspace* of $\mathcal{W}$). We will eventually generalize the notion of eigenvalue to "singular values" when $\mathcal{V} \neq \mathcal{W}$.

We will also see that eigenvalues reveal vastly useful data about linear operators, and we will develop some methods for computing eigenvalues from scratch. It is frustrating that the eigenvalue problem is really overdetermined, for we want to solve the single equation $\mathcal{T}v = \lambda v$ with the two unknowns λ and v. For now, we focus on basic calculations and a variety of examples.

7.9 Example. We claim that 2 is an eigenvalue of the linear operator induced by

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

with corresponding eigenvector

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

We check this by computing

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

Of course, we also talk about eigenvalues and eigenvectors of matrices. The following is mostly a rehash of Definition 7.8.

7.10 Definition. *Let $A \in \mathbb{F}^{n \times n}$. A scalar $\lambda \in \mathbb{F}$ is an* EIGENVALUE *of A if there exists $\mathbf{v} \in \mathbb{F}^n \setminus \{\mathbf{0}\}$ such that*

$$A\mathbf{v} = \lambda \mathbf{v}.$$

Such a vector $\mathbf{v} \neq \mathbf{0}$ is an EIGENVECTOR *of A corresponding to λ (equivalently, λ is an eigenvalue of A "corresponding" to $\mathbf{v}$), and the ordered pair $(\lambda, \mathbf{v})$ is an* EIGENPAIR *of A.*

7.11 Problem. Let $A \in \mathbb{F}^{n \times n}$. Check that Definitions 7.8 and 7.10 are the same in the sense that $(\lambda, \mathbf{v})$ is an eigenpair for A (according to Definition 7.10) if and only if $(\lambda, \mathbf{v})$

is an eigenpair of the the linear operator induced by A (according to Definition 7.10 and Theorem 7.6).

We will see that $A \in \mathbb{F}^{n \times n}$ always has at least one eigenvalue (in $\mathbb{C}$, not necessarily in $\mathbb{R}$) and at most n distinct eigenvalues. However, operators more generally need not have so restrictive—or so generous—an amount of eigenvalues.

7.12 Example. For an interval $I \subseteq \mathbb{R}$, denote by $\mathcal{C}^\infty(I)$ the set of all infinitely differentiable functions on I. That is, $f \in \mathcal{C}^\infty(I)$ if and only if each derivative $f^{(k)}$ exists for all $k \geq 1$. This, unsurprisingly, is a vector space, and the differentiation operator

$$\mathcal{T} \colon \mathcal{C}^\infty(\mathbb{R}) \to \mathcal{C}^\infty(\mathbb{R}) \colon f \mapsto f'$$

is defined and linear.

We compute its eigenvalues: $\mathcal{T}f = \lambda f$ with $f \in \mathcal{C}^\infty(\mathbb{R}) \setminus \{0\}$ if and only if

$$f'(x) = \lambda f(x) \text{ for all } x \in \mathbb{R} \quad \text{and} \quad f(x) \neq 0 \text{ for at least one } x \in \mathbb{R}.$$

The first condition is the pointwise equality that defines $\mathcal{T}f = \lambda f$, and the second equality is the pointwise condition that means $f \neq 0$. That first condition means that f is a function whose derivative is a multiple of itself; we know from calculus that such functions are multiples of exponentials. Specifically, $f'(x) = \lambda f(x)$ for all $x \in \mathbb{R}$ if and only if

$$f(x) = f(0)e^{\lambda x},$$

where we are free to choose $f(0) \in \mathbb{R}$ to be any value.

Taking $f(0) = 1$, we see that $f(x) := e^{\lambda x}$ is an eigenvalue (one might say, "eigenfunction") of $\mathcal{T}$, and so every real number is an eigenvalue of $\mathcal{T}$. If we allow our functions to be complex-valued and accept that the calculus of complex-valued functions of a real variable is the same as the real-valued calculus that we know and love, and if we have a definition of e^z for $z \in \mathbb{C}$ that preserves the derivative identity $f'(x) = \lambda e^{\lambda x}$ when $\lambda \in \mathbb{C}$ and $x \in \mathbb{R}$, then every complex number would be an eigenvalue of $\mathcal{T}$.

7.13 Problem. Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. What are the eigenvalues of the zero operator on $\mathcal{V}$? The identity operator on $\mathcal{V}$? (These operators were defined in Problem 6.7.)

Day 8: Wednesday, August 28.

Here is an eigenvalue example that illustrates how the choice of vector space—context!—matters.

8.1 Example. **(i)** Consider the shift operator from Example 7.1:

$$\mathcal{T}: \mathbb{R}^\infty \to \mathbb{R}^\infty : (a_k) \mapsto (a_{k+1}).$$

To search for eigenvalues and eigenvectors, we study the equation

$$\mathcal{T}(a_k) = \lambda(a_k)$$

with $(a_k) \neq 0$. That is, we want $a_k \neq 0$ for at least one k and

$$(a_{k+1}) = (\lambda a_k).$$

Since sequences are equal if and only if their corresponding terms are equal, we want

$$a_{k+1} = \lambda a_k \tag{8.1}$$

for all integers $k \geq 1$. We see what this means for a few small values of k:

$$\begin{aligned} a_2 &= a_{1+1} = \lambda a_1 \\ a_3 &= a_{2+1} = \lambda a_2 = \lambda(\lambda a_1) = \lambda^2 a_1 \\ a_4 &= a_{3+1} = \lambda a_3 = \lambda(\lambda^2)a_1 = \lambda^3 a_1. \end{aligned}$$

It looks like

$$a_{k+1} = \lambda^k a_1$$

for all k, equivalently,

$$a_k = \lambda^{k-1} a_1 \tag{8.2}$$

for all k. We could prove this by induction on k from the relation (8.2), but we could also just take (8.2) as a *candidate* for an eigenvector with eigenvalue λ and check. We compute

$$\mathcal{T}(\lambda^{k-1}a_1) = (\lambda^{(k-1)+1}a_1) = \lambda(\lambda^{k-1}a_1).$$

Thus $(\lambda^{k-1}a_1)$ is an eigenvector for λ provided that $(\lambda^{k-1}a_1) \neq 0$.

If $a_1 \neq 0$ and $\lambda \neq 0$, then $(\lambda^{k-1}a_1)$ is definitely not the zero sequence, so any $\lambda \in \mathbb{R} \setminus \{0\}$ is an eigenvalue. We might want to be more careful with $\lambda = 0$, as there $\lambda^{k-1}a_1 = 0$ for $k \geq 2$, regardless of the choice of a_1. At $k = 1$, if we interpret $0^0 = 1$, then the sequence (v_k) defined by

$$v_k := \begin{cases} 1, & k \neq 0 \\ 0, & k = 0 \end{cases}$$

is not the zero sequence and

$$v_{k+1} = 0 = 0 \cdot v_k$$

for all $k \geq 1$. Then

$$\mathcal{T}(v_k) = (v_{k+1}) = 0,$$

and so (v_k) is an eigenvector corresponding to the eigenvalue 0.

(ii) Consider $\mathcal{T}$ now as an operator from ℓ^∞ to ℓ^∞, where ℓ^∞ was defined in Example 5.5. That $\mathcal{T}(a_k) \in \ell^\infty$ for any $(a_k) \in \ell^\infty$ is easy: if there is $M > 0$ such that $|a_k| \le M$ for all k, then certainly $|a_{k+1}| \le M$ for all k, too. Above we showed that if $\mathcal{T}(a_k) = \lambda(a_k)$, then $a_k = \lambda^{k-1}a_1$ for some $a_1 \in \mathbb{R}$. For (a_k) to be an eigenvector, we need $a_1 \ne 0$. But now the sequence $(\lambda^{k-1}a_1)$ need not be bounded. Indeed, if $|r| > 1$, then the sequence (r^k) is unbounded—it satisfies $\lim_{k\to\infty} r^k = \infty$ if $r > 0$, while if $r < 0$, the terms of the sequence (r^k) grow unboundedly large as $k \to \infty$, i.e., $\lim_{k\to\infty} r^{2k} = \infty$ and $\lim_{k\to\infty} r^{2k+1} = -\infty$. Consequently, $(\lambda^{k-1}a_1)$ is only bounded when $|\lambda| \le 1$, and so the eigenvalues of $\mathcal{T}$ as a linear operator from ℓ^∞ to ℓ^∞ are not all of $\mathbb{R}$ but only the interval $[-1, 1]$.

Here is an operator that has no eigenvalues.

8.2 Example. Define $\mathcal{T} \colon \mathcal{C}([0,1]) \to \mathcal{C}([0,1])$ by $(\mathcal{T}f)(x) = xf(x)$. That is, $\mathcal{T}$ is the (unimaginatively named) "multiplication by x" operator. Suppose that $\mathcal{T}f = \lambda f$ for some $\lambda \in \mathbb{R}$ and nonzero $f \in \mathcal{C}([0,1])$. By "nonzero" we mean that $f(x) \ne 0$ for at least one $x \in [0,1]$.

Pointwise, we have $\mathcal{T}f = \lambda f$ if and only if $(\mathcal{T}f)(x) = \lambda f(x)$ for all $x \in [0,1]$, thus if and only if

$$xf(x) = \lambda f(x), \ 0 \le x \le 1.$$

This is equivalent to

$$(x - \lambda)f(x) = 0, \ 0 \le x \le 1,$$

and so, for each $x \in [0,1]$, either

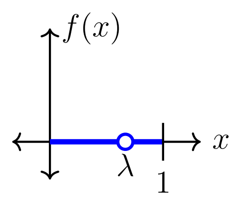
$$x - \lambda = 0 \quad \text{or} \quad f(x) = 0, \tag{8.3}$$

or possibly both.

If $x - \lambda = 0$, that means $x = \lambda$. But this is only possible if $\lambda \in [0,1]$. So, we consider two cases on λ.

1. $\lambda \in \mathbb{R} \setminus [0,1]$. That is, $\lambda < 0$ or $\lambda > 1$. Then in (8.3), it can never be the case that $x - \lambda = 0$ for some $x \in [0,1]$, and so it must be the case that $f(x) = 0$ for all x. But then $f = 0$, which is not allowed for an eigenvector. So, no $\lambda \in \mathbb{R} \setminus [0,1]$ is an eigenvalue.

2. $\lambda \in [0,1]$. Then for $x \in [0,1] \setminus \{\lambda\}$, we have from (8.3) that $f(x) = 0$. That is, f is 0 for all but one point in $[0,1]$. Here is the graph of f when $0 < \lambda < 1$.



Since f is continuous at λ, we have

$$f(\lambda) = \lim_{x\to\lambda} f(x) = \lim_{x\to\lambda} 0 = 0.$$

But then $f(x) = 0$ for all $x \in [0, 1]$, which is not allowed for an eigenvector. A similar argument with left or right limits, when $\lambda = 0$ or $\lambda = 1$, respectively, shows that $f = 0$ in those two cases as well. Thus no point in $[0, 1]$ is an eigenvalue.

Working with continuous complex-valued functions does not change the situation in the previous example, as the same continuity arguments would show that no $\lambda \in \mathbb{C}$ could be an eigenvalue of the "multiply by x" operator. However, here is a situation in which changing the field from $\mathbb{R}$ to $\mathbb{C}$ does improve the eigenvalue situation.

8.3 Example. Let

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

(i) Define

$$\mathcal{T} \colon \mathbb{R}^2 \to \mathbb{R}^2 \colon \mathbf{v} \mapsto A\mathbf{v}.$$

Here we consider $\mathbb{R}^2$ as a vector space over $\mathbb{R}$. (It is definitely not a vector space over $\mathbb{C}$, as $i\mathbf{v} \notin \mathbb{R}^2$ for $\mathbf{v} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$.)

We have $\mathcal{T}\mathbf{v} = \lambda\mathbf{v}$ if and only if

$$\begin{cases} -v_2 &= \lambda v_1 \\ v_1 &= \lambda v_2. \end{cases}$$

Like all eigenvalue-eigenvector problems, this is still overdetermined (two equations in the three unknowns λ, v_1, and v_2), but we can substitute the formula for v_1 from the second equation into the first to find

$$-v_2 = \lambda(\lambda v_2) = \lambda^2 v_2,$$

thus

$$(\lambda^2 + 1)v_2 = 0.$$

If $v_2 = 0$, then the second equation implies $v_1 = 0$ and so $\mathbf{v} = \mathbf{0}$, which is not permissible. So, to solve the eigenvalue-eigenvector problem, we need

$$\lambda^2 + 1 = 0,$$

thus $\lambda = \pm i \notin \mathbb{R}$.

Recall from Definition 7.8 that if $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ is a linear operator and $\mathcal{V}$ is a vector space over $\mathbb{F}$, then an eigenvalue λ must belong to $\mathbb{F}$. Here $\mathbb{F} = \mathbb{R}$, so this "multiply by A" operator has no eigenvalues.

(ii) Now define

$$\mathcal{T} \colon \mathbb{C}^2 \to \mathbb{C}^2 \colon \mathbf{v} \mapsto A\mathbf{v},$$

where we consider $\mathbb{C}^2$ as a vector space over $\mathbb{C}$. (It is also a vector space over $\mathbb{R}$.) The "action" of this operator is exactly the same as in the previous part (multiply by A), but the domain of this operator is different (and larger). All of the previous work shows that $\mathcal{T}\mathbf{v} = \lambda\mathbf{v}$ only if $\lambda = \pm i$, and now we are considering $\mathbb{C}^2$ as a vector space over $\mathbb{C}$. So, the (putative) eigenvalues do belong to the field.

8.4 Problem. Why "putative" at the end of the example above? We did not show the existence of $\mathbf{v} \in \mathbb{C}^2 \setminus \{\mathbf{0}\}$ such that

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{v} = i\mathbf{v}.$$

Do that. Do the same for $-i$. How, if at all, are the eigenvectors for i and $-i$ related?

The previous example encourages us to reconsider an aspect of our definition of matrix eigenvalues.

8.5 Remark. *Definition 7.10 requires that an eigenvalue of $A \in \mathbb{F}^{n \times n}$ belong to $\mathbb{F}$. This may introduce an ambiguity for $A \in \mathbb{R}^{n \times n}$, as for such A we also have $A \in \mathbb{C}^{n \times n}$. From now on, we say that $\lambda \in \mathbb{C}$ is an* EIGENVALUE *of $A \in \mathbb{R}^{n \times n}$ if there is $\mathbf{v} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. While we will be very interested in when a matrix (with real or complex, nonreal entries) has real eigenvalues, we will not discount a complex, nonreal number as an eigenvalue of a matrix with strictly real entries.*

We have claimed that eigenvalues reveal valuable data about linear operators beyond when they act simply as scalar multiplication. To describe and evaluate such data, we need more tools, and so we shift our focus back to more general properties of linear operators.

The recent work on matrix operators and the result of Theorem 7.6 beg a converse question. We know that if $A \in \mathbb{F}^{n \times n}$, then A induces a linear operator $\mathcal{T} \colon \mathbb{F}^n \to \mathbb{F}^n$ given by multiplying by A. What if $\mathcal{T} \colon \mathbb{F}^n \to \mathbb{F}^n$ is *any* linear operator? Is $\mathcal{T}$ really matrix-vector multiplication?

Here is a suggestive example.

8.6 Example. The map

$$\mathcal{T} \colon \mathbb{R}^3 \to \mathbb{R}^3 \colon \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \mapsto \begin{bmatrix} v_1 \\ v_2 - 2v_1 \\ v_3 \end{bmatrix}$$

is linear; this is straightforward but tedious to check from the definition. (The experienced reader might note that $\mathcal{T}$ acts on $\mathbf{v} \in \mathbb{R}^3$ by subtracting twice the first row of $\mathbf{v}$ from the second; this is an elementary row operation, the likes of which we shall see frequently soon.) Some clever "backwards" algebra reveals

$$\begin{bmatrix} v_1 \\ v_2 - 2v_1 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ -2v_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ v_3 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

We recognize the expression on the right as a linear combination—specifically, the matrix-vector multiplication

$$v_1 \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

That is, $\mathcal{T}$ is the linear operator induced by the matrix above (and so we did not even have to check from the definition that $\mathcal{T}$ was linear).

Day 9: Friday, August 30.

No class.

Day 10: Wednesday, September 4.

No class.

Day 11: Friday, September 6.

Material from *Linear Algebra* by Meckes & Meckes

See the various proof references within today's material below. Read the eigenvalue examples on pp.121–122.

Do Quick Exercises #23 and #24 in Section 2.5

Vocabulary from today

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

Eigenspace, range of a linear operator, kernel of a linear operator, null space of a matrix

And we're back.

Now we show that every linear operator from $\mathbb{F}^n$ to $\mathbb{F}^m$ is really matrix-vector multiplication. To do this, we need one new piece of technology that will later become a regular favorite.

11.1 Definition. *Let $n \geq 1$ be an integer and $1 \leq j \leq n$ be an integer. The* jTH STANDARD BASIS VECTOR IN $\mathbb{F}^n$ *is the vector $\mathbf{e}_j \in \mathbb{F}^n$ whose jth entry is 1 and whose other entries are all 0.*

The notation $\mathbf{e}_j$ is fairly standard, but it does not indicate the dimension n; that is usually clear from context.

11.2 Example. **(i)** In $\mathbb{F}^3$, the standard basis vector are

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

(ii) For

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{F}^3,$$

we have

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

That is,

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3.$$

(iii) More generally, for $\mathbf{v} \in \mathbb{F}^n$, we have

$$\mathbf{v} = \sum_{j=1}^{n} v_j \mathbf{e}_j,$$

where v_j is, of course, the jth entry of $\mathbf{v}$. This is one of the essential "basis" properties of the vectors $\mathbf{e}_j$ that we will explore at length later.

Now we are ready for our result about linear operators on Euclidean spaces.

11.3 Theorem. *Let $\mathcal{T} \colon \mathbb{F}^n \to \mathbb{F}^m$ be linear. Then there exists $A \in \mathbb{F}^{m \times n}$ such that $\mathcal{T}\mathbf{v} = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$.*

Proof. Theorem 2.8 in the textbook. ■

11.4 Problem. Represent each of the following linear operators (which are "elementary row operations") by matrix-vector multiplication.

(i) The row interchange

$$\mathbb{F}^2 \to \mathbb{F}^2 \colon \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \mapsto \begin{bmatrix} v_2 \\ v_1 \end{bmatrix}.$$

(ii) The scaling

$$\mathbb{F}^2 \to \mathbb{F}^2 \colon \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \mapsto \begin{bmatrix} d_1 v_1 \\ d_2 v_2 \end{bmatrix},$$

where d_1, $d_2 \in \mathbb{F}$ are given.

This begs a new question. Every matrix $A \in \mathbb{F}^{m \times n}$ gives rise to a linear operator from $\mathbb{F}^n$ to $\mathbb{F}^m$. And every linear operator from $\mathbb{F}^n$ to $\mathbb{F}^m$ is incarnated by matrix-vector multiplication against a very specific matrix. But matrices in $\mathbb{F}^{m \times n}$ and linear operators from $\mathbb{F}^n$ to $\mathbb{F}^m$ are not the same; both, strictly speaking, are functions, but with very different domains and codomains. Yet they are acting in the same way on elements of $\mathbb{F}^n$. Is there a precise mathematical way of expressing this "sameness" while also maintaining this distinction between matrices and linear operators?

There is, and the answer is more linear operators, and more vector spaces. We now build some new machinery to encode "sameness," and this starts with examining more closely how a linear operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ interacts with the spaces $\mathcal{V}$ and $\mathcal{W}$. Recall that with $\mathcal{T}$ as a function (Definition 2.5), $\mathcal{V}$ is the **DOMAIN** of $\mathcal{T}$ and $\mathcal{W}$ is the **CODOMAIN**. And as a function, the **RANGE** of $\mathcal{T}$ is the set

$$\mathcal{T}(\mathcal{V}) := \{\mathcal{T}v \mid v \in \mathcal{V}\}.$$

However, the range is more than just a set.

11.5 Theorem. *Let $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ be linear. Then the range $\mathcal{T}(\mathcal{V})$ is a subspace of $\mathcal{W}$.*

Proof. Theorem 2.30 in the book. ■

The range tells us how much of $\mathcal{W}$ a linear operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ can "reach." If we want to solve the all-important linear equation $\mathcal{T}v = w$ for $v \in \mathcal{V}$ given $w \in \mathcal{W}$, we want the range to be as large as possible, probably ideally $\mathcal{T}(V) = \mathcal{W}$.

11.6 Problem. Use the fundamental theorem of calculus to prove that the range of

$$\mathcal{T} \colon \mathcal{C}^1([0,1]) \to \mathcal{C}([0,1]) \colon f \mapsto f'$$

is $\mathcal{C}([0,1])$.

Another space associated with $\mathcal{T}$ gives equally critical data.

11.7 Definition. *The* **KERNEL** *of a linear operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ is the space*

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

We might want to emphasize which zero vector is in play here: $v \in \ker(\mathcal{T})$ if and only if $\mathcal{T}v = 0_{\mathcal{W}}$. The kernel is often called the **NULL SPACE** *in the context of the matrix-vector multiplication operator, i.e., the* **NULL SPACE** *of $A \in \mathbb{F}^{m \times n}$ is $\{\mathbf{v} \in \mathbb{F}^n \mid A\mathbf{v} = 0\}$.*

11.8 Theorem. *Let $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ be linear. Then the kernel $\ker(\mathcal{T})$ is a subspace of $\mathcal{V}$.*

Proof. Theorem 2.36 in the book. ■

11.9 Problem. Let

$$\mathcal{V} := \left\{ f \in \mathcal{C}^1([0,1]) \;\middle|\; f(0) = 0 \right\}.$$

Show that the kernel of

$$\mathcal{T} \colon \mathcal{V} \to \mathcal{C}([0,1]) \colon f \mapsto f'$$

is **TRIVIAL** in the sense that $\mathcal{T}f = 0$ if and only if $f = 0$.

11.10 Problem. Show that the kernel controls uniqueness of solutions to the all-important linear equation $\mathcal{T}v = w$ in the following sense. Let $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ be linear and suppose that $v_0 \in \ker(\mathcal{T}) \setminus \{0\}$. Let $w \in \mathcal{W}$ and suppose that $\mathcal{T}v_\star = w$ for some $v_\star \in \mathcal{V}$. Show that $\mathcal{T}(v_\star + \alpha v_0) = w$ for all $\alpha \in \mathbb{F}$ and so the problem $\mathcal{T}v = w$ has infinitely many solutions.

A third kind of subspace associated to a linear operator arises when $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ has an eigenvalue. The eigenvectors from Examples 7.12 and 8.1 were really *multiples* of one particular eigenvector. This might call to mind the notion of span (Definition 6.3), and spans are subspaces (Theorem 6.4).

11.11 Theorem. *Let $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ be a linear operator with the eigenvalue λ. Then the* **EIGENSPACE**

$$\mathcal{E}_\lambda(\mathcal{T}) := \{ v \in \mathcal{V} \mid \mathcal{T}v = \lambda v \}$$

is a subspace of $\mathcal{V}$.

Proof. Corollary 2.40 in the book. ■

11.12 Problem. Explain why $\mathcal{E}_\lambda(\mathcal{T})$ is not the set of all eigenvectors of $\mathcal{T}$ corresponding to λ.

The proof of Theorem 11.11 hinged on manipulating the eigenvalue-eigenvector equation $\mathcal{T}v = \lambda v$ into the “kernel”-type equation $(\mathcal{T} - \lambda I)v = 0$. This was a perfectly natural (and naive) symbolic manipulation, and it bears more examination. We can add vectors and multiply them by scalars; linear operators act on vectors, so pointwise (vectorwise?) we should be able to add linear operators and multiply them by scalars.

11.13 Definition. *Let $\mathcal{T}_1$, $\mathcal{T}_2 \colon \mathcal{V} \to \mathcal{W}$ be linear operators and $\alpha \in \mathbb{F}$. We define the operators $\mathcal{T}_1 + \mathcal{T}_2$ and $\alpha\mathcal{T}_1$ “pointwise” by*

$$(\mathcal{T}_1 + \mathcal{T}_2)v := \mathcal{T}_1 v + \mathcal{T}_2 v \quad \text{and} \quad (\alpha\mathcal{T}_1)v := \alpha(\mathcal{T}_1 v) \tag{11.1}$$

for $v \in \mathcal{V}$.

On the left in (11.1), the sum $\mathcal{T}_1 + \mathcal{T}_2$ is the name for the new operator that pairs $v \in \mathcal{V}$ with the vector addition $\mathcal{T}_1 v + \mathcal{T}_2 v \in \mathcal{W}$, and likewise $\alpha\mathcal{T}_1$ is the new operator that pairs

$v \in \mathcal{V}$ with the scalar multiplication $\alpha(\mathcal{T}_1 v) \in \mathcal{W}$. This is *exactly* how we defined addition and scalar multiplication of functions from an arbitrary set X into $\mathbb{R}$.

With this operator arithmetic, we can define a new vector space of linear operators.

11.14 Theorem. *Let $\mathbf{L}(\mathcal{V}, \mathcal{W})$ denote the set of all linear operators from the vector space $\mathcal{V}$ to the vector space $\mathcal{W}$. Then $\mathbf{L}(\mathcal{V}, \mathcal{W})$ is a vector space with addition and scalar multiplication defined in Definition 11.13.*

Proof. Theorem 2.5 in the book. ■

Marvel at how far we have come: we started with spaces of column vectors and functions, generalized their fundamental properties to vector spaces, built the machinery of linear operators to connect vector spaces, and now we have created a new vector space out of linear operators.

Day 12: Monday, September 9.

Material from *Linear Algebra* by Meckes & Meckes

The composition of linear operators is defined on p.81. See Theorem 2.6 for distribution.

Do Quick Exercises #8 and #10 in Section 2.2.

Vocabulary from today

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

Composition of linear operators (how do you define it?)

We recall that if $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, then the set $\mathbf{L}(\mathcal{V}, \mathcal{W})$ of linear operators from $\mathcal{V}$ to $\mathcal{W}$ is a vector space with the "pointwise" algebraic operations of

$$\mathcal{T}_1 + \mathcal{T}_2 \colon \mathcal{V} \to \mathcal{W} \colon v \mapsto \mathcal{T}_1 v + \mathcal{T}_2 v \quad \text{and} \quad \alpha\mathcal{T} \colon \mathcal{V} \to \mathcal{W} \colon v \mapsto \alpha(\mathcal{T} v)$$

We consider a number of examples of this space of operators.

12.1 Example. **(i)** We have previously identified each operator $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ with a matrix $A \in \mathbb{F}^{m \times n}$ in the sense that $\mathcal{T}\mathbf{v} = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$. While $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathbb{F}^{m \times n}$ are different vector spaces (their elements are different vectors!), one of our lingering goals is to address precisely how they are "the same."

(ii) Possibly the simplest nontrivial operator space is $\mathbf{L}(\mathbb{F}, \mathcal{W})$, where $\mathcal{W}$ is any vector space. This is because the field $\mathbb{F}$, considered as a vector space over itself, is probably the simplest vector space, other than the trivial space $\{0\}$. Let $\mathcal{T} \in \mathbf{L}(\mathbb{F}, \mathcal{W})$. Then for any

$v \in \mathbb{F}$, we have $v = v \cdot 1$, and so

$$\mathcal{T}v = \mathcal{T}(v \cdot 1) = v\mathcal{T}(1),$$

where the second equality is due to the linearity of $\mathcal{T}$ and the assumption $v \in \mathbb{F}$. That is, $\mathcal{T}v$ is just the scalar multiplication of v against the vector $\mathcal{T}(1)$, and so all operators in $\mathbf{L}(\mathbb{F}, \mathcal{W})$ are really scalar multiplication.

(iii) Consider the situation opposite to the one above: $\mathbf{L}(\mathcal{V}, \mathbb{F})$, where $\mathcal{V}$ is any vector space. We call this the **(ALGEBRAIC) DUAL SPACE** of $\mathcal{V}$ and write $\mathcal{V}' := \mathbf{L}(\mathcal{V}, \mathbb{F})$. We may just call this the dual space, but be aware that there is another kind of dual space when $\mathcal{V}$ has a norm, and that space is usually denoted $\mathcal{V}^*$. We call an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathbb{F})$ a **LINEAR FUNCTIONAL** on $\mathcal{V}$. It turns out that many interesting properties of $\mathcal{V}$ can be encoded via the linear functionals on $\mathcal{V}$.

For a concrete example, consider $\mathbf{L}(\mathbb{F}^3, \mathbb{F})$. We know that any $\mathcal{T} \in \mathbf{L}(\mathbb{F}^3, \mathbb{F})$ has the form $\mathcal{T}\mathbf{v} = A\mathbf{v}$ for some $A \in \mathbb{F}^{1\times 3}$, say, $A = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$. Then

$$\mathcal{T}\mathbf{v} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = a_1v_1 + a_2v_2 + a_3v_3.$$

This is, of course, the **DOT PRODUCT** of the column vectors

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

That is, every linear functional on $\mathbb{F}^3$ is given by taking the dot product against some fixed vector, and the same is true for $\mathbb{F}^n$.

(iv) If $\mathcal{V} = \mathcal{W}$, then we write $\mathbf{L}(\mathcal{V}) := \mathbf{L}(\mathcal{V}, \mathcal{V})$, and we say that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is a **LINEAR OPERATOR ON** $\mathcal{V}$. The space $\mathbf{L}(\mathcal{C}^\infty(\mathbb{R}))$ is immense and includes all differential and integral operators. The set

$$\mathcal{V} := \{\mathcal{T} \in \mathbf{L}(\mathcal{C}^\infty(\mathbb{R})) \mid \mathcal{T}f = af'' + bf' + cf \text{ for some } a,\ b,\ c \in \mathbb{R}\}$$

is a subspace and is arguably the central object of study in an ODE class.

We live in vector spaces, and we move between them via linear operators; together, linear operators and vector spaces encode the all-important problem of solving $\mathcal{T}v = w$ for an operator $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$ with $\mathcal{V}$ and $\mathcal{W}$ vector spaces and $w \in \mathcal{W}$ given. But we can move among multiple vector spaces with successive operators.

12.2 Theorem. *Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces and let $\mathcal{T}_1 \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{T}_2 \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Fix $u \in \mathcal{U}$. Then $\mathcal{T}_1u \in \mathcal{V}$, and so $\mathcal{T}_2(\mathcal{T}_1u) \in \mathcal{W}$. We can therefore define a map*

$$\mathcal{T}_2\mathcal{T}_1 \colon \mathcal{U} \to \mathcal{W} \colon u \mapsto \mathcal{T}_2(\mathcal{T}_1u).$$

This map $\mathcal{T}_2\mathcal{T}_1$ is the **COMPOSITION** *of $\mathcal{T}_2$ and $\mathcal{T}_1$, and it is linear: $\mathcal{T}_2\mathcal{T}_1 \in \mathbf{L}(\mathcal{U}, \mathcal{W})$. PIC*

Proof. Proposition 2.4 in the book. ■

Given operators $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$, we probably formally expect distributive laws like

$$\mathcal{T}_3(\mathcal{T}_1 + \mathcal{T}_2) = \mathcal{T}_3\mathcal{T}_1 + \mathcal{T}_3\mathcal{T}_2 \quad \text{and} \quad (\mathcal{T}_1 + \mathcal{T}_2)\mathcal{T}_3 = \mathcal{T}_1\mathcal{T}_3 + \mathcal{T}_2\mathcal{T}_3.$$

What do these mean? Since we want to add $\mathcal{T}_1$ and $\mathcal{T}_2$, both should be elements of the same operator space $\mathbf{L}(\mathcal{U}, \mathcal{V})$, so $\mathcal{T}_1 + \mathcal{T}_2 \in \mathbf{L}(\mathcal{U}, \mathcal{V})$. Since we want to compose $\mathcal{T}_3$ with $\mathcal{T}_1 + \mathcal{T}_2$, and since $\mathcal{T}_1 + \mathcal{T}_2$ maps into $\mathcal{V}$, the operator $\mathcal{T}_3$ should have domain $\mathcal{V}$, thus $\mathcal{T}_3 \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Since operators are functions, we need pointwise (vectorwise?) equalities here:

$$\mathcal{T}_3(\mathcal{T}_1 + \mathcal{T}_2)u = (\mathcal{T}_3\mathcal{T}_1 + \mathcal{T}_3\mathcal{T}_2)u \tag{12.1}$$

for all $u \in \mathcal{U}$. What does each side of this equality mean?

The composition on the left of (12.1) is

$$\mathcal{T}_3(\mathcal{T}_1 + \mathcal{T}_2)u = \mathcal{T}_3[(\mathcal{T}_1 + \mathcal{T}_2)u],$$

and then the sum here is, by definition of the sum of operators,

$$(\mathcal{T}_1 + \mathcal{T}_2)u = \mathcal{T}_1 u + \mathcal{T}_2 u.$$

By linearity, the composition is

$$\mathcal{T}_3(\mathcal{T}_1 u + \mathcal{T}_2 u) = \mathcal{T}_3(\mathcal{T}_1 u) + \mathcal{T}_3(\mathcal{T}_2 u). \tag{12.2}$$

The sum on the right of (12.1) is, by definition of the sum of operators,

$$(\mathcal{T}_3\mathcal{T}_1 + \mathcal{T}_3\mathcal{T}_2)u = \mathcal{T}_3\mathcal{T}_1 u + \mathcal{T}_3\mathcal{T}_2 u,$$

and by definition of composition, this is

$$\mathcal{T}_3\mathcal{T}_1 u + \mathcal{T}_3\mathcal{T}_2 u = \mathcal{T}_3(\mathcal{T}_1 u) + \mathcal{T}_3(\mathcal{T}_2 u).$$

And that is exactly (12.2).

The point of this discussion was not to convince us that operator composition distributes over operator addition—of course it does, or we would not be writing composition via juxtaposition like multiplication or using the symbol +. Rather, the point above was to practice definitions: what does equality mean (pointwise/vectorwise), what does composition mean, what does addition mean. *What does it all mean?*

Here is a concrete example of operator composition, which is really matrix multiplication in disguise.

12.3 Example. Define

$$\mathcal{T}_1 \colon \mathbb{R}^2 \to \mathbb{R}^3 \colon \mathbf{u} \mapsto \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{u} \quad \text{and} \quad \mathcal{T}_2 \colon \mathbb{R}^3 \to \mathbb{R}^4 \colon \mathbf{v} \mapsto \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{v}.$$

Then

$$\mathcal{T}_1\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix}, \qquad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

and so

$$\mathcal{T}_2(\mathcal{T}_1\mathbf{u}) = \mathcal{T}_2 \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix} = \begin{bmatrix} u_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Later we will view this more succinctly as a product of matrices.

The space $\mathbf{L}(\mathcal{V})$ has the additional structure of operator composition, in addition to addition of operators and scalar multiplication of operators. That is, if $\mathcal{T}_1$, $\mathcal{T}_2 \in \mathbf{L}(\mathcal{V})$, then the compositions $\mathcal{T}_1\mathcal{T}_2$ and $\mathcal{T}_2\mathcal{T}_1$ are both defined.

12.4 Problem. Why is that not true more generally in $\mathbf{L}(\mathcal{V}, \mathcal{W})$?

However, there is no guarantee that $\mathcal{T}_1\mathcal{T}_2 = \mathcal{T}_2\mathcal{T}_1$ for $\mathcal{T}_1$, $\mathcal{T}_2 \in \mathbf{L}(\mathcal{V})$ in general. That is, operator composition is not necessarily commutative.

12.5 Example. (i) Define

$$\mathcal{T}_1 \colon \mathbb{R}^2 \to \mathbb{R}^2 \colon \mathbf{v} \mapsto \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix} \quad \text{and} \quad \mathcal{T}_2 \colon \mathbb{R}^2 \to \mathbb{R}^2 \colon \mathbf{v} \mapsto \begin{bmatrix} 3v_1 \\ v_2 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Of course, $\mathcal{T}_1$ and $\mathcal{T}_2$ encode elementary row operations. We compute

$$\mathcal{T}_2(\mathcal{T}_1\mathbf{v}) = \mathcal{T}_2 \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix} = \begin{bmatrix} 3v_1 \\ v_2 - 2v_1 \end{bmatrix} \quad \text{but} \quad \mathcal{T}_1(\mathcal{T}_2\mathbf{v}) = \mathcal{T}_1 \begin{bmatrix} 3v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 \\ v_2 - 6v_1 \end{bmatrix}.$$

We expect $v_2 - 2v_1 \neq v_2 - 6v_1$ in general; indeed, these numbers are equal only when $-2v_1 = -6v_1$, thus only if $v_1 = 0$. So, $\mathcal{T}_1\mathcal{T}_2\mathbf{v} = \mathcal{T}_2\mathcal{T}_1\mathbf{v}$ only for

$$\mathbf{v} = \begin{bmatrix} 0 \\ v_2 \end{bmatrix}.$$

(ii) Let $\mathcal{T}_1 \in \mathbf{L}(\mathcal{C}^\infty(\mathbb{R}))$ be the multiplication operator $(\mathcal{T}_1 f)(x) := xf(x)$, and let $\mathcal{T}_2$ be differentiation. Then

$$(\mathcal{T}_2(\mathcal{T}_1 f))(x) = (\mathcal{T}_1 f)'(x) = (xf(x))' = f(x) + xf'(x)$$

but

$$(\mathcal{T}_1(\mathcal{T}_2 f))(x) = (\mathcal{T}_1 f')(x) = xf'(x).$$

We have $f(x) + xf'(x) = xf'(x)$ only when $f(x) = 0$, so $\mathcal{T}_1\mathcal{T}_2 f = \mathcal{T}_2\mathcal{T}_1 f$ only for $f = 0$.

12.6 Problem. Explain why the last result is not surprising by checking that

$$0 \in \{v \in \mathcal{V} \mid \mathcal{T}_1\mathcal{T}_2 v = \mathcal{T}_2\mathcal{T}_1 v\}$$

for any $\mathcal{T}_1$, $\mathcal{T}_2 \in \mathbf{L}(\mathcal{V})$. Is this set a subspace of $\mathcal{V}$?

Day 13: Wednesday, September 11.

Material from *Linear Algebra* by Meckes & Meckes

Pages 78–79 present basic properties of isomorphisms, and Theorem 2.9 proves that $\mathbb{F}^{m \times n}$ and $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ are isomorphic.

Do Quick Exercise #7 in Section 2.2.

Vocabulary from today

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

Injective/one-to-one linear operator (N), surjective/onto linear operator (N), isomorphism (N), isomorphic vector spaces (N)

We finally have the tools that we need to encode "sameness" of vector spaces. We start with an illustrative example.

13.1 Example. Of course,

$$\mathbb{R}^3 = \left\{ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \;\middle|\; v_1,\ v_2,\ v_3 \in \mathbb{R} \right\} = \mathbb{R}^{1,2,3},$$

and we let

$$\mathbb{P}_2(\mathbb{R}) := \left\{ f \in \mathbb{R}^{\mathbb{R}} \;\middle|\; f(x) = ax^2 + bx + c \text{ for some } a,\ b,\ c \in \mathbb{R} \right\}$$

be the set of all "at-most quadratic" functions on $\mathbb{R}$. We claim that $\mathbb{R}^3$ and $\mathbb{P}_2(\mathbb{R})$ are "the same" in that their vector space operations behave "the same." Any vector in either space is controlled by exactly three real numbers; we add vectors in $\mathbb{R}^3$ componentwise, while we add functions in $\mathbb{P}_2(\mathbb{R})$ by combining "like terms," i.e., the same powers of x, which is a kind of componentwise addition.

The spaces $\mathbb{R}^3$ and $\mathbb{P}_2(\mathbb{R})$ really appear to be "the same" when viewed through the lens of a special linear operator. Any linear operator $\mathcal{T}\colon \mathbb{R}^3 \to \mathbb{P}_2(\mathbb{R})$ maps column vectors to functions, so to define $\mathcal{T}\mathbf{v}$ for $\mathbf{v} \in \mathbb{R}^3$, we need to specify the values of $(\mathcal{T}\mathbf{v})(x)$ for $x \in \mathbb{R}$. Put

$$(\mathcal{T}\mathbf{v})(x) := v_1x^2 + v_2x + v_3 = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \cdot \begin{bmatrix} x^2 \\ x \\ 1 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

It is not too hard to show that $\mathcal{T}$ is linear. For example, to show $\mathcal{T}(\mathbf{v}+\mathbf{w}) = \mathcal{T}\mathbf{v}+\mathcal{T}\mathbf{w}$, we need to show $(\mathcal{T}(\mathbf{v}+\mathbf{w}))(x) = (\mathcal{T}\mathbf{v})(x) + (\mathcal{T}\mathbf{w})(x)$, and that amounts to checking

$$(v_1 + w_1)x^2 + (v_2 + w_2)x + (v_3 + w_3) = (v_1x^2 + v_2x + v_3) + (w_1x^2 + w_2x + w_3)$$

for all $x \in \mathbb{R}$. Of course this is true.

What is more interesting is how $\mathcal{T}$ relates elements of $\mathbb{R}^3$ and $\mathbb{P}_2(\mathbb{R})$. First, given any $f \in \mathbb{P}_2(\mathbb{R})$, if we write $f(x) = ax^2 + bx + c$, then

$$f = \mathcal{T}\begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

That is, for every $f \in \mathbb{P}_2(\mathbb{R})$, there is $\mathbf{v} \in \mathbb{R}^3$ such that $\mathcal{T}\mathbf{v} = f$. In other words, we can represent any $f \in \mathbb{P}_2(\mathbb{R})$ by some $\mathbf{v} \in \mathbb{R}^3$ "under the lens" of $\mathcal{T}$.

Next, this representation is unique. Suppose $\mathcal{T}\mathbf{v} = \mathcal{T}\mathbf{w}$ for some $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^3$. Here equality means equality of functions, so $(\mathcal{T}\mathbf{v})(x) = (\mathcal{T}\mathbf{w})(x)$ for all $x \in \mathbb{R}$. By definition of $\mathcal{T}$, this implies

$$v_1x^2 + v_2x + v_3 = w_1x^2 + w_2x + w_3, \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \ \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}.$$

This equality holds if and only if $v_1 = w_1$, $v_2 = w_2$, and $v_3 = w_3$. (Why this is true is actually a linear algebra problem of linear systems; we will just take it as a fact that two polynomials are equal on $\mathbb{R}$ if and only if the coefficients on their corresponding powers are equal.) Thus $\mathbf{v} = \mathbf{w}$.

The operator from the previous example is a special case of a much more general, but still special, kind of operator.

13.2 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$.*

(i) *The operator $\mathcal{T}$ is* **SURJECTIVE** *or* **ONTO** *if for each $w \in \mathcal{W}$ there exists $v \in \mathcal{V}$ such that $\mathcal{T}v = w$.*

(ii) *The operator $\mathcal{T}$ is* **INJECTIVE** *or* **ONE-TO-ONE** *if whenever $\mathcal{T}v_1 = \mathcal{T}v_2$ for v_1, $v_2 \in \mathcal{V}$, then $v_1 = v_2$.*

(iii) *The operator $\mathcal{T}$ is an* **ISOMORPHISM**, *and the spaces $\mathcal{V}$ and $\mathcal{W}$ are* **ISOMORPHIC**, *if $\mathcal{T}$ is both surjective and injective.*

13.3 Problem. **(i)** Explain why

$$\mathcal{T}_1 \colon \mathcal{C}^1([0,1]) \to \mathcal{C}([0,1]) \colon f \mapsto f'$$

is surjective but not injective.

(ii) Put

$$\mathcal{V} := \left\{ f \in \mathcal{C}^1([0,1]) \;\middle|\; f(0) = 0 \right\}.$$

Explain why

$$\mathcal{T}_2 \colon \mathcal{V} \to \mathcal{C}([0,1]) \colon f \mapsto f'$$

is both surjective and injective and thus an isomorphism.

13.4 Problem. Suppose that $\mathcal{V}$ and $\mathcal{W}$ are isomorphic vector spaces. Is every operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ an isomorphism?

The fundamental problem of linear algebra, that of solving $\mathcal{T}v = w$ uniquely for all $w \in\in \mathcal{W}$ with $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, is solvable precisely when $\mathcal{T}$ is an isomorphism. There is no glamorous way of checking surjectivity, but there is a shortcut to injectivity. We have

$$\mathcal{T}v_1 = \mathcal{T}v_2 \iff \mathcal{T}v_1 - \mathcal{T}v_2 = 0 \iff \mathcal{T}(v_1 - v_2) = 0 \iff v_1 - v_2 \in \ker(\mathcal{T}).$$

Since injectivity demands $v_1 = v_2$, equivalently, $v_1 - v_2 = 0$, and since $0 \in \ker(\mathcal{T})$ always, this calculation suggests that the following is true, and it is.

13.5 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Then $\mathcal{T}$ is injective if and only if* $\ker(\mathcal{T}) = \{0\}$.

Proof. Theorem 2.37 in the book. ■

The notion of isomorphism helps us qualify how matrices and linear operators on Euclidean spaces are "the same."

13.6 Theorem. *The spaces $\mathbb{F}^{m \times n}$ and $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ are isomorphic. Specifically, the operator $\mathcal{T} \colon \mathbb{F}^{m \times n} \to \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ such that $(\mathcal{T}A)\mathbf{v} = A\mathbf{v}$ for all $A \in \mathbb{F}^{m \times n}$ and $\mathbf{v} \in \mathbb{F}^n$, is an isomorphism.*

Proof. We have already proved surjectivity in Theorem 11.3. To see this from the definition, let $\mathcal{S} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. We need to find $A \in \mathbb{F}^{m \times n}$ such that $\mathcal{T}A = \mathcal{S}$. This equality is equality of operators on $\mathbb{F}^n$, so we need $(\mathcal{T}A)\mathbf{v} = \mathcal{S}\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$. By definition of $\mathcal{T}$, the matrix A needs to satisfy $A\mathbf{v} = \mathcal{S}\mathbf{v}$. The proof of Theorem 11.3 tells us how to construct A: take $A = \begin{bmatrix} \mathcal{S}\mathbf{e}_1 & \cdots & \mathcal{S}\mathbf{e}_n \end{bmatrix}$, where $\mathbf{e}_j$ is the jth standard basis vector for $\mathbb{F}^n$ (Definition 11.1).

For injectivity, we show $\ker(\mathcal{T}) = \{0\}$. Here 0 is the zero matrix in $\mathbb{F}^{m \times n}$, i.e., the $m \times n$ matrix whose entries are all 0 (i.e., the number 0). We already have $0 \in \ker(\mathcal{T})$, so we just need to show that if $\mathcal{T}A = 0$, then $A = 0$. In the equality $\mathcal{T}A = 0$, the symbol 0 represents the zero operator from $\mathbb{F}^n$ to $\mathbb{F}^m$. (The poor symbol 0 is getting quite a workout here.)

So, assume that $A \in \mathbb{F}^{m \times n}$ with $\mathcal{T}A = 0$. Then $(\mathcal{T}A)\mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in \mathbb{F}^n$, where now $\mathbf{0}$ is the zero vector in $\mathbb{F}^m$. Since this is true for all $\mathbf{v}$, we can take $\mathbf{v}$ conveniently: let $\mathbf{v} = \mathbf{e}_j$ for $j = 1, \ldots, n$. It is a fact that $A\mathbf{e}_j = \mathbf{a}_j$, where $\mathbf{a}_j$ is the jth column of A. Then $\mathbf{a}_j = \mathbf{0}$ for each j, and therefore each column of A is the zero vector (in $\mathbb{F}^m$). Consequently, A is the zero matrix in $\mathbb{F}^{m \times n}$, as desired. ■

13.7 Problem. Check that fact: $A\mathbf{e}_j = \mathbf{a}_j$ for each j.

As an exercise in baroque notation, the operator $\mathcal{T}$ from the previous theorem is an element of $\mathbf{L}(\mathbb{F}^{m \times n}, \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m))$!

Day 14: Friday, September 13.

We consider some concrete problems for injectivity and surjectivity.

14.1 Example. Let $\mathcal{T} \colon \mathcal{C}([0,1]) \to \mathcal{C}^1([0,1])$ be the "antidifferentiation" operator given by

$$(\mathcal{T}f)(x) := \int_0^x f(s)\ ds, \ \ f \in \mathcal{C}([0,1]).$$

To check injectivity, we assume $\mathcal{T}f = 0$, and we want to show $f = 0$. That is, pointwise, we have $(\mathcal{T}f)(x) = 0$ for all $x \in [0,1]$, and the goal is $f(x) = 0$ for all $x \in [0,1]$. The definition of $\mathcal{T}$ implies

$$\int_0^x f(s)\ ds = 0, \ \ x \in [0,1].$$

One way to extract information about a function f from an integral involving f is to try to differentiate that integral. Since the equality above is true *for all* x, we have

$$\frac{d}{dx}\left[\int_0^x f(s)\ ds\right] = \frac{d}{dx}[0].$$

Using the fundamental theorem of calculus on the left, we get

$$f(x) = 0$$

for all x. Thus $f = 0$, and $\mathcal{T}$ is injective.

For surjectivity, we want to take $g \in \mathcal{C}^1([0,1])$ and find $f \in \mathcal{C}([0,1])$ such that $\mathcal{T}f = g$, i.e., such that $(\mathcal{T}f)(x) = g(x)$ for all x. Can we do this? We need

$$\int_0^x f(s)\ ds = g(x), \ \ 0 \le x \le 1. \tag{14.1}$$

We could try the trick above of differentiating both sides to get $f(x) = g'(x)$. But if we check our work, we find (by the fundamental theorem of calculus)

$$(\mathcal{T}g')(x) = \int_0^x g'(s)\, ds = g(x) - g(0), \tag{14.2}$$

and this does not equal $g(x)$ unless $g(0) = 0$. Indeed, taking $x = 0$ in (14.1) implies

$$g(0) = \int_0^0 f(s)\, ds = 0,$$

and so we have a "solvability condition": if $\mathcal{T}f = g$, then $g(0) = 0$. There are plenty of functions in $\mathcal{C}^1([0,1])$ that do not meet this; take $g(x) = \cos(x)$ or $g(x) = 1$. And so $\mathcal{T}$ is not surjective: there is no $f \in \mathcal{C}([0,1])$ such that $(\mathcal{T}f)(x) = 1$ for all x.

But this points to at least a characterization of the range of $\mathcal{T}$: we might conjecture

$$\mathcal{T}\big(\mathcal{C}([0,1])\big) = \big\{g \in \mathcal{C}^1([0,1]) \bigm| g(0) = 0\big\}.$$

We already know that $\mathcal{T}f$ satisfies $(\mathcal{T}f)(0) = 0$, so assume $g \in \mathcal{C}^1([0,1])$ with $g(0) = 0$. The calculation in (14.2) implies $\mathcal{T}f = g$, and so g is in the range of $\mathcal{T}$.

14.2 Example. We study the shift operator

$$\mathcal{T} \colon \mathbb{R}^\infty \to \mathbb{R}^\infty \colon (a_k) \mapsto (a_{k+1}).$$

To check injectivity, we assume $\mathcal{T}(a_k) = 0$, where 0 is the zero sequence. Then $(a_{k+1}) = 0$, so $a_{k+1} = 0$ for all $k \geq 1$ (where the second 0 is the scalar 0). That is, $a_j = 0$ for $j \geq 2$, but this says nothing about a_1. Indeed, if we put

$$a_k = \begin{cases} 1, & k = 1 \\ 0, & k \geq 2, \end{cases}$$

then we have $(a_k) \neq 0$ but $\mathcal{T}(a_k) = 0$, so $\mathcal{T}$ is not injective.

For surjectivity, let $(b_k) \in \mathbb{R}^\infty$. We want $(a_k) \in \mathbb{R}^\infty$ such that $\mathcal{T}(a_k) = (b_k)$, so we want $(a_{k+1}) = (b_k)$. Termwise, this means $a_{k+1} = b_k$, and we could reindex this to $a_j = b_{j-1}$ for $j \geq 2$. This tells us nothing about a_1, however, and so we could simply set

$$a_k = \begin{cases} 0, & k = 1 \\ b_{k-1}, & k \geq 2 \end{cases} \tag{14.3}$$

to conclude $\mathcal{T}(a_k) = (b_k)$.

14.3 Problem. In the context of surjectivity of the previous example, let

$$z_k = \begin{cases} 1, & k = 1 \\ 0, & k \neq 0 \end{cases}$$

and show that $\mathcal{T}\big[(a_k)+\alpha(z_k)\big] = (b_k)$ with (a_k) defined by (14.3). Interpret this calculation in the context of Problem 11.10.

14.4 Problem. Prove that the operator

$$\mathcal{T}\colon \mathbb{R}^\infty \to \mathbb{R}^\infty \colon (a_k) \mapsto (ka_k)$$

is an isomorphism.

Day 15: Monday, September 16.

Material from *Linear Algebra* by Meckes & Meckes

Pages 78–80 discuss invertible linear operators. Proposition 2.2 proves that inverses are linear, and the example on p.80 gives a finite-dimensional inverse calculation. See also pp.380–382 in Appendix A for inverses of functions more generally.

Vocabulary from today

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

Invertible linear operator (N), inverse of a linear operator

Previously we have checked injectivity and surjectivity of an operator separately. This was a good idea, since our recent examples have failed to be both! However, it can be more efficient to test for both simultaneously.

15.1 Problem. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces. Prove that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is an isomorphism if and only if for all $w \in \mathcal{W}$, there exists a unique $v \in \mathcal{V}$ such that $\mathcal{T}v = w$.

15.2 Example. We check if

$$\mathcal{T}\colon \mathbb{R}^2 \to \mathbb{R}^2 \colon \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \mapsto \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix}$$

is an isomorphism. For $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^2$, we calculate

$$\mathcal{T}\mathbf{v} = \mathbf{w} \iff \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\iff \begin{cases} v_1 = w_1 \\ v_2 - 2v_1 = w_2 \end{cases}$$

$$\iff \begin{cases} v_1 = w_1 \\ v_2 = w_2 + 2w_1 \end{cases} .$$

Yes, $\mathcal{T}$ is an isomorphism: for all $\mathbf{w} \in \mathbb{R}^2$, there is a unique vector $\mathbf{v} \in \mathbb{R}^2$ such that $\mathcal{T}\mathbf{v} = \mathbf{w}$, and this vector $\mathbf{v}$ is given by $\mathbf{v} = \mathcal{S}\mathbf{w}$, where $\mathcal{S}$ is the linear operator

$$\mathcal{S} \colon \mathbb{R}^2 \to \mathbb{R}^2 \colon \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \mapsto \begin{bmatrix} w_1 \\ w_2 + 2w_1 \end{bmatrix}.$$

We are not actually going to check that $\mathcal{S}$ is linear, but by now hopefully it is easy to represent $\mathcal{S}$ as matrix-vector multiplication. The calculation above shows that, for $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^2$, we have

$$\mathcal{T}\mathbf{v} = \mathbf{w} \iff \mathbf{v} = \mathcal{S}\mathbf{w}. \tag{15.1}$$

Knowing this alone actually proves that $\mathcal{T}$ is an isomorphism, even without the formula for $\mathcal{S}$. Indeed, for surjectivity, let $\mathbf{w} \in \mathbb{R}^2$ and compute, by the defining property (15.1) of $\mathcal{S}$, that $\mathcal{T}(\mathcal{S}\mathbf{w}) = \mathbf{w}$ (take $\mathbf{v} = \mathcal{S}\mathbf{w}$). For injectivity, suppose $\mathcal{T}\mathbf{v} = \mathbf{0}$. Then the defining property (15.1) of $\mathcal{S}$ implies $\mathbf{v} = \mathcal{S}\mathbf{0} = \mathbf{0}$ since $\mathcal{S}$ is linear (here we take $\mathbf{w} = \mathbf{0}$).

15.3 Problem. Generalize the result at the end of this example as follows. Suppose that $\mathcal{V}$ and $\mathcal{W}$ are vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. If there exists $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that $\mathcal{T}v = w$ if and only if $v = \mathcal{S}w$ for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$, show that $\mathcal{T}$ is an isomorphism.

Here is the abstraction of the previous example (which also proves that the result in the problem can be strengthened to an "if and only if" statement). Let $\mathcal{V}$ and $\mathcal{W}$ be isomorphic vector spaces with isomorphism $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$. So, $\mathcal{T}$ is a linear operator that is injective and surjective. Then for each $w \in \mathcal{W}$, there is a unique $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. Now define a map (not necessarily a linear operator!) $\mathcal{S} \colon \mathcal{W} \to \mathcal{V}$ by setting, for $w \in \mathcal{W}$, the vector $\mathcal{S}w \in \mathcal{V}$ to be the unique vector $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. That is,

$$\mathcal{T}v = w \iff \mathcal{S}w = v \quad \text{and so} \quad \mathcal{T}(\mathcal{S}w) = w. \tag{15.2}$$

Both injectivity and surjectivity of $\mathcal{T}$ are critical for $\mathcal{S}$ to be a function from $\mathcal{W}$ to $\mathcal{V}$: surjectivity guarantees the existence of at least one $v \in \mathcal{V}$ such that $\mathcal{T}v = w$, given $w \in \mathcal{W}$, while injectivity guarantees the uniqueness of v.

15.4 Problem. Since $\mathcal{S} \colon \mathcal{W} \to \mathcal{V}$ is a function, $\mathcal{S}$ really is a set of ordered pairs. Prove that

$$\mathcal{S} = \{(w, v) \mid w \in \mathcal{W} \text{ and } v \in \mathcal{V} \text{ satisfy } \mathcal{T}v = w\}.$$

Of course, $\mathcal{S}$ will not be just any function but a linear operator from $\mathcal{W}$ to $\mathcal{V}$. First we check that if $\alpha \in \mathbb{F}$ and $w \in \mathcal{W}$, then $\mathcal{S}(\alpha w) = \alpha\mathcal{S}w$. We know that $\mathcal{S}w$ is the unique vector in $\mathcal{V}$ to satisfy $\mathcal{T}(\mathcal{S}w) = w$, and likewise $\mathcal{S}(\alpha w)$ is the unique vector in $\mathcal{V}$ to satisfy $\mathcal{T}(\mathcal{S}(\alpha w)) = \alpha w$. Then we have $\alpha w = \alpha\mathcal{T}(\mathcal{S}w)$, and because $\mathcal{T}$ is linear this becomes

$$\alpha w = \alpha\mathcal{T}(\mathcal{S}w) = \mathcal{T}(\alpha\mathcal{S}w).$$

Thus

$$\mathcal{T}(\alpha\mathcal{S}w) = \alpha w = \mathcal{T}(\mathcal{S}(\alpha w)),$$

so by the injectivity of $\mathcal{T}$ we have $\alpha\mathcal{S}w = \mathcal{S}(\alpha w)$.

Now let $w_1, w_2 \in \mathcal{W}$. We want to show $\mathcal{S}(w_1+w_2) = \mathcal{S}w_1+\mathcal{S}w_2$. We know that $\mathcal{S}(w_1+w_2)$ is the unique vector in $\mathcal{V}$ to satisfy $\mathcal{T}[\mathcal{S}(w_1 + w_2)] = w_1 + w_2$, $\mathcal{S}w_1$ is the unique vector in $\mathcal{V}$ to satisfy $\mathcal{T}(\mathcal{S}w_1) = w_1$, and $\mathcal{S}w_2$ is the unique vector in $\mathcal{V}$ to satisfy $\mathcal{T}(\mathcal{S}w_2) = w_2$. The linearity of $\mathcal{T}$ implies

$$w_1 + w_2 = \mathcal{T}(\mathcal{S}w_1) + \mathcal{T}(\mathcal{S}w_2) = \mathcal{T}(\mathcal{S}w_1 + \mathcal{S}w_2),$$

and so

$$\mathcal{T}(\mathcal{S}w_1 + \mathcal{S}w_2) = w_1 + w_2 = \mathcal{T}[\mathcal{S}(w_1 + w_2)].$$

The injectivity of $\mathcal{T}$ implies $\mathcal{S}w_1 + \mathcal{S}w_2 = \mathcal{S}(w_1 + w_2)$.

Of course, we want to call $\mathcal{S}$ the inverse of $\mathcal{T}$ and write $\mathcal{S} = \mathcal{T}^{-1}$. But the definite article "the" needs justification—is $\mathcal{S}$ really the only operator to satisfy (15.2)? Suppose there are two: let $\mathcal{S}_1, \mathcal{S}_2 \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that

$$\mathcal{T}v = w \iff \mathcal{S}_1 w = v \quad \text{and} \quad \mathcal{T}v = w \iff \mathcal{S}_2 w = v \tag{15.3}$$

for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$. Our goal is $\mathcal{S}_1 = \mathcal{S}_2$, i.e., $\mathcal{S}_1 w = \mathcal{S}_2 w$ for all $w \in \mathcal{W}$.

Fix $w \in \mathcal{W}$. Take $v = \mathcal{S}_1 w$ and then $v = \mathcal{S}_2$ in (15.3) to obtain, respectively, $\mathcal{T}(\mathcal{S}_1 w) = w$ and $\mathcal{T}(\mathcal{S}_2 w) = w$. That is, for all $w \in \mathcal{W}$, we have $\mathcal{T}(\mathcal{S}_1 w) = \mathcal{T}(\mathcal{S}_2 w)$, and so injectivity of $\mathcal{T}$ forces $\mathcal{S}_1 w = \mathcal{S}_2 w$.

The defining property (15.2) of $\mathcal{S}$ means that $\mathcal{T}(\mathcal{S}w) = w$ for all $w \in \mathcal{W}$. That is, $\mathcal{T}\mathcal{S} = I_{\mathcal{W}}$, where

$$I_{\mathcal{W}} \colon \mathcal{W} \to \mathcal{W} \colon w \mapsto w$$

is the identity operator on $\mathcal{W}$. We claim that we also have $\mathcal{S}(\mathcal{T}v) = v$ for all $v \in \mathcal{V}$. Indeed, if we fix $v \in \mathcal{V}$ and take $w = \mathcal{T}v$, then (15.2) implies $v = \mathcal{S}w$. Substituting $\mathcal{T}v$ for w gives $v = \mathcal{S}(\mathcal{T}v)$, and so $\mathcal{S}\mathcal{T} = I_{\mathcal{V}}$, where

$$I_{\mathcal{V}} \colon \mathcal{V} \to \mathcal{V} \colon v \mapsto v$$

is the identity operator on $\mathcal{V}$.

15.5 Problem. Do $\mathcal{T}$ and $\mathcal{S}$ commute?

We celebrate with a theorem.

15.6 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ be an isomorphism. Then the map $\mathcal{S}\colon \mathcal{W} \to \mathcal{V}$ defined by taking $\mathcal{S}w$ for $w \in \mathcal{W}$ to be the unique vector $v \in \mathcal{V}$ such that $\mathcal{T}v = w$ is a linear operator from $\mathcal{W}$ to $\mathcal{V}$. Moreover, $\mathcal{S}$ is the only operator from $\mathcal{W}$ to $\mathcal{V}$ to satisfy*

$$\mathcal{T}v = w \iff \mathcal{S}w = v. \tag{15.4}$$

We call $\mathcal{S}$ the **INVERSE** *of $\mathcal{T}$, say that $\mathcal{T}$ is* **INVERTIBLE**, *and write $\mathcal{S} = \mathcal{T}^{-1}$. We have*

$$\mathcal{T}(\mathcal{S}w) = w \text{ for all } w \in \mathcal{W} \qquad \text{and} \qquad \mathcal{S}(\mathcal{T}v) = v \text{ for all } v \in \mathcal{V}.$$

15.7 Problem. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ be an isomorphism. Prove that $\mathcal{T}^{-1}$ is also an isomorphism (from $\mathcal{W}$ to $\mathcal{V}$) and that $(\mathcal{T}^{-1})^{-1} = \mathcal{T}$.

15.8 Example. A variety of previous problems and examples lead us to conclude that the operator

$$\mathcal{T}\colon \mathcal{C}^1([0,1]) \to \mathcal{C}([0,1])\colon f \mapsto f'$$

is not an isomorphism, because $\mathcal{T}$ is not injective ($\mathcal{T}f = 0$ for any constant f). This does not say that $\mathcal{C}^1([0,1])$ and $\mathcal{C}([0,1])$ are not isomorphic, merely that if they are, it will not be through $\mathcal{T}$.

If we define

$$\mathcal{V} := \left\{ f \in \mathcal{C}^1([0,1]) \;\middle|\; f(0) = 0 \right\} \qquad \text{and} \qquad \mathcal{T}_0\colon \mathcal{V} \to \mathcal{C}([0,1])\colon f \mapsto f',$$

then we expect that $\mathcal{T}_0$ is an isomorphism. (Note that $\mathcal{T} \neq \mathcal{T}_0$, since the domains of these operators are different. Rather, $\mathcal{T}_0$ is the "restriction" of $\mathcal{T}$ to $\mathcal{V}$ in the sense that $\mathcal{V} \subseteq \mathcal{C}^1([0,1])$ and $\mathcal{T}_0 f = \mathcal{T}f$ for all $f \in \mathcal{V}$.)

Here is the calculation: for $f \in \mathcal{V}$ and $g \in \mathcal{C}([0,1])$, we have

$$\begin{aligned}
\mathcal{T}_0 f = g &\iff (\mathcal{T}_0 f)(x) = g(x) \text{ for all } x \in [0,1] \\
&\iff f'(x) = g(x) \text{ for all } x \in [0,1] \\
&\iff \int_0^x f'(s)\, ds = \int_0^x g(s)\, ds \text{ for all } x \in [0,1] \\
&\iff f(x) - f(0) = \int_0^x g(s)\, ds \text{ for all } x \in [0,1] \\
&\iff f(x) = \int_0^x g(s)\, ds \text{ for all } x \in [0,1], \text{ since } f(0) = 0.
\end{aligned}$$

The third if and only if merits expansion. Certainly if $f' = g$ on $[0,1]$, then all of the integrals $\int_0^x f'(s)\, ds$ and $\int_0^x g(s)\, ds$ are equal. Conversely, if these integrals are equal for all x, we may differentiate both sides of the equality $\int_0^x f'(s)\, ds = \int_0^x g(s)\, ds$ to obtain $f' = g$.

We conclude that $\mathcal{T}_0$ is an isomorphism with

$$(\mathcal{T}_0^{-1}g)(x) = \int_0^x g(s)\ ds,\ 0 \le x \le 1.$$

15.9 Problem. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. A LEFT INVERSE of $\mathcal{T}$ is a linear operator $\mathcal{L} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that $\mathcal{L}\mathcal{T}v = v$ for all $v \in \mathcal{V}$, and a RIGHT INVERSE of $\mathcal{T}$ is an linear operator $\mathcal{R} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that $\mathcal{T}\mathcal{R}w = w$ for all $w \in \mathcal{W}$.

(i) Prove that if $\mathcal{T}$ has a left inverse, then $\mathcal{T}$ is injective.

(ii) Prove that if $\mathcal{T}$ has a right inverse, then $\mathcal{T}$ is surjective.

(iii) Prove that $\mathcal{T}$ is invertible if and only if $\mathcal{T}$ has both a left inverse $\mathcal{L}$ and a right inverse $\mathcal{R}$, in which case $\mathcal{L} = \mathcal{R} = \mathcal{T}^{-1}$.

Day 16: Wednesday, September 18.

Material from *Linear Algebra* by Meckes & Meckes

Pages 90–91 motivate the definition of matrix-matrix multiplication in terms of operator composition. Equation (2.11) gives an entrywise definition of this product. Lemma 2.11 proves our "columnwise" definition from that entrywise definition. Study the box at the bottom of p.91 carefully. Finally, Lemma 2.14 gives the familiar definition of AB as "the (i, j)-entry of AB is the dot product of the ith row of A with the jth column of B."

Do Quick Exercises #11 and #14 in Section 2.3.

We have devoted significant effort to understanding linear operators between arbitrary vector spaces. Few of our results have specified exactly what the vector spaces were. Now we have the tools to appreciate more operators on Euclidean spaces (spaces of column vectors: $\mathbb{F}^n$) and their connections to matrices.

Theorem 13.6 tells us that all linear operators between Euclidean spaces are given by matrix-vector multiplication. Specifically, let $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. Then the MATRIX REPRESENTATION OF $\mathcal{T}$ is the matrix

$$[\mathcal{T}] := \begin{bmatrix} \mathcal{T}\mathbf{e}_1 & \cdots & \mathcal{T}\mathbf{e}_n \end{bmatrix} \in \mathbb{F}^{m \times n},$$

where $\mathbf{e}_j$ is the jth standard basis vector for $\mathbb{F}^n$ (Definition 11.1). The matrix $[\mathcal{T}]$ satisfies

$$\mathcal{T}\mathbf{v} = [\mathcal{T}]\mathbf{v}$$

for any $\mathbf{v} \in \mathbb{F}^n$; on the left, we just have the abstract application of $\mathcal{T}$ to $\mathbf{v}$ (i.e., the evaluation of the function $\mathcal{T}$ at $\mathbf{v}$), while on the right, this application is given "concretely" by the matrix-vector product $[\mathcal{T}]\mathbf{v}$.

16.1 Example. For the linear operator

$$\mathcal{T} \colon \mathbb{R}^2 \to \mathbb{R}^2 \colon \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \mapsto \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix},$$

we compute

$$\mathcal{T}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 - 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

and

$$\mathcal{T}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 - 2 \cdot 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

thus

$$[\mathcal{T}] = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}.$$

Indeed,

$$[\mathcal{T}]\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix} = \mathcal{T}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Now let $\mathcal{T}_1 \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^p)$ and $\mathcal{T}_2 \in \mathbf{L}(\mathbb{F}^p, \mathbb{F}^m)$. Their composition $\mathcal{T}_2\mathcal{T}_1$, defined by $(\mathcal{T}_2\mathcal{T}_1)\mathbf{u} = \mathcal{T}_2(\mathcal{T}_1\mathbf{u})$, satisfies $\mathcal{T}_2\mathcal{T}_1 \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. Thus $[\mathcal{T}_1] \in \mathbb{F}^{p \times n}$, $[\mathcal{T}_2] \in \mathbb{F}^{m \times p}$, and $[\mathcal{T}_2\mathcal{T}_1] \in \mathbb{F}^{m \times n}$. Moreover,

$$\mathcal{T}_1\mathbf{u} = [\mathcal{T}_1]\mathbf{u}, \ \mathbf{u} \in \mathbb{F}^n, \quad \text{and} \quad \mathcal{T}_2\mathbf{v} = [\mathcal{T}_2]\mathbf{v}, \ \mathbf{v} \in \mathbb{F}^p,$$

and we compute twice

$$(\mathcal{T}_2\mathcal{T}_1)\mathbf{u} = [\mathcal{T}_2\mathcal{T}_1]\mathbf{u} \quad \text{and also} \quad (\mathcal{T}_2\mathcal{T}_1)\mathbf{u} = \mathcal{T}_2(\mathcal{T}_1\mathbf{u}) = \mathcal{T}_2\big([\mathcal{T}_1]\mathbf{u}\big) = [\mathcal{T}_2]\big([\mathcal{T}_1]\mathbf{u}\big)$$

for all $\mathbf{u} \in \mathbb{F}^n$. That is,

$$[\mathcal{T}_2\mathcal{T}_1]\mathbf{u} = [\mathcal{T}_2]\big([\mathcal{T}_1]\mathbf{u}\big) \tag{16.1}$$

for all $\mathbf{u} \in \mathbb{F}^n$.

Since (16.1) is true for all $\mathbf{u} \in \mathbb{F}^n$, we are free to take $\mathbf{u} = \mathbf{e}_j$ as the jth standard basis vector, so

$$[\mathcal{T}_2\mathcal{T}_1]\mathbf{e}_j = [\mathcal{T}_2]\big([\mathcal{T}_1]\mathbf{e}_j\big) \tag{16.2}$$

for each j. It is a fact that if $A \in \mathbb{F}^{m \times n}$ and $\mathbf{e}_j$ is the jth standard basis vector for $\mathbb{F}^n$, then $A\mathbf{e}_j$ is the jth column of A. That is, if $A = \begin{bmatrix} \mathbf{a}_1 & \cdots \mathbf{a}_n \end{bmatrix}$, then

$$A\mathbf{e}_j = \mathbf{a}_j.$$

16.2 Problem. Prove it.

In words, then, (16.2) says that the jth column of the matrix $[\mathcal{T}_2\mathcal{T}_1]$ is the matrix-vector product of the matrix $[\mathcal{T}_2]$ and the jth column of $[\mathcal{T}_1]$.

16.3 Problem. Stare at (16.2) until you fully believe the sentence above.

This suggests a meaningful way of defining the product of two matrices. Let $A \in \mathbb{F}^{m \times p}$ and $B \in \mathbb{F}^{p \times n}$. Then the operators

$$\mathcal{T}_A \colon \mathbb{F}^p \to \mathbb{F}^m \colon \mathbf{v} \mapsto A\mathbf{v} \quad \text{and} \quad \mathcal{T}_B \colon \mathbb{F}^n \to \mathbb{F}^p \colon \mathbf{u} \mapsto B\mathbf{u}$$

have matrix representations

$$[\mathcal{T}_A] = A \quad \text{and} \quad [\mathcal{T}_B] = B.$$

16.4 Problem. Why?

We would like to define the matrix product AB so that the composition

$$\mathcal{T}_A\mathcal{T}_B \colon \mathbb{F}^n \to \mathbb{F}^m \colon \mathbf{u} \mapsto A(B\mathbf{u})$$

has matrix representation

$$[\mathcal{T}_A\mathcal{T}_B] = AB. \tag{16.3}$$

How should we define this symbol AB? The work before Problem 16.3 suggests that we want the jth column of AB to be the matrix-vector product of A and the jth column of B. That is, if $B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix}$, then we should define

$$AB := \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_n \end{bmatrix}. \tag{16.4}$$

16.5 Problem. With this definition of AB and still assuming $A \in \mathbb{F}^{m \times p}$ and $B \in \mathbb{F}^{p \times n}$, explain why $AB \in \mathbb{F}^{m \times n}$. Then check that (16.3) is indeed true, as desired.

Day 17: Friday, September 20.

We took Exam 1.

Day 18: Monday, September 23.

Material from *Linear Algebra* by Meckes & Meckes

Pages 97–100 discuss matrix inverses. We are beginning to discuss an operator-theoretic approach to row reduction and Gaussian elimination. You should already be familiar with all of Section 1.2 on elimination.

Do Quick Exercise #15 in Section 2.3

Vocabulary from today

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

Invertible matrix (N), matrix inverse, identity matrix

18.1 Example. We compute the matrix product

$$\begin{bmatrix}1 & 3\\ 2 & 4\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \tag{18.1}$$

using the "columnwise" definition. First, we compute the matrix-vector products

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

There are many ways to do this, including Definition 7.3 and Problem 16.2. Thus

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

The result is that we have interchanged the columns of the matrix on the left in (18.1); this is no accident, as the matrix on the right is a permutation matrix, which we will study later.

18.2 Problem. Compute

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

and conclude that matrix multiplication need not commute. Describe in words the result of this multiplication and contrast it to the result of the previous example.

Now that we have a way of multiplying matrices that corresponds to operator composition, we might ask what properties of operator composition extend to matrix multiplication. One such property is inverting operators. What is the right way to invert matrices?

Before doing that, we take a detour into inverting operator *compositions*. Suppose that $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ are vector spaces and

$$\mathcal{T}_1 \colon \mathcal{U} \to \mathcal{V} \quad \text{and} \quad \mathcal{T}_2 \colon \mathcal{V} \to \mathcal{W}$$

are both invertible. Experience suggests that $\mathcal{T}_2\mathcal{T}_1 \colon \mathcal{U} \to \mathcal{W}$ is also invertible, and that $(\mathcal{T}_2\mathcal{T}_1)^{-1} = \mathcal{T}_1^{-1}\mathcal{T}_2^{-1}$. Is it?

First, note that since $\mathcal{T}_2^{-1} \in \mathbf{L}(\mathcal{W},\mathcal{V})$ and $\mathcal{T}_1^{-1} \in \mathbf{L}(\mathcal{V},\mathcal{U})$, we have $\mathcal{T}_1^{-1}\mathcal{T}_2^{-1} \in \mathbf{L}(\mathcal{W},\mathcal{U})$. A consequence of Problem 15.9 is that we will have $(\mathcal{T}_2\mathcal{T}_1)^{-1} = \mathcal{T}_1^{-1}\mathcal{T}_2^{-1}$ if

$$(\mathcal{T}_2\mathcal{T}_1)(\mathcal{T}_1^{-1}\mathcal{T}_2^{-1}) = I_{\mathcal{W}} \quad \text{and} \quad (\mathcal{T}_1^{-1}\mathcal{T}_2^{-1})(\mathcal{T}_2\mathcal{T}_1) = I_{\mathcal{V}}. \tag{18.2}$$

We work through the first of these equalities "pointwise." Fix $w \in \mathcal{W}$. Then, successively applying the definition of operator composition multiple times, we have

$$\big((\mathcal{T}_2\mathcal{T}_1)(\mathcal{T}_1^{-1}\mathcal{T}_2^{-1})\big)w = \mathcal{T}_2(\mathcal{T}_1\mathcal{T}_1^{-1})\mathcal{T}_2^{-1}w = \mathcal{T}_2 I_{\mathcal{V}}\mathcal{T}_2^{-1}w = \mathcal{T}_2\mathcal{T}_2^{-1}w = I_{\mathcal{W}}w = w.$$

18.3 Problem. Check the second equality in (18.2).

We conclude what we expected.

18.4 Theorem. *Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces with $\mathcal{T}_1 \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{T}_2 \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ invertible. Then $\mathcal{T}_2\mathcal{T}_1 \in \mathbf{L}(\mathcal{U}, \mathcal{W})$ is invertible and*

$$(\mathcal{T}_2\mathcal{T}_1)^{-1} = \mathcal{T}_1^{-1}\mathcal{T}_2^{-1}.$$

Now we focus on matrices. Let $A \in \mathbb{F}^{m \times n}$. What *should* it mean for A to be invertible?

Experience suggests that we want a matrix S so that the products AS and SA are defined and equal identity matrices.

18.5 Problem. Check that $[I_{\mathbb{F}^n}] = I_n$, where I_n is the **IDENTITY MATRIX**, whose jth column is the standard basis vector (Definition 11.1) $\mathbf{e}_j$, i.e.,

$$I_n = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix}.$$

But what should the dimensions of S and those identity matrices be? Here it is helpful to remember the slogan that *what things do defines what things are.* Every matrix induces a linear operator. With $A \in \mathbb{F}^{m \times n}$, define

$$\mathcal{T}_A \colon \mathbb{F}^n \to \mathbb{F}^m \colon \mathbf{v} \mapsto A\mathbf{v},$$

so $A = [\mathcal{T}_A]$. We know what it means for $\mathcal{T}_A$ to be invertible: there must exist $\mathcal{S} \in \mathbf{L}(\mathbb{F}^m, \mathbb{F}^n)$ such that

$$\mathcal{T}_A\mathcal{S} = I_{\mathbb{F}^m} \quad \text{and} \quad \mathcal{S}\mathcal{T}_A = I_{\mathbb{F}^n}, \tag{18.3}$$

where $\mathcal{S}\mathbf{w} = [\mathcal{S}]\mathbf{w}$ with $[\mathcal{S}] \in \mathbb{F}^{n \times m}$. Recall that $I_{\mathbb{F}^n}$ is the identity operator on $\mathbb{F}^n$, i.e., $I_{\mathbb{F}^n} \in \mathbf{L}(\mathbb{F}^n)$ with $I_{\mathbb{F}^n}\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$.

We connect these equalities to matrices as follows. The equalities (18.3) say

$$[\mathcal{T}_A\mathcal{S}] = [I_{\mathbb{F}^m}] \quad \text{and} \quad [\mathcal{S}\mathcal{T}_A] = [I_{\mathbb{F}^n}],$$

equivalently,

$$[\mathcal{T}_A][\mathcal{S}] = I_m \quad \text{and} \quad \mathcal{S}[\mathcal{T}_A] = I_n,$$

and, last,

$$A[\mathcal{S}] = I_m \quad \text{and} \quad [\mathcal{S}]A = I_n.$$

This suggests that we should define $A \in \mathbb{F}^{m \times n}$ to be invertible if there is $S \in \mathbb{F}^{n \times m}$ such that $AS = I_m$ and $SA = I_n$. And this is what we will do.

There is just one problem: no such S can exist if $m \neq n$. That is, according to the rules of operator inverses, it is impossible for a nonsquare matrix to be invertible. More generally, the following negative result is true, although we do not have the tools to prove it yet.

18.6 Theorem. *Let m, $n \geq 1$ be integers with $m \neq n$ and let $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. Then $\mathcal{T}$ is not invertible.*

More positively, we have the following definition.

18.7 Definition. *A matrix $A \in \mathbb{F}^{n \times n}$ is* **INVERTIBLE** *if either of the following (equivalent) conditions holds:*

(i) *The operator $\mathcal{T}_A \colon \mathbb{F}^n \to \mathbb{F}^m \colon \mathbf{v} \mapsto A\mathbf{v}$ is invertible (in the sense of Theorem 15.6).*

(ii) *There exists $S \in \mathbb{F}^{n \times n}$ such that $AS = I_n$ and $SA = I_n$, where I_n is the $n \times n$ identity matrix (Problem 18.5). This matrix S is the* **INVERSE MATRIX** *of A.*

In fact, the second condition has a redundancy. Around the time that we prove Theorem 18.6, we will show that $S \in \mathbb{F}^{n \times n}$ satisfies $AS = I_n$ if and only if S also satisfies $SA = I_n$. That is, we only need to check one of those equalities for the other to hold. This result, like Theorem 18.6, is a powerful consequence of dimension counting arguments.

18.8 Problem (Optional, maybe annoying, definitely useful). Actually computing a matrix inverse is often computationally expensive and irrelevant; knowing the existence of the inverse is more important. One case that often arises in practice is the inverse of a 2×2 matrix. Show that

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

is invertible if and only if $ad - bc \neq 0$, in which case the inverse is

$$\frac{1}{ad - bc}\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}.$$

18.9 Problem. Prove that if $A \in \mathbb{F}^{n \times n}$ is invertible, its inverse is unique. That is, show that if there are matrices $B, C \in \mathbb{F}^{n \times n}$ such that

$$AB = BA = I_n \quad \text{and} \quad AC = CA = I_n,$$

then $B = C$. [Hint: *try doing this two ways. First, use results about the uniqueness of operator inverses. Second, start with $B = BI_n$. How do we get C to show up in an equality with I_n?*]

We now use our wealth of operator-theoretic and matrix-theoretic tools to study in detail linear systems of equations—recall our very first problem in Example 1.1. (This usually comes first in a first course on linear algebra: linear systems, then matrices, then linear operators. But we have a more evolved sensibility in this second course!) Remember as well that the fundamental problem of linear algebra is, arguably (and we should feel free to argue about this), to solve $\mathcal{T}v = w$ with $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, $\mathcal{V}$ and $\mathcal{W}$ vector spaces, and $w \in \mathcal{W}$ given. Here we address this problem in the form $A\mathbf{x} = \mathbf{b}$, with $A \in \mathbb{F}^{m \times n}$ and $\mathbf{b} \in \mathbb{F}^m$.

We begin with a very simple toy problem:

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

There are three "elementary row operations" that we can perform on this system to transform it into an "equivalent" system—that is, x_1 and x_2 solve the original system if and only if they solve the new one. First is the "interchange." The order in which we list equations does not matter:

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

Second is "scaling." If $\alpha \neq 0$, then we can multiply both sides of any equation by α and get an equivalent system. (This is because of the fundamental relationship that $x = y$ if and only if $\alpha x = \alpha y$ for $\alpha \neq 0$.) So, for example,

$$\begin{cases} x_1 - 2x_2 = 1 \\ 3x_1 + 2x_2 = 11. \end{cases} \iff \begin{cases} -3x_1 + 6x_2 = -3 \\ 3x_1 + 2x_2 = 11. \end{cases}$$

Third, we can add (a multiple of) one equation to another and get an equivalent system. This is possibly the strangest property, but we know something like $x = y$ if and only if $x + z = y + z$. Thus, for example,

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

To be really pedantic, in the second equation we took $x = 3x_1 + 2x_2$, $y = 11$, and $z = -3(x_1 - 2x_2)$. But from the first equation we could also say $z = -3$.

Day 19: Wednesday, September 25.

Material from *Linear Algebra* by Meckes & Meckes

Pages 12–14 perform row reduction on a 4×3 system. Pages 102–104 introduce the three kinds of elementary matrices.

Do Quick Exercises #5 and #6 in Section 1.2.

We study the system

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

exhaustively via row operations under several different lenses. First, we subtract 3 times the first equation (E1) from the second (E2) to find

$$\begin{cases} x_1 - 2x_2 = 1 \\ 3x_1 + 2x_2 = 11 \end{cases} \xrightarrow{\text{E2} \mapsto \text{E2}-3\times\text{E1}} \begin{cases} x_1 - 2x_2 = 1 \\ 8x_2 = 8. \end{cases}$$

The idiosyncratic "pseudocode" $(\text{E2}) \mapsto (\text{E2}) - 3 \times (\text{E1})$ is meant to suggest that the second equation on the left is replaced by that original second equation minus three times the first equation.

Next, we multiply both sides of the (new) second equation (E2) by 1/8 to find

$$\begin{cases} x_1 - 2x_2 = 1 \\ 8x_2 = 8 \end{cases} \xrightarrow{\text{E2} \mapsto 1/8\times\text{E2}} \begin{cases} x_1 - 2x_2 = 1 \\ x_2 = 1. \end{cases}$$

This tells us immediately the value of x_2: we have $x_2 = 1$. We substitute this into the first equation to find $x_1 - 2 = 1$, thus $x_1 = 3$. We have solved the system.

Here is how we could view this at the level of matrices. Put

$$A := \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} := \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

Then the original system is equivalent to $A\mathbf{x} = \mathbf{b}$. We introduce the "augmented matrix"

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

Then we subtract 3 times the first row (R1) from the second (R2):

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} \xrightarrow{\text{R2} \mapsto \text{R2}-3\times\text{R1}} \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 8 & 8 \end{array}\right].$$

Next we scale the (new) second row (R2) by 1/8:

$$\left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 8 & 8 \end{array}\right] \xrightarrow{\text{R2} \mapsto 1/8\times\text{R2}} \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 1 & 1 \end{array}\right].$$

This last matrix is the augmented matrix for the system

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

which we previously solved.

We could simplify this system even further by manipulating the first equation. If we add 2 times the second equation to the first (equivalently, and pedantically, subtract -2 times the second equation from the first), we have

$$\begin{cases} x_1 - 2x_2 = 1 \\ x_2 = 1 \end{cases} \xrightarrow{\text{E1} \mapsto \text{E1}-(-2)\times\text{E2}} \begin{cases} x_1 = 3 \\ x_2 = 1, \end{cases}$$

and there is nothing more to do. At the level of matrices, we have added 2 times the second row to the first (equivalently, and still pedantically, subtracted -2 times the second equation from the first):

$$\left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 1 & 1 \end{array}\right] \xrightarrow{\text{R1} \mapsto \text{R1}-(-2)\times\text{R2}} \left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 1 \end{array}\right]$$

One of the virtues of matrices is that they encode these "elementary row operations" via fairly "elementary" matrices. Put

$$E_{21} := \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}, \qquad D_{22} := \begin{bmatrix} 1 & 0 \\ 0 & 1/8 \end{bmatrix}, \quad \text{and} \quad E_{12} := \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}.$$

Each of these matrices is a variation on the 2×2 identity matrix; the subscript tells us what entry is somehow changed (although the notation does not tell us what the new value there is). The symbol E indicates that the matrix "eliminates" something, while D is a "diagonal" matrix that "scales" a row. The subscripts on E, read backward, tell us what row is subtracted from what; the matrix E_{21} causes (a multiple of) the first row to be subtracted from the second.

19.1 Problem. Let
$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$
Compute $E_{21}\mathbf{v}$, $D_{22}\mathbf{v}$, and $E_{12}\mathbf{v}$ and express the results in terms of v_1 and v_2. Describe in words how each matrix-vector multiplication changes the rows of $\mathbf{v}$.

The result is that
$$E_{21}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 8 & 8 \end{array}\right],$$
$$D_{22}E_{21}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 1 & 1 \end{array}\right],$$
and
$$E_{12}D_{22}E_{21}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 1 \end{array}\right]. \qquad (19.1)$$

Abbreviate
$$M := E_{12}D_{22}E_{21}.$$
Since we multiply matrices columnwise, we have
$$E_{12}D_{22}E_{21}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = M\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} MA & M\mathbf{b} \end{bmatrix}.$$
But by (19.1), this also reads
$$\begin{bmatrix} MA & M\mathbf{b} \end{bmatrix} = \begin{bmatrix} I_2 & \mathbf{c} \end{bmatrix}, \qquad \mathbf{c} := \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$
with I_2, as usual, as the 2×2 identity matrix. Since two matrices are equal if and only if their corresponding columns are equal, we get $MA = I_2$. Thus A is invertible and $A^{-1} = M$ (which we could have checked by Problem 18.8).

Here is what we should take away from this systematic analysis of this overly simple problem.

- The elementary row operations of subtracting a multiple of one row from another and scaling a row by a nonzero number can transform a "complicated" system into an equivalent "simpler" one whose solution, if it exists, we can more or less read off from the structure of the "simpler" system.

- These two elementary row operations can be encoded by matrix multiplication, where the "elementary" matrices involved are identity matrices with one entry replaced by the multiplier factor (or the negative thereof).

- If $A \in \mathbb{F}^{n \times n}$ is invertible, then we can write A^{-1} as the product of these "elementary" matrices.

Strictly speaking, we have proved none of these statements. Also, missing is any discussion of interchanging rows/equations. We will see an example later where that is necessary. For now, we do another, larger example for a specific $A \in \mathbb{F}^{3\times 3}$ in which we construct a product M of elementary matrices such that MA is the identity matrix. In principle, this would allow us to solve $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{F}^3$. (Computationally, this is "expensive" and a bad idea.)

19.2 Example. Let

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

We multiply A by a series of elimination and scaling elementary matrices so that the final product is the identity matrix:

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{bmatrix} \xrightarrow[E_{21}]{\text{R2} \mapsto \text{R2}-2\times\text{R1}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 8 & 7 & 9 \end{bmatrix}, \qquad E_{21} := \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

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

$$\xrightarrow[D_{33}]{\text{R3} \mapsto (1/2)\times\text{R3}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad D_{33} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$$

$$\xrightarrow[E_{23}]{\text{R2} \mapsto \text{R2}-1\times\text{R3}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{23} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[E_{13}]{\text{R1} \mapsto \text{R1}-1\times\text{R3}} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_{13} := \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

$$\xrightarrow[D_{11}]{\text{R1} \mapsto (1/2)\times\text{R1}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad D_{11} := \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It follows that

$$D_{11}E_{13}E_{23}D_{33}E_{32}E_{31}E_{21}A = I_3,$$

and so

$$A^{-1} = D_{11}E_{13}E_{23}D_{33}E_{32}E_{31}E_{21}.$$

19.3 Problem (Optional, tedious, maybe worthwhile). Check that each of the elementary matrices in the previous example "does what it should." For example $E_{31}\mathbf{v}$ subtracts 4 times the first row of $\mathbf{v}$ from the third row.

We did not bother multiplying all of those elementary matrices out to find the actual formula for A^{-1}. If a formula is absolutely necessary, a tried-and-true algorithm for hand computation of A^{-1} for invertible $A \in \mathbb{F}^{n\times n}$ is to work on the "identity-augmented matrix" $\begin{bmatrix} A & I_n \end{bmatrix}$ and perform on this entire matrix the row operations that reduce A to I_n. Collecting the elementary matrices that perform this matrix into a single matrix M, we have

$$M \begin{bmatrix} A & I_n \end{bmatrix} = \begin{bmatrix} I_n & B \end{bmatrix},$$

where B is "whatever happens to I_n." More precisely, of course, $MA = I_n$ and $MI_n = B$, thus, again, $M = A^{-1}$.

Day 20: Friday, September 27.

No class due to university closure.

Day 21: Monday, September 30.

We continue trying to solve (and, more importantly, trying to understand) the linear problem $A\mathbf{x} = \mathbf{b}$ by performing elementary row operations on A (and $\mathbf{b}$) to reduce A to the identity matrix. The strategy is that we use the "leading nonzero entry" (the "pivot") in a row to "zero out" the column below that entry and keep doing so, moving left to right, until A is "transformed" into an upper-triangular matrix. Then we "work upwards" and create zeros above the "pivots," along with rescaling the pivots to be 1. This process is called **Gauss–Jordan elimination** (the downwards step is just **Gaussian elimination**).

Here is a situation in which elimination and scaling alone are not enough.

21.1 Example. Let

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

As before, we create zeros in the first column below the $(1,1)$-entry by multiplying by elimination matrices:

$$E_{31}E_{21}A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 3 & 5 \end{bmatrix}.$$

The problem is the new $(2,2)$-entry of 0: that is not allowed on the diagonal of the identity matrix. (The other problem is the new $(3,2)$-entry of 3, which is also not allowed.)

The right idea is to swap the second and third rows. That is, we want to multiply $E_{31}E_{21}A$ by a matrix $P \in \mathbb{R}^3$ such that

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

There are several ways of determining such P, including rearranging algebraically the right side of this desired equality or taking the vectors on which P acts to be the standard basis vectors. However we do it, the right P is

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

This is a "permutation" matrix, since its columns (and rows) are rearrangements (permutations) of the columns (and rows) of the identity matrix. We find

$$PE_{31}E_{21}A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 5 \\ 0 & 0 & 1 \end{bmatrix},$$

and from here we can find a product M of "elementary" matrices such that $MPE_{31}E_{21}A = I_3$.

21.2 Problem. Find that product M.

21.3 Problem. Do as we said and not as we did and define

$$\mathcal{T} \colon \mathbb{R}^3 \to \mathbb{R}^3 \colon \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \mapsto \begin{bmatrix} v_1 \\ v_3 \\ v_2 \end{bmatrix}.$$

Find the matrix representation of $\mathcal{T}$ and check that it is P as stated above in the previous example.

We formalize the notation of permutation matrix.

21.4 Definition. *A matrix $P \in \mathbb{R}^{n \times n}$ is a* **PERMUTATION MATRIX** *if the columns of P are the columns of the $n \times n$ identity matrix I_n. That is, the columns of P are the standard basis vectors for $\mathbb{R}^n$.*

21.5 Problem. Let $n \geq 1$ be an integer and let $1 \leq i, \ j \leq n$. Let P_{ij} be the permutation matrix whose ith column is the standard basis vector $\mathbf{e}_j$ and whose jth column is the standard basis vector $\mathbf{e}_i$.

(i) Let $\mathbf{v} \in \mathbb{F}^n$. For $k = 1, \ldots, n$, what is the kth row of $P_{ij}\mathbf{v}$ in terms of the rows of $\mathbf{v}$?

(ii) Let $A \in \mathbb{F}^{n \times n}$. Compare and contrast the effects of multiplying $P_{ij}A$ versus AP_{ij}.

(iii) Why is P_{ij} invertible? What is its inverse?

Now let $P \in \mathbb{F}^{n \times n}$ be *any* permutation matrix. Answer the questions above with P_{ij} replaced by P.

21.6 Problem. Let $A \in \mathbb{F}^{m \times n}$, let $1 \leq p \leq n$, and let $1 \leq j_1 < \cdots < j_p \leq n$. Suppose that we want to "select" columns j_1 through j_p of A and put them in a matrix in the order in which they appeared in A. This will result in an $m \times p$ matrix. Find a matrix $S \in \mathbb{F}^{n \times p}$ such that AS has this structure. [Hint: *revisit Problem 13.7. As needed, work with some small matrices until you see the right pattern in S in general.*]

Here is a situation in which A is not invertible, and so we will fail to solve $A\mathbf{x} = \mathbf{b}$ for all $\mathbf{b}$. We first discuss how Gauss–Jordan elimination fails, and then we seek to *understand* what this failure says about the extent to which we can solve $A\mathbf{x} = \mathbf{b}$.

21.7 Example. Let

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

We might notice that the second row of A is double the first row, and that should lead us to expect problems. Indeed, they arise via the elimination

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

The problem is that row of zeros. There is no way to multiply A by any other "elementary" matrices so that a 1 appears in the $(2,2)$-entry of that product.

Here is what elimination says about solving $A\mathbf{x} = \mathbf{b}$. Suppose that this equality is true. Then $E_{21}A\mathbf{x} = E_{21}\mathbf{b}$. That is, if we can solve $A\mathbf{x} = \mathbf{b}$, we must have

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix} \mathbf{x} = E_{21}\mathbf{b}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Computing each side, we arrive at (return to?) the linear system

$$\begin{cases} x_1 + 2x_2 + 3x_3 = b_1 \\ 0 = b_2 - 2b_1 \\ 5x_3 = b_3. \end{cases}$$

The second equation is the killer. It says $b_2 - 2b_1 = 0$, so $b_2 = 2b_1$. This is a "solvability condition" for the problem: if there is a solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$, then the entries of $\mathbf{b}$ must meet $b_2 = 2b_1$. (To be fair, this is only a condition relating the first two entries of $\mathbf{b}$; it says nothing about b_3.) If $b_2 \neq 2b_1$, then no solution can exist, and there is no point in trying to solve the problem.

What if the solvability condition is met, and $b_2 - 2b_1 = 0$? There are still two other equations in play. The third immediately gives $x_3 = b_3/5$, which we substitute into the first to find that x_1 and x_2 must meet

$$x_1 + 2x_2 + \frac{3b_3}{5} = b_1.$$

This is one equation in two unknowns—not a recipe for unique solutions. One approach is to solve for x_1 in terms of x_2:

$$x_1 = b_1 - \frac{3b_3}{5} - 2x_2.$$

Then for every choice of x_2, we get a new x_1. Together, they form a solution to $A\mathbf{x} = \mathbf{b}$.

In the language of vectors, we have

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 - 3b_3/5 - 2x_2 \\ x_2 \\ b_3/5 \end{bmatrix} = \begin{bmatrix} b_1 - 3b_3/5 \\ 0 \\ b_3/5 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

This should remind us of Problem 11.10. (Does it?) On one hand, we have infinitely many solutions to $A\mathbf{x} = \mathbf{b}$ given the solvability condition $b_2 = 2b_1$. On the other, all solutions have a very similar form, and there is only one "degree of freedom" given by that multiplier x_2.

This problem raises (at least) two questions, neither of which we have the tools to answer just yet.

1. Is there a more meaningful way to describe the size of the solution set to $A\mathbf{x} = \mathbf{b}$ than "infinite"? Yes: we need bases.

2. Is there a way to "rig the game" so that we can always solve $A\mathbf{x} = \mathbf{b}$ uniquely if we specify $\mathbf{x}$ and $\mathbf{b}$ correctly? That is, are there subspaces $\mathcal{V}, \mathcal{W} \subseteq \mathbb{R}^3$ such that the operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{W} \colon \mathbf{v} \mapsto A\mathbf{v}$ is an isomorphism? Yes: we need the geometry of inner products and orthogonal complements.

21.8 Problem. We could have gone further with elementary row operations in the previous example. Find a matrix $E \in \mathbb{R}^{3\times 3}$ such that

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

Express E as the product of "elementary" matrices; you do not need to multiply all of that product out. The first factor in that product (on the right) will be E_{21} from the example; a permutation matrix will also be necessary in the mix.

We finally study a nonsquare matrix. Theorem 18.6 should lead us to believe that we will somehow fail to solve $A\mathbf{x} = \mathbf{b}$ in this case—either no $\mathbf{x}$ will exist, or it will not be unique. For now, we just focus on arithmetic and matrix multiplication: we want to make A look as much like the identity matrix as possible.

21.9 Example. Let

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

so $A \in \mathbb{R}^{3\times 4}$. If we want to multiply A by some matrix B and have the product BA defined, we need $B \in \mathbb{R}^{p\times 3}$ for some p. And if we want B to be invertible, we need $p = 3$. So, the elementary matrices that we apply to A here will continue to be 3×3.

Here we go:

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

$$\xrightarrow[D_{33}]{\mathsf{R3} \mapsto (1/2)\times\mathsf{R3}} \begin{bmatrix}1 & 2 & 1 & 7\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 4\end{bmatrix}, \qquad D_{33} := \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1/2\end{bmatrix}$$

$$\xrightarrow[E_{13}]{\mathsf{R1} \mapsto \mathsf{R1}-\mathsf{R3}} \begin{bmatrix}1 & 2 & 0 & 3\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 4\end{bmatrix}, \qquad E_{13} := \begin{bmatrix}1 & 0 & -1\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$\xrightarrow[P_{23}]{\mathsf{R2} \mapsto \mathsf{R3},\ \mathsf{R3} \mapsto \mathsf{R2}} \begin{bmatrix}1 & 2 & 0 & 3\\ 0 & 0 & 1 & 4\\ 0 & 0 & 0 & 0\end{bmatrix}, \qquad P_{23} := \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}.$$

We have done all the row operations that worked in the past to convert a square A to the identity matrix. How exactly should we interpret the result here?

Day 22: Wednesday, October 2.

Material from *Linear Algebra* by Meckes & Meckes

All of Section 1.2 is worth reading line by line. Theorem 1.1 provides the algorithm for Gaussian elimination. Pages 102–104 discuss elementary matrices. Theorems 2.22 and 2.24 give factorizations of the RREF.

Do all of the Quick Exercises in Section 1.2. Do Quick Exercises #16 and #17 in Section 2.4

Here is a summary of what we have been calculating. We have worked on matrices A_1, $A_2 \in \mathbb{R}^{3\times 3}$ and $A_3 \in \mathbb{R}^{3\times 4}$ and found matrices E_1, E_2, $E_3 \in \mathbb{R}^{3\times 3}$ such that

$$E_1A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$E_2A_2 = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

and

$$E_3A_3 = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Such matrices A_1 appeared in both Examples 19.2 and 21.1 with A_2 in Example 21.7 and Problem 21.8 and A_3 in Example 21.9. The matrices E_1, E_2, and E_3 were products of what we called, euphemistically, "elementary matrices." After a lot of experience, we now define this concept precisely.

22.1 Definition. *An* ELEMENTARY MATRIX *is one of the following three kinds of $n \times n$ matrices.*

(i) *An* ELIMINATION MATRIX E_{ji} *(where $i \neq j$) such that $E_{ji}A$ is formed by subtracting α times row i of A from row j of A. The matrix E_{ji} is formed by replacing the (i, j)-entry of the $n \times n$ identity matrix I_n with $-\alpha$.*

(ii) *A* SCALING MATRIX D_{ii} *such that $D_{ii}A$ is formed by multiplying row i of A by α. The matrix D_{ii} is formed by replacing the (i, i)-diagonal entry of I_n by α.*

(iii) *A* PERMUTATION MATRIX P_{ij} *such that $P_{ij}A$ is formed by interchanging rows i and j of A. The matrix P_{ij} is formed by interchanging columns i and j of I_n. (This is a special case of Definition 21.4.)*

22.2 Problem. Prove that any elementary matrix is invertible in at least two ways. First, explain in words what the inverse of each kind of elementary matrix should *do*. Then describe (in words, like the previous definition) how to construct each inverse out of I_n. (You do not have to verify that the product of these putative inverses and the original elementary matrices is I_n.) Optionally, explain why the multiplication operators induced by these elementary matrices (e.g., $\mathcal{T}_{E_{ij}}\mathbf{v} = E_{ij}\mathbf{v}$) is invertible by considering injectivity and surjectivity.

Now we observe the special structures of the matrices $E_k A_k$ above, $k = 1, 2, 3$. These matrices have what we will call the REDUCED ROW ECHELON FORM **(RREF)** structure.

RREF1. Any row whose entries are all 0 is below every row with nonzero entries:

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

RREF2. The first nonzero entry of any row is 1:

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

We euphemistically call such an entry of 1 a LEADING 1.

RREF3. Any leading 1 is the only nonzero entry in its column.

RREF4. Each of these matrices contains columns of identity matrices. We first highlight (in blue) the columns of I_3:

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

Now we highlight (still in blue) the columns of I_2 in the second and third matrices:

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

(Strictly speaking, the columns of I_2 similarly appear in the first matrix, but more interesting will be the case when we do row operations and do not get an identity matrix.) Both kinds of columns appear in the order that they do in I_2 and I_3, although not all columns of I_3 are present.

We can make the matrices $E_2 A_2$ and $E_3 A_3$ a bit nicer by permuting their *columns*. Now is a good time to look at Problem 21.5. We have

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I_2 & F \\ 0 & 0 \end{bmatrix}, \qquad F := \begin{bmatrix} 2 \\ 0 \end{bmatrix} \tag{22.1}$$

and

$$\begin{bmatrix}1&2&0&3\\0&0&1&4\\0&0&0&0\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}=\begin{bmatrix}1&0&2&3\\0&1&0&4\\0&0&0&0\end{bmatrix}=\begin{bmatrix}I_2&F\\0&0\end{bmatrix},\qquad F:=\begin{bmatrix}2&3\\0&4\end{bmatrix}. \tag{22.2}$$

In the last matrices on the right in each calculation, the bottom rows of 0 represent not the scalar 0 but zero *matrices*. (What are their sizes?)

Up to a permutation matrix, the matrices E_2A_2 and E_3A_3 have a special "block" structure: the top left block is an identity matrix, the bottom two blocks are all 0, and the top left block is "junk." Every matrix (except the zero matrix) can be written in this form.

This leads us to two interpretations of the RREF. First, it is a "canonical form" satisfying the four conditions above. Second, it is a special "factorization" of a matrix. Both are useful interpretations.

22.3 Theorem (RREF: existence and uniqueness). *Let $A \in \mathbb{F}^{m\times n}$. There exists $E \in \mathbb{F}^{m\times m}$ such that E is a product of elementary matrices and $R := EA$ is in* REDUCED ROW ECHELON FORM (RREF) *in the sense that it satisfies the following.*

Row Property 1. *Any nonzero row of R is below any row with nonzero entries.*

Row Property 2. *If a row contains nonzero entries, the first nonzero entry of that row is 1, called the* LEADING *1 or the* PIVOT *for that row.*

Column Property 1. *The other entries of any column containing a leading 1 are 0. That is, a column containing a leading 1 is a column of the $m\times m$ identity matrix I_m, equivalently, a standard basis vector for $\mathbb{F}^m$. Such a column is called a* PIVOT COLUMN.

Column Property 2. *If columns i and j of I_m appear in R and $i < j$, then the first appearance of column i must be to the left of column j.*

The matrix R is unique in the sense that if EA and $\widetilde{E}A$ both satisfy the four properties above with E and $\widetilde{E}$ invertible, then $EA = \widetilde{E}A$. We sometimes write $R = \mathsf{rref}(A)$.

Proof. Both existence and uniqueness are a byproduct of the "proof by algorithm" given in Theorem 1.1 in the book. Existence should be obvious: just do the elimination. For uniqueness, the key point of the algorithm is that we can run it *in the same way every time*. In the language of the book, when using row operation R3, if necessary, to "get a nonzero entry at the top of the column" or "in the second row of [the] column," just select the *first* nonzero entry in the column under consideration. When using row operation R1, work strictly downward (eliminate in row i before row j when $i < j$). When using row operation R1 "to make all the entries in the same column as any pivot 0" work strictly upward (eliminate in row j before row i when $i < j$).

This ensures that we always reach the RREF of A in the same way. (This is nice theoretically but may be problematic numerically; there are good reasons not to "pivot" with just the first nonzero entry of a column all the time.) If we then *define* the RREF of A to be the matrix that we reach via this specific sequence of elementary row operations, then that matrix will have the four properties above.

However, it may be less clear that no matter how we do the elementary row operations, we always reach the same RREF, or that if we can write $EA = R$ with E invertible (if necessary, specifying that E is a product of elementary matrices) and R in RREF, then there is only one choice for R. (There will definitely *not* be only one choice for E, as we certainly *can* do the elementary row operations in different orders.)

We give a proof of uniqueness (which in particular only relies on E being invertible, not the product of elementary matrices) later, after building more intuition. ■

22.4 Problem. Explain *all* of the reasons why

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

is not in RREF.

22.5 Problem. By considering the matrix

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

explain the importance of the adjective "first" in Column Property 2 of Theorem 22.3.

The RREF factorizations that appeared in (22.1) and (22.2) can be hugely useful.

22.6 Theorem (RREF: factored version). *Let $A \in \mathbb{F}^{m \times n} \setminus \{0\}$. Then there exist an integer r with $1 \le r \le \min\{m, n\}$ and invertible matrices $E \in \mathbb{F}^{m \times m}$ and $P \in \mathbb{F}^{n \times n}$ such that $EA = \mathsf{rref}(A)$ and*

$$EA = \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P. \tag{22.3}$$

The integer r is unique.

Proof. Write $R = \mathsf{rref}(A)$, so $R = EA$ for some invertible E. (It is possible that there are multiple such E.) Let r be the number of columns in R that are columns of the $m \times m$ identity matrix I_m. (Equivalently, and importantly, r is the number of rows of R with a leading 1.) Let P be a permutation matrix such that the first r of RP columns are the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r$ for $\mathbb{F}^m$. (It is possible that there are multiple such P.) Then RP has the form on the right in (22.3).

For the uniqueness of r, if $EA = \mathsf{rref}(A)$ and EA also has the factorization in (22.3), then r is the number of columns of the $m \times m$ identity matrix that appear in $\mathsf{rref}(A)$. This ensures that there is only one possible value for r in that factorization. ■

22.7 Example. Here is how we interpret the special structure of the block matrix in (22.3), which we hereafter call B, so $\mathsf{rref}(A) = BP$. (It is important to remember that B is not necessarily $\mathsf{rref}(A)$.) First,

$$B := \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} \in \mathbb{F}^{m \times n}.$$

This is necessary for the product on the right of (22.3) to be defined and to have the same dimensions as the product on the left, given the specified sizes of A, E, and P. The zero blocks and F may or may not be present.

(i) We allow the case $r = n < m$, in which case

$$B = \begin{bmatrix} I_n \\ 0 \end{bmatrix}.$$

The block F is no longer present, as otherwise B would have more than n columns. The matrix zero block at the bottom must be present, as B has m rows, but I_n has $n < m$ rows. Also, $\mathsf{rref}(A) = B$ and here $P = I_n$, as otherwise Column Property 2 would not hold. This shows that not every permutation matrix is allowed in (22.3).

For example, we could have $r = n = 2$ and $m = 3$ and

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

(ii) We allow the case $r = m < n$, in which case

$$B = \begin{bmatrix} I_m & F \end{bmatrix}.$$

The block F must be present (it could be anything—maybe all 0), but if F is not present, then $B = I_m$, which violates the inequality $m < n$. The zero blocks cannot be present, as I_m has m rows, so with zero blocks B would have more than m rows.

For example, we could have $r = m = 2$ and $n = 3$ and

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

(iii) We allow the case $r = m = n$, in which case all matrices are square and $R = I_n = I_m$. Otherwise, with F present, B would have more than n columns, and with zero blocks, B would have more than m rows.

(iv) If $r < m$ and $r < n$, then B must have both zero blocks and the F block. Without the zero blocks, it would be the case that B would have $r < m$ rows; without the F block, it would be the case that B would have $r < n$ columns.

For example, we could have $r = 2$, $m = 3$, and $n = 4$ and use what will be one of our favorite recurring examples:

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

22.8 Problem. Example 21.9 constructs $E \in \mathbb{R}^{3\times3}$ such that

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

(i) By revisiting the elementary row operations in that example, explain why E in Theorem 22.6 might not be unique. [Hint: *could P_{23} have appeared earlier? Could D_{33} have appeared later?*]

(ii) Find a permutation matrix $\widetilde{P}$ such that

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

Contrast this with (22.2) and explain how this shows that P and F from Theorem 22.6 may not be unique.

(iii) Explain why there cannot exist a matrix $A \in \mathbb{R}^{3\times4}$ such that

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

Conclude (as noted in part (i) of Example 22.7) that not every permutation matrix P can appear in (22.3).

22.9 Problem. Find a matrix $A \in \mathbb{R}^{3\times4}$ whose entries are all nonzero such that

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

Provide a matrix $E \in \mathbb{R}^{3\times3}$ such that $EA = \mathsf{rref}(A)$; you may express E as a product of elementary matrices, and you do not have to multiply that product out.

22.10 Example. We find all $R \in \mathbb{F}^{3\times5}$ that are in RREF with a leading 1 in columns 2 and 4 only. Throughout, we denote by $*$ an entry whose value may be an arbitrary number in $\mathbb{F}$. We break our reasoning into the following steps.

1. The first column of R must be $\mathbf{0}$. Otherwise, that column would contain a nonzero entry, and that nonzero entry would be the leading nonzero entry in its row. Then that entry would have to be 1, but the first column of R does not contain a leading 1.

2. Since the second column of R must contain a leading 1, we therefore have three possibilities for R currently:

$$\begin{bmatrix} 0 & 1 & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 & * & * & * \\ 0 & 1 & * & * & * \\ 0 & 0 & * & * & * \end{bmatrix}, \qquad \text{or} \qquad \begin{bmatrix} 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 1 & * & * & * \end{bmatrix}. \tag{22.4}$$

3. We claim that the second and third cases are impossible. In the second case, the first row cannot be all 0, as then we would have a row with entries all 0 (row 1) above a row with some nonzero entries (row 2). Any nonzero entry in the first row would be the leading nonzero entry in that row and therefore 1, and so the rest of that column would be 0. But then the standard basis vector $\mathbf{e}_2$ for $\mathbb{F}^3$ would appear for the first time (in column 2) before $\mathbf{e}_1$ (in columns 3, 4, or 5).

4. The same contradiction results in the third case: either the first row is 0 or $\mathbf{e}_1$ appears for the first time only after $\mathbf{e}_3$.

5. So, R must have the form of the first matrix in (22.4). The RREF properties in Theorem 22.3 say nothing about the $(1,3)$-entry of this matrix: it is not a leading nonzero entry in row 1 (and so it does not have to be 1), and column 3 does not contain a leading 1 (and so this entry does not have to be 0). So, we leave it arbitrary. However, the other entries of column 3 must be 0, as otherwise they would be leading nonzero entries in rows 2 and 3. So, R has the form

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

6. Since column 4 contains a leading 1, and since there is already a leading 1 in row 1, this leading 1 in column 4 must appear in rows 2 or 3. So, there are two possibilities for R now:

$$\begin{bmatrix} 0 & 1 & * & 0 & * \\ 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & * \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} 0 & 1 & * & 0 & * \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 1 & * \end{bmatrix}.$$

7. We claim that the third case is impossible. For in this case the $(2,5)$-entry of row 2 must be 0, as otherwise there would be a leading nonzero entry in column 5. Then the entries of row 2 are all 0, but row 3 has a nonzero entry.

8. We are down to

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

as the only possible form of R. The RREF properties say nothing about the $(1,5)$-entry, as it is not a leading nonzero entry in row 1, and there is no leading 1 in column 5. The same holds for the $(2,5)$-entry. But the $(3,5)$-entry must be 0, as otherwise there would be a leading nonzero entry in column 5, which would have to be a leading 1. So, R has the form

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

9. That is, all matrices $R \in \mathbb{F}^{3\times 5}$ that are in RREF with a leading 1 in columns 2 and 4 only have the form

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

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

22.11 Problem. Let $\mathcal{V}$ be the set of all matrices R described at the end of Example 22.9. Is $\mathcal{V}$ a subspace of $\mathbb{F}^{3\times 5}$?

22.12 Problem. Generalize one of the arguments in Example 22.9 as follows. Let $A \in \mathbb{F}^{m\times n}$ and $EA = R$, where $E \in \mathbb{F}^{m\times m}$ is invertible and R is in RREF. Prove that the first column of R is either the zero vector $\mathbf{0} \in \mathbb{F}^m$ or the first standard basis vector $\mathbf{e}_1 \in \mathbb{F}^m$. Go further and argue that the first *nonzero* column of R is $\mathbf{e}_1$.

The following example is designed to be used in the proof of uniqueness for the RREF.

22.13 Example. Here another properties of the RREF that we might tease out of the experience of building Example 22.9. Let $R \in \mathbb{F}^{m\times n}$ be in RREF.

(i) *The (j,j)-entry of R is either 0 or 1 for $j = 1, \ldots, \min\{m,n\}$. That is, on the "diagonal" R is either 0 or 1. Moreover, the entries of column j are 0 in rows $j+1$ and below.*

Here is why. By Problem 22.12, the first column of R is either $\mathbf{0}$ or $\mathbf{e}_1$; that takes care of the $(1,1)$-entry. Suppose the $(2,2)$-entry is nonzero; then because the $(2,1)$-entry is 0 (it is the second entry of $\mathbf{e}_1$), the $(2,2)$-entry must be the leading nonzero entry row 2 and thus 1. Then every entry in column 2 that is not in row 2 must be 0; in particular, the $(3,2)$-entry is 0. And also the $(3,1)$-entry is 0 since the first column is either $\mathbf{0}$ or $\mathbf{e}_1$. Do it again: if the $(3,3)$-entry is nonzero, then since the $(3,1)$- and $(3,2)$-entries are 0, the $(3,3)$-entry is the leading nonzero entry in row 3 and thus 1.

Now turn the crank: if we know that the entries of column j, $1 \le j \le k$, are 0 in rows

$j+1$ and below, then the $(k+1, j)$-entries are 0 for $j = 1, \ldots, k$. So, if the $(k+1, k+1)$-entry is nonzero, it must be the leading nonzero entry in row $k+1$ and thus 1.

(ii) *If $\mathbf{e}_j$ is a column of R and $i < j$, then $\mathbf{e}_i$ must appear as a column of R at least once before $\mathbf{e}_j$.* Here is why—suppose $\mathbf{e}_i$ does not appear at all. Then the entries of row i of R are all 0; otherwise, some nonzero entry in row i would be the leading nonzero entry in that row, and then $\mathbf{e}_i$ would appear. So, $\mathbf{e}_i$ appears as *some* column of R, and so it must appear *before* $\mathbf{e}_j$.

(iii) *Any column of R that appears before the first appearance of $\mathbf{e}_i$ has zeros in rows i and below.* Let $\mathbf{r}$ be such a column and suppose that an entry of $\mathbf{r}$ in row i or below is nonzero. Since $\mathbf{e}_i$ has not yet appeared, all entries in row i in the columns before $\mathbf{r}$ must be 0; otherwise, such an entry would be the leading nonzero entry in row i, which would force a prior appearance of $\mathbf{e}_i$. So, if $\mathbf{r}$ has a nonzero entry in row i, that is the leading nonzero entry in row i, and therefore $\mathbf{r} = \mathbf{e}_i$. But $\mathbf{e}_i$ has not yet appeared.

This nonzero entry of $\mathbf{r}$ must therefore be in row $i+1$ or below. If this entry is the leading nonzero entry of row $i+1$, then $\mathbf{r} = \mathbf{e}_{i+1}$. But then $\mathbf{e}_{i+1}$ appears for the first time before $\mathbf{e}_i$. If this entry is not the leading nonzero entry of row $i+1$, then row $i+1$ has its leading nonzero entry in some column before $\mathbf{r}$. And that column is then $\mathbf{e}_{i+1}$, which again appears before $\mathbf{e}_i$.

22.14 Problem. **(i)** Determine all possible forms of a matrix $R \in \mathbb{F}^{1\times n}$ in RREF.

(ii) Determine all possible forms of a matrix $R \in \mathbb{F}^{m\times 1}$ $(= \mathbb{F}^m)$ in RREF.

22.15 Problem. Let $A \in \mathbb{F}^{m\times n}$ and $EA = R$, where $E \in \mathbb{F}^{m\times m}$ is invertible and R is in RREF. Prove that the jth column of A is $\mathbf{0}$ if and only if the jth column of R is $\mathbf{0}$. [Hint: *think about the matrix products $EA = R$ and $A = E^{-1}R$ using the definition* (16.4) *of matrix multiplication—what happens when one column in the factor on the right is $\mathbf{0}$?*]

Here is the proof of uniqueness of the RREF of $A \in \mathbb{F}^{m\times n}$, as claimed in Theorem 22.3. This is optional reading. The argument is based on a method of Holzmann, available here:

`https : //www.cs.uleth.ca/ holzmann/notes/reduceduniq.pdf.`

Assume that $A \neq 0$, as otherwise $R = \widetilde{R} = 0$. Let $EA = R$ and $\widetilde{E}A = \widetilde{R}$, where E, $\widetilde{E} \in \mathbb{F}^{m\times m}$ are invertible and R, $\widetilde{R} \in \mathbb{F}^{m\times n}$ are in RREF. By Problem 22.12, the first nonzero column of R is $\mathbf{e}_1$. By Problem 22.15, any zero column of R corresponds to a zero column of A and thus to a zero column of $\widetilde{R}$. We conclude that R and $\widetilde{R}$ have the same number of leading zero columns (if any), and that the first nonzero column of each is $\mathbf{e}_1$, and $\mathbf{e}_1$ occurs for the first time in the same column in both R and $\widetilde{R}$.

Suppose that the first p columns of R and $\widetilde{R}$ starting with this first occurrence of $\mathbf{e}_1$ agree (so $p \geq 1$). Denote by $\mathbf{r}$ and $\widetilde{\mathbf{r}}$ the $(p+1)$st column after this first occurrence of $\mathbf{e}_1$. Suppose that the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_j$, $j \geq 1$, appear within these first p columns of R and $\widetilde{R}$ starting with the first occurrence of $\mathbf{e}_1$ (since these p columns agree, the same standard basis vectors must appear among them).

Let $S \in \mathbb{F}^{n\times(j+1)}$ be the selection matrix (Problem 21.6) that selects $\mathbf{e}_1, \ldots, \mathbf{e}_j$ from within these first p columns and also the $(p+1)$st column. Then

$$EAS = RS = \left[\begin{array}{c|c} I_j & \mathbf{r} \\ 0 & \end{array}\right] \quad \text{and} \quad \widetilde{E}AS = \widetilde{R}S = \left[\begin{array}{c|c} I_j & \widetilde{\mathbf{r}} \\ 0 & \end{array}\right]. \tag{22.5}$$

That is, the first p columns of EAS and $\widetilde{E}AS$ are the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_p$ for $\mathbb{F}^m$, while the last column is $\mathbf{r}$ and $\widetilde{\mathbf{r}}$, respectively.

For example, if

$$R = \begin{bmatrix} 0 & 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \widetilde{R} = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

then $p = 4$ and $j = 2$ and we would choose S to select columns 2, 3, and 4 from R and $\widetilde{R}$, thus

$$EAS = RS = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \widetilde{E}AS = \widetilde{R}S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Now we examine the structure of $\mathbf{r}$ and $\widetilde{\mathbf{r}}$. Since the only standard basis vectors to the left of $\mathbf{r}$ in R are $\mathbf{e}_1, \ldots, \mathbf{e}_j$, either $\mathbf{r} = \mathbf{e}_{j+1}$ or $\mathbf{r}$ is 0 in rows $j+1$ and below. This is a result from Example 22.13. Consequently, we can refine the structure in (22.5) to read

$$\underbrace{EAS = \left[\begin{array}{c|c} I_j & \mathbf{r}_0 \\ 0 & 0 \end{array}\right]}_{I} \quad \text{or} \quad \underbrace{EAS = \left[\begin{array}{c|c} I_j & \mathbf{e}_{j+1} \\ 0 & \end{array}\right]}_{II}$$

and

$$\underbrace{\widetilde{E}AS = \left[\begin{array}{c|c} I_j & \widetilde{\mathbf{r}}_0 \\ 0 & \end{array}\right]}_{\widetilde{I}} \quad \text{or} \quad \underbrace{\widetilde{E}AS = \left[\begin{array}{c|c} I_j & \mathbf{e}_{j+1} \\ 0 & \end{array}\right]}_{\widetilde{II}},$$

where $\mathbf{r}_0$ and $\widetilde{\mathbf{r}}_0$ contain the first j rows of $\mathbf{r}$ and $\widetilde{\mathbf{r}}$.

If both cases II and $\widetilde{II}$ hold, then $\mathbf{r} = \mathbf{e}_{j+1} = \widetilde{\mathbf{r}}$. Suppose that cases I and $\widetilde{I}$ hold. The idea is now to view EAS and $\widetilde{E}AS$ as augmented matrices. Specifically, let $\widehat{S}$ be the selection matrix that selects just the columns of EA and $\widetilde{E}A$ containing $\mathbf{e}_1, \ldots, \mathbf{e}_j$ up to column p, so

$$EA\widehat{S} = \begin{bmatrix} I_j \\ 0 \end{bmatrix} = \widetilde{E}A\widehat{S}.$$

Consider the linear system $A\widehat{S}\mathbf{x} = AS\mathbf{e}_{j+1}$; we have selected this right side because $\mathbf{r}$ is the $(j+1)$st column of EAS. Applying E, this system is equivalent to

$$A\widehat{S}\mathbf{x} = AS\mathbf{e}_{j+1} \iff EA\widehat{S}\mathbf{x} = EAS\mathbf{e}_{j+1} \iff \begin{bmatrix} I_j \\ 0 \end{bmatrix}\mathbf{x} = \begin{bmatrix} \mathbf{r}_0 \\ 0 \end{bmatrix} \iff \mathbf{x} = \mathbf{r}_0.$$

Applying $\widetilde{E}$ gives $\mathbf{x} = \widetilde{\mathbf{r}}_0$ as well, thus $\mathbf{r} = \widetilde{\mathbf{r}}$.

Last, suppose that cases I and $\widetilde{II}$ hold (the same argument will work if cases II and $\widetilde{I}$ hold). As above, we can take $\mathbf{x} = \mathbf{r}_0$ to solve $A\widehat{S}\mathbf{x} = AS\mathbf{e}_{j+1}$. But then

$$A\widehat{S}\mathbf{r}_0 = AS\mathbf{e}_{j+1} \Longrightarrow \widetilde{E}A\widehat{S}\mathbf{r}_0 = \widetilde{E}AS\mathbf{e}_{j+1} \Longrightarrow \begin{bmatrix} I_j \\ 0 \end{bmatrix} \mathbf{r}_0 = \mathbf{e}_{j+1}.$$

But the entries of

$$\begin{bmatrix} I_j \\ 0 \end{bmatrix} \mathbf{x}$$

are 0 in rows $j+1$ and below, and so no such product can equal $\mathbf{e}_{j+1}$. This shows that cases I and $\widetilde{II}$ cannot simultaneously hold.

Day 23: Friday, October 4.

Today was really just a retreat of Day 22, so read the (copious) notes there.

Day 24: Monday, October 7.

Material from *Linear Algebra* by Meckes & Meckes

The column space is defined on p.116 in.

Vocabulary from today

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

Column space

We have touted the RREF as an important factorization of a matrix. What good does it actually do?

Suppose that $A \in \mathbb{F}^{m \times n}$ has the RREF

$$EA = \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P$$

with $1 \le r < n$. Since $r < n$, Example 22.7 reminds us that both the block of F is genuinely present in the RREF (although the blocks of 0 may not be present). Of course, E and P are invertible.

We claim that A has a nontrivial kernel (null space), and we can "parametrize" it as follows. We have $A\mathbf{x} = \mathbf{0}_n$ (it will be useful to emphasize the size of the zero vector here and there) if and only if

$$E^{-1} \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P\mathbf{x} = \mathbf{0}_n.$$

Multiply both sides by E and abbreviate $\mathbf{y} = P\mathbf{x}$, so $A\mathbf{x} = \mathbf{0}$ if and only if

$$\begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} \mathbf{y} = \mathbf{0}_n. \tag{24.1}$$

It might be easier to understand the structure here if we work with a very small matrix, say $m = n = 3$ and $r = 2$. Then the situation is

$$\begin{bmatrix} 1 & 0 & f_1 \\ 0 & 1 & f_2 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y} = \mathbf{0}_3.$$

This is the same as the linear system

$$\begin{cases} y_1 \quad + f_1 y_3 = 0 \\ \quad y_2 + f_2 y_3 = 0 \\ \qquad\quad 0 = 0, \end{cases}$$

and so

$$y_1 = -f_1 y_3 \quad \text{and} \quad y_2 = -f_2 y_3.$$

Thus

$$\mathbf{y} = \begin{bmatrix} -f_1 y_3 \\ -f_2 y_3 \\ y_3 \end{bmatrix} = y_3 \begin{bmatrix} -f_1 \\ -f_2 \\ 1 \end{bmatrix}.$$

Here is how this structure shows up in (24.1). Denote by $\mathbf{y}^{(r)}$ the first r rows of $\mathbf{y}$ and by $\mathbf{y}^{(n-r)}$ the last $n - r$ rows of $\mathbf{y}$. Then (24.1) is equivalent to

$$\begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}^{(r)} \\ \mathbf{y}^{(n-r)} \end{bmatrix} = \mathbf{0}_n, \tag{24.2}$$

and that in turn is equivalent to

$$I_r \mathbf{y}^{(r)} + F\mathbf{y}^{(n-r)} = \mathbf{0}_r. \tag{24.3}$$

24.1 Problem. By considering carefully the sizes of the 0 blocks in the RREF above (if they are even present), explain why (24.2) and (24.4) really are equivalent. (Doing the block multiplication shows that (24.2) implies (24.4), but why does (24.4) imply (24.2)?)

Then (24.4) is equivalent to

$$\mathbf{y}^{(r)} = -F\mathbf{y}^{(n-r)}, \tag{24.4}$$

and from that we have

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}^{(r)} \\ \mathbf{y}^{(n-r)} \end{bmatrix} = \begin{bmatrix} -F\mathbf{y}^{(n-r)} \\ \mathbf{y}^{(n-r)} \end{bmatrix} = \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{y}^{(n-r)}.$$

Now recall that $\mathbf{y} = P\mathbf{x}$ and rewrite $\mathbf{z} = \mathbf{y}^{(n-r)}$. So here is what we have shown: $A\mathbf{x} = \mathbf{0}_n$ if and only if

$$\mathbf{x} = P^{-1}\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\mathbf{z}$$

for some $\mathbf{z} \in \mathbb{F}^{n-r}$. This formula for $\mathbf{x}$ is the key to controlling $\ker(A)$.

For now, suppose $\mathbf{z} \neq \mathbf{0}_{n-r}$. Then

$$\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\mathbf{z} = \begin{bmatrix} -F\mathbf{z} \\ \mathbf{z} \end{bmatrix} \neq \mathbf{0}_n,$$

since the bottom $n - r$ rows are $\mathbf{z} \neq \mathbf{0}_{n-r}$. Since P is invertible,

$$P^{-1}\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\mathbf{z} \neq \mathbf{0}.$$

Here is what we have proved.

24.2 Theorem. *Let $A \in \mathbb{F}^{m\times n}$ and suppose that*

$$\mathsf{rref}(A) = \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P,$$

where $1 \leq r < n$ and $P \in \mathbb{F}^{n\times n}$ is invertible (the 0 blocks may or may not be present). Then

$$\ker(A) = \left\{ P^{-1}\begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\mathbf{z} \;\middle|\; \mathbf{z} \in \mathbb{F}^{n-r} \right\}.$$

In particular, $\ker(A)$ *is nontrivial.*

This number r is, of course, the rank of A, and it will be hugely important in controlling the behavior of solutions to $A\mathbf{x} = \mathbf{b}$. For now, having $r < n$ destroys uniqueness of solutions if they exist.

24.3 Problem. Let $n > m \geq 1$. Prove that no linear operator $\mathcal{T}: \mathbb{F}^n \to \mathbb{F}^m$ is injective. [Hint: *consider the RREF of the matrix representation of $\mathcal{T}$ and recall $r \leq \min\{m, n\}$. What is $\min\{m, n\}$ here?*]

Theorem 24.2 is particularly helpful in understanding more about invertible matrices.

24.4 Theorem. *A matrix $A \in \mathbb{F}^{n\times n}$ is invertible if and only if* $\mathsf{rref}(A) = I_n$.

Proof. ($\Longleftarrow$) This direction is (slightly) easier, so we do it first. If $\mathsf{rref}(A) = I_n$, then there is $E \in \mathbb{F}^{n\times n}$ invertible such that $EA = I_n$, thus $A = E^{-1}$. Since E^{-1} is also invertible, so is A.

($\Longrightarrow$) If A is invertible, then there is $B \in \mathbb{F}^{n\times n}$ invertible such that $BA = I_n$. The matrix I_n is already in RREF, so since B is invertible, the uniqueness of the RREF forces $\mathsf{rref}(A) = I_n$.

Here is a different proof that does not rely on the uniqueness of the RREF (which was not easy to establish). Suppose $\mathsf{rref}(A) \neq I_n$. Then

$$\mathsf{rref}(A) = \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P$$

with $1 \le r < n$ and P invertible. (Since $\mathsf{rref}(A)$ has n rows as well as n columns, those 0 blocks are genuinely present—otherwise I_r would have to have n rows and then $r = n$. However, the presence or absence of the 0 blocks is not really important here.) Theorem 24.2 implies that $\ker(A) \neq \{\mathbf{0}\}$, and so there is $\mathbf{x} \in \mathbb{F}^n \setminus \{\mathbf{0}\}$ such that $A\mathbf{x} = I_n$. But then A is not invertible. ■

Our last immediate consequence of the RREF is a relaxation of what needs to be checked to ensure that a matrix is invertible. Our current definition of an invertible matrix $A \in \mathbb{F}^{n \times n}$ requires the existence of $B \in \mathbb{F}^{n \times n}$ such that both $AB = I_n$ and $BA = I_n$. We claim that only one such equality needs to hold.

24.5 Theorem. *Let $A \in \mathbb{F}^{n \times n}$ and suppose that there exists $B \in \mathbb{F}^{n \times n}$ such that either $AB = I_n$ or $BA = I_n$. Then A is invertible.*

Proof. Be very careful in that we are not assuming that B is invertible, as otherwise we would be done.

Suppose first that $BA = I_n$. By Theorem 24.4, we will be done if we can show that $\mathsf{rref}(A) = I_n$. Motivated by what might go wrong if A is not invertible, we study $\ker(A)$. Let $A\mathbf{x} = \mathbf{0}$, so $BA\mathbf{x} = \mathbf{0}$ as well. But also $BA\mathbf{x} = \mathbf{x}$, so $\mathbf{x} = \mathbf{0}$. Thus $\ker(A)$ is trivial. If $\mathsf{rref}(A) \neq I_n$, then Theorem 24.2 implies that $\ker(A)$ is nontrivial. So $\mathsf{rref}(A) = I_n$, as desired.

Now suppose that $AB = I_n$. The previous paragraph shows that if $CD = I_n$ for some C, $D \in \mathbb{F}^{n \times n}$, then D is invertible; thus (with $C = A$ and $D = B$) B is invertible. Then $ABB^{-1} = I_nB^{-1}$, so $A = B^{-1}$, which is invertible. ■

We now turn to counting problems. Theorem 24.2 shows us how $\mathbb{F}^{n-r}$ effectively "parametrizes" $\ker(A)$ for $A \in \mathbb{F}^{m \times n}$ when the block I_r shows up in $\mathsf{rref}(A)$. What does this number $n - r$ say about the "size" of the kernel? What, more generally, can the RREF possibly say about solving $A\mathbf{x} = \mathbf{b}$?

Here is what it does not say.

24.6 Problem. Let $\mathcal{V}$ be a nontrivial vector space, i.e., $\mathcal{V} \neq \{0\}$ (where 0 is the zero vector for $\mathcal{V}$). Show that $\mathcal{V}$ contains infinitely many elements.

This result destroys the possibility that counting the number of elements in a vector space will ever yield meaningful information, outside of the trivial space $\{0\}$. It turns out that "counting" the "size" of the range of a linear operator will be more effective than starting with the kernel.

24.7 Definition. *Let $A \in \mathbb{F}^{m \times n}$. The* COLUMN SPACE *of A is*

$$\mathsf{col}(A) := \{A\mathbf{v} \mid \mathbf{v} \in \mathbb{F}^n\}.$$

24.8 Problem. Check that the column space of $A \in \mathbb{F}^{m \times n}$ is the range of $\mathcal{T}_A \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ given by $\mathcal{T}_A\mathbf{v} = A\mathbf{v}$, and so $\mathsf{col}(A)$ is a subspace of $\mathbb{F}^n$.

For a tractable example, let

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

We have $\mathbf{w} \in \mathsf{col}(A)$ if and only if $\mathbf{w} = A\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^4$, thus

$$\mathbf{w} = v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3\mathbf{a}_3 + v_4\mathbf{a}_4.$$

Now, observe that

$$\mathbf{a}_2 = 2\mathbf{a}_1 \quad \text{and} \quad \mathbf{a}_4 = 3\mathbf{a}_1 + 4\mathbf{a}_3.$$

Thus any $\mathbf{w} \in \mathsf{col}(A)$ has the form

$$\mathbf{w} = v_1\mathbf{a}_1 + v_2(2\mathbf{a}_1) + v_3\mathbf{a}_3 + v_4(3\mathbf{a}_1 + 4\mathbf{a}_3) = (v_1 + 2v_2 + 3v_4)\mathbf{a}_1 + (v_3 + 4v_4)\mathbf{a}_3.$$

That is, the vectors $\mathbf{a}_2$ and $\mathbf{a}_4$ are redundant in describing $\mathsf{col}(A)$, and that is because they are linear combinations of the other columns of A.

This toy example illustrates two general principles that we will develop much further: many interesting subspaces are given by spans (we have already seen this with some concrete eigenspaces), and sometimes some of the vectors in a span are redundant. How do we quantify and qualify redundancy to be as efficient as possible in working with spans?

Day 25: Wednesday, October 9.

Material from *Linear Algebra* by Meckes & Meckes

Pages 140–141 discuss redundancy and linear combinations, and pp.141–142 give two equivalent definitions of linear (in)dependence. Proposition 3.1 relates linear (in)dependence of column vectors to matrix kernels.

Do Quick Exercises #1, #2, and #3 in Section 3.1

Vocabulary from today

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

List/finite sequence of length n in a set X, linearly dependent list in a vector space $\mathcal{V}$, linearly independent list in a vector space $\mathcal{V}$

We return to the matrix

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

We can consider the redundancies among its columns at three levels.

1. *One column is a linear combination of the others.* Here it is the case that

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

This "singles out one column for blame."

2. *A nontrivial linear combination of the columns is the zero vector.* We rewrite the equalities above as

$$2\mathbf{a}_1 + (-1)\mathbf{a}_2 + 0\mathbf{a}_3 + 0\mathbf{a}_4 = \mathbf{0}_3 \quad \text{and} \quad 3\mathbf{a}_1 + 0\mathbf{a}_2 + 4\mathbf{a}_3 + (-1)\mathbf{a}_4 = \mathbf{0}_3.$$

These linear combinations are "nontrivial" because some of the scalar coefficients in them are nonzero. Here no vector is "guiltier" than another.

3. *The kernel is nontrivial.* The equalities above show

$$A \begin{bmatrix} 2 \\ -1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}_4 \quad \text{and} \quad A \begin{bmatrix} 3 \\ 0 \\ 4 \\ -1 \end{bmatrix} = \mathbf{0}_4.$$

So, there are nonzero vectors in $\ker(A)$.

The first and second properties are equivalent in any vector space, not just $\mathbb{F}^3$. First, we need a slight variation on sigma notation for finite sums. Let $\mathcal{V}$ be a vector space and $v_1, \ldots, v_n \in \mathcal{V}$ and $1 \le j \le n$. Then

$$\sum_{\substack{k=1 \\ k \ne j}}^n v_k := \begin{cases} \sum_{k=2}^n v_k, & j = 1 \\ \sum_{k=1}^{j-1} v_k + \sum_{k=j+1}^n v_k, & 2 \le j \le n-1 \\ \sum_{k=1}^{n-1} v_k, & j = n. \end{cases}$$

25.1 Lemma. *Let $\mathcal{V}$ be a vector space and $v_1, \ldots, v_n \in \mathcal{V}$. The following are equivalent:*

(i) *One of the v_k is a linear combination of the others: there exists j such that*

$$v_j = \sum_{\substack{k=1 \\ k \ne j}}^n \alpha_k v_k$$

for some $\alpha_k \in \mathbb{F}$.

(ii) *A nontrivial linear combination of the* v_k *is* 0*: there exist* $\beta_1, \ldots, \beta_n \in \mathbb{F}$ *such that*

$$\sum_{k=1}^{n} \beta_k v_k = 0,$$

and at least one of the β_k *is nonzero.*

Proof. (i) $\implies$ (ii) Just rewrite

$$-v_j + \sum_{\substack{k=1 \\ k \neq j}}^{n} \alpha_k v_k = 0$$

and define

$$\beta_k := \begin{cases} \alpha_k, & k \neq j \\ -1, & k = j. \end{cases}$$

Then $\sum_{k=1}^{n} \beta_k v_k = 0$ and at the very least $\beta_j = -1 \neq 0$.

(ii) $\implies$ (i) Say $\beta_j \neq 0$. Then

$$0 = \sum_{k=1}^{n} \beta_k v_k = \beta_j v_j + \sum_{\substack{k=1 \\ k \neq j}} \beta_k v_k,$$

and so

$$v_j = \sum_{\substack{k=1 \\ k \neq j}} \left(-\frac{\beta_k}{\beta_j} \right) v_k.$$

Here it is important that $\beta_j \neq 0$ so we can divide. ■

We probably want to say that either of the conditions in the previous lemma means linear dependence. But *what object* should we call linearly dependent?

The columns of the matrix A above certainly should be linearly dependent, from this lemma. We also probably want to say that the columns of

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

are linearly dependent, since the nontrivial linear combination

$$(1) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + (0) \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + (-1) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}_3$$

is the zero vector. But we should *not* say that the *set* of columns of this matrix is linearly dependent, for that set is

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

The problem here is that set conventions ignore repetition of elements, while linear dependence relations are very much affected by repetition.

The answer is in our statement of the previous lemma: when we take vectors $v_1, \ldots, v_n \in \mathcal{V}$, we are allowing repetition (maybe $v_1 = v_n$, which certainly happens if $n = 1$), but in addition, we are also encoding a sense of order. We formalize this as follows.

25.2 Definition. *Let X be a set and $n \geq 1$ be an integer. A* **LIST OF LENGTH** n **IN** X *or a* **FINITE SEQUENCE IN** X **OF LENGTH** n *is a function from $\{1, \ldots, n\}$ to X. If $f \colon \{1, \ldots, n\} \to X$ is such a function with $x_k := f(k)$, then we often write $f = (x_1, \ldots, x_n)$. If $n = 1$, then we interpret $(x_1, \ldots, x_1) = (x_1)$.*

A list of length n in X is sometimes called an **ORDERED** n**-TUPLE WITH ENTRIES IN** X. Note that even though we say "in X" in the definition above, a list *in* X is not an element *of* X.

25.3 Example. The lists $(1, 2, 1)$ and $(1, 1, 2)$ in $\mathbb{R}$ are not the same. The first list is the function

$$(1, 2, 1) = \{(1, 1), (2, 2), (3, 1)\},$$

while the second list is the function

$$(1, 1, 2) = \{(1, 1), (2, 1), (3, 2)\}.$$

As functions, their domains and ranges are the same: both domains are $\{1, 2, 3\}$ and both ranges are $\{1, 2\}$. But pointwise these functions are different.

25.4 Problem. Explain why $\mathbb{F}^n$ is the set of all lists of length n in $\mathbb{F}$, and so it is correct to say

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = (v_1, \ldots, v_n).$$

If a matrix $A \in \mathbb{F}^{m \times n}$ is a function from $\{(i, j) \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}$, and if the columns of A are $\mathbf{a}_1, \ldots, \mathbf{a}_n$, is it really correct to say $A = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$?

Embiggened with lists, we now rigorously define linear dependence and independence.

25.5 Definition. *A list $(v_1, \ldots, v_n)$ in a vector space $\mathcal{V}$ is* **LINEARLY DEPENDENT** *if either condition in Lemma 25.1 holds and* **LINEARLY INDEPENDENT** *if it is not linearly dependent.*

So, linear dependence is an *existential* condition: we must show that either there exists a nontrivial linear combination of the vectors in the list that adds up to the zero vector, or that there exists a vector in the list that is a linear combination of the others. But linear independence is a *universal* condition: we must show that no linear combination of the vectors in the list is zero except the trivial combination ($\sum_{k=1}^{n}\alpha_k v_k \implies \forall k : \alpha_k = 0$) or that no vector in the list is a linear combination of the others in the list.

25.6 Example. **(i)** Consider a list of length 1 in the vector space $\mathcal{V}$: this has the form (v) for some $v \in \mathcal{V}$. This list is linearly dependent if and only if there is $\alpha \in \mathbb{F} \setminus \{0\}$ such that $\alpha v = 0$. If $v \neq 0$, $\alpha v = 0$ forces $\alpha = 0$. So, every list of length 1 is linearly independent, except for (0).

(ii) Consider a list of length 2 in the vector space $\mathcal{V}$: this has the form (v_1, v_2) for some v_1, $v_2 \in \mathcal{V}$. This list is linearly dependent if and only if there are α_1, $\alpha_2 \in \mathbb{F}$, not both 0, such that $\alpha_1 v_1 + \alpha_2 v_2 = 0$. Say $\alpha_1 \neq 0$. Then $v_1 = -(\alpha_2/\alpha_1)v_2$. So, a list of length 2 is linearly dependent if and only if one of the vectors in the list is a scalar multiple of the other.

(iii) Let $(v_1, \ldots, v_n)$ be a list in $\mathcal{V}$ and suppose that one vector in the list is 0, say, $v_j = 0$. Put

$$\alpha_k := \begin{cases} 1, & k = j \\ 0, & k \neq j. \end{cases}$$

Then

$$\sum_{k=1}^{n} \alpha_k v_k = (1)v_j + \sum_{\substack{k=1 \\ k\neq j}} 0v_k = 0 + 0,$$

so this list is linearly dependent. So, any list containing the zero vector is linearly dependent (we saw this above with the case $n = 1$).

(iv) Let $(v_1, \ldots, v_n)$ be a list in $\mathcal{V}$ of length $n \geq 2$. Suppose that $v_{j_1} = v_{j_2}$ where $1 \leq j_1 < j_2 \leq n$. Put

$$\alpha_k := \begin{cases} 1, & k = j_1 \\ -1, & k = j_2 \\ 0, & k \neq j_1,\ j_2. \end{cases}$$

Then

$$\sum_{k=1}^{n} \alpha_k v_k = (1)v_{j_1} + (-1)v_{j_2} + \sum_{\substack{k=1 \\ k\neq j_1,\ j_2}}^{n} 0v_k = v_{j_1} - v_{j_1} + 0 = 0.$$

so this list is linearly dependent. So, any list of two or more vectors with a repeated vector is linearly dependent.

(v) A list $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ in $\mathbb{F}^m$ is linearly dependent if and only if there are $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ not all 0 such that $\sum_{k=1}^{n}\alpha_k \mathbf{v}_k = \mathbf{0}_m$; this is equivalent to

$$\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \alpha_n \end{bmatrix} = \mathbf{0}_m,$$

which in turn is equivalent to a nontrivial kernel for $\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$. So, a list in $\mathbb{F}^m$ is linearly dependent if and only if the matrix whose columns are that list has a nontrivial kernel. (We will eventually develop the tools to isolate the "guilty" columns of that matrix and see exactly how they are linear combinations of the other columns.)

(vi) Recall the notion of algebraic dual space from Example 12.1 and define linear functionals on $\mathcal{C}(\mathbb{R})$ as follows. For $x \in \mathbb{R}$, let φ_x be the "evaluate at x" functional such that $\varphi_x(f) := f(x)$ for $f \in \mathcal{C}(\mathbb{R})$. For $x_1, \ldots, x_n \in \mathbb{R}$, let $(\varphi_{x_1}, \ldots, \varphi_{x_n})$ be a list of such functionals in $(\mathcal{C}(\mathbb{R}))'$. This list is automatically linearly dependent if any terms are the same, which here is equivalent to $x_j = x_k$ for some $j \neq k$. So, assume that all of the x_k are distinct. We claim this list is linearly independent.

Here is why. Suppose that $\sum_{k=1}^n \alpha_k \varphi_{x_k} = 0$ for some $\alpha_k \in \mathbb{F}$. This means that $\sum_{k=1}^n \alpha_k \varphi_{x_k}(f) = 0$ for all $f \in \mathcal{C}(\mathbb{R})$, and so $\sum_{k=1}^n \alpha_k f(x_k) = 0$ for all $f \in \mathcal{C}(\mathbb{R})$. Since this is true for *all* f, we can pick f to be any function that we like. In particular, we could "interpolate" and choose f to be 0 at each x_k except for one value of k. Say that $f_j \in \mathcal{C}(\mathbb{R})$ with

$$f_j(x_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

Then $0 = \sum_{k=1}^n \alpha_k f_j(x_k) = \alpha_j$, and so each α_k is 0.

Does such a function f_j really exist? Here is how to construct it for small n, say, $n = 3$. Just put

$$f_1(x) = \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)}, \qquad f_2(x) = \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)},$$

$$\text{and} \qquad f_3(x) = \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)}.$$

There are no problems with division by zero because all of the x_k are distinct.

25.7 Problem. Check that the "evaluate at x" functional really is a linear functional on $\mathcal{C}(\mathbb{R})$.

25.8 Problem. Use the definition of linear independence that the list $(\mathbf{e}_1, \ldots, \mathbf{e}_m)$ of standard basis vectors in $\mathbb{F}^m$ is linearly independent.

25.9 Problem. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{F}^{n \times n}$ be invertible. Prove that $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ is linearly independent. (This is understood to be the list of columns of A with no repetitions among the entries of the list.)

Day 26: Friday, October 11.

Material from *Linear Algebra* by Meckes & Meckes

The "spanning lemma" is Theorem 3.6 (and Corollary 3.7), which the Meckeses call the "linear dependence lemma" and use frequently. Our lemma adds a little detail on linear independence. The eigenvector result is Theorem 3.8. Results on linear (in)dependence in $\mathbb{F}^m$ appear in Corollaries 3.2, 3.3, and 3.5 and Algorithm 3.4.

Do Quick Exercises #4 and #5 in Section 3.1.

It follows from the definition of linear independence that linear independence prevents redundancy: if a list $(v_1, \ldots, v_n)$ in $\mathcal{V}$ is linearly independent, then no vector in that list is a linear combination of the others. This also removes ambiguity: there is only one way to write a vector $v \in \operatorname{span}(v_1, \ldots, v_n)$ as a linear combination of the v_k.

26.1 Lemma (Unique representation). *Let $\mathcal{V}$ be a vector space and let $(v_1, \ldots, v_n)$ be a linearly independent list in $\mathcal{V}$. Suppose that $\sum_{k=1}^n \alpha_k v_k = \sum_{k=1}^n \beta_k v_k$ for some α_k, $\beta_k \in \mathbb{F}$. Then $\alpha_k = \beta_k$ for all k.*

Proof. Subtract to find

$$\sum_{k=1}^n (\alpha_k - \beta_k) v_k = 0,$$

so $\alpha_k - \beta_k = 0$ for all k. ■

So, if $v \in \operatorname{span}(v_1, \ldots, v_n)$, then by definition of span we can write v as a linear combination of the v_k, and by the previous result, there is only one way to do this.

Here is another useful result about spans and linear (in)dependence. It says that any linearly dependent list whose first term is not the zero vector can be pared down to a linearly independent list that respects certain "order" properties of the list. (Shortly we will see a similar result that chops up the order a bit.)

26.2 Lemma (Spanning). **(i)** *Let $(v_1, \ldots, v_n)$ be a linearly dependent list in the vector space $\mathcal{V}$ with $n \geq 2$. Suppose $v_1 \neq 0$. Then there exists $j \geq 2$ such that $v_j \in (v_1, \ldots, v_{j-1})$. Moreover, j can be chosen so that $(v_1, \ldots, v_{j-1})$ is linearly independent.*

(ii) *Let $(v_1, \ldots, v_n)$ be a list in $\mathcal{V}$ with $n \geq 2$ such that $v_j \notin \operatorname{span}(v_1, \ldots, v_{j-1})$ for $j = 1, \ldots, n$. Then $(v_1, \ldots, v_n)$ is linearly independent.*

Proof. **(i)** Since $(v_1, \ldots, v_n)$ is linearly dependent, there exist $\alpha_k \in \mathbb{F}$ not all 0 such that $\sum_{k=1}^n \alpha_k v_k = 0$. Let $m \geq 1$ be the smallest index such that $\alpha_m \neq 0$. At worst $m = n$, but at least $m \geq 2$, for if $\alpha_k = 0$ for $k \geq 2$ but $\alpha_1 \neq 0$, then $\alpha_1 v_1 = 0$, and then $v_1 = 0$. So, we have

$$0 = \sum_{k=1}^n \alpha_k v_k = \alpha_m v_m + \sum_{k=1}^{m-1} \alpha_k v_k,$$

and therefore, since $\alpha_m \neq 0$,

$$v_m = \sum_{k=1}^{m-1} \left(-\frac{\alpha_k}{\alpha_m}\right) v_k \in \operatorname{span}(v_1, \ldots, v_{m-1}).$$

Now let j be the smallest of these m, i.e., the smallest index of the integers $m \in \{2, \ldots, n\}$ such that $v_m \in (v_1, \ldots, v_{m-1})$. The work above shows that at least one such m exists. We claim that $(v_1, \ldots, v_{j-1})$ is linearly independent. Certainly this is true if $j = 2$, since $v_1 \neq 0$, so (v_1) is linearly independent. Otherwise, if $j \geq 3$, suppose $\sum_{k=1}^{j-1} \alpha_k v_k = 0$, and let ℓ be the smallest index such that $\alpha_\ell \neq 0$. As before, we must have $\ell \geq 2$. Then

$$0 = \sum_{j=1}^{j-1} \alpha_k v_k = \sum_{k=1}^{\ell} \alpha_k v_k,$$

so, just as before,

$$v_\ell = \sum_{k=1}^{\ell-1} \left(-\frac{\alpha_k}{\alpha_p}\right) v_k \in \operatorname{span}(v_1, \ldots, v_{\ell-1}).$$

But $p \leq j - 1 < j$, which contradicts the minimality of j.

(ii) This is the contrapositive of the first result above:

$$\big(\text{Linear dependence} \Longrightarrow \exists j : v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})\big)$$
$$\iff \big(\forall j : v_j \notin \operatorname{span}(v_1, \ldots, v_{j-1}) \Longrightarrow \text{Linear independence}\big). \quad \blacksquare$$

26.3 Problem. Use the spanning lemma to give another proof that the list of standard basis vectors $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ in $\mathbb{F}^n$ is linearly independent.

Experience might suggest that eigenvectors corresponding to distinct eigenvalues are linearly independent (this is hopefully obvious for a diagonal matrix, where the standard basis vectors show up as eigenvectors). Here is how we see this with only a handful of eigenvectors. Let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ have the distinct eigenvalues λ_1, $\lambda_2 \in \mathbb{F}$, so $\lambda_1 \neq \lambda_2$, and there are v_1, $v_2 \in \mathcal{V} \setminus \{0\}$ such that

$$\mathcal{T} v_1 = \lambda_1 v_1 \quad \text{and} \quad \mathcal{T} v_2 = \lambda_2 v_2.$$

Suppose $\alpha_1 v_1 + \alpha_2 v_2 = 0$. We want to show $\alpha_1 = \alpha_2 = 0$. We can get the eigenvalues to show up in that linear combination by applying $\mathcal{T}$ to both sides:

$$0 = \mathcal{T}(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 \mathcal{T} v_1 + \alpha_2 \mathcal{T} v_2 = \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2.$$

It now looks like we have a system of linear equations for α_1 and α_2, except the "coefficients" are not scalars in $\mathbb{F}$ but vectors:

$$\begin{cases} \alpha_1 v_1 + \alpha_2 v_2 = 0 \\ \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 = 0. \end{cases}$$

We can make λ_1 and λ_2 interact via elementary row operations: multiply both sides of $\alpha_1 v_1 + \alpha_2 v_2 = 0$ by $-\lambda_2$ and add to the result above to get

$$0 = \big(\alpha_1\lambda_1 v_1 + \alpha_2\lambda_2 v_2\big) + \big(-\lambda_2\alpha_1 v_1 - \lambda_2\alpha_2 v_2\big) = \alpha_1(\lambda_1 - \lambda_2)v_1.$$

Since $v_1 \neq 0$, we have $\alpha_1(\lambda_1 - \lambda_2) = 0$, and since $\lambda_1 \neq \lambda_2$, we have $\alpha_1 = 0$. Back to the original linear combination, we reduce to $\alpha_2 v_2 = 0$, so $\alpha_2 = 0$ since $v_2 \neq 0$.

Here is how this works in general: let $\mathcal{T} v_k = \lambda_k v_k$, with each $v_k \neq 0$ and the λ_k distinct. Consider the list $(v_1, \ldots, v_n)$ of eigenvectors: $v_1 \neq 0$, so if the list is linearly dependent, the spanning lemma gives j such that $v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$ and $(v_1, \ldots, v_{j-1})$ is linearly independent.

26.4 Problem. Explain why, in the $n = 2$ case, we really had $j = 2$.

Write $v_j = \sum_{k=1}^{j-1} \alpha_k v_k$. Apply $\mathcal{T}$ to find $\mathcal{T} v_j = \sum_{k=1}^{j-1} \alpha_k \mathcal{T} v_k$. Use the definition of eigenvalue(vector) to find $\lambda_j v_j = \sum_{k=1}^{j-1} \alpha_k \lambda_k v_k$. We now have the "linear system"

$$\begin{cases} v_j = \sum_{k=1}^{j-1} \alpha_k v_k \\ \lambda_j v_j = \sum_{k=1}^{j-1} \alpha_k \lambda_k v_k. \end{cases}$$

Multiply both sides of $v_j = \sum_{k=1}^{j-1} \alpha_k v_k$ by $-\lambda_j$ and add:

$$\lambda_j v_j - \lambda_j v_j = \sum_{k=1}^{j-1} \alpha_k \lambda_k v_k - \sum_{k=1}^{j-1} \alpha_k \lambda_j v_k,$$

thus

$$\sum_{k=1}^{j-1} \alpha_k (\lambda_k - \lambda_j) v_k = 0.$$

By linear independence of $(v_1, \ldots, v_{j-1})$, $\alpha_k(\lambda_k - \lambda_j) = 0$, and since the eigenvalues are distinct and $\lambda_k - \lambda_j \neq 0$, we must have $\alpha_k = 0$.

At this point we may not be sure where we are. We are trying to derive a contradiction from the linear dependence of $(v_1, \ldots, v_n)$. Yet nowhere have we considered a linear combination of the form $\sum_{k=1}^{n} \beta_k v_k$ that involves all of the vectors in this list—surely that should be part of the argument?

Not necessarily. Our work above shows $\alpha_k = 0$ for $k = 1, \ldots, j-1$, where $v_j = \sum_{k=1}^{j-1} \alpha_k v_k$. But then $v_j = \sum_{k=1}^{j-1} 0 v_k = 0$, and the zero vector is not an eigenvector. That is the contradiction.

26.5 Theorem. *Eigenvectors of a linear operator corresponding to distinct eigenvalues are linearly independent. More precisely, if (λ_k, v_k) is an eigenpair of $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ for $k = 1, \ldots, n$ and $\lambda_j \neq \lambda_k$ for $j \neq k$, then $(v_1, \ldots, v_n)$ is linearly independent.*

A good exercise for the ardent apprentice linear algebraist is to redo the argument above for arbitrary n until it feels completely natural. Starting with $n = 3$ concretely may make things more transparent.

Day 27: Monday, October 14.

Material from *Linear Algebra* by Meckes & Meckes

Algorithm 3.4 describes how to show that a list in $\mathbb{F}^m$ is linearly independent.

Do Quick Exercise #4 in Section 3.1.

Here is the other side of the spanning lemma: we can "reduce" or "winnow down" any linearly dependent list into a linearly independent one with the same span. This relies on the following fact.

27.1 Problem. Let $(v_1, \ldots, v_n)$ be a list in the vector space $\mathcal{V}$ with $n \geq 2$ and suppose that some entry v_j of the list is a linear combination of the other vectors in the list. Show that $\operatorname{span}(v_1, \ldots, v_n) = \operatorname{span}(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)$. (If $j = 1$, we interpret $(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n) = (v_2, \ldots, v_n)$, and if $j = n$, this is $(v_1, \ldots, v_{n-1})$.

Consider the following list in $\mathbb{R}^4$:

$$(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6) = \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix} \right).$$

The zero vector contributes nothing to the span, so we can ignore that (and we should remove it from the list to protect linear independence anyway). Since $\mathbf{v}_2 \neq \mathbf{0}_4$, we may as well try to include $\mathbf{v}_2$ in the span. Since $\mathbf{v}_3 = 2\mathbf{v}_2$, $\mathbf{v}_3$ contributes nothing new to a span of vectors already containing $\mathbf{v}_2$. So, right now, $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \operatorname{span}(\mathbf{v}_2)$. Next, $\mathbf{v}_4$ is definitely not a scalar multiple of $\mathbf{v}_2$, so we include it. If we think a bit we can find α_2 and α_4 such that $\alpha_2\mathbf{v}_2 + \alpha_4\mathbf{v}_4 = \mathbf{v}_5$, so we should exclude $\mathbf{v}_5$ from the span. Right now, $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_5) = \operatorname{span}(\mathbf{v}_2, \mathbf{v}_4)$.

27.2 Problem. Think a bit and find them.

Last, $\mathbf{v}_6$ is not in the span of $\mathbf{v}_2$ and $\mathbf{v}_4$, since no linear combination of these three vectors adds to $\mathbf{0}_4$ except the trivial one.

27.3 Problem. Check that: if $\alpha_2\mathbf{v}_2 + \alpha_4\mathbf{v}_4 + \alpha_6\mathbf{v}_6 = \mathbf{0}_4$, then $\alpha_2 = \alpha_4 = \alpha_6 = 0$. [Hint: *look at rows* 2 *and* 4 *first.*]

We conclude $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_6) = \operatorname{span}(\mathbf{v}_2, \mathbf{v}_4, \mathbf{v}_6)$. We probably want to call $(\mathbf{v}_2, \mathbf{v}_4, \mathbf{v}_6)$ a "sublist" of $(\mathbf{v}_1, \ldots, \mathbf{v}_6)$, since the entries of the former appear in the latter and in the same order. This is somewhat annoying to define precisely.

27.4 Definition. *Let X be a set and let $(x_1, \dots, x_n)$ be a list of length n in X. Let $r \le n$. A list $(y_1, \dots, y_r)$ is a* SUBLIST *of $(x_1, \dots, x_n)$ if there is a map $\sigma\colon \{1, \dots, r\} \to \{1, \dots, n\}$ that is strictly increasing (in the sense that $\sigma(j) < \sigma(k)$ for $j < k$) with $y_k = x_{\sigma(k)}$ for each k.*

In the concrete example above, the sublist of $(\mathbf{v}_1, \dots, \mathbf{v}_6)$ of interest was $(\mathbf{v}_2, \mathbf{v}_4, \mathbf{v}_6)$. We could call this sublist $(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$ and put $\sigma(1) = 2$, $\sigma(2) = 4$, and $\sigma(3) = 6$.

We generalize this "reduction" procedure as follows.

27.5 Lemma (Reduction). *Let $(v_1, \dots, v_n)$ be a linearly dependent list in the vector space $\mathcal{V}$ with at least one of the v_k nonzero. There exists a linearly independent sublist $(v_{k_1}, \dots, v_{k_r})$ of $(v_1, \dots, v_n)$ such that $\operatorname{span}(v_{k_1}, \dots, v_{k_r}) = \operatorname{span}(v_1, \dots, v_n)$.*

Proof. First, we require at least one entry in the list to be nonzero, as otherwise the span is just $\{0\}$, and there is nothing interesting here. Next, we require the list to be linearly dependent, as otherwise the list is linearly independent, and that is the best kind of list.

We reduce the list as follows. Let v_{k_1} be the first nonzero vector in the list. (At least one exists.) So $\operatorname{span}(v_1, \dots, v_{k_1}) = \operatorname{span}(v_{k_1})$. Also, (v_{k_1}) is linearly independent because $v_{k_1} \neq 0$.

Let v_{k_2} be the first vector in the list that is not a scalar multiple of v_{k_1}. So $\operatorname{span}(v_1, \dots, v_{k_2}) = \operatorname{span}(v_{k_1}, v_{k_2})$. Also, (v_{k_1}, v_{k_2}) is linearly independent because neither entry is a scalar multiple of the other (or by the spanning lemma, since $v_{k_2} \notin \operatorname{span}(v_{k_1})$).

Let v_{k_3} be the first vector in the list that is not in $\operatorname{span}(v_{k_1}, v_{k_2})$. So $\operatorname{span}(v_1, \dots, v_{k_3}) = \operatorname{span}(v_{k_1}, v_{k_2}, v_{k_3})$. Also, $(v_{k_1}, v_{k_2}, v_{k_3})$ is linearly independent by the spanning lemma.

Now turn the crank and keep going: eventually we run out of vectors in the list. ■

The reduction lemma is an existential result, and the algorithm within requires an annoying entry-by-entry examination of the list. In the very important case of lists of column vectors, there are much easier, and more meaningful, methods of determining which vectors in a list are linearly independent and preserve the span. These hinge on the RREF.

Here is an illustrative example of a much more general phenomenon. We have previously shown the existence of an invertible matrix $E \in \mathbb{R}^{3\times 3}$ such that

$$EA = \mathsf{rref}(A) =: R, \qquad A := \begin{bmatrix} 1 & 2 & 1 & 7 \\ 2 & 4 & 2 & 14 \\ 0 & 0 & 2 & 8 \end{bmatrix}, \qquad R := \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now we introduce a new piece of terminology.

27.6 Definition. **(i)** *Let $R \in \mathbb{F}^{m\times n}$ be in RREF. Column j of R is a* PIVOT COLUMN *of R if column j contains a leading 1 (i.e., if column j is the first appearance in R of a standard basis vector for $\mathbb{F}^m$).*

(ii) *Let $A \in \mathbb{F}^{m\times n}$. Column j of A is a* PIVOT COLUMN *of A if the jth column of $\mathsf{rref}(A)$ is a pivot column.*

So, with A and R as above, columns 1 and 3 are the pivot columns. With $\mathbf{r}_j$ as the jth column of R, we have also previously seen that

$$\mathbf{r}_2 = 2\mathbf{r}_1 \quad \text{and} \quad \mathbf{r}_4 = 3\mathbf{r}_1 + 4\mathbf{r}_3.$$

That is, the nonpivot columns of R_0 are linear combinations of the pivot columns. The same is true of the nonpivot columns of A, although this may be less obvious: $\mathbf{a}_2 = 2\mathbf{a}_1$ (well, that should be obvious) and also $\mathbf{a}_4 = 3\mathbf{a}_4 + 4\mathbf{a}_3$.

It should be obvious that $(\mathbf{r}_1, \mathbf{r}_3)$ is linearly independent, since the entries of this list are (nonrepeated) standard basis vectors. It may be less obvious that $(\mathbf{a}_1, \mathbf{a}_3)$ is linearly independent.

27.7 Problem. Check that.

The linear (in)dependence relations among the columns of A appear to coincide with those among the columns of R. This is no accident: $A\mathbf{x} = \mathbf{0}_3$ if and only if $EA\mathbf{x} = \mathbf{0}_3$, since E is invertible, and so this holds if and only if $R_0\mathbf{x} = \mathbf{0}_3$.

Here is the more general (but also more precise) situation.

27.8 Theorem. *Let $A \in \mathbb{F}^{m \times n} \setminus \{0\}$.*

(i) *The pivot columns of A are linearly independent. More precisely, if $\mathbf{a}_{j_k}$ is a pivot column of A for $k = 1, \ldots, r$ with $1 \le j_k < j_{k+1} \le r \le \min\{m, n\}$, then $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$ is linearly independent.*

(ii) *Any nonpivot column of A is a linear combination of the pivot columns of A.*

(iii) $\mathsf{col}(A)$ *is the span of the pivot columns.*

Proof. **(i)** Say that columns $j_1, \ldots, j_r$ are the pivot columns of A, where $r \le \min\{m, n\}$, and $1 \le j_1 < \cdots < j_r \le n$. (Since $A \ne 0$, A has at least one pivot column, and A cannot have more than $\min\{m, n\}$ pivot columns because (1) A has at most n columns and (2) at most m columns in $\mathsf{rref}(A)$ can contain a leading 1.) Suppose $\sum_{k=1}^r \alpha_k \mathbf{a}_{j_k} = \mathbf{0}_m$ for some $\alpha_k \in \mathbb{F}$.

Let $E \in \mathbb{F}^{m \times m}$ be invertible with $EA = \mathsf{rref}(A)$. Then $E\big(\sum_{k=1}^r \alpha_k \mathbf{a}_{j_k}\big) = \mathbf{0}_m$. Since $E\mathbf{a}_{j_k}$ is a pivot column of $\mathsf{rref}(A)$, $E\mathbf{a}_{j_k}$ must be one of the standard basis vectors for $\mathbb{F}^m$. Specifically, $E\mathbf{a}_{j_k} = \mathbf{e}_k$, since the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r$ must appear for the first time "in order" in $\mathsf{rref}(A)$. Thus $\sum_{k=1}^r \alpha_k \mathbf{e}_k = \mathbf{0}_m$, so $\alpha_k = 0$ for all k.

(ii) Let $\mathbf{a}_j$ be a nonpivot column of A. So $E\mathbf{a}_j$ is a nonpivot column of $\mathsf{rref}(A)$. Our idea is that the nonpivot columns of $\mathsf{rref}(A)$ are linear combinations of the pivot columns, which are $\mathbf{e}_1, \ldots, \mathbf{e}_r$. Thus $E\mathbf{a}_j = \sum_{k=1}^r \alpha_k \mathbf{e}_k$ for some α_k, equivalently, $\mathbf{a}_j = \sum_{k=1}^r \alpha_k E^{-1}\mathbf{e}_k$. And each $E^{-1}\mathbf{e}_k$ is a pivot column of A.

Here we give slightly more detail on that "idea." If A has a nonpivot column, it must be the case that

$$\mathsf{rref}(A) = \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P$$

with $r < n$ and $F \in \mathbb{F}^{r \times (n-r)}$. (As usual, the zero blocks may or may not be present, and P is a permutation matrix.) Any column of F is a linear combination of the columns of I_r, so any nonpivot column of $\mathsf{rref}(A)$ is a linear combination of $\mathbf{e}_1, \ldots, \mathbf{e}_r$.

(iii) Apply Problem 27.1 repeatedly to all of the nonpivot columns. ■

This corollary provides an explicit recipe for finding the linearly independent columns of a matrix and controlling its column space: just look at the RREF. It also provides a quick opinion on linear dependence of a list.

27.9 Corollary. *Any list of vectors in $\mathbb{F}^m$ of length $n > m$ is linearly dependent.*

27.10 Problem. Prove this corollary. [Hint: *how many pivot columns, at the most, can the matrix whose columns are the vectors in that list have?*]

That is,

$$\text{More columns than rows} \Longrightarrow \text{Linearly dependent.}$$