MATH 4260: LINEAR ALGEBRA II

Daily Log for Lectures and Readings

Timothy E. Faver

March 6, 2026

Contents

How to Use This Daily Log

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

The document contains several classes of problems, which interact intimately with the material and which supplement (but certainly do not replace) the problems in the textbook.

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

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

(+) Problems marked (+) are meant to be more challenging than the (!)- and ($\star$)-problems and will take you deeper into calculations and proofs and make connections to concepts across and beyond the course. It will not be necessary to do any (+)-problems to master the essential material of the course, but your experience may be richer (and more meaningful, and more fun) by considering them. If you have done all of the (!)- and ($\star$)-problems, and the required and recommended problems from the textbook, and if you're still feeling bored or wondering if something is "missing," check out the (+)-problems. Sometimes a (!)- or ($\star$)-problem will reference a (+)-problem; you should read the statement of that (+)-problem, but feel no obligation to do it.

Day 1: Monday, January 12.

Linear algebra is *everywhere*. It arises naturally in every branch of mathematics—pure, applied, computational—and in problems in all STEM fields, especially in today's most popular (and sultry) field of *data science*. This course is a second course in linear algebra from a more abstract, general, and proofs-based perspective. We will both assume familiarity with many "classical" topics and techniques from a standard first course in linear algebra (such as matrix-vector multiplication), but we will also revisit those topics in greater depth and to a broader extent. The following two problems exemplify the kinds of questions that we will ask, and often answer, in this course.

The first problem hopefully feels very familiar from a first course in linear algebra.

1.1 Example. Let

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

For what vectors $\mathbf{b} \in \mathbb{R}^3$ can we find $\mathbf{x} \in \mathbb{R}^4$ such that $A\mathbf{x} = \mathbf{b}$?

Before proceeding, we are presuming familiarity with the Euclidean spaces $\mathbb{R}^n$ of column vectors and matrix-vector multiplication. We will revisit these topics. Here

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

with $b_1, b_2, b_3 \in \mathbb{R}^3$, and for convenience we will also write $\mathbf{b} = (b_1, b_2, b_3)$. We will elaborate on this notation later.

We might first note that if $\mathbf{x} = (x_1, x_2, x_3, x_4)$, then

$$A\mathbf{x} = \begin{bmatrix} x_1 + 2x_2 + x_3 + 7x_4 \\ 2(x_1 + 2x_2 + x_3 + 7x_4) \\ 2x_3 + 8x_4 \end{bmatrix}, \tag{1.1}$$

and so if $\mathbf{x} \in \mathbb{R}^4$ and $\mathbf{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$ satisfy $A\mathbf{x} = \mathbf{b}$, then $b_2 = 2b_1$. This is a "solvability condition" for the problem $A\mathbf{x} = \mathbf{b}$, and so we will not be able to solve it for all $\mathbf{b} \in \mathbb{R}^3$; take $\mathbf{b} = (0, 1, 0)$, for example.

It turns out that this solvability condition is both necessary and sufficient for being able to solve the problem. Elementary row operations show that $A\mathbf{x} = \mathbf{b}$ if and only if $R\mathbf{x} = \mathbf{c}$, where

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} b_1 - b_3/2 \\ b_3/2 \\ b_2 - 2b_1 \end{bmatrix}. \tag{1.2}$$

In turn, the problem $R\mathbf{x} = \mathbf{c}$ is equivalent to the system

$$\begin{cases} x_1 + 2x_2 + 3x_4 = b_1 - b_3/2 \\ x_3 + 4x_4 = b_3/2 \\ 0 = b_2 - 2b_1. \end{cases} \tag{1.3}$$

The third equation is $b_2 - 2b_1 = 0$, which is the solvability condition. In the first two equations, we can solve for the pivot variables x_1 and x_3 in terms of the free variables x_2 and x_4 to represent the solution $\mathbf{x}$ "parametrically" as

$$\mathbf{x} = \begin{bmatrix} b_1 - b_3/2 \\ 0 \\ b_3/2 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}. \tag{1.4}$$

This tells us all solutions when the solvability condition is met and in particular that solutions are not unique.

All of this should be reasonably familiar from a first course in linear algebra. Here are some follow-up questions, which may well also have been addressed in that first course.

1. Is there a meaningful, natural way to "force" uniqueness of the solution? Can we impose some extra conditions on $\mathbf{x}$ to guarantee that, if $\mathbf{b}$ meets the solvability condition, then there is exactly one $\mathbf{x}$ that meets $A\mathbf{x} = \mathbf{b}$? Perhaps we could "minimize" $\mathbf{x}$ relative to some norm (which need not be achieved by taking $x_2 = x_4 = 0$).

2. We know that we can solve $A\mathbf{x} = \mathbf{b}$ precisely when $b_2 = 2b_1$. How does this solvability condition affect, or determine, the "structure" of $\mathbb{R}^3$? If a vector in $\mathbb{R}^3$ does not meet the solvability condition, how is it related to vectors that *do* meet it?

3. What happens if $\mathbf{b}$ does not meet the solvability condition? Is there a meaningful, natural way to "approximate" $\mathbf{b}$ by some other $\widehat{\mathbf{b}} \in \mathbb{R}^3$ for which the "approximate" problem $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$ *does* have a (possibly nonunique) solution? (Here the unknown $\widehat{\mathbf{x}}$ is just meant to emphasize that the problem $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$ is not the same as $A\mathbf{x} = \mathbf{b}$.)

Content from *Linear Algebra* by Meckes & Meckes. Sections 1.1, 1.2, and 1.3 review linear systems, Gaussian elimination, and the RREF (in system form). I will not talk about this in class, and I expect that you are very comfortable with this from Linear I, or will get comfortable soon. Page 67 defines matrix-vector multiplication, but I prefer (2.3) as the definition over (2.2). I also expect that you are comfortable with matrix-matrix multiplication, as treated on pp. 90–96. (Skip transposes and inverses for now.) Finally, you should be familiar with how elementary matrices perform elementary row operations; see pp. 102–104 up to and including Theorem 2.22. We will consider matrix-vector and matrix-matrix multiplication from more abstract perspectives in this class later, and we will also revisit the RREF, albeit somewhat briefly, when we study dimension and rank.

1.2 Problem (!). This is an opportunity to review some of the underlying techniques that were used in the previous example.

(i) Carry out the matrix-vector multiplication that gave (1.1).

(ii) Convince yourself that if $A\mathbf{x} = \mathbf{b}$, then $b_2 = 2b_1$.

(iii) Rewrite the solvability condition in the form $\mathbf{b} \cdot \mathbf{z} = 0$ for some $\mathbf{z} \in \mathbb{R}^3$.

1.3 Problem ($\star$). That example hinged on the equivalence of the problems $A\mathbf{x} = \mathbf{b}$ and $R\mathbf{x} = \mathbf{c}$.

(i) Carry out in detail those elementary row operations that establish this equivalence. That is, convert the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to its reduced row-echelon form $\begin{bmatrix} R & \mathbf{c} \end{bmatrix}$, where R and $\mathbf{c}$ have the forms given in (1.2).

(ii) Convince yourself that the problem $R\mathbf{x} = \mathbf{c}$ is equivalent to the system (1.3).

(iii) Solve (1.3) and explain why the solution has the "parametric" form (1.4).

Our second problem is really the fundamental problem of calculus.

1.4 Example. **(i)** For what continuous functions $g\colon \mathbb{R} \to \mathbb{R}$ is there a differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ such that $f' = g$? Unlike in Example 1.1, we have no qualms about existence. The fundamental theorem of calculus asserts that any continuous function g has an antiderivative. Just take

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

where we can allow $f(0)$ to have any value that we want.

That freedom, however, is the downfall of uniqueness. Without prescribing the "initial condition" $f(0)$ further, any continuous function has infinitely many antiderivatives.

(ii) An initial condition is a "pointwise" phenomenon, as it considers the behavior of a function at a single point. There are various more "global" considerations that we could take. For example, we might restrict consideration to periodic functions. A function $g\colon \mathbb{R} \to \mathbb{R}$ is **1-PERIODIC** if $g(x+1) = g(x)$ for all $x \in \mathbb{R}$. Then we might ask if every 1-periodic function g has a 1-periodic antiderivative. The answer is immediately no: consider $g(x) = 1$.

However, we can work backwards and learn something. Suppose that $g\colon \mathbb{R} \to \mathbb{R}$ is 1-periodic and has a 1-periodic antiderivative f. Then f satisfies both

$$f(x) = f(0) + \int_0^1 g(s)\,ds \quad \text{and} \quad f(x+1) = f(x),\ x \in \mathbb{R}.$$

Taking $x = 1$, we have

$$f(1) = f(0+1) = f(0) \quad \text{and} \quad f(1) = f(0) + \int_0^1 g(s)\,ds,$$

thus

$$f(0) = f(0) + \int_0^1 g(s)\,ds,$$

and therefore
$$\int_0^1 g(s)\ ds = 0.$$
We have discovered a solvability condition for our problem. For $g\colon \mathbb{R} \to \mathbb{R}$ to be 1-periodic and to have a 1-periodic antiderivative, it must be the case that $\int_0^1 g(s)\ ds = 0$.

The natural immediate question is if the logic goes in the other direction. Does the condition $\int_0^1 g(s)\ ds = 0$ imply that putting $f(x) = f(0) + \int_0^1 g(s)\ ds$ for an arbitrary value of $f(0)$ gives a periodic function f? Yes, but this requires some additional properties of integrals. And this does not resolve uniqueness, as $f(0)$ can have an arbitrary value.

It turns out that a way to force uniqueness is to require a similar condition on f. If $g\colon \mathbb{R} \to \mathbb{R}$ is continuous and 1-periodic with $\int_0^1 g(s)\ ds = 0$, then there exists only one differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ such that $f' = g$ and $\int_0^1 f(s)\ ds = 0$.

1.5 Remark. *Calculus teaches us the following about integrals. Let $I \subseteq \mathbb{R}$ be an interval. For each a, $b \in I$, there is a real number $\int_a^b f(s)\ ds$ with the following properties.*

($\int$1) $\displaystyle\int_a^b 1\ ds = b - a.$

($\int$2) $\displaystyle\int_a^c f(s)\ ds + \int_c^b f(s)\ ds = \int_a^b f(s)\ ds$ *for any* $c \in I$.

($\int$3) $\displaystyle\int_a^b (f(s)+g(s))\ ds = \int_a^b f(s)\ ds + \int_a^b g(s)\ ds$ *and* $\displaystyle\int_a^b (\alpha f(s))\ ds = \alpha \int_a^b f(s)\ ds$ *for any continuous functions* f, $g\colon I \to \mathbb{R}$ *and* $\alpha \in \mathbb{R}$.

($\int$4) $\displaystyle\int_a^b f(s)\ ds \ge 0$ *if* $a \le b$ *and* $f(s) \ge 0$ *for* $a \le s \le b$.

Knowing these four properties of the integral alone is enough to establish the **TRIANGLE INEQUALITY***:*
$$\left|\int_a^b f(s)\ ds\right| \le \int_a^b |f(s)|\ ds.$$
From that one can prove the fundamental theorem of calculus: given $f \in \mathcal{V}$, the function
$$F\colon I \to \mathbb{R}\colon x \mapsto \int_0^x f(s)\ ds$$
is differentiable with $F' = f$. Also, if $f\colon I \to \mathbb{R}$ is differentiable and f' is continuous on I, then
$$\int_a^b f'(s)\ ds = f(b) - f(a).$$

1.6 Problem ($\star$). This problem fills in some of the details from part (ii) of Example 1.4. Suppose that $g\colon \mathbb{R} \to \mathbb{R}$ is continuous and 1-periodic with $\int_0^1 g(s)\ ds = 0$.

(i) Put

$$f(x) := \int_0^x g(s)\ ds.$$

Use integral property ($\int 2$) from Remark 1.5 to show that if $G(x) := \int_x^{x+1} g(s)\ ds = 0$ for all $x \in \mathbb{R}$, then f is 1-periodic. Then show that $G' = 0$ and use that to conclude $G = 0$.

(ii) Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is differentiable with $f' = g$ and that $\int_0^1 f(s)\ ds = 0$ as well. (Since $\int_0^1 g(s)\ ds = 0$, we know that f is 1-periodic, too, although that will not play a role here.) Show that $f(0) = \int_0^1 xg(x)\ dx$ and conclude that f is unique. [Hint: *compute* $\int_0^1 f(s)\ ds$ *using the formula* $f(s) = f(0) + \int_0^s g(x)\ dx$ *and interchange the order of integration in the resulting double integral.*]

1.7 Problem ($\star$). Let $g\colon [0,\infty) \to \mathbb{R}$ be continuous. Show that for the problem

$$\begin{cases} f' = g \\ \lim\limits_{x\to\infty} f(x) = 0 \end{cases}$$

to have a solution, it must be the case that g is improperly integrable on $[0,\infty)$, i.e., that $\int_0^\infty g(s)\ ds$ converges. In this case, explain why the solution f is unique.

Here is a more complicated version of Example 1.4.

1.8 Example. Let $g\colon \mathbb{R} \to \mathbb{R}$ be a continuous function. What functions $f\colon \mathbb{R} \to \mathbb{R}$ are twice-continuously differentiable (so f' and f'' exist and f'' is continuous) and solve the ordinary differential equation (ODE)

$$f'' + f = g?$$

That is, we want $f''(x) + f(x) = g(x)$ for all $x \in \mathbb{R}$. This is a version of the ODE that governs the motion of a simple harmonic oscillator (a mass-spring system); the mass here is 1, the spring constant is 1, and there is no friction (because there is no term with f'), while g encapsulates all external forces acting on the oscillator.

The answer turns out to be any function f of the form

$$f(x) = f(0)\cos(x) + f'(0)\sin(x) + (\mathcal{S}g)(x), \tag{1.5}$$

where

$$(\mathcal{S}g)(x) := \sin(x)\int_0^x \cos(y)g(y)\ dy - \cos(x)\int_0^x \sin(y)g(y)\ dy. \tag{1.6}$$

This is the dreaded method of variation of parameters. Our focus here is not deriving this formula (*Having a formula for something is not the same as understanding that thing*) but in exploiting it.

Our first conclusion should be that this ODE *always* has a solution, unlike the previous linear system, and that, like the previous linear system, this solution is never unique without some additional constraints. For example, we might fix the initial conditions to be $f(0) = y_0$ and $f'(0) = y_1$ for some given y_0, $y_1 \in \mathbb{R}$, and that specifies exactly what f is.

We might also add some "qualitative" constraints to the problem. Since g represents an external force, we could look at *periodic* forcing: say that g is 2π-periodic, so $g(x + 2\pi) = g(x)$ for all $x \in \mathbb{R}$. Will f also be 2π-periodic? It turns out that if f in the form (1.5) is 2π-periodic, then g must meet

$$\int_0^{2\pi} \cos(y)g(y)\, dy = 0 \quad \text{and} \quad \int_0^{2\pi} \sin(y)g(y)\, dy = 0. \tag{1.7}$$

These are "solvability conditions" for the problem $f''+f = g$ when we work with 2π-periodic functions. Not every 2π-periodic g will meet these conditions, and so we cannot always solve the problem now. However, the solvability conditions here *do* guarantee existence: if g meets (1.7), then $\mathcal{S}g$ is 2π-periodic, and so f as defined by (1.5) is 2π-periodic, too. This is just like how the solvability condition for the linear system was necessary and sufficient for existence of solutions.

We are still faced with a lack of uniqueness in this 2π-periodic setting. We could impose initial conditions, but it also turns out that requiring the solution f to meet the solvability conditions

$$\int_0^{2\pi} \cos(y)f(y)\, dy = 0 \quad \text{and} \quad \int_0^{2\pi} \sin(y)f(y)\, dy = 0 \tag{1.8}$$

guarantees uniqueness.

What if g does not meet the solvability conditions? Can we approximate g by some $\widehat{g}$ that does and then solve $\widehat{f}'' + \widehat{f} = \widehat{g}$? (There is an unfortunate conflict of notation with Fourier coefficients here, by the way.) We are working with functions on $\mathbb{R}$, not vectors, and there are probably many meaningful ways to approximate a function by other functions—by polynomials (such as, but not limited to, Taylor polynomials), by trigonometric polynomials (as in partial sums of Fourier series). What is best? What is the right thing to do?

1.9 Problem (+). **(i)** Check that any function of the form (1.5) does solve $f'' + f = g$. [Hint: *this is not the same as proving that any solution to $f'' + f = g$ has the form (1.5); plug this formula into the ODE and calculate away. The most complicated part will be differentiating $\mathcal{S}g$, which will require the fundamental theorem of calculus.*]

(ii) Use a trigonometric addition formula to explain why

$$(\mathcal{S}g)(x) = \int_0^{x} \sin(x - y)g(y)\, dy. \tag{1.9}$$

(iii) Suppose that f, $g \colon \mathbb{R} \to \mathbb{R}$ are 2π-periodic with f twice-continuously differentiable, g continuous, and $f'' + f = g$. Integrate by parts to show that g must meet the solvability conditions (1.7). [Hint: *use what f and g do to rewrite*

$$\int_0^{2\pi} \sin(y)g(y)\ dy = \int_0^{2\pi} \sin(y)(f''(y) + f(y))\ dy.$$

Integrate by parts in $\int_0^{2\pi} \sin(y)f''(y)\ dy$. *How does this help? Repeat this work on the integral with cosine.*]

(iv) Show that the function f defined by (1.5) is 2π-periodic if and only if $\mathcal{S}g$ is 2π-periodic.

(v) Suppose that $g\colon \mathbb{R} \to \mathbb{R}$ is continuous and 2π-periodic and meets the solvability conditions (1.7). Show that $\mathcal{S}g$ is also 2π-periodic. [Hint: *use a trigonometric addition formula to rewrite $\mathcal{I}(x)$ in such a way that the solvability conditions from* (1.7) *appear.*]

(vi) Suppose that f_1, f_2, $g\colon \mathbb{R} \to \mathbb{R}$ are 2π-periodic with f_1 and f_2 twice-continuously differentiable, g continuous, and $f_j'' + f_j = g$ for $j = 1, 2$. Suppose also that f_1 and f_2 meet the solvability condition (1.8). Prove that $f_1 = f_2$. [Hint: *show that $f := f_1 - f_2$ solves $f'' + f = 0$. Use* (1.5) *to find a formula for f. Compute* $\int_0^{2\pi} \sin(y)f(y)\ dy$ *and* $\int_0^{2\pi} \cos(y)f(y)\ dy$ *from this formula and, using the fact that these integrals are* 0 *by* (1.8), *show that* $f(0) = f'(0) = 0$.]

Day 2: Wednesday, January 14.

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 of a function, codomain of a function, range of a function, image of a set under a function, $\mathbb{R}^n$ (as a set of functions), $\mathbb{R}^{m\times n}$ (as a set of functions)

The problems of Examples 1.1 and 1.4 should look cosmetically different, but they are really asking many of the same questions. Example 1.1 compresses into the following problem: with $\mathcal{V} = \mathbb{R}^4$, $\mathcal{W} = \mathbb{R}^3$, and

$$\mathcal{T}\colon \mathcal{V} \to \mathcal{W}\colon \mathbf{v} \mapsto A\mathbf{v},$$

where $A \in \mathbb{R}^{3\times 4}$ was given in that example, find all $\mathbf{w} \in \mathcal{W}$ for which the problem $\mathcal{T}\mathbf{v} = \mathbf{w}$ has a solution. This is fundamentally a *linear problem*: $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, and $\mathcal{T}$ is a linear operator from $\mathcal{V}$ to $\mathcal{W}$. We will define these terms in detail shortly, but the idea is that we can add elements of $\mathcal{V}$ and multiple them by real numbers (and the same for $\mathcal{W}$), and addition and multiplication behave exactly as we expect. This is a consequence of how we define vector addition and scalar multiplication in $\mathbb{R}^n$, which is really a *componentwise* definition. As for $\mathcal{T}$, it is a function from $\mathcal{V}$ to $\mathcal{W}$ (a way of pairing elements of $\mathcal{V}$ uniquely with elements of $\mathcal{W}$) that "respects" the behavior of addition and multiplication in $\mathcal{V}$ and $\mathcal{W}$. This is a consequence of how we define matrix-vector multiplication.

Example 1.4 compresses into a similar problem: with

$$\mathcal{V} = \{f \colon \mathbb{R} \to \mathbb{R} \mid f \text{ is differentiable}, \ f' \text{ is continuous}, f \text{ is 1-periodic}\}$$

and

$$\mathcal{W} = \{g \colon \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}, f \text{ is 1-periodic}\},$$

and

$$\mathcal{T} \colon \mathcal{V} \to \mathcal{W} \colon f \mapsto f',$$

find all $g \in \mathcal{W}$ for which the problem $\mathcal{T}f = g$ has a solution. Again, this is linear, because $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, and $\mathcal{T}$ is a linear operator from $\mathcal{V}$ to $\mathcal{W}$. These are consequences of how we define function addition and multiplication (which is really a *pointwise* definition) and of the fundamental linearity of the limit (and thus of continuity and derivatives) from calculus.

These kinds of problems are the central object of study in this course. Specifically, we will consider vector spaces $\mathcal{V}$ and $\mathcal{W}$ and a linear operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$. (Many first courses in linear algebra consider this already, and many do not. We will review notions of vector spaces and linear operators from scratch.) Given a vector $w \in \mathcal{W}$, we ask if we can find $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. If we can, we then ask if v is unique; if it is, then we write $v = \mathcal{T}^{-1}w$. If v is not unique, we ask if there is a natural, meaningful way to choose v to be unique (what is "natural" and "meaningful" will depend on the precise context of what $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{T}$ are). If we cannot solve $\mathcal{T}v = w$, then we ask if there is a natural, meaningful way to "approximate" w by some other $\widehat{w} \in \mathcal{W}$ such that the problem $\mathcal{T}v = \widehat{w}$ *does* have a solution. In short, *how do we characterize and understand the range of a linear operator?*

Addressing these questions will involve several overlapping, interlaced areas of focus. We will study vector spaces and their properties—in particular, subspaces, bases, and dimension. We will study linear operators and their properties—in particular, range, kernel, composition, invertibility, eigenvalues, and their interactions with properties of vector spaces. And we will study geometric aspects of vector spaces and linear operators that arise in natural, meaningful ways for many problems—in particular, inner products, norms, orthogonality, orthonormal bases, and adjoints of linear operators. In particular, we will see how bases for finite-dimensional vector spaces afford us powerful control over these spaces and the operators between them by serving as unique coordinate systems, and we will see how geometry offers versatile characterizations of structures associated with linear operators. Although we will sometimes stray from a direct focus on the problem $\mathcal{T}v = w$, that will always be our goal: *what more can we understand about linear operators?* When possible, we will learn how to factor operators into products of simpler, meaningful operators that store different types of data relevant to the original operator.

That being said, we will start small, with functions. 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. And even the most precise definition of basis is ultimately couched in the notion of function.

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 (s, t) equals (x, y) if and only if $x = s$ and $y = t$. 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 A and B be sets. The* ORDERED PAIR *whose first component is $x \in A$ and whose second component is $y \in B$ is the set*

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

The CARTESIAN PRODUCT *of A and B is the set $A \times B$ of all ordered pairs with first component in A and second component in B:*

$$A \times B = \{(x, y) \mid x \in A,\ y \in B\}.$$

2.4 Example. Let $A = \{1, 2, 3\}$ and $B = \{4, 5\}$. Then

$$A \times B = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}.$$

2.5 Problem (!). Let A and B be as in Example 2.4. Determine the elements of the following sets.

(i) $B \times A$

(ii) $\varnothing \times A$

2.6 Definition. *Let A and B be sets. A* **FUNCTION** $f\colon A \to B$ **FROM** A **TO** B *is a set $f \subseteq A \times B$ such that for every $x \in A$, there is a unique $y \in B$ such that $(x, y) \in f$. We use the following additional terminology and notation.*

(i) *If $(x, y) \in f$, then we write $y = f(x)$.*

(ii) *The set A is the* **DOMAIN** *of f, and the set B is the* **CODOMAIN** *of f.*

(iii) *The* **RANGE** *of f (sometimes the* **IMAGE** *of f) is the set*

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

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

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

That a function is a set of ordered pairs encodes the act of pairing: elements of A are paired with elements of B as ordered pairs. The more precise quantified statement that each $x \in A$ is paired with precisely one $y \in B$ encodes the uniqueness of this pairing. In calculus we perhaps more often think of the set $\{(x, f(x)) \mid x \in A\}$ as the **GRAPH** of f, but for us this set really *is what f is.*

2.7 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.)

Content from *Linear Algebra* by Meckes & Meckes. Pages 379–380 define functions. The definition on p. 379 does not use ordered pairs and is perfectly sufficient for almost all encounters with functions that we will have in this course. We will not do much with function composition and inversion for *arbitrary* functions as on pp. 380–382, but this will reinforce work with linear operators later.

2.8 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 $\mathbb{R}$?

2.9 Problem (!). Suppose that A and B are sets, $x \in A$, and $f\colon A \to B$ is a function. Which, if any, of the objects x, $\{x\}$, $f(x)$, $f(\{x\})$, and $\{f(x)\}$ are equal?

2.10 Problem (+). 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.

2.11 Definition. *If A and B are sets, we denote by*

$$B^A$$

the set of all functions from A to B.

2.12 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,

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

2.13 Problem ($\star$). What are all of the elements of $\{1, -1\}^{\{1,2,3,4\}}$?

We will mostly consider functions whose codomains are $\mathbb{R}$ (and sometimes the complex numbers $\mathbb{C}$) or functions that are linear operators between vector spaces. For the former, the additional algebraic structure of the codomain ensures that we can do algebra (and arithmetic) on functions.

2.14 Example. Let $f, g \in \mathbb{R}^{\mathbb{R}}$, where we are taking the domain to be $\mathbb{R}$ right now just for simplicity. Then we have a natural notion of adding $f + g$, which should be that $(f+g)(x) = f(x) + g(x)$. Be careful: we want $f + g \in \mathbb{R}^{\mathbb{R}}$, too, but for any $x \in \mathbb{R}$, we have $f(x) + g(x) \in \mathbb{R}$. That is,

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

We might also write more formulaically

$$f + g \colon \mathbb{R} \to \mathbb{R} \colon x \mapsto f(x) + g(x).$$

We emphasize here that $f + g \in \mathbb{R}^{\mathbb{R}}$, whereas $f(x) + g(x) \in \mathbb{R}$ for each $x \in \mathbb{R}$.

Likewise, for $\alpha \in \mathbb{R}$, we have a natural notion of what αf should be: it is the function on $\mathbb{R}$ given pointwise by $(\alpha f)(x) = \alpha f(x)$. Again, $\alpha f \in \mathbb{R}^{\mathbb{R}}$, but $\alpha f(x) \in \mathbb{R}$ for any $x \in \mathbb{R}$.

Of course, we can also multiply functions $f, g \in \mathbb{R}^{\mathbb{R}}$ to get a natural product $fg \in \mathbb{R}^{\mathbb{R}}$, but this operation turns out to be somewhat less important in linear algebra than in calculus. Also, the set $\mathbb{R}^{\mathbb{R}}$ is far too big for daily use; in calculus we restrict ourselves to much nicer functions, chief among them the continuous, differentiable, and integrable functions.

The careful reader will note that none of the function arithmetic above involved the domain. We did not need to be able to add *inputs* to f and g to be able to add their *outputs*. So, if X is any set, and if $f, g \in \mathbb{R}^X$ and $\alpha \in \mathbb{R}$, then we can define $f + g$ and αf pointwise as above. In fact, we do not really need to restrict to real numbers.

2.15 Definition. *We denote by $\mathbb{F}$ either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$.*

2.16 Definition. *Let X be a set, $f, g \in \mathbb{F}^X$, and $\alpha \in \mathbb{F}$. We define $f + g$, $\alpha f \in \mathbb{F}^X$ by*

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

and

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

All of this is exactly how we defined arithmetic in $\mathbb{R}^n$ in a first course in linear algebra, except there we probably used the word “componentwise” instead of pointwise. For example, in $\mathbb{R}^2$, we add

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

We can think of these vectors as functions defined on the "discrete" domain $\{1, 2\}$, and their "componentwise" addition is really their *pointwise* addition.

This leads us to a rigorous (if not often useful) definition of $\mathbb{R}^n$ and column vectors: they are functions on the set of integers from 1 to n. There is no need, however, to stop with $\mathbb{R}$ here.

2.17 Definition. *Let $n \geq 1$. Then*

$$\{1, \ldots, n\} := \{k \in \mathbb{N} \mid 1 \leq k \leq n\} = [1, n] \cap \mathbb{N}.$$

2.18 Definition. $\mathbb{F}^n := \mathbb{F}^{\{1,\ldots,n\}}$ *for $n \geq 2$ and $\mathbb{F}^1 := \mathbb{F}$.*

Of course, we really do not think of functions in $\mathbb{F}^n$ as actual functions. Typically, if $f \in \mathbb{F}^n$, then we put $v_k := f(k)$ for $k = 1, \ldots, n$ and declare the two symbols

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad \text{and} \quad (v_1, \ldots, v_n)$$

to be equal to both each other and to f. Strictly speaking,

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = (v_1, \ldots, v_n) = \{(k, v_k)\}_{k=1}^n.$$

2.19 Remark. *This can lead to some awkwardness when $n = 2$, as then we have*

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = (v_1, v_2) = \{(1, v_1), (2, v_2)\}.$$

That is, for $n = 2$ we are unfortunately overworking the notation of ordered pair. For consistency, we still prefer to think of vectors in $\mathbb{R}^2$ as functions in $\mathbb{R}^{\{1,2\}}$, and context will guide us as to the meaning of (v_1, v_2).

We can also think of matrices rigorously (if uselessly) as functions. Intuitively, an $m \times n$ matrix is a rectangular array of numbers with m rows and n columns. This perspective no doubt carried us through our first course in linear algebra, and most of the time it will do so here. However, this second pass at the subject is the time for thinking precisely. A matrix like

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

is really a way of associating each of the six entries with a real number. We want to specify where the entry falls with respect to both rows and columns, so we need two coordinates for the entry.

2.20 Definition. **(i)** $\mathbb{F}^{m\times n} := \mathbb{F}^{\{1,\ldots,m\}\times\{1,\ldots,n\}}$ *for* $m \geq 1$ *and* $n \geq 2$.

(ii) $\mathbb{F}^{m\times 1} := \mathbb{F}^m$ *for any* $m \geq 1$.

(iii) $\mathbb{F}^{1\times 1} := \mathbb{F}$.

(iv) *We do not identify* $\mathbb{F}^{1\times n}$ *and* $\mathbb{F}^n$ *for* $n \geq 2$.

2.21 Problem (!). How are $\mathbb{F}^{1\times n}$ and $\mathbb{F}^n$ different? [Hint: *think about their elements as functions. What are the domains?*]

Here is a slight generalization of the notion of $\mathbb{F}^n$ that will be useful from time to time, especially when we speak precisely about bases.

2.22 Definition. *Let* Y *be a set. A* LIST *of length* $n \geq 1$ *in* Y *is a function in* $Y^{\{1,\ldots,n\}}$. *If* $f \in Y^{\{1,\ldots,n\}}$ *with* $f(k) = y_k$ *for* $k = 1, \ldots, n$, *then we define* $(y_1, \ldots, y_n) := f$. *That is,*

$$(y_1, \ldots, y_n) = \{(k, y_k)\}_{k=1}^n.$$

For example, vectors in $\mathbb{F}^n$ are lists of length n in $\mathbb{F}$. Again, this can be awkward for $n = 2$: is (y_1, y_2) an ordered pair in $Y \times Y$ or the list $(y_1, y_2) = \{(1, y_1), (2, y_2)\}$? Context will make this clear. Right now, the primary value of the concept of list will be that it makes the concept of vector space excruciatingly precise.

Day 3: Friday, January 16.

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 for a vector space (what does it do?), additive inverse for a vector space (what does it do?)

Possibly most of the meaningful examples of vectors and vector spaces in a first course in linear algebra are at the level of column vectors. We can think of these as functions with finite, discrete domains: the n integers in the set $\{1, \ldots, n\}$, each of which is separated from its neighbors by a distance of 1. At the other extremes are functions in $\mathbb{R}^{\mathbb{R}}$ or $\mathbb{R}^I$ for $I \subseteq \mathbb{R}$ an interval. Here the domains are "continuous" because intervals are unbroken (strictly speaking, connected) and infinite (uncountably infinite, actually). Even when I is a proper subinterval of $\mathbb{R}$, the set $\mathbb{R}^I$ is far too large for calculus purposes.

3.1 Definition. *Let $I \subseteq \mathbb{R}$ be an interval.*

(i) *The set $\mathcal{C}(I)$ consists of all real-valued continuous functions on I.*

(ii) *A function $f: I \to \mathbb{R}$ is* CONTINUOUSLY DIFFERENTIABLE *on I if f is differentiable on I and if $f' \in \mathcal{C}(I)$.*

(iii) *Let $r \geq 1$ be an integer. The set $\mathcal{C}^r(I)$ consists of all r-times continuously differentiable functions on I. That is, $f \in \mathcal{C}^r(I)$ if and only if the r derivatives $f', \ldots, f^{(r)}$ exist on I with $f^{(r)} \in \mathcal{C}(I)$.*

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

(v) $\mathcal{C}^\infty(I) := \cap_{r=0}^\infty \mathcal{C}^r(I)$. *The functions in $\mathcal{C}^\infty(I)$ are* INFINITELY DIFFERENTIABLE.

We will only talk about calculus in the context of real numbers, although there is much to be said when the domain or codomain of a function is complex.

3.2 Example. Define

$$f: \mathbb{R} \to \mathbb{R}: x \mapsto |x|.$$

Then $f \in \mathcal{C}(\mathbb{R})$ but $f \notin \mathcal{C}^r(\mathbb{R})$ for any $r \geq 1$.

Most of the functions that we meet in calculus courses and for which we have "familiar" formulas are infinitely differentiable or at worst piecewise continuous (which we have not specified here). Often in differential equations one studies an rth-order differential equation and desires solutions that are r-times continuously differentiable; the idea is to have some extra control over the rth derivative beyond its existence. Nonetheless, for each $r \geq 1$, it is possible to construct $f \in \mathcal{C}^r(I)$ such that $f \notin \mathcal{C}^{r-1}(I)$ and to construct an r-times differentiable function whose rth derivative is not continuous.

The sets $\mathbb{R}^n$ and $\mathcal{C}^r(I)$ may look very different, but algebraically they have much in common. We list some similar properties below for $r = 0$ and, for convenience, $I = [0, 1]$.

1. *Addition.* We can add vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ componentwise and get a new vector $\mathbf{v}+\mathbf{w} \in \mathbb{R}^n$. This is just a consequence of how we define vector addition in $\mathbb{R}^n$ really as function addition in $\mathbb{R}^{\{1,\ldots,n\}}$. Likewise, we can add functions $f, g \in \mathcal{C}([0,1])$ pointwise and get a new function $f + g \in \mathcal{C}([0,1])$. This is a little deeper: we show in calculus that defining $f + g$ by $(f + g)(x) = f(x) + g(x)$ does yield a new continuous function $f + g$ when f and g are both continuous.

2. *Scalar multiplication.* We can multiply $\mathbf{v} \in \mathbb{R}^n$ by $\alpha \in \mathbb{R}$ componentwise and get a new vector $\alpha\mathbf{v} \in \mathbb{R}^n$. Again, this is a consequence of how we define multiplication by a scalar in $\mathbb{R}^{\{1,\ldots,n\}}$. Likewise, we can multiply $\alpha \in \mathbb{R}$ and $f \in \mathcal{C}([0,1])$ pointwise and get a new function $\alpha f \in \mathcal{C}([0,1])$. Again, the extra step here is proving continuity of αf.

3. *Arithmetic works as it should.* We have identities like $\mathbf{v}+\mathbf{w} = \mathbf{w}+\mathbf{v}$ in $\mathbb{R}^n$ and $(\alpha+\beta)f = \alpha f + \beta f$ in $\mathcal{C}([0,1])$. This mostly boils down to componentwise or pointwise definitions and how arithmetic works in $\mathbb{R}$.

4. *Additive identity.* Denote by $\mathbf{0}_n$ the vector in $\mathbb{R}^n$ whose entries are all $0 \in \mathbb{R}$. Then $\mathbf{v} + \mathbf{0}_n = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. Denote, annoyingly, by 0 the function from $[0,1]$ to $\mathbb{R}$ whose value at any $x \in [0,1]$ is 0. That is,

$$0 \colon [0,1] \to \mathbb{R} \colon x \mapsto 0.$$

Then $f + 0 = f$ for all $f \in \mathcal{C}([0,1])$.

5. *Additive inverse.* For $\mathbf{v} \in \mathbb{R}^n$, the vector $(-1)\mathbf{v}$ satisfies $\mathbf{v} + (-1)\mathbf{v} = \mathbf{0}_n$. Of course, we usually just write $-\mathbf{v}$, not $(-1)\mathbf{v}$, but $(-1)\mathbf{v}$ emphasizes that we are multiplying each entry of $\mathbf{v}$ by -1. Likewise, for $f \in \mathcal{C}([0,1])$, we have $f + (-1)f = 0$, where $((-1)f)(x) = -f(x)$.

Content from *Linear Algebra* by Meckes & Meckes. The examples on p. 50 discuss these similarities.

These properties are (some of) the fundamental ways that a vector space behaves. It is possible to talk about vector spaces over very general fields; we will do so only for the real and complex numbers.

3.3 Definition. *The symbol $\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$ and will always mean the same in a given context. We denote addition in $\mathbb{F}$ by $+$ as usual, so for α, $\beta \in \mathbb{F}$, we have $\alpha + \beta \in \mathbb{F}$. We denote scalar multiplication in $\mathbb{F}$ by juxtaposition, so the product of α, $\beta \in \mathbb{F}$ is $\alpha\beta$.*

Content from *Linear Algebra* by Meckes & Meckes. Section 1.4 presents fields in the abstract. This is wholly optional reading. Maybe the most important point is that from a (relatively) small list of axioms (p. 39), you can prove all of the familiar properties of arithmetic in a field. See pp. 40–43 and the "Bottom Line" boxes on pp. 40 and 43. The rest of the section revisits linear systems of equations and Gaussian elimination in the context of a field.

3.4 Definition. *A* VECTOR SPACE OVER $\mathbb{F}$ *is a list of length 4 of the form $(\mathcal{V}, \mathbb{F}, +_{\mathcal{V}}, \cdot)$, where $\mathcal{V}$, $+_{\mathcal{V}}$, and $\cdot$ satisfy the following.*

- *$\mathcal{V}$ is a nonempty set.*
- *$+_{\mathcal{V}} \colon \mathcal{V} \times \mathcal{V} \to \mathcal{V} \colon (v, w) \mapsto v +_{\mathcal{V}} w$ is a function that satisfies the axioms below.*
- *$\cdot \colon \mathbb{F} \times \mathcal{V} \to \mathcal{V} \colon (\alpha, v) \mapsto \alpha \cdot v$ is a function that satisfies the axioms below.*

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* $w \in \mathcal{V}$ *such that* $v +_{\mathcal{V}} w = 0_{\mathcal{V}}$.

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* $v, w \in \mathcal{V}$.

Content from *Linear Algebra* by Meckes & Meckes. These axioms appear on p. 51; I have taken their grouping from Strang's *Introduction to Linear Algebra* (Sixth Edition). The important thing to consider is the "Bottom Line" box on p. 51 and the notational remarks on p. 52. Do Quick Exercise #21 on p. 52.

3.5 Remark. **(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 additive identity* $0_{\mathcal{V}}$ *is unique and therefore merits a special symbol; of course we call this the* ZERO VECTOR *for* $\mathcal{V}$.

(iv) *We can also show that the additive inverse is unique and therefore merits the special symbol* $-v$. *That is, for each* $v \in \mathcal{V}$, *the vector* $-v \in \mathcal{V}$ *that satisfies* $v + (-v) = 0_{\mathcal{V}}$. *We will shortly show as well that* $-v = (-1) \cdot v$, *and so 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* $\alpha, \beta \in \mathbb{F}$ *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* $\mathcal{V}$*.*

(vii) *Typically we do not feel the need to denote vector addition in* $\mathcal{V}$ *by the special symbol* $+_{\mathcal{V}}$ *but will use the ordinary* $+$ *as in* $\mathbb{F}$*; context will make clear what kind of addition is occurring. Likewise, we usually write* αv*, not* $\alpha \cdot v$*, outside of the special emphases in these axioms. (In these axioms, we are writing* $\cdot$*, not* $\cdot_{\mathcal{V}}$*, as we already have a different notation available for multiplication in* $\mathbb{F}$*: juxtaposition.)*

Here is possibly the most important "general" example of a "concrete" vector space that we will encounter.

3.6 Example. Let X be a set. The function set $\mathbb{F}^X$ is a vector space over $\mathbb{F}$ when we define addition and scalar multiplication in the natural ways:

$$f +_{\mathbb{F}^X} g \colon X \to \mathbb{F} \colon x \mapsto f(x) + g(x) \tag{3.1}$$

and

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

This is mostly a good exercise in reading notation. For the definition of vector addition, in (3.1), we want to pair any $f, g \in \mathbb{F}^X$ as a new function $f +_{\mathbb{F}^X} g \in \mathbb{F}^X$, which means that for each $x \in X$, we have to define an element $(f +_{\mathbb{F}^X} g)(x)$. We need to do the same for any $\alpha \in \mathbb{F}$ and $f \in \mathbb{F}^X$.

That $+_{\mathbb{F}^X}$ and $\cdot$ satisfy the vector space axioms is mostly a consequence of these pointwise definitions and the arithmetic and algebraic properties of $\mathbb{F}$. For the additive identity, the function

$$0_{\mathbb{F}^X} \colon X \to \mathbb{F} \colon x \mapsto 0$$

satisfies $f +_{\mathbb{F}^X} 0_{\mathbb{F}^X} = f$, while for the additive inverse, the function

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

satisfies $f +_{\mathbb{F}^X} (-f) = 0$. To be clear, here we are using the convention that $-\alpha = (-1)\alpha$ for any $\alpha \in \mathbb{F}$.

This shows that $\mathbb{F}^n$ and $\mathbb{F}^{m \times n}$ are vector spaces over $\mathbb{F}$. Most of the vector spaces that we use will either be $\mathbb{F}^X$ for a good choice of X (like $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$) or a *subspace* of $\mathbb{F}^I$ for some interval $I \subseteq \mathbb{R}$ (like $\mathcal{C}([0,1])$). We will rarely, if ever, use the baroque notation $+_{\mathbb{F}^X}$ and $0_{\mathbb{F}^X}$ after this.

3.7 Problem (!). Using the definitions in the example above, rewrite $f +_{\mathbb{F}^X} g$, $\alpha \cdot f$, $0_{\mathbb{F}^X}$, and $-f$ as sets of ordered pairs.

It is rarely challenging to show that given candidate functions for vector addition and scalar multiplication on a set satisfy the algebraic axioms of Definition 3.4. This is because those functions are usually defined transparently in terms of operations on $\mathbb{F}$ or, more generally, a previously established vector space. Rather, the challenge is usually more subtle: do the candidates actually map back into the proposed vector space? Are their codomains really the purported vector space? This challenge will become more apparent when we consider the concept of subspace, which is how most interesting vector spaces arise.

3.8 Problem (+). Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and let X be a set. Define vector addition and scalar multiplication on $\mathcal{V}^X$ in terms of vector addition and scalar multiplication on $\mathcal{V}$ so that $\mathcal{V}^X$ becomes a vector space with these operations.

The vector space axioms should be unsurprising: things works the way that they should work. What might be surprising is that these axioms are *all* that we need to show that things works the way that they should work. The following shows how from those axioms we can derive important (if unsurprising) consequences. We select these consequences to illustrate both the use of the axioms and some common proof techniques.

3.9 Example. Let $\mathcal{V}$ be a vector space over $\mathbb{F}$.

(i) *The additive identity is unique.* This proof illustrates the slogan that "what things do defines what things are." Suppose that w, $\widetilde{w} \in \mathcal{V}$ both "do the job" of the additive identity:

$$(1)\ v + w = v \text{ for all } v \in \mathcal{V}. \quad \text{and} \quad (2)\ v + \widetilde{w} = v \text{ for all } v \in \mathcal{V}.$$

To show uniqueness, we want to prove that $w = \widetilde{w}$. This is how many uniqueness proofs go: assume that two things do the job and show that the two things are the same.

The trick here is to exploit the "for all" quantifier by introducing the objects that we care about. In (1), put $v = \widetilde{w}$ to get $\widetilde{w} + w = \widetilde{w}$. In (2), put $v = w$ to get $w + \widetilde{w} = w$. Since addition is commutative,

$$\widetilde{w} = \widetilde{w} + w = w + \widetilde{w} = w.$$

(ii) $0v = 0_{\mathcal{V}}$ *for all* $v \in \mathcal{V}$. On the left the symbol $0 \in \mathbb{F}$ is the additive identity in $\mathbb{F}$; here it may be helpful to distinguish the zero vector as $0_{\mathcal{V}}$. We could try proving this via the slogan "what things do defines what things are" and attempt to show that $0v + w = w$ for any $w \in \mathcal{W}$. This might be hard, however, as we do not really know how v and w would interact.

Instead, the trick is algebra in $\mathbb{F}$:

$$0v = (0 + 0)v = 0v + 0v,$$

where we have used distribution. Then we may add the additive inverse of $0v$ to both sides:

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

We then have

$$0_{\mathcal{V}} = 0v$$

as desired; on the right we used associativity of addition to regroup as

$$(0v + 0v) + (-0v) = 0v + ((0v) + (-0v)) = 0v + 0_{\mathcal{V}} = 0v.$$

(iii) *$\alpha 0_{\mathcal{V}} = 0_{\mathcal{V}}$ for all $\alpha \in \mathcal{V}$.* This illustrates proof by cases. First, if $\alpha = 0$, then we now know that $00_{\mathcal{V}} = 0_{\mathcal{V}}$. Next, if $\alpha \neq 0$, then we can show that $\alpha 0_{\mathcal{V}}$ does the job of the zero vector. Let $v \in \mathcal{V}$. Then

$$\alpha 0_{\mathcal{V}} + v = \alpha(0_{\mathcal{V}} + \alpha^{-1}v) = \alpha(\alpha^{-1}v) = (\alpha\alpha^{-1})v = v.$$

Here we used what the zero vector does in the third equality and associativity of multiplication.

(iv) *If $\alpha v = 0_{\mathcal{V}}$ for some $\alpha \in \mathbb{F}$ and $v \in \mathcal{V}$, then $\alpha = 0$ or $v = 0_{\mathcal{V}}$.* Here we use the equivalence of the statements $P \Longrightarrow (Q \text{ or } R)$ and $(P \text{ and not } Q) \Longrightarrow R$. Specifically, we show that if $\alpha v = 0_{\mathcal{V}}$ and $\alpha \neq 0$, then $v = 0_{\mathcal{V}}$. Since $\alpha \neq 0$, we may divide: $\alpha v = 0_{\mathcal{V}}$ implies $\alpha^{-1}(\alpha v) = \alpha^{-1}0_{\mathcal{V}}$, thus $v = 0_{\mathcal{V}}$.

(v) *The additive inverse is unique.* Let $v \in \mathcal{V}$ and suppose that $w, \widetilde{w} \in \mathcal{V}$ "do the job" of being the additive inverse of v. That is, $v + w = 0_{\mathcal{V}}$ and $v + \widetilde{w} = 0_{\mathcal{V}}$. We want to show that $w = \widetilde{w}$. This is an example of the proof technique of leaving the proof to the reader.

(vi) *$-v = (-1)v$ for all $v \in \mathcal{V}$.* We emphasize here that $-v$ is merely the symbol for the additive inverse of v, and it is defined by what it does: $v + (-v) = 0_{\mathcal{V}}$. We want to show that the vector $(-1)v$ also does this. That is, the goal is $v + (-1)v = 0_{\mathcal{V}}$. We can achieve this by factoring:

$$v + (-1)v = 1v + (-1)v = (1 + (-1))v = 0v = 0_{\mathcal{V}}.$$

3.10 Problem (!). Complete the proof of the uniqueness of the additive inverse begun in part (v) of Example 3.9. [Hint: *subtract the two equations $v + w = 0_{\mathcal{V}}$ and $v + \widetilde{w} = 0_{\mathcal{V}}$.*]

Content from *Linear Algebra* by Meckes & Meckes. This example is largely the content of Theorem 1.11 on pp. 57–58. Some related techniques appear in the proof of Theorem 5 for field arithmetic on pp. 42–43.

Day 4: Wednesday, January 21.

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 of a vector space (N), polynomial

A nice example of a vector space that is both a function space and that, morally, sits between spaces like $\mathbb{R}^n$ and $\mathbb{R}^{\mathbb{R}}$ is the space of sequences.

4.1 Definition. *Denote the natural numbers (the positive integers) by $\mathbb{N}$ and put*

$$\mathbb{F}^\infty := \mathbb{F}^{\mathbb{N}}.$$

A **SEQUENCE** *in $\mathbb{F}$ is a vector in $\mathbb{F}^\infty$. If $f \in \mathbb{F}^\infty$ and $f(k) = a_k$ for $k \in \mathbb{N}$, then we write $f = (a_k)$.*

Of course, $\mathbb{F}^\infty$ is a vector space with the usual componentwise addition and scalar multiplication.

4.2 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 $\mathbb{C}$.

As fundamental an example of a vector space as the function space $\mathbb{F}^X$ is, it is not sufficient by itself. Spaces like $\mathbb{R}^{\mathbb{R}}$ or $\mathbb{R}^{[0,1]}$ or even $\mathbb{R}^\infty$ are just "too large" to be useful in calculus. Most interesting vector spaces really arise as *subspaces* of some larger ambient space.

4.3 Definition. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and let $\mathcal{U} \subseteq \mathcal{V}$. Then $\mathcal{U}$ is a* **SUBSPACE** *of $\mathcal{V}$ if the following hold:*

(i) *Presence of the zero vector: $0_{\mathcal{V}} \in \mathcal{U}$.*

(ii) *Closure under vector addition: if v, $w \in \mathcal{U}$, then $v + w \in \mathcal{U}$.*

(iii) *Closure under scalar multiplication: if $\alpha \in \mathbb{F}$ and $v \in \mathcal{U}$, then $\alpha v \in \mathcal{U}$.*

4.4 Problem (!). Prove that if $\mathcal{V}$ is a vector space, then $\{0_{\mathcal{V}}\}$ is a subspace of $\mathcal{V}$.

If $\mathcal{U}$ is a subspace of $\mathcal{V}$, then $\mathcal{U}$ is also a vector space over $\mathbb{F}$ with the operations of addition and scalar multiplication restricted to $\mathcal{U}$. More technically, if $\mathcal{U}$ is a subspace of $(\mathcal{V}, \mathbb{F}, +_{\mathcal{V}}, \cdot_{\mathcal{V}})$, then $(\mathcal{U}, \mathbb{F}, +_{\mathcal{U}}, \cdot_{\mathcal{U}})$ is also a vector space, where $v +_{\mathcal{U}} w := v +_{\mathcal{V}} w$ and $\alpha \cdot_{\mathcal{U}} v := \alpha \cdot_{\mathcal{V}} v$ for v, $w \in \mathcal{U}$ and $\alpha \in \mathbb{F}$. The upshot is that verifying the subspace axioms automatically implies that $\mathcal{U}$ is a vector space in this way. In practice, because so many interesting vector spaces are subspaces of $\mathbb{F}^X$ for a well-chosen X, we can avoid a lot of boring work by inheriting the pointwise vector space structure of $\mathbb{F}^X$.

4.5 Problem (!). Explain why $\mathcal{U}$ is still a subspace of $\mathcal{V}$ if the first axiom is replaced by the condition that $\mathcal{U} \neq \varnothing$.

4.6 Example. Here are some simple situations in $\mathbb{R}^2$ to practice with the subspace axioms.

(i) The set

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

is a subspace of $\mathbb{R}^2$. First we show that $\mathbf{0}_2 = (0,0) \in \mathcal{U}$; this is true by taking $x = 0$. Next, suppose that $\mathbf{v}$, $\mathbf{w} \in \mathcal{U}$; we need to show that $\mathbf{v} + \mathbf{w} \in \mathcal{U}$. Since $\mathbf{v} = (v, 0)$ and $\mathbf{w} = (w, 0)$ for some v, $w \in \mathbb{R}$, we have $\mathbf{v} + \mathbf{w} = (v + w, 0) \in \mathcal{U}$. Finally, if $\alpha \in \mathbb{R}$ and $\mathbf{v} \in \mathcal{U}$, then $\alpha\mathbf{v} = (\alpha v, 0) \in \mathcal{U}$.

(ii) The set

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

is not a subspace of $\mathbb{R}^2$. We only need to break one of the axioms, but we show that many fail. First, $\mathbf{0}_2 \notin \mathcal{W}$ because $\mathbf{0}_2 = (0,0) \neq (x,1)$ for any $x \in \mathbb{R}$.

Next, 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:

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

Finally, 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:

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

Note, however, that $1\mathbf{v} \in \mathcal{W}$ for all $\mathbf{v} \in \mathcal{W}$.

4.7 Example. Calculus teaches us that sums and products of continuous functions are continuous and that constant functions are continuous. Thus for any interval $I \subseteq \mathbb{R}$ and f, $g \in \mathcal{C}(I)$, we have $f + g \in \mathcal{C}(I)$ and $\alpha f \in \mathcal{C}(I)$ for all $\alpha \in \mathbb{R}$; certainly $0 \in \mathcal{C}(I)$, too. And so $\mathcal{C}(I)$ is a subspace of $\mathbb{R}^I$. More generally, $\mathcal{C}^r(I)$ is a subspace of $\mathbb{R}^I$, too, by linearity of the derivative. Also, $\mathcal{C}^{r+1}(I)$ is a subspace of $\mathcal{C}^r(I)$ for all r. (It is true as well that if

$f,\ g \in \mathcal{C}^r(I)$, then $fg \in \mathcal{C}^r(I)$. That is, we can multiply two vectors in $\mathcal{C}^r(I)$ and obtain another vector in $\mathcal{C}^r(I)$. An abstract vector space does not necessarily have a "natural" notion of multiplying two vectors—a space with such "vector" multiplication is called an **ALGEBRA**, and we will see examples of that later.)

4.8 Problem ($\star$). Let $\mathcal{V}$ be a subspace of $\mathbb{R}^{\mathbb{R}}$ and let

$$\mathcal{U} := \{f \in \mathcal{V} \mid f(x) = f(-x) \text{ for all } x \in \mathbb{R}\}.$$

That is, $\mathcal{U}$ is the set of all **EVEN** functions in $\mathcal{V}$. Prove that $\mathcal{U}$ is a subspace of $\mathcal{V}$.

Much as we do not want to study *all* functions on a real interval in calculus, we also prefer sequences with nice behaviors. Here are some of them.

4.9 Example. (i) Denote by ℓ^∞ the set of all **BOUNDED** sequences in $\mathbb{F}$:

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

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 subspace of $\mathbb{F}^\infty$. The zero sequence (0) is certainly an element of ℓ^∞, since $|0| < 1$. (Remember that (0) is the map $\mathbb{N} \to \mathbb{F} \colon k \mapsto 0$.)

Next 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 the **TRIANGLE INEQUALITY**:

$$|\alpha + \beta| \le |\alpha| + |\beta|,\ \alpha,\ \beta \in \mathbb{F}.$$

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.

Last, we check scalar multiplication. Let $\alpha \in \mathbb{F}$ 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 $C > 0$ such that $|\alpha a_k| \le C$ for all k. Since $(a_k) \in \ell^\infty$, we know there is $M > 0$ such that $|a_k| \le M$ for all k. Since

$$|\alpha\beta| = |\alpha||\beta|,\ \alpha,\ \beta \in \mathbb{F},$$

we have $|\alpha a_k| = |\alpha||a_k| \le |\alpha| M$. Taking $C = |\alpha| M$ is the bound we want.

(ii) Denote by $\mathbb{F}_c^\infty$ the set of all convergent sequences in $\mathbb{F}$:

$$\mathbb{F}_c^\infty := \left\{(a_k) \in \mathbb{F}^\infty \;\middle|\; \lim_{k\to\infty} a_k \text{ exists}\right\}.$$

Then $0 \in \mathbb{F}_c^\infty$, since $\lim_{k\to\infty} 0 = 0$, and $\mathbb{F}_c^\infty$ is closed under addition and scalar multiplication because of "how limits work." For example, if (a_k) and (b_k) are convergent sequences with $\lim_{k\to\infty} a_k = L_1$ and $\lim_{k\to\infty} b_k = L_2$, then $(a_k + b_k)$ is convergent.

Content from *Linear Algebra* by Meckes & Meckes. I am presuming familiarity with the calculus of sequences and the modulus for complex numbers. Example 4 on p. 56 reviews limit arithmetic for sequences. You should be familiar with the properties of complex numbers on p. 382 of Appendix A.2. (Basically, $i^2 = -1$, and all of the arithmetic is going to work as you think it should.)

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

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

is a subspace of $\mathbb{F}^\infty$ (the notation c_0 is unfortunate, as it looks like a coefficient in some sum, but traditional).

(ii) Prove that

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

is not a subspace of $\mathbb{F}^\infty$ when $\alpha \neq 0$. Explain *all* of the ways in which $\mathcal{V}_\alpha$ fails to be a subspace.

Content from *Linear Algebra* by Meckes & Meckes. Pages 55–57 discuss subspaces. The book uses the notation $D^k[a,b]$ for what we would call $\mathcal{C}^k([a,b])$.

Here is a nice subspace that is effectively $\mathbb{F}^{n+1}$ in disguise but is also a function space.

4.11 Definition. *A* **POLYNOMIAL** *with coefficients in* $\mathbb{F}$ *is a function* $p \in \mathbb{F}^\mathbb{F}$ *of the form*

$$p(s) = \sum_{k=0}^{n} a_k s^k, \;\; s \in \mathbb{F},$$

for some $a_k \in \mathbb{F}$, $k = 0, \ldots, n$. *If* $a_n \neq 0$, *then the* **DEGREE** *of* p *is* $\deg(p) := n$, *and we define* $\deg(0) := 0$.

4.12 Example. Denote by $\mathbb{P}^n$ the set of all polynomials in $\mathbb{F}^\mathbb{R}$ of degree less than or equal to n. (This notation does not indicate whether the coefficients are in $\mathbb{R}$ or $\mathbb{C}$; usually we will not care.) Then $\mathbb{P}^n$ is a subspace of $\mathbb{F}^\mathbb{R}$. We can explain this quickly in words: the zero function is a polynomial of degree 0, adding polynomials of degree at most n results in a polynomial of degree at most n, and scaling a polynomial does not increase its degree.

In more symbols, consider the relatively simple case of $n = 2$. Then $0 = 0x^2 + 0x + 0$, so $0 \in \mathbb{P}^2$. (The first four appearances of 0 in that sentence were the scalar $0 \in \mathbb{F}$, while the fifth was the function $0 \in \mathbb{F}^\mathbb{R}$.) If p, $q \in \mathbb{P}^2$, write $p(x) = a_2x^2 + a_1x + a_0$ and $q(x) = b_2x^2 + b_1x + b_0$, so

$$(p+q)(x) = p(x) + q(x) = (a_2 + b_2)x^2 + (a_1 + b_1)x + (a_0 + b_0),$$

thus $p + q \in \mathbb{P}^2$. And if $\alpha \in \mathbb{F}$, then

$$(\alpha p)(x) = \alpha p(x) = \alpha(a_2x^2 + a_1x + a_0) = (\alpha a_2)x^2 + (\alpha a_1)x + (\alpha a_0),$$

thus $\alpha p \in \mathbb{P}^2$.

4.13 Problem (!). Why is $\mathcal{U} := \{p \in \mathbb{P}^2 \mid \deg(p) = 2\}$ not a subspace of $\mathbb{F}^{\mathbb{R}}$?

Content from *Linear Algebra* by Meckes & Meckes. This is a very different perspective on polynomials from the "formal polynomial" approach taken on p. 57.

4.14 Remark. *Many of our forthcoming examples will take place in one of the four spaces* $\mathbb{F}^n$, $\mathbb{P}^n$, $\mathbb{F}^\infty$, *or* $\mathcal{C}^r([0,1])$, *or some subspace of those four. We place three of these four spaces on a "continuum of complexity" of spaces related to* $\mathbb{F}^X$*:*

$\mathbb{F}^n$	$\mathbb{F}^\infty$	$\mathcal{C}^r([0,1])$
finite, discrete domain	*infinite, discrete domain*	*infinite, continuous domain*

$\longrightarrow$

Day 5: Friday, January 23.

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."

Linear combination of a list of vectors, span of a list of vectors, linear operator (N), identity operator on a vector space, zero operator between vector spaces

The spaces $\mathbb{F}^n$ and $\mathbb{P}^n$ can both be described efficiently by only a "few" vectors. Namely, if $\mathbf{e}_j \in \mathbb{F}^n$ is the vector whose jth component is 1 and whose components are otherwise 0, then any $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{F}^n$ has the form

$$\mathbf{v} = \sum_{j=1}^{n} v_j \mathbf{e}_j. \tag{5.1}$$

And if $p_j(x) := x^j$, then any $p \in \mathbb{P}^n$ has the form

$$p = \sum_{j=0}^{n} a_j p_j \tag{5.2}$$

for some $a_j \in \mathbb{F}$. This is our first encounter with *bases*, which give "unique coordinate systems" for vector spaces. We will study bases in detail later (and we point out that the basis given here for $\mathbb{F}^n$ is nicer than the one for $\mathbb{P}^n$ because the former is *orthonormal* and talks to the dot product very nicely—in particular, $v_j = \mathbf{v} \cdot \mathbf{e}_j$, and so we get an easy way of extracting the coefficients of $\mathbf{v} \in \mathbb{F}^n$ relative to this basis, unlike the a_j in that polynomial basis).

For now, we focus on the linear structure of (5.1) and (5.2).

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

(i) *A vector $v \in \mathcal{V}$ is a* **LINEAR COMBINATION** *of some vectors $v_1, \ldots, v_n \in \mathcal{V}$ if there exist $\alpha_j \in \mathbb{F}$ such that $v = \sum_{j=1}^n \alpha_j v_j$. If $n = 1$ and $v = \alpha_1 v_1$, then we say that v is a* **SCALAR MULTIPLE** *of v_1.*

(ii) *The set of all linear combinations of $v_1, \ldots, v_n \in \mathcal{V}$ is their* **SPAN***:*

$$\operatorname{span}(v_1, \ldots, v_n) := \left\{ \sum_{j=1}^n \alpha_j v_j \;\middle|\; \alpha_1, \ldots, \alpha_n \in \mathbb{F} \right\}.$$

We allow repetition among the v_j.

(iii) *If $\mathcal{B} \subseteq \mathcal{V}$ (here $\mathcal{B}$ need not be finite), the* **SPAN** *of $\mathcal{B}$ is the set of all linear combinations of vectors in $\mathcal{B}$ (which are necessarily finite sums):*

$$\operatorname{span}(\mathcal{B}) := \left\{ \sum_{j=1}^n \alpha_j v_j \;\middle|\; \alpha_1, \ldots, \alpha_n \in \mathbb{F},\ v_1, \ldots, v_n \in \mathcal{B},\ n \geq 1 \right\}.$$

We are writing $\operatorname{span}(\{v_1, \ldots, v_n\}) = \operatorname{span}(v_1, \ldots, v_n)$.

Content from *Linear Algebra* by Meckes & Meckes. Linear combinations and spans are defined on p. 53. We will not use the notation $\langle v_1, \ldots, v_n \rangle$ for the span of $v_1, \ldots, v_n$, as that too much resembles the (forthcoming) notation for an inner product.

5.2 Example. Every span is a subspace. We show this only for the case of $\mathcal{U} = \operatorname{span}(v_1, v_2)$, where v_1, $v_2 \in \mathcal{V}$ for some vector space $\mathcal{V}$ over $\mathbb{F}$.

First we want to show that $0 \in \mathcal{U}$. That is, we need to write $0 = \alpha_1 v_1 + \alpha_2 v_2$ for some α_1, $\alpha_2 \in \mathbb{F}$. We can do this by taking $\alpha_1 = \alpha_2 = 0$.

Next, suppose v, $w \in \mathcal{U}$. We want to show that $v + w \in \mathcal{U}$. We know that we can write $v = \alpha_1 v_1 + \alpha_2 v_2$ and $w = \beta_1 v_1 + \beta_2 v_2$ for some α_1, α_2, β_1, $\beta_2 \in \mathbb{F}$. Then

$$v + w = (\alpha_1 + \beta_1) v_1 + (\alpha_2 + \beta_2) v_2 \in \mathcal{U}.$$

Last, if $\alpha \in \mathbb{F}$ and $v \in \mathcal{U}$, then

$$\alpha v = \alpha(\alpha_1 v_1 + \alpha_2 v_2) = (\alpha\alpha_1) v_1 + (\alpha\alpha_2) v_2 \in \mathcal{U}.$$

This is about all that we need to say about vector spaces right now. (We will have lots more to say in the future.) Vector spaces are the worlds in which our problems live, but we pass between those worlds via linear operators: the special functions between vector spaces that "respect linearity."

5.3 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ (so both over $\mathbb{R}$ or both over $\mathbb{C}$). A* **LINEAR OPERATOR** *from $\mathcal{V}$ to $\mathcal{W}$ is a map $\mathcal{T} \in \mathcal{W}^{\mathcal{V}}$ such that*

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

for all v, $w \in \mathcal{V}$ and $\alpha \in \mathbb{F}$. If $\mathcal{V} = \mathcal{W}$, then we sometimes say that $\mathcal{T}$ is a linear operator **ON** *$\mathcal{V}$.*

5.4 Remark. **(i)** *When no confusion will result, we typically write $\mathcal{T}v := \mathcal{T}(v)$.*

(ii) *Synonyms for "linear operator" include "linear map" and "linear transformation." The latter is often used in a first course in linear algebra but rarely outside that. Sometimes "linear operator" is reserved for a linear map whose domain and codomain are the same (so $\mathcal{V} = \mathcal{W}$). While we will often be interested in that situation, we will use "linear operator" even when $\mathcal{V} \neq \mathcal{W}$.*

(iii) *Often (though not always) it will be obvious that a map $\mathcal{T} \in \mathcal{W}^{\mathcal{V}}$ between vector spaces is linear. What may be less obvious is that $\mathcal{T}$ does indeed map $\mathcal{V}$ to $\mathcal{W}$, as this will depend on the exact properties of these spaces. And even less obvious will be a precise characterization of the range of $\mathcal{T}$, which is largely the point of the course.*

(iv) *We always assume that $\mathcal{V}$ and $\mathcal{W}$ are vector spaces over the same field, so either $\mathbb{F} = \mathbb{R}$ in both cases or $\mathbb{F} = \mathbb{C}$ in both cases. It would be challenging to interpret $\mathcal{T}(\alpha v) = \alpha\mathcal{T}v$ if $\mathcal{V}$ is a vector space over $\mathbb{C}$ but $\mathcal{W}$ is a vector space over only $\mathbb{R}$.*

5.5 Example. Let $\mathcal{V}$ be a vector space over $\mathbb{F}$.

(i) Define

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

Then $\mathcal{I}_{\mathcal{V}}(v+w) = v + w = \mathcal{I}_{\mathcal{V}}v + \mathcal{I}_{\mathcal{V}}w$ and $\mathcal{I}_{\mathcal{V}}(\alpha v) = \alpha v = \alpha\mathcal{I}_{\mathcal{V}}v$ for all v, $w \in \mathcal{V}$ and $\alpha \in \mathbb{F}$, so $\mathcal{I}_{\mathcal{V}}$ is linear. We call $\mathcal{I}_{\mathcal{V}}$ the **IDENTITY OPERATOR** for $\mathcal{V}$.

(ii) Scalar multiplication gives a particularly simple kind of linear operator on any vector space. Fix $\lambda \in \mathbb{F}$ and define

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

The vector space axioms show that $\mathcal{T}$ is linear:

$$\mathcal{T}(v+w) = \lambda(v+w) = \lambda v + \lambda w = \mathcal{T}v + \mathcal{T}w \quad \text{and} \quad \mathcal{T}(\alpha v) = \lambda(\alpha v) = \alpha(\lambda v) = \alpha\mathcal{T}v.$$

Later we will define arithmetic for linear operators to see that $\mathcal{T} = \lambda\mathcal{I}_{\mathcal{V}}$.

5.6 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and let $\mathcal{T}: \mathcal{V} \to \mathcal{W}$ be a linear operator. Prove that $\mathcal{T}0_{\mathcal{V}} = 0_{\mathcal{W}}$. This gives an easy way to check that a map $\mathcal{T} \in \mathcal{W}^{\mathcal{V}}$ is not linear: show $\mathcal{T}0_{\mathcal{V}} \neq 0_{\mathcal{W}}$. [Hint: *try proving this in two ways. First, what is* $\mathcal{T}(0_{\mathcal{V}} + 0_{\mathcal{V}})$*? Next, what is* $\mathcal{T}(0v)$ *for any* $v \in \mathcal{V}$*?*]

5.7 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces.

(i) Prove that the zero map

$$0_{\mathcal{V}\to\mathcal{W}} : \mathcal{V} \to \mathcal{W} : v \mapsto 0_{\mathcal{W}}$$

from $\mathcal{V}$ to $\mathcal{W}$ is linear. We unsurprisingly call $0_{\mathcal{V}\to\mathcal{W}}$ the **ZERO OPERATOR** from $\mathcal{V}$ to $\mathcal{W}$.

(ii) Fix $w_0 \in \mathcal{W}$. Is the map $\mathcal{T}: \mathcal{V} \to \mathcal{W}: v \mapsto w_0$ ever linear?

5.8 Remark. *We will severely overwork the symbol* 0*. Given vector spaces* $\mathcal{V}$ *and* $\mathcal{W}$ *over the field* $\mathbb{F}$*, we might have* $0 \in \mathbb{F}$, $0 = 0_{\mathcal{V}} \in \mathcal{V}$, $0 = 0_{\mathcal{W}} \in \mathcal{W}$, *or* $0 = 0_{\mathcal{V}\to\mathcal{W}} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$*. At least in Euclidean space we will write* $\mathbf{0}_n \in \mathbb{F}^n$*.*

Content from *Linear Algebra* by Meckes & Meckes. Page 64 defines linear operators. Read in particular the paragraph below that definition and the "more substantial example" after that. You may find the geometric perspectives on pp. 65–66 helpful. More examples of linear operators appear on pp. 86–88. Theorem 2.7 on p. 82 extends the action of $\mathcal{T}$ to a finite sum, not just a sum of two vectors.

The following quotes perhaps illustrate an interesting historical evolution of the point of linear algebra. Hoffman and Kunze's classic *Linear Algebra* (1971) states that "Loosely speaking, linear algebra is that branch of mathematics which treats the common properties of algebraic systems which consist of a set, together with a reasonable notion of a 'linear combination' of elements in the set" (p. 28). Axler's groundbreaking *Linear Algebra Done Right* (2025), however, argues that "No one gets excited about vector spaces. The interesting part of linear algebra is the subject to which we now turn—linear maps" (p. 51). And the Meckeses unambiguously state at the start that "Our perspective is that mathematicians invented vector spaces so that they could talk about linear maps" (p. xiii). The latter two quotes indicate our priorities in this course: understanding the problem $\mathcal{T}v = w$, with vector spaces playing (major) auxiliary roles.

5.9 Example. For $f \in \mathcal{C}([0,1])$, define the new function $\mathcal{T}f$ pointwise by

$$(\mathcal{T}f)(x) := xf(x).$$

That is, with $m(x) := x$, we have put $\mathcal{T}f = mf$. Since $m, f \in \mathcal{C}([0,1])$ and the product of continuous functions is continuous, we have $mf \in \mathcal{C}([0,1])$. That is,

$$\mathcal{T}: \mathcal{C}([0,1]) \to \mathcal{C}([0,1]): f \mapsto mf$$

is a function. (So are m, f, and mf. But m, f, and mf have domain and codomain equal to $[0,1]$, whereas $\mathcal{T}$ has domain and codomain equal to $\mathcal{C}([0,1])$.)

Now we check that $\mathcal{T}$ is linear. We want to show $\mathcal{T}(f+g) = \mathcal{T}f + \mathcal{T}g$; that is, we want to show the equality of the functions $\mathcal{T}(f+g)$ and $\mathcal{T}f + \mathcal{T}g$. (Remember that f, g, $\mathcal{T}(f+g)$, $\mathcal{T}f$, $\mathcal{T}g$, and $\mathcal{T}f + \mathcal{T}g$ are all functions.) We prove this function equality by checking the pointwise equality

$$(\mathcal{T}(f+g))(x) = (\mathcal{T}f + \mathcal{T}g)(x)$$

in $\mathbb{R}$.

On the left, we have

$$(\mathcal{T}(f+g))(x) = x((f+g)(x)) = x(f(x)+g(x)) = xf(x) + m(x)g(x).$$

The first equality here is the definition of $\mathcal{T}$ applied to the function $f+g$, the second equality is the pointwise definition of the sum $f+g$, and the third equality is arithmetic in $\mathbb{R}$. On the right, we have

$$(\mathcal{T}f + \mathcal{T}g)(x) = (\mathcal{T}f)(x) + (\mathcal{T}g)(x) = xf(x) + xg(x).$$

The first equality is now the pointwise definition of the sum $\mathcal{T}f + \mathcal{T}g$, and the second equality is the definition of $\mathcal{T}$. Together, the left and the right are equal, so $\mathcal{T}(f+g) = \mathcal{T}f + \mathcal{T}g$.

Last, we want to show that $\mathcal{T}(\alpha f) = \alpha\mathcal{T}f$; that is, we want to show the equality of the functions $\mathcal{T}(\alpha f)$ and $\alpha\mathcal{T}f$. (Remember that f, αf, $\mathcal{T}(\alpha f)$, and $\alpha\mathcal{T}f$ are all functions.) We prove this function equality by checking the pointwise equality

$$(\mathcal{T}(\alpha f))(x) = (\alpha\mathcal{T}f)(x)$$

in $\mathbb{R}$.

On the left, we have

$$(\mathcal{T}(\alpha f))(x) = x((\alpha f)(x)) = x\alpha f(x) = \alpha(m(x)f(x)).$$

The first equality here is the definition of $\mathcal{T}$ applied to the function αf, the second equality is the pointwise definition of αf, and the third equality is arithmetic in $\mathbb{R}$.

That "multiply by m" is a linear operator is probably not surprising. The important thing to value in this argument is the parenthesis juggling: what does each and every object mean?

5.10 Example. Differentiation is inherently linear because limits are linear:

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

for any differentiable functions f and g and any $\alpha \in \mathbb{R}$. Here we point out that changing (co)domains changes linear operators, even if the "formula" for the operator does not change. The following are all linear.

(i) $\mathcal{T}_1 \colon \mathcal{C}^1([0,1]) \to \mathcal{C}([0,1]) \colon f \mapsto f'$. What is important here is that if $f \in \mathcal{C}^1([0,1])$, then f' is continuous.

(ii) $\mathcal{T}_2 \colon \mathbb{P}^2 \to \mathbb{P}^1 \colon f \mapsto f'$. What is important here is that "differentiation lowers the degree by 1."

(iii) $\mathcal{T}_3 \colon \mathbb{P}^2 \to \mathbb{P}^2 \colon f \mapsto f'$. What is important here is that $\mathbb{P}^1 \subseteq \mathbb{P}^2$.

Day 6: Monday, January 26.

No class.

Day 7: Wednesday, January 28.

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."

Linear functional on a vector space, multiplication operator induced by a matrix, eigenvalue of a linear operator or matrix, eigenvector of a linear operator or matrix

7.1 Example. Let $(a_k) \in \mathbb{F}^\infty$. Write, euphemistically, $(a_k) = (a_1, a_2, a_3, \ldots)$, so we can "see" its terms. Put $\mathcal{T}(a_k) := (0, a_1, a_2, a_3, \ldots)$. That is, if $f \in \mathbb{F}^\infty$, then

$$(\mathcal{T}f)(k) = \begin{cases} 0, & k = 1 \\ f(k-1), & k \geq 2. \end{cases}$$

Then $\mathcal{T}$ is a linear operator on $\mathbb{F}^\infty$:

$$\begin{aligned} \mathcal{T}((a_k) + (b_k)) = \mathcal{T}(a_k + b_k) &= (0, a_1 + b_1, a_2 + b_2, a_3 + b_3, \ldots) \\ &= (0, a_1, a_2, a_3) + (0, b_1, b_2, b_3, \ldots) = \mathcal{T}(a_k) + \mathcal{T}(b_k) \end{aligned}$$

and

$$\mathcal{T}(\alpha(a_k)) = \mathcal{T}(\alpha a_k) = (0, \alpha a_1, \alpha a_2, \alpha a_3, \ldots) = \alpha(0, a_1, a_2, a_3, \ldots) = \alpha\mathcal{T}(a_k).$$

7.2 Problem (!). Is $\mathcal{T}(a_k) := (\lambda, a_1, a_2, a_3, \ldots)$ linear when $\lambda \neq 0$?

Our examples of linear operators so far have mapped to what we probably think of as "actual" vector spaces (spaces with dimension at least 2). But the underlying field of scalars $\mathbb{F} = \mathbb{F}^1$ is still a vector space over $\mathbb{F}$.

7.3 Example. The map

$$\mathcal{T}: \mathcal{C}([0,1]) \to \mathbb{R}: f \mapsto \int_0^1 f(s)\, ds$$

is linear. This follows from the linearity of the definite integral.

(i) The "evaluate at 0" map

$$\mathcal{T}: \mathcal{C}([0,1]) \to \mathbb{R}: f \mapsto f(0)$$

is linear by properties of function arithmetic. Specifically,

$$\mathcal{T}(f+g) = (f+g)(0) = f(0) + g(0)$$

and

$$\mathcal{T}(\alpha f) = (\alpha f)(0) = \alpha f(0).$$

Linear operators that map to the field of scalars have a special name.

7.4 Definition. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. A* **linear functional** *on $\mathcal{V}$ is a linear operator from $\mathcal{V}$ to $\mathbb{F}$. Unlike linear operators, we usually denote linear functionals by lowercase Greek letters and denote pointwise evaluation with parentheses: if $\varphi: \mathcal{V} \to \mathbb{F}$ is linear, we write $\varphi(v)$, not φv.*

A linear functional is one of the simplest possible linear operators with domain $\mathcal{V}$, since its codomain is so "tame" as a vector space. We will see that linear functionals control and measure a great deal of information about vector spaces and linear operators; they are excellent instruments for extracting data about vectors and operators.

We have not yet introduced the most fundamental linear operator from a first course in linear algebra: matrix-vector multiplication. We motivate the definition of this multiplication by starting with a toy linear system and rewriting it in several ways. The first three equalities are just componentwise equalities from our definitions of arithmetic in $\mathbb{F}^2$: We have

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

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

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

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

Equality (1) is the componentwise definition of vector equality. Equality (2) is the componentwise definition of vector addition. Equality (3) is the componentwise definition of scalar multiplication. And equality (4) is how we choose to define matrix-vector multiplication.

7.5 Definition. *The* MATRIX-VECTOR PRODUCT *of*

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{F}^{m \times n} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{F}^n$$

is

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

That is, $A\mathbf{v}$ is the linear combination of the columns of A weighted by the entries of $\mathbf{v}$.

If we show that

$$A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w} \quad \text{and} \quad A(\alpha\mathbf{v}) = \alpha A\mathbf{v} \tag{7.1}$$

for all $\mathbf{v}, \mathbf{w} \in \mathbb{F}^n$ and $\alpha \in \mathbb{F}$, then multiplication by A is a linear operator.

7.6 Problem ($\star$). Show that (7.1) is true.

7.7 Theorem. *Let $A \in \mathbb{F}^{m \times n}$. Then the map*

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

is a linear operator, which we call the linear operator INDUCED *by A.*

7.8 Problem (!). Let $A \in \mathbb{F}^{m \times n}$. Explain how $A \neq \mathcal{M}_A$ as functions. In particular, comment on domains and codomains.

7.9 Remark. *Let $A \in \mathbb{R}^{m \times n}$. If $\mathbf{v} \in \mathbb{R}^n$, then $A\mathbf{v} \in \mathbb{R}^m \subseteq \mathbb{C}^m$, and so we could view $\mathcal{M}_A$ as a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^m$, or from $\mathbb{R}^n$ to $\mathbb{C}^m$. And if $\mathbf{w} \in \mathbb{C}^n$, then $A\mathbf{w} \in \mathbb{C}^m$, too, so $\mathcal{M}_A$ could also be interpreted as a linear operator from $\mathbb{C}^n$ to $\mathbb{C}^m$. Our notation in Theorem 7.7 does not indicate any of this; if it matters, context will make it clear.*

7.10 Problem ($\star$). The definition of matrix-vector multiplication from Definition 7.5 may not be the fastest way to compute matrix-vector products for "small" matrices and vectors by hand. Recall that the DOT PRODUCT of $\mathbf{v}, \mathbf{w} \in \mathbb{F}^n$ is $\mathbf{v} \cdot \mathbf{w} = \sum_{j=1}^n v_j w_j$. (Here we are *not* conjugating if $\mathbb{F} = \mathbb{C}$.) Use Definition 7.5 to prove that the ith entry of $A\mathbf{v}$ is the dot product of $\mathbf{v}$ with the ith row of A viewed as a vector in $\mathbb{F}^n$.

Content from *Linear Algebra* by Meckes & Meckes. Page 67 defines matrix-vector multiplication (using dot products). Some meaningful examples of products are on pp. 68–69. Pages 73–75 review how to compress a linear system as a matrix-vector equation.

7.11 Problem ($\star$). 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. (You do not have to prove this.) Find a matrix $A \in \mathbb{R}^{3\times 3}$ such that $\mathcal{T} = \mathcal{M}_A$. Describe in words the action of $\mathcal{M}_A$ on a vector $\mathbf{v}$. We will eventually show that any linear operator $\mathcal{T}\colon \mathbb{F}^n \to \mathbb{F}^m$ has the form $\mathcal{T} = \mathcal{M}_A$ for (a unique) $A \in \mathbb{F}^{m\times n}$.

We have now met the majority of the linear operators that will serve as examples in the rest of the course. As we prepare to tackle the overarching problem of solving, or at least understanding, the linear equation $\mathcal{T}v = w$, we might ask what are the "simplest" kinds of linear operators between vector spaces. Starting simple is always a good idea.

We saw some very simple operators in Example 5.5. The identity map $\mathcal{T}v = v$ is not that complicated, but it does require $\mathcal{V} = \mathcal{W}$, or at least $\mathcal{V} \subseteq \mathcal{W}$. The same requirement shows up with the scalar multiplication operator.

For this reason, we specialize to $\mathcal{V} = \mathcal{W}$ and take the perspective that the simplest linear operator on $\mathcal{V}$ is scalar multiplication. (If we get to choose the codomain, the simplest linear operator is probably a linear functional; here we are trying to keep the codomain as general as possible.) Many linear operators are certainly not scalar multiplication, but sometimes they act like scalar multiplication. When?

7.12 Definition. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{T}\colon \mathcal{V} \to \mathcal{V}$ be linear. A vector $v \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$ is an* EIGENVECTOR *of $\mathcal{T}$ corresponding to the* EIGENVALUE *$\lambda \in \mathbb{F}$ if*

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

So, an operator $\mathcal{T}$ acts like scalar multiplication by λ on all of its eigenvectors corresponding to λ. We exclude $0_{\mathcal{V}} \in \mathcal{V}$ as an eigenvector because then $\mathcal{T}0_{\mathcal{V}} = \lambda 0_{\mathcal{V}}$ for any $\lambda \in \mathbb{F}$; that is uselessly generous. However, $0 \in \mathbb{F}$ may well be an eigenvalue.

We will see that knowing the eigenvalues and eigenvectors of a linear operator affords us tremendous control over that operator. Indeed, this "eigenequation" $\mathcal{T}v = \lambda v$ is just a particular case of our fundamental problem $\mathcal{T}v = w$ (with the added specification that the domain and codomain of $\mathcal{T}$ be the same). Nonetheless, this equation really involves two unknowns, λ and v, and so in some sense it is more complicated—or at least solutions are often "less likely" to exist. Often if one is assured of the existence of the eigen*value*, then computing the eigen*vector* is a more feasible task, since it is then really a version of solving $\mathcal{T}v = w$, with $\mathcal{T}$ replaced by $\mathcal{T} - \lambda\mathcal{I}$ and $w = 0_{\mathcal{V}}$.

Content from *Linear Algebra* by Meckes & Meckes. Page 69 defines eigenvalues and eigenvectors for both operators and matrices. The remarks at the bottom of p. 69 and the examples on pp. 70–71 interpret eigenvalues and eigenvectors geometrically. The additional examples on pp. 72–73 compute eigenvalues for matrices without using determinants.

Because of the German origins of the words "eigenvalue" and "eigenvector" (see the footnote on p. 69), Tefethen and Bau's excellent *Numerical Linear Algebra* suggests (p. 180 of T&B) abbreviating "eigenvector" by "ev" and eigenvalue" by "ew." That book (pp. 181–182 of T&B) goes on to say

> "Eigenvalue problems have a very different character from the problems involving square or rectangular systems of linear equations...To ask about the eigenvalues of a [nonsquare matrix] A would be meaningless. Eigenvalue problems make sense only when the [matrix is square]. This reflects the fact that in applications, eigenvalues are generally used when a matrix is to be compounded iteratively...
>
> Broadly speaking, eigenvalues and eigenvectors are useful for two reasons, one algorithmic, the other physical. Algorithmically, eigenvalue analysis can simplify solutions of certain problems by reducing a coupled system to a collection of scalar problems. Physically, eigenvalue analysis can give insight into the behavior of evolving systems governed by linear equations. The most familiar examples in this latter class are the study of *resonance* (e.g., of musical instruments when struck or plucked or bowed) and of *stability* (e.g., of fluid flows subjected to small perturbations). In such cases eigenvalues tend to be particularly useful for analyzing behavior for large times t."

For now, we focus on computing some eigenvalues and eigenvectors (and in the process getting more practice with linear operators).

7.13 Example. Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad \text{and} \qquad \mathcal{M}_A \colon \mathbb{R}^2 \to \mathbb{R}^2 \colon \mathbf{v} \mapsto A\mathbf{v}.$$

Since $A\mathbf{e}_1 = \mathbf{e}_1 = 1\mathbf{e}_1$, with $\mathbf{e}_1 = (1, 0)$, the vector $\mathbf{e}_1$ is an eigenvector of $\mathcal{M}_A$ corresponding to the eigenvalue 1.

This is something that we just "saw" from the structure of A and $\mathcal{M}_A$ but that a first course in linear algebra would teach us to expect. Eventually we will develop some more systematic procedures for computing eigenvalues.

7.14 Problem ($\star$). Let A be as in the previous example.

(i) Show that 2 is an eigenvalue of $\mathcal{M}_A$ by finding an eigenvector $\mathbf{v}$ and checking $\mathcal{M}_A\mathbf{v} = 2\mathbf{v}$.

(ii) Let $\lambda \in \mathbb{R}$ be an eigenvalue of $\mathcal{M}_A$. Show that $\lambda = 1$ or $\lambda = 2$. *Do not use any facts about determinants. Try to do this "from scratch" using only the definition that* $\mathcal{M}_A\mathbf{v} = \lambda\mathbf{v}$ *for some* $\mathbf{v} \in \mathbb{R}^2 \setminus \{\mathbf{0}_2\}$.

Here is an eigenvalue example in which the choice of field matters.

7.15 Example. Let

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

(i) We consider (recall Remark 7.9) $\mathcal{M}_A\mathbf{v} = A\mathbf{v}$ as a linear operator on $\mathbb{R}^2$. We have $\mathcal{M}_A\mathbf{v} = \lambda\mathbf{v}$ 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 7.12 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 the operator $\mathcal{M}_A$ has no eigenvalues.

(ii) Now consider $\mathcal{M}_A$ as a linear operator on $\mathbb{C}^2$, with $\mathbb{C}^2$ 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 $\mathcal{M}_A\mathbf{v} = \lambda\mathbf{v}$ only if $\lambda = \pm i$, and now we are considering $\mathbb{C}^2$ as a vector space over $\mathbb{C}$. So, $\mathcal{M}_A$ does have eigenvalues now.

7.16 Problem (!). Find eigenvectors for the eigenvalues $\pm i$ in the previous example.

We will develop conditions that guarantee the existence of eigenvalues (namely, finite-dimensionality of the domain), and we will see then that taking the field to be $\mathbb{C}$ really is essential. When we think about matrix multiplication operators and eigenvalues, we will more or less always work over $\mathbb{C}$.

7.17 Definition. *Let $A \in \mathbb{C}^{n \times n}$. A vector $\mathbf{v} \in \mathbb{C}^n \setminus \{\mathbf{0}_n\}$ is an* EIGENVECTOR *of A corresponding to the* EIGENVALUE *$\lambda \in \mathbb{C}$ if $A\mathbf{v} = \lambda\mathbf{v}$.*

7.18 Problem (!). Let $A \in \mathbb{C}^{n \times n}$. Check that if $\mathbf{v} \in \mathbb{C}^n$ is an eigenvector of A with eigenvalue $\lambda \in \mathbb{C}$, then $\mathbf{v}$ is an eigenvector of $\mathcal{M}_A$ with eigenvalue λ.

7.19 Remark. *The eigenvalue problem is one of the most important times in our narrative when the choice of the field as $\mathbb{C}$, not $\mathbb{R}$, will very much matter. We always allow a matrix $A \in \mathbb{F}^{n \times n}$ to have (complex) eigenvalues, regardless of whether $A \in \mathbb{R}^{n \times n}$ or not, and we will prove that any matrix has at least one (complex) eigenvalue. However, the associated matrix multiplication operator $\mathcal{M}_A \colon \mathbb{F}^n \to \mathbb{F}^n$ may not have eigenvalues if $A \in \mathbb{R}^{n \times n}$ and we think of $\mathcal{M}_A$ as an operator on a real space.*

7.20 Problem (!). Show that a vector cannot be an eigenvector for two different eigenvalues. That is, let $\mathcal{V}$ be a vector space over $\mathbb{F}$, $\mathcal{T} \in \mathbf{L}(\mathcal{V})$, λ_1, $\lambda_2 \in \mathbb{F}$, and $v \in \mathcal{V} \setminus \{0\}$. If $\mathcal{T}v = \lambda_1 v$ and $\mathcal{T}v = \lambda_2 v$, explain why $\lambda_1 = \lambda_2$. This encodes the notion that the action of $\mathcal{T}$ on an eigenvector is to stretch or shrink that vector, and $\mathcal{T}$ should not stretch or shrink a vector in two different ways.

Day 8: Friday, January 30.

Here is an operator for which every (real) scalar is an eigenvalue.

8.1 Example. Let $\mathcal{V} = \mathcal{C}^\infty([0,1])$ and define

$$\mathcal{T} \colon \mathcal{V} \to \mathcal{V} \colon f \mapsto f'.$$

A function $f \in \mathcal{V} \setminus \{0\}$ is an eigenvector for $\mathcal{T}$ with eigenvalue $\lambda \in \mathbb{R}$ if $\mathcal{T}f = \lambda f$, equivalently, if $f' = \lambda f$. Pointwise, this means $f'(x) = \lambda f(x)$. Calculus tells us that all such functions have the form $f(x) = f(0)e^{\lambda x}$. Consequently, if we are given $\lambda \in \mathbb{R}$, the function $f(x) = e^{\lambda x}$ will be an eigenvector for λ. And so every scalar in $\mathbb{R}$ is an eigenvalue of $\mathcal{T}$.

8.2 Problem ($\star$). For $f \in \mathcal{C}([0,1])$, let

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

so $\mathcal{T}$ is a linear operator on $\mathcal{C}([0,1])$. Use the following to show that $\mathcal{T}$ has no eigenvalues.

(i) Suppose $\mathcal{T}f = 0$. Differentiate both sides. What does this tell you about f?

(ii) Suppose $\mathcal{T}f = \lambda f$ with $\lambda \neq 0$. Since $f = \lambda^{-1}\mathcal{T}f$, conclude that f is differentiable and that f satisfies the ODE $f' = \lambda^{-1}f$. Obtain $f(x) = f(0)e^{x/\lambda}$. Substitute this into $\mathcal{T}f = \lambda f$, evaluate the integral, and conclude $f(0) = 0$. What does this tell you about f?

Here is an operator that has no eigenvalues and for which changing the field would not help us claw back eigenvalues.

8.3 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 $f(x) \neq 0$ 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

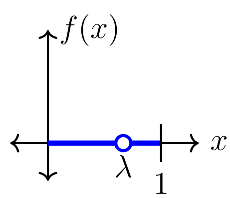
$$x - \lambda = 0 \quad \text{or} \quad f(x) = 0, \tag{8.1}$$

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 (8.1), 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 (8.1) 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.

8.4 Problem (!). Let $\mathcal{V} = \mathbb{C}^{[0,1]}$, consider $\mathcal{V}$ as a complex vector space, and define $\mathcal{T}: \mathcal{V} \to \mathcal{V}$ by $(\mathcal{T}f)(x) = xf(x)$. That is, $\mathcal{T}$ has the same action as in Example 8.3, but the functions in $\mathcal{V}$ need not be continuous. Prove that no $\lambda \in \mathbb{C} \setminus \mathbb{R}$ is an eigenvalue of $\mathcal{T}$.

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

8.5 Example. **(i)** Put

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

That is,

$$\mathcal{T}(a_1, a_2, a_3, \ldots) = (a_2, a_3, a_4, \ldots).$$

Then $\mathcal{T}$ is the "shift by 1" operator on $\mathbb{F}^\infty$. Indeed, $(\mathcal{T}f)(k) = f(k+1)$, and this formula should make it easy to check that $\mathcal{T}$ is linear.

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

$$a_{k+1} = \lambda a_k \tag{8.2}$$

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,

$$a_k = \lambda^{k-1} a_1 \tag{8.3}$$

for all k. We could prove this by induction on k from the relation (8.2).

This is the classical mathematical technique of working backwards: if $\mathcal{T}(a_k) = \lambda(a_k)$, then λ and (a_k) must satisfy (8.3). Does the logic go the other way? If λ and (a_k) satisfy (8.3), is λ an eigenvalue of $\mathcal{T}$ with eigenvector (a_k)?

We need to check two things. First, we compute

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

Next, if (a_k) meets (8.3), do we have $(a_k) \neq (0)$? First, we need $a_1 \neq 0$; this is nonnegotiable. Then if $\lambda \neq 0$, then $(\lambda^{k-1} a_1)$ is definitely not the zero sequence, so any

$\lambda \in \mathbb{R} \setminus \{0\}$ is an eigenvalue with eigenvector $(\lambda^{-1}a_1)$. 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 $(a_1, 0, 0, \ldots) = (0^{k-1}a_1)$ is not the zero sequence and still has the form (8.3), so it is an eigenvector for the eigenvalue 0.

We conclude that any $\lambda \in \mathbb{F}$ is an eigenvalue for $\mathcal{T}$.

(ii) Consider $\mathcal{T}$ now as an operator from ℓ^∞ to ℓ^∞, where ℓ^∞ was defined in Example 4.9. That $\mathcal{T}(a_k) \in \ell^\infty$ for any $(a_k) \in \ell^\infty$ is easy: if there is $M > 0$ such that $|a_k| \leq M$ for all k, then certainly $|a_{k+1}| \leq M$ 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 $a_1 \in \mathbb{F}$. That is still true here, as we have not changed the "formula" for T. As before, for (a_k) to be an eigenvector, we need $a_1 \neq 0$.

Do we have $(\lambda^{k-1}a_1) \in \ell^\infty$ for all λ, $a_1 \in \mathbb{F}$? If so, then there is $M > 0$ such that $|\lambda^{k-1}a_1| \leq M$ for all k, equivalently, $|\lambda|^k < M/|a_1|$ for all k, and so the sequence of powers $(|\lambda|^k)$ must be bounded. Conversely, if $(|\lambda|^k) \in \ell^\infty$, then $(\lambda^{k-1}a_1) \in \ell^\infty$.

It is a fact from calculus that the sequence $(|\lambda|^k)$ is bounded if and only if $|\lambda| \leq 1$; for $|\lambda| > 1$, we have $\lim_{k\to\infty} |\lambda|^k = \infty$. So, $(\lambda^{k-1}a_1) \in \ell^\infty$ precisely when $|\lambda| \leq 1$, and therefore the only eigenvalues of $\mathcal{T}$ are those scalars $\lambda \in \mathbb{F}$ with $|\lambda| \leq 1$. The lesson is that restricting the domain and codomain of $\mathcal{T}$ vastly changed the eigenvalue behavior.

Content from *Linear Algebra* by Meckes & Meckes. For another perspective on the mapping properties of this shift operator, see Example 4 on pp. 87–88.

8.6 Problem ($\star$). We saw in Example 7.1 that the map

$$\mathcal{T}\colon \mathbb{F}^\infty \to \mathbb{F}^\infty \colon (a_1, a_2, a_3, \ldots) \mapsto (0, a_1, a_2, a_3, \ldots)$$

is linear. We show here that $\mathcal{T}$ has no eigenvalues. Suppose that $\mathcal{T}(a_k) = \lambda(a_k)$ for some $(a_k) \in \mathbb{F}^\infty$ and $\lambda \in \mathbb{F}$.

(i) Show that

$$0 = \lambda a_1 \quad \text{and} \quad a_k = \lambda a_{k+1}, \ k \geq 1.$$

[Hint: *it suffices to establish the second equality by matching components for some small values of k, say, up to $k = 3$. A rigorous proof uses induction.*]

(ii) If $\lambda = 0$, explain why $(a_k) = (0)$, so 0 is not an eigenvalue.

(iii) If $\lambda \neq 0$, explain (in a slightly different way) why $(a_k) = (0)$, so λ is not an eigenvalue.

8.7 Problem (!). Let $\Lambda \in \mathbb{F}^{n \times n}$ be diagonal. Prove that the eigenvalues of Λ (equivalently, of $\mathcal{M}_\Lambda \in \mathbf{L}(\mathbb{F}^n)$) are the diagonal entries of Λ. (You need to do two things here: show that every diagonal entry of Λ is an eigenvalue, and show that every eigenvalue equals some diagonal entry of Λ.)

8.8 Problem (!). Find all of the eigenvalues and corresponding eigenvectors for $\mathcal{T} \colon \mathbb{P}^1 \to \mathbb{P}^1 \colon p \mapsto p'$.

Day 9: Monday, February 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."

Algebraic dual space, composition of linear operators

We now possess a (hopefully reasonable) command over the manipulation of particular, individual linear operators. Our work in this course is always in service to the operator equation $\mathcal{T}v = w$ for $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ linear and $w \in \mathcal{W}$ given. While only one operator appears in that equation, understanding that equation more deeply will result from understanding how operators interact with each other, not just individual vectors.

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. Problem 3.8 offered an opportunity to show that the set of all functions $\mathcal{W}^{\mathcal{V}}$ from $\mathcal{V}$ to $\mathcal{W}$ is a vector space over $\mathbb{F}$ via the expected pointwise operations. Linear operators form a subspace of this larger function space.

9.1 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. Let $\mathcal{T}, \mathcal{S} \colon \mathcal{V} \to \mathcal{W}$ be linear operators from $\mathcal{V}$ to $\mathcal{W}$ and let $\alpha \in \mathbb{F}$. Define*

$$\mathcal{T} + \mathcal{S} \colon \mathcal{V} \to \mathcal{W} \colon v \mapsto \mathcal{T}v + \mathcal{S}v \quad \text{and} \quad \alpha\mathcal{T} \colon \mathcal{V} \to \mathcal{W} \colon v \mapsto \alpha\mathcal{T}v.$$

Then $\mathcal{T} + \mathcal{S}$ and $\alpha\mathcal{T}$ are linear operators from $\mathcal{V}$ to $\mathcal{W}$.

Proof. What we need to prove here is not that $\mathcal{T} + \mathcal{S}$ and $\alpha\mathcal{T}$ are *functions* from $\mathcal{V}$ to $\mathcal{W}$ but rather that they are functions with the linear properties of a linear operator. We prove just one thing: that $(\mathcal{T}+\mathcal{S})(v_1+v_2) = (\mathcal{T}+\mathcal{S})v_1+(\mathcal{T}+\mathcal{S})v_2$. This is mostly just an exercise in juggling parentheses:

$$\begin{aligned}(\mathcal{T}+\mathcal{S})(v_1+v_2) &= \mathcal{T}(v_1+v_2)+\mathcal{S}(v_1+v_2) \text{ by definition of } \mathcal{T}+\mathcal{S} \\ &= \mathcal{T}v_1+\mathcal{T}v_2+\mathcal{S}v_1+\mathcal{S}v_2 \text{ by the linearity of } \mathcal{T} \text{ and } \mathcal{S} \\ &= (\mathcal{T}v_1+\mathcal{S}v_1)+(\mathcal{T}v_2+\mathcal{S}v_2) \text{ by commutativity of addition in } \mathcal{W} \\ &= (\mathcal{T}+\mathcal{S})v_1+(\mathcal{T}+\mathcal{S})v_2 \text{ by definition, again, of } \mathcal{T}+\mathcal{S}.\end{aligned} \tag{9.1}$$

The rest of the proof follows from similar manipulations. ■

9.2 Problem (!). Prove the rest of the theorem.

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 2.5 on pp. 81–82 (its proof is an exercise in the book).

These pointwise operations turn the set of all linear operators from $\mathcal{V}$ to $\mathcal{W}$ into a subspace of $\mathcal{W}^{\mathcal{V}}$, and so we now have a dual view of a linear operator. Sometimes it is a function that acts on vectors, and sometimes it is a vector itself.

9.3 Corollary. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. The set $\mathbf{L}(\mathcal{V}, \mathcal{W})$ of all linear operators from $\mathcal{V}$ to $\mathcal{W}$ is a subspace of $\mathcal{W}^{\mathcal{V}}$.*

Proof. Theorem 9.1 says that if $\mathcal{T}$, $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and $\alpha \in \mathbb{F}$, then $\mathcal{T} + \mathcal{S}$, $\alpha\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. The zero "vector" in $\mathcal{W}^{\mathcal{V}}$ is the map $0_{\mathcal{V}\to\mathcal{W}} \colon \mathcal{V} \to \mathcal{W} \colon v \mapsto 0_{\mathcal{W}}$, which is linear by Problem 5.7. ■

Depending on the codomain $\mathcal{W}$, there are two special cases of the operator space $\mathbf{L}(\mathcal{V}, \mathcal{W})$.

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

(i) $\mathbf{L}(\mathcal{V}) := \mathbf{L}(\mathcal{V}, \mathcal{V})$. *Recall that a linear operator in $\mathbf{L}(\mathcal{V})$ is sometimes called an operator* **ON** $\mathcal{V}$.

(ii) $\mathcal{V}' := \mathbf{L}(\mathcal{V}, \mathbb{F})$. *This is the* **(ALGEBRAIC) DUAL SPACE** *of $\mathcal{V}$. Recall that a linear operator in $\mathcal{V}'$ is usually called a* **LINEAR FUNCTIONAL** *on $\mathcal{V}$.*

We sometimes call $\mathcal{V}'$ the *algebraic* dual space to distinguish it from another meaningful (sub)space of linear functionals on certain vector spaces, which we will meet later. That forthcoming space will be denoted by $\mathcal{V}^*$, and we will not use $\mathcal{V}^*$ to refer to $\mathbf{L}(\mathcal{V}, \mathbb{F})$.

9.5 Problem ($\star$). Here is a third special case of the operator space. Let $\mathcal{W}$ be a vector space over $\mathbb{F}$ and let $\mathcal{T} \in \mathbf{L}(\mathbb{F}, \mathcal{W})$. Put $w_1 := \mathcal{T}1$. Show that $\mathcal{T}\alpha = \alpha w_1$ for all $\alpha \in \mathbb{F}$.

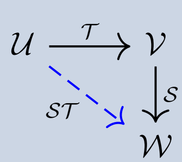
Content from *Linear Algebra* by Meckes & Meckes. Pages 81–82 discuss the vector space structure of $\mathbf{L}(\mathcal{V}, \mathcal{W})$.

We have now outlined how linear operators between two vector spaces (unsurprisingly) interact with each other. When three or more vector spaces are involved, there is another operator interaction.

9.6 Theorem. *Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. Given $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, define*

$$\mathcal{S}\mathcal{T} \colon \mathcal{U} \to \mathcal{W} \colon u \mapsto \mathcal{S}(\mathcal{T}u).$$

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



For $u \in \mathcal{U}$, we often write $\mathcal{S}\mathcal{T}u := (\mathcal{S}\mathcal{T})u$. For $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ and an integer $k \geq 0$, we put

$$\mathcal{T}^k := \begin{cases} \mathcal{I}_{\mathcal{V}}, \ k = 0 \\ \mathcal{T}, \ k = 1 \\ \mathcal{T}^{k-1}\mathcal{T}, \ k \geq 2. \end{cases}$$

Proof. The proof is, again, parenthesis juggling. We show only that $(\mathcal{S}\mathcal{T})(u_1 + u_2) = (\mathcal{S}\mathcal{T})u_1 + (\mathcal{S}\mathcal{T})u_2$. The parentheses so far emphasize that $\mathcal{S}\mathcal{T}$ is a single object, which maps from $\mathcal{U}$ to $\mathcal{W}$. We have

$$\begin{aligned} (\mathcal{S}\mathcal{T})(u_1 + u_2) &= \mathcal{S}(\mathcal{T}(u_1 + u_2)) \\ &= \mathcal{S}(\mathcal{T}u_1 + \mathcal{T}u_2) \\ &= \mathcal{S}(\mathcal{T}u_1) + \mathcal{S}(\mathcal{T}u_2) \\ &= (\mathcal{S}\mathcal{T})u_1 + (\mathcal{S}\mathcal{T})u_2. \end{aligned} \tag{9.2}$$

The rest of the proof follows by mostly similar calculations. ■

9.7 Problem (!). **(i)** Justify each equality in (9.2). Compare your justifications to the ones given for (9.1) in the proof of Theorem 9.1.

(ii) Finish the proof.

Content from *Linear Algebra* by Meckes & Meckes. Proposition 2.4 on p. 81 discusses operator composition (its proof is an exercise in the book).

When $\mathcal{V} = \mathcal{W}$ and $\mathcal{S}, \mathcal{T} \in \mathbf{L}(\mathcal{V})$, then both operator "products" $\mathcal{S}\mathcal{T}$ and $\mathcal{T}\mathcal{S}$ are defined and belong to $\mathbf{L}(\mathcal{V})$. However, we should not expect that they are equal, i.e., typically $\mathcal{S}\mathcal{T} \neq \mathcal{T}\mathcal{S}$, and so operator composition is not COMMUTATIVE.

9.8 Example. Let

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

Then A and B encode elementary row operations. While we have not yet rigorously defined the matrix product AB, we can still compute, compare, and contrast $\mathcal{M}_A\mathcal{M}_B$ and $\mathcal{M}_B\mathcal{M}_A$ (which will, of course, have the effect of multiplying by AB and BA, i.e., of doing $\mathcal{M}_{AB}$ and $\mathcal{M}_{BA}$).

For $\mathbf{v} = (v_1, v_2)$, we have

$$(\mathcal{M}_A\mathcal{M}_B)\mathbf{v} = \mathcal{M}_A(\mathcal{M}_B\mathbf{v}) = \mathcal{M}_A\left(\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}\right) = \mathcal{M}_A\begin{bmatrix} 3v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 3v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 \\ v_2 - 6v_1 \end{bmatrix}$$

but

$$(\mathcal{M}_B\mathcal{M}_A)\mathbf{v} = \mathcal{M}_B(\mathcal{M}_A\mathbf{v}) = \mathcal{M}_B\left(\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}\right) = \mathcal{M}_B\begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix}$$
$$= \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 - 2v_1 \end{bmatrix} = \begin{bmatrix} 3v_1 \\ v_2 - 2v_1 \end{bmatrix}.$$

So, we have $\mathcal{M}_A\mathcal{M}_B\mathbf{v} = \mathcal{M}_B\mathcal{M}_A\mathbf{v}$ if and only if $v_2 - 6v_1 = v_2 - 2v_1$, which happens precisely when $v_1 = 0$.

9.9 Example. Let $\mathcal{V} = \mathcal{C}^\infty([0,1])$.

(i) Let $\mathcal{T}f = f'$ and $(\mathcal{S}f)(x) = xf(x)$. Experience with the product rule should suggest that $\mathcal{T}$ and $\mathcal{S}$ will not commute. Note that for $f \in \mathcal{V}$,

$$\mathcal{S}\mathcal{T}f = (\mathcal{S}\mathcal{T})f = \mathcal{S}(\mathcal{T}f) \quad \text{and} \quad \mathcal{T}\mathcal{S}f = (\mathcal{T}\mathcal{S})f = \mathcal{T}(\mathcal{S}f)$$

are functions, while for $x \in \mathbb{R}$,

$$((\mathcal{S}\mathcal{T})f)(x) = (\mathcal{S}(\mathcal{T}f))(x) \quad \text{and} \quad ((\mathcal{T}\mathcal{S})f)(x) = (\mathcal{T}(\mathcal{S}f))(x)$$

are real numbers, and we are going to see if $((\mathcal{S}\mathcal{T})f)(x)$ and $((\mathcal{T}\mathcal{S})f)(x)$ are equal.

So, we compute

$$(\mathcal{S}(\mathcal{T}f))(x) = x(\mathcal{T}f)(x) = xf'(x)$$

and

$$(\mathcal{T}(\mathcal{S}f))(x) = (\mathcal{S}f)'(x) = f(x) + xf'(x).$$

Then $\mathcal{S}\mathcal{T}f = \mathcal{T}\mathcal{S}f$ only when $f = 0$, which should not be surprising (in the sense that $\mathcal{S}\mathcal{T}0_{\mathcal{V}} = \mathcal{T}\mathcal{S}0_{\mathcal{V}} = 0_{\mathcal{V}}$ for any operators $\mathcal{S}$, $\mathcal{T}$ on any vector space $\mathcal{V}$).

(ii) Calculus teaches us that differentiation and integration are "inverse" processes that "undo" each other. Are they? Let $\mathcal{T}f = f'$ and and $(\mathcal{S}f)(x) = \int_0^x f(s)\ ds$, so $\mathcal{T}$, $\mathcal{S} \in \mathbf{L}(\mathcal{V})$.

We compute (minding our parentheses carefully)

$$(\mathcal{S}\mathcal{T}f)(x) = (\mathcal{S}(\mathcal{T}f))(x) = \int_0^x (\mathcal{T}f)(s)\ ds = \int_0^x f'(s)\ ds = f(x) - f(0)$$

and

$$(\mathcal{T}\mathcal{S}f)(x) = (\mathcal{T}(\mathcal{S}f))(x) = (\mathcal{S}f)'(x) = f(x).$$

So, $(\mathcal{S}\mathcal{T})f = (\mathcal{T}\mathcal{S})f$ only when $f(0) = 0$. Differentiation and integration do not commute, at least without further restrictions on the functions involved, and the constant of integration really is important.

It is worth noting some notation common to both situations above: we have $\mathcal{T}f = f'$, so, regardless of what $\mathcal{S}$ was, we have $\mathcal{T}\mathcal{S}f = (\mathcal{S}f)'$. But $\mathcal{S}\mathcal{T}f = \mathcal{S}f'$. Parentheses make a big difference: $(\mathcal{S}f)' \neq \mathcal{S}f'$.

9.10 Problem (!). Here we use the notation for the zero operator from the proof of Corollary 9.3. Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{U},\mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V},\mathcal{W})$. What are $0_{\mathcal{V}\to\mathcal{W}}\mathcal{T}$ and $\mathcal{S}0_{\mathcal{U}\to\mathcal{V}}$?

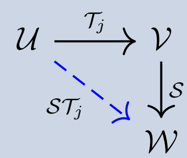
9.11 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V},\mathcal{W})$. What are $\mathcal{I}_{\mathcal{W}}\mathcal{T}$ and $\mathcal{T}\mathcal{I}_{\mathcal{V}}$?

Operator composition exhibits distributivity and associativity properties similar to those of vector arithmetic.

9.12 Theorem. *Let $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{X}$ be vector spaces over $\mathbb{F}$.*

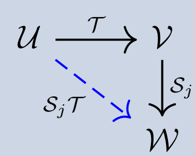
(i) *If $\mathcal{T}_1$, $\mathcal{T}_2 \in \mathbf{L}(\mathcal{U},\mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V},\mathcal{W})$, then*

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



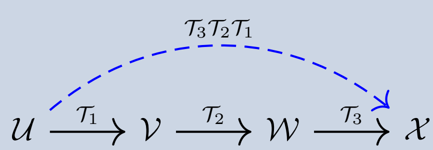
(ii) *If $\mathcal{T} \in \mathbf{L}(\mathcal{U},\mathcal{V})$ and $\mathcal{S}_1$, $\mathcal{S}_2 \in \mathbf{L}(\mathcal{V},\mathcal{W})$, then*

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



(iii) *If $\mathcal{T}_1 \in \mathbf{L}(\mathcal{U},\mathcal{V})$, $\mathcal{T}_2 \in \mathbf{L}(\mathcal{V},\mathcal{W})$, and $\mathcal{T}_3 \in \mathbf{L}(\mathcal{W},\mathcal{X})$, then*

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



and so we usually just write $\mathcal{T}_3\mathcal{T}_2\mathcal{T}_1$.

Proof. Once again, the proof is mostly parenthesis juggling and the correct definition of "equals." We prove only part of the first part: we will show

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

for all $u \in \mathcal{U}$. We have

$$\begin{aligned}(\mathcal{S}(\mathcal{T}_1 + \mathcal{T}_2))u &= \mathcal{S}((\mathcal{T}_1 + \mathcal{T}_2)u) \text{ by definition of } \mathcal{S}(\mathcal{T}_1 + \mathcal{T}_2)\\ &= \mathcal{S}(\mathcal{T}_1 u + \mathcal{T}_2 u) \text{ by definition of } \mathcal{T}_1 + \mathcal{T}_2\\ &= \mathcal{S}(\mathcal{T}_1 u) + \mathcal{S}(\mathcal{T}_2 u) \text{ by the linearity of } \mathcal{S}\\ &= (\mathcal{S}\mathcal{T}_1)u + (\mathcal{S}\mathcal{T}_2 u) \text{ by definition of } \mathcal{S}\mathcal{T}_j\end{aligned}$$

$$= (\mathcal{S}\mathcal{T}_1 + \mathcal{S}\mathcal{T}_2)u \text{ by definition of } \mathcal{S}\mathcal{T}_1 + \mathcal{S}\mathcal{T}_2.$$

The rest of the proof is just mostly similar calculations. ■

9.13 Problem ($\star$). **(i)** Prove the rest of this theorem.

(ii) Why do we really need both of the first two statements in the theorem, when they both appear to be saying the same thing? Articulate in words and as few symbols as possible why they are *not* saying the same thing.

Content from *Linear Algebra* by Meckes & Meckes. Pages 81–82 discuss operator composition. In particular, Theorem 2.6 on p. 82 contains the distributivity properties. Do Quick Exercise #8 on p. 82.

The operator space $\mathbf{L}(\mathcal{V})$ has slightly more structure than the more general space $\mathbf{L}(\mathcal{V}, \mathcal{W})$. In $\mathbf{L}(\mathcal{V})$, we can compose operators and obtain another operator in $\mathbf{L}(\mathcal{V})$, and operator composition interacts with operator addition and scalar multiplication in pretty much the ways that we expect. Such composition is not available in $\mathbf{L}(\mathcal{V}, \mathcal{W})$.

Here is the more general structure of $\mathbf{L}(\mathcal{V})$. We resume counting axioms from Definition 3.4.

9.14 Definition. *An* ALGEBRA *over* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *is a list of length 5 of the form* $(\mathcal{V}, \mathbb{F}, +, \cdot, \star)$, *where* $(\mathcal{V}, \mathbb{F}, +, \cdot)$ *is a vector space over* $\mathbb{F}$, *and the* VECTOR MULTIPLICATION *map* $\star\colon \mathcal{V} \times \mathcal{V} \to \mathcal{V}\colon (v, w) \mapsto v \star w$ *satisfies the following.*

Axiom for vector multiplication.

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

Axioms relating vector addition and multiplication.

11. *Right distribution:* $v \star (w + u) = (v \star w) + (v \star u)$ *for all* u, v, $w \in \mathcal{V}$.

12. *Left distribution:* $(v + w) \star u = (v \star u) + (w \star u)$ *for all* u, v, $w \in \mathcal{V}$.

Axiom relating scalar and vector multiplication.

13. *Distribution:* $\alpha(v \star w) = (\alpha v) \star w = v \star (\alpha w)$ *for all* $\alpha \in \mathbb{F}$ *and* v, $w \in \mathcal{V}$.

Of course, we would usually just call $\mathcal{V}$ the algebra.

9.15 Example. **(i)** If $\mathcal{V}$ is a vector space, then $\mathbf{L}(\mathcal{V})$ is an algebra. It is UNITAL because the identity operator $\mathcal{I}_{\mathcal{V}}$ satisfies $\mathcal{I}_{\mathcal{V}}\mathcal{T} = \mathcal{T}\mathcal{I}_{\mathcal{V}} = \mathcal{T}$ for any $\mathcal{T} \in \mathbf{L}(\mathcal{V})$.

(ii) The function space $\mathbb{F}^X$ is an algebra for any set X with multiplication given pointwise: $(fg)(x) := f(x)g(x)$. This is because of how multiplication works in $\mathbb{F}$. It is a COMMUTA-

TIVE algebra because $fg = gf$, since multiplication is commutative in $\mathbb{F}$. It is also unital because if $\mathbf{1}(x) := 1$, then $\mathbf{1}f = f$ for all $f \in \mathbb{F}^X$.

(iii) The space $\mathcal{C}([0,1])$ is an algebra with pointwise multiplication of functions, since the (pointwise) product of continuous functions is continuous.

9.16 Remark. **(i)** *An algebra is effectively a vector space in which we can multiply vectors and get another vector, and multiplication interacts with vector addition and scalar multiplication in pretty much the ways that we would expect. We do not assume that vector multiplication is commutative:* $v \star w \neq w \star v$ *in general. Indeed, in the prime example of an algebra,* $\mathbf{L}(\mathcal{V})$*, multiplication (= operator composition) is typically not commutative.*

(ii) *We do not assume that an algebra* $\mathcal{V}$ *is* **UNITAL**: *that there exists* $\mathbf{1} \in \mathcal{V}$ *such that* $v \star \mathbf{1} = \mathbf{1} \star v = v$.

(iii) *The triple* $(\mathcal{V}, +, \star)$ *is a ring.*

Day 10: Wednesday, February 4.

Vocabulary from today

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

Matrix representation of a linear operator in $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$, invertible linear operator (N), inverse of a linear operator, identity matrix

Operator composition on Euclidean space goes hand-in-hand with matrix multiplication. Let $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{n \times p}$. We expect that the matrix product AB is defined with $AB \in \mathbb{F}^{m \times p}$, and specifically we expect that if $B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_p \end{bmatrix}$, then $AB = \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_p \end{bmatrix}$. Here is how operator composition motivates this.

We have $\mathcal{M}_A \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathcal{M}_B \in \mathbf{L}(\mathbb{F}^p, \mathbb{F}^n)$, where $\mathcal{M}_A\mathbf{v} = A\mathbf{v}$ and $\mathcal{M}_B\mathbf{w} = B\mathbf{w}$. Then the composition $\mathcal{M}_A\mathcal{M}_B$ is defined with $\mathcal{M}_A\mathcal{M}_B \in \mathbf{L}(\mathbb{F}^p, \mathbb{F}^m)$. We want the product $AB \in \mathbb{F}^{m \times p}$ to satisfy $\mathcal{M}_{AB} = \mathcal{M}_A\mathcal{M}_B$. That is, for $\mathbf{v} \in \mathbb{F}^p$, we want

$$(AB)\mathbf{v} = \mathcal{M}_{AB}\mathbf{v} = (\mathcal{M}_A\mathcal{M}_B)\mathbf{v} = \mathcal{M}_A(\mathcal{M}_B\mathbf{v}) = \mathcal{M}_A(B\mathbf{v}) = A(B\mathbf{v}).$$

Since $B \in \mathbb{F}^{n \times p}$ and $\mathbf{v} \in \mathbb{F}^p$, we have $B\mathbf{v} \in \mathbb{F}^n$, and then since $A \in \mathbb{F}^{m \times n}$, we have $A(B\mathbf{v}) \in \mathbb{F}^m$.

So, all of this should seem reasonable, and the question is what the right way is to define AB so that

$$(AB)\mathbf{v} = A(B\mathbf{v}).$$

This is a good instance of the principle that *what things do defines what things are*. What a matrix *is* is an array of data, but what a matrix *does* is multiply vectors (and other matrices). And that multiplication defines what the matrix is.

10.1 Problem (!). Let $A \in \mathbb{F}^{m \times n}$ and let $\mathbf{e}_j \in \mathbb{F}^n$ be the jth standard basis vector for $\mathbb{F}^n$. That is, $\mathbf{e}_j$ is 1 in row j and 0 in all other rows. Show that $A\mathbf{e}_j$ is the jth column of A. That is, if $A = \begin{bmatrix} \mathbf{a}_1 & \cdots \mathbf{a}_n \end{bmatrix}$, then $A\mathbf{e}_j = \mathbf{a}_j$.

If we want $(AB)\mathbf{v} = A(B\mathbf{v})$ for all $\mathbf{v} \in \mathbb{F}^p$, then in particular this should hold at the jth standard basis vector $\mathbf{e}_j$. So, the jth column of AB should be

$$(AB)\mathbf{e}_j = A(B\mathbf{e}_j) = A\mathbf{b}_j,$$

as we expect.

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

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

10.3 Theorem. *Let $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{n \times p}$. Then $AB \in \mathbb{F}^{m \times p}$ and*

$$\mathcal{M}_{AB} = \mathcal{M}_A \mathcal{M}_B.$$

10.4 Problem (!). Everything preceding the statement of the theorem was working backwards (a pretty good way to work when you need to figure things out). Prove the theorem directly.

Content from *Linear Algebra* by Meckes & Meckes. Pages 91–96 discuss matrix multiplication from a variety of perspectives. Much of this should be familiar from a first course in linear algebra. Make sure that you can do Quick Exercises #11 (p. 92), #12 (p. 93), #13 (p. 95), and # 14 (p. 96) without hesitation. We will skip transposes for now.

10.5 Problem ($\star$). Give an example of vector spaces $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ and linear operators $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ such that $\mathcal{S}\mathcal{T} = 0_{\mathcal{U} \to \mathcal{W}}$ but $\mathcal{T} \neq 0_{\mathcal{U} \to \mathcal{V}}$ and $\mathcal{S} \neq 0_{\mathcal{V} \to \mathcal{W}}$. [Hint: *work with* 2×2 *matrices.*]

Every matrix in $\mathbb{F}^{m \times n}$ induces a linear operator $\mathcal{M}_A \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. The reverse turns out to be true. Let $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. Is there $A \in \mathbb{F}^{m \times n}$ such that $\mathcal{T} = \mathcal{M}_A$? That is, do we have $\mathcal{T}\mathbf{v} = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$? Is every operator from $\mathbb{F}^n$ to $\mathbb{F}^m$ really just matrix-vector multiplication?

If so, then in particular we have $\mathcal{T}\mathbf{e}_j = A\mathbf{e}_j = \mathbf{a}_j$. That is, $\mathcal{T}\mathbf{e}_j$ must be the jth column of A, and so the only choice for A is

$$A = \begin{bmatrix} \mathcal{T}\mathbf{e}_1 & \cdots & \mathcal{T}\mathbf{e}_n \end{bmatrix}.$$

10.6 Definition. *Let $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. The* MATRIX REPRESENTATION *of $\mathcal{T}$ with respect* *to the standard bases for $\mathbb{F}^n$ and $\mathbb{F}^m$ is the matrix*

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

10.7 Problem (!). Let $\mathcal{T}$ be the operator from Problem 7.11. How did that problem show how to find $[\mathcal{T}]$?

10.8 Problem (!). The $n \times n$ **IDENTITY MATRIX** is $I_n := [\mathcal{I}_{\mathbb{F}^n}]$. What is the jth column of I_n?

Content from *Linear Algebra* by Meckes & Meckes. Theorem 2.8 on p. 83 proves the existence of the matrix representation.

Now, does $[\mathcal{T}]$ do what we want? Do we have $\mathcal{T} = \mathcal{M}_{[\mathcal{T}]}$? Let $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{F}^n$, so $\mathbf{v} = \sum_{j=1}^n v_j \mathbf{e}_j$. Then

$$\mathcal{T}\mathbf{v} = \mathcal{T}\left(\sum_{j=1}^n v_j \mathbf{e}_j\right) = \sum_{j=1}^n v_j \mathcal{T}\mathbf{e}_j = \begin{bmatrix} \mathcal{T}\mathbf{e}_1 & \cdots & \mathcal{T}\mathbf{e}_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = [\mathcal{T}]\mathbf{v} = \mathcal{M}_{[\mathcal{T}]}\mathbf{v}. \tag{10.1}$$

So, yes, $\mathcal{T} = \mathcal{M}_{[\mathcal{T}]}$.

Now consider the maps

$$\mathcal{S}_1 \colon \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m) \to \mathbb{F}^{m \times n} \colon \mathcal{T} \mapsto [\mathcal{T}]$$

and

$$\mathcal{S}_2 \colon \mathbb{F}^{m \times n} \colon A \mapsto \mathcal{M}_A.$$

10.9 Problem ($\star$). Prove that $\mathcal{S}_1 \in \mathbf{L}(\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m), \mathbb{F}^{m \times n})$ and $\mathcal{S}_2 \in \mathbf{L}(\mathbb{F}^{m \times n}, \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m))$. Take a moment to marvel at our progress: we are working with linear operators whose domains or codomains are spaces of linear operators! [Hint: *for the linearity of $\mathcal{S}_1$, use the componentwise—or, now, maybe columnwise—definition of matrix addition from viewing $\mathbb{F}^{m \times n}$ as a function space. For the linearity of $\mathcal{S}_2$, show something like $\mathcal{M}_{A+B} = \mathcal{M}_A + \mathcal{M}_B$ by using what equality means here: the pointwise—or now, maybe, vectorwise—equality $\mathcal{M}_{A+B}\mathbf{v} = \mathcal{M}_A\mathbf{v} + \mathcal{M}_B\mathbf{v}$.*]

For $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$, we just showed

$$\mathcal{S}_2\mathcal{S}_1\mathcal{T} = \mathcal{S}_2[\mathcal{T}] = \mathcal{M}_{[\mathcal{T}]} = \mathcal{T}. \tag{10.2}$$

And for $A \in \mathbb{F}^{m \times n}$, we claim that

$$A = [\mathcal{M}_A].$$

10.10 Problem. Check this. [Hint: *compute $A\mathbf{e}_j$ and $[\mathcal{M}_A]\mathbf{e}_j$.*]

Thus

$$\mathcal{S}_1\mathcal{S}_2A = \mathcal{S}_1\mathcal{M}_A = [\mathcal{M}_A] = A. \tag{10.3}$$

The actions of the operators $\mathcal{S}_1$ and $\mathcal{S}_2$ appear to "undo" each other. This almost resembles the situation of differentiation and integration in part (ii) of Example 9.9, except the "undoing" did not quite work out because of the constant of integration.

10.11 Example. Let

$$\mathcal{V} := \{f \in \mathcal{C}^1([0,1]) \mid f(0) = 0\} \qquad \text{and} \qquad \mathcal{W} := \mathcal{C}([0,1]).$$

For $f \in \mathcal{V}$ and $g \in \mathcal{W}$, put $\mathcal{T}f = f'$ and $(\mathcal{S}g)(x) = \int_0^x g(s)\ ds$. Then $(\mathcal{T}\mathcal{S}g)(x) = g(x)$ by the fundamental theorem of calculus, as in part (ii) of Example 9.9, but now

$$(\mathcal{S}\mathcal{T}f)(x) = \int_0^x f'(s)\ ds = f(x) - f(0) = f(x),$$

since now $f(0) = 0$. So, $\mathcal{S}\mathcal{T}f = f$ and $\mathcal{T}\mathcal{S}g = g$ for all $f \in \mathcal{V}$ and $g \in \mathcal{W}$. Note that we do not say $\mathcal{S}\mathcal{T} = \mathcal{T}\mathcal{S}$, since the domain of $\mathcal{S}\mathcal{T}$ is $\mathcal{V}$ and the domain of $\mathcal{T}\mathcal{S}$ is $\mathcal{W} \neq \mathcal{V}$.

We are seeing a very special kind of operator behavior, and it is closely related to the existence and uniqueness of solutions of our fundamental problem $\mathcal{T}v = w$. Perhaps the ideal situation is that given $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and $w \in \mathcal{W}$, there is a unique $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. Suppose that there is $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that

$$\mathcal{S}\mathcal{T}v = v \text{ for all } v \in \mathcal{V} \qquad \text{and} \qquad \mathcal{T}\mathcal{S}w = w \text{ for all } w \in \mathcal{W}. \tag{10.4}$$

Then given $w \in \mathcal{W}$, we just take $v = \mathcal{S}w$ to solve $\mathcal{T}v = w$. And if $v_1, v_2 \in \mathcal{V}$ with $\mathcal{T}v_1 = \mathcal{T}v_2$, then $\mathcal{S}\mathcal{T}v_1 = \mathcal{S}\mathcal{T}v_2$, thus $v_1 = v_2$. Of course, we want to call $\mathcal{S}$ the inverse of $\mathcal{T}$ and write $\mathcal{S} = \mathcal{T}^{-1}$.

The immediate sticky point is the definite article "the." Is there only one $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ that satisfies (10.4)? Suppose that $\mathcal{S}_1, \mathcal{S}_2 \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ both meet (10.4). Then

$$\begin{aligned} \mathcal{S}_1 w &= \mathcal{S}_1(\mathcal{T}\mathcal{S}_2 w) \text{ because } \mathcal{T}\mathcal{S}_2 w = w\ \forall w \in \mathcal{W} \\ &= \mathcal{S}_1(\mathcal{T}(\mathcal{S}_2 w)) \text{ by definition of } \mathcal{T}\mathcal{S}_2 \\ &= (\mathcal{S}_1\mathcal{T})(\mathcal{S}_2 w) \text{ by associativity of operator composition} \\ &= \mathcal{S}_2 w \text{ because } \mathcal{S}\mathcal{T}_1 v = v\ \forall v \in \mathcal{V}. \end{aligned} \tag{10.5}$$

By the way, we did not use $\mathcal{T}\mathcal{S}_1 w = w$ or $\mathcal{S}_2\mathcal{T}v = v$ here.

10.12 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is* INVERTIBLE *if there is $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that*

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

This operator $\mathcal{S}$ is the INVERSE *of $\mathcal{T}$, and we write $\mathcal{T}^{-1} := \mathcal{S}$.*

10.13 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. Prove that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is invertible if and only if there is $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that

$$\mathcal{S}\mathcal{T} = \mathcal{I}_{\mathcal{V}} \quad \text{and} \quad \mathcal{T}\mathcal{S} = \mathcal{I}_{\mathcal{W}}.$$

[Hint: *this is just a repackaging of the definition.*]

10.14 Problem (⋆). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. Show that if $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is invertible, then so is $\mathcal{T}^{-1}$. What is $(\mathcal{T}^{-1})^{-1}$? [Hint: *show that the natural candidate for* $(\mathcal{T}^{-1})^{-1}$ *does what it should do, per the definition.*]

The work above did not use the linearity of $\mathcal{S}_1$ or $\mathcal{S}_2$ at all, just the associativity of composition. This argument works more generally to show that *function* inverses are unique when we are thinking of composition of functions between arbitrary sets. It turns out that if a linear operator has an inverse in the set-theoretic sense, then that inverse is unique.

More precisely, let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and let $f \in \mathcal{W}^{\mathcal{V}}$ such that $\mathcal{T}f(w) = w$ for all $w \in \mathcal{W}$ and $f(\mathcal{T}v) = v$ for all $v \in \mathcal{V}$. This notation should feel strange, since usually in this course we do not compose a linear operator with a function that is *not* a linear operator. Nonetheless, it turns out that $f \in \mathbf{L}(\mathcal{W}, \mathcal{V})$. Here is why. Let w_1, $w_2 \in \mathcal{W}$. We want to show that $f(w_1 + w_2) = f(w_1) + f(w_2)$. The trick is to rewrite

$$\begin{aligned} f(w_1 + w_2) &= f(\mathcal{T}f(w_1) + \mathcal{T}f(w_2)) \text{ because } \mathcal{T}f(w_1) = w_1,\ \mathcal{T}f(w_2) = w_2 \\ &= f(\mathcal{T}(f(w_1) + f(w_2))) \text{ as } \mathcal{T} \text{ is linear: } \mathcal{T}f(w_1) + \mathcal{T}f(w_2) = \mathcal{T}(f(w_1) + f(w_2)) \\ &= f(w_1) + f(w_2) \text{ since } f(\mathcal{T}v) = v. \end{aligned}$$

10.15 Problem (⋆). Adapt the work above to show $f(\alpha w) = \alpha f(w)$ for all $\alpha \in \mathbb{F}$ and $w \in \mathcal{W}$.

Here is what we conclude.

10.16 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. Suppose that $f \in \mathcal{W}^{\mathcal{V}}$ with $\mathcal{T}f(w) = w$ for all $w \in \mathcal{W}$ and $f(\mathcal{T}v) = v$ for all $v \in \mathcal{V}$. Then $f \in \mathbf{L}(\mathcal{W}, \mathcal{V})$. In particular, $\mathcal{T}$ is invertible and $f = \mathcal{T}^{-1}$.*

Day 11: Friday, February 6.

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."

Surjective (or onto) linear operator (N), injective (or one-to-one) linear operator (N), bijective linear operator (N), invertible matrix (N), inverse of an invertible matrix

We saw above that for $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, the existence of the inverse operator $\mathcal{T}^{-1}$ proved the unique solvability of our central problem $\mathcal{T}v = w$. The reverse is true: suppose that for all $w \in \mathcal{W}$, there is a unique $v \in \mathcal{V}$ such that $\mathcal{T}v = w$.

11.1 Problem (!). This may sound suspiciously like the definition of a function. It is not: discuss how the following two quantified statements are different.

(i) $\forall v \in \mathcal{V}\ \exists! w \in \mathcal{W} : (v, w) \in \mathcal{T}$.

(ii) $\forall w \in \mathcal{W}\ \exists! v \in \mathcal{V} : (v, w) \in \mathcal{T}$.

Can one be true and the other be false?

Content from *Linear Algebra* by Meckes & Meckes. Pages 380–382 of Appendix A.1 review function composition and inversion from a much more general perspective.

We claim that putting

$$\mathcal{S} := \{(w, v) \in \mathcal{W} \times \mathcal{V} \mid (v, w) \in \mathcal{T}\}$$

gives a linear operator in $\mathbf{L}(\mathcal{W}, \mathcal{V})$ such that $\mathcal{T}\mathcal{S}(w) = w$ for all $w \in \mathcal{W}$ and $\mathcal{S}(\mathcal{T}v) = v$ for all $v \in \mathcal{V}$. First we need to check that $\mathcal{S}$ is in fact a function in $\mathcal{V}^{\mathcal{W}}$: given $w \in \mathcal{W}$, is there a unique $v \in \mathcal{V}$ such that $(w, v) \in \mathcal{S}$? Certainly: let $v \in \mathcal{V}$ satisfy $\mathcal{T}v = w$, equivalently, $(v, w) \in \mathcal{T}$. So, given $w \in \mathcal{W}$, there is $v \in \mathcal{V}$ such that $(w, v) \in \mathcal{S}$. For uniqueness, if (w, v_1), $(w, v_2) \in \mathcal{S}$, then (v_1, w), $(v_2, w) \in \mathcal{T}$. So $\mathcal{T}v_1 = \mathcal{T}v_2$, and therefore by the unique solvability of $\mathcal{T}v = w$, we have $v_1 = v_2$.

Next, we check that $\mathcal{S}(\mathcal{T}v) = v$ and $\mathcal{T}\mathcal{S}(w) = w$ for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$. (Note our extra use of parentheses right now in applying $\mathcal{S}$: we have not yet proved that it is linear.) if $v \in \mathcal{V}$, we have $(v, \mathcal{T}v) \in \mathcal{T}$, so $(\mathcal{T}v, v) \in \mathcal{S}$. That is, $\mathcal{S}(\mathcal{T}v) = v$.

11.2 Problem (!). Use a similar argument to show that $\mathcal{T}(\mathcal{S}w) = w$ for all $w \in \mathcal{W}$.

One way to prove the linearity of $\mathcal{S}$ uses the following more general result about linear operators.

11.3 Problem (!). This problem and the next outline another way of proving the linearity of inverses. Let $\mathcal{X}$ and $\mathcal{Y}$ be vector spaces over $\mathbb{F}$. Prove that a map $\mathcal{T} \in \mathcal{Y}^{\mathcal{X}}$ is linear if and only if both of the following hold.

(i) If (x_1, y_1), $(x_2, y_2) \in \mathcal{T}$, then $(x_1 + x_2, y_1 + y_2) \in \mathcal{T}$.

(ii) If $\alpha \in \mathbb{F}$ and $(x, y) \in \mathcal{T}$, then $(\alpha x, \alpha y) \in \mathcal{T}$.

We start with (w_1, v_1), $(w_2, v_2) \in \mathcal{S}$. Then (v_1, w_1), $(v_2, w_2) \in \mathcal{T}$, and since $\mathcal{T}$ is linear, we have $(v_1 + v_2, w_1 + w_2) \in \mathcal{T}$. And so $(w_1 + w_2, v_1 + v_2) \in \mathcal{S}$.

11.4 Problem (!). Check that if $\alpha \in \mathbb{F}$ and $(w, v) \in \mathcal{S}$, then $(\alpha w, \alpha v) \in \mathcal{S}$.

We can wrap everything up in a neat little package.

11.5 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. The following are equivalent:*

(i) *For all $w \in \mathcal{W}$, there exists a unique $v \in \mathcal{V}$ such that $\mathcal{T}v = w$.*

(ii) *There exists $f \in \mathcal{V}^{\mathcal{W}}$ such that*

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

(iii) *There exists $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that*

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

The map f above is necessarily linear, and both f and $\mathcal{S}$ are unique.

11.6 Example. Let

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

We check if $\mathcal{M}_A$ is invertible: for $\mathbf{v} = (v_1, v_w)$, $\mathbf{w} = (w_1, w_2) \in \mathbb{F}$, we have

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

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

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

$$\iff \mathbf{v} = \begin{bmatrix} w_1 \\ w_2 + 2w_1 \end{bmatrix}$$

$$\iff \mathbf{v} = \mathcal{M}_B\mathbf{w}, \qquad B := \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

These logically equivalent statements say that given $\mathbf{w} \in \mathbb{F}$, if $\mathbf{v} = \mathcal{M}_B\mathbf{w}$, then $\mathcal{M}_A\mathbf{v} = \mathbf{w}$ (solutions exist), while if $\mathcal{M}_A\mathbf{v} = \mathbf{w}$ for some $\mathbf{v}$, then $\mathbf{v} = \mathcal{M}_B\mathbf{w}$ (solutions are unique). So, yes, $\mathcal{M}_A$ is invertible and $\mathcal{M}_A^{-1} = \mathcal{M}_B$.

This inverse should be what we should expect: $\mathcal{M}_A$ enacts the elementary row operation of subtracting 2 times row 1 from row 2, so $\mathcal{M}_A^{-1}$ should undo that by adding 2 times row 1 to row 2.

This example of course motivates the definition of invertible matrices and their inverses.

11.7 Definition. *A matrix $A \in \mathbb{F}^{n \times n}$ is* **INVERTIBLE** *if there exists $B \in \mathbb{F}^{n \times n}$ such that $AB = BA = I_n$. The matrix B is the* **INVERSE** *of A and we write $B = A^{-1}$.*

11.8 Problem ($\star$). Let $A \in \mathbb{F}^{n \times n}$.

(i) Prove that A is invertible if and only if $\mathcal{M}_A$ is invertible.

(ii) If A is invertible, show that its inverse is unique. [Hint: *just appeal to the result about $\mathcal{M}_A$.*]

(iii) Show that if A is invertible, then $\mathcal{M}_A^{-1} = \mathcal{M}_{A^{-1}}$.

It turns out that if $A \in \mathbb{F}^{n \times n}$, we only need to check one of the conditions $AB = I_n$ or $BA = I_n$ to conclude that A is invertible. This is a consequence of either some careful manipulations with elementary row operations and upper-triangular matrices, or some abstract arguments with dimension counting. Either way, it is nontrivial.

Content from *Linear Algebra* by Meckes & Meckes. Pages 97–100 review matrix inverses. All of this should be familiar from a first course in linear algebra. Do Quick Exercise #15 on p. 98.

It can be profitable to decouple the existence and uniqueness problems for $\mathcal{T}v = w$ and consider each separately.

11.9 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$.*

(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 $v_1, v_2 \in \mathcal{V}$, we have $v_1 = v_2$.*

(iii) *The operator $\mathcal{T}$ is* **BIJECTIVE** *it it is both injective and surjective.*

Bijectivity is equivalent to invertibility by Theorem 11.5. A common misconception is that injectivity is equivalent to the uniqueness condition in the definition of a function. It is not. Because any $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is a function in $\mathcal{W}^{\mathcal{V}}$, if (v, w_1), $(v, w_2) \in \mathcal{T}$, then $w_1 = w_2$. Injectivity, however, asks if (v_1, w), $(v_2, w) \in \mathcal{T}$ forces $v_1 = v_2$.

Content from *Linear Algebra* by Meckes & Meckes. Pages 380–382 of Appendix A.1 also review injectivity, surjectivity, and bijectivity for functions from a much more general perspective.

11.10 Example. Let $\mathcal{V} = \mathcal{W} = \mathbb{F}^\infty$.

(i) The operator $\mathcal{T}(a_1, a_2, a_3 \ldots) := (0, a_1, a_2, \ldots)$ is injective but not surjective. The failure of surjectivity comes from the first coordinate being set to 0: if $\mathcal{T}(a_k) = (b_k)$, then $b_1 = 0$. For example, there is no (a_k) such that $\mathcal{T}(a_k) = (1, 0, 0, \ldots)$.

For injectivity, if $\mathcal{T}(a_k) = \mathcal{T}(b_k)$, then $(0, a_1, a_2, \ldots) = (0, b_1, b_2, \ldots)$, so equating the kth coordinate for $k \geq 2$ gives $a_{k-1} = b_{k-1}$, thus $(a_k) = (b_k)$.

(ii) The operator $\mathcal{T}(a_1, a_2, a_3, \ldots) := (a_2, a_3, a_4, \ldots)$ is surjective but not injective. Given (b_k), we could put $(a_k) = (0, b_1, b_2, \ldots)$, or take any value for the first coordinate, really. Then $\mathcal{T}(a_k) = (b_1, b_2, b_3, \ldots) = (b_k)$.

The freedom of choice above, however, destroys injectivity. Since the first coordinate is irrelevant, we have things like $\mathcal{T}(1, 0, 0, \ldots) = \mathcal{T}(0, 0, 0, \ldots) = (0)$.

11.11 Problem ($\star$). **(i)** Prove that $\mathcal{T}\colon \mathcal{C}^1([0,1]) \to \mathcal{C}([0,1])\colon f \mapsto f'$ is surjective but not injective.

(ii) Prove that $(\mathcal{T}f)(x) := \int_0^x f(s)\, ds$ is injective but not surjective. [Hint: *If* $\int_0^x f(s)\, ds = 0$ *for all* x*, differentiate both sides.*]

11.12 Problem ($\star$). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$.

(i) An operator $\mathcal{L} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ is a **LEFT INVERSE** of $\mathcal{T}$ if $\mathcal{L}\mathcal{T}v = v$ for all $v \in \mathcal{V}$. Prove that if $\mathcal{T}$ has a left inverse, then $\mathcal{T}$ is injective.

(ii) An operator $\mathcal{R} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ is a **RIGHT INVERSE** of $\mathcal{T}$ if $\mathcal{T}\mathcal{R}w = w$ for all $w \in \mathcal{W}$. Prove that if $\mathcal{T}$ has a right inverse, then $\mathcal{T}$ is surjective.

(iii) Prove that if $\mathcal{T}$ has both a left and a right inverse, then $\mathcal{T}$ is invertible and $\mathcal{T}^{-1} = \mathcal{L} = \mathcal{R}$. [Hint: *for the latter, think about* (10.5).]

(iv) In proving Theorem 11.5, we established that the set-theoretic function inverse of a linear operator is itself linear. This need not be true for a left or right inverse. Let $\mathcal{V} = \mathcal{C}^1([0,1])$ and $\mathcal{W} = \mathcal{C}([0,1])$ and put $\mathcal{T}f = f'$ and $(\mathcal{G}f)(x) := \int_0^x f(s)\, ds + (f(0))^2$. Prove that $\mathcal{T}\mathcal{G}(f) = f$ for all $f \in \mathcal{V}$ but that $\mathcal{G}$ is not a linear operator.

We will later prove the reverse implications: injectivity (surjectivity) implies the existence of a left (right) inverse. This will require bases.

Day 12: Monday, February 9.

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."

Range of a linear operator, kernel of a linear operator, isomorphism (N), isomorphic vector spaces (N), eigenspace of a linear operator corresponding to an eigenvalue, null space of a matrix, column space of a matrix

Injectivity, surjectivity, and bijectivity are all properties of functions in general, not just linear operators. However, linearity helps us characterize these properties in ways that are not available outside of the vector space structure.

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. The RANGE of $\mathcal{T}$ is the same as for functions in general in Definition 2.6:

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

Surjectivity is just saying that $\mathcal{T}(\mathcal{V}) = \mathcal{W}$. What is new here is that the range inherits the linear structure of $\mathcal{V}$.

12.1 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Then $\mathcal{T}(\mathcal{V})$ is a subspace of $\mathcal{W}$.*

Proof. We check the three subspace axioms (Definition 4.3). First, let w_1, $w_2 \in \mathcal{T}(\mathcal{V})$. We want to show that $w_1 + w_2 \in \mathcal{T}(\mathcal{V})$, and so we need to find $v \in \mathcal{V}$ such that $\mathcal{T}v = w_1 + w_2$. By definition, there are v_1, $v_2 \in \mathcal{V}$ such that $\mathcal{T}v_1 = w_1$ and $\mathcal{T}v_2 = w_2$. By the linearity of $\mathcal{T}$,

$$w_1 + w_2 = \mathcal{T}v_1 + \mathcal{T}v_2 = \mathcal{T}(v_1 + v_2) \in \mathcal{T}(\mathcal{V}).$$

Next, let $\alpha \in \mathbb{F}$ and $w \in \mathcal{T}(\mathcal{V})$. We want to show that $\alpha w \in \mathcal{T}(\mathcal{V})$, and so we need to find $u \in \mathcal{V}$ such that $\mathcal{T}u = \alpha w$. By definition, there is $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. By the linearity of $\mathcal{T}$,

$$\alpha w = \alpha\mathcal{T}v = \mathcal{T}(\alpha v) \in \mathcal{T}(\mathcal{V}).$$

Finally, we want to show that $0_{\mathcal{W}} \in \mathcal{T}(\mathcal{V})$, and so we need to find $v \in \mathcal{V}$ such that $\mathcal{T}v = 0_{\mathcal{W}}$. We know $\mathcal{T}0_{\mathcal{V}} = 0_{\mathcal{W}}$ since $\mathcal{T}$ is a linear operator, so $0_{\mathcal{W}} = \mathcal{T}0_{\mathcal{V}} \in \mathcal{T}(\mathcal{V})$. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 2.30 on p. 115. Pages 114–118 discuss the range of a linear operator. You should be familiar with all of the results for matrices on these pages. Do Quick Exercises #20 (p. 115), #21 (p. 117), and #22 (p. 117).

12.2 Problem (!). Adapt the proof of Theorem 12.1 to show that if $\mathcal{U} \subseteq \mathcal{V}$ is a subspace of $\mathcal{V}$, then $\mathcal{T}(\mathcal{U}) = \{\mathcal{T}u \mid u \in \mathcal{U}\}$ is a subspace of $\mathcal{W}$. Give an example to show that if $\mathcal{U}$ is only a subset, and not a subspace, of $\mathcal{V}$, then $\mathcal{T}(\mathcal{U})$ need not be a subspace of $\mathcal{W}$.

12.3 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, and $v_1, \ldots, v_n \in \mathcal{V}$. Prove that $\mathcal{T}(\operatorname{span}(v_1, \ldots, v_n)) \subseteq \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$.

Injectivity interacts with linearity in the following way. We start by assuming $\mathcal{T}v_1 = \mathcal{T}v_2$, and we want to know if $v_1 = v_2$, equivalently, if $v_1 - v_2 = 0$. We have $\mathcal{T}v_1 = \mathcal{T}v_2$ if and only if $\mathcal{T}(v_1 - v_2) = 0$. If $v = 0$ whenever $\mathcal{T}v = 0$, then $\mathcal{T}$ will be injective, because then $v_1 - v_2 = 0$. Conversely, if $\mathcal{T}$ is injective, then the only solution to $\mathcal{T}v = 0$ is $v = 0$.

12.4 Definition. *The* KERNEL *of* $\mathcal{T}$ *is*

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

12.5 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$,*

(i) $\ker(\mathcal{T})$ *is a subspace of* $\mathcal{V}$.

(ii) $\mathcal{T}$ *is injective if and only if* $\ker(\mathcal{T}) = \{0_{\mathcal{V}}\}$.

Proof. **(i)** We check the three subspace axioms (Definition 4.3). First let v_1, $v_2 \in \ker(\mathcal{T})$. We want to show that $v_1 + v_2 \in \ker(\mathcal{T})$, and so we need to check that $\mathcal{T}(v_1 + v_2) = 0_{\mathcal{W}}$. By definition, $\mathcal{T}v_1 = \mathcal{T}v_2 = 0_{\mathcal{W}}$, and then the linearity of $\mathcal{T}$ implies

$$\mathcal{T}(v_1 + v_2) = \mathcal{T}v_1 + \mathcal{T}v_2 = 0_{\mathcal{W}} + 0_{\mathcal{W}} = 0_{\mathcal{W}}.$$

Next let $\alpha \in \mathbb{F}$ and $v \in \ker(\mathcal{T})$. We want to show that $\alpha v \in \ker(\mathcal{T})$, and so we need to check that $\mathcal{T}(\alpha v) = 0_{\mathcal{W}}$. By definition, $\mathcal{T}v = 0_{\mathcal{W}}$, and then the linearity of $\mathcal{T}$ implies

$$\mathcal{T}(\alpha v) = \alpha \mathcal{T}v = \alpha 0_{\mathcal{W}} = 0_{\mathcal{W}}.$$

Finally, we want to show that $0_{\mathcal{V}} \in \ker(\mathcal{T})$. That is, we need $\mathcal{T}0_{\mathcal{V}} = 0_{\mathcal{W}}$, and this is true by the linearity of $\mathcal{T}$.

(ii) This was proved in the paragraph before Definition 12.4. ■

12.6 Remark. *The range is an explicitly defined subspace: every vector in $\mathcal{T}(\mathcal{V})$ is given by the formula $\mathcal{T}v$ for some $v \in \mathcal{V}$. The kernel is an implicitly defined subspace: every vector in $\ker(\mathcal{T})$ solves the homogeneous probulem $\mathcal{T}v = 0_{\mathcal{W}}$, but this does not necessarily give us an explicit formula for such v. The proofs that the range and the kernel are subspaces should reflect this. The proof for the range repeatedly requires us to summon up a formula for something, while the proof for the kernel uses how the vectors involved solve the homogeneous problem.*

Content from *Linear Algebra* by Meckes & Meckes. This is Theorems 2.36 on p. 119 and 2.37 on p. 120.

If $\ker(\mathcal{T}) \neq \{0_{\mathcal{V}}\}$, then solutions to our fundamental problem $\mathcal{T}v = w$, if they exist, cannot be unique. Indeed, suppose that $\mathcal{T}v = w$ and $z \in \ker(\mathcal{T})$ with $z \neq 0_{\mathcal{V}}$. Then $\mathcal{T}(v + \alpha z) = w$ for all $\alpha \in \mathbb{F}$, and since $z \neq 0_{\mathcal{V}}$, when $\alpha_1 \neq \alpha_2$, we have $v + \alpha_1 z \neq v + \alpha_2 z$, thus every α gives a different solution.

12.7 Problem (!). Convince yourself that every part of the last sentence above is true. Then explain why if the problem $\mathcal{T}v = w$ has two different solutions, it has infinitely many solutions.

The tools of bases and dimension will help us quantify more precisely what happens when the problem $\mathcal{T}v = w$ has infinitely many solutions.

12.8 Problem ($\star$). Let $\mathcal{V}$ be a vector space over $\mathbb{F}$.

(i) Let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ and $\lambda \in \mathbb{F}$. Put

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

Prove that $\mathcal{E}_\lambda(\mathcal{T})$ is a subspace of $\mathcal{V}$, which we call the **EIGENSPACE** of $\mathcal{T}$ corresponding to λ.

(ii) Let $\mathcal{T}, \mathcal{S} \in \mathbf{L}(\mathcal{V})$ and let

$$\mathcal{U} := \{v \in \mathcal{V} \mid \mathcal{S}\mathcal{T}v = \mathcal{T}\mathcal{S}v\}.$$

Prove that $\mathcal{U}$ is a subspace of $\mathcal{V}$.

[Hint: *view both sets as kernels.*]

Content from *Linear Algebra* by Meckes & Meckes. Pages 120–122 discuss eigenspaces. Do Quick Exercise #24 on p. 122.

12.9 Problem (!). Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$.

(i) Prove that $\mathcal{T}$ is injective if and only if 0 is not an eigenvalue of $\mathcal{T}$. (Later we will consider what conditions on the eigenvalues could guarantee surjectivity.)

(ii) Prove that $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathcal{T}$ if and only if $\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}}$ is not injective.

12.10 Problem ($\star$). **(i)** Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ be invertible. Prove that $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathcal{T}$ if and only if λ^{-1} is an eigenvalue of $\mathcal{T}^{-1}$. (By

Problem 12.9, we know $\lambda \neq 0$, so λ^{-1} makes sense here.) How is an eigenvector for λ as an eigenvalue of $\mathcal{T}$ related to an eigenvector for λ^{-1} as an eigenvalue of $\mathcal{T}^{-1}$?

(ii) Let $\mathcal{V} = \mathcal{C}^\infty([0,1])$ and $\mathcal{W} = \{f \in \mathcal{V}) \mid f(0) = 0\}$. Example 8.1 showed that every $\lambda \in \mathbb{R}$ is an eigenvalue of the differentiation operator on $\mathcal{V}$. Problem 8.2 showed that the antidifferentiation operator $(\mathcal{S}f)(x) := \int_0^x f(s)\, ds$ has no eigenvalues as an operator on $\mathcal{V}$ (since $\mathcal{V}$ is a subspace of $\mathcal{C}([0,1])$, and $\mathcal{S}$ has no eigenvalues as an operator on that larger space). And Example 10.11 showed that $\mathcal{T}$ and $\mathcal{S}$ are each other's inverses on $\mathcal{W}$. Is there any contradiction here? If $\mathcal{T}$ has eigenvalues and $\mathcal{S} = \mathcal{T}^{-1}$, why does this not mean that $\mathcal{S}$ has eigenvalues? [Hint: *think about domains.*]

12.11 Problem ($\star$). Recall from Example 8.3 that the operator $(\mathcal{T}f)(x) := xf(x)$ on $\mathcal{C}([0,1])$ has no eigenvalues, and so $\mathcal{T} - \lambda\mathcal{I}$ is injective by Problem 12.9. Show that if $0 \leq \lambda \leq 1$, then $\mathcal{T} - \lambda\mathcal{I}$ is not surjective. [Hint: *if $(\mathcal{T} - \lambda\mathcal{I})f = g$, what is $g(\lambda)$?*] This suggests a generalization of eigenvalue: for an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$, a scalar $\lambda \in \mathbb{F}$ is a SPECTRAL VALUE of $\mathcal{T}$ if $\mathcal{T} - \lambda\mathcal{I}$ is not invertible. Every eigenvalue is a spectral value.

12.12 Problem ($\star$). Let $\mathcal{V}$ be a vector space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. A subspace $\mathcal{U}$ of $\mathcal{V}$ is INVARIANT UNDER $\mathcal{T}$ if $\mathcal{T}(\mathcal{U}) \subseteq \mathcal{U}$. Let $\lambda \in \mathbb{F}$. Prove that the kernel and range of $\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}}$ are invariant under $\mathcal{T}$.

There is some special terminology for matrix multiplication operators that is worth remembering.

12.13 Remark. *Let $A \in \mathbb{F}^{m\times n}$ The kernel of the matrix multiplication operator $\mathcal{M}_A$ is often called the* NULL SPACE *of A:*

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

The RANGE *of $\mathcal{M}_A$ is often called the* COLUMN SPACE *of A:*

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

Content from *Linear Algebra* by Meckes & Meckes. Pages 118–122 discuss kernels. You should be familiar with the results for matrices. Do Quick Exercise #23 on p. 119.

We previously observed that linear operators are the "natural" functions between vector spaces to study because they "respect" the linear structure of those spaces. (Since linear operators arise naturally in many problems, one might say that vector spaces are the natural domains for linear operators because they "respect" the linear structure of those operators!) When a linear operator $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$ is invertible, or bijective, then it does more than "respect" the linear structure of $\mathcal{V}$ and $\mathcal{W}$: it "preserves" the behavior of $\mathcal{V}$ in $\mathcal{W}$ and the behavior of $\mathcal{W}$ in $\mathcal{V}$. Under the lens of an invertible linear operator $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$, the spaces $\mathcal{V}$ and $\mathcal{W}$ are "the same."

12.14 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. The spaces $\mathcal{V}$ and $\mathcal{W}$ are* **ISOMORPHIC** *if there exists an invertible $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, and such an invertible $\mathcal{T}$ is an* **ISOMORPHISM**.

12.15 Example. **(i)** $\mathbb{P}^n$ and $\mathbb{F}^{n+1}$ are isomorphic. We show this just for $n = 1$. Recall that $p \in \mathbb{P}^1$ is a function of the form $p(x) = a_1 x + a_0$ for a_1, $a_0 \in \mathbb{F}$. This suggests associating p with $(a_1, a_0) \in \mathbb{F}^2$. If we want to be precise, we could note that $a_0 = p(0)$ and $a_1 = p'(0)$; these are really Taylor coefficients. That is, if $p \in \mathbb{P}^1$, then

$$p(x) = p'(0)x + p(0).$$

And so we define

$$\mathcal{T} : \mathbb{P}^1 \to \mathbb{F}^2 : p \mapsto (p'(0), p(0)).$$

Linearity follows from linearity of the derivative and pointwise evaluation of functions. For injectivity, if $\mathcal{T}p = 0$, then $p'(0) = p(0) = 0$, so $p(x) = 0x + 0 = 0$. Thus $p = 0$. For surjectivity, let $(a_1, a_0) \in \mathbb{F}^2$ and put $p(x) = a_1 x + a_0$, so $p(0) = a_0$ and $p'(0) = a_1$. Then $\mathcal{T}p = (a_1, a_0)$.

(ii) $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathbb{F}^{m \times n}$ are isomorphic. The crux of this is the calculations in (10.2) and (10.3), which we review here. Define

$$\mathcal{S} : \mathbb{F}^{m \times n} \to \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m) : A \mapsto \mathcal{M}_A.$$

To check linearity, we want to show that $\mathcal{S}(A + B) = \mathcal{S}A + \mathcal{S}B$, equivalently, $\mathcal{M}_{A+B} = \mathcal{M}_A + \mathcal{M}_B$. This is an equality of functions, so we want $\mathcal{M}_{A+B}\mathbf{v} = (\mathcal{M}_A + \mathcal{M}_B)\mathbf{v}$ for $\mathbf{v} \in \mathbb{F}^n$. Equivalently, we want $(A + B)\mathbf{v} = \mathcal{M}_A\mathbf{v} + \mathcal{M}_B\mathbf{v}$; in turn, $A\mathbf{v} + B\mathbf{v} = A\mathbf{v} + B\mathbf{v}$, which is of course true. That $\mathcal{S}(\alpha A) = \alpha\mathcal{S}A$ is proved similarly.

For injectivity, suppose $\mathcal{S}A = 0_{\mathbb{F}^n \to \mathbb{F}^m}$. Then $\mathcal{M}_A\mathbf{v} = \mathbf{0}_m$ for all $\mathbf{v} \in \mathbb{F}^n$. In particular, we have $\mathbf{0}_m = \mathcal{M}_A\mathbf{e}_j = A\mathbf{e}_j$, and so each column of A is the zero vector; thus A is the zero matrix.

For surjectivity, let $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. Put $[\mathcal{T}] = \begin{bmatrix} \mathcal{T}\mathbf{e}_1 & \cdots & \mathcal{T}\mathbf{e}_n \end{bmatrix}$. The calculation in (10.1) shows $\mathcal{T} = \mathcal{M}_{[\mathcal{T}]} = \mathcal{S}[\mathcal{T}]$.

Content from *Linear Algebra* by Meckes & Meckes. Pages 78–80 discuss isomorphisms and operator inverses. Do Quick Exercise #7 on p. 78. Theorem 2.9 on pp. 85–86 proves the isomorphism of $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathbb{F}^{m \times n}$.

12.16 Problem ($\star$). Prove that $\{f \in \mathcal{C}^1([0,1]) \mid f(0) = 0\}$ and $\mathcal{C}([0,1])$ are isomorphic. [Hint: *feel free to cite some prior results.*]

Day 13: Wednesday, February 11.

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."

Finite-dimensional vector space, infinite-dimensional vector space

Here is an exercise in diagram-chasing.

13.1 Theorem. *Let $\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{W}_1$, $\mathcal{W}_2$ be vector spaces over $\mathbb{F}$ such that V_1 and V_2 are isomorphic and $\mathcal{W}_1$ and $\mathcal{W}_2$ are isomorphic. Then $\mathbf{L}(\mathcal{V}_1, \mathcal{W}_1)$ and $\mathbf{L}(\mathcal{V}_2, \mathcal{W}_2)$ are isomorphic.*

Proof. Say that $\mathcal{T} \in \mathbf{L}(\mathcal{V}_1, \mathcal{V}_2)$ and $\mathcal{S} \in \mathbf{L}(\mathcal{W}_1, \mathcal{W}_2)$ are isomorphisms. We want to construct a bijection from $\mathbf{L}(\mathcal{V}_1, \mathcal{W}_1)$ to $\mathbf{L}(\mathcal{V}_2, \mathcal{W}_2)$. One way to do this is to start with an operator $\mathcal{A} \in \mathbf{L}(\mathcal{V}_1, \mathcal{W}_1)$ and try to associate it in a "natural" way with an operator in $\mathbf{L}(\mathcal{V}_2, \mathcal{W}_2)$. Such an association would probably involve $\mathcal{T}$ and $\mathcal{S}$, and so we draw the following picture.

$$\begin{array}{ccc} \mathcal{V}_1 & \xrightarrow{\mathcal{A}} & \mathcal{W}_1 \\ {\scriptstyle \mathcal{T}}\downarrow & & \downarrow{\scriptstyle \mathcal{S}} \\ \mathcal{V}_2 & \underset{?}{\dashrightarrow} & \mathcal{W}_2 \end{array}$$

If we just reverse $\mathcal{T}$, then we will have an operator from $\mathcal{V}_2$ to $\mathcal{W}_2$.

$$\begin{array}{ccc} \mathcal{V}_1 & \xrightarrow{\mathcal{A}} & \mathcal{W}_1 \\ {\scriptstyle \mathcal{T}^{-1}}\uparrow & & \downarrow{\scriptstyle \mathcal{S}} \\ \mathcal{V}_2 & \underset{\mathcal{S}\mathcal{A}\mathcal{T}^{-1}}{\dashrightarrow} & \mathcal{W}_2 \end{array}$$

So, we are going to map $\mathcal{A}$ to $\mathcal{S}\mathcal{A}\mathcal{T}^{-1}$. This composition is indeed defined and in $\mathbf{L}(\mathcal{V}_2, \mathcal{W}_2)$, since $\mathcal{T}^{-1} \in \mathbf{L}(\mathcal{V}_2, \mathcal{V}_1)$, $\mathcal{A} \in \mathbf{L}(\mathcal{V}_1, \mathcal{W}_1)$, and $\mathcal{S} \in \mathbf{L}(\mathcal{W}_1, \mathcal{W}_2)$.

We are running a bit short on letters, so we put

$$\mathcal{L}\colon \mathbf{L}(\mathcal{V}_1, \mathcal{W}_1) \to \mathbf{L}(\mathcal{V}_2, \mathcal{W}_2)\colon \mathcal{A} \mapsto \mathcal{S}\mathcal{A}\mathcal{T}^{-1}$$

We need to check that $\mathcal{L}$ is linear and bijective. For linearity, we compute

$$\mathcal{L}(\mathcal{A}_1 + \mathcal{A}_2) = \mathcal{S}(\mathcal{A}_1 + \mathcal{A}_2)\mathcal{T}^{-1} = \mathcal{S}\mathcal{A}_1\mathcal{T}^{-1} + \mathcal{S}\mathcal{A}_2\mathcal{T}^{-1} = \mathcal{L}\mathcal{A}_1 + \mathcal{L}\mathcal{A}_2.$$

This is both kinds of distribution for operator composition. We leave checking $\mathcal{L}(\alpha\mathcal{A}) = \alpha\mathcal{L}\mathcal{A}$ as an exercise.

Now we check injectivity. If $\mathcal{L}\mathcal{A} = 0_{\mathcal{V}_2 \to \mathcal{W}_2}$, then $\mathcal{S}\mathcal{A}\mathcal{T}^{-1} = 0_{\mathcal{V}_2 \to \mathcal{W}_2}$. Thus $\mathcal{A} = \mathcal{S}^{-1}0_{\mathcal{V}_2 \to \mathcal{W}_2}\mathcal{T} = 0_{\mathcal{V}_1 \to \mathcal{W}_1}$. This proves injectivity. We emphasize that $0_{\mathcal{V}_1 \to \mathcal{W}_1}$ is the zero vector in $\mathbf{L}(\mathcal{V}_1, \mathcal{W}_1)$ and $0_{\mathcal{V}_2 \to \mathcal{W}_2}$ is the zero vector in $\mathbf{L}(\mathcal{V}_2, \mathcal{W}_2)$.

Last, for surjectivity, let $\widetilde{\mathcal{A}} \in \mathbf{L}(\mathcal{V}_2, \mathcal{W}_2)$. We want to find $\mathcal{A} \in \mathbf{L}(\mathcal{V}_1, \mathcal{W}_1)$ such that $\mathcal{L}\mathcal{A} = \widetilde{\mathcal{A}}$.

$$\begin{array}{ccc} \mathcal{V}_1 & \overset{\mathcal{A}?}{\dashrightarrow} & \mathcal{W}_1 \\ \mathcal{T}\big\downarrow & & \big\downarrow\mathcal{S} \\ \mathcal{V}_2 & \underset{\widetilde{\mathcal{A}}}{\longrightarrow} & \mathcal{W}_2 \end{array}$$

By definition of $\mathcal{L}$, this happens if and only if $\mathcal{S}\mathcal{A}\mathcal{T}^{-1} = \widetilde{\mathcal{A}}$, which is equivalent to $\mathcal{A} = \mathcal{S}^{-1}\widetilde{\mathcal{A}}\mathcal{T}$.

$$\begin{array}{ccc} \mathcal{V}_1 & \overset{\mathcal{S}^{-1}\widetilde{\mathcal{A}}\mathcal{T}}{\dashrightarrow} & \mathcal{W}_1 \\ \mathcal{T}\big\downarrow & & \big\uparrow\mathcal{S}^{-1} \\ \mathcal{V}_2 & \underset{\widetilde{\mathcal{A}}}{\longrightarrow} & \mathcal{W}_2 \end{array}$$

This proves surjectivity. (By the way, the surjectivity proof really shows $\mathcal{L}\mathcal{A} = \widetilde{\mathcal{A}}$ if and only if $\mathcal{A} = \mathcal{S}^{-1}\widetilde{\mathcal{A}}\mathcal{T}$. This really bundles injectivity and surjectivity together, per Theorem 11.5, and so we could have skipped the injectivity proof above.) ■

We conclude our discussion of operator inverses by considering the inverse of a product. Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ be invertible. We probably expect that $\mathcal{S}\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{W})$ is invertible, and if we think of $\mathcal{S}\mathcal{T}$ as "doing $\mathcal{T}$ first, then doing $\mathcal{S}$ to the result of that," then we might expect that $(\mathcal{S}\mathcal{T})^{-1}$ as "undoing $\mathcal{S}$ first, then undoing $\mathcal{T}$." That is, we should conjecture $(\mathcal{S}\mathcal{T})^{-1} = \mathcal{T}^{-1}\mathcal{S}^{-1}$.

13.2 Problem (!). Check that the composition $\mathcal{T}^{-1}\mathcal{S}^{-1}$ is defined with $\mathcal{T}^{-1}\mathcal{S}^{-1} \in \mathbf{L}(\mathcal{W}, \mathcal{U})$.

Proving that $(\mathcal{S}\mathcal{T})^{-1} = \mathcal{T}^{-1}\mathcal{S}^{-1}$ is just a calculation: by Problem 10.13, we need to show

$$(\mathcal{S}\mathcal{T})(\mathcal{T}^{-1}\mathcal{S}^{-1}) = \mathcal{I}_{\mathcal{W}} \quad \text{and} \quad (\mathcal{T}^{-1}\mathcal{S}^{-1})(\mathcal{S}\mathcal{T}) = \mathcal{I}_{\mathcal{V}}.$$

We prove only the first equality, and its proof is mostly associativity of operator composition. Note that $\mathcal{T}\mathcal{T}^{-1} = \mathcal{I}_{\mathcal{V}}$ and $\mathcal{S}\mathcal{S}^{-1} = \mathcal{I}_{\mathcal{W}}$. Then

$$(\mathcal{S}\mathcal{T})(\mathcal{T}^{-1}\mathcal{S}^{-1}) = \mathcal{S}(\mathcal{T}\mathcal{T}^{-1})\mathcal{S}^{-1} = \mathcal{S}\mathcal{I}_{\mathcal{V}}\mathcal{S}^{-1} = \mathcal{S}\mathcal{S}^{-1} = \mathcal{I}_{\mathcal{W}},$$

as desired.

13.3 Problem (!). Prove the other equality: $(\mathcal{T}^{-1}\mathcal{S}^{-1})(\mathcal{S}\mathcal{T}) = \mathcal{I}_{\mathcal{V}}$.

We summarize this formally.

13.4 Theorem. *Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ be invertible. Then $\mathcal{S}\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{W})$ is invertible with inverse*

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

13.5 Problem (!). Operator inverses interact nicely with composition, which is unsurprising—inverses are designed to work with composition! Inverses interact less nicely with arithmetic. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces.

(i) If $\mathcal{S}$, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ are invertible, is $\mathcal{S} + \mathcal{T}$ invertible?

(ii) If $\alpha \in \mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is invertible, is $\alpha\mathcal{T}$ invertible?

13.6 Problem ($\star$). Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces over $\mathbb{F}$. Suppose that $\mathcal{U}$ and $\mathcal{V}$ are isomorphic and $\mathcal{V}$ and $\mathcal{W}$ are isomorphic. Prove that $\mathcal{U}$ and $\mathcal{W}$ are isomorphic. What is the isomorphism? (This, by the way, is one step in showing that isomorphism is an equivalence relation on any set of vector spaces.)

Our fundamental problem is understanding, and maybe solving, the operator equation $\mathcal{T}v = w$, where $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ for vector spaces $\mathcal{V}$ and $\mathcal{W}$ over $\mathbb{F}$, and $w \in \mathcal{W}$. Existence and uniqueness of solutions is guaranteed if $\mathcal{T}$ is invertible, or bijective, or an isomorphism, but checking that is the whole challenge.

The range of a linear operator controls the existence of solutions to our fundamental problem, while the kernel controls uniqueness of those solutions (if they exist). If we want to be able to solve our fundamental problem in as many instances as possible, then we want the range to be as "large" as possible. And if we want solutions to be "as unique as possible," then we want the kernel to be as "small" as possible. We can achieve such quantifiable results on the sizes of ranges and kernels if we specialize to the natural situation of finite-dimensional vector spaces.

Another approach is to obtain more "qualitative" characterizations of range and kernel in terms of other structural aspects of $\mathcal{T}$, $\mathcal{V}$, and/or $\mathcal{W}$. We will do this via geometry and the tools of inner products and norms, which can be available in infinite-dimensional vector spaces, too. Both approaches—dimension, geometry—employ tools that very naturally arise in many problems.

Problems often become simpler (relatively speaking) when we impose more structure. A natural structure to impose on a vector space is that it can be written as the span of finitely many of its vectors; this leads to great control over the vector space, as most questions boil down to an analysis of those finitely many vectors (or fewer!), in particular questions about *operators* on such spaces.

Both $\mathbb{F}^n$ and $\mathbb{P}^n$ have this structure:

$$\mathbb{F}^n = \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_n) \quad \text{and} \quad \mathbb{P}^n = \operatorname{span}(p_0, p_1, \ldots, p_n), \tag{13.1}$$

where $\mathbf{e}_j$ is 1 in its jth entry and 0 elsewhere, and $p_j(x) = x^j$. Even better, a vector in $\mathbb{F}^n$ or $\mathbb{P}^n$ has a *unique* representation in the respective span: there is only one way to choose the coefficients in the linear combination giving that vector. (Best of all, in $\mathbb{F}^n$ we can extract those coefficients using dot products—not quite so in $\mathbb{P}^n$.)

Because we will rely so much on spans, we briefly review some technology associated with them. A list of length $n \geq 1$ in a set Y is a function in $Y^{\{1,\ldots,n\}}$. If $f \in Y^{\{1,\ldots,n\}}$, we write $(f(1), \ldots, f(n)) := f$, and we say that $f(j)$ is the jth entry (or term, or component) of the list. If (v_1) is a list of length 1, sometimes we write $(v_1) = (v_1, \ldots, v_1)$. This should not be confused with a list of length greater than 1 whose entries are all v_1.

Remember that the point of lists is to encode order and allow repetition. In $\mathbb{F}^3$, the lists

$$(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3), \qquad (\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3), \qquad \text{and} \qquad (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1),$$

are all different, but the sets

$$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}, \qquad \{\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3\}, \qquad \text{and} \qquad \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1\}$$

are all the same. (Specifically, these sets are just $\{\mathbf{e}_1, \mathbf{e}_2\}$.)

Now, we can talk about the span of either a list of vectors or a set of vectors. The span of a list is the set of all linear combinations of vectors in that list; the span of a set is the set of all linear combinations of vectors in that set.

13.7 Example. In $\mathbb{F}^3$, the set of entries in the list $(\mathbf{e}_1, \mathbf{e}_2)$ is $\{\mathbf{e}_1, \mathbf{e}_2\}$. The span of the list is

$$\operatorname{span}(\mathbf{e}_1, \mathbf{e}_2) = \left\{ \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} \;\middle|\; v_1,\ v_2 \in \mathbb{F} \right\},$$

and the span of the set is the same:

$$\operatorname{span}(\{\mathbf{e}_1, \mathbf{e}_2\}) = \left\{ \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} \;\middle|\; v_1,\ v_2 \in \mathbb{F} \right\}.$$

But we also have

$$\operatorname{span}(\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1\}) = \operatorname{span}(\{\mathbf{e}_2, \mathbf{e}_1\}) = \operatorname{span}(\{\mathbf{e}_1, \mathbf{e}_2\}).$$

If a vector is in a span, it will be convenient to associate that vector with a set of "coordinates": the coefficients needed to make that vector out of a linear combination. Because lists encode order and prevent redundant repetition, it is easier to perform this association for spans of lists than spans of sets. Consequently, we will only discuss spans of lists. This is at least partially a matter of taste, for which there is no rigorous accounting.

We will need a particular structure associated with a list that we have not yet used. Let (y_1, y_2, y_3, y_4) be a list in a set Y.

13.8 Example. Hopefully, it is intuitively clear why we would want to call the lists (y_1, y_2, y_3), (y_2, y_4), and (y_1) "sublists" of the original list: they are lists whose terms appear in the original list *in the same order*. We would not call (y_2, y_1, y_3) or (y_1, y_1, y_2) sublists of the original list.

The precise definition of sublist is worth considering, but practically speaking we will not need to use it much.

13.9 Definition. *Let $f \in Y^{\{1,\ldots,n\}}$ be a list in Y and let $1 \le m \le n$. A function $g \in Y^{\{1,\ldots,m\}}$ is a* SUBLIST *of f if there exists a strictly increasing function $h \colon \{1,\ldots,m\} \to \{1,\ldots,n\}$ such that $g(j) = f(h(j))$ for each j. (By "strictly increasing," we mean $h(j) < h(j+1)$ for $j = 1\ldots,m-1$.)*

13.10 Example. Consider the list (y_1, y_2, y_3, y_4) in some set Y.

(i) We want to say that (y_1, y_2) is a sublist of (y_1, y_2, y_3, y_4). Here

$$f = \{(1,y_1),(2,y_2),(3,y_3),(4,y_4)\} \quad \text{and} \quad g = \{(1,y_1),(2,y_2)\},$$

with $n = 4$ and $m = 2$. So we want $h(1) = 1$ and $h(2) = 2$.

(ii) If the sublist is (y_2, y_4), then we want $g = \{(1,y_2),(2,y_4)\}$ and $h(1) = 2$, $h(2) = 4$.

Now here is the "simple" kind of vector space structure that we will consider in detail.

13.11 Definition. *A vector space $\mathcal{V}$ is* FINITE-DIMENSIONAL *if if is spanned by a finite list: there exists a list $(v_1,\ldots,v_n)$ in $\mathcal{V}$ such that $\mathcal{V} = \operatorname{span}(v_1,\ldots,v_n)$. If $\mathcal{V}$ is not finite-dimensional, then it is* INFINITE-DIMENSIONAL.

We will soon quantify much more precisely just how "finite" the "dimension" of a finite-dimensional vector space can be. Quantifying infinite-dimensional spaces is trickier, and not always worthwhile.

13.12 Example. The spaces $\mathbb{F}^n$ and $\mathbb{P}^{n+1}$ are finite-dimensional by (13.1). We will prove, with some nontrivial technology, that $\mathbb{F}^\infty$ and $\mathcal{C}^r([0,1])$ are infinite-dimensional.

Definition 13.11 allows some unfortunate inefficiency in writing a vector space as a span.

13.13 Example. **(i)** Let

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

Then the range of $\mathcal{M}_A$ is both the span of all four columns of A and the span of just the first and third columns of A, but not the span of any one column of A.

(ii) With $p_j(x) = x^j$, we have both $\mathbb{P}^2 = \operatorname{span}(p_0, p_1, p_2)$ and $\mathbb{P}^2 = \operatorname{span}(p_0, p_1, p_2, 2p_0)$, but the former span is more efficient (a shorter list) than the latter.

13.14 Problem (!). Prove all of the claims in Example 13.13.

13.15 Problem ($\star$). Here is a generalization of this redundancy (which will actually sometimes be helpful). Let $\mathcal{V}$ be a finite-dimensional vector space with $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$ for some list $(v_1, \ldots, v_n)$ in $\mathcal{V}$. Let $(w_1, \ldots, w_m)$ be another list in $\mathcal{V}$. Prove that $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n, w_1, \ldots, w_m)$.

Day 14: Friday, February 13.

Vocabulary from today

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

Basis for a (finite-dimensional) vector space, basis operator for a vector space associated with a basis, linearly independent list (N), linearly dependent list (N)

We should try to avoid the "linear redundancy" that Definition 13.11 permits by writing a finite-dimensional vector space as the span of "just enough" vectors—enough vectors to span the space, not too many to be unnecessary. Such a "just right" list is a basis: a *unique* "coordinate system" for the space.

14.1 Definition. *Let $\mathcal{V}$ be a finite-dimensional vector space over $\mathbb{F}$. A list $(v_1, \ldots, v_n)$ is a* BASIS *for $\mathcal{V}$ if each vector $v \in \mathcal{V}$ can be written uniquely as the span of $(v_1, \ldots, v_n)$. That is, for each $v \in \mathcal{V}$, there is a unique list $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$ such that $v = \sum_{j=1}^n \alpha_j v_j$.*

14.2 Example. The standard basis vectors $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ better be a basis for $\mathbb{F}^n$. They are: if $\mathbf{v} = (v_1, \ldots, v_n) \in \mathcal{V}$, then doing the arithmetic shows $\mathbf{v} = \sum_{j=1}^n v_j \mathbf{e}_j$. So $\mathbf{v} \in \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. And if $\mathbf{v} = \sum_{j=1}^n w_j \mathbf{e}_j$, then some more arithmetic shows $\sum_{j=1}^n (v_j - w_j)\mathbf{e}_j = \mathbf{0}_n$. Componentwise equalities then force $v_j - w_j = 0$, so $v_j = w_j$. Thus the representation of $\mathbf{v}$ as a span of the list $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is unique.

Here is an important operator-theoretic perspective on bases.

14.3 Problem (!). Let $\mathcal{V}$ be a vector space. Show that the list $(v_1, \ldots, v_n)$ in $\mathcal{V}$ is a basis for $\mathcal{V}$ if and only if the operator

$$\mathcal{B} \colon \mathbb{F}^n \to \mathcal{V} \colon (\alpha_1, \ldots, \alpha_n) \mapsto \sum_{j=1}^n \alpha_j v_j \tag{14.1}$$

is an isomorphism. Also, check that $v_j = \mathcal{B}\mathbf{e}_j$.

14.4 Definition. *The operator $\mathcal{B}$ from* (14.1) *is the* BASIS OPERATOR FOR $\mathcal{V}$ ASSOCIATED WITH THE BASIS $(v_1, \ldots, v_n)$. *The* COORDINATE VECTOR *of $v \in \mathcal{V}$ with respect* *to the basis $(v_1, \ldots, v_n)$ is $[v]_{\mathcal{V}_{\mathcal{B}}} := \mathcal{B}^{-1}v \in \mathbb{F}^n$. This notation does not indicate dependence on the actual basis $(v_1, \ldots, v_n)$, which can usually be discerned from context.*

14.5 Problem (!). Let $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ be invertible. Show that $\mathcal{B}$ is the basis operator for $\mathcal{V}$ associated with the basis $(\mathcal{B}\mathbf{e}_1, \ldots, \mathcal{B}\mathbf{e}_n)$ in the sense that

$$\mathcal{B}(\alpha_1, \ldots, \alpha_n) = \sum_{j=1}^{n} \alpha_j \mathcal{B}\mathbf{e}_j.$$

Content from *Linear Algebra* by Meckes & Meckes. Pages 150–152 review spans and introduce finite-dimensional spaces and bases (the latter from a slightly different point of view). What we (and the Meckeses) call a basis is sometimes called an "ordered" basis in other sources (those sources preferring to think of our bases as sets, not lists).

We will eventually prove that every finite-dimensional vector space has a basis and that all bases are the same length. This length is, of course, the dimension of the space. It can also be shown that *every* vector space, finite- or infinite-dimensional, has a basis (with some adjustments in the definition of basis for the infinite-dimensional case) and that all bases for a space, finite- or infinite-dimensional, have the same length (with length appropriately defined for the infinite-dimensional case). However, a basis for an infinite-dimensional is often just "not that useful"—assume more natural structures on the space, there are better relatives of basis to use.

While our definition of basis encodes the most important idea that a basis is a unique coordinate system (what things do defines what things are), it is often convenient to decouple the "coordinate" aspect of a basis from the uniqueness. This is exactly how we worked through Example 14.2.

14.6 Theorem. *Let $\mathcal{V}$ be a finite-dimensional vector space. A list $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$ if and only if both of the following hold.*

(i) $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$

(ii) *If $\sum_{j=1}^{n} \alpha_j v_j = 0_{\mathcal{V}}$ for $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$, then $\alpha_j = 0$ for all j.*

Proof. $(\Longrightarrow)$ Definition 14.1 of basis immediately implies (i). For (ii), we have $0_{\mathcal{V}} = \sum_{j=1}^{n} 0 v_j$ already, so if $0_{\mathcal{V}} = \sum_{j=1}^{n} \alpha_j v_j$ for some $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$, then by uniqueness $\alpha_j = 0$.

$(\Longleftarrow)$ Condition (i) implies that any $v \in \mathcal{V}$ can be written in the form $v = \sum_{j=1}^{n} \alpha_j v_j$ for some $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$. We need to prove uniqueness. Suppose also that $v = \sum_{j=1}^{n} \beta_j v_j$. Some arithmetic implies $0_{\mathcal{V}} = \sum_{j=1}^{n} (\alpha_j - \beta_j) v_j$. Condition (ii) then forces $\alpha_j - \beta_j = 0$. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.10 on pp. 152–153.

Condition (ii) in the above characterization of basis (possibly the surprising one—spans should not be surprising) is hugely important by itself.

14.7 Definition. *Let $\mathcal{V}$ be a vector space (not necessarily finite-dimensional) and let $(v_1, \ldots, v_n)$ be a list in $\mathcal{V}$.*

(i) *The list is* **(LINEARLY) INDEPENDENT** *if $\sum_{j=1}^{n} \alpha_j v_j = 0_{\mathcal{V}}$ for $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$, then $\alpha_j = 0$ for all j. In symbols,*

$$\forall(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n : \sum_{j=1}^{n} \alpha_j v_j = 0_{\mathcal{V}} \Longrightarrow \forall j : \alpha_j = 0.$$

(ii) *The list is* **(LINEARLY) DEPENDENT** *if it is not independent. In symbols,*

$$\exists(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n : \sum_{j=1}^{n} \alpha_j v_j = 0_{\mathcal{V}} \text{ and } \exists j : \alpha_j \neq 0.$$

14.8 Example. We discuss (in)dependence in several concrete contexts. Throughout, it is important to think about what "being equal to the zero vector" means in different vector spaces.

(i) The list $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ of standard basis vectors in $\mathbb{F}^n$ is independent: if $\sum_{j=1}^{n} \alpha_j \mathbf{e}_j = \mathbf{0}_n$, then doing the arithmetic yields $(\alpha_1, \ldots, \alpha_n) = \mathbf{0}_n$. Thus $\alpha_j = 0$ for all j. (These two sentences illustrate two different uses of list notation: the list of vectors $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ in $\mathbb{F}^n$, which is really a function in $(\mathbb{F}^n)^{1,\ldots,n}$, per the definition of list, and the list $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n = \mathbb{F}^{\{1,\ldots,n\}}$.) Here "being equal to the zero vector" means componentwise equality to the scalar 0.

(ii) Let $p_0(x) = 1$ and $p_1(x) = x$. Then the list (p_0, p_1) is independent in $\mathcal{C}([0,1])$. Suppose $\alpha_0 p_0 + \alpha_1 p_1 = 0$. This is a function equality, so it means that $\alpha_0 p_0(x) + \alpha_1 p_1(x) = 0$ for all $x \in [0,1]$. That is, $\alpha_0 + \alpha_1 x = 0$ for all x.

Intuitively, we want to show that the only line that lies on the x-axis has x-intercept 0 and slope 0. Because $\alpha_0 + \alpha_1 x = 0$ for all $x \in [0,1]$, we can pick any x that we like, so we may as well choose "easy" values of x. At $x = 0$, we have $\alpha_0 = 0$. Then it is the case that $\alpha_1 x = 0$ for all $x \in [0,1]$. Taking another "easy" value, at $x = 1$ we conclude $\alpha_1 = 0$.

(iii) Fix $x \in [0,1]$. The "evaluate at x" map $\varphi_x \colon \mathcal{C}([0,1]) \to \mathbb{R} \colon f \mapsto f(x)$ is a linear functional, so $\varphi_x \in (\mathcal{C}([0,1]))'$. Note that $\varphi(f) = f(x)$. Let $x_1, \ldots, x_n \in [0,1]$ be distinct. We claim that $(\varphi_{x_1}, \ldots, \varphi_{x_n})$ is independent. To show this, suppose that $\sum_{j=1}^{n} \alpha_j \varphi_{x_j} = 0$ for some $(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$. Here "equals 0" means pointwise evaluation—on functions in $\mathcal{C}([0,1])$. That is, we are assuming

$$\sum_{j=1}^{n} \alpha_j \varphi_{x_j}(f) = 0$$

for all $f \in \mathcal{C}([0,1])$. In turn, this means

$$\sum_{j=1}^{n} \alpha_j f(x_j) = 0$$

for all $f \in \mathcal{C}([0,1])$.

Here is the trick: because $x_1, \ldots, x_n$ are distinct, we can "interpolate" them by functions $f_1, \ldots, f_n \in \mathcal{C}([0,1])$ such that

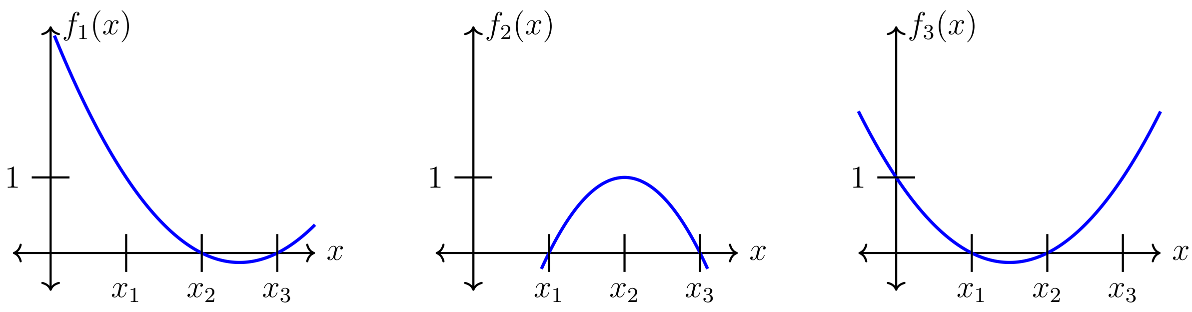
$$f_j(x_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

Then for each k,

$$0 = \sum_{j=1}^{n} \alpha_j f_k(x_j) = \alpha_k,$$

thus $\alpha_k = 0$ for all k.

To figure out how to construct these f_j, we draw some pictures when $n = 3$ (for simplicity):



This suggests taking

$$f_1(x) := \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)}, \qquad f_2(x) := \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)},$$

$$\text{and} \qquad f_3(x) := \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)}.$$

Similar, but more complicated, formulas work for the case of a general n.

14.9 Problem (!). Show that the list $(1, \sin(\cdot), \cos(\cdot))$ is independent in $\mathcal{C}([0,1])$.

Here is an easier situation.

14.10 Lemma. *A list of length 1 is dependent if and only if its only entry is the zero vector. That is, (v_1) is dependent if and only if $v_1 = 0$.*

Proof. ($\Longrightarrow$) If (v_1) is dependent, then (v_1) is not independent, so there is $\alpha_1 \in \mathbb{F}$ such that $\alpha_1 v_1 = 0_{\mathcal{V}}$ and $\alpha_1 \neq 0$. Thus $v_1 = 0_{\mathcal{V}}$.

($\Longleftarrow$) If $v_1 = 0_{\mathcal{V}}$, then $1 \cdot 0_{\mathcal{V}} = 0_{\mathcal{V}}$, so $(0_{\mathcal{V}})$ is not independent. ■

14.11 Problem ($\star$). Prove the following easy ways of checking the dependence of a list.

(i) A list (of length at least 2) with a repeated vector is dependent.

(ii) A list with the zero vector is dependent.

(iii) A list of length 2 is dependent if and only if one vector is a scalar multiple of the other. (Must the first be a scalar multiple of the second?)

14.12 Problem ($\star$). The field matters when thinking about independence. Prove that the list $(1, i)$ is independent in the vector space $\mathcal{V} = \mathbb{C}$ considered over the field $\mathbb{F} = \mathbb{R}$ but dependent in the vector space $\mathcal{W} = \mathbb{C}$ considered over the field $\mathbb{F} = \mathbb{C}$.

14.13 Problem (!). Let $(v_1, \ldots, v_n)$ be an independent list in the space $\mathcal{V}$, and let $(\alpha_1, \ldots, \alpha_n)$ be a list of nonnegative numbers. Prove that if $\alpha_k > 0$ for at least one k, then $\sum_{j=1}^{n} \alpha_j v_j \neq 0_{\mathcal{V}}$.

Here is an important characterization of a dependent list.

14.14 Theorem (Linear dependence is linear redundancy). *A list of length at least 2 is dependent if and only if one vector in the list is a linear combination of the others. That is, if $n \geq 2$, then $(v_1, \ldots, v_n)$ is dependent if and only if there exists k such that*

$$v_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} \alpha_j v_j \tag{14.2}$$

for some $\alpha_j \in \mathbb{F}^n$, $j \neq k$.

Proof. ($\Longrightarrow$) If $(v_1, \ldots, v_n)$ is dependent, then there exists $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n \setminus \{\mathbf{0}_n\}$ such that $\sum_{j=1}^{n} \alpha_j v_j = 0_{\mathcal{V}}$ with $\alpha_j \neq 0$ for at least one j. Let k be the largest integer such that $\alpha_k \neq 0$. If $k = 1$, then $\alpha_2 = \cdots = \alpha_n = 0$, so $\alpha_1 v_1 = 0_{\mathcal{V}}$. Since $\alpha_1 \neq 0$, we have

$$v_1 = 0_{\mathcal{V}} = \sum_{j=2}^{n} 0 v_j.$$

If $k = n$, then

$$v_n = \sum_{j=1}^{n-1} \left(-\frac{\alpha_j}{\alpha_n} \right) v_j.$$

Finally, if $1 < k < n$ (which, incidentally, implies $n \geq 3$), we have $\sum_{j=1}^{k} \alpha_j v_j = 0_{\mathcal{V}}$; the sum stops at $j = k$ since $\alpha_{k+1} = \cdots = \alpha_n = 0$. Then

$$v_k = \sum_{j=1}^{k-1} \left(-\frac{\alpha_j}{\alpha_k}\right) v_j.$$

The case $k = n$ is really just a special case of this one, as is the case $k = 1$ (which we just singled out to avoid the awkward expression $\sum_{j=2}^{0}$).

($\Longleftarrow$) Conversely, (14.2) rearranges into

$$\sum_{j=1}^{n} \beta_j v_j = 0_{\mathcal{V}}, \qquad \beta_j := \begin{cases} \alpha_j, & j \neq k \\ -1, & j = k, \end{cases}$$

and we note that $\beta_k \neq 0$. ■

While important, this characterization puts a "burden of guilt" on one particular vector in a list for dependence. If the list is long, it may be hard to spot which vector is a linear combination of the others. Our original definition of dependence is more "democratic": all vectors are "equally guilty."

Content from *Linear Algebra* by Meckes & Meckes. Pages 140–143 introduce linear (in)dependence. The material on pp. 143–145 should be familiar from a first course in linear algebra.

14.15 Problem (!). Let $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ be a list in $\mathbb{F}^m$ and let $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$. Prove that the following are equivalent.

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

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

(iii) $\mathcal{M}_A \colon \mathbb{F}^n \to \mathbb{F}^m$ is injective.

Linear redundancy, while ineffective, is not terribly hard to overcome.

14.16 Lemma (Removing linearly redundant vectors preserves spans). *If a list of length $n \geq 2$ is dependent, then a sublist of length $n-1$ has the same span. That is, if $(v_1, \ldots, v_n)$ is dependent, then there exists a sublist $(v_{j_1}, \ldots, v_{j_{n-1}})$ of this list such that* $\operatorname{span}(v_1, \ldots, v_n) = \operatorname{span}(v_{j_1}, \ldots, v_{j_{n-1}})$. *More precisely, if v_k is a linear combination of the other v_j, then the sublist formed by removing v_k from $(v_1, \ldots, v_n)$ has the same span as the original list $(v_1, \ldots, v_n)$.*

Proof. Since the list is dependent, an entry v_k is a linear combination of the others:

$$v_k = \sum_{\substack{j=1 \\ j \neq k}} \alpha_j v_j.$$

Then any linear combination of the original list $(v_1, \ldots, v_n)$ has the form

$$\sum_{j=1}^{n} \beta_j v_j = \sum_{\substack{j=1 \\ j \neq k}} \beta_j v_j + \beta_k \sum_{\substack{j=1 \\ j \neq k}} \alpha_j v_j = \sum_{\substack{j=1 \\ j \neq k}} (\beta_j + \beta_k \alpha_j) v_j.$$

The desired sublist is the original list with v_k removed. Conversely, any linear combination of the sublist is certainly in the span of the original list:

$$\sum_{\substack{j=1 \\ j \neq k}}^{n} \gamma_j v_j = \sum_{j=1}^{n} \mu_j v_j, \qquad \mu_j = \begin{cases} \gamma_j, & j \neq k \\ 0, & j = k. \end{cases}$$

■

There is yet another way of packaging dependence that nicely encodes the idea of sweeping the columns of a matrix from left to right. For various contemporary cultural reasons, this particular result has earned the title of "linear (in)dependence lemma," even though most of our current results involve linear (in)dependence.

14.17 Lemma (Linear (in)dependence lemma). *Let $(v_1, \ldots, v_n)$ be a list of length $n \geq 2$ with $v_1 \neq 0_{\mathcal{V}}$. Then $(v_1, \ldots, v_n)$ is dependent if and only if there exists $k \geq 2$ such that $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$. Equivalently, when $n \geq 2$ and $v_1 \neq 0_{\mathcal{V}}$, the list $(v_1, \ldots, v_n)$ is independent if and only if $v_j \notin \operatorname{span}(v_1, \ldots, v_{j-1})$ for each $j \geq 2$.*

Proof. ($\Longrightarrow$) If $(v_1, \ldots, v_n)$ is dependent, then there exists $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n \setminus \{\mathbf{0}_n\}$ such that $\sum_{j=1}^{n} \alpha_j v_j = 0_{\mathcal{V}}$ with $\alpha_j \neq 0$ for at least one j. Let k be the largest integer such that $\alpha_k \neq 0$. If $k = 1$, then $\alpha_2 = \cdots = \alpha_n = 0$, and then $\alpha_1 v_1 = 0_{\mathcal{V}}$. Since $\alpha_1 \neq 0$, we have $v_1 = 0_{\mathcal{V}}$. So, $k \geq 2$. Then $\sum_{j=1}^{k} \alpha_j v_j = 0_{\mathcal{V}}$, and this rearranges to

$$v_k = \sum_{\substack{j=1 \\ j \neq k}}^{k-1} \left(-\frac{\alpha_j}{\alpha_k} \right) v_j \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$

($\Longleftarrow$) Conversely, if $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$, then $v_k = \sum_{j=1}^{k-1} \alpha_j v_j$ for some $(\alpha_1, \ldots, \alpha_{k-1}) \in \mathbb{F}^{k-1}$. Then

$$\sum_{j=1}^{n} \beta_j v_j = 0_{\mathcal{V}}, \qquad \beta_j := \begin{cases} \alpha_j, & 1 \leq j \leq k-1 \\ -1, & j = k \\ 0, & k+1 \leq j \leq n. \end{cases}$$

Since $\beta_j = -1 \neq 0$, the list $(v_1, \ldots, v_n)$ is therefore dependent. ■

14.18 Remark. *Here is why it so important to assume that $v_1 \neq 0_{\mathcal{V}}$ in the linear (in)dependence lemma. If $(v_2, \ldots, v_n)$ is an independent list, then $(0_{\mathcal{V}}, v_2, \ldots, v_n)$ is dependent. We still have $v_j \notin \operatorname{span}(0_{\mathcal{V}}, v_2, \ldots, v_{j-1})$ for $j \geq 2$ since $(v_2, \ldots, v_n)$ is independent.*

Content from *Linear Algebra* by Meckes & Meckes. The linear (in)dependence lemma is Theorem 3.6 and Corollary 3.7 on p. 145. Do Quick Exercise #5 on p. 146. Try doing Quick Exercise #4 on p. 144 using Theorem 3.6.

14.19 Problem (!). Let $\mathcal{V}$ be a vector space and suppose that the list $(v_1, \ldots, v_n)$ in $\mathcal{V}$ is independent. Let $w \in \mathcal{V} \setminus \operatorname{span}(v_1, \ldots, v_n)$. Use the linear independence lemma to prove that $(v_1, \ldots, v_n, w)$ is independent.

Day 15: Monday, February 16.

Now we prove that every finite-dimensional vector space has a basis. Here is the intuitive idea of the proof. Any finite-dimensional vector space $\mathcal{V}$ is spanned by a finite list: $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$. If that list is independent, stop: it already is a basis. Otherwise, some vector in that list is a linear combination of the others (Theorem 14.14); remove that vector, and call the new list $(v_{j_1}, \ldots, v_{j_{n-1}})$. By Lemma 14.16, we still have $\mathcal{V} = \operatorname{span}(v_{j_1}, \ldots, v_{j_{n-1}})$. If $(v_{j_1}, \ldots, v_{j_{n-1}})$ is independent, stop: we now have a basis. Otherwise, remove another vector and keep going.

Whenever a proof scheme has this idea of "keep going"—or of iterating a process, or of "turning the crank"—this is an indication that the proof involves induction. Here, the rigorous part of the proof of the existence of a basis is knowing that the removal process eventually ends. Why do we not "keep going" indefinitely? (Intuitive answer: we would run out of vectors in the spanning list.)

Content from *Linear Algebra* by Meckes & Meckes. Pages 388–389 in Appendix A review proof by induction. Here is another summary. Suppose that for each integer $n \geq 1$, $P(n)$ is a statement that can be true or false. Suppose that $P(1)$ is true and that if $P(n)$ is true, then $P(n+1)$ is true. Then $P(n)$ is true for all n. Here is why: if $P(m)$ is false for some m, let $F = \{n \geq 1 \mid P(n) \text{ is false}\}$. Then $m \in F$, so $F \neq \varnothing$. That is, F is a nonempty set of positive integers, and so F has a least element n_0. This number n_0 satisfies $n_0 \in F$ and $n \geq n_0$ for all $n \in F$. Since $P(1)$ is true, $1 \notin F$, and so $n_0 \geq 2$. But then $1 \leq n_0 - 1$ and $n_0 - 1 \notin F$, so $P(n_0 - 1)$ is true, thus $P(n_0)$ is true after all, a contradiction.

15.1 Theorem (Any spanning list can be reduced to a basis). *Let $\mathcal{V}$ be a nonzero vector space with $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$, then there exists an independent sublist $(v_{j_1}, \ldots, v_{j_r})$ of $(v_1, \ldots, v_n)$ such that $\mathcal{V} = \operatorname{span}(v_{j_1}, \ldots, v_{j_r})$ as well.*

Proof. We induct on n. For $n = 1$, $\mathcal{V} = \operatorname{span}(v_1)$ and since $\mathcal{V}$ is nonzero, $v_1 \neq 0_{\mathcal{V}}$, so the list (v_1) is independent. Since it spans $\mathcal{V}$ already, it is a basis.

Assume that the result is true for some $n \geq 1$ and suppose now that $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n, v_{n+1})$. Let $\mathcal{V}_n = \operatorname{span}(v_1, \ldots, v_n)$. By the induction hypothesis, $\mathcal{V}_n = \operatorname{span}(v_{j_1}, \ldots, v_{j_r})$ for an independent sublist $(v_{j_1}, \ldots, v_{j_r})$ of $(v_1, \ldots, v_n)$. Then

$$\operatorname{span}(v_{j_1}, \ldots, v_{j_r}, v_{n+1}) = \operatorname{span}(v_1, \ldots, v_n, v_{n+1}) = \mathcal{V}.$$

Now consider the following two cases. First, if $v_{n+1} \in \mathcal{V}_n$, then $v_{n+1} \in \operatorname{span}(v_{j_1}, \ldots, v_{j_r})$. By the "removal" process from Lemma 14.16, we have

$$\operatorname{span}(v_{j_1}, \ldots, v_{j_r}, v_{n+1}) = \operatorname{span}(v_{j_1}, \ldots, v_{j_r}),$$

and so $\mathcal{V} = \operatorname{span}(v_{j_1}, \ldots, v_{j_r})$. The list $(v_{j_1}, \ldots, v_{j_r})$ is still an independent sublist of $(v_1, \ldots, v_n)$ and thus an independent sublist of $(v_1, \ldots, v_n, v_{n+1})$.

Second, if $v_{n+1} \notin \operatorname{span}(v_{j_1}, \ldots, v_{j_r})$, then the list $(v_{j_1}, \ldots, v_{j_r}, v_{n+1})$ is independent. And so $(v_{j_1}, \ldots, v_{j_r}, v_{n+1})$ is an independent list that spans $\mathcal{V}$. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.11 on p. 153, with a different proof (specifically, one that uses Lemmas 14.16 and 14.17).

15.2 Theorem. *Every nonzero finite-dimensional vector space has a basis.*

Proof. Let $\mathcal{V}$ be a nonzero finite-dimensional vector space. By definition, $\mathcal{V}$ is spanned by some finite list, and so Theorem 15.1 allows us to reduce that list to an independent sublist with the same span. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Corollary 3.12 on p. 154. Algorithm 3.13 on that page should be familiar from a first course in linear algebra.

We now know that every finite-dimensional vector space has a basis: a coordinate system with no redundancy (i.e., the coordinates of a vector are unique). It would be terribly inefficient, and maybe redundant, if two bases could have different lengths. This cannot happen: all bases for a finite-dimensional space have the same length, and of course this length is the dimension of the space. Proving that dimension is "well-defined" is a major result in linear algebra, and every proof is, arguably, at least somewhat challenging. We follow a recent approach due to Jochen Glück, available here:

https://mathoverflow.net/questions/499774/alternative-proofs-that-two-bases-of-a-vector-space-have-the-same-size.

This takes some preparation, but the ancillary results that we develop along the way have multiple uses, including an immediate application to understanding the operator problem $\mathcal{T}v = w$.

We encourage some warm-up exercises to refresh awareness of injectivity, surjectivity, spans, and independence.

15.3 Problem (!). Let $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$ be vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{V})$, $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$.

(i) Prove that if $\mathcal{T}$ and $\mathcal{S}$ are injective, then $\mathcal{S}\mathcal{T}$ is also injective.

(ii) Prove that if $\mathcal{T}$ and $\mathcal{S}$ are surjective, then $\mathcal{S}\mathcal{T}$ is also surjective.

15.4 Problem ($\star$). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces. Suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is not injective and $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$. Show that the list $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ is dependent. [Hint: *let $v \in \ker(\mathcal{T}) \setminus \{0_\mathcal{V}\}$ and expand v as a linear combination of the list $(v_1, \ldots, v_n)$. Apply $\mathcal{T}$.*]

15.5 Problem ($\star$). **(i)** Show that injections preserve independence. That is, prove that if $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is injective, and $(v_1, \ldots, v_n)$ is an independent list in $\mathcal{V}$, then $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ is an independent list in $\mathcal{W}$.

(ii) Give an example to show how this result may fail if $\mathcal{T}$ is not injective.

15.6 Problem ($\star$). **(i)** Show that surjections preserve spans. That is, prove that if $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is surjective, and $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$, then $\mathcal{W} = \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$.

(ii) Show that this result is false if $\mathcal{T}$ is not surjective. That is, give an example of $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ such that $\mathcal{T}$ is not surjective, $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$, but $\mathcal{W} \neq \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$.

Here is our first tool toward the proof that dimension is well-defined. It partially encodes the idea that injectivity requires "breathing space": an operator is injective if it maps any two distinct inputs to distinct outputs, and so there must be enough "space" in the codomain relative to the "amount" of inputs in the domain. If an operator is injective and not surjective, then there should be "extra space" left over in the codomain compared to the domain. We begin with the only spaces whose "dimension" we definitely know: Euclidean space.

15.7 Lemma. *If $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is injective and not surjective, then $m \geq 2$ and there is an injection $\mathcal{S} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^{m-1})$.*

Proof. We leave the proof that $m \geq 2$ as an exercise. We claim that there must be k such that $\mathcal{T}\mathbf{v} \neq \mathbf{e}_k^{(m)}$ for all $\mathbf{v} \in \mathbb{F}^n$, where $\mathbf{e}_k^{(m)}$ is the kth standard basis vector in $\mathbb{F}^m$. Let $\mathcal{P}_{m,k} \colon \mathbb{F}^m \to \mathbb{F}^{m-1}$ be the operator that "removes the kth coordinate." For example, if $m = 3$ and $k = 2$, then $\mathcal{P}_{3,2}(1, 2, 3) = (1, 3)$. More generally,

$$\mathcal{P}_{m,k}\mathbf{w} = \sum_{j=1}^{k-1} w_j \mathbf{e}_j^{(m-1)} + \sum_{j=k+1}^{m} w_j \mathbf{e}_{j-1}^{(m-1)}.$$

Here $\mathbf{e}_j^{(m-1)}$ is the jth standard basis vector in $\mathbb{F}^{m-1}$.

We first consider $\ker(\mathcal{P}_{m,k})$. If $\mathcal{P}_{m,k}\mathbf{w} = \mathbf{0}_{m-1}$, then by linear independence $w_j = 0$ for $j \neq k$, and so $\mathbf{w} \in \operatorname{span}(\mathbf{e}_k^{(m)})$. (By the way, $\mathcal{P}_{m,k}$ is not injective.)

Now we consider $\ker(\mathcal{P}_{m,k}\mathcal{T})$. If $\mathcal{P}_{m,k}\mathcal{T}\mathbf{v} = \mathbf{0}_{m-1}$, then $\mathcal{T}\mathbf{v} \in \operatorname{span}(\mathbf{e}_k)$. If $\mathcal{T}\mathbf{v} \neq \mathbf{0}_m$, then $\mathcal{T}\mathbf{v} = \alpha\mathbf{e}_k^{(m)}$ for some $\alpha \neq 0$, thus $\mathcal{T}(\alpha^{-1}\mathbf{v}) = \mathbf{e}_k^{(m)}$, a contradiction. Hence $\mathcal{T}\mathbf{v} = \mathbf{0}_m$, so by the injectivity of $\mathcal{T}$, we have $\mathbf{v} = \mathbf{0}_n$. Thus $\ker(\mathcal{P}_{m,k}\mathcal{T}) = \{\mathbf{0}_n\}$. The desired injection is therefore $\mathcal{S} := \mathcal{P}_{m,k}\mathcal{T}$. (By the way, this shows that the composition of an injective operator and a noninjective operator can still be injective; in particular, we did not, and could not, use Problem 15.3.) ■

15.8 Problem (!). **(i)** Prove the claim in Lemma 15.7 that there must be k such that $\mathcal{T}\mathbf{v} \neq \mathbf{e}_k^{(m)}$ for all $\mathbf{v} \in \mathbb{F}^n$.

(ii) Prove the claim in Lemma 15.7 that $m \geq 2$.

Day 16: Wednesday, February 18.

Now here is the rigorous verification of the notion that "injectivity implies breathing space."

16.1 Theorem. *If $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is injective, then $n \leq m$.*

Proof. We induct on n. If $n = 1$, then since $m \geq 1$ anyway, the result is immediate.

Suppose that for some $n \geq 1$ and all $m \geq 1$, if $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is injective, then $n \leq m$. Now let $\mathcal{T} \in \mathbf{L}(\mathbb{F}^{n+1}, \mathbb{F}^m)$ be injective. We want to show that $n + 1 \leq m$, equivalently, $n \leq m - 1$. By the induction hypothesis (replacing m with $m - 1$), it therefore suffices to find an injection in $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^{m-1})$. And by Lemma 15.7, we can do that by finding an injection in $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ that is not a surjection.

We accomplish this by working with the "natural embedding" of $\mathbb{F}^n$ into $\mathbb{F}^{n+1}$ given by

$$\mathcal{J}_n : \mathbb{F}^n \to \mathbb{F}^{n+1} : \mathbf{v} \mapsto (\mathbf{v}, 0).$$

Then $\mathcal{J}_n$ is injective, and so $\mathcal{T}\mathcal{J}_n \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is injective by Problem 15.3.

However, $\mathcal{T}\mathcal{J}_n$ is not surjective; the most natural way to establish this is to work with what $\mathcal{J}_n$ "misses," which is all (nonzero) scalar multiples of $\mathbf{e}_{n+1}$. (Here $\mathbf{e}_{n+1}$ is the $(n+1)$st standard basis vector in $\mathbb{F}^{n+1}$.) Put $\mathbf{w} := \mathcal{T}\mathbf{e}_{n+1} \in \mathbb{F}^m$. If there is $\mathbf{v} \in \mathbb{F}^n$ such that $\mathcal{T}\mathcal{J}_n\mathbf{v} = \mathbf{w}$, then $\mathcal{T}(\mathcal{J}_n\mathbf{v}) = \mathcal{T}\mathbf{e}_{n+1}$. By the injectivity of $\mathcal{T}$, we conclude $\mathcal{J}_n\mathbf{v} = \mathbf{e}_{n+1}$. This is impossible from the definition of $\mathcal{J}_n$.

So, $\mathcal{T}\mathcal{J}_n : \mathbb{F}^n \to \mathbb{F}^m$ is injective and not surjective. By Lemma 15.7, $m \geq 2$ and there is an injection $\mathcal{S} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^{m-1})$. By the induction hypothesis, $n \leq m - 1$, and so $n + 1 \leq m$. ■

Next we consider surjectivity and reversing the previous inequality to get $m \leq n$. The following partially encodes the idea that surjectivity requires "sufficient resources": an operator is surjective if there is an input for every possible output, and so there must be enough "space" in the domain relative to the "amount" of outputs in the codomain. If an operator is surjective and not injective, then there should be "overly" sufficient resources in the domain compared to the codomain.

16.2 Lemma. *If $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is surjective and not injective, then $n \geq 2$ and there is a surjection $\mathcal{S} \in \mathbf{L}(\mathbb{F}^{n-1}, \mathbb{F}^m)$.*

Proof. We leave the proof that $n \geq 2$ as an exercise. Let $\mathbf{z} \in \ker(\mathcal{T}) \setminus \{\mathbf{0}_n\}$ and suppose that $z_k \neq 0$. Since $\mathcal{T}\mathbf{v} = \mathbf{0}_m$, we obtain

$$\mathcal{T}\mathbf{e}_k = \sum_{\substack{j=1 \\ j\neq k}}^{n} \left(-\frac{z_j}{z_k}\right)\mathbf{e}_j.$$

(Here $\mathbf{e}_j$ is the jth standard basis vector in $\mathbb{F}^n$.) And since $\mathcal{T}$ is surjective, for each $\mathbf{w} \in \mathbb{F}^m$, there is $\mathbf{v} \in \mathbb{F}^n$ such that

$$\mathbf{w} = \mathcal{T}\mathbf{v} = \sum_{j=1}^{n} v_j\mathcal{T}\mathbf{e}_j = \sum_{j=1}^{n} \alpha_j\mathcal{T}\mathbf{e}_j \tag{16.1}$$

for an appropriate choice of $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$ with $\alpha_k = 0$. Motivated by this, we define

$$\mathcal{S}\colon \mathbb{F}^{n-1} \to \mathbb{F}^m\colon (\alpha_1, \ldots, \alpha_{n-1}) \mapsto \sum_{j=1}^{k-1} \alpha_j\mathcal{T}\mathbf{e}_j + \sum_{j=k+1}^{n} \alpha_{j-1}\mathcal{T}\mathbf{e}_j.$$

The calculation (16.1) shows that $\mathcal{S}$ is surjective, and so the composition $\mathcal{T}\mathcal{S} \in \mathbf{L}(\mathbb{F}^{n-1}, \mathbb{F}^m)$ is also surjective by Problem 15.3. ∎

16.3 Problem (!). Prove the claim in Lemma 16.2 that $n \geq 2$.

Day 17: Friday, February 20.

You took Exam 1.

Day 18: Monday, February 23.

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."

Dimension of a finite-dimensional vector space

18.1 Theorem. *If $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is surjective, then $n \geq m$.*

Proof. We induct on m. If $m = 1$, then the result is immediate.

Suppose that for some $m \geq 1$ and all n, if $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is surjective, then $n \geq m$. Now let $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^{m+1})$ be surjective. Let $\mathcal{P}_m$ be the "natural projection" of $\mathbb{F}^{m+1}$ onto $\mathbb{F}^m$ given by

$$\mathcal{P}_m\colon \mathbb{F}^{m+1} \to \mathbb{F}^m\colon (w_1, \ldots, w_m, w_{m+1}) \mapsto (w_1, \ldots, w_m).$$

Then $\mathcal{P}_m$ is surjective, and so $\mathcal{P}_m\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is surjective.

However, $\mathcal{P}_m\mathcal{T}$ is not injective. Since $\mathcal{T}$ is surjective, there is $\mathbf{v} \in \mathbb{F}^n$ such that $\mathcal{T}\mathbf{v} = \mathbf{e}_{m+1}$; in particular, $\mathbf{v} \neq \mathbf{0}_n$. Then $\mathcal{P}_m\mathcal{T}\mathbf{v} = \mathcal{P}_m\mathbf{e}_{m+1} = \mathbf{0}_m$. By Lemma 16.2, $n \geq 2$, and there is a surjection $\mathcal{S} \in \mathbf{L}(\mathbb{F}^{n-1}, \mathbb{F}^m)$. By the induction hypothesis, $n-1 \leq m$, and so $n \leq m+1$. ■

18.2 Problem (!). Show that if there is an isomorphism $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$, then $n = m$.

Here is the major result toward which we have been building.

18.3 Theorem (Dimension is well-defined). *All bases for a finite-dimensional vector space have the same length.*

Proof. Let $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$ be bases for $\mathcal{V}$. Let $\mathcal{B}_n \colon \mathbb{F}^n \to \mathcal{V}$ and $\mathcal{B}_m \colon \mathbb{F}^m \to \mathcal{V}$ be the basis operators, so $\mathcal{B}_n$ and $\mathcal{B}_m$ are isomorphisms. Then $\mathcal{B}_m^{-1}\mathcal{B}_n \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is an isomorphism, so $n = m$. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.19 on p. 164, albeit with a different proof.

Since all bases for a finite-dimensional space have the same length, we can associate a single quantity with that length.

18.4 Definition. *Let $\mathcal{V}$ be a finite-dimensional vector space. The* DIMENSION *of $\mathcal{V}$ is 0 if $\mathcal{V} = \{0_{\mathcal{V}}\}$, and otherwise it is the length of any basis for $\mathcal{V}$. We denote the dimension of $\mathcal{V}$ by $\dim(\mathcal{V})$. If we are discussing a vector space and refer to $\dim(\mathcal{V})$, we are tacitly assuming that $\mathcal{V}$ is finite-dimensional. We do not adopt the occasional convention that if $\mathcal{V}$ is infinite-dimensional, then $\dim(\mathcal{V}) = \infty$.*

Content from *Linear Algebra* by Meckes & Meckes. Page 164 defines dimension. Read carefully the paragraph after that definition. Algorithm 3.25 on p. 166 should be familiar from a first course in linear algebra.

18.5 Example. **(i)** Since $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is a basis for $\mathbb{F}^n$, unsurprisingly $\dim(\mathbb{F}^n) = n$.

(ii) Since $(p_0, \ldots, p_n)$ is a basis for $\mathbb{P}^n$, with $p_j(x) = x^j$, and since this list has length $n+1$, $\dim(\mathbb{P}^n) = n + 1$.

(iii) For $i = 1, \ldots, m$ and $j = 1, \ldots, n$, let E_{ij} be the $m \times n$ matrix whose (i, j)-entry is 1 and whose other entries are 0. For example, if $m = 2$ and $n = 3$, then

$$E_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad E_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It should be conceptually unsurprising to accept, but perhaps notationally annoying to prove, that the list of these E_{ij} is a basis for $\mathbb{F}^{m \times n}$. There are mn such E_{ij}, and so $\dim(\mathbb{F}^{m \times n}) = mn$.

18.6 Problem ($\star$). Dimension gives the "correct" notion of "size" for a vector space, at least when the space is finite-dimensional. Let $\mathcal{V}$ be a nonzero vector space. Prove that $\mathcal{V}$ contains infinitely many vectors.

18.7 Problem (!). Let $\mathcal{V}$ be a one-dimensional vector space over $\mathbb{F}$ and let $\mathcal{W}$ be a vector space. Let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Find a "formula" for $\mathcal{T}$ that resembles the result of Problem 9.5.

Now we apply our notion of dimension to the operator problem $\mathcal{T}v = w$. We will be interested in the contrapositives of the following results.

18.8 Theorem (Dimension, injectivity, and surjectivity). *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$.*

(i) *If $\mathcal{T}$ is injective and $\mathcal{V}$ is finite-dimensional, then either $\mathcal{W}$ is infinite-dimensional or $\mathcal{W}$ is finite-dimensional with $\dim(\mathcal{V}) \le \dim(\mathcal{W})$.*

(ii) *If $\mathcal{T}$ is surjective and $\mathcal{W}$ is finite-dimensional, then either $\mathcal{V}$ is infinite-dimensional or $\mathcal{V}$ is finite-dimensional with $\dim(\mathcal{V}) \ge \dim(\mathcal{W})$.*

(iii) *If $\mathcal{T}$ is bijective and one of $\mathcal{V}$ or $\mathcal{W}$ is finite-dimensional, then both $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional and $\dim(\mathcal{V}) = \dim(\mathcal{W})$.*

Proof. (i) Suppose that $\mathcal{W}$ is not infinite-dimensional, so $\mathcal{W}$ is finite-dimensional. Let $n = \dim(\mathcal{V})$ and $m = \dim(\mathcal{W})$. Let $\mathcal{B}_n \colon \mathbb{F}^n \to \mathcal{V}$ and $\mathcal{B}_m \colon \mathbb{F}^m \to \mathcal{W}$ be isomorphisms. Then $\mathcal{B}_m^{-1}\mathcal{T}\mathcal{B}_n \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is injective, so $n \le m$ by Theorem 16.1.

(ii) Suppose that $\mathcal{V}$ is not infinite-dimensional, so $\mathcal{V}$ is finite-dimensional. Let $n = \dim(\mathcal{V})$ and $m = \dim(\mathcal{W})$. Let $\mathcal{B}_n \colon \mathbb{F}^n \to \mathcal{V}$ and $\mathcal{B}_m \colon \mathbb{F}^m \to \mathcal{W}$ be isomorphisms. Then $\mathcal{B}_m^{-1}\mathcal{T}\mathcal{B}_n \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ is surjective, so $n \le m$ by Theorem 18.1.

(iii) This follows by combining the previous two parts. ■

The contrapositives of the first two results of Theorem 18.8 are quite useful, in a negative sense, for understanding the fundamental problem $\mathcal{T}v = w$. The first part says that for $\mathcal{T}$ to be injective, $\mathcal{W}$ needs to be "large enough" relative to $\mathcal{V}$: an injection "spreads out" all of $\mathcal{V}$ into $\mathcal{W}$. If $\dim(\mathcal{W}) < \dim(\mathcal{V})$, then no operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ can be injective, and so uniqueness always fails in the problem $\mathcal{T}v = w$. (Existence may also fail, too.)

The second part says that for $\mathcal{T}$ to be surjective, $\mathcal{V}$ needs to be "large enough" relative to $\mathcal{W}$: a surjection has to "cover" all of $\mathcal{W}$. If $\dim(\mathcal{V}) < \dim(\mathcal{W})$, then no operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ can be surjective, and so existence will sometimes fail in the problem $\mathcal{T}v = w$. (Uniqueness may also fail, too.)

By the way, these results should feel familiar from considering how injectivity and surjectivity interact with set cardinality. But in the context of vector spaces, dimension replaces cardinality as the correct and useful measurement of "size" of a space.

18.9 Problem (!). Let $\mathcal{V}$ be a finite-dimensional vector space, let $\mathcal{W}$ be a vector space, and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is not injective. Let $(z_1, \ldots, z_p)$ be a basis for $\ker(\mathcal{T})$, with $1 \le p \le \dim(\mathcal{V})$. Last, let $w \in \mathcal{W}$ and $v_0 \in \mathcal{V}$ such that $\mathcal{T}v_0 = w$. Prove that any other $v \in \mathcal{V}$ with $\mathcal{T}v = w$ has the form

$$v = v_0 + \sum_{j=1}^{p} \alpha_j z_j$$

for some $(\alpha_1, \ldots, \alpha_p) \in \mathbb{F}^p$. This makes more precise our earlier observation (preceding Problem 12.7) that if $\mathcal{T}$ is not injective and $\mathcal{T}v = w$ has *a* solution, then the problem has infinitely many solutions.

We can prove a stronger result than part (iii) of Theorem 18.8 offers: an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is bijective precisely when $\mathcal{V}$ and $\mathcal{W}$ have the same dimension.

18.10 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces over $\mathbb{F}$. Then* $\dim(\mathcal{V}) = \dim(\mathcal{W})$ *if and only if $\mathcal{V}$ and $\mathcal{W}$ are isomorphic.*

Proof. ($\Longrightarrow$) Let $n = \dim(\mathcal{V}) = \dim(\mathcal{W})$. By Problem 14.3, $\mathcal{V}$ and $\mathcal{W}$ are each isomorphic to $\mathbb{F}^n$. That is, there are isomorphisms $\mathcal{B}_{\mathcal{V}} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ and $\mathcal{B}_{\mathcal{W}} \in \mathbf{L}(\mathbb{F}^n, \mathcal{W})$. Then $\mathcal{B}_{\mathcal{W}}\mathcal{B}_{\mathcal{V}}^{-1} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is an isomorphism.

($\Longleftarrow$) This is part (iii) of Theorem 18.8. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.23 on p. 165. Theorem 3.15 on p. 156 gives another perspective on the isomorphism operator. Corollary 3.16 on p. 158 should be familiar from a first course in linear algebra.

18.11 Example. **(i)** Example 18.5 gave $\dim(\mathbb{P}^n) = n + 1$ directly by exhibiting a basis for $\mathbb{P}^n$. Part (i) of Example 12.15 showed that $\mathbb{P}^n$ and $\mathbb{F}^{n+1}$ are isomorphic, which gives another proof that $\dim(\mathbb{P}^n) = \dim(\mathbb{F}^{n+1}) = n + 1$.

(ii) Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces with $\dim(\mathcal{V}) = n$ and $\dim(\mathcal{W}) = m$. Then $\mathcal{V}$ and $\mathbb{F}^n$ are isomorphic, as are $\mathcal{W}$ and $\mathbb{F}^m$. By Theorem 13.1, $\mathbf{L}(\mathcal{V}, \mathcal{W})$ and $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ are isomorphic. And by part (ii) of Example 12.15, $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathbb{F}^{m \times n}$ are isomorphic. Thus $\mathbf{L}(\mathcal{V}, \mathcal{W})$ and $\mathbb{F}^{m \times n}$ are isomorphic, by Problem 13.6, and so

$$\dim(\mathbf{L}(\mathcal{V}, \mathcal{W})) = \dim(\mathbb{F}^{m \times n}) = mn$$

by part (iii) of Example 18.5.

18.12 Problem (!). Let $\mathcal{V}$ be a vector space.

(i) Suppose that $\mathcal{V}$ is finite-dimensional. Prove that $\mathcal{V}'$ is also finite-dimensional with

$\dim(\mathcal{V}) = \dim(\mathcal{V}')$.

(ii) Prove that if $\mathcal{V}'$ is infinite-dimensional, then so is $\mathcal{V}$.

The proof of Theorem 18.10 says that if $(v_1, \ldots, v_n)$ is a basis for the finite-dimensional space $\mathcal{V}$ and $(w_1, \ldots, w_n)$ is a basis for the finite-dimensional space $\mathcal{W}$, then the isomorphism $\mathcal{T} := \mathcal{B}_{\mathcal{W}}\mathcal{B}_{\mathcal{V}}^{-1}$ has the formula

$$\mathcal{T}\left(\sum_{j=1}^{n} \alpha_j v_j\right) = \sum_{j=1}^{n} \alpha_j w_j. \tag{18.1}$$

It turns out that given a basis $(v_1, \ldots, v_n)$ for a finite-dimensional space $\mathcal{V}$ and given a list $(w_1, \ldots, w_n)$ for the space $\mathcal{W}$, where the list $(w_1, \ldots, w_n)$ need not be a basis (or a spanning list, or independent) and $\mathcal{W}$ need not be finite-dimensional, then we can always construct $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ using the formula (18.1). This is an example of how bases *discretize* problems: they reduce consideration from an entire vector space to just a (very special) list of vectors.

18.13 Theorem (Extension by linearity: bases determine operators). *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$, with $\mathcal{V}$ finite-dimensional. Let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$ and let $(w_1, \ldots, w_n)$ be a list. There is a unique operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ such that $\mathcal{T}v_j = w_j$ for each j. Specifically, $\mathcal{T}$ is the operator*

$$\mathcal{T}\sum_{j=1}^{n} \alpha_j v_j = \sum_{j=1}^{n} \alpha_j w_j. \tag{18.2}$$

Proof. **1.** *Uniqueness.* First we should see why (18.2) is the "right" definition of $\mathcal{T}$. If $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ satisfies $\mathcal{T}v_j = w_j$, and if $v \in \mathcal{V}$ has the expansion $v = \sum_{j=1}^{n} \alpha_j v_j$, then

$$\begin{aligned} \mathcal{T}v &= \mathcal{T}\sum_{j=1}^{n} \alpha_j v_j \text{ by definition of } v \\ &= \sum_{j=1}^{n} \alpha_j \mathcal{T}v_j \text{ by the linearity of } \mathcal{T} \\ &= \sum_{j=1}^{n} \alpha_j w_j \text{ by the assumption on } \mathcal{T}. \end{aligned}$$

This also basically leads to a proof of uniqueness for $\mathcal{T}$. Suppose that $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ also satisfies $\mathcal{S}v_j = w_j$ for each j. We want to show that $\mathcal{T} = \mathcal{S}$, equivalently, that $\mathcal{T}v = \mathcal{S}v$ for each $v \in \mathcal{V}$. So, fix $v \in \mathcal{V}$ and write $v = \sum_{j=1}^{n} \alpha_j v_j$. Then

$$\begin{aligned} \mathcal{T}v &= \sum_{j=1}^{n} \alpha_j w_j \text{ by definition of } \mathcal{T} \\ &= \sum_{j=1}^{n} \alpha_j \mathcal{S}v_j \text{ by the assumption on } \mathcal{S} \end{aligned}$$

$$= \mathcal{S}\sum_{j=1}^n \alpha_j v_j \text{ by the linearity of } \mathcal{S}$$
$$= \mathcal{S}v \text{ by definition of } v.$$

This calculation did not use the linearity of $\mathcal{T}$ as defined in (18.2) This is good, because we have not yet established the linearity of $\mathcal{T}$, but it also suggests that doing so should be "easy." It "is."

2. *Existence: $\mathcal{T}$ is a function.* First, however, we need to check that (18.2) actually gives a function in $\mathcal{W}^{\mathcal{V}}$. We are saying that

$$\mathcal{T} := \left\{ \left(\sum_{j=1}^n \alpha_j v_j, \sum_{j=1}^n \alpha_j w_j \right) \;\middle|\; (\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n \right\}. \tag{18.3}$$

Does this give $\mathcal{T} \in \mathcal{W}^{\mathcal{V}}$? Be careful in that we are defining $\mathcal{T}v$ based on a "choice" from v: we are *choosing* to represent v as a linear combination of the list $(v_1, \ldots, v_n)$, and then we are using this choice of representation—the coefficients on v_j in that linear combination—to define $\mathcal{T}v$.

We first check that if $v \in \mathcal{V}$, then there is $w \in \mathcal{W}$ such that $(v, w) \in \mathcal{T}$. This is true because $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$. Given $v \in \mathcal{V}$, write $v = \sum_{j=1}^n \alpha_j v_j$, and then put $w = \sum_{j=1}^n \alpha_j w_j$. Then $(v, w) \in \mathcal{T}$.

Now suppose $(v, w_1), (v, w_2) \in \mathcal{T}$. We check that $w_1 = w_2$. Since $(v, w_1) \in \mathcal{T}$, by definition we have $v = \sum_{j=1}^n \alpha_j v_j$ and $w_1 = \sum_{j=1}^n \alpha_j w_j$ for some $\alpha_j \in \mathbb{F}$. And since $(v, w_2) \in \mathcal{T}$, we also have $v = \sum_{j=1}^n \beta_j v_j$ and $w_2 = \sum_{j=1}^n \beta_j w_j$ for some $\beta_j \in \mathbb{F}$. Since $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$, we have $\alpha_j = \beta_j$ for each j, and so $w_1 = w_2$.

3. *Linearity.* Finally, we check (part of) linearity. Let v, $\widetilde{v} \in \mathcal{V}$ with $v = \sum_{j=1}^n \alpha_j v_j$ and $\widetilde{v} = \sum_{j=1}^n \beta_j v_j$. Then $v + \widetilde{v} = \sum_{j=1}^n (\alpha_j + \beta_j) v_j$, so

$$\mathcal{T}(v + \widetilde{v}) = \sum_{j=1}^n (\alpha_j + \beta_j) w_j = \sum_{j=1}^n \alpha_j w_j + \sum_{j=1}^n \beta_j w_j = \mathcal{T}v + \mathcal{T}\widetilde{v}.$$

We leave the proof that $\mathcal{T}(\alpha v) = \alpha \mathcal{T} v$ as an exercise. ■

18.14 Problem (!). Finish the proof.

18.15 Problem (!). If the work showing that $\mathcal{T}$ defined by (18.3) feels like overkill, let $\mathcal{V} = \mathcal{W} = \mathbb{R}^2$ and put

$$\mathcal{T} = \left\{ (\alpha_1 \mathbf{e}_1 + \alpha_2(2\mathbf{e}_1), \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2) \;\middle|\; (\alpha_1, \alpha_2) \in \mathbb{F}^2 \right\}.$$

Explain all of the reasons why $\mathcal{T} \notin \mathcal{W}^{\mathcal{V}}$.

Content from *Linear Algebra* by Meckes & Meckes. "Extension by linearity" is Theorem 3.14 on p. 155.

18.16 Problem (!). Here is another manifestation of the notion that "bases determine operators." Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and $\mathcal{S}$, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$. Prove that $\mathcal{S} = \mathcal{T}$ if and only if $\mathcal{S}v_j = \mathcal{T}v_j$ for all j.

Day 19: Wednesday, February 25.

Now we develop some further properties of bases.

19.1 Theorem (Lists that cannot be bases). *Let $\mathcal{V}$ be a vector space.*

(i) *A list of length less than $\dim(\mathcal{V})$ cannot span $\mathcal{V}$. In particular, a list of length less than $\dim(\mathcal{V})$ cannot be a basis for $\mathcal{V}$.*

(ii) *Any independent list has length at most $\dim(\mathcal{V})$. Equivalently, any list of length greater than $\dim(\mathcal{V})$ is dependent. In particular, a list of length greater than $\dim(\mathcal{V})$ cannot be a basis for $\mathcal{V}$.*

Proof. **(i)** If a list of length $m < \dim(\mathcal{V})$ spans $\mathcal{V}$, then it can be reduced to a basis for $\mathcal{V}$. This reduced list has length $\dim(\mathcal{V}) \le m$, a contradiction.

(ii) Let $\dim(\mathcal{V}) = n$. Let $(u_1, \ldots, u_r)$ be an independent list in $\mathcal{V}$. Put $\mathcal{U} = \operatorname{span}(u_1, \ldots, u_r)$. Then $(u_1, \ldots, u_r)$ is a basis for $\mathcal{U}$. Let $\mathcal{B}_r \colon \mathbb{F}^r \to \mathcal{U}$ and $\mathcal{B}_n \colon \mathbb{F}^n \to \mathcal{V}$ be basis operators. Let $\mathcal{J} \colon \mathcal{U} \to \mathcal{V} \colon u \mapsto u$ be the "natural embedding" of $\mathcal{U}$ into $\mathcal{V}$. Then $\mathcal{B}_n^{-1}\mathcal{J}\mathcal{B}_r \in \mathbf{L}(\mathbb{F}^r, \mathbb{F}^n)$ is injective, so $r \le n$. ■

19.2 Problem (!). Let $\mathcal{V}$ be a vector space. State the following precisely and prove it: any independent list in $\mathcal{V}$ is no longer than any spanning list for $\mathcal{V}$.

Content from *Linear Algebra* by Meckes & Meckes. This problem and the preceding theorem are, effectively, Propositions 3.20 and 3.21 on p. 165.

19.3 Theorem (Lists that always are bases). *Let $\mathcal{V}$ be a vector space.*

(i) *An independent list of length equal to $\dim(\mathcal{V})$ is a basis for $\mathcal{V}$.*

(ii) *A spanning list of length equal to $\dim(\mathcal{V})$ is a basis for $\mathcal{V}$.*

Proof. **(i)** If such a list $(v_1, \ldots, v_n)$ with $n = \dim(\mathcal{V})$ is independent but not a basis, there is $v_{n+1} \in \mathcal{V} \setminus \operatorname{span}(v_1, \ldots, v_n)$. Then the list $(v_1, \ldots, v_n, v_{n+1})$ is independent with length $n + 1 > \dim(\mathcal{V})$, which cannot happen.

(ii) If this list is not a basis, it is dependent, and so it can be reduced to an independent list of length less than $\dim(\mathcal{V})$ with the same span: $\mathcal{V}$. This reduced list is therefore a basis for $\mathcal{V}$ and so has length $\dim(\mathcal{V})$, a contradiction. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.26 on p. 167 and Theorem 3.28 on p. 168. Read the example on p. 167.

19.4 Theorem (Characterization of infinite-dimensional spaces). *A vector space $\mathcal{V}$ is infinite-dimensional if and only if for each integer $n \ge 1$, there is an independent list of length n in $\mathcal{V}$.*

Proof. ($\Longrightarrow$) We induct on n. For $n = 1$, if $\mathcal{V} = \{0_{\mathcal{V}}\}$, then $\mathcal{V}$ is finite-dimensional. So, there is $v_1 \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$, and the list (v_1) is independent. Assume that for some $n \ge 1$, there is an independent list in $\mathcal{V}$ of length n. Now we show the existence of an independent list of length $n+1$. Say that the independent list of length n is $(v_1, \ldots, v_n)$. If $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$, then $\mathcal{V}$ is finite-dimensional. So, there is $v_{n+1} \in \mathcal{V} \setminus \operatorname{span}(v_1, \ldots, v_n)$. That is, $v_{n+1} \notin \operatorname{span}(v_1, \ldots, v_n)$, and so the list $(v_1, \ldots, v_n, v_{n+1})$ is independent.

($\Longleftarrow$) Suppose instead that $\mathcal{V}$ is finite-dimensional. If $\mathcal{V} = \{0_{\mathcal{V}}\}$, then there is no independent list in $\mathcal{V}$. If $m := \dim(\mathcal{V}) \ge 1$, then the hypothesis provides an independent list of length $m+1$ in $\mathcal{V}$. This is impossible, because any independent list in $\mathcal{V}$ has length at most m. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.18 on p. 163.

19.5 Problem ($\star$). For $j \in \mathbb{N}$, define $e_j \in \mathbb{F}^\infty$ by

$$e_j(k) := \begin{cases} 1, & j = k \\ 0, & j \ne k. \end{cases}$$

Prove that for any $n \ge 1$ the list $(e_1, \ldots, e_n)$ is independent and therefore $\mathbb{F}^\infty$ is infinite-dimensional. [Hint: *if $f := \sum_{j=1}^n \alpha_j e_j$, what is $f(k)$ for $1 \le k \le n$?*]

19.6 Problem ($\star$). Let $\mathcal{V}$ be a vector space such that the longest independent list in $\mathcal{V}$ has length n. Show that $\mathcal{V}$ is finite-dimensional and $\dim(\mathcal{V}) = n$.

19.7 Problem ($\star$). This problem expands the results of Theorem 18.8. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces.

(i) Let $\mathcal{V}$ be infinite-dimensional and suppose that there exists an injective linear operator $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$. Prove that $\mathcal{W}$ is infinite-dimensional.

(ii) Let $\mathcal{W}$ be infinite-dimensional and suppose that there exists a surjective linear operator $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$. Prove that $\mathcal{V}$ is infinite-dimensional.

19.8 Theorem (Any independent list can be extended to a basis). *Let $\mathcal{V}$ be a vector space and let $(v_1, \ldots, v_r)$ be an independent list in $\mathcal{V}$. Then one, and only one, of the following is true.*

(i) $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_r)$.

(ii) *There exist $v_{r+1}, \ldots, v_n \in \mathcal{V}$ such that $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$.*

(iii) *$\mathcal{V}$ is infinite-dimensional.*

Proof. Suppose that (i) and (ii) are false. We prove (iii) by showing that for any $m \geq 1$, there exist $v_{r+1}, \ldots, v_m \in \mathcal{V}$ such that the list $(v_1, \ldots, v_r, v_{r+1}, \ldots, v_m)$ is independent. And we do this by inducting on m.

For the base case $m = 1$, since (i) is false, we know that $\mathcal{V} \neq \operatorname{span}(v_1, \ldots, v_r)$, and so there is $v_{r+1} \in \mathcal{V} \setminus \operatorname{span}(v_1, \ldots, v_r)$. Then $(v_1, \ldots, v_r, v_{r+1})$ is independent. Now assume that for some $m \geq 1$, there are $v_{r+1}, \ldots, v_m \in \mathcal{V}$ such that $(v_1, \ldots, v_r, v_{r+1}, \ldots, v_m)$ is independent. Since (ii) is false, this list $(v_1, \ldots, v_r, v_{r+1}, \ldots, v_m)$ is not a basis for $\mathcal{V}$; since this list is already independent, we have $\mathcal{V} \neq \operatorname{span}(v_1, \ldots, v_r, v_{r+1}, \ldots, v_m)$. Then there is $v_{m+1} \in \mathcal{V} \setminus \operatorname{span}(v_1, \ldots, v_r, v_{r+1}, \ldots, v_m)$, and so $\operatorname{span}(v_1, \ldots, v_r, v_{r+1}, \ldots, v_m, v_{m+1})$ is independent. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.27 on p. 167.

19.9 Problem ($\star$). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces with $\mathcal{V}$ finite-dimensional and nonzero. Suppose that $\mathcal{U}$ is a subspace of $\mathcal{V}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{W})$. Prove that $\mathcal{T}$ can be "extended" to a linear operator $\widetilde{\mathcal{T}} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ such that $\widetilde{\mathcal{T}}u = \mathcal{T}u$ for all $u \in \mathcal{U}$. Is this extension unique? [Hint: *if $\mathcal{U} \neq \mathcal{V}$, start by extending a basis of $\mathcal{U}$ to a basis of $\mathcal{V}$.*]

Day 20: Friday, February 27.

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."

Dual basis (relative to a given basis), double (algebraic) dual space of a vector space

20.1 Theorem (Basis, dimension, and subspaces). *Let $\mathcal{V}$ be a finite-dimensional vector space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Then $\mathcal{U}$ is finite-dimensional and* $\dim(\mathcal{U}) \leq \dim(\mathcal{V})$ *with equality if and only if $\mathcal{U} = \mathcal{V}$.*

Proof. If $\mathcal{V} = \{0_{\mathcal{V}}\}$, then $\mathcal{U} = \mathcal{V}$, so $\mathcal{U}$ is finite-dimensional and $\dim(\mathcal{U}) = \dim(\mathcal{V})$. Let $\mathcal{V}$ be nonzero, so $\dim(\mathcal{V}) \geq 1$. Suppose that $\mathcal{U}$ is infinite-dimensional. Then there is an independent list in $\mathcal{U}$ of length $\dim(\mathcal{V}) + 1$, so there is an independent list in $\mathcal{V}$ of length

$\dim(\mathcal{V})+1$. This is impossible. Hence $\mathcal{U}$ is finite-dimensional and so has a basis $(u_1, \ldots, u_r)$ with $r = \dim(\mathcal{U})$. This basis is an independent list in $\mathcal{V}$, so $r \le \dim(\mathcal{V})$. If $r = \dim(\mathcal{V})$, then $(u_1, \ldots, u_r)$ is an independent list in $\mathcal{V}$ of length $\dim(\mathcal{V})$, so $(u_1, \ldots, u_r)$ is a basis for $\mathcal{V}$ as well. Then $\mathcal{U} = \operatorname{span}(u_1, \ldots, u_r) = \mathcal{V}$. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.29 on p. 168.

20.2 Problem (!). What is wrong with the following attempt to prove Theorem 20.1? Let $\mathcal{V}$ be a finite-dimensional nonzero vector space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Then $\mathcal{V}$ has a basis $(v_1, \ldots, v_n)$. Since $\mathcal{U} \subseteq \mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$, by Definition 13.11, $\mathcal{U}$ is finite-dimensional.

The following gives us practice with linear functionals and dual spaces via dimension counting. Let $\mathcal{V}$ be a finite-dimensional vector space. Then $\dim(\mathcal{V}) = \dim(\mathcal{V}')$ by Problem 18.12, so $\mathcal{V}$ and $\mathcal{V}'$ are isomorphic. We explore the consequences of this isomorphism.

1. If $\dim(\mathcal{V}) = n$, then there are bases $(v_1, \ldots, v_n)$ for $\mathcal{V}$ and $(\varphi_1, \ldots, \varphi_n)$ for $\mathcal{V}'$ of the same length. Any $v \in \mathcal{V}$ can be written as $v = \sum_{j=1}^n \alpha_j v_j$, while any $\varphi \in \mathcal{V}'$ has the form $\varphi = \sum_{k=1}^n \beta_k \varphi_k$. Then

$$\varphi(v) = \varphi\left(\sum_{j=1}^n \alpha_j v_j\right) = \sum_{j=1}^n \alpha_j \varphi(v_j) = \sum_{j=1}^n \alpha_j \left(\sum_{k=1}^n \beta_k \varphi_k(v_j)\right) = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \beta_k \varphi_k(v_j). \tag{20.1}$$

This is a pretty awful expression; uncharacteristically, bases have not made things simpler.

The problem is that we have not chosen the *right* bases here, or more precisely the right basis for $\mathcal{V}'$. Things *would* be much simpler if we had better control over $\varphi_k(v_j)$.

2. Start again with the basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$. Let $\mathcal{B}\colon \mathbb{F}^n \to \mathcal{V}$ be the associated basis operator. If $v \in \mathcal{V}$ has the form $v = \sum_{j=1}^n \alpha_j v_j$, then $\mathcal{B}^{-1}v = (\alpha_1, \ldots, \alpha_n)$. We can extract the jth component of this vector by taking dot products: $\alpha_j = (\mathcal{B}^{-1}v) \cdot \mathbf{e}_j$. Define

$$\varphi_j\colon \mathcal{V} \to \mathbb{F}\colon v \mapsto (\mathcal{B}^{-1}v) \cdot \mathbf{e}_j.$$

Then $\varphi_j \in \mathcal{V}'$ (why?) and any $v \in \mathcal{V}$ can be written as

$$v = \sum_{j=1}^n \varphi_j(v) v_j.$$

From the point of view of actually computing the coefficients of a vector with respect to this basis, we have learned nothing really new. From the point of view of language, now we can just say $v = \sum_{j=1}^n \varphi_j(v) v_j$ without specifying α_j, and that is sometimes faster.

We do know the values of these functionals on the basis vectors: since $\mathcal{B}^{-1}v_j = \mathbf{e}_j$ (why?), we have

$$\varphi_k(v_j) = (\mathcal{B}^{-1}v_j) \cdot \mathbf{e}_k = \mathbf{e}_j \cdot \mathbf{e}_k = \begin{cases} 1, & j = k \\ 0, & j \ne k. \end{cases}$$

Moreover, since $\dim(\mathcal{V}') = \dim(\mathcal{V}) = n$, and since $(\varphi_1, \ldots, \varphi_n)$ is a list of length n in $\mathcal{V}'$, we might wonder if $(\varphi_1, \ldots, \varphi_n)$ is a basis for $\mathcal{V}'$. It is.

3. To show that this list is a basis, we only need to check either its independence or that it spans $\mathcal{V}'$. (Why?) We check spanning here: given $\varphi \in \mathcal{V}'$, we want to write $\varphi = \sum_{k=1}^n \beta_k \varphi_k$. That is, we want $\varphi(v) = \sum_{k=1}^n \beta_k \varphi_k(v)$ for all $v \in \mathcal{V}$. Given $v \in \mathcal{V}$, write $v = \sum_{j=1}^n \alpha_j v_j$, so $\varphi_k(v) = \alpha_k$. Then, as in (20.1),

$$\varphi(v) = \sum_{j=1}^n \alpha_j \varphi(v_j) = \sum_{j=1}^n \varphi_j(v)\varphi(v_j).$$

So, $\varphi = \sum_{j=1}^n \varphi(v_j)\varphi_j$, as desired.

We have mostly proved the following result.

20.3 Theorem. *Let $\mathcal{V}$ be a finite-dimensional vector space with basis $(v_1, \ldots, v_n)$. There exists a unique basis $(\varphi_1, \ldots, \varphi_n)$ for $\mathcal{V}'$ such that*

$$\varphi_k(v_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j. \end{cases} \tag{20.2}$$

This basis $(\varphi_1, \ldots, \varphi_n)$ is the **DUAL BASIS FOR $\mathcal{V}'$ RELATIVE TO THE BASIS** $(v_1, \ldots, v_n)$ **FOR** $\mathcal{V}$. *Additionally, if $v \in \mathcal{V}$ and $\varphi \in \mathcal{V}'$, then*

$$v = \sum_{j=1}^n \varphi_j(v) v_j \quad \text{and} \quad \varphi = \sum_{j=1}^n \varphi(v_j)\varphi_j. \tag{20.3}$$

20.4 Problem (!). Prove that the dual basis is unique. That is, under the hypotheses and notation of Theorem 20.3, suppose that $(\psi_1, \ldots, \psi_n)$ is another basis for $\mathcal{V}'$ such that

$$\psi_k(v_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j. \end{cases}$$

Prove that $\varphi_k = \psi_k$ for all j. [Hint: *use Problem 18.16.*]

20.5 Problem ($\star$). Prove that the dual basis is an independent list (from scratch—without using that it is a basis). [Hint: *if $\sum_{j=1}^n \gamma_j \varphi_j = 0_{\mathcal{V}'}$, evaluate the left side at v_k for $k = 1, \ldots, n$.*]

20.6 Problem ($\star$). Here is a concrete example of a dual basis. Let $\mathcal{V} = \mathbb{P}^1$. For $p \in \mathcal{V}$, put

$$\varphi_1(p) := p(0) \quad \text{and} \quad \varphi_2(p) := \int_0^1 p(x)\, dx,$$

so φ_1, $\varphi_2 \in \mathcal{V}'$ (you do not have to prove this).

(i) Prove that (φ_1, φ_2) is independent and therefore (why?) a basis for $\mathcal{V}'$. [Hint: *if* α_1, $\alpha_2 \in \mathbb{F}$ *are such that* $\alpha_1\varphi_1(p) + \alpha_2\varphi_2(p) = 0$ *for all* $p \in \mathbb{P}^1$, *pick* p *as simply as possible.*]

(ii) Find a basis (p_1, p_2) for $\mathcal{V}$ such that (φ_1, φ_2) is the dual basis relative to that basis. [Hint: *the goal is that* $\varphi_j(p_k) = 1$ *for* $j = k$ *and* 0 *for* $j \neq k$*; this gives four equations, which nicely match the four (why?) unknowns that control the basis* (p_1, p_2).]

We have previously said that linear functionals can tell us a great deal of information about a vector space, and sometimes we can think of them as "instruments" that we apply to vectors in a space. Here is one such instance of this claim.

20.7 Lemma. *Let* $\mathcal{V}$ *be a finite-dimensional vector space and* $v \in \mathcal{V}$. *Then* $v = 0_{\mathcal{V}}$ *if and only if* $\varphi(v) = 0$ *for all* $\varphi \in \mathcal{V}'$.

Proof. First, if $v = 0_{\mathcal{V}}$, then $\varphi(v) = \varphi(0_{\mathcal{V}}) = 0$ for all $\varphi \in \mathcal{V}'$. Conversely, suppose that $v \in \mathcal{V}$ satisfies $\varphi(v) = 0$ for all $\varphi \in \mathcal{V}'$. If $\dim(\mathcal{V}) = 0$, then $\mathcal{V} = \{0_{\mathcal{V}}\}$, so $v = 0_{\mathcal{V}}$, and there is nothing to prove. For $\dim(\mathcal{V}) = n \geq 1$, let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$, and let $(\varphi_1, \ldots, \varphi_n)$ be the dual basis for $\mathcal{V}'$ relative to $(v_1, \ldots, v_n)$. In particular, then, $\varphi_j(v) = 0$ for each j, and so $v = \sum_{j=1}^{n} \varphi_j(v)v_j = 0_{\mathcal{V}}$. ■

The dual basis for the algebraic dual space of a finite-dimensional vector space is the "natural" way to relate the original space to its dual via a basis for the original space. However, that relation does not really inform us explicitly of what every functional in the dual space is; the second formula in (20.3) ultimately defines $\varphi \in \mathcal{V}'$ in terms of itself. We can obtain better control over a different, related space.

20.8 Definition. *Let* $\mathcal{V}$ *be a vector space. The* **DOUBLE (ALGEBRAIC) DUAL SPACE** *of* $\mathcal{V}$ *is* $\mathcal{V}'' := (\mathcal{V}')' = \mathbf{L}(\mathcal{V}', \mathbb{F})$.

If $\mathcal{V}$ is finite-dimensional, then we have $\dim(\mathcal{V}'') = \dim(\mathcal{V}') = \dim(\mathcal{V})$, and so $\mathcal{V}$ and $\mathcal{V}''$ are also isomorphic. What is interesting is not *that* $\mathcal{V}$ and $\mathcal{V}''$ are isomorphic but *how* they are isomorphic. There is a particular isomorphism that is in some sense "the best," and we study that now. Unsurprisingly, it relies on dual bases.

First, it will be helpful to paraphrase Theorem 20.3 with new notation. (When struggling with mathematical communication, sometimes a change in notation to avoid overworking a certain symbol is all that is needed to make things better.)

20.9 Theorem (Restatement of Theorem 20.3). *Let* $\mathcal{W}$ *be a finite-dimensional vector space with basis* $(w_1, \ldots, w_n)$. *There exists a unique basis* $(\psi_1, \ldots, \psi_n)$ *for* $\mathcal{W}'$ *such that*

$$\psi_k(w_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j. \end{cases} \tag{20.4}$$

Additionally, if $w \in \mathcal{W}$*, then*

$$w = \sum_{j=1}^{n} \psi_j(w) w_j, \tag{20.5}$$

and if $\psi \in \mathcal{W}'$*, then*

$$\psi = \sum_{j=1}^{n} \psi(w_j) \psi_j. \tag{20.6}$$

Here we will adopt the occasional custom of denoting an element of $\mathcal{V}'$ by φ' and of $\mathcal{V}''$ by φ''. So, $\varphi'(v) \in \mathbb{F}$ is defined for $v \in \mathcal{V}$, and likewise $\varphi''(\varphi') \in \mathbb{F}$ is defined for $\varphi' \in \mathcal{V}'$ and $\varphi'' \in \mathcal{V}''$. (The primes have nothing to do with derivatives.)

Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim(\mathcal{V}) = n \geq 1$. Start with a basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$. Then let $(\varphi_1', \ldots, \varphi_n')$ be the dual basis for $\mathcal{V}'$ relative to $(v_1, \ldots, v_n)$, and let $(\varphi_1'', \ldots, \varphi_n'')$ be the dual basis for $\mathcal{V}''$ relative to $(\varphi_1', \ldots, \varphi_n')$. So we have the identities

$$\varphi_k'(v_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \qquad \text{and} \qquad \varphi_\ell''(\varphi_k') = \begin{cases} 1, & \ell = k \\ 0, & \ell \neq k \end{cases} \tag{20.7}$$

as well as the representation

$$v = \sum_{j=1}^{n} \varphi_j'(v) v_j, \ v \in \mathcal{V}. \tag{20.8}$$

This, of course, is (20.5) with $w = v$, $\psi_j = \varphi_j'$, and $w_j = v_j$.

But for $\varphi' \in \mathcal{V}'$, we have two representations. The first follows from (20.6) by taking $\psi = \varphi'$, $w_j = v_j$, and $\psi_j = \varphi_j'$:

$$\varphi' = \sum_{j=1}^{n} \varphi'(v_j) \varphi_j'. \tag{20.9}$$

We should think of this representation of φ' as telling us *dynamically* what φ' *does*: it is a linear functional that acts on the space $\mathcal{V}$. After all, we obtained (20.6) (or, more precisely, its progenitor in (20.3) from Theorem 20.3) by evaluating φ' pointwise on the representation (20.8).

The second representation follows from (20.5) by taking $w = \varphi'$, $\psi_j = \varphi_j''$, and $w_j = \varphi_j'$:

$$\varphi' = \sum_{j=1}^{n} \varphi_j''(\varphi') \varphi_j'. \tag{20.10}$$

We should think of this representation of φ' as telling us *statically* what φ' *is*: it is a vector in the space $\mathcal{V}'$, and the list $(\varphi_1'', \ldots, \varphi_n'')$ contains the coordinate functionals for the basis $(\varphi_1', \ldots, \varphi_n')$ of $\mathcal{V}'$. Of course, what things do defines what things are, and this is not the first time that we have seen an object both dynamically and statically, for what it both does and is—much more generally than linear functionals, we think of linear operators as both acting on a vector space and belonging to a vector space of operators.

By the way, if it looks as though we are defining v and φ' in terms of themselves in (20.8), (20.9), and (20.10), we are, from a certain point of view. This is not wholly dissimilar from Taylor series—say, if $p(x) = \sum_{j=0}^{n} a_j x^j$, then since $a_j = p^{(j)}(0)/j!$, we also have $p(x) = \sum_{j=0}^{n}(p^{(j)}(0)/j!)x^j$, and so p is morally "defined in terms of itself."

Equating (20.9) and (20.10) gives two representations of $\varphi' \in \mathcal{V}$ in the span of the independent list $(\varphi_1', \ldots, \varphi_n')$:

$$\sum_{j=1}^{n} \varphi'(v_j)\varphi_j' = \sum_{j=1}^{n} \varphi_j''(\varphi')\varphi_j'.$$

Consequently, the coefficients are equal: $\varphi'(v_j) = \varphi_j''(\varphi')$. We record this for future use.

20.10 Lemma. *Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim(\mathcal{V}) = n \geq 1$. Let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$, let $(\varphi_1', \ldots, \varphi_n')$ be the dual basis for $\mathcal{V}'$ relative to $(v_1, \ldots, v_n)$, and let $(\varphi_1'', \ldots, \varphi_n'')$ be the dual basis for $\mathcal{V}''$ relative to $(\varphi_1', \ldots, \varphi_n')$. Then*

$$\varphi_j''(\varphi') = \varphi'(v_j), \ \varphi' \in \mathcal{V}'.$$

We can think of each v_j as "representing" φ_j'': the action of φ_j'' on a functional $\varphi' \in \mathcal{V}$ is just "evaluate at v_j." As linear functionals acting on linear functionals go, this is a fairly transparent action. It turns out that when $\mathcal{V}$ is finite-dimensional, all $\varphi'' \in \mathcal{V}''$ are "evaluate at" functionals.

Day 21: Monday, March 2.

First we check that the "evaluate at" functional is indeed a linear functional on $\mathcal{V}'$.

21.1 Lemma. *Let $\mathcal{V}$ be a vector space (not necessarily finite-dimensional). For $v \in \mathcal{V}$ and $\varphi' \in \mathcal{V}'$, put*

$$\varphi_v''(\varphi') := \varphi'(v).$$

Then $\varphi_v'' \in \mathcal{V}''$ and the map

$$\mathcal{J} \colon \mathcal{V} \to \mathcal{V}'' \colon v \mapsto \varphi_v'' \tag{21.1}$$

is linear.

Proof. **1.** First we show that $\mathcal{J}v \in \mathcal{V}'$ for any $v \in \mathcal{V}$. If φ_1', $\varphi_2' \in \mathcal{V}'$, then

$$\begin{aligned}(\mathcal{J}v)(\varphi_1' + \varphi_2') = \varphi_v''(\varphi_1' + \varphi_2') = (\varphi_1' + \varphi_2')(v) = \varphi_1'(v) + \varphi_2'(v) &= \varphi_v''(\varphi_1') + \varphi_v''(\varphi_2') \\ &= (\mathcal{J}v))(\varphi_1') + (\mathcal{J}v)(\varphi_2').\end{aligned}$$

We are using slightly more parentheses here than usual to emphasize that $\mathcal{J}v$ is a single functional in $\mathcal{V}''$. That $\varphi_v''(\alpha\varphi') = \alpha\varphi_v''(\varphi')$ is similar—all of this, really, is just how we define pointwise addition and scalar multiplication in $\mathbb{F}^{\mathcal{V}}$. Thus $\mathcal{J}v \in \mathcal{V}'$ for any $v \in \mathcal{V}$.

2. Now we check that $\mathcal{J} \in \mathbf{L}(\mathcal{V}, \mathcal{V}'')$. For v_1, $v_2 \in \mathcal{V}$, we want to show $\mathcal{J}(v_1+v_2) = \mathcal{J}v_1+\mathcal{J}v_2$. (Convince yourself that the previous work showing $(\mathcal{J}v)(\varphi_1' + \varphi_2') = (\mathcal{J}v))(\varphi_1') + (\mathcal{J}v)(\varphi_2')$ for $v \in \mathcal{V}$ and φ_1', $\varphi_2' \in \mathcal{V}'$ is not the same as this.) Recall what equality means here: we

want
$$(\mathcal{J}(v_1+v_2))(\varphi') = (\mathcal{J}v_1+\mathcal{J}v_2)(\varphi')$$
for all $\varphi' \in \mathcal{V}'$. On the left, we have
$$(\mathcal{J}(v_1+v_2))(\varphi') = \varphi'(v_1+v_2) = \varphi'(v_1)+\varphi'(v_2),$$
while on the right
$$(\mathcal{J}v_1+\mathcal{J}v_2)(\varphi') = (\mathcal{J}v_1)(\varphi')+(\mathcal{J}v_2)(\varphi') = \varphi'(v_1)+\varphi'(v_2).$$
This gives the desired equality, and showing $\mathcal{J}(\alpha v) = \alpha\mathcal{J}v$ is similar. ■

21.2 Problem (!). Complete the previous proof by showing that $(\mathcal{J}v)(\alpha\varphi') = \alpha(\mathcal{J}v)(\varphi')$ and $\mathcal{J}(\alpha v) = \alpha v$ for all $\alpha \in \mathbb{F}$, $v \in \mathcal{V}$, and $\varphi' \in \mathcal{V}$. (Convince yourself that these are actually distinct tasks.)

Now we prove that pairing vectors $v \in \mathcal{V}$ with the "evaluate at v" functional on $\mathcal{V}'$ does give an isomorphism between $\mathcal{V}$ and $\mathcal{V}''$ when $\mathcal{V}$ is finite-dimensional. There are other isomorphisms between these two spaces, but this is the most "natural" one.

21.3 Theorem. *Let $\mathcal{V}$ be a finite-dimensional vector space. The map $\mathcal{J} \in \mathbf{L}(\mathcal{V},\mathcal{V}'')$ defined in* (21.1) *is an isomorphism, called the* CANONICAL ISOMORPHISM *between $\mathcal{V}$ and $\mathcal{V}''$.*

Proof. 1. First we check that $\mathcal{J}$ is injective. Suppose $\mathcal{J}v = 0_{\mathcal{V}''} = 0_{\mathcal{V}'\to\mathbb{F}}$. We want to show that $v = 0_{\mathcal{V}}$. Let $\varphi' \in \mathcal{V}'$. Then
$$(\mathcal{J}\varphi')(v) = 0_{\mathcal{V}'\to\mathbb{F}}\varphi' = 0,$$
Also,
$$(\mathcal{J}\varphi')(v) = \varphi'(v),$$
by definition of $\mathcal{J}$. Hence
$$\varphi'(v) = 0$$
for any $\varphi' \in \mathcal{V}$. By Problem 20.7, this implies $v = 0_{\mathcal{V}}$.

2. Next we check that $\mathcal{J}$ is surjective. This is possibly the hardest step right now, and we will shortly find an easier way to do it. Let $\varphi'' \in \mathcal{V}''$. We want to find $v \in \mathcal{V}$ such that $\varphi'' = \mathcal{J}v$. That is, we want v to satisfy $\varphi''(\varphi') = \varphi'(v)$ for all $\varphi' \in \mathcal{V}'$. What v could do this?

Here it is helpful to introduce bases. Fix a basis $(v_1,\ldots,v_n)$ for $\mathcal{V}$. Let $(\varphi_1',\ldots,\varphi_n')$ be the dual basis for $\mathcal{V}'$ relative to $(v_1,\ldots,v_n)$. And let $(\varphi_1'',\ldots,\varphi_n'')$ be the dual basis for $\mathcal{V}''$ relative to $(\varphi_1',\ldots,\varphi_n')$.

We start by working backwards: let $\varphi'' \in \mathcal{V}''$ and suppose that there exists $v \in \mathcal{V}$ such that $\varphi''(\varphi') = \varphi'(v)$ for all $\varphi' \in \mathcal{V}'$. Such a v, like any vector in $\mathcal{V}$, has the expansion $v = \sum_{j=1}^n \varphi_j'(v)v_j$, and so we just need to determine $\varphi_j'(v_j)$. Since $\varphi'(v) = \varphi''(\varphi)$ for all $\varphi' \in \mathcal{V}'$, we are free to take $\varphi' = \varphi_j'$ to conclude $\varphi_j'(v) = \varphi''(\varphi_j')$. Our desired v is therefore $v = \sum_{j=1}^n \varphi''(\varphi_j')v_j$.

3. Here is the actual proof of surjectivity. Let $\varphi'' \in \mathcal{V}''$ and put $v = \sum_{j=1}^{n} \varphi''(\varphi_j')v_j$, so in particular $\varphi_j'(v) = \varphi''(\varphi_j')$. Let $\varphi' \in \mathcal{V}'$. We show that $\varphi''(\varphi') = \varphi'(v)$ by computing

$$\varphi''(\varphi') = \varphi''\left(\sum_{j=1}^{n} \varphi_j''(\varphi')\varphi_j'\right) \text{ by the representation } \varphi' = \sum_{j=1}^{n} \varphi_j''(\varphi')\varphi_j' \text{ from (20.10)}$$

$$= \sum_{j=1}^{n} \varphi_j''(\varphi')\varphi''(\varphi_j') \text{ by the linearity of } \varphi''$$

$$= \sum_{j=1}^{n} \varphi_j''(\varphi')\varphi_j'(v) \text{ by definition of } v$$

$$= \varphi'(v) \text{ again by the representation of } \varphi' \text{ from (20.10).}$$

■

21.4 Problem ($\star$). Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim(\mathcal{V}) = n \ge 1$, and let $(\varphi_1', \ldots, \varphi_n')$ be a basis for $\mathcal{V}'$. Prove that there exists a basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$ such that $(\varphi_1', \ldots, \varphi_n')$ is the dual basis relative to this basis for $\mathcal{V}$. That is, construct a basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$ such that

$$\varphi_j(v_k) = \begin{cases} 1, & j = k \\ 0, & j \ne k. \end{cases}$$

[Hint: *let* $(\varphi_1'', \ldots, \varphi_n'')$ *be the dual basis for* $\mathcal{V}''$ *relative to* $(\varphi_1', \ldots, \varphi_n')$*. Let* $v_k \in \mathcal{V}$ *satisfy* $\varphi_k''(\varphi') = \varphi'(v_k)$ *for any* $\varphi' \in \mathcal{V}'$*. Use the fact that the canonical isomorphism between* $\mathcal{V}$ *and* $\mathcal{V}''$ *is an isomorphism to prove that* $(v_1, \ldots, v_n)$ *is independent and therefore a basis for* $\mathcal{V}$*. For extra, optional practice, show directly that* $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$.]

Day 22: Wednesday, March 4.

It turns out that we can always work backward from an independent list $(\varphi_1, \ldots, \varphi_n)$ in $\mathcal{V}'$ to an independent list $(v_1, \ldots, v_n)$ in $\mathcal{V}$ such that the two lists meet the defining property (20.4) of a dual basis—except that we do not need to require that $\mathcal{V}$ is finite-dimensional. This requires an auxiliary lemma about linear functionals that will be useful later, too. And this lemma requires a representation of the most accessible and transparent functionals on finite-dimensional spaces possible.

22.1 Problem (!). Let $\varphi \in (\mathbb{F}^n)'$. Prove that there exists $\mathbf{w} \in \mathbb{F}^n$ such that $\varphi(\mathbf{v}) = \mathbf{v} \cdot \mathbf{w}$ for all $\mathbf{v} \in \mathbb{F}^n$. [Hint: *recall that* $\mathbf{v} = \sum_{j=1}^{n} (\mathbf{v} \cdot \mathbf{e}_j)\mathbf{e}_j$.]

22.2 Lemma. *Let $\mathcal{V}$ be a vector space and $\varphi_1, \ldots, \varphi_n \in \mathcal{V}'$. Suppose that $\cap_{j=1}^n \ker(\varphi_j) \subseteq \ker(\varphi)$ for some $\varphi \in \mathcal{V}'$. Then $\varphi = \sum_{j=1}^n \alpha_j \varphi_j$ for some $\alpha_j \in \mathbb{F}$.*

Proof. We prove the $n = 1$ and $n \geq 2$ cases separately. There is no real need for this separation, but a technique that arises in the $n = 1$ case will reappear later in other contexts, and it is worth seeing it early.

1. $n = 1$. If $\varphi_2 = 0_{\mathcal{V} \to \mathbb{F}}$, take $\alpha = 0$. Otherwise, let $v_0 \in \mathcal{V}$ with $\varphi_2(v_0) \neq 0$. Then $\varphi_1(v_0) \neq 0$ as well, as otherwise $v_0 \in \ker(\varphi_1) \subseteq \ker(\varphi_2)$. Also, $\varphi_2(v_0) = \alpha\varphi_1(v_0)$, and so, since $\varphi_1(v_0) \neq 0$, the only choice for α is

$$\alpha = \frac{\varphi_2(v_0)}{\varphi_1(v_0)}.$$

That is, we want to show

$$\varphi_2(v) = \frac{\varphi_2(v_0)}{\varphi_1(v_0)}\varphi_1(v)$$

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

This is equivalent to

$$\varphi_2(v) = \varphi_2\left(\frac{\varphi_1(v)}{\varphi_1(v_0)}v_0\right),$$

which in turn is equivalent to

$$\varphi_2\left(v - \frac{\varphi_1(v)}{\varphi_1(v_0)}v_0\right) = 0,$$

and that is equivalent to

$$v - \frac{\varphi_1(v)}{\varphi_1(v_0)}v_0 \in \ker(\varphi_2).$$

Since the vector under consideration only involves φ_1, and since $\ker(\varphi_1) \subseteq \ker(\varphi_2)$, it is natural to compute

$$\varphi_1\left(v - \frac{\varphi_1(v)}{\varphi_1(v_0)}v_0\right) = \varphi_1(v) - \frac{\varphi_1(v)}{\varphi_1(v_0)}\varphi_1(v_0) = \varphi_1(v) - \varphi_1(v) = 0.$$

Working backward, we have shown $\varphi_2 = \alpha\varphi_1$ with $\alpha = \varphi_2(v_0)/\varphi_1(v_0)$.

2. $n \geq 2$. Our goal is the representation $\varphi(v) = \sum_{j=1}^n \alpha_j \varphi_j(v)$ for each $v \in \mathcal{V}$. If we put $\mathbf{w} := (\overline{\alpha_1}, \ldots, \overline{\alpha_n}) \in \mathbb{F}^n$ and

$$\mathcal{T} \colon \mathcal{V} \to \mathbb{F}^n \colon v \mapsto (\varphi_1(v), \ldots, \varphi_n(v)),$$

then this representation of φ is equivalent to the identity $\varphi(v) = \mathcal{T}v \cdot \mathbf{w}$. The goal is then to find $\mathbf{w} \in \mathbb{F}^n$ that makes this identity true.

Since we know that we can represent functionals on all of $\mathbb{F}^n$ by the dot product, and since $\mathcal{T}(\mathcal{V})$ is a subspace of $\mathbb{F}^n$, we might ask if there is a functional $\psi \in (\mathbb{F}^n)'$ such that

$\psi(\mathcal{T}v) = \varphi(v)$. Since there would be $\mathbf{w} \in \mathbb{F}^n$ such that $\psi(\mathbf{v}) = \mathbf{v} \cdot \mathbf{w}$ for all $\mathbf{v} \in \mathbb{F}^n$, this would give us $\varphi(v) = \psi(\mathcal{T}v) = \mathcal{T}v \cdot \mathbf{w}$ for all $v \in \mathcal{V}$, as desired.

We construct this $\mathbf{w}$ by first defining a functional ψ on the subspace $\mathcal{T}(\mathcal{V})$ alone by

$$\psi\colon \mathcal{T}(\mathcal{V}) \to \mathbb{F}\colon \mathcal{T}v \mapsto \varphi(v). \tag{22.1}$$

The challenge here is that an element $w \in \mathcal{T}(\mathcal{V})$ might have the different representations $w = \mathcal{T}v_1$ and $w = \mathcal{T}v_2$ for different $v_1, v_2 \in \mathcal{V}$. How do we know, then, that $\varphi(v_1) = \varphi(v_2)$?

What we are really doing is defining

$$\psi = \{(\mathcal{T}v, \varphi(v)) \mid v \in \mathcal{V}\}. \tag{22.2}$$

One first shows that $\psi \in \mathbb{F}^{\mathcal{T}(\mathcal{V})}$ and then that $\psi \in (\mathcal{T}(\mathcal{V}))'$. We leave this as an exercise. Once this is established, by Problem 19.9 we may extend ψ to a functional on all of $\mathbb{F}^n$ and then represent this functional by the dot product, which gives the desired $\mathbf{w}$. ■

22.3 Problem ($\star$). Under the notation and hypotheses of Lemma 22.2, prove that ψ defined in (22.1) and (22.2) is a linear functional on $\mathcal{T}(\mathcal{V})$ as follows.

(i) Show that $\ker(\mathcal{T}) = \cap_{j=1}^{n} \ker(\varphi_j)$, so $\ker(\mathcal{T}) \subseteq \ker(\varphi)$.

(ii) Show that $\psi \in \mathbb{F}^{\mathcal{T}(\mathcal{V})}$. [Hint: *check that ψ meets the definition of a function—assume (w, α_1), $(w, \alpha_2) \in \psi$ and show that $\alpha_1 = \alpha_2$.*]

(iii) Show that $\psi \in (\mathcal{T}(\mathcal{V}))'$.

Day 23: Friday, March 6.

23.1 Theorem. *Let $(\varphi_1, \ldots, \varphi_n)$ be an independent list in $\mathcal{V}'$. There exists a list $(v_1, \ldots, v_n)$ in $\mathcal{V}$ such that*

$$\varphi_j(v_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases} \tag{23.1}$$

Proof. We induct on n.

1. *The base case $n = 1$.* Here the list (φ_1) is independent, so $\varphi_1 \neq 0_{\mathcal{V}\to\mathbb{F}}$, and so there is $v \in \mathcal{V}$ such that $\varphi_1(v) \neq 0$. Put $v_1 := v/\varphi_1(v)$ to find $\varphi_1(v_1) = 1$.

2. *The induction hypothesis and step.* Assume that the result is true for some $n \geq 1$, and now let $(\varphi_1, \ldots, \varphi_n, \varphi_{n+1})$ be an independent list in $\mathcal{V}'$. Apply the induction hypothesis to the independent list $(\varphi_1, \ldots, \varphi_n)$ in $\mathcal{V}'$ to produce a list $(v_1, \ldots, v_n)$ in $\mathcal{V}$ satisfying (23.1).

We just need to find $v_{n+1} \in \mathcal{V}$ satisfying $\varphi_{n+1}(v_{n+1}) = 1$ and $\varphi_j(v_{n+1}) = 0$ for $j = 1, \ldots, n$. It really suffices to find $w \in \mathcal{V}$ such that $\varphi_{n+1}(w) \neq 0$ and $\varphi_j(w) = 0$ for $j = 1, \ldots, n$, and then we can take $v_{n+1} = w/\varphi_{n+1}(w)$. Suppose that we cannot: what if for each $v \in \mathcal{V}$ with $\varphi_j(v) = 0$ for each j, we also have $\varphi_{n+1}(v) = 0$? Then $\cap_{j=1}^{n} \ker(\varphi_j) \subseteq \ker(\varphi_{n+1})$, and so the list $(\varphi_1, \ldots, \varphi_n, \varphi_{n+1})$ is dependent. ■

23.2 Problem (!). Under the notation and hypotheses of Theorem 23.1, prove that the list $(v_1, \ldots, v_n)$ is independent.

23.3 Problem ($\star$). Let $\mathcal{V}$ be a vector space. Problem 18.12 shows that if $\mathcal{V}'$ is infinite-dimensional, then so is $\mathcal{V}$. Use Theorem 23.1 and Problem 23.2 to give another proof of this result.

23.4 Problem (+). We have shown that if $\mathcal{V}$ is a finite-dimensional vector space, then $\mathcal{V}'$ is also finite-dimensional, and now we have two proofs that if $\mathcal{V}'$ is infinite-dimensional, then so is $\mathcal{V}$. It is also true that if $\mathcal{V}$ is infinite-dimensional, then $\mathcal{V}'$ is infinite-dimensional. (Take a moment to think about why none of our prior results immediately imply this.) However, proving this requires a notion of basis for infinite-dimensional spaces, which we will not pursue right now. Instead, suppose that $\mathcal{V}$ is infinite-dimensional and that the result of Problem 20.7 is still true: if $v \in \mathcal{V}$, then $v = 0_{\mathcal{V}}$ if and only if $\varphi(v) = 0$ for all $\varphi \in \mathcal{V}'$. (This can be shown to be true for any vector space, assuming an adequate notion of basis.) Assuming this, prove that $\mathcal{V}'$ is infinite-dimensional. [Hint: *as in the proof of Theorem 21.3, check that the canonical embedding $\mathcal{J}\colon \mathcal{V} \to \mathcal{V}''$ is injective. Use Problem 19.7 to conclude that $\mathcal{V}''$ is infinite-dimensional. What does this say about $\mathcal{V}'$?*]

Eigenvalues, eigenvectors, linear independence, and dimension are closely related. Here is a classical deployment of the linear independence lemma, which in its fullest form uses induction.

23.5 Example. A list of eigenvectors corresponding to distinct eigenvalues is independent. That is, assume $n \geq 2$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ have the distinct eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ with corresponding eigenvectors $v_1, \ldots, v_n \in \mathcal{V}$. So, $\lambda_j \neq \lambda_k$ for $j \neq k$, and $\mathcal{T}v_j = \lambda_j v_j$. Then the list $(v_1, \ldots, v_n)$ is independent.

1. We first show this for the simple case of $n = 3$. What if the list is dependent? Since the entries of the list are eigenvectors, none is $0_{\mathcal{V}}$, so in particular $v_1 \neq 0_{\mathcal{V}}$. So, it must be the case that either $v_2 \in \operatorname{span}(v_1)$ or $v_3 \in \operatorname{span}(v_1, v_2)$.

(i) In the former, we have $v_2 = \alpha_1 v_1$. In particular, $\alpha_1 \neq 0$, as otherwise $v_2 = 0_{\mathcal{V}}$. The only other thing that we know about v_1 and v_2 is how they talk with $\mathcal{T}$, so we apply $\mathcal{T}$ to both sides to get $\mathcal{T}v_2 = \alpha_1 \mathcal{T}v_1$, and thus $\lambda_2 v_2 = \alpha_1 \lambda_1 v_1$. Substitute $v_2 = \alpha_1 v_1$ on the left to find $\lambda_2 \alpha_1 v_1 = \alpha_1 \lambda_1 v_1$. Since $\alpha_1 \neq 0$, we may divide to find $\lambda_2 v_1 = \lambda_1 v_1$, thus $(\lambda_1 - \lambda_2)v_1 = 0_{\mathcal{V}}$. Since $\lambda_1 \neq \lambda_2$, we have $v_1 = 0_{\mathcal{V}}$, a contradiction. So, $v_2 \notin \operatorname{span}(v_1)$.

(ii) What if $v_3 \in \operatorname{span}(v_1, v_2)$? Then $v_3 = \alpha_1 v_1 + \alpha_2 v_2$. Apply $\mathcal{T}$ to both sides to obtain

$$\lambda_3 v_3 = \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2.$$

Substitute $v_3 = \alpha_1 v_1 + \alpha_2 v_2$ to obtain

$$\lambda_3 \alpha_1 v_1 + \lambda_3 \alpha_2 v_2 = \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2.$$

Rearrange to get

$$\alpha_1(\lambda_1 - \lambda_3)v_1 + \alpha_2(\lambda_2 - \lambda_3)v_2 = 0_{\mathcal{V}}.$$

We know $v_2 \notin \operatorname{span}(v_1)$ and $v_1 \neq 0_{\mathcal{V}}$, so the list (v_1, v_2) is independent. Thus

$$\alpha_1(\lambda_1 - \lambda_3) = \alpha_2(\lambda_2 - \lambda_3) = 0,$$

and since $\lambda_1 \neq \lambda_3$ and $\lambda_2 \neq \lambda_3$, we must have $\alpha_1 = \alpha_2 = 0$. Thus $v_3 = 0_{\mathcal{V}}$, another contradiction. So, $v_3 \notin \operatorname{span}(v_1, v_2)$.

(iii) We know that $v_1 \neq 0_{\mathcal{V}}$, $v_2 \notin \operatorname{span}(v_1)$, and $v_3 \notin \operatorname{span}(v_1, v_2)$. This proves the independence of the list (v_1, v_2, v_3).

2. Here is how the argument works in general. We induct on n. For the base case $n = 1$, the the list (v_1) has only one entry, which is nonzero, since v_1 is an eigenvector, and so this list is independent.

Assume that the result is true for some $n \geq 1$. Now let $(v_1, \ldots, v_n, v_{n+1})$ be a list of eigenvectors of $\mathcal{T}$ corresponding to distinct eigenvalues. If the whole list $(v_1, \ldots, v_n, v_{n+1})$ is dependent, then since $v_1 \neq 0_{\mathcal{V}}$ (again, because v_1 is an eigenvector), the linear independence lemma says that $v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$ for some $j \geq 2$. But $(v_1, \ldots, v_n)$ is a list of n eigenvectors of $\mathcal{T}$ corresponding to distinct eigenvalues, so by the induction hypothesis it is independent, and therefore $v_j \notin \operatorname{span}(v_1, \ldots, v_{j-1})$ for $2 \leq j \leq n$. The only possibility is that $v_{n+1} \in \operatorname{span}(v_1, \ldots, v_n)$.

Write $v_{n+1} = \sum_{j=1}^{n} \alpha_j v_j$, thus

$$\lambda_{n+1} v_{n+1} = \mathcal{T} v_{n+1} = \mathcal{T} \sum_{j=1}^{n} \alpha_j v_j = \sum_{j=1}^{n} \alpha_j \mathcal{T} v_j = \sum_{j=1}^{n} \alpha_j \lambda_j v_j.$$

Substitute $v_{n+1} = \sum_{j=1}^{n} \alpha_j v_j$ to find

$$\lambda_{n+1} \sum_{j=1}^{n} \alpha_j v_j = \sum_{j=1}^{n} \alpha_j \lambda_j v_j.$$

Rearrange to find

$$\sum_{j=1}^{n} \alpha_j (\lambda_j - \lambda_{n+1}) v_j = 0_{\mathcal{V}}.$$

By the independence of $(v_1, \ldots, v_n)$, we have $\alpha_j(\lambda_j - \lambda_{n+1}) = 0$. Since $\lambda_j \neq \lambda_{n+1}$, this implies $\alpha_j = 0$ for all j, thus $v_{n+1} = 0_{\mathcal{V}}$, a contradiction.

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.8 on pp. 146–147. Pages 388–389 in Appendix A.3 review proof by induction.

23.6 Example. Let $n \geq 1$ and $p_j(x) := x^j$ for $0 \leq j \leq n$. We can show the independence of the list $(p_0, \ldots, p_n)$ in $\mathcal{C}^\infty([0,1])$ by recognizing each polynomial as an eigenvector of one particular operator corresponding to distinct eigenvalues. What should this operator be?

Perhaps the operator that immediately comes to mind is differentiation, but

$$p_j'(x) = jx^{j-1} = jp_{j-1}(x)$$

for $1 \leq j \leq n$. Actually, this is still true at $j = 0$, but none of this gives p_j' as a scalar multiple of p_j.

However, if we multiply by that missing factor of x, we get $xp_j'(x) = jp_j(x)$ for $j \geq 1$. So, put $(\mathcal{T}f)(x) := xf'(x)$. For $1 \leq j \leq n$, we have

$$(\mathcal{T}p_j)(x) = xp_j'(x) = xjx^{j-1} = jx^j = jp_j(x),$$

while at $j = 0$ we have

$$(\mathcal{T}p_0)(x) = xp_0'(x) = x \cdot 0 = 0 = 0p_0(x).$$

Thus each p_j is an eigenvector of $\mathcal{T}$ corresponding to the eigenvalue j; these eigenvalues are distinct, and so the list is independent.

23.7 Problem (!). Prove that $\mathcal{C}^r([0,1])$ is infinite-dimensional for $1 \leq r \leq \infty$.

Content from *Linear Algebra* by Meckes & Meckes. For another perspective on this example, see the example on p. 146.

We have previously seen that an operator on a vector space over $\mathbb{R}$ need not have eigenvalues (Example 7.15), and likewise an operator on an infinite-dimensional space also need not have eigenvalues (Example 8.3, Problems 8.2 and 8.6), but also that an operator on an infinite-dimensional space can have infinitely many eigenvalues (Examples 8.1 and 8.5). None of this can cannot happen on a finite-dimensional vector space over $\mathbb{C}$.

23.8 Problem (!). Let $\mathcal{V}$ be be a finite-dimensional vector space over $\mathbb{F}$ (here we do not require $\mathbb{F} = \mathbb{C}$) and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. How many distinct eigenvalues can $\mathcal{T}$ have?

This is an upper bound on eigenvalues. Here is the lower bound: when the field is complex, an operator on a finite-dimensional vector always has at least one eigenvalue (in $\mathbb{C}$). To prove this, we need some results about polynomials.

Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Recall that we can take powers of $\mathcal{T}$: for integers $k \geq 0$, put

$$\mathcal{T}^k := \begin{cases} \mathcal{I}_{\mathcal{V}}, \ k = 0 \\ \mathcal{T}, \ k = 1 \\ \mathcal{T}^{k-1}\mathcal{T}, \ k \geq 2. \end{cases}$$

Then we can define an "operator-valued" polynomial. If $p(x) := \sum_{k=1}^{n} a_k x^k$ is a polynomial with coefficients in k, put

$$p(\mathcal{T}) := \sum_{k=1}^{n} a_k \mathcal{T}^k.$$

23.9 Problem ($\star$). Let $\mathbb{P}(\mathbb{F})$ denote the vector space of all polynomials (of any degree) with coefficients in $\mathbb{F}$. Let $\mathcal{V}$ be any vector space over $\mathbb{F}$ and fix $\mathcal{T} \in \mathcal{L}(V)$. Show that the map

$$f_{\mathcal{T}} \colon \mathbb{P}(\mathbb{F}) \to \mathbf{L}(\mathcal{V}) \colon p \mapsto p(\mathcal{T})$$

is linear.

Content from *Linear Algebra* by Meckes & Meckes. Page 217 discusses operator polynomials.

There is another useful way to express polynomials, and that nicely carries over to operator polynomials. Here we need product notation: if $w_1, \ldots, w_n \in \mathbb{C}$, then

$$\prod_{j=1}^{n} w_j := \begin{cases} w_1, & n = 1 \\ \left(\prod_{j=1}^{n-1} w_j\right) w_n, & n \geq 2. \end{cases}$$

With this notation, we state the fundamental theorem of algebra: every polynomial with complex coefficients factors into a product of linear factors with complex coefficients.

23.10 Theorem (Fundamental theorem of algebra). *Let $p(z) = \sum_{k=0}^{n} a_k z^k$ be a polynomial of degree n with coefficients in $\mathbb{C}$: $a_k \in \mathbb{C}$, $a_n \neq 0$. There is a list $(z_1, \ldots, z_n) \in \mathbb{C}^n$ such that*

$$p(z) = a_n \prod_{j=1}^{n} (z - z_j).$$

Content from *Linear Algebra* by Meckes & Meckes. Page 217 discusses the FTA.

For example, $z^2+1 = (z+i)(z-i)$. Thus every polynomial p has (at least) two expressions: the Taylor expansion $p(z) = \sum_{k=0}^{n} a_k z^k$ and the factored form above. The key difference is that even though all of the coefficients a_k may be real, some or all of the roots z_j may be complex. Just consider $p(z) = z^2 + 1$.

We can also consider arbitrary operator products. If $\mathcal{V}$ is a vector space and $(\mathcal{S}_1, \ldots, \mathcal{S}_n)$ is a list in $\mathbf{L}(\mathcal{V})$, we put

$$\prod_{j=1}^{n} \mathcal{S}_j := \begin{cases} \mathcal{S}_1, & n = 1 \\ \left(\prod_{j=1}^{n-1} \mathcal{S}_j\right), & n \geq 2. \end{cases}$$

Now let $\mathcal{V}$ be a vector space over $\mathbb{C}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. If a polynomial $p(z) = \sum_{k=0}^{n} a_k z^k$ with coefficients in $\mathbb{C}$ factors as

$$p(z) = a_n \prod_{j=1}^{n} (z - z_j),$$

do we have

$$p(\mathcal{T}) = a_n \prod_{j=1}^{n} (\mathcal{T} - z_j I)$$

as well?

23.11 Lemma. *Yes.*

Proof. We induct on n. When $n = 1$, we have $p(z) = a_1 z + a_0$ with $a_1 \neq 0$, thus $p(z) = a_1(z - (-a_0/a_1))$ as well. The same algebra shows

$$a_1 \mathcal{T} + a_0 \mathcal{I}_{\mathcal{V}} = a_1 \left(\mathcal{T} - \left(\frac{a_0}{a_1} \right) \mathcal{I}_{\mathcal{V}} \right).$$

Suppose the result is true for some $n \geq 1$. Now let p be a polynomial of degree $n+1$ and write p in two ways:

$$p(z) = \sum_{k=0}^{n+1} a_k z^k = a_{n+1} \prod_{j=1}^{n+1} (z - z_j).$$

Let

$$q(z) := a_{n+1} \prod_{j=1}^{n} (z - z_j),$$

so q is a polynomial of degree n, and therefore we can write

$$q(z) = \sum_{k=0}^{n} b_k z^k.$$

for some $b_k \in \mathbb{C}$. The induction hypothesis then implies

$$q(\mathcal{T}) = \sum_{k=0}^{n} b_k \mathcal{T}^k = a_{n+1} \prod_{j=1}^{n} (\mathcal{T} - z_j \mathcal{I}_{\mathcal{V}}),$$

and so

$$q(\mathcal{T})(\mathcal{T} - z_{n+1}\mathcal{I}_{\mathcal{V}}) = \left(a_{n+1} \prod_{j=1}^{n} (\mathcal{T} - z_j \mathcal{I}_{\mathcal{V}}) \right) (\mathcal{T} - z_{n+1}\mathcal{I}_{\mathcal{V}}) = a_{n+1} \prod_{j=1}^{n+1} (\mathcal{T} - z_j \mathcal{I}_{\mathcal{V}}).$$

If we can show that

$$p(\mathcal{T}) = q(\mathcal{T})(\mathcal{T} - z_{n+1}\mathcal{I}_{\mathcal{V}}), \tag{23.2}$$

then we will be done.

We do this in two passes. First, we rewrite

$$
\begin{aligned}
p(z) &= q(z)(z - z_{n+1}) \\
&= \sum_{k=0}^{n} b_k z^k (z - z_{n+1}) \\
&= \sum_{k=0}^{n} (b_k z^{k+1} - b_k z_{n+1} z^k) \\
&= \sum_{k=0}^{n} b_k z^{k+1} - \sum_{k=0}^{n} b_k z_{n+1} z^k \\
&= \sum_{k=1}^{n+1} b_{k-1} z^k - \sum_{k=0}^{n} b_k z_{n+1} z^k \\
&= -b_0 z_{n+1} + \sum_{k=1}^{n} (b_{k-1} - b_k z_{n+1}) z^k + b_n z^{n+1}.
\end{aligned}
\tag{23.3}
$$

Put

$$
c_k = \begin{cases} -b_0 z_{n+1}, \ k = 0 \\ b_{k-1} - b_k z_{n+1}, \ 1 \le k \le n \\ b_n, \ k = n+1, \end{cases}
$$

so we have shown

$$
\sum_{k=0}^{n+1} a_k z^k = p(z) = \sum_{k=0}^{n+1} c_k z^k.
$$

By uniqueness of a polynomial's coefficients, we have $a_k = c_k$. Thus

$$
p(\mathcal{T}) = \sum_{k=0}^{n+1} c_k \mathcal{T}^k.
$$

Second, the same algebra from (23.3) with z replaced by $\mathcal{T}$ shows

$$
\begin{aligned}
q(\mathcal{T})(\mathcal{T} - z_{n+1}\mathcal{I}_{\mathcal{V}}) &= \sum_{k=0}^{n} b_k \mathcal{T}^k (\mathcal{T} - z_{n+1}\mathcal{I}_{\mathcal{V}}) \\
&= -b_0 z_{n+1} I + \sum_{k=1}^{n} (b_{k-1} - b_k z_{n+1}) \mathcal{T}^k + b_n \mathcal{T}^{n+1} \ \text{(this is the fruit of (23.3))} \\
&= \sum_{k=0}^{n+1} c_k \mathcal{T}^k = \sum_{k=0}^{n+1} a_k \mathcal{T}^k = p(\mathcal{T}).
\end{aligned}
$$

This is the desired equality (23.2). ■

23.12 Problem (+). **(i)** Let $\mathcal{V}$ be a vector space and $\mathcal{S}, \mathcal{T} \in \mathbf{L}(\mathcal{V})$. Suppose that $\mathcal{S}$ and $\mathcal{T}$ commute: $\mathcal{S}\mathcal{T} = \mathcal{T}\mathcal{S}$. Prove that $\mathcal{S}\mathcal{T}$ is invertible if and only if both $\mathcal{S}$ and $\mathcal{T}$ are invertible. [Hint: *for any* $\mathcal{A} \in \mathbf{L}(\mathcal{V})$, *we have* $\mathcal{AST} = \mathcal{ATS}$ *and* $\mathcal{STA} = \mathcal{TSA}$.]

(ii) Let p be a polynomial, $\mathcal{V}$ be a finite-dimensional vector space over $\mathbb{C}$, and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Prove the **POLYNOMIAL SPECTRAL MAPPING THEOREM**: $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathcal{T}$ if and only if $p(\lambda)$ is an eigenvalue of $p(\mathcal{T})$. [Hint: *if p is constant, then* $p(\mathcal{T}) = p(0)\mathcal{I}_{\mathcal{V}}$. *Otherwise, let* $\lambda \in \mathbb{C}$, *factor* $p(z) - p(\lambda) = a\prod_{j=1}^{n}(z - z_j)$, *where* $n = \deg(p)$. *Explain why* $z_j = \lambda$ *for at least one* j. *Then explain why the following are equivalent: (i)* $p(\mathcal{T}) - \lambda\mathcal{I}_{\mathcal{V}}$ *is invertible, (ii)* $\prod_{j=1}^{n}(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})$ *is invertible, and (iii)* $\mathcal{T} - z_jI$ *is invertible for each* $1 \le j \le n$.]

Now here is why we care about operator polynomials: they are the key to proving that any linear operator on a finite-dimensional space has an eigenvalue. The proof of this fact is an abstraction of the following concrete situation.

23.13 Example. Let

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

We show that the linear operator $\mathcal{M}_A \colon \mathbb{C}^2 \to \mathbb{C}^2$ has eigenvalues in $\mathbb{C}$ (without using determinants).

Here is the trick. The list $(\mathbf{v}, \mathcal{M}_A\mathbf{v}, \mathcal{M}_A^2\mathbf{v})$ is linearly dependent in $\mathbb{C}^2$ for any $\mathbf{v} \in \mathbb{C}^2$, since the list has three entries, but $\dim(\mathbb{C}^2) = 2$, of course (when we consider $\mathbb{C}^2$ as a vector space over $\mathbb{C}$). For simplicity, we pick $\mathbf{v} = \mathbf{e}_1$, and we compute

$$A\mathbf{e}_1 = \mathbf{e}_2 \quad \text{and} \quad A^2\mathbf{e}_1 = A\mathbf{e}_2 = -\mathbf{e}_1.$$

Then the list is $(\mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_1)$, and the (hopefully obvious) linear dependence relationship is

$$1\mathbf{e}_1 + 0\mathbf{e}_2 + 1(-\mathbf{e}_1) = \mathbf{0}_2.$$

That is, we have the matrix-vector equation

$$A^2\mathbf{e}_1 + I_2\mathbf{e}_1 = \mathbf{0}_2,$$

and this is the same as the (somewhat more cumbersomely notated) operator-vector equation

$$(\mathcal{M}_A^2 + \mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 = \mathbf{0}_2.$$

Put $p(z) = z^2 + 1$. Then $p(\mathcal{M}_A)\mathbf{e}_1 = \mathbf{0}_2$, and since p factors as $p(z) = (z + i)(z - i)$, this also says that

$$(\mathcal{M}_A + i\mathcal{I}_{\mathbb{C}^2})(\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 = \mathbf{0}_2. \tag{23.4}$$

Now we consider cases.

First, if $(\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 = \mathbf{0}_2$, then $\mathcal{M}_A\mathbf{e}_1 = i\mathbf{e}_1$, so $\mathbf{e}_1$ would be an eigenvector of $\mathcal{M}_A$ corresponding to the eigenvalue i. Second, if $\mathbf{w} := (\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 \neq \mathbf{0}_2$, then (23.4) forces $(\mathcal{M}_A + i\mathcal{I}_{\mathbb{C}^2})[(\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1] = \mathbf{0}_2$. That is, $(\mathcal{M}_A + i\mathcal{I}_{\mathbb{C}^2})\mathbf{w} = \mathbf{0}_2$ and $\mathbf{w} \neq \mathbf{0}_2$, thus $\mathbf{w}$ is an eigenvector of $\mathcal{M}_A$ corresponding to the eigenvalue $-i$.

23.14 Problem (!). **(i)** Which is which? Compute $(\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1$ and decide if it $\mathbf{e}_1$ is an eigenvector corresponding to i, or if $(\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1$ is an eigenvector corresponding to $-i$.

(ii) Use the approach above to find the other eigenvalue. [Hint: *try* $\mathbf{v} = \mathbf{e}_2$.]

Content from *Linear Algebra* by Meckes & Meckes. Read the example on p. 219 and do Quick Exercise #32 on that page.