MATH 4260: LINEAR ALGEBRA II

Daily Log for Lectures and Readings

Timothy E. Faver

May 5, 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

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

and

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

and

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

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 $\mathcal{T} : \mathcal{V} \to \mathcal{W}$. (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 $f = \big\{(x,x^2) \bigm| x \in \mathbb{R}\big\}$. Let $I = [0,\infty)$. Show that $f(I) = I$.

(iii) Why is $\big\{(x,y) \bigm| x,\ y \in \mathbb{R} \text{ and } y^2 = x\big\}$ 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,

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

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 $f\colon I \to \mathbb{R}$ 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

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

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.

3.9 Remark. *The function that is* **IDENTICALLY ZERO** *is the zero vector in* $\mathbb{F}^X$. *This function* $0_{\mathbb{F}^X}$ *satisfies* $0_{\mathbb{F}^X}(x) = 0$ *for all* $x \in \mathcal{X}$. *A function* $f \colon X \to \mathbb{F}$ *is* **NONZERO** *if* $f(x) \neq 0$ *for at least one* $x \in X$*; such a function* f *may satisfy* $f(y) = 0$ *for some other* $y \in X$. *This is the same as how we say that a vector or matrix is* **NONZERO** *if at least one entry of that vector or matrix is nonzero, although not all entries need to be nonzero.*

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.

3.10 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:

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

3.11 Problem (!). Complete the proof of the uniqueness of the additive inverse begun in part (v) of Example 3.10. [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 (⋆). 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 $p(x) = a_2 x^2 + a_1 x + a_0$ and $q(x) = b_2 x^2 + b_1 x + b_0$, 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*

→

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

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

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*

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

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:

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

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 $\mathbb{P}^1 \subseteq \mathbb{P}^2$.

5.11 Problem (!). Fix $g \in \mathcal{C}([0,1])$ and define

$$\mathcal{T} \colon \mathcal{C}([0,1]) \to \mathcal{C}([0,1]) \colon f \mapsto f \circ g.$$

Prove that f is linear. (Such an operator $\mathcal{T}$ is called a COMPOSITION OPERATOR.) Is $\mathcal{S}f := g \circ f$ still linear?

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

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

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

(i) The "evaluate at 0" map

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

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 $\varphi \colon \mathcal{V} \to \mathbb{F}$ 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

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

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

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

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

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

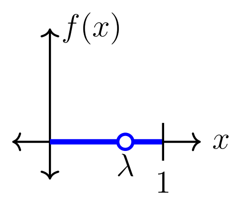
$$x - \lambda = 0 \quad \text{or} \quad f(x) = 0, \tag{8.1}$$

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 $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ 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

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

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

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

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 $\alpha \in \mathbb{F}$.

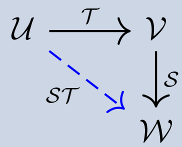
Content from *Linear Algebra* by Meckes & Meckes. Pages 81–82 discuss the vector space structure of $\mathbf{L}(\mathcal{V}, \mathcal{W})$.

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 $\mathcal{S}0_{\mathcal{U}\to\mathcal{V}}$?

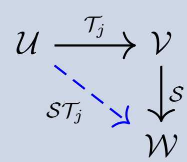
9.11 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathbb{F}$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. What are $\mathcal{I}_{\mathcal{W}}\mathcal{T}$ and $\mathcal{T}\mathcal{I}_{\mathcal{V}}$?

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 $\mathbb{F}$.*

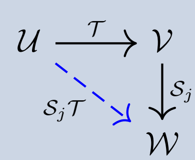
(i) *If $\mathcal{T}_1$, $\mathcal{T}_2 \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, then*

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



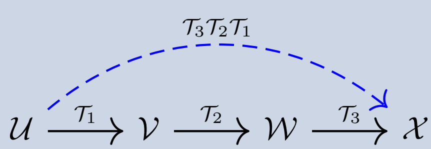
(ii) *If $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{S}_1$, $\mathcal{S}_2 \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, then*

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



(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

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

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

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 COMMUTATIVE 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

$$\mathcal{S}_1\mathcal{S}_2 A = \mathcal{S}_1\mathcal{M}_A = [\mathcal{M}_A] = A. \tag{10.3}$$

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

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

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*

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

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 ($\star$). 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 ($\star$). 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 (⋆). **(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 (⋆). 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 (⋆). 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 $p(x) = a_1x + a_0$ 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

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

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 $p(x) = a_1x + a_0$, 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

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

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 $\mathcal{T} = \mathcal{M}_{[\mathcal{T}]} = \mathcal{S}[\mathcal{T}]$.

The fundamental utility of an isomorphism is that it relates a given vector space to another, possibly simpler or "friendlier," vector space in a one-to-one, onto manner. The given space has exactly the same behavior as the other space, but cosmetically one may be more tractable than the other—contrast the ostensible variety of possible operators in $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ with the comfortingly familiar structure of matrices in $\mathbb{F}^{m\times n}$.

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.*]

12.17 Problem ($\star$). Given the frequency with which we employ the differentiation and antidifferentiation operators as examples, we might wonder how else they could interact. Let $\mathcal{V} = \mathcal{C}^\infty([0,1])$. Define $\mathcal{T}\colon \mathcal{V} \to \mathcal{V}$ by

$$(\mathcal{T}f)(x) := f'(x) + \int_0^x f(s)\, ds.$$

(i) Prove that if $\mathcal{T}f = g$ for some f, $g \in \mathcal{V}$, then $f'(0) = g(0)$.

(ii) Prove that if $\mathcal{T}f = g$, then f solves the ordinary differential equation

$$f'' + f = g',$$

and this can be solved for f in terms of $f(0)$, $f'(0)$, and g' via the dreaded method of variation of parameters. Conversely, show that any solution f to this ODE meeting $f'(0) = g(0)$ solves $\mathcal{T}f = g$. Conclude that $\mathcal{T}$ is surjective.

(iii) Let $f_0(x) = \sin(x)$. Prove that $\ker(\mathcal{T}) = \operatorname{span}(f_0)$. (Along the way, explain why the cosine is not in the kernel.) By thinking of $\ker(\mathcal{T})$ as the set of functions "whose derivatives are the negatives of their antiderivatives," explain why this result is unsurprising.

Day 13: Wednesday, February 11.

Vocabulary from today

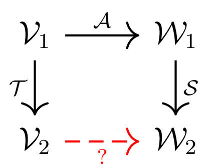
You should memorize the definition of each term, phrase, or concept below and be able to provide a concrete example of each and a nonexample for those marked "N."

Finite-dimensional vector space, infinite-dimensional vector space

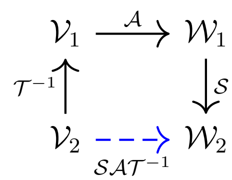
Here is an exercise in diagram-chasing.

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.



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



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

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

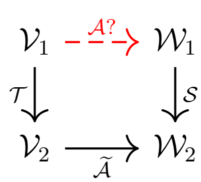
We need to check that $\mathcal{L}$ is linear and bijective. For linearity, we compute

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

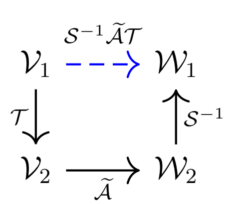
This is both kinds of distribution for operator composition. We leave checking $\mathcal{L}(\alpha\mathcal{A}) = \alpha\mathcal{L}\mathcal{A}$ as an exercise.

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 $\mathcal{L}\mathcal{A} = \widetilde{\mathcal{A}}$.



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 $\mathcal{A} = \mathcal{S}^{-1}\widetilde{\mathcal{A}}\mathcal{T}$.



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.

13.13 Example. Here is an important abstract example of how one finite-dimensional space induces another. Let $\mathcal{V}$ be a finite-dimensional vector space, $\mathcal{W}$ be any vector space, and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. We claim that $\mathcal{T}(\mathcal{V})$ is finite-dimensional. Write $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$ for some list $(v_1, \ldots, v_n)$ in $\mathcal{V}$. Then if $v \in \mathcal{V}$, we have $v = \sum_{j=1}^n \alpha_j v_j$ for some $\alpha_j \in \mathbb{F}$, and so

$$\mathcal{T}v = \mathcal{T}\sum_{j=1}^n \alpha_j v_j = \sum_{j=1}^n \alpha_j \mathcal{T}v_j.$$

Thus $\mathcal{T}v \in \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$, and so $\mathcal{T}(\mathcal{V}) \subseteq \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$. Conversely, since $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ is a list in the subspace $\mathcal{T}(\mathcal{V})$, we have $\operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n) \subseteq \mathcal{T}(\mathcal{V})$. We conclude $\mathcal{T}(\mathcal{V}) = \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$.

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

13.14 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.

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

13.16 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 $[v]_{\mathcal{B}} := \mathcal{B}^{-1} v \in \mathbb{F}^n$. 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

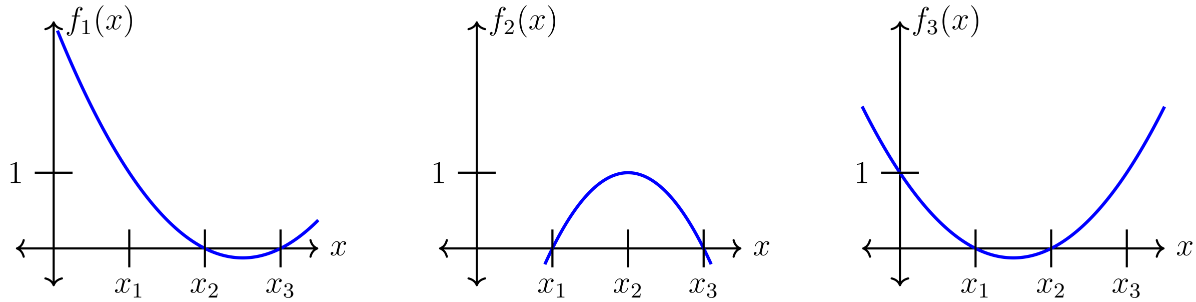
$$f_j(x_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

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.

14.20 Problem (!). Let $\mathcal{V}$ be a vector space and suppose that $(\mathcal{T}_1, \ldots, \mathcal{T}_n)$ is a list in $\mathbf{L}(\mathcal{V})$ such that $\mathcal{T}_j\mathcal{T}_k = 0_{\mathcal{V}\to\mathcal{V}}$ when $j \neq k$ and $\mathcal{T}_k^2 \neq 0_{\mathcal{V}\to\mathcal{V}}$ for all k. Prove that $(\mathcal{T}_1, \ldots, \mathcal{T}_n)$ is independent. [Hint: *if* $\sum_{j=1}^n \alpha_j \mathcal{T}_j = 0_{\mathcal{V}\to\mathcal{V}}$, *what is* $\mathcal{T}_k \sum_{j=1}^n \alpha_j \mathcal{T}_j = 0_{\mathcal{V}\to\mathcal{V}}$ *for any* $k = 1, \ldots, n$*?*]

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.

(iii) Now suppose that $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, and $(w_1, \ldots, w_n)$ is an independent list in $\mathcal{W}$. Suppose also that for each $j = 1, \ldots, n$, there is $v_j \in \mathcal{V}$ such that $\mathcal{T}v_j = w_j$. Prove that the list $(v_1, \ldots, v_n)$ is independent in $\mathcal{V}$.

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 $\mathcal{W} \neq \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$.

15.7 Problem ($\star$). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Suppose that $(w_1, \ldots, w_n)$ is an independent list in $\mathcal{W}$ and that for each $j = 1, \ldots, n$, there is $v_j \in \mathcal{V}$ such that $\mathcal{T}v_j = w_j$. Prove that the list $(v_1, \ldots, v_n)$ is independent in $\mathcal{V}$.

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.

15.8 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.) ■

15.9 Problem (!). **(i)** Prove the claim in Lemma 15.8 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 15.8 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 15.8, 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

$$\mathcal{J}_n \colon \mathbb{F}^n \to \mathbb{F}^{n+1} \colon \mathbf{v} \mapsto (\mathbf{v}, 0).$$

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, $\mathcal{T}\mathcal{J}_n \colon \mathbb{F}^n \to \mathbb{F}^m$ is injective and not surjective. By Lemma 15.8, $m \geq 2$ 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

$$\begin{aligned} \mathcal{T}v &= \sum_{j=1}^{n} \alpha_j w_j \text{ by definition of } \mathcal{T} \\ &= \sum_{j=1}^{n} \alpha_j \mathcal{S}v_j \text{ by the assumption on } \mathcal{S} \\ &= \mathcal{S}\sum_{j=1}^{n} \alpha_j v_j \text{ by the linearity of } \mathcal{S} \\ &= \mathcal{S}v \text{ by definition of } v. \end{aligned}$$

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 $n \geq 1$, 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 $n \geq 1$, 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 $m := \dim(\mathcal{V}) \geq 1$, 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

$$e_j(k) := \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

Prove that for any $n \geq 1$ 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* $1 \leq k \leq n$*?*]

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* $\dim(\mathcal{U}) \le \dim(\mathcal{V})$ *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 $\dim(\mathcal{V}) \ge 1$. 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 $\mathcal{U} = \mathrm{span}(u_1, \ldots, u_r) = \mathcal{V}$. ■

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 $\mathcal{U} \subseteq \mathcal{V} = \mathrm{span}(v_1, \ldots, v_n)$, 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

$$\varphi(v) = \varphi\left(\sum_{j=1}^n \alpha_j v_j\right) = \sum_{j=1}^n \alpha_j \varphi(v_j) = \sum_{j=1}^n \alpha_j \left(\sum_{k=1}^n \beta_k \varphi_k(v_j)\right) = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \beta_k \varphi_k(v_j). \quad (20.1)$$

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

$$\varphi_j\colon \mathcal{V} \to \mathbb{F}\colon v \mapsto (\mathcal{B}^{-1}v) \cdot \mathbf{e}_j. \tag{20.2}$$

That is,

$$\mathcal{B}^{-1}v = (\varphi_1(v), \ldots, \varphi_n(v)).$$

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

$$\varphi_k(v_j) = (\mathcal{B}^{-1}v_j) \cdot \mathbf{e}_k = \mathbf{e}_j \cdot \mathbf{e}_k = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

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*

$$\varphi_k(v_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j. \end{cases} \tag{20.3}$$

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*

$$v = \sum_{j=1}^{n} \varphi_j(v)v_j \quad \text{and} \quad \varphi = \sum_{j=1}^{n} \varphi(v_j)\varphi_j. \tag{20.4}$$

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).*]

20.7 Problem ($\star$). Let $\mathcal{V}$ be a finite-dimensional vector space with basis $(v_1, \ldots, v_n)$, and let $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ be the associated basis operator. Show that the associated dual basis $(\varphi_1, \ldots, \varphi_n)$ satisfies $\varphi_j(v) = \mathcal{B}^{-1}v \cdot \mathbf{e}_j$.

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.

20.8 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 (20.4) ultimately defines $\varphi \in \mathcal{V}'$ in terms of itself. We can obtain better control over a different, related space.

20.9 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.)

20.10 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*

$$\psi_k(w_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j. \end{cases} \tag{20.5}$$

Additionally, if $w \in \mathcal{W}$, then

$$w = \sum_{j=1}^{n} \psi_j(w) w_j, \tag{20.6}$$

and if $\psi \in \mathcal{W}'$, then

$$\psi = \sum_{j=1}^{n} \psi(w_j) \psi_j. \tag{20.7}$$

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

$$\varphi_k'(v_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \qquad \text{and} \qquad \varphi_\ell''(\varphi_k') = \begin{cases} 1, & \ell = k \\ 0, & \ell \neq k \end{cases} \tag{20.8}$$

as well as the representation

$$v = \sum_{j=1}^n \varphi_j'(v) v_j, \ v \in \mathcal{V}. \tag{20.9}$$

This, of course, is (20.6) 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 (20.7) by taking $\psi = \varphi'$, $w_j = v_j$, and $\psi_j = \varphi_j'$:

$$\varphi' = \sum_{j=1}^n \varphi'(v_j)\varphi_j'. \tag{20.10}$$

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 (20.7) (or, more precisely, its progenitor in (20.4) from Theorem 20.3) by evaluating φ' pointwise on the representation (20.9).

The second representation follows from (20.6) by taking $w = \varphi'$, $\psi_j = \varphi_j''$, and $w_j = \varphi_j'$:

$$\varphi' = \sum_{j=1}^n \varphi_j''(\varphi')\varphi_j'. \tag{20.11}$$

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 (20.9), (20.10), and (20.11), 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 (20.10) and (20.11) 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.

20.11 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 20.8, 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

$$\varphi''(\varphi') = \varphi''\left(\sum_{j=1}^n \varphi_j''(\varphi')\varphi_j'\right) \text{ by the representation } \varphi' = \sum_{j=1}^n \varphi_j''(\varphi')\varphi_j' \text{ from (20.11)}$$

$$= \sum_{j=1}^n \varphi_j''(\varphi')\varphi''(\varphi_j') \text{ by the linearity of } \varphi''$$

$$= \sum_{j=1}^{n} \varphi_j''(\varphi')\varphi_j'(v) \text{ by definition of } v$$

$$= \varphi'(v) \text{ again by the representation of } \varphi' \text{ from (20.11).}$$ ■

21.4 Problem ($\star$). Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim(\mathcal{V}) = n \geq 1$, 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

$$\varphi_j(v_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

[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 (20.5) 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. ∎

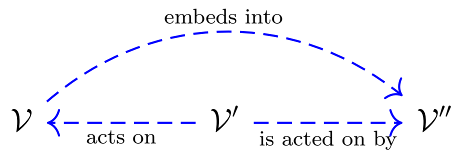
23.2 Problem (!). Under the notation and hypotheses of Theorem 23.1, prove that the list $(v_1, \ldots, v_n)$ is independent.

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 20.8 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* $\mathcal{V}'$*?*]

The "duality" of the dual space $\mathcal{V}'$ comes in part from its extensive "dual" role as a space of linear operators acting on $\mathcal{V}$ and as a space of vectors on which $\mathcal{V}''$ acts.



We have seen this duality before—the space of operators $\mathbf{L}(\mathcal{V}, \mathcal{W})$ between vector spaces $\mathcal{V}$ and $\mathcal{W}$ is "dynamic" in that its elements act on $\mathcal{V}$ and "static" in that its elements are vectors themselves. (We might call this the **HOT DOG PRINCIPLE**: hot dogs consist of chopped meat encased in animal intestines, so when we eat hot dogs, they stay hot dogs, and we ourselves become hot dogs.)

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 $v_2 \notin \text{span}(v_1)$ 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, $v_3 \notin \text{span}(v_1, v_2)$.

(iii) We know that $v_1 \neq 0_{\mathcal{V}}$, $v_2 \notin \text{span}(v_1)$, and $v_3 \notin \text{span}(v_1, v_2)$. 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 $v_j \in \text{span}(v_1, \ldots, v_{j-1})$ 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 $v_j \notin \text{span}(v_1, \ldots, v_{j-1})$ for $2 \leq j \leq n$. The only possibility is that $v_{n+1} \in \text{span}(v_1, \ldots, v_n)$.

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

$$\prod_{j=1}^{n} \mathcal{S}_j := \begin{cases} \mathcal{S}_1, \ n = 1 \\ (\prod_{j=1}^{n-1} \mathcal{S}_j), \ n \geq 2. \end{cases}$$

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 $\mathcal{A}\mathcal{S}\mathcal{T} = \mathcal{A}\mathcal{T}\mathcal{S}$ and $\mathcal{S}\mathcal{T}\mathcal{A} = \mathcal{T}\mathcal{S}\mathcal{A}$.*]

(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.

Day 24: Monday, March 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."

Finite-rank operator (N), rank of an operator

The trick of Example 23.13 generalizes substantially.

24.1 Theorem. *Let $\mathcal{V}$ be a nonzero finite-dimensional vector space over $\mathbb{C}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Then $\mathcal{T}$ has an eigenvalue: there exist $\lambda \in \mathbb{C}$ and $v \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$ such that $\mathcal{T}v = \lambda v$.*

Proof. Let $v \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$ and let $n = \dim(\mathcal{V})$. Consider the list $(v, \mathcal{T}v, \mathcal{T}^2 v, \ldots, \mathcal{T}^n v)$ of length $n+1$ whose jth entry is $\mathcal{T}^j v$. Since this list has length $n+1 > \dim(\mathcal{V})$, it must be linearly dependent. Then there is a list $(\alpha_0, \ldots, \alpha_{n+1}) \in \mathbb{F}^{n+1} \setminus \{\mathbf{0}_{n+1}\}$ such that

$$\sum_{j=0}^{n+1} \alpha_j \mathcal{T}^j v = 0_{\mathcal{V}}.$$

Let $m \le n+1$ be the largest index such that $\alpha_m \ne 0$. That is, $\alpha_m \ne 0$ and if $j \ge m+1$, then $\alpha_j = 0$. (If $m = n+1$, then this second condition is irrelevant.) This gives

$$\sum_{j=0}^{m} \alpha_j \mathcal{T}^j v = 0_{\mathcal{V}},$$

too. Write

$$p(z) = \sum_{j=0}^{m} \alpha_j z^j,$$

so $p(\mathcal{T})v = 0_{\mathcal{V}}$, and factor

$$p(z) = \alpha_m \prod_{j=1}^{m} (z - z_j)$$

for some $z_j \in \mathbb{C}$. Since $\alpha_m \neq 0$, we have

$$\prod_{j=1}^{m}(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})v = 0_{\mathcal{V}}.$$

There are now two possibilities. First, $(\mathcal{T} - z_m\mathcal{I}_{\mathcal{V}})v = 0_{\mathcal{V}}$. Then $\mathcal{T}v = z_m v$, and since $v \neq 0_{\mathcal{V}}$, this shows that v is an eigenvector of $\mathcal{T}$ with eigenvalue λ.

Otherwise, $(\mathcal{T} - z_m\mathcal{I}_{\mathcal{V}})v \neq 0_{\mathcal{V}}$. Let $r \geq 1$ be the largest index such that

$$\prod_{j=r}^{m}(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})v \neq 0_{\mathcal{V}}.$$

Such an index exists because $(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})v \neq 0_{\mathcal{V}}$. Certainly $r \leq m$, but also $r \geq 2$, as otherwise we would have

$$\prod_{j=1}^{m}(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})v \neq 0_{\mathcal{V}}.$$

Since r is the largest such index, we have

$$\prod_{j=r-1}^{m}(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})v = 0_{\mathcal{V}}.$$

Then

$$0_{\mathcal{V}} = \prod_{j=r-1}^{m}(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})v = (\mathcal{T} - z_{r-1}\mathcal{I}_{\mathcal{V}})\prod_{j=r}^{m}(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})v.$$

This shows that $\prod_{j=r}^{m}(\mathcal{T} - z_j\mathcal{I}_{\mathcal{V}})v$ is an eigenvector for $\mathcal{T}$ with eigenvalue z_{r-1}. ■

Content from *Linear Algebra* by Meckes & Meckes. Proposition 3.66 in the book. Axler opines in *Linear Algebra Done Right* (2025) that "The main reason that a richer theory exists for operators [in $\mathbf{L}(\mathcal{V})$] than for more general linear maps [in $\mathbf{L}(\mathcal{V}, \mathcal{W})$] is that operators [in $\mathbf{L}(\mathcal{V})$] can be raised to powers" (p. 137). Being able to raise $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ to nonnegative integer powers was key to the construction of eigenvalues.

24.2 Problem ($\star$). **(i)** Let $(v_1, \ldots, v_n)$ be a basis for the vector space $\mathcal{V}$ over $\mathbb{F}$ and let $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$. Define a linear operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ by setting $\mathcal{T}v_j = \lambda_j v_j$ and extending $\mathcal{T}$ to $\mathcal{V}$ by linearity. Prove that the eigenvalues of $\mathcal{T}$ are the scalars $\lambda_1, \ldots, \lambda_n$.

(ii) Prove that if the λ_j are all distinct (that is, $\lambda_j \neq \lambda_k$ for $k \neq j$), then the eigenspace corresponding to λ_k (Problem 12.8) is $\operatorname{span}(v_k)$. [Hint: *if* $\mathcal{T}v = \lambda_k v$ *with* $v \neq 0_{\mathcal{V}}$, *write* $v = \sum_{j=1}^{n}\alpha_j v_j$. *Obtain* $\sum_{j=1}^{n}\alpha_j(\lambda_j - \lambda_k)v_j = 0_{\mathcal{V}}$. *How does this help?*]

24.3 Problem ($\star$). Let $\mathcal{V}$ be a finite-dimensional vector space over $\mathbb{F}$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Suppose that each $v \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$ is an eigenvector for $\mathcal{T}$. (We are not right now assuming that each v is an eigenvector for the *same* eigenvalue.) Prove that $\mathcal{T} = \lambda\mathcal{I}_{\mathcal{V}}$ for some $\lambda \in \mathbb{F}$. [Hint: *consider the action of* $\mathcal{T}$ *on each basis vector in a basis* $(v_1, \ldots, v_n)$ *for* $\mathcal{V}$ *and then on the vector* $\sum_{j=1}^{n} v_j$.]

Dimension counting helps predict success or failure with the fundamental problem when it is posed on finite-dimensional spaces. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ cannot be surjective if $\dim(\mathcal{V}) < \dim(\mathcal{W})$ or injective if $\dim(\mathcal{V}) > \dim(\mathcal{W})$.

This informs us when existence or uniqueness fails. What else can we do with failure? How can we understand failure better? If $\mathcal{V}$ is finite-dimensional, there is a relationship among $\dim(\ker(\mathcal{T}))$, $\dim(\mathcal{T}(\mathcal{V}))$, and $\dim(\mathcal{V})$ which, if we know any two of these values, allows us to know the third. In particular, this allows us to quantify how existence failure and uniqueness failure interact.

24.4 Definition. *Let* $\mathcal{V}$ *and* $\mathcal{W}$ *be vector spaces.*

(i) *An operator* $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ *is* ***FINITE-RANK*** *if* $\mathcal{T}(\mathcal{V})$ *is finite-dimensional, and the* ***RANK*** *of* $\mathcal{T}$ *is* $\operatorname{rank}(\mathcal{T}) := \dim(\mathcal{T}(\mathcal{V}))$.

(ii) *For* $r \geq 1$, *an operator* $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ *is* ***RANK-****r* *if* $\mathcal{T}$ *is finite-rank and* $\operatorname{rank}(\mathcal{T}) = r$.

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

(i) Prove that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is 0-rank if and only if $\mathcal{T} = 0_{\mathcal{V}\to\mathcal{W}}$.

(ii) Suppose that $\mathcal{V}$ or $\mathcal{W}$ is finite-dimensional. Prove that every operator in $\mathbf{L}(\mathcal{V}, \mathcal{W})$ is finite-rank and relate $\operatorname{rank}(\mathcal{T})$ to $\dim(\mathcal{V})$ or $\dim(\mathcal{W})$.

(iii) Let $\mathcal{V}$ be an infinite-dimensional vector space. Prove that the identity operator $\mathcal{I}_{\mathcal{V}}$ is not finite-rank.

24.6 Problem ($\star$). Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces. Let $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Suppose that $\mathcal{T}$ is finite-rank. Prove that $\mathcal{S}\mathcal{T}$ is also finite-rank. How are $\operatorname{rank}(\mathcal{T})$ and $\operatorname{rank}(\mathcal{S}\mathcal{T})$ related? Are they ever equal?

Rank-1 operators are particularly simple, and every finite-rank operator is a linear combination of some list of them. We begin with some preparatory results on rank-1 operators.

24.7 Lemma. *Let* $\mathcal{V}$ *and* $\mathcal{W}$ *be vector spaces. An operator* $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ *is rank-1 if and only if there exist* $\varphi \in \mathcal{V}' \setminus \{0_{\mathcal{V}\to\mathbb{F}}\}$ *and* $w \in \mathcal{W} \setminus \{0_{\mathcal{W}}\}$ *such that* $\mathcal{T}v = \varphi(v)w$ *for all* $v \in \mathcal{V}$.

Proof. ($\Longrightarrow$) If $\mathcal{T}$ is rank-1, then $\mathcal{T}(\mathcal{V}) = \operatorname{span}(w)$ for some $w \in \mathcal{W} \setminus \{0_{\mathcal{W}}\}$. Put

$$\varphi = \{(v, \alpha) \mid \mathcal{T}v = \alpha w\}. \tag{24.1}$$

We leave it as an exercise to show that $\varphi \in \mathcal{V}' \setminus \{0_{\mathcal{V}\to\mathbb{F}}\}$.

($\Longleftarrow$) Here $\mathcal{T}(\mathcal{V}) \subseteq \operatorname{span}(w)$, and since $\varphi \neq 0_{\mathcal{V}\to\mathbb{F}}$, there is $v \in \mathcal{V}$ such that $\varphi(v) \neq 0$. Then for any $\alpha \in \mathbb{F}$,

$$\alpha w = \frac{\alpha}{\varphi(v)}(\varphi(v)w) = \frac{\alpha}{\varphi(v)}\mathcal{T}v = \mathcal{T}\left(\frac{\alpha}{\varphi(v)}v\right) \in \mathcal{T}(\mathcal{V}),$$

so $\operatorname{span}(w) = \mathcal{T}(\mathcal{V})$. Since $w \neq 0_{\mathcal{W}}$, we have $\dim(\operatorname{span}(w)) = 1$. ■

24.8 Problem ($\star$). Let φ be defined as in (24.1). First show that $\varphi \in \mathbb{F}^{\mathcal{V}}$. [Hint: *if* (v,α), $(v,\beta) \in \varphi$, *why do we have* $\alpha = \beta$*?*] Then show that $\varphi \in \mathcal{V}'$. [Hint: *use Problem 11.3.*] Finally, check that $\varphi \neq 0_{\mathcal{V}\to\mathbb{F}}$.

24.9 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, and let $(\mathcal{T}_1,\ldots,\mathcal{T}_n)$ be a list in $\mathbf{L}(\mathcal{V},\mathcal{W})$. If each $\mathcal{T}_j$ is rank-1, then there are $\varphi_j \in \mathcal{V}'$ and $w_j \in \mathcal{W}$ such that $\mathcal{T}_j v = \varphi_j(v)w_j$. This problem shows that the list $(\mathcal{T}_1,\ldots,\mathcal{T}_n)$ may be independent even if one or both of the lists $(\varphi_1,\ldots,\varphi_n)$ or $(w_1,\ldots,w_n)$ are dependent.

(i) Let $\mathcal{V} = \mathcal{W} = \mathbb{F}^2$. Let $\varphi_1(\mathbf{v}) = \mathbf{v}\cdot\mathbf{e}_1$, $\varphi_2(\mathbf{v}) = \mathbf{v}\cdot\mathbf{e}_2$, and $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{e}_1$. Put $\mathcal{T}_j\mathbf{v} = \varphi_j(\mathbf{v})\mathbf{w}_j$ for $j = 1$, 2. What do you know about the lists (φ_1,φ_2), $(\mathbf{w}_1,\mathbf{w}_2)$, and $(\mathcal{T}_1,\mathcal{T}_2)$?

(ii) Let $\mathcal{V} = \mathcal{W} = \mathbb{F}^2$. Let $\varphi_1(\mathbf{v}) = \mathbf{v}\cdot\mathbf{e}_1$, $\varphi_2(\mathbf{v}) = \mathbf{v}\cdot\mathbf{e}_1$, $\mathbf{w}_1 = \mathbf{e}_1$, and $\mathbf{w}_2 = \mathbf{e}_2$. Put $\mathcal{T}_j\mathbf{v} = \varphi_j(\mathbf{v})\mathbf{w}_j$ for $j = 1$, 2. What do you know about the lists (φ_1,φ_2), $(\mathbf{w}_1,\mathbf{w}_2)$, and $(\mathcal{T}_1,\mathcal{T}_2)$?

(iii) Let $\mathcal{V} = \mathcal{W} = \mathbb{F}^2$. Let $\varphi_1(\mathbf{v}) = \mathbf{v}\cdot\mathbf{e}_1$, $\varphi_2(\mathbf{v}) = \mathbf{v}\cdot\mathbf{e}_2$, $\varphi_3 = \varphi_1+\varphi_2$, $\mathbf{w}_1 = \mathbf{e}_1$, $\mathbf{w}_2 = \mathbf{e}_2$, and $\mathbf{w}_3 = \mathbf{w}_1+\mathbf{w}_2$. Put $\mathcal{T}_j\mathbf{v} = \varphi_j(\mathbf{v})\mathbf{w}_j$ for $j = 1$, 2, 3. What do you know about the lists $(\varphi_1,\varphi_2,\varphi_3)$, $(\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3)$, and $(\mathcal{T}_1,\mathcal{T}_2,\mathcal{T}_3)$?

24.10 Problem ($\star$). However, suppose that $\mathcal{V}$ and $\mathcal{W}$ are vector spaces and that the lists $(\varphi_1,\ldots,\varphi_r)$ in $\mathcal{V}'$ and $(w_1,\ldots,w_r)$ in $\mathcal{V}$ are independent. Let $\mathcal{T}_j v = \varphi_j(v)w_j$.

(i) Show that the list $(\mathcal{T}_1,\ldots,\mathcal{T}_r)$ is independent. [Hint: *use a third list* $(v_1,\ldots,v_r)$ *in* $\mathcal{V}$ *satisfying* (23.1).]

(ii) Let $\alpha_1,\ldots,\alpha_r \in \mathbb{F}\setminus\{0\}$ and $\mathcal{T} := \sum_{j=1}^r \alpha_j\mathcal{T}_j$. Prove that $\operatorname{rank}(\mathcal{T}) = r$.

24.11 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces with $\mathcal{V}$ finite-dimensional, and let $\mathcal{T} \in \mathbf{L}(\mathcal{V},\mathcal{W})$ be nonzero. Prove that $\mathcal{T} = \sum_{j=1}^s \mathcal{T}_j$ for a list $(\mathcal{T}_1,\ldots,\mathcal{T}_s)$ of rank-1 operators in $\mathbf{L}(\mathcal{V},\mathcal{W})$. [Hint: *take a basis for* $\mathcal{V}$ *and use the dual basis to rewrite* $\mathcal{T}v$ *as an appropriate linear combination.*]

We can sharpen the result of this problem and relax the hypothesis that $\mathcal{V}$ is finite-dimensional.

24.12 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. If $\mathcal{T}$ is finite-rank with $r := \text{rank}(\mathcal{T}) \geq 1$, then there is an independent list $(\mathcal{T}_1, \ldots, \mathcal{T}_r)$ of rank-1 operators in $\mathbf{L}(\mathcal{V}, \mathcal{W})$ such that $\mathcal{T} = \sum_{j=1}^r \mathcal{T}_j$.*

Proof. Let $(w_1, \ldots, w_r)$ be a basis for $\mathcal{T}(\mathcal{V})$, and let $(\varphi_1, \ldots, \varphi_r)$ be the corresponding dual basis for $\mathcal{T}(\mathcal{V})'$. Then $\mathcal{T}v = \sum_{j=1}^r \varphi_j(\mathcal{T}v)w_j$ for each $v \in \mathcal{V}$; this is the first identity in (20.4). This representation suggests that we put $\mathcal{T}_j v := \varphi_j(\mathcal{T}v)w_j$. Then $\mathcal{T}_j \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and $\mathcal{T} = \sum_{j=1}^r \mathcal{T}_j$. By independence, $\varphi_j \neq 0_{\mathcal{T}(\mathcal{V}) \to \mathbb{F}}$ and $w_j \neq 0_{\mathcal{W}}$, and so $\text{rank}(\mathcal{T}_j) = 1$ for all j by Lemma 24.7.

We check the independence of the list $(\mathcal{T}_1, \ldots, \mathcal{T}_r)$. Put $\psi_j(v) := \varphi_j(\mathcal{T}v)$, so $\psi_j \in \mathcal{V}'$ and $\mathcal{T}_j v = \psi_j(v)w_j$. We already know that the list $(w_1, \ldots, w_r)$ is independent. If we can show that $(\psi_1, \ldots, \psi_r)$ is independent as well, then Problem 24.10 will show that $(\mathcal{T}_1, \ldots, \mathcal{T}_r)$ is independent. So, suppose that $\sum_{j=1}^r \alpha_j \psi_j = 0_{\mathcal{V} \to \mathbb{F}}$. Then $\sum_{j=1}^r \alpha_j \psi_j(v) = 0$ for all $v \in \mathcal{V}$. Fix $1 \leq k \leq r$ and take $v = v_k$, where $v_k \in \mathcal{V}$ satisfies $\mathcal{T}v_k = w_k$. The fundamental identity (20.5) for the dual basis then gives

$$\psi_j(v_k) = \varphi_j(\mathcal{T}v_k) = \varphi_j(w_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

Thus

$$0 = \sum_{j=1}^r \alpha_j \psi_j(v_k) = \alpha_k.$$

Since k was arbitrary, we have $\alpha_k = 0$ for all $k = 1, \ldots, r$. ■

Day 25: Wednesday, March 18.

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."

Coset modulo a subspace, quotient space of a vector space modulo a subspace

We can achieve substantial control over the dimensions of the kernel and range of a finite-rank operator and, in doing so, quantify precisely how existence and/or uniqueness might fail for the fundamental problem $\mathcal{T}v = w$. We know that the rank of a linear operator is the dimension of its range; the "nullity" of a linear operator is a somewhat old-fashioned way of referring to the dimension of $\ker(\mathcal{T})$.

Content from *Linear Algebra* by Meckes & Meckes. Pages 172–175 discuss rank and nullity. We will discuss the rank of matrices later (and the rank of a transpose still later after that), but you should be familiar with the results in Theorem 3.32, Algorithm 3.33, and Theorem 3.34 from a first course in linear algebra. Do Quick Exercises #16 and #17 on p. 173 and #18 and #19 on p. 175.

Here is how these two dimensions are related.

25.1 Theorem (Rank–nullity). *Let $\mathcal{V}$ be a finite-dimensional vector space, let $\mathcal{W}$ be a vector space (not necessarily finite-dimensional), and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Then*

$$\dim(\ker(\mathcal{T})) + \dim(\mathcal{T}(\mathcal{V})) = \dim(\mathcal{V}).$$

Proof. Let $n = \dim(\mathcal{V})$. We leave the cases $\dim(\ker(\mathcal{T})) = 0$ and $\dim(\ker(\mathcal{T})) = n$ as exercises. So, put $p := \dim(\ker(\mathcal{T}))$ and assume $1 \le p < n$. Let $(v_1, \ldots, v_p)$ be a basis for $\ker(\mathcal{T})$, and extend this list to a basis $(v_1, \ldots, v_p, v_{p+1}, \ldots, v_n)$ for $\mathcal{V}$. Note that $\mathcal{T}v_j = 0$ for $j = 1, \ldots, p$.

We claim that $(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n)$ is a basis for $\mathcal{T}(\mathcal{V})$. If so, then since this list contains $n - p$ entries, we will have $\dim(\mathcal{T}(\mathcal{V})) = n - p = \dim(\mathcal{V}) - \dim(\ker(\mathcal{T}))$.

First we check that $\operatorname{span}(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n) = \mathcal{T}(\mathcal{V})$. Since $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$, for any $v \in \mathcal{V}$, we can write $v = \sum_{j=1}^{n} \alpha_j v_j$. Then

$$\mathcal{T}v = \sum_{j=1}^{n} \alpha_j \mathcal{T}v_j = \sum_{j=p+1}^{n} \alpha_j \mathcal{T}v_j.$$

Thus $\mathcal{T}v \in \operatorname{span}(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n)$, and certainly this span is contained in $\mathcal{T}(\mathcal{V})$. That is, $\mathcal{T}(\mathcal{V}) = \operatorname{span}(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n)$.

Next we check that $(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n)$ is independent. Suppose that $\sum_{j=p+1}^{n} \alpha_j \mathcal{T}v_j = 0_{\mathcal{W}}$. We will show that $\alpha_j = 0$. Then $\mathcal{T}\sum_{j=p+1}^{n} \alpha_j v_j = 0_{\mathcal{W}}$, so $\sum_{j=p+1}^{n} \alpha_j v_j \in \ker(\mathcal{T})$. Since $(v_1, \ldots, v_p)$ is a basis for $\ker(\mathcal{T})$, we can write $\sum_{j=p+1}^{n} \alpha_j v_j = \sum_{j=1}^{p} \beta_j v_j$ for some $\beta_j \in \mathbb{F}$, thus

$$\sum_{j=1}^{p} (-\beta_j) v_j + \sum_{j=p+1}^{n} \alpha_j v_j = 0_{\mathcal{V}}.$$

Since $(v_1, \ldots, v_n)$ is a basis, this list is independent, thus $-\beta_j = \alpha_j = 0$. ■

25.2 Problem (!). The following questions revisit the proof of rank–nullity.

(i) Finish the proof by treating the cases $\dim(\ker(\mathcal{T})) = 0$ and $\dim(\ker(\mathcal{T})) = n = \dim(\mathcal{V})$.

(ii) Why did we need to check both spanning and independence to conclude that the list $(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n)$ was a basis for $\mathcal{T}(\mathcal{V})$? That is, why did it not suffice to check just one of spanning or independence?

(iii) Since $(v_{p+1}, \ldots, v_n)$ is a sublist of the independent list $(v_1, \ldots, v_n)$, it is independent. Why does this not immediately imply that $(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n)$ is independent?

25.3 Problem (+). Here is a different proof of rank–nullity that starts with a basis for the range, not the kernel. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces with $\mathcal{V}$ finite-dimensional and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Let $\dim(\mathcal{V}) = n$ and $\dim(\mathcal{T}(\mathcal{V})) = r$.

(i) Prove that if $r = 0$, then $\ker(\mathcal{T}) = \mathcal{V}$ and so $\dim(\ker(\mathcal{T})) = n$.

(ii) Prove that if $r = n$, then given a basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$, the list $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ is a basis for $\mathcal{T}(\mathcal{V})$. Conclude that if $\mathcal{T}v = 0_{\mathcal{W}}$, then $v = 0_{\mathcal{V}}$, hence $\dim(\ker(\mathcal{T})) = 0$.

(iii) Suppose that $1 \le r \le n-1$ and let $(w_1, \ldots, w_r)$ be a basis for $\mathcal{T}(\mathcal{V})$. Let $v_j \in \mathcal{V}$ satisfy $\mathcal{T}v_j = w_j$. Check that the list $(v_1, \ldots, v_r)$ is independent. Now let $(z_1, \ldots, z_p)$ be a basis for $\ker(\mathcal{T})$; we will show $p = n-r$ by demonstrating that the list $(v_1, \ldots, v_r, z_1, \ldots, z_p)$ is a basis for $\mathcal{V}$. Check that this list is independent by assuming $\sum_{j=1}^r \alpha_j v_j + \sum_{j=1}^p \beta_j z_j = 0_{\mathcal{V}}$. Obtain $\sum_{j=1}^r \alpha_j \mathcal{T}v_j = 0_{\mathcal{W}}$ and conclude $\alpha_j = 0$. What does this imply about β_j? Next, check that this list spans $\mathcal{V}$: for $v \in \mathcal{V}$, write $\mathcal{T}v = \sum_{j=1}^r \gamma_j w_j$ and conclude $v - \sum_{j=1}^r \gamma_j v_j \in \ker(\mathcal{T})$. How does this help?

Content from *Linear Algebra* by Meckes & Meckes. Theorem 3.35 on p. 175 is rank–nullity. We will prove the matrix version later.

Here is the interpretation of rank–nullity. If $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ fails to be invertible (or an isomorphism, or a bijection), then either injectivity or surjectivity fails. Rank–nullity allows us to quantify this failure.

First, if existence fails and $\mathcal{T}(\mathcal{V}) \neq \mathcal{W}$, then $\dim(\mathcal{T}(\mathcal{V}))$ is not as "large" as it could be. If $\dim(\mathcal{T}(\mathcal{V}))$ is "too small," then $\dim(\ker(\mathcal{T}))$ will have to be "large enough' to make $\dim(\ker(\mathcal{T})) + \dim(\mathcal{T}(\mathcal{V})) = \dim(\mathcal{V})$ true. This means that if existence fails "sufficiently much," then uniqueness also fails with a "certain large degree of freedom."

Second, if uniqueness fails and $\ker(\mathcal{T}) \neq \{0_{\mathcal{V}}\}$, then $\dim(\ker(\mathcal{T}))$ is not as "small" as it could be. If $\dim(\ker(\mathcal{T}))$ is "too big," then $\dim(\mathcal{T}(\mathcal{V}))$ will have to be "small enough" to make, again, $\dim(\ker(\mathcal{T})) + \dim(\mathcal{T}(\mathcal{V})) = \dim(\mathcal{V})$ true. This means that if uniqueness fails with a "certain large degree of freedom," then uniqueness also fails "sufficiently much."

25.4 Example. We can use rank–nullity to prove a curious result about linear functionals. Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim(\mathcal{V}) = n \ge 1$. Let $\varphi \in \mathcal{V}'$ be a nonzero functional, so there is at least one $v_0 \in \mathcal{V}$ such that $\varphi(v_0) \neq 0$. (Since we are not working with elements of $\mathcal{V}''$, we are not decorating φ as φ' to distinguish it from elements of $\mathcal{V}''$.)

We claim that φ is surjective, so $\varphi(\mathcal{V}) = \mathbb{F}$, and therefore $\dim(\varphi(\mathcal{V})) = 1$. By rank–nullity, $\dim(\ker(\varphi)) = \dim(\mathbb{F}) - \dim(\varphi(\mathcal{V})) = n - 1$. Thus the kernel of a nontrivial functional is quite large. Indeed, if it were any larger, then we would have $\dim(\ker(\varphi)) = n$, and then $\varphi = 0$ would be the zero functional after all.

Here is the proof of surjectivity. Let $\alpha \in \mathbb{F}$. Since $\varphi(v_0) \neq 0$, we have

$$\alpha = \frac{\alpha\varphi(v_0)}{\varphi(v_0)} = \varphi\left(\frac{\alpha}{\varphi(v_0)}v_0\right) \in \varphi(\mathcal{V}).$$

The first equality is algebra (or arithmetic) in $\mathbb{F}$ via $\alpha = \alpha \cdot 1$, and the second equality is the linearity of φ, since $\alpha/\varphi(v_0) \in \mathbb{F}$ and $v_0 \in \mathcal{V}$.

25.5 Problem (!). Use the rank–nullity theorem to explain why we could have stopped the proof of Theorem 21.3 after proving the injectivity of $\mathcal{J}$ (which was probably easier than proving surjectivity).

25.6 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces. Use rank–nullity to prove the contrapositives of parts (i) and (ii) of Theorem 18.8. (To be clear, these contrapositives are true because we already proved the theorem. Here you are giving a *different* proof using rank–nullity.)

(i) Show that if $\dim(\mathcal{V}) < \dim(\mathcal{W})$, then no operator from $\mathcal{V}$ to $\mathcal{W}$ is surjective. [Hint: *if there is a surjection, obtain the contradiction* $\dim(\mathcal{V}) \geq \dim(\mathcal{W})$.]

(ii) Show that if $\dim(\mathcal{W}) < \dim(\mathcal{V})$, then no operator from $\mathcal{V}$ to $\mathcal{W}$ is injective. [Hint: *if there is an injection, obtain the contradiction* $\dim(\mathcal{V}) < \dim(\mathcal{V})$.]

Here is the most important consequence of rank–nullity *when the dimensions of the domain and codomain are the same.*

25.7 Problem (!). Show that if $\dim(\mathcal{V}) = \dim(\mathcal{W})$, then an operator is injective if and only if it is surjective (and so we only need to check one condition for invertibility).

Content from *Linear Algebra* by Meckes & Meckes. Corollary 3.36 on p. 178 and Proposition 3.39 on p. 180 give related versions of the results in this problem. Read and work through the proof of Corollary 3.37 on p. 179 as an exercise. The material about matrices on the bottom of p. 180/top of p. 181 should be familiar from a first course in linear algebra. Read (hopefully again) the geometric perspectives in the paragraph on p. 181.

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Both $\mathcal{V}$ and $\mathcal{W}$ may have some "redundant" information relative to $\mathcal{T}$. From the point of view of $\mathcal{V}$, any vector in $\ker(\mathcal{T})$ "contributes nothing" to the action of $\mathcal{T}$, since $\mathcal{T}z = 0_{\mathcal{W}}$ for all $z \in \ker(\mathcal{T})$. From the point of view of $\mathcal{W}$, any vector not in $\mathcal{T}(\mathcal{V})$ does not "interact" with $\mathcal{T}$ in any (ostensibly) meaningful way. We might wonder if there is a way to view $\mathcal{T}$ differently so that only the information in $\mathcal{V}$ and $\mathcal{W}$ "relevant" to $\mathcal{T}$ enters the perspective.

25.8 Problem (⋆). Here is one way to do this when $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional. Let $n = \dim(\mathcal{V})$, $m = \dim(\mathcal{W})$, and $r = \operatorname{rank}(\mathcal{T})$, and assume $1 \leq r < \min\{m, n\}$. Let $(z_1, \ldots, z_{n-r})$ be a basis for $\ker(\mathcal{T})$ and extend it to a basis $(z_1, \ldots, z_{n-r}, v_1, \ldots, v_r)$ for $\mathcal{V}$. Let $(w_1, \ldots, w_r)$ be a basis for $\mathcal{T}(\mathcal{V})$. Prove that

$$\widetilde{\mathcal{T}} \colon \operatorname{span}(v_1, \ldots, v_r) \to \mathcal{T}(\mathcal{V}) \colon v \mapsto \mathcal{T}v$$

is an isomorphism.

Here is another way, which does not rely on dimension. When does $\mathcal{T}$ "act the same" on vectors v_1, $v_2 \in \mathcal{V}$? When $\mathcal{T}v_1 = \mathcal{T}v_2$, equivalently (why?), when $v_2 - v_1 \in \ker(\mathcal{T})$. If $z := v_2 - v_1 \in \ker(\mathcal{T})$, then

$$v_2 = v_1 + (v_2 - v_1) = v_1 + z,$$

and if $\mathcal{T}$ "acts the same" on vectors v_1 and v_2, then v_2 is a "perturbation" from v_1 by a vector in $\ker(\mathcal{T})$. (Conversely, if $v_2 = v_1 + z$ for some $z \in \ker(\mathcal{T})$, then certainly $\mathcal{T}v_1 = \mathcal{T}v_2$.)

Thus given $v \in \mathcal{V}$, the operator $\mathcal{T}$ "acts the same" on all vectors in the set

$$\{v + z \mid z \in \ker(\mathcal{T})\}.$$

25.9 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, and $v \in V$. Prove that

$$\mathcal{T}(\{v + z \mid z \in \ker(\mathcal{T})\}) = \{\mathcal{T}v\}.$$

The arithmetic involved in the set $\{v + z \mid z \in \ker(\mathcal{T})\}$ can be generalized substantially.

25.10 Definition. *Let $\mathcal{V}$ be a vector space and $\mathcal{U}_1$, $\mathcal{U}_2 \subseteq \mathcal{V}$. (We are not assuming that either $\mathcal{U}_1$ or $\mathcal{U}_2$ is a subspace of $\mathcal{V}$.) The* ***SUM*** *of $\mathcal{U}_1$ and $\mathcal{U}_2$ is*

$$\mathcal{U}_1 + \mathcal{U}_2 := \{u_1 + u_2 \mid u_1 \in \mathcal{U}_1, \ u_2 \in \mathcal{U}_2\}.$$

25.11 Problem (!). **(i)** Prove that $\mathbb{F}^2 = \operatorname{span}(\mathbf{e}_1) + \operatorname{span}(\mathbf{e}_2)$.

(ii) If $\mathcal{V}$ is a vector space and $\mathcal{U}_1$, $\mathcal{U}_2 \subseteq \mathcal{V}$ are subspaces, prove that $\mathcal{U}_1 + \mathcal{U}_2$ is also a subspace.

(iii) Give an example to show that $\mathcal{U}_1 + \mathcal{U}_2$ may be a subspace even if both $\mathcal{U}_1$, $\mathcal{U}_2$ are not subspaces.

(iv) If $\mathcal{V}$ is a vector space and $\mathcal{U} \subseteq \mathcal{V}$ is a subspace, describe $\mathcal{U} + \mathcal{U}$ as precisely as possible.

So, if $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, and $v \in \mathcal{V}$, then $\mathcal{T}$ "acts the same" on

$$\{v\} + \ker(\mathcal{T}) = \{v + z \mid z \in \ker(\mathcal{T})\}.$$

In this case, we just write

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

25.12 Example. Consider $\mathcal{V} = \mathcal{W} = \mathbb{R}^2$ and $\mathcal{T} = \mathcal{M}_A$, where

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

Then $\ker(\mathcal{M}_A) = \operatorname{span}(\mathbf{e}_2)$ and, for $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$,

$$\mathbf{v} + \ker(\mathcal{M}_A) = \{\mathbf{v} + \alpha\mathbf{e}_2 \mid \alpha \in \mathbb{R}\} = \left\{ \begin{bmatrix} v_1 \\ v_2 + \alpha \end{bmatrix} \;\middle|\; \alpha \in \mathbb{R} \right\}.$$

Graphically, $\mathbf{v} + \ker(\mathcal{M}_A)$ is the set of all points in $\mathbb{R}^2$ whose first coordinate equals the first coordinate of $\mathbf{v}$. If we decide that two points in the plane "are the same" if they have the same first coordinate, then the vertical line through $\mathbf{v} = (v_1, v_2)$ is the set of all points that "are the same" as $\mathbf{v}$, and $\mathcal{M}_A$ acts the same" on all of these points.

25.13 Problem (!). Let $\mathbf{v}_0 := (1, 1)$. Sketch the sets $\mathbf{v} + \operatorname{span}(\mathbf{v}_0)$ in $\mathbb{R}^2$ for $\mathbf{v} = \pm\mathbf{e}_1$, $\pm\mathbf{e}_2$, and $\pm(1, -1)$.

The right way to extract only the essential behavior of $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is to consider the subsets of $\mathcal{V}$ on which $\mathcal{T}$ "acts the same" and endow them with an algebraic structure that makes them into a vector space. (This idea of taking sets and thinking of them as vectors, or numbers, appears elsewhere in mathematics, e.g., when completing a metric space.) Recall that $\mathcal{T}$ "acts the same" on v_1, $v_2 \in \mathcal{V}$ if $v_2 - v_1 \in \ker(\mathcal{T})$. It will be helpful to move beyond a fixed operator $\mathcal{T}$ and say that, given a subspace $\mathcal{U}$ of $\mathcal{V}$, vectors in $\mathcal{V}$ "are the same *with respect to* $\mathcal{U}$" if their difference is in $\mathcal{U}$.

25.14 Definition. *Let $\mathcal{V}$ be a vector space and $\mathcal{U} \subseteq \mathcal{V}$ be a subspace of $\mathcal{V}$.*

(i) *The* **COSET** *of $v \in \mathcal{V}$ with respect to $\mathcal{U}$ is the set*

$$v + \mathcal{U} := \{v + u \mid u \in \mathcal{U}\}.$$

(ii) *The* **QUOTIENT OF** $\mathcal{V}$ **MODULO** $\mathcal{U}$ *is the set*

$$\mathcal{V}/\mathcal{U} := \{v + \mathcal{U} \mid v \in \mathcal{V}\}.$$

We typically pronounce $\mathcal{V}/\mathcal{U}$ as "$\mathcal{V}$ mod $\mathcal{U}$." Note that $\mathcal{V}/\mathcal{U} \neq \mathcal{V} \setminus \mathcal{U}$. In particular, $\mathcal{V}/\mathcal{U} \not\subseteq \mathcal{V}$, but rather $\mathcal{V}/\mathcal{U}$ is a collection of subsets of $\mathcal{V}$.

25.15 Problem ($\star$). Let $\mathcal{V}$ be a vector space and $\mathcal{U} \subseteq \mathcal{V}$ be a subspace of $\mathcal{V}$. For what $v \in \mathcal{V}$ is $v + \mathcal{U}$ a subspace of $\mathcal{V}$?

25.16 Problem (!). Let $\mathcal{V}$ be a vector space. Describe as precisely as possible $\mathcal{V}/\{0_{\mathcal{V}}\}$ and $\mathcal{V}/\mathcal{V}$.

Given a vector space $\mathcal{V}$ and a subspace $\mathcal{U}$, if v, $w \in \mathcal{V}$ and $w \in v + \mathcal{U}$, then $w = v + u$ for some $u \in \mathcal{U}$, hence $w - v = u \in \mathcal{U}$. Conversely, if $w - v \in \mathcal{U}$, then $w = v + (w - v) \in v + \mathcal{U}$. That is,

$$v + \mathcal{U} = \{w \in \mathcal{V} \mid w - v \in \mathcal{U}\}.$$

This leads to a bit of a challenge of representation. Any element $\mathcal{Z} \in \mathcal{V}/\mathcal{U}$ is a set of the form $\mathcal{Z} = v + \mathcal{U}$ for some $v \in \mathcal{V}$, but this v is not unique. We can have $v_1 + \mathcal{U} = v_2 + \mathcal{U}$ for $v_1 \neq v_2$.

25.17 Problem ($\star$). Let $\mathcal{V}$ be a vector space, $\mathcal{U} \subseteq \mathcal{V}$ be a subspace of $\mathcal{V}$, and v_1, $v_2 \in \mathcal{V}$. Prove that either $v_1 + \mathcal{U} = v_2 + \mathcal{U}$ or $(v_1 + \mathcal{U}) \cap (v_2 + \mathcal{U}) = \varnothing$, and give necessary and sufficient conditions on v_1 and v_2 for each case to hold.

Day 26: Friday, March 20.

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."

Quotient space of a vector space modulo a subspace (be able to define the vectors in the quotient space, the zero vector, and the operations of vector addition and scalar multiplication)

26.1 Example. We work in the vector space $\mathcal{V} = \mathcal{C}([0,1])$. Let

$$\mathcal{U} := \left\{ h \in \mathcal{C}([0,1]) \;\middle|\; \int_0^1 h(x)\, dx = 0 \right\}.$$

For $f \in \mathcal{C}([0,1])$, we have $g \in f + \mathcal{U}$ if and only if $g - f \in \mathcal{U}$, which is equivalent to

$$\int_0^1 f(x)\, dx = \int_0^1 g(x)\, dx.$$

That is, the functions in $f + \mathcal{U}$ are the functions in $\mathcal{C}([0,1])$ that have the same "net area under their graphs" as f does.

If, say, $f(x) = 1$ for all x, then $\int_0^1 f(x)\, dx = 1$, and so

$$f + \mathcal{U} = \left\{ g \in \mathcal{C}([0,1]) \;\middle|\; \int_0^1 g(x)\, dx = 1 \right\}.$$

Then the following functions are in $f + \mathcal{U}$:

$$g_1(x) = 2x, \qquad g_2(x) = x + \frac{1}{2}, \qquad g_3(x) = 3x^2.$$

26.2 Problem (!). Consider the vector space $\mathcal{V} = \mathcal{C}([0,1])$. Let

$$\mathcal{U} := \{h \in \mathcal{C}([0,1]) \mid h(0) = 0\}.$$

For each $f_j \in \mathcal{C}([0,1])$ below, find a function $g_j \in f_j + \mathcal{U}$ such that $g_j \neq f_j$:

$$f_1(x) = \sin(x), \qquad f_2(x) = e^x, \qquad f_3(x) = x^2 - 1.$$

26.3 Problem (!). Let

$$\mathcal{V} := \left\{(a_k) \in \mathbb{F}^\infty \;\middle|\; \lim_{k\to\infty} a_k \text{ exists}\right\} \quad \text{and} \quad \mathcal{U} := \left\{(b_k) \in \mathbb{F}^\infty \;\middle|\; \lim_{k\to\infty} b_k = 0\right\}.$$

(i) For $(a_k) \in \mathbb{F}^\infty$, show that

$$(a_k) + \mathcal{U} = \left\{(c_k) \in \mathcal{V} \;\middle|\; \lim_{k\to\infty} a_k = \lim_{k\to\infty} c_k\right\}.$$

(ii) Give an example of two different elements of $(1/k+1) + \mathcal{U}$.

Now we begin to endow $\mathcal{V}/\mathcal{U}$ with algebraic structures that will turn it into a vector space. We will eventually return to $\mathcal{U} = \ker(\mathcal{T})$ for an operator $\mathcal{T}$, but for now that extra structure will not help us. Say that $\mathcal{Z}_1$, $\mathcal{Z}_2 \in \mathcal{V}/\mathcal{U}$. We know how to add $\mathcal{Z}_1 + \mathcal{Z}_2$ from Definition 25.10, so that might seem like a reasonable definition of vector addition in $\mathcal{V}/\mathcal{U}$. However, we need to be sure that this definition gives $\mathcal{Z}_1 + \mathcal{Z}_2 \in \mathcal{V}/\mathcal{U}$. That is, if v_1, $v_2 \in \mathcal{V}$, is there $v_3 \in \mathcal{V}$ such that $(v_1 + \mathcal{U}) + (v_2 + \mathcal{U}) = v_3 + \mathcal{U}$? Yes.

26.4 Lemma. *Let $\mathcal{V}$ be a vector space, $\mathcal{U} \subseteq \mathcal{V}$ be a subspace of $\mathcal{V}$, and v_1, $v_2 \in \mathcal{V}$. Then*

$$(v_1 + \mathcal{U}) + (v_2 + \mathcal{U}) = (v_1 + v_2) + \mathcal{U}.$$

Proof. By Definition 25.10, we have

$$(v_1+\mathcal{U})+(v_2+\mathcal{U}) = \{z_1 + z_2 \mid z_1 \in v_1 + \mathcal{U},\ z_2 \in v_2 + \mathcal{U}\} = \{(v_1 + u_1) + (v_2 + u_2) \mid u_1,\ u_2 \in \mathcal{U}\}.$$

Here we allow u_1 and u_2 to be different, since having $z_j \in v_j + \mathcal{U}$ just means $z_j = v_j + u_j$ for some $u_j \in \mathcal{U}$. Then

$$(v_1 + \mathcal{U}) + (v_2 + \mathcal{U}) = \{(v_1 + v_2) + (u_1 + u_2) \mid u_1,\ u_2 \in \mathcal{U}\}.$$

If we can "change variables" with $u = u_1 + u_2$, and if

$$\{(v_1 + v_2) + (u_1 + u_2) \mid u_1,\ u_2 \in \mathcal{U}\} = \{(v_1 + v_2) + u \mid u \in \mathcal{U}\},$$

then we are done (why?). Here is why this last equality holds. Certainly taking $u = u_1 + u_2$ gives

$$(v_1 + v_2) + (u_1 + u_2) \in \{(v_1 + v_2) + u \mid u \in \mathcal{U}\}$$

for any u_1, $u_2 \in \mathcal{U}$. Conversely, if $u \in \mathcal{U}$, then taking $u_1 = u$ and $u_2 = 0_{\mathcal{V}}$ gives

$$(v_1 + v_2) + u \in \{(v_1 + v_2) + (u_1 + u_2) \mid u_1,\ u_2 \in \mathcal{U}\}.$$

Thus

$$(v_1 + \mathcal{U}) + (v_2 + \mathcal{U}) = \{(v_1 + v_2) + u \mid u \in \mathcal{U}\} = (v_1 + v_2) + \mathcal{U}.$$ ■

If we want $\mathcal{V}/\mathcal{U}$ to be a vector space with vector addition given by Definition 25.10, then we will need a compatible zero vector. Say that $\mathcal{Z}_0 \in \mathcal{V}/\mathcal{U}$ satisfies $\mathcal{Z} + \mathcal{Z}_0 = \mathcal{Z}$ for all $\mathcal{Z} \in \mathcal{V}/\mathcal{U}$. Writing $\mathcal{Z}_0 = v_0 + \mathcal{U}$, we have

$$v + \mathcal{U} = (v + \mathcal{U}) + \mathcal{Z} = (v + \mathcal{U}) + (v_0 + \mathcal{U}) = (v + v_0) + \mathcal{U}$$

for all $v \in \mathcal{V}$. In particular, taking $v = 0_{\mathcal{V}}$, we have

$$0_{\mathcal{V}} + \mathcal{U} = v_0 + \mathcal{U}.$$

26.5 Problem (!). Check that $0_{\mathcal{V}} + \mathcal{U} = \mathcal{U}$. Then find a condition on $v \in \mathcal{V}$ that is necessary and sufficient to have $v + \mathcal{U} = \mathcal{U}$. [Hint: *use your solution to Problem 25.17.*]

So, if $v_0 \in \mathcal{V}$ satisfies $(v + \mathcal{U}) + (v_0 + \mathcal{U}) = v + \mathcal{U}$ for all $v \in \mathcal{V}$, then

$$v_0 + \mathcal{U} = \mathcal{U}.$$

Since $v_0 = v_0 + 0_{\mathcal{V}} \in \mathcal{U}$, we conclude $v_0 \in \mathcal{U}$, and so $v_0 + \mathcal{U} = \mathcal{U}$ after all. Our zero vector in $\mathcal{V}/\mathcal{U}$ should therefore be $\mathcal{U}$.

26.6 Lemma. *Let $\mathcal{V}$ be a vector space and $\mathcal{U} \subseteq \mathcal{V}$ be a subspace of $\mathcal{V}$. Then*

$$\mathcal{U} + \mathcal{Z} = \mathcal{Z}$$

for all $\mathcal{Z} \in \mathcal{V}/\mathcal{U}$.

26.7 Problem (!). Give a succinct proof of this lemma.

Now we define scalar multiplication in $\mathcal{V}/\mathcal{U}$. From our success with using Definition 25.10 to obtain vector addition, we might try to define scalar multiplication as

$$\alpha\mathcal{Z} := \{\alpha z \mid z \in \mathcal{Z}\} \tag{26.1}$$

for $\alpha \in \mathbb{F}$ and $\mathcal{Z} \in \mathcal{V}/\mathcal{U}$. This almost works. However, when $\alpha = 0$, we want $0\mathcal{Z}$ to be the zero vector in $\mathcal{V}/\mathcal{U}$, so we want

$$0\mathcal{Z} = \mathcal{U}.$$

But (26.2) would have

$$0\mathcal{Z} = \{0_{\mathcal{V}}\}.$$

26.8 Problem (⋆). If $\{0_{\mathcal{V}}\} \in \mathcal{V}/\mathcal{U}$, what do you know about $\mathcal{V}$ and/or $\mathcal{U}$?

The right idea is a piecewise definition. For $\alpha \in \mathbb{F}$ and $\mathcal{Z} \in \mathcal{V}/\mathcal{U}$, put

$$\alpha\mathcal{Z} := \begin{cases} \{\alpha z \mid z \in \mathcal{Z}\}, & \alpha \neq 0 \\ \mathcal{U}, & \alpha = 0. \end{cases} \tag{26.2}$$

26.9 Problem (⋆). Let $\mathcal{V}$ be a vector space, $\mathcal{U} \subseteq \mathcal{V}$ be a subspace, $v \in \mathcal{V}$, and $\alpha \in \mathbb{F}$. Use (26.2) to show that

$$\alpha(v + \mathcal{U}) = (\alpha v) + \mathcal{U}.$$

With the right definitions in place, $\mathcal{V}/\mathcal{U}$ is indeed a vector space. Informally, we can view $\mathcal{V}/\mathcal{U}$ as "forget about differences up to belonging to $\mathcal{U}$." We might also say that two vectors v_1, $v_2 \in \mathcal{V}$ are the same "modulo $\mathcal{U}$" if $v_1 - v_2 \in \mathcal{U}$, and that we "mod out by $\mathcal{U}$" or "quotient out by $\mathcal{U}$" to form $\mathcal{V}/\mathcal{U}$. By the way, the space $\mathcal{V}/\mathcal{U}$ is probably the most abstract vector space that we have yet encountered! After all, most of our "concrete" spaces have been function spaces of the form $\mathbb{F}^X$ for a well-chosen X, or subspaces thereof, and even the operator spaces $\mathbf{L}(\mathcal{V}, \mathcal{W})$, $\mathbf{L}(\mathcal{V})$, and $\mathcal{V}'$ (and $\mathcal{V}''$) are still function spaces, of sorts.

26.10 Problem (+). Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{U} \subseteq \mathcal{V}$ be a subspace. With $\mathcal{V}/\mathcal{U}$ defined in Definition 25.14, check that vector addition as defined in Definition 25.10 and scalar multiplication as defined in (26.2) make $\mathcal{V}/\mathcal{U}$ a vector space over $\mathbb{F}$. Explicitly cite where you are using these definitions and where you are using the vector space properties of $\mathcal{V}$. We call $\mathcal{V}/\mathcal{U}$ the **QUOTIENT SPACE OF $\mathcal{V}$ MODULO $\mathcal{U}$** (which is a bit of an upgrade from the phrasing of Definition 25.14, where we could not yet use the word "space" yet).

Since the point of the quotient space $\mathcal{V}/\mathcal{U}$ is to help us "forget about" $\mathcal{U}$, we might expect that when $\mathcal{V}$ is finite-dimensional, so is $\mathcal{V}/\mathcal{U}$, and $\dim(\mathcal{V}/\mathcal{U})$ should equal what is left in $\mathcal{V}$ when we "forget about" $\mathcal{U}$.

26.11 Theorem. *Let $\mathcal{V}$ be a finite-dimensional vector space and $\mathcal{U} \subseteq \mathcal{V}$ be a subspace of $\mathcal{V}$. Then $\mathcal{V}/\mathcal{U}$ is also finite-dimensional with*

$$\dim(\mathcal{V}/\mathcal{U}) = \dim(\mathcal{V}) - \dim(\mathcal{U}).$$

The quantity $\dim(\mathcal{V}/\mathcal{U})$ *is the* ***CODIMENSION OF $\mathcal{U}$ IN $\mathcal{V}$.***

Proof. If $\mathcal{U} = \{0_{\mathcal{V}}\}$ or $\mathcal{U} = \mathcal{V}$, this follows from Problem 25.16. (Why?) Otherwise, suppose that $\mathcal{U} \neq \{0_{\mathcal{V}}\}$ and $\mathcal{U} \neq \mathcal{V}$. Since everything about $\mathcal{V}/\mathcal{U}$ hinges on $\mathcal{U}$, and since now $1 \le \dim(\mathcal{U}) < \dim(\mathcal{V}) =: n$, it is reasonable to start with a basis $(u_1, \ldots, u_r)$ and extend it to a basis $(u_1, \ldots, u_r, v_1, \ldots, v_{n-r})$ for $\mathcal{V}$. It should then feel natural to consider how the list $(u_1 + \mathcal{U}, \ldots, u_r + \mathcal{U}, v_1 + \mathcal{U}, \ldots, v_{n-r} + \mathcal{U})$ interacts with $\mathcal{V}/\mathcal{U}$.

Let $\mathcal{Z} \in \mathcal{V}/\mathcal{U}$ and take $v \in \mathcal{V}$ such that $\mathcal{Z} = v + \mathcal{U}$. Since $(u_1, \ldots, u_r, v_1, \ldots, v_{n-r})$ spans $\mathcal{V}$, we can write $v = \sum_{j=1}^{r} \alpha_j u_j + \sum_{j=1}^{n-r} \beta_j v_j$ for some α_j, $\beta_j \in \mathbb{F}$. Then

$$v + \mathcal{U} = \left(\sum_{j=1}^{r} \alpha_j u_j + \sum_{j=1}^{n-r} \beta_j v_j\right) + \mathcal{U} = \sum_{j=1}^{r} \alpha_j (u_j + \mathcal{U}) + \sum_{j=1}^{n-r} \beta_j (v_j + \mathcal{U}) = \sum_{j=1}^{n-r} \beta_j (v_j + \mathcal{U})$$

since $u_j + \mathcal{U} = \mathcal{U}$ for $j = 1, \ldots, n$. Thus

$$\mathcal{V}/\mathcal{U} = \operatorname{span}(v_1 + \mathcal{U}, \ldots, v_{n-r} + \mathcal{U}),$$

and so $\mathcal{V}/\mathcal{U}$ is finite-dimensional.

The list $(v_1+\mathcal{U},\ldots,v_{n-r}+\mathcal{U})$ will therefore be a basis for $\mathcal{V}/\mathcal{U}$ if we can show that it is independent. Suppose that

$$\sum_{j=1}^{n-r}\alpha_j(v_j+\mathcal{U})=\mathcal{U}$$

for some $\alpha_j\in\mathbb{F}$. The goal, of course, is that $\alpha_j=0$. We jave

$$\left(\sum_{j=1}^{n-r}\alpha_jv_j\right)+\mathcal{U}=\mathcal{U},$$

thus $\sum_{j=1}^{n-r}\alpha_jv_j\in\mathcal{U}$. Then we can write $\sum_{j=1}^{n-r}\alpha_jv_j=\sum_{j=1}^{r}\beta_ju_j$ for some $\beta_j\in\mathbb{F}$, and so

$$\sum_{j=1}^{n-r}\alpha_jv_j+\sum_{j=1}^{r}(-\beta_j)u_j=0_{\mathcal{V}}.$$

Since $(u_1,\ldots,u_r,v_1,\ldots,v_{n-r})$ is independent, we have $\alpha_j=\beta_j=0$ for all j.

The list $(v_1+\mathcal{U},\ldots,v_{n-r}+\mathcal{U})$ is therefore a basis for $\mathcal{V}/\mathcal{U}$, and so we have

$$\dim(\mathcal{V}/\mathcal{U})=n-r=\dim(\mathcal{V})-\dim(\mathcal{U}).$$ ■

26.12 Problem (!). What similarities do you notice between the proof of this theorem and the proof of rank–nullity?

Now we return to our original motivation for quotient spaces. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T}\in\mathbf{L}(\mathcal{V},\mathcal{W})$. What is an "efficient" way of viewing $\mathcal{T}$ with the least amount of irrelevant information in the domain and codomain? The answer is the so-called **FIRST ISOMORPHISM THEOREM**.

26.13 Theorem (First isomorphism theorem). *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T}\in\mathbf{L}(\mathcal{V},\mathcal{W})$. The map*

$$\widetilde{\mathcal{T}}\colon\mathcal{V}/\ker(\mathcal{T})\to\mathcal{T}(\mathcal{V})\colon v+\ker(\mathcal{T})\mapsto\mathcal{T}(\mathcal{V})$$

is an isomorphism.

Proof. First we need to check that $\widetilde{\mathcal{T}}\in\mathcal{T}(\mathcal{V})^{\mathcal{V}/\ker(\mathcal{T})}$. This might seem more obvious than it is, as we ostensibly have a "formula" for $\widetilde{\mathcal{T}}(v+\ker(\mathcal{T}))$. However, set-theoretically we really have

$$\widetilde{\mathcal{T}}:=\{(\mathcal{Z},\mathcal{T}v)\mid\mathcal{Z}=v+\mathcal{U}\text{ for some }v\in\mathcal{V}\}.$$

Certainly for all $\mathcal{Z}\in\mathcal{V}/\ker(\mathcal{T})$, there is $w\in\mathcal{T}(\mathcal{V})$ such that $(\mathcal{Z},w)\in\widetilde{\mathcal{T}}$.

But if $(\mathcal{Z},w_1)$, $(\mathcal{Z},w_2)\in\widetilde{\mathcal{T}}$, do we have $w_1=w_2$? If $(\mathcal{Z},w_1)\in\widetilde{\mathcal{T}}$, then $\mathcal{Z}=v_1+\ker(\mathcal{T})$ for some $v_1\in\mathcal{V}$, and $w_1=\mathcal{T}v_1$. And if $(\mathcal{Z},w_2)\in\widetilde{\mathcal{T}}$, then $\mathcal{Z}=v_2+\ker(\mathcal{T})$ for some $v_2\in\mathcal{V}$, and $w_2 = \mathcal{T}v_2$. There is no reason to assume $v_1 = v_2$. Nonetheless, since $v_1 + \ker(\mathcal{T}) = \mathcal{Z} = v_2 + \ker(\mathcal{T})$, we have $v_2 - v_1 \in \ker(\mathcal{T})$, and so

$$w_1 = \mathcal{T}v_1 = \mathcal{T}v_2 - \mathcal{T}(v_2 - v_1) = \mathcal{T}v_2 = w_2.$$

So, $\widetilde{\mathcal{T}} \colon \mathcal{V}/\ker(\mathcal{T}) \to \mathcal{T}(\mathcal{V})$ is a function.

Checking linearity is a relatively straightforward exercise (do it). For injectivity, if $\widetilde{\mathcal{T}}\mathcal{Z} = 0_{\mathcal{W}}$ for some $\mathcal{Z} \in \mathcal{V}/\ker(\mathcal{T})$, then taking $\mathcal{Z} = v + \ker(\mathcal{T})$ for some $v \in \mathcal{V}$ gives

$$0_{\mathcal{W}} = \widetilde{\mathcal{T}}(v + \ker(\mathcal{T})) = \mathcal{T}v,$$

thus $v \in \ker(\mathcal{T})$, and so $v + \ker(\mathcal{T}) = \ker(\mathcal{T}) = 0_{\mathcal{V}/\ker(\mathcal{T})}$.

Finally, for surjectivity, let $w \in \mathcal{T}(\mathcal{V})$, and take $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. Then

$$w = \mathcal{T}v = \widetilde{\mathcal{T}}(v + \ker(\mathcal{T})). \qquad \blacksquare$$

26.14 Problem (!). Let $\mathcal{V}$ be a vector space and $\varphi \in \mathcal{V}' \setminus \{0_{\mathcal{V}\to\mathbb{F}}\}$. Prove that $\mathcal{V}/\ker(\varphi)$ and $\mathbb{F}$ are isomorphic. Explain how this generalizes the result of Example 25.4.

26.15 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$.

(i) Use the first isomorphism theorem and Theorem 26.11 to prove the rank–nullity theorem.

(ii) Use the rank–nullity theorem and Theorem 26.11 to prove that $\mathcal{V}/\ker(\mathcal{T})$ and $\mathcal{T}(\mathcal{V})$ are isomorphic. Does this give another proof of the first isomorphism theorem?

26.16 Problem ($\star$). Put

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

In language accessible to a calculus student who understands the fundamental theorem of calculus but has not taken any linear algebra, explain the structure and behavior of a coset $f + \ker(\mathcal{T})$ and the operator

$$\widetilde{\mathcal{T}} \colon \mathcal{C}^1([0,1])/\ker(\mathcal{T}) \to \mathcal{C}([0,1]) \colon f + \ker(\mathcal{T}) \mapsto \mathcal{T}f.$$

Day 27: Monday, March 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."

L^2-inner product on $\mathcal{C}([a,b])$, orthogonal list (N)

The fundamental question of existence and uniqueness for the operator problem $\mathcal{T}v = w$ boils down set-theoretically to knowledge of range and kernel. In finite dimensions, rank–nullity provides precise *quantitative* information about range and kernel. Such information is not available, or even meaningful, in infinite dimensions. While many problems naturally involve a finite-dimensional setting, in infinite dimensions (or even finite dimensions), there is often another tool available (which also interacts well with lists, independence, bases, and dimension, so we are not totally losing our good prior results). This is the tool of geometry made manifest through the inner product.

Toward the operator problem, we will use inner products to give a *qualitative* characterization of the range that is not limited to a finite-dimensional setting and to approximate the unsolvable problem $\mathcal{T}v = w$ when w is not in the range of $\mathcal{T}$ by a solvable (and meaningful) problem $\mathcal{T}v = \widehat{w}$ for an appropriately chosen $\widehat{w}$ that *is* in the range of $\mathcal{T}$. We will build a number of structural results for inner product spaces first, so, as with finite-dimensional spaces, it will be some time before operators appear. Along the way, even when no linear operator is present (a fearful situation!), we will gain significant insight into the structure of a vector space that possess an inner product, as inner products help us compare vectors, measure vectors, represent vectors, and extract meaningful data about vectors.

Many vector spaces are naturally equipped with an inner product that reflects our intuitive notions of geometry in two- and three-dimensional space. Here are two of the three most important examples of inner products.

27.1 Example. For $\mathbf{v}$, $\mathbf{w} \in \mathbb{F}^n$, define

$$\langle \mathbf{v}, \mathbf{w} \rangle := \sum_{j=1}^{n} v_j \overline{w_j}.$$

Here, for $x + iy \in \mathbb{F}$ with x, $y \in \mathbb{R}$, the scalar $\overline{x+iy} := x - iy$ is the **CONJUGATE** of $x + iy$, and we review some of its properties along the way in the context of essential properties of $\langle \cdot, \cdot \rangle$. Of course, $\langle \mathbf{v}, \mathbf{w} \rangle$ is the **DOT PRODUCT** of $\mathbf{v}$ and $\mathbf{w}$, with conjugation necessary in the case $\mathbb{F} = \mathbb{C}$.

1. We have

$$\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \sum_{j=1}^{n} (u_j + v_j)\overline{w}_j = \sum_{j=1}^{n} (u_j \overline{w}_j + v_j \overline{w}_j) = \sum_{j=1}^{n} u_j \overline{w}_j + \sum_{j=1}^{n} v_j \overline{w}_j = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$$

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

2. Likewise, because of how arithmetic works, we have

$$\langle \alpha \mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle$$

for all $\alpha \in \mathbb{F}$, $\mathbf{v}$, $\mathbf{w} \in \mathbb{F}^n$.

3. If we reverse the order of things, we have

$$\langle \mathbf{w}, \mathbf{v} \rangle = \sum_{j=1}^n w_j \overline{v_j} = \sum_{j=1}^n \overline{\overline{w_j} v_j} = \overline{\sum_{j=1}^n \overline{w_j} v_j} = \overline{\langle \mathbf{v}, \mathbf{w} \rangle}.$$

Here we have used the properties $\overline{\overline{\alpha}} = \alpha$, $\overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta}$, and $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$.

4. If we make both slots the same, we have

$$\langle \mathbf{v}, \mathbf{v} \rangle = \sum_{k=1}^n v_j \overline{v_j} = \sum_{k=1}^n |v_j|^2 \geq 0.$$

Here we are using the property that $\alpha\overline{\alpha} \geq 0$ for all $\alpha \in \mathbb{F}$; indeed, if $\alpha = x + iy$ with x, $y \in \mathbb{R}$, then $\alpha\overline{\alpha} = x^2 + y^2 \geq 0$.

5. What if equality is achieved above and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$? Then $\sum_{j=1}^n |v_j|^2 = 0$. Let $1 \leq k \leq n$. Then

$$0 \leq |v_k|^2 \leq \sum_{j=1}^n |v_j|^2 = 0,$$

so $|v_k|^2 = 0$, thus $|v_k| = 0$, and therefore $v_k = 0$. Then $\mathbf{v} = \mathbf{0}_n$. Here we are using the definition $|\alpha| := \sqrt{\alpha\overline{\alpha}}$ and the consequent property that $|\alpha| = 0$ if and only if $\alpha = 0$ for any $\alpha \in \mathbb{F}$.

Content from *Linear Algebra* by Meckes & Meckes. Pages 225–226 discuss the dot product as an inner product. This continues in the examples on p. 227.

27.2 Example. Let $\mathcal{V} = \mathcal{C}([0,1])$; recall that functions in $\mathcal{V}$ are real-valued (we could develop the following for complex-valued functions, but that requires a little too much calculus for complex-valued functions of a real variable than we care to pursue). For f, $g \in \mathcal{V}$, put

$$\langle f, g \rangle := \int_0^1 f(x)g(x)\ dx.$$

This integral is defined because the product of continuous functions is continuous and therefore integrable. We compare the properties of $\langle \cdot, \cdot \rangle$ here to the previous example with the intuitive idea that the integral is a "continuous sum" (i.e., a limit of Riemann sums, which are finite sums of products and thus morally dot products). Nothing in the following would change if we worked on an arbitrary interval $[a, b]$, and with a little more work all of this would be valid for improper integrals, too.

1. Linearity of the integral implies

$$\langle f+g,h\rangle = \int_0^1 \big(f(x)+g(x)\big)h(x)\ dx = \int_0^1 f(x)h(x)\ dx + \int_0^1 g(x)h(x)\ dx = \langle f,h\rangle + \langle g,h\rangle$$

for all f, g, $h \in \mathcal{V}$.

2. More linearity of the integral implies

$$\langle \alpha f, g\rangle = \int_0^1 \alpha f(x)g(x)\ dx = \alpha \int_0^1 f(x)g(x)\ dx = \alpha \langle f,g\rangle .$$

3. Since f, $g \in \mathcal{V}$ are real-valued, $\langle f,g\rangle \in \mathbb{R}$, and so $\overline{\langle f,g\rangle} = \langle f,g\rangle$. But also

$$\langle g,f\rangle = \int_0^1 g(x)f(x)\ dx = \int_0^1 f(x)g(x)\ dx = \langle f,g\rangle = \overline{\langle f,g\rangle}.$$

4. We compute

$$\langle f,f\rangle = \int_0^1 f(x)f(x)\ dx = \int_0^1 [f(x)]^2\ dx \geq 0,$$

since $[f(x)]^2 \geq 0$ for all $x \in [0,1]$. Here we are using the **MONOTONICITY OF THE INTEGRAL**: if g, $h \in \mathcal{V}$ with $g(x) \leq h(x)$ for all $x \in [0,1]$, then $\int_0^1 g(x)\ dx \leq \int_0^1 h(x)\ dx$. (This is best understood in terms of areas: if $w(x) \geq 0$ for $0 \leq x \leq 1$, then $\int_0^1 w(x)\ dx \geq 0$. Take $w = g - h$.)

5. Suppose $\langle f,f\rangle = 0$, so $\int_0^1 |f(x)|^2\ dx = 0$. (Of course here $|f(x)|^2 = (f(x))^2$; we are just using absolute value for notational cleanliness.) By analogy with the dot product, we might conjecture that this forces $f(x) = 0$ for all $x \in [0,1]$.

What if $f(x_0) \neq 0$ for some $x_0 \in [0,1]$? Continuity implies the existence of $\delta > 0$ such that $|f(x)| > |f(x_0)|/2 > 0$ for $x \in (x_0-\delta, x_0+\delta) \cap [0,1]$. If $0 < x_0 < 1$, then

$$\begin{aligned} 0 < \frac{\delta|f(x_0)|^2}{2} = ((x_0+\delta)-(x_0-\delta))\frac{|f(x_0)|^2}{4} &= \int_{x_0-\delta}^{x_0+\delta} \frac{|f(x_0)|^2}{4}\ dx \leq \int_{x_0-\delta}^{x_0+\delta} |f(x)|^2\ dx \\ &\leq \int_0^1 |f(x)|^2\ dx < \int_{x_0-\delta}^{x_0+\delta} [f(x)]^2\ dx \leq \int_0^1 [f(x)]^2\ dx = 0. \end{aligned}$$

More precisely, the first nonstrict inequality with integrals is monotonicity of the integral, while the second is the inequality $\int_c^d g(x)\ dx \leq \int_0^1 g(x)\ dx$ for $0 \leq c \leq d \leq 1$ and $g(x) \geq 0$ on $[0,1]$. We have reached the contradiction $0 < 0$, and so it must be the case that $f(x_0) = 0$. Then $f(x) = 0$ for $0 < x < 1$, so by continuity $f(0) = \lim_{x\to 0^+} f(x) = 0$, and likewise $f(1) = 0$. Thus $f = 0$.

We codify the properties of the structures $\langle \cdot, \cdot \rangle$ above into a definition. Here it is important to note explicitly what the underlying field is.

27.3 Definition. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. An* **INNER PRODUCT** *on $\mathcal{V}$ is a function*

$$\langle \cdot, \cdot \rangle : \{(v, w) \mid v,\ w \in \mathcal{V}\} \to \mathbb{F}$$

such that the following hold.

1. **[Additivity]** $\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle$ *for all* $v_1,\ v_2,\ w \in \mathcal{V}$.

2. **[Homogeneity]** $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$ *for all* $\alpha \in \mathbb{F}$ *and* $v,\ w \in \mathcal{V}$.

3. **[Conjugacy]** $\overline{\langle v, w \rangle} = \langle w, v \rangle$ *for all* $v,\ w \in \mathcal{V}$. *(This is trivially true if $\mathbb{F} = \mathbb{R}$.)*

4. **[Nonnegativity]** $\langle v, v \rangle \geq 0$ *for all* $v \in \mathcal{V}$.

5. **[Definiteness]** *If* $\langle v, v \rangle = 0$, *then* $v = 0$.

An **INNER PRODUCT SPACE** *is an ordered list $(\mathcal{V}, \mathbb{F}, +, \cdot, \langle \cdot, \cdot \rangle)$, where $(\mathcal{V}, \mathbb{F}, +, \cdot)$ is a vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $\langle \cdot, \cdot \rangle$ is an inner product on $\mathcal{V}$. Of course, we usually just refer to the space as $\mathcal{V}$.*

While a vector space can be defined over more general fields than $\mathbb{R}$ or $\mathbb{C}$ (and we are not doing that in this course), inner product spaces require real or complex fields.

Content from *Linear Algebra* by Meckes & Meckes. Pages 226–227 define inner products and inner product spaces. We will need to be slightly more careful about the role of complex scalars here than we have before, so rereading Appendix A.2, mostly Lemma A.3 on p. 382, is a good idea.

27.4 Example. **(i)** The map from Example 27.1 is an inner product on $\mathbb{F}^n$, and of course we usually call it the **DOT PRODUCT**, or sometimes the **EUCLIDEAN INNER PRODUCT**, and write it as $\mathbf{v} \cdot \mathbf{w}$, not as $\langle \mathbf{v}, \mathbf{w} \rangle$.

(ii) Let $\mathbf{q} = (q_1, \ldots, q_n) \in \mathbb{R}^n$ with $q_j > 0$ for each j. Then

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{q}} := \sum_{j=1}^{n} (q_j v_j) \overline{w_j} \tag{27.1}$$

is an inner product (a "weighted" inner product) on $\mathbb{F}^n$. We check only nonnegativity and definiteness:

$$\langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{q}} = \sum_{j=1}^{n} q_j (v_j \overline{v_j}) = \sum_{j=1}^{n} q_j |v_j|^2 \geq 0,$$

since $q_j > 0$ for all j. And if $\langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{q}} = 0$, then

$$0 \leq q_k |v_k|^2 \leq \langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{q}} = 0$$

for all k, since $q_k > 0$. Thus $q_k|v_k|^2 = 0$, and so we may solve for $|v_k|^2 = 0$ because, once again, $q_k > 0$ (and here is the only time that we really used the strict inequality $q_k > 0$).

(iii) The map from Example 27.2 is an inner product on $\mathcal{C}([0,1])$, which we often call (for various historical and cultural reasons) the **L^2-INNER PRODUCT ON** $\mathcal{C}([0,1])$. More generally,

$$\langle f, g \rangle := \int_a^b f(x)g(x)\ dx, \ \ f, \ \ g \in \mathcal{C}([a,b])$$

gives an inner product on $\mathcal{C}([a,b])$, which we also call the **L^2-INNER PRODUCT ON** $\mathcal{C}([a,b])$. (Mostly we will take $a = 0$ and $b = 1$, but sometimes we will work over symmetric intervals like $[-1,1]$ for geometric reasons or intervals of length 2π for trigonometric convenience.)

Content from *Linear Algebra* by Meckes & Meckes. The examples on pp. 233–234 (skip #4 on p. 235 for now) offer more inner product spaces. Example 1 is our weighted dot product.

27.5 Problem (!). Give an example of $\mathbf{q} \in \mathbb{R}^n$ such that the weighted inner product $\langle \cdot, \cdot \rangle_{\mathbf{q}}$ from (27.1) is not an inner product on $\mathbb{F}^n$. Discuss exactly what goes wrong.

27.6 Problem (!). Why does defining

$$\langle f, g \rangle := \int_0^1 f'(x)g(x)\ dx$$

not give an inner product on $\mathcal{C}^1([0,1])$?

27.7 Problem ($\star$). Let $\mathcal{V}$ be an inner product space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{V}}$. What conditions on $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ guarantee that

$$\langle v, w \rangle_{\mathcal{T}} := \langle \mathcal{T}v, w \rangle_{\mathcal{V}}$$

defines an inner product on $\mathcal{V}$?

We have done quite some work with the linear functionals on an arbitrary (often finite-dimensional) vector space. The inner product induces a very special, very important kind of linear functional.

27.8 Example. Let $\mathcal{V}$ be an inner product space. The map

$$\varphi \colon \mathcal{V} \to \mathbb{F} \colon v \mapsto \langle v, w \rangle$$

is linear. There is really not much to the proof: it is just additivity and homogeneity of the inner product. We will denote this functional by $\langle \cdot, w \rangle$. That is,

$$\langle \cdot, w \rangle = \{(v, \langle v, w \rangle) \mid v \in \mathcal{V}\} \quad \text{and} \quad \langle \cdot, w \rangle \in \mathcal{V}'.$$

A natural question, then, is whether *all* functionals on an inner product space can be so "represented" by the inner product. We will address this in detail. (Short answer: "Yes" with an "if." Long answer: "No" with a "but.")

27.9 Problem (!). Let $\mathcal{V}$ be an inner product space.

(i) Fix w_1, $w_2 \in \mathcal{V}$. Prove that the map

$$\mathcal{T} \colon \mathcal{V} \to \mathcal{V} \colon \langle v, w_1 \rangle\, w_2$$

is linear.

(ii) Prove that the inner product is **ANTILINEAR** in the second slot in the sense that

$$\langle v, w_1 + w_2 \rangle = \langle v, w_1 \rangle + \langle v, w_2 \rangle \quad \text{and} \quad \langle v, \alpha w \rangle = \overline{\alpha}\, \langle v, w \rangle$$

for all v, w, w_1, $w_2 \in \mathcal{V}$ and $\alpha \in \mathbb{F}$.

Here is one of the most important ways in which we use inner products (and the functionals defined by them) as instruments for extracting data about vectors.

27.10 Theorem. *Let $\mathcal{V}$ be an inner product space and $v \in \mathcal{V}$. The following are equivalent.*

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

(ii) $\langle v, w \rangle = 0$ *for all* $w \in \mathcal{V}$.

(iii) $\langle w, v \rangle = 0$ *for all* $w \in \mathcal{V}$.

Proof. The second and third parts are equivalent because $\langle w, v \rangle = \overline{\langle v, w \rangle}$, and, for $z \in \mathbb{F}$, we have $z = 0$ if and only if $\overline{z} = 0$.

We work on the equivalence of the first two parts. If $v = 0_V$, then

$$\langle v, w \rangle = \langle 0_{\mathcal{V}}, w \rangle = \langle 0 \cdot 0_{\mathcal{V}}, w \rangle = 0\, \langle 0_{\mathcal{V}}, w \rangle = 0.$$

Conversely, suppose that $\langle v, w \rangle = 0$ for all $w \in \mathcal{V}$. We know a special property of inner products and 0 when both inputs to the inner product are the same, and we are allowed to pick any $w \in \mathcal{V}$ here, so we set $w = v$ and compute $0 = \langle v, v \rangle$. The axioms for an inner product then imply $v = 0_{\mathcal{V}}$. ■

This result is an excellent example of the slogan "What things do defines what things are." What the zero vector does with respect to the inner product is that taking the inner product against (the) zero (vector) always returns (the) zero (scalar).

Content from *Linear Algebra* by Meckes & Meckes. Proposition 4.2 on p. 228 contains some of these nice algebraic properties of inner products.

27.11 Problem ($\star$). Let $\mathcal{V}$ be a vector space and $\mathcal{W}$ be an inner product space with inner product $\langle \cdot, \cdot \rangle$. Suppose that both $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional with bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$, respectively. Let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Prove that $\mathcal{T} = 0$ if and only if $\langle \mathcal{T} v_j, w_k \rangle = 0$ for $j = 1, \ldots, n$ and $k = 1, \ldots, m$. [Hint: *recall from Theorem 18.13 that it suffices to know the values of $\mathcal{T} v_j$; use Theorem 27.10 and algebraic properties of the inner product to reduce this to knowledge of $\langle \mathcal{T} v_j, w_k \rangle$.*] Explain why checking that $\mathcal{T}_1 = \mathcal{T}_2$ for $\mathcal{T}_1, \mathcal{T}_2 \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ amounts to doing only mn calculations involving scalars (although, to be fair, the same could be done involving the n calculations $(\mathcal{T}_1 - \mathcal{T}_2)v_j = 0$ with vectors).

Geometry in inner product spaces mostly involves angles and lengths. There is really only one important angle: the right angle.

27.12 Example. **(i)** Consider the inner product space $\mathbb{F}^n$ with inner product given by the dot product, as usual. Let $\mathbf{v} \in \mathbb{F}^n$. The kth component of $\mathbf{v}$ is $\langle \mathbf{v}, \mathbf{e}_k \rangle$ and so

$$\mathbf{v} = \sum_{j=1}^{n} v_j \mathbf{e}_j = \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{e}_j \rangle \, \mathbf{e}_j. \tag{27.2}$$

The inner product thus provides a convenient way of expressing $\mathbf{v}$ as a linear combination of the standard basis vectors. Consequently, we can also calculate inner products via inner products:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{j=1}^{n} v_j \overline{w_j} = \sum_{k=1}^{n} \langle \mathbf{v}, \mathbf{e}_j \rangle \, \overline{\langle \mathbf{w}, \mathbf{e}_j \rangle}. \tag{27.3}$$

(ii) Now let both $\mathbb{F}^n$ and $\mathbb{F}^m$ have the dot product(s) as their inner products and let $A \in \mathbb{F}^{m \times n}$. Let $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be the standard basis for $\mathbb{F}^n$ and $(\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_m)$ be the standard basis for $\mathbb{F}^m$. For example, if $n = 3$ and $m = 2$, then

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{but} \quad \widetilde{\mathbf{e}}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Then the jth column of A is $A\mathbf{e}_j$, and the kth component of this vector is $\langle A\mathbf{e}_j, \widetilde{\mathbf{e}}_k \rangle$. Changing one letter, more colloquially this says that the (i, j)-entry of A is $\langle A\mathbf{e}_j, \widetilde{\mathbf{e}}_i \rangle$.

What makes the standard basis vectors so special with respect to the dot product is not really their componentwise formulas (although that *is* what everything ultimately relies on) but rather their "orthonormality":

$$\langle \mathbf{e}_j, \mathbf{e}_k \rangle = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases} \tag{27.4}$$

We abstract and exploit this calculation in much more general contexts.

27.13 Definition. *Let $\mathcal{V}$ be an inner product space. A list $(v_1, \ldots, v_n)$ in $\mathcal{V}$ is* **ORTHOGONAL** *if $\langle v_j, v_k \rangle = 0$ for all $j \neq k$.*

Content from *Linear Algebra* by Meckes & Meckes. Orthogonality is defined at the bottom of p. 229.

27.14 Example. **(i)** Certainly the list of standard basis vectors in $\mathbb{F}^n$ is orthogonal.

(ii) The list

$$\left(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-3\\0\\1\end{bmatrix}\right)$$

is orthogonal in $\mathbb{F}^3$, since

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

(iii) The list $(\sin(\cdot), \cos(\cdot)$ is orthogonal in $\mathcal{C}([-\pi, \pi])$, since

$$\langle \sin(\cdot), \cos(\cdot) \rangle = \int_{-\pi}^{\pi} \sin(x)\cos(x)\ dx = 0$$

after substituting $u = \cos(x)$.

27.15 Problem (!). Let x, $y \in \mathbb{R}$. Show that

$$\left(\begin{bmatrix}x\\y\end{bmatrix}, \begin{bmatrix}y\\-x\end{bmatrix}\right)$$

is orthogonal in $\mathbb{F}^2$. Draw a picture that illustrates how this corresponds to the usual geometric notion of orthogonality (= perpendicularity).

27.16 Problem (!). Let $(v_1, \ldots, v_n)$ be an orthogonal list in an inner product space $\mathcal{V}$, and let $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$. Prove that the list $(\alpha_1 v_1, \ldots, \alpha_n v_n)$ is also orthogonal.

Of course the list of standard basis vectors in $\mathbb{F}^n$ is linearly independent; we now show that this is an easy consequence of orthogonality. First, it is straightforward to represent the coefficients of a vector in the span of an orthogonal list.

27.17 Lemma. *Let $(v_1, \ldots, v_n)$ be an orthogonal list in the inner product space $\mathcal{V}$ with $v_j \neq 0$ for all j. If $v \in \operatorname{span}(v_1, \ldots, v_n)$, then*

$$v = \sum_{j=1}^{n} \frac{\langle v, v_j \rangle}{\langle v_j, v_j \rangle} v_j. \tag{27.5}$$

Proof. Write $v = \sum_{j=1}^{n} \alpha_j v_j$. We know that $\langle v_k, v_j \rangle = 0$ for $k \neq j$, so to make this identity show up in our assumption, we fix k, take the inner product of the linear combination against v_j, and use algebraic properties of the inner product:

$$\langle v, v_k \rangle = \left\langle \sum_{j=1}^{n} \alpha_j v_j, v_k \right\rangle = \sum_{j=1}^{n} \langle \alpha_j v_j, v_k \rangle = \sum_{j=1}^{n} \alpha_j \langle v_j, v_k \rangle = \alpha_k \langle v_k, v_k \rangle . \tag{27.6}$$

The last equality is the identity $\langle v_j, v_k \rangle = 0$ for $k \neq j$. Since $v_j \neq 0$ by our hypotheses on this list, the inner product axioms imply $\langle v_j, v_j \rangle > 0$, thus $\alpha_j = 0$. ■

Content from *Linear Algebra* by Meckes & Meckes. This result is Theorem 4.3 on p. 230. Do Quick Exercise #3 on that page.

27.18 Remark. *The calculation in (27.6) is an important example of good mathematical grammar: we are using j as the index of summation, and so we should not overwork it by using j in the second slot. That is why we wrote $\left\langle \sum_{j=1}^{n} \alpha_j v_j, v_k \right\rangle$, not $\left\langle \sum_{j=1}^{n} \alpha_j v_j, v_j \right\rangle$. Indeed, the latter would have us calculate*

$$\left\langle \sum_{j=1}^{n} \alpha_j v_j, v_j \right\rangle = \sum_{j=1}^{n} \alpha_j \langle v_j, v_j \rangle ,$$

which is useless, because it does not bring the orthogonality of the list into play, and it does not "extract" any particular coefficient from the sum.

With this lemma, we can prove a very nice property of orthogonal lists.

27.19 Theorem. *An orthogonal list of nonzero vectors is independent.*

Proof. Let $(v_1, \ldots, v_n)$ be an orthogonal list in the inner product space $\mathcal{V}$ with $v_j \neq 0_{\mathcal{V}}$ for all j. Suppose $\sum_{j=1}^{n} \alpha_j v_j = 0_{\mathcal{V}}$. Then $\alpha_j = \langle v_j, 0_{\mathcal{V}} \rangle / \langle v_j, v_j \rangle = 0$. ■

Day 28: Wednesday, March 25.

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."

Orthonormal list in an inner product space (N), orthonormal basis for a finite-dimensional inner product space

Orthogonality by itself does not account for the excellent behavior of the standard basis vectors in (27.4). In particular, orthogonality says nothing about the behavior of the inner products $\langle v, v\rangle$ when both vectors are the same.

28.1 Definition. *Let $\mathcal{V}$ be an inner product space. A list $(u_1, \ldots, u_n)$ in $\mathcal{V}$ is* ***ORTHONORMAL*** *if*

$$\langle u_j, u_k\rangle = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

A vector $u \in \mathcal{V}$ such that $\langle u, u\rangle = 1$ is a ***UNIT VECTOR***.

Content from *Linear Algebra* by Meckes & Meckes. Page 239 defines orthonormality. Read the examples there and on p. 240 and do Quick Exercise #6. The phrasing of Example #3 might be reworded to say that the list there is not an orthonormal basis simply because not every function in the space is a trigonometric polynomial; the dimension has less to do with it. Read in particular Example #6 on p. 240. Go back to Example 1 on p. 233 and Figure 4.3 on p. 234 for an illustration of a unit vector in the context of a weighted dot product.

So, every orthonormal list is orthogonal. Thus we have the (strict) inclusions

$$\text{orthonormal lists} \subsetneq \text{orthogonal lists} \subsetneq \text{independent lists}.$$

28.2 Problem (!). Give an example of each of the following.

(i) An independent list in an inner product space that is not orthogonal.

(ii) An orthogonal list in an inner product space that is not orthonormal.

We can adapt the calculation in (27.6) to generalize the expansion (27.2). We can paraphrase this via the slogan "Inner products extract coefficients from linear combinations."

28.3 Theorem. *Let $(u_1, \ldots, u_n)$ be an orthonormal list in the inner product space $\mathcal{V}$, and* *let* $v \in \operatorname{span}(u_1, \ldots, u_n)$*. Then*

$$v = \sum_{j=1}^{n} \langle v, u_j \rangle u_j. \tag{28.1}$$

The scalars in the list $(\langle v, u_1 \rangle, \ldots, \langle v, u_n \rangle)$ *are the* ***Fourier coefficients*** *of* v *with respect to the orthonormal list* $(u_1, \ldots, u_n)$*.*

Proof. Write $v = \sum_{j=1}^{n} \alpha_j u_j$. We compute

$$\langle v, u_j k \rangle = \left\langle \sum_{j=1}^{n} \alpha_j u_j, u_k \right\rangle = \sum_{j=1}^{n} \alpha_j \langle u_j, u_k \rangle = \alpha_j.$$

The third equality is orthonormality with $\langle u_j, u_k \rangle = 1$ for $k = j$ and 0 otherwise. ■

Now we generalize (27.3).

28.4 Theorem. *Let* $(u_1, \ldots, u_n)$ *be an orthonormal list in the inner product space* $\mathcal{V}$*, and let* v*,* $w \in \operatorname{span}(u_1, \ldots, u_n)$*. Then*

$$\langle v, w \rangle = \sum_{j=1}^{n} \langle v, u_j \rangle \overline{\langle w, u_j \rangle}. \tag{28.2}$$

Proof. We use (28.1) to expand

$$\langle v, w \rangle = \left\langle \sum_{j=1}^{n} \langle v, u_j \rangle u_j, w \right\rangle. \tag{28.3}$$

Note that we are only expanding v via (28.1), not w. Then we use linearity in the first slot of the inner product to obtain

$$\left\langle \sum_{j=1}^{n} \langle v, u_j \rangle u_j, w \right\rangle = \sum_{j=1}^{n} \langle \langle v, u_j \rangle u_j, w \rangle = \sum_{j=1}^{n} \langle v, u_j \rangle \langle u_j, w \rangle. \tag{28.4}$$

Finally, we use conjugacy to find

$$\sum_{j=1}^{n} \langle v, u_j \rangle \langle u_j, w \rangle = \sum_{j=1}^{n} \langle v, u_j \rangle \overline{\langle w, u_j \rangle}. \tag{28.5}$$

Combining (28.3), (28.4), and (28.5) gives (28.1). ■

The identities (27.5), (28.1), and (28.2) are examples of our initial claim that inner products extract data about vectors and measure properties of vectors: the inner product, so far, has extracted coefficients of vectors in spans of orthogonal and orthonormal lists, and then we used those coefficients to calculate other inner products.

Content from *Linear Algebra* by Meckes & Meckes. The example on pp. 243–244 does these kinds of calculations (ignore the reference to Theorem 4.9).

Throughout, quantities of the form $\langle v, v\rangle$ have shown up, and we know such quantities are nonnegative and positive for all but the zero vector.

28.5 Definition. *Let $\mathcal{V}$ be an inner product space with inner product $\langle\cdot,\cdot\rangle$. The* **NORM** *induced by $\langle\cdot,\cdot\rangle$ is the map*

$$\|\cdot\| : \mathcal{V} \to \mathbb{R} : v \mapsto \sqrt{\langle v, v\rangle}.$$

28.6 Example. (i) The Euclidean norm on $\mathbb{F}^n$ is

$$\|\mathbf{v}\| = \left(\sum_{j=1}^{n} |v_j|^2\right)^{1/2},$$

which, of course, is exactly how we typically think about lengths of vectors.

(ii) The L^2-norm on $\mathcal{C}([0,1])$ is

$$\|f\| = \left(\int_0^1 |f(x)|^2 \, dx\right)^{1/2}.$$

The fundamental idea of a norm is that it should measure the size or the length of a vector in a meaningful way. The Euclidean norm already does this intuitively; the L^2-norm is a means of capturing the total area under the graph, and we might think of the squaring in the integrand as penalizing "small" parts of f less and "large" parts of f more. We will eventually meet many other norms that are not induced by inner products (indeed, that *cannot* be induced by inner products), and we will see how those norms provide other meaningful, and useful, interpretations of size and length.

The norm induced by the inner product has many useful properties; here are two straightforward ones.

28.7 Theorem. *Let $\mathcal{V}$ be an inner product space.*

(i) $\|\alpha v\| = |\alpha| \, \|v\|$ *for any $\alpha \in \mathbb{F}$ and $v \in \mathcal{V}$.*

(ii) $\|v\| \geq 0$ *for all $v \in \mathcal{V}$.*

(iii) $\|v\| = 0$ *if and only if $v = 0_{\mathcal{V}}$.*

28.8 Problem (!). Prove it.

We will explore some deeper properties of the norm shortly. For now, we spend some more time with orthonormal lists. Any finite-dimensional inner product space has a basis (because it is a finite-dimensional vector space), but there is no guarantee that this basis "talks" well to the inner product. (On a not unrelated note, we saw something similar with the dual basis when we tried to relate arbitrary bases for finite-dimensional $\mathcal{V}$ and $\mathcal{V}'$, which, as (20.1) indicated, did not exactly go well.) It would be nice if we could guarantee that any finite-dimensional inner product space has a basis consisting of orthonormal vectors.

28.9 Definition. *A basis for a finite-dimensional inner product space is an* ***ORTHONORMAL BASIS*** *if this basis is also an orthonormal list.*

28.10 Example. The standard basis vectors $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ form an orthonormal basis for $\mathbb{F}^n$.

We can indeed guarantee this, and a bit more, in a rather general context. Let $\mathcal{V}$ be an inner product space and let $(v_1, \ldots, v_n)$ be an independent list in $\mathcal{V}$. We are going to find an orthonormal list $(u_1, \ldots, u_n)$ that "preserves spans" relative to $(v_1, \ldots, v_n)$: for $j = 1, \ldots, n$, we will have

$$\operatorname{span}(v_1, \ldots, v_j) = \operatorname{span}(u_1, \ldots, u_j). \tag{28.6}$$

This is what we want for $j = n$, but our procedure gives (28.6) for all j to boot.

There is a fairly transparent, recursive algorithm for constructing the u_j. We look at a few small cases of n to see how to develop this algorithm from scratch.

1. $n = 1$. We start with the very small independent list (v_1), so $v_1 \neq 0_{\mathcal{V}}$, and we want the very small list (u_1) to be orthonormal with $\operatorname{span}(v_1) = \operatorname{span}(u_1)$. We therefore want $\|u_1\| = 1$, and we will get the spanning property if $u_1 = \alpha_1 v_1$ for some $\alpha_1 \in \mathbb{F} \setminus \{0\}$. How should we choose α_1? All we know is that

$$1 = \|u_1\| = \|\alpha_1 v_1\| = |\alpha_1| \, \|v_1\| \, .$$

Since $v_1 \neq 0_{\mathcal{V}}$, we have $\|v_1\| \neq 0$. Then we may solve for

$$|\alpha_1| = \frac{1}{\|v_1\|}.$$

This suggests taking

$$\alpha_1 = \frac{1}{\|v_1\|},$$

so

$$u_1 := \frac{v_1}{\|v_1\|}$$

satisfies $\|u_1\| = 1$ and $\operatorname{span}(v_1) = \operatorname{span}(u_1)$.

2. $n = 2$. We start with the independent list (v_1, v_2), and we want an orthonormal list (u_1, u_2) with $\operatorname{span}(v_1, v_2) = \operatorname{span}(u_1, u_2)$. Since declaring $u_1 := v_1 / \|v_1\|$ was enough in the $n = 1$ case, we might try doing that here. (As before, $v_1 \neq 0_{\mathcal{V}}$ since v_1 is a vector in a basis.) Then we want

$$u_2 \in \operatorname{span}(u_1, u_2) = \operatorname{span}(v_1, v_2),$$

and so we want to be able to write

$$u_2 = \alpha_1 v_1 + \alpha_2 v_2$$

for some α_1, $\alpha_2 \in \mathbb{F}$. We also want

$$\langle u_2, u_1 \rangle = 0 \quad \text{and} \quad \|u_2\| = 1.$$

There are two equations here, and we have two unknowns (α_1 and α_2), so we might feel optimistic.

We first need

$$0 = \langle u_2, u_1 \rangle = \langle \alpha_1 v_1 + \alpha_2 v_2, u_1 \rangle = \alpha_1 \langle v_1, u_1 \rangle + \alpha_2 \langle v_2, u_1 \rangle .$$

We have

$$\langle v_1, u_1 \rangle = \left\langle v_1, \frac{v_1}{\|v_1\|} \right\rangle = \frac{\langle v_1, v_1 \rangle}{\|v_1\|} = \|v_1\| ,$$

and so α_1 and α_2 must satisfy

$$\alpha_1 \|v_1\| + \alpha_2 \langle v_2, u_1 \rangle = 0.$$

This is one equation with two unknowns (not a recipe for success), but remember that we have not yet used the other desired condition $\|u_2\| = 1$. Since $\|v_1\| \neq 0$, we can solve for α_1 in terms of α_2 as

$$\alpha_1 = -\alpha_2 \frac{\langle v_2, u_1 \rangle}{\|v_1\|}.$$

Then

$$u_2 = \alpha_1 v_1 + \alpha_2 v_2 = -\alpha_2 \left(\frac{\langle v_2, u_1 \rangle}{\|v_1\|} \right) v_1 + \alpha_2 v_2 = \alpha_2 \big(v_2 - \langle v_2, u_1 \rangle u_1 \big).$$

We can achieve $\|u_2\| = 1$ by taking

$$\alpha_2 = \frac{1}{\|v_2 - \langle v_2, u_1 \rangle u_1\|}$$

if $\|v_2 - \langle v_2, u_1 \rangle u_1\| \neq 0$, equivalently, if $v_2 - \langle v_2, u_1 \rangle u_1 \neq 0_{\mathcal{V}}$. This is true, for otherwise

$$0_{\mathcal{V}} = v_2 - \langle v_2, u_1 \rangle u_1 = \left(-\frac{\langle v_2, u_1 \rangle}{\|v_1\|} \right) v_1 + v_2,$$

which implies the dependence of (v_1, v_2).

To summarize, we put

$$\begin{cases} u_1 := v_1 / \|v_1\| \\ w_2 := v_2 - \langle v_2, u_1 \rangle u_1 \\ u_2 := w_2 / \|w_2\| \end{cases} \tag{28.7}$$

to get an orthonormal list (u_1, u_2) with the desired property that $\operatorname{span}(v_1, v_2) = \operatorname{span}(u_1, u_2)$, and the added bonus that $\operatorname{span}(u_1) = \operatorname{span}(u_1)$. The intermediate vector w_2 is just there for notational convenience.

3. $n = 3$. We want (u_1, u_2, u_3) to be orthonormal and $\text{span}(v_1, v_2, v_3) = \text{span}(u_1, u_2, u_3)$. Based on our prior success, we might use (28.7) to define u_1 and u_2. Then (check this) we would have $\text{span}(v_1, v_2, v_3) = \text{span}(u_1, u_2, v_3)$. We therefore want $u_3 \in \text{span}(u_1, u_2, v_3)$, and so we need $u_3 = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 v_3$. So, we want $u_3 = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 v_3$. To have $\langle u_3, u_1 \rangle = 0$, we need

$$0 = \langle u_3, u_1 \rangle = \alpha_1 \langle u_1, u_1 \rangle + \alpha_2 \langle u_2, u_1 \rangle + \alpha_3 \langle v_3, u_1 \rangle = \alpha_1 + \alpha_3 \langle v_3, u_1 \rangle$$

since $\langle u_1, u_1 \rangle = \|u_1\|^2 = 1$ and $\langle u_2, u_1 \rangle = 0$. Similarly, we want

$$0 = \langle u_3, u_2 \rangle = \alpha_1 \langle u_1, u_2 \rangle + \alpha_2 \langle u_2, u_2 \rangle + \alpha_3 \langle v_3, u_2 \rangle = \alpha_2 + \alpha_3 \langle v_3, u_2 \rangle$$

since $\langle u_1, u_2 \rangle = 0$ and $\langle u_2, u_2 \rangle = \|u_2\|^2 = 1$. Then we need

$$\alpha_1 = -\alpha_3 \langle v_3, u_1 \rangle \qquad \text{and} \qquad \alpha_2 = -\alpha_3 \langle v_3, u_2 \rangle .$$

So, u_3 must have the form

$$u_3 = -\alpha_3 \langle v_3, u_1 \rangle u_1 - \alpha_3 \langle v_3, u_2 \rangle u_2 + \alpha_3 v_3 = \alpha_3 (v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2).$$

Put $w_3 = v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2$. If $w_3 = 0_{\mathcal{V}}$, then $v_3 \in \text{span}(u_1, u_2) = \text{span}(v_1, v_2)$, which contradicts the independence of (v_1, v_2, v_3). So, $w_3 \neq 0_{\mathcal{V}}$, and therefore to have $\|u_3\| = 1$, we could take $\alpha_3 = 1/\|w_3\|$.

To summarize, we put

$$\begin{cases} u_1 := v_1/\|v_1\| \\ w_2 := v_2 - \langle v_2, u_1 \rangle u_1 \\ u_2 := w_2/\|w_2\| \\ w_3 := v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2 \\ u_3 := w_3/\|w_3\| \end{cases}$$

to get an orthonormal list u_1, u_2, u_3 with the desired property that $\text{span}(v_1, v_2, v_3) = \text{span}(u_1, u_2, u_3)$, and the added bonus that $\text{span}(v_1) = \text{span}(u_1)$ and $\text{span}(v_1, v_2) = \text{span}(u_1, u_2)$.

Here is how this works in general.

28.11 Theorem (Gram–Schmidt process). *Let $\mathcal{V}$ be an inner product space and let $(v_1, \ldots, v_n)$ be an independent list in $\mathcal{V}$. Define*

$$u_j := \begin{cases} \dfrac{v_1}{\|v_1\|}, & j = 1 \\[2ex] \dfrac{w_j}{\|w_j\|}, & j \geq 2, \end{cases} \qquad w_j := v_j - \sum_{k=1}^{j-1} \langle v_j, u_k \rangle u_k, \ j \geq 2. \tag{28.8}$$

Then $(u_1, \ldots, u_n)$ is an orthonormal list with

$$\text{span}(v_1, \ldots, v_j) = \text{span}(u_1, \ldots, u_j), \ j = 1, \ldots, n. \tag{28.9}$$

Proof. We induct on n.

1. *The case* $n = 1$. Since the list $(v_1, \ldots, v_n)$ is independent, $v_1 \neq 0_{\mathcal{V}}$, and so we may define $u_1 := v_1 / \|v_1\|$. Then $\|u_1\| = 1$ and $\text{span}(v_1) = \text{span}(u_1)$, since u_1 is just a nonzero scalar multiple of v_1.

2. *The induction hypothesis and step.* Suppose that the result is true for some $n \geq 1$ and consider an independent list $(v_1, \ldots, v_n, v_{n+1})$. Define the list $(u_1, \ldots, u_n)$ by (28.8). Then $(u_1, \ldots, u_n)$ is orthonormal and preserves spans in the sense of (28.9).

Consider now the vector

$$w_{n+1} := v_{n+1} - \sum_{k=1}^{n} \langle v_{n+1}, u_k \rangle u_k.$$

Let $1 \leq j \leq n$. We compute

$$\langle w_{n+1}, u_j \rangle = \langle v_{n+1}, u_j \rangle - \left\langle \sum_{k=1}^{n} \langle v_{n+1}, u_k \rangle u_k, u_j \right\rangle,$$

where

$$\left\langle \sum_{k=1}^{n} \langle v_{n+1}, u_k \rangle u_k, u_j \right\rangle = \sum_{k=1}^{n} \langle v_{n+1}, u_k \rangle \langle u_k, u_j \rangle = \langle v_{n+1}, u_j \rangle$$

since $\langle u_j, u_j \rangle = 1$ and $\langle u_k, u_j \rangle = 0$ for $k \neq j$. Thus

$$\langle w_{n+1}, u_j \rangle = \langle v_{n+1}, u_j \rangle - \langle v_{n+1}, u_j \rangle = 0,$$

and so the list $(u_1, \ldots, u_n, w_{n+1})$ is orthogonal (by Problem 27.16).

If $w_{n+1} = 0_{\mathcal{V}}$, then

$$v_{n+1} = \sum_{k=1}^{n} \langle v_n, u_k \rangle u_k \in \text{span}(u_1, \ldots, u_n) = \text{span}(v_1, \ldots, v_n),$$

and that contradicts the independence of $(v_1, \ldots, v_n, v_{n+1})$. We take $u_{n+1} := w_{n+1} / \|w_{n+1}\|$ to find that $(u_1, \ldots, u_n, u_{n+1})$ is orthonormal. ■

Content from *Linear Algebra* by Meckes & Meckes. Algorithm 4.11 on pp. 244–245 is Gram–Schmidt. The examples on pp. 245–246 show the sort of laborious calculations typically involved in doing Gram–Schmidt on an actual set of vectors. This is good practice, but it's even better to know the pseudocode for the algorithm and how to rediscover that in the first place.

Here are the good consequences of Gram–Schmidt for bases.

28.12 Problem ($\star$). Let $\mathcal{V}$ be a finite-dimensional inner product space. Prove the following.

(i) $\mathcal{V}$ has an orthonormal basis.

(ii) Any orthonormal list in $\mathcal{V}$ can be extended to an orthonormal basis for $\mathcal{V}$.

Content from *Linear Algebra* by Meckes & Meckes. This problem is Corollary 4.12 on p. 246 and Corollary 4.13 on p. 247.

As bases go, orthonormal bases really are the best: every finite-dimensional space has one, and we can easily read off the coefficients of a vector in the basis from the inner product—this is (28.1). In that sense, orthonormal bases are closely related to the coordinate functionals from the corresponding dual basis.

28.13 Problem (!). Let $(u_1, \ldots, u_n)$ be an orthonormal basis for the finite-dimensional inner product space $\mathcal{V}$. Show that the corresponding dual basis is $(\langle \cdot, u_1 \rangle, \ldots, \langle \cdot, u_n \rangle)$.

28.14 Problem ($\star$). Let $\mathcal{V}$ be a finite-dimensional vector space with basis $(v_1, \ldots, v_n)$, and let $(\varphi_1, \ldots, \varphi_n)$ be the associated dual basis. For v, $w \in \mathcal{V}$, define

$$\langle v, w \rangle := \sum_{j=1}^{n} \varphi_j(v)\overline{\varphi_j(w)}.$$

(i) Show that $\langle \cdot, \cdot \rangle$ is an inner product on $\mathcal{V}$.

(ii) Show that the original basis $(v_1, \ldots, v_n)$ is orthonormal with respect to this inner product.

Day 29: Friday, March 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."

Projection onto a vector, Cauchy–Schwarz inequality

The development of Gram–Schmidt subtly contained a tool (specifically, at the preliminary $j = 2$ step) that will help us develop more properties of the norm induced by the inner product. Here is that tool.

29.1 Definition. *Let $\mathcal{V}$ be an inner product space and $w \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$. The* **PROJECTION**

ONTO w *is the linear operator*

$$\mathcal{P}_w \colon \mathcal{V} \to \mathcal{V} \colon v \mapsto \frac{\langle v, w \rangle}{\|w\|^2} w. \tag{29.1}$$

If we revisit (28.7) in the motivational discussion for Gram–Schmidt, then (in the notation of that discussion) we should recognize

$$v_2 - \langle v_2, u_1 \rangle u_1 = v_2 - \left\langle v_2, \frac{v_1}{\|v_1\|} \right\rangle \frac{v_1}{\|v_1\|} = v_2 - \frac{\langle v_2, v_1 \rangle}{\|v_1\|^2} v_1 = v_2 - \mathcal{P}_{v_1} v_2.$$

The idea from Gram–Schmidt, then, was that $v_2 - \mathcal{P}_{v_1} v_2$ and v_1 are orthogonal; since $\mathcal{P}_{v_1} v_2$ is a scalar multiple of v_1, this also says that $v_2 - \mathcal{P}_{v_1} v_2$ and $\mathcal{P}_{v_1} v_2$ are orthogonal.

29.2 Problem (!). Prove the following about $\mathcal{P}_w$.

(i) $\mathcal{P}_w v = v$ if and only if $v \in \operatorname{span}(w)$.

(ii) $\langle v - \mathcal{P}_w v, w \rangle = \langle \mathcal{P}_w v, v - \mathcal{P}_w v \rangle = 0$.

(iii) $\mathcal{P}_w^2 = \mathcal{P}_w$.

(iv) $\operatorname{rank}(\mathcal{P}_w) = 1$.

29.3 Problem (!). Let $\mathcal{V} = \mathbb{R}^2$ with the usual Euclidean inner product. Let $\mathbf{v} = (2, 3)$ and $\mathbf{w} = (4, 0)$. Draw $\mathbf{v}$, $\mathbf{w}$, $\mathcal{P}_{\mathbf{w}} \mathbf{v}$, and $\mathbf{v} - \mathcal{P}_{\mathbf{w}} \mathbf{v}$ in the same picture. Are you surprised?

Content from *Linear Algebra* by Meckes & Meckes. Lemma 4.5 on p. 231 and the hugely useful Figure 4.1 on that page give somewhat different perspectives on $\mathcal{P}_w$. In the notation of that lemma, think of a as giving $aw = \mathcal{P}_w v$.

Here is a further application of projections onto vectors. Experience teaches us that going from point A to point B, and then (on the weekends) from point B to point C, should take longer than just going from point A to point C. This is the **TRIANGLE INEQUALITY**: $\|v + w\| \le \|v\| + \|w\|$ for all v, $w \in \mathcal{V}$, where $\|\cdot\|$ is the norm on $\mathcal{V}$ induced by its inner product..

29.4 Problem (!). Draw a picture illustrating the triangle inequality in $\mathbb{R}^2$. [Hint: *think about "tip-to-tail" addition of vectors.*]

Content from *Linear Algebra* by Meckes & Meckes. Figure 4.2 on p. 233 is that picture.

We would like to show that

$$\|v + w\| \le \|v\| + \|w\|.$$

Square roots are challenging, so we might try proving this inequality by showing instead

$$\|v+w\|^2 \le \big(\,\|v\| + \|w\|\,\big)^2.$$

29.5 Problem (!). Show that this is equivalent to

$$\mathrm{Re}[\langle v, w\rangle] \le \|v\|\,\|w\|\,. \tag{29.2}$$

[Hint: *use* $\|u\|^2 = \langle u, u\rangle$*, algebraic properties of the norm, and the identity* $z+\overline{z} = 2\,\mathrm{Re}(z)$*, valid for all* $z \in \mathbb{C}$*.*]

So, do we have this peculiar inequality (29.2)? Yes. In fact, we get something a bit stronger.

29.6 Problem (!). To motivate the following more general result, consider the dot product on $\mathbb{R}^2$. Compute

$$(\mathbf{v}\cdot\mathbf{w})^2 = v_1^2w_1^2 + 2(v_1w_2)(w_1v_2) + v_2^2w_2^2.$$

Use the inequality $2ab \le a^2+b^2$, valid for all a, $b \in \mathbb{R}$ (since $0 \le (a-b)^2$), to find

$$(\mathbf{v}\cdot\mathbf{w})^2 \le v_1^2w_1^2 + v_1^2w_2^2 + w_1^2v_2^2 + v_2^2w_2^2.$$

Factor the right side as

$$(v_1^2+v_2^2)(w_1^2+w_2^2) = \|\mathbf{v}\|^2\,\|\mathbf{w}\|^2\,.$$

29.7 Problem ($\star$). Here are some steps to prepare for the following general result.

(i) Prove that if $\mathcal{V}$ is an inner product space and v, $w \in \mathcal{V}$ with $\langle v, w\rangle = 0$, then

$$\|v+w\|^2 = \|v\|^2 + \|w\|^2\,. \tag{29.3}$$

[Hint: *just expand everything.*]

(ii) Draw a picture illustrating this in $\mathbb{R}^2$ and explain why it makes you think about right triangles and why we might call the equality (29.3) the **Pythagorean identity**.

(iii) Why does the Pythagorean identity not imply a triangle "equality" for orthogonal vectors? That is, why do we not have $\|v+w\| = \|v\| + \|w\|$ when $\langle v, w\rangle = 0$? [Hint: $\sqrt{a^2+b^2} \ne |a|+|b|$ *in general.*]

(iv) Generalize (29.3) to the case where $(v_1, \ldots, v_n)$ is an orthogonal list:

$$\left\|\sum_{j=1}^n v_j\right\|^2 = \sum_{j=1}^n \|v_j\|^2\,. \tag{29.4}$$

29.8 Problem (!). Let $(u_1, \ldots, u_n)$ be an orthonormal list in an inner product space $\mathcal{V}$ and $v \in \operatorname{span}(u_1, \ldots, u_n)$. Prove that

$$\|v\|^2 = \sum_{j=1}^{n} |\langle v, u_j \rangle|^2$$

in two ways: first using (28.2) and then with (29.4).

Content from *Linear Algebra* by Meckes & Meckes. Theorem 4.4 on p. 230 is the Pythagorean identity for n.

The following result will give us the desired inequality (29.2), and much more.

29.9 Theorem (Cauchy–Schwarz inequality). *Let $\mathcal{V}$ be an inner product space and v, $w \in \mathcal{V}$. Then*

$$|\langle v, w \rangle| \le \|v\| \|w\|.$$

Proof. There are many proofs of this inequality, none of them quite obvious. Perhaps the easiest place to start is the case $\langle v, w \rangle = 0$; then since $\|v\| \ge 0$ and $\|w\| \ge 0$, we have $0 \le \|v\| \|w\|$. So, when v and w are orthogonal, the Cauchy–Schwarz inequality is true. If v and w are not orthogonal, we might think about how we could make v "be as orthogonal to w as possible."

Here is where such a line of thought could lead us. First, if $w = 0_{\mathcal{V}}$, then certainly equality holds in the Cauchy–Schwarz inequality. Assume $w \ne 0_{\mathcal{V}}$, so we can introduce the projection operator $\mathcal{P}_w$ from (29.1). It is straightforward to check that the Cauchy–Schwarz inequality is equivalent to $\|\mathcal{P}_w v\| \le \|v\|$, and drawing pictures in $\mathbb{R}^2$ alone should convince us that this latter inequality is true.

Here is the actual proof. Adding zero, we have

$$v = \mathcal{P}_w v + (v - \mathcal{P}_w v),$$

and also

$$\langle \mathcal{P}_w v, v - \mathcal{P}_w v \rangle = 0$$

by Problem 29.2. Then the Pythagorean identity (29.3) implies

$$\|v\|^2 = \|\mathcal{P}_w v + (v - \mathcal{P}_w v)\|^2 = \|\mathcal{P}_w v\|^2 + \|v - \mathcal{P}_w v\|^2.$$

Since $\|v - P_w v\|^2 \ge 0$, we have

$$\|v\|^2 \ge \|\mathcal{P}_w v\|^2.$$

Thus $\|\mathcal{P}_w v\| \le \|v\|$, as desired. ■

29.10 Problem ($\star$). Reread the proof of the Cauchy–Schwarz inequality and find exactly where an inequality appears for the first time. What would make that inequality an equality? Determine a condition on v and w that is equivalent to *equality* in the Cauchy–

Schwarz *in*equality. [Hint: *by part (i) of Problem 29.2,* $\mathcal{P}_w v = v$ *if and only if* $v \in \operatorname{span}(w)$.]

29.11 Problem (+). Here is another, equally nonintuitive, proof of Cauchy–Schwarz that introduces a useful technique. Let $\mathcal{V}$ be an inner product space and v, $w \in \mathcal{V}$.

(i) First suppose that (v, w) is dependent. Explain why the Cauchy–Schwarz inequality immediately follows if one of v or w is $0_{\mathcal{V}}$. Then assume that both are nonzero and prove the inequality.

(ii) Now assume that (v, w) is independent. For $\alpha \in \mathbb{F}$, put

$$p(\alpha) = \langle v + \alpha w, v + \alpha w \rangle$$

and show that

$$p(\alpha) = \|v\|^2 + 2\operatorname{Re}[\overline{\alpha}\langle v, w\rangle] + |\alpha|^2 \|w\|^2 .$$

(iii) If $\langle v, w\rangle = 0$, then the Cauchy–Schwarz inequality already holds, so assume that $\langle v, w\rangle \neq 0$. For $t \in \mathbb{F}$, put

$$q(t) := p\left(\frac{t\langle v, w\rangle}{|\langle v, w\rangle|}\right).$$

Continue to assume that (v, w) is independent and conclude $q(t) > 0$ for all t. Use the discriminant of q to obtain the Cauchy–Schwarz inequality.

(iv) Last, suppose that equality in the Cauchy–Schwarz inequality holds: $|\langle v, w\rangle| = \|v\|\|w\|$. If $\langle v, w\rangle = 0$, conclude that (v, w) is dependent. If $\langle v, w\rangle \neq 0$, show that $q(t) = 0$ for some $t \in \mathbb{R}$. From this, obtain that $p(\alpha) = 0$ for some $\alpha \in \mathbb{F}$, and use that to conclude that (v, w) is, again, dependent.

Content from *Linear Algebra* by Meckes & Meckes. Theorem 4.6 on p. 232 is Cauchy–Schwarz. Read the footnote, preferably aloud.

Now we can return to proving the triangle inequality. We want to show $\|v + w\| \leq \|v\| + \|w\|$, and we will have this if we can show

$$\operatorname{Re}[\langle v, w\rangle] \leq \|v\|\|w\| .$$

Recall that for any $z \in \mathbb{C}$, we have $\operatorname{Re}(z) \leq |\operatorname{Re}(z)| \leq |z|$. Thus

$$\operatorname{Re}[\langle v, w\rangle] \leq |\langle v, w\rangle| \leq \|v\|\|w\|$$

by the Cauchy–Schwarz inequality. This completes the proof of the triangle inequality for the norm induced by the inner product.

29.12 Problem (!). Let $\mathcal{V} = \mathcal{C}([0, 1])$ with the usual inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$. Let $f(x) = 1$ and $g(x) = x$. Compute $\|f\|$, $\|g\|$, $\|f + g\|$, and $\langle f, g\rangle$. Check that the

Cauchy–Schwarz and triangle inequalities hold. Are there any equalities?

29.13 Problem ($\star$). Reread the proof of the triangle inequality above, which was interrupted by several digressions involving the Cauchy–Schwarz inequality. (It might help to rewrite the proof of the triangle inequality without those digressions.) Determine a condition on v and w that is equivalent to equality in the triangle inequality. [Hint: *think about Problem 29.5 and when equality holds in the Cauchy–Schwarz inequality.*]

Content from *Linear Algebra* by Meckes & Meckes. Theorem 4.7 on p. 232 is the triangle inequality, with a much more concise proof than the discussion here.

The Cauchy–Schwarz inequality and a little calculus allow us to establish a third major example of an inner product space. This is a sequence space and, as such, it sits nicely between $\mathbb{F}^n$ with the dot product and $\mathcal{C}([0,1])$ with the L^2-inner product.

29.14 Example. Let

$$\ell^2 := \left\{ (a_k) \in \mathbb{F}^\infty \;\middle|\; \sum_{k=1}^\infty |a_k|^2 < \infty \right\}.$$

We pronounce the symbol ℓ^2 as "little ell two" or, sometimes, just "ell two," and we say that sequences in ℓ^2 are **SQUARE SUMMABLE**.

1. We first show that ℓ^2 is a subspace of $\mathbb{F}^\infty$. Certainly the zero sequence is square summable. Next, let $\alpha \in \mathbb{F}$ and $(a_k) \in \ell^2$. The series $\sum_{k=1}^\infty |a_k|^2$ converges, and therefore $\sum_{k=1}^\infty |\alpha|^2|a_k|^2$ also converges. And $|\alpha|^2|a_k|^2 = |\alpha a_k|^2$.

Finally, let (a_k), $(b_k) \in \ell^2$, so the series $\sum_{k=1}^\infty |a_k|^2$ and $\sum_{k=1}^\infty |b_k|^2$ converge. Why does the series $\sum_{k=1}^\infty |a_k + b_k|^2$ converge?

First, the triangle inequality says $|a_k + b_k| \le |a_k| + |b_k|$. Squaring, we have

$$|a_k + b_k|^2 \le |a_k|^2 + 2|a_k b_k| + |b_k|^2 =: c_k.$$

We already know that the series $\sum_{k=1}^\infty |a_k|^2$ and $\sum_{k=1}^\infty |b_k|^2$ converge. If we can also show that the series $\sum_{k=1}^\infty |a_k b_k|$ converges, then the series $\sum_{k=1}^\infty c_k$ will converge, and so the comparison test will guarantee that $\sum_{k=1}^\infty |a_k + b_k|^2$ converges.

Here we need a calculus result: a sequence that is increasing and bounded above converges. We apply this to series by supposing that (d_k) is a sequence for which there is $C > 0$ such that $\sum_{k=1}^n |d_k| \le C$ for all $n \ge 1$. Then this "monotone convergence" result implies the convergence of $\sum_{k=1}^\infty |d_k|$.

Accepting this to be true, we use the Cauchy–Schwarz inequality in $\mathbb{F}^n$ to estimate

$$\sum_{k=1}^n |a_k b_k| = |\mathbf{a}_n \cdot \mathbf{b}_n| \le \|\mathbf{a}_n\|_n \|\mathbf{b}_n\|_n,$$

where

$$\mathbf{a}_n := (|a_1|, \ldots, |a_n|), \qquad \mathbf{b}_n := (|b_1|, \ldots, |b_n|), \qquad \|\mathbf{v}\|_n := \sqrt{\mathbf{v} \cdot \mathbf{v}}, \ \mathbf{v} \in \mathbb{F}^n.$$

Observe that

$$\|\mathbf{a}_n\|_n^2 = \sum_{k=1}^{n} |a_k|^2 \le \sum_{k=1}^{\infty} |a_k|^2,$$

and so for each n we have

$$\sum_{k=1}^{n} |a_k b_k| \le \left(\sum_{k=1}^{\infty} |a_k|^2\right)^{1/2} \left(\sum_{k=1}^{\infty} |b_k|^2\right)^{1/2} =: C.$$

We can therefore use that calculus result to ensure the convergence of $\sum_{k=1}^{\infty} |a_k b_k|$, and from that we are done.

2. We define an inner product on ℓ^2 by

$$\langle (a_k), (b_k) \rangle := \sum_{k=1}^{\infty} a_k \overline{b_k}.$$

We saw above that if (a_k), $(b_k) \in \ell^2$, then the series $\sum_{k=1}^{\infty} |a_k b_k|$ converges. Since

$$|a_k \overline{b_k}| = |a_k||\overline{b_k}| = |a_k||b_k| = |a_k b_k|,$$

the series $\sum_{k=1}^{\infty} |a_k \overline{b_k}|$ converges, and so this inner product is indeed defined. That it satisfies the properties of an inner product is the same sort of arithmetic and reasoning as in the proof for the dot product on Euclidean space in Example 27.1.

29.15 Problem (!). Explain why $(1/k) \in \ell^2$. [Hint: *p-series.*]

Content from *Linear Algebra* by Meckes & Meckes. Example 3 on pp. 227–228 and Example 4 on p. 235 discuss ℓ^2.

Day 30: Monday, March 30.

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."

Orthogonal complement

Often we are interested in a very particular subspace of a vector space—frequently, in the context of the operator equation $\mathcal{T}v = w$, the space is the range or kernel, but also perhaps because a subspace consists of particularly interesting and tame vectors (say, the at-most-degree-n polynomials in the vast space $\mathcal{C}([0,1])$). We might wonder how a given subspace "sits" within the larger ambient space. What is missing from the subspace to fill out the whole space? How can we characterize membership in the subspace?

Here is a motivating example.

30.1 Example. Consider the inner product space $\mathbb{F}^3$ with the dot product. Let $\mathcal{U}_1 := \operatorname{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $\mathcal{U}_2 := \operatorname{span}(\mathbf{e}_3)$. Then any $\mathbf{v} \in \mathbb{F}^3$ can be written uniquely in the form $\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2$ for some $\mathbf{u}_1 \in \mathcal{U}_1$ and $\mathbf{u}_2 \in \mathcal{U}_2$. This is hopefully obvious from physical experience with three-dimensional space, or from arithmetic:

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

Also, if $\mathbf{u}_1 \in \mathcal{U}_1$ and $\mathbf{u}_2 \in \mathcal{U}_2$, we have $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$. In fact,

$$\mathcal{U}_1 = \left\{ \mathbf{v} \in \mathbb{F}^3 \;\middle|\; \mathbf{v} \cdot \mathbf{u} = 0 \text{ for all } \mathbf{u} \in \mathcal{U}_2 \right\}$$

and

$$\mathcal{U}_2 = \left\{ \mathbf{w} \in \mathbb{F}^3 \;\middle|\; \mathbf{w} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{v} \in \mathcal{U}_1 \right\}.$$

We can therefore think that $\mathcal{U}_1$ and $\mathcal{U}_2$ induce an "orthogonal decomposition" of $\mathbb{F}^3$: each vector in $\mathbb{F}^3$ can be written uniquely as a sum of vectors in $\mathcal{U}_1$ and $\mathcal{U}_2$, and those vectors are orthogonal. Moreover, $\mathcal{U}_1$ contains all vectors that are orthogonal to all vectors in $\mathcal{U}_2$, and $\mathcal{U}_2$ contains all vectors that are orthogonal to all vectors in $\mathcal{U}_1$.

Here is how $\mathcal{U}_1$ and $\mathcal{U}_2$ in the previous example "complement" each other.

30.2 Definition. *Let $\mathcal{V}$ be an inner product space and $\mathcal{U} \subseteq \mathcal{V}$. The* ***ORTHOGONAL COMPLEMENT*** *of $\mathcal{U}$ is*

$$\mathcal{U}^\perp := \{ v \in \mathcal{V} \mid \langle u, v \rangle = 0 \text{ for all } u \in \mathcal{U} \}.$$

Content from *Linear Algebra* by Meckes & Meckes. Page 252 defines orthogonal complements. See the examples on that page, in particular Figure 4.4.

30.3 Example. **(i)** Returning to Example 30.1, if

$$\mathcal{U} = \left\{ \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} \in \mathbb{F}^3 \;\middle|\; v_1,\ v_2 \in \mathbb{F} \right\},$$

then

$$\mathcal{U}^\perp = \left\{ \begin{bmatrix} 0 \\ 0 \\ v_3 \end{bmatrix} \in \mathbb{F}^3 \;\middle|\; v_3 \in \mathbb{F} \right\},$$

and we have $(\mathcal{U}^\perp)^\perp = \mathcal{U}$.

(ii) Consider $\mathcal{C}([0,1])$ with the L^2-inner product. Denote by 1 the function $1(x) := 1$ and let $\mathcal{U} = \operatorname{span}(1)$, so $\mathcal{U}$ consists of all constant functions on $[0,1]$. We have $f \in \{1\}^\perp$ if and only if $\langle f, 1\rangle = 0$, thus if and only if $\int_0^1 f(x)\ dx = 0$. Then

$$\mathcal{U}^\perp = \left\{ f \in \mathcal{C}([0,1]) \;\middle|\; \int_0^1 f(x)\ dx = 0 \right\}.$$

Do we have $(\mathcal{U}^\perp)^\perp = \mathcal{U}$? That is, $g \in (\mathcal{U}^\perp)^\perp$, then $\int_0^1 g(x)f(x)\ dx = 0$ for all $f \in \mathcal{C}([0,1])$ such that $\int_0^1 f(x)\ dx = 0$, does this mean that g is constant? The answer turns out to be yes, but showing this just using calculus alone might be hard. We will do this with linear algebra.

30.4 Problem (!). Consider the inner product space $\mathcal{C}([-1,1])$ with the L^2-inner product. Let

$$\mathcal{U} := \{ f \in \mathcal{C}([-1,1]) \mid f(x) = f(-x) \text{ for all } x \in [-1,1] \}.$$

Describe as precisely as possible all functions in $\mathcal{U}^\perp$. [Hint: *If $a > 0$ and $h \in \mathcal{C}([-a,a])$ is odd, then $\int_{-a}^a h(x)\ dx = 0$. Also, every $h \in \mathcal{C}([-a,a])$ can be written in the form $h = h_e + h_o$, where*

$$h_e(x) = \frac{h(x) + h - x)}{2} \quad \text{and} \quad h_o(x) = \frac{h(x) - h(-x)}{2}.$$

For $g \in \mathcal{U}^\perp$, obtain $\langle g_e, f\rangle = 0$ for all $f \in \mathcal{U}$, and then make a useful choice for f.]

The following properties of orthogonal complements follow mostly from the definition and will be very useful.

30.5 Problem (⋆). Let $\mathcal{V}$ be an inner product space and $\mathcal{U} \subseteq \mathcal{V}$.

(i) Prove that $\mathcal{U}^\perp$ is a subspace of $\mathcal{V}$. (Your proof should not require that $\mathcal{U}$ be a subspace.)

(ii) Prove that if $\mathcal{U}$ is a subspace of $\mathcal{V}$, then $\mathcal{U} \cap \mathcal{U}^\perp = \{0_\mathcal{V}\}$. What happens if we do not assume that $\mathcal{U}$ is a subspace here?

(iii) What are $\mathcal{V}^\perp$ and $\{0_\mathcal{V}\}^\perp$?

(iv) If $v_1, \ldots, v_n \in \mathcal{V}$, prove that

$$(\operatorname{span}(v_1, \ldots, v_n))^\perp = \{ w \in \mathcal{V} \mid \langle w, v_j\rangle = 0,\ j = 1, \ldots, n \}.$$

30.6 Problem (+). Let $\mathcal{V}$ be an inner product space and $\mathcal{U}$ be a subspace of $\mathcal{V}$. The orthogonal complement can be characterized completely in terms of the norm:

$$\mathcal{U}^\perp = \{ v \in \mathcal{V} \mid \|v - u\| \geq \|v\| \text{ for all } u \in \mathcal{U} \}. \tag{30.1}$$

(i) Let $\mathcal{V} = \mathbb{R}^2$ with the dot product and $\mathcal{U} = \text{span}(\mathbf{e}_1)$. Draw a picture illustrating what this result is saying about $\mathcal{U}^\perp = \text{span}(\mathbf{e}_2)$.

(ii) Suppose that $v \in \mathcal{U}^\perp$ and $u \in \mathcal{U}$. Use the Pythagorean identity to prove that $\|v - u\| \geq \|v\|$.

(iii) Now suppose that $v \in \mathcal{V}$ satisfies $\|v\| \leq \|v - u\|$ for all $u \in \mathcal{U}$. Fix $u \in \mathcal{U}$; the goal is $\langle v, u \rangle = 0$. Inspired by Problem 29.11, for $\alpha \in \mathbb{F}$, put $p(\alpha) := \|v - \alpha u\|^2$. Conclude that

$$\text{Re}[\overline{\alpha}\,\langle v, u \rangle] \leq \frac{|\alpha|^2\,\|u\|^2}{2}.$$

By taking $\alpha \in \mathbb{R}$ and considering the limit as $\alpha \to 0^\pm$ of this inequality (consider the left and right limits separately), conclude $\text{Re}[\langle v, u \rangle] = 0$. Then take $\alpha = i\beta$ with $\beta \in \mathbb{R}$, consider the limit as $\beta \to 0^\pm$, and conclude $\text{Im}[\langle v, u \rangle] = 0$, too.

Here is a slightly trickier property that partially addresses the question raised in part (ii) of Example 30.3.

30.7 Lemma. *Let $\mathcal{U}$ be a subspace of the inner product space $\mathcal{V}$. Then $\mathcal{U} \subseteq (\mathcal{U}^\perp)^\perp$.*

Proof. Let $u \in \mathcal{U}$. We want to show that $u \in (\mathcal{U}^\perp)^\perp$. That is, we want to show that $\langle u, v \rangle = 0$ for all $v \in \mathcal{U}^\perp$. So, suppose that $v \in \mathcal{U}^\perp$. Because $u \in \mathcal{U}$, we have $\langle v, u \rangle = 0$, and therefore $\langle u, v \rangle = 0$, as desired. ■

Content from *Linear Algebra* by Meckes & Meckes. This proof is the first paragraph of the proof of Proposition 4.15 on p. 254.

From now on we will write

$$\mathcal{U}^{\perp\perp} := (\mathcal{U}^\perp)^\perp.$$

Here is a situation in which $\mathcal{U}^{\perp\perp}$ is much larger than $\mathcal{U}$.

30.8 Example. Consider the inner product space ℓ^2 of Example 29.14 with $\langle f, g \rangle = \sum_{k=1}^\infty f(k)\overline{g(k)}$ for $f,\ g \in \ell^2$. It will be notationally convenient to write vectors in ℓ^2 as functions on $\mathbb{N}$ (which they are). Put

$$\mathcal{U} := \left\{ f \in \ell^2 \;\middle|\; f(k) = 0 \text{ for all but finitely many } k \right\}.$$

So, for example, $(1/k) \notin \mathcal{U}$, but $(1, 0, 0, \ldots) \in \mathcal{U}$. In fact, with

$$e_j \colon \mathbb{N} \to \mathbb{F} \colon k \mapsto \begin{cases} 1, & j = k \\ 0, & j \neq k, \end{cases} \tag{30.2}$$

we have $e_j \in \mathcal{U}$ for all j.

Now let $g \in \mathcal{U}^\perp$. Then $\langle f, g \rangle = 0$ for all $f \in \mathcal{U}$. In particular, $0 = \langle e_j, g \rangle = g(j)$ for each $j \in \mathbb{N}$, so $g = 0$. That is, $\mathcal{U}^\perp = \{0\}$, and so $\mathcal{U}^{\perp\perp} = \{0\}^\perp = \ell^2$. But certainly $\mathcal{U} \neq \ell^2$.

We will develop conditions on an inner product space $\mathcal{V}$ and a subspace $\mathcal{U}$ that guarantee $\mathcal{U} = \mathcal{U}^{\perp\perp}$. This turns out to be closely related to a more pressing question: for what inner product spaces $\mathcal{V}$ and subspaces $\mathcal{U}$ does the situation of Example 30.1 hold—that for each $v \in \mathcal{V}$, there exist $u \in \mathcal{U}$ and $u^{\perp} \in \mathcal{U}^{\perp}$ such that $v = u + u^{\perp}$? Such u and $u^{\perp}$, if they do exist, are necessarily unique.

30.9 Lemma. *Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. If u_1, $u_2 \in \mathcal{U}$ and $u_1^{\perp}$, $u_2^{\perp} \in \mathcal{U}^{\perp}$ such that $u_1 + u_1^{\perp} = u_2 + u_2^{\perp}$, then $u_1 = u_2$ and $u_1^{\perp} = u_2^{\perp}$.*

Proof. If $u_1 + u_1^{\perp} = u_2 + u_2^{\perp}$, then $u_1 - u_2 = u_2^{\perp} - u_1^{\perp} \in \mathcal{U}^{\perp}$. But u_1, $u_2 \in \mathcal{U}$, so $u_1 - u_2 \in \mathcal{U}$. Hence $u_1 - u_2 \in \mathcal{U} \cap \mathcal{U}^{\perp} = \{0_{\mathcal{V}}\}$, so $u_1 = u_2$. The same reasoning (invoking the fact that $\mathcal{U}^{\perp}$ is a subspace) shows $u_2^{\perp} - u_1^{\perp}$. ■

Content from *Linear Algebra* by Meckes & Meckes. This proof is the last paragraph of the proof of Theorem 4.14 on p. 253.

We do have this orthogonal decomposition when $\mathcal{U}$ is finite-dimensional.

30.10 Theorem. *Let $\mathcal{V}$ be an inner product space (not necessarily finite-dimensional) and let $\mathcal{U}$ be a finite-dimensional subspace. For each $v \in \mathcal{V}$, there exist $u \in \mathcal{U}$ and $u^{\perp} \in \mathcal{U}^{\perp}$ such that $v = u + u^{\perp}$.*

Proof. Uniqueness is Lemma 30.9, so we prove only existence. If we know u, then $u^{\perp}$ must be $u^{\perp} = v - u$, so we focus on finding u first. If $\mathcal{U} = \{0_{\mathcal{V}}\}$, then $\mathcal{U}^{\perp} = \mathcal{V}$, so we can take $u = 0$ and $u^{\perp} = v$.

Otherwise, let $\dim(\mathcal{U}) = n \geq 1$ and take an orthonormal basis $(u_1, \ldots, u_n)$ for $\mathcal{V}$. Any $u \in \mathcal{U}$ has the form $u = \sum_{j=1}^{n} \langle u, u_j \rangle u_j$. This is true but not helpful here, as it does not tell us what u is in terms of v. Instead, we just write $u = \sum_{j=1}^{n} \alpha_j u_j$, and we use what u and $u^{\perp}$ do to extract the coefficients α_j. What they do is satisfy $v = u + u^{\perp} = \sum_{j=1}^{n} \alpha_j u_j + u^{\perp}$ with $\langle u^{\perp}, u_k \rangle = 0$ for all k, since $u_k \in \mathcal{U}$. Thus we must have

$$\langle v, u_k \rangle = \left\langle \sum_{j=1}^{n} \alpha_j u_j + u^{\perp}, u_k \right\rangle = \sum_{j=1}^{n} \alpha_j \langle u_j, u_k \rangle + \langle u^{\perp}, u_k \rangle = \alpha_j.$$

Here we have used the orthonormality of the basis.

This gives us our *candidate* for u:

$$u = \sum_{j=1}^{n} \langle v, u_j \rangle u_j.$$

We definitely have $u \in \mathcal{U}$, and so we just need to check that $u^{\perp} := v - u \in \mathcal{U}^{\perp}$. That is, we need

$$\left\langle v - \sum_{j=1}^{n} \langle v, u_j \rangle u_j, w \right\rangle = 0$$

for all $w \in \mathcal{U}$. By part (iv) of Problem 30.5, it suffices to prove this just for $w = u_j$, $j = 1, \ldots, n$, and that is a quick calculation. ■

30.11 Problem (!). Do it.

Content from *Linear Algebra* by Meckes & Meckes. Theorem 30.10 is Theorem 4.14 on p. 253.

Day 31: Wednesday, April 1.

Vocabulary from today

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

Orthogonal direct sum of subspaces, orthogonal projection onto a subspace

We give a precise name and notation to this decomposition phenomenon.

31.1 Definition. *Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}_1$ and $\mathcal{U}_2$ be subspaces of $\mathcal{V}$. Then $\mathcal{V}$ is the* **ORTHOGONAL DIRECT SUM** *of $\mathcal{U}_1$ and $\mathcal{U}_2$, written $\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2$, if for all $v \in \mathcal{V}$, there exist unique $u_1 \in \mathcal{U}_1$ and $u_2 \in \mathcal{U}_2$ such that $v = u_1 + u_2$ and $\langle u_1, u_2 \rangle = 0$.*

The uniqueness condition in this definition is unnecessary.

31.2 Problem (!). Adapt the proof of Lemma 30.9 to show that if $\mathcal{V}$ is an inner product space and u_1, u_2, $\widetilde{u}_1$, $\widetilde{u}_2 \in \mathcal{V}$ with $u_1 + u_2 = \widetilde{u}_1 + \widetilde{u}_2$ and $\langle u_1, u_2 \rangle = \langle \widetilde{u}_1, \widetilde{u}_2 \rangle = 0$, then $u_1 = \widetilde{u}_1$ and $u_2 = \widetilde{u}_2$.

31.3 Example. $\mathbb{F}^3 = \operatorname{span}(\mathbf{e}_1, \mathbf{e}_2) \oplus \operatorname{span}(\mathbf{e}_3)$.

Content from *Linear Algebra* by Meckes & Meckes. See the definition, and notational remarks, on p. 254, and then read the examples after that.

31.4 Problem ($\star$). Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}$ be a nontrivial proper (i.e., $\mathcal{U} \neq \{0_{\mathcal{V}}\}$ and $\mathcal{U} \neq \mathcal{V}$) subspace of $\mathcal{V}$.

(i) Why should we expect $\mathcal{V} \neq \mathcal{U} \cup \mathcal{U}^{\perp}$? [Hint: *if $u \in \mathcal{U} \setminus \{0_{\mathcal{V}}\}$ and $u^{\perp} \in \mathcal{U}^{\perp} \setminus \{0_{\mathcal{V}}\}$, do we have $u + u^{\perp}$ in that union?*] With $\mathcal{V} = \mathbb{R}^2$ (and the inner product being, as always, the dot product) and $\mathcal{U} = \operatorname{span}(\mathbf{e}_1)$, draw a picture illustrating why $\mathbb{R}^2 \neq \operatorname{span}(\mathbf{e}_1) \cup \operatorname{span}(\mathbf{e}_1)^{\perp}$.

(ii) Suppose that $\mathcal{V}$ is finite-dimensional, which means that both $\mathcal{U}$ and $\mathcal{U}^{\perp}$ are finite-dimensional, too, and also $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$. Let $(u_1, \ldots, u_n)$ be a basis for $\mathcal{U}$ and $(u_1^{\perp}, \ldots, u_m^{\perp})$ be a basis for $\mathcal{U}^{\perp}$. Prove that $(u_1, \ldots, u_n, u_1^{\perp}, \ldots, u_m^{\perp})$ is a basis for $\mathcal{V}$ and thus $\dim(\mathcal{V}) = \dim(\mathcal{U}) + \dim(\mathcal{U}^{\perp})$. [Hint: *for spanning, use* $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$*; for independence, consider what happens when* $\sum_{k=1}^{n} \alpha_k u_k = \sum_{k=1}^{m} \beta_k u_k^{\perp}$ *for* α_k, $\beta_k \in \mathbb{F}$.]

(iii) Here is a variation on part (ii). Say that $(w_1, \ldots, w_p)$ is an orthonormal basis for $\mathcal{V}$ and $1 \le r < p$. Prove that $(w_{r+1}, \ldots, w_p)$ is an orthonormal basis for $\operatorname{span}(w_1, \ldots, w_p)^{\perp}$.

Our eventual goal will be to develop a more general condition on an inner product space $\mathcal{V}$ and subspaces $\mathcal{U}$ that permits the orthogonal direct sum decomposition $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$ without requiring $\mathcal{U}$ to be finite-dimensional. For now, we explore some consequences of this decomposition.

So, suppose that $\mathcal{V}$ is an inner product space and $\mathcal{U}$ is a subspace of $\mathcal{V}$ with $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$. Then for each $v \in \mathcal{V}$, there are unique $u \in \mathcal{U}$ and $u^{\perp} \in \mathcal{U}^{\perp}$ such that $v = u + u^{\perp}$. Given v, then, we define $\mathcal{P}_{\mathcal{U}}(v) := u$. That is, $\mathcal{P}_{\mathcal{U}}(v)$ satisfies $\mathcal{P}_{\mathcal{U}}(v) \in \mathcal{U}$ and $v - \mathcal{P}_{\mathcal{U}}(v) \in \mathcal{U}^{\perp}$; we could also say

$$\mathcal{P}_{\mathcal{U}} = \left\{ (v, u) \;\middle|\; u \in \mathcal{U},\ v - u \in \mathcal{U}^{\perp} \right\}. \tag{31.1}$$

This $\mathcal{P}_{\mathcal{U}}(v)$ is unique, and so $\mathcal{P}_{\mathcal{U}} \colon \mathcal{V} \to \mathcal{U}$ really is a function. What is more is that $\mathcal{P}_{\mathcal{U}}$ is linear.

31.5 Theorem. *The map* $\mathcal{P}_{\mathcal{U}}$ *defined in* (31.1) *has the following properties.*

(i) $\mathcal{P}_{\mathcal{U}} \in \mathbf{L}(\mathcal{V}, \mathcal{U})$.

(ii) $\mathcal{P}_{\mathcal{U}}^2 = \mathcal{P}_{\mathcal{U}}$.

(iii) *Let* $v \in \mathcal{V}$*. Then* $v \in \mathcal{U}$ *if and only if* $v = \mathcal{P}_{\mathcal{U}} v$.

This operator $\mathcal{P}_{\mathcal{U}}$ *is the* ***ORTHOGONAL PROJECTION*** *onto* $\mathcal{U}$.

Proof. **(i)** Let v_1, $v_2 \in \mathcal{V}$. Write $v_1 = u_1 + u_1^{\perp}$ and $v_2 = u_2 + u_2^{\perp}$ for u_1, $u_2 \in \mathcal{U}$ and $u_1^{\perp}$, $u_2^{\perp} \in \mathcal{U}^{\perp}$. Then $v_1 + v_2 = (u_1 + u_2) + (u_1^{\perp} + u_2^{\perp})$. Since $\mathcal{U}$ and $\mathcal{U}^{\perp}$ are subspaces of $\mathcal{V}$, we have $u_1 + u_2 \in \mathcal{U}$ and $u_1^{\perp} + u_2^{\perp} \in \mathcal{U}^{\perp}$. By the uniqueness of the orthogonal decomposition, $\mathcal{P}_{\mathcal{U}}(v_1 + v_2) = u_1 + u_2 = \mathcal{P}_{\mathcal{U}}(v_1) + \mathcal{P}_{\mathcal{U}}(v_2)$. A similar proof shows $\mathcal{P}_{\mathcal{U}}(\alpha v) = \alpha \mathcal{P}_{\mathcal{U}}(v)$.

(ii) Let $v \in \mathcal{V}$ and write $v = u + u^{\perp}$ with $u \in \mathcal{U}$ and $u^{\perp} \in \mathcal{U}^{\perp}$. Then $\mathcal{P}_{\mathcal{U}} v = u$, so $\mathcal{P}_{\mathcal{U}}^2 v = \mathcal{P}_{\mathcal{U}} u$. Now, $u = u + 0_{\mathcal{V}}$, where $u \in \mathcal{U}$ and $0_{\mathcal{V}} \in \mathcal{U}^{\perp}$, so $\mathcal{P}_{\mathcal{U}} u = u$. ■

31.6 Problem (!). **(i)** Finish the proof that $\mathcal{P}_{\mathcal{U}}$ is linear by showing $\mathcal{P}_{\mathcal{U}}(\alpha v) = \alpha \mathcal{P}_{\mathcal{U}}(v)$.

(ii) Prove part (iii) of the theorem. [Hint: *if* $v \in \mathcal{U}$*, write* $v = v + 0_{\mathcal{V}}$*, and then use uniqueness of the decomposition.*]

An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ such that $\mathcal{T}^2 = \mathcal{T}$ is called **IDEMPOTENT**. The idempotency of $\mathcal{P}_{\mathcal{U}}$ means that the subspace $\mathcal{P}_{\mathcal{U}}(\mathcal{V})$ is **INVARIANT** under $\mathcal{P}_{\mathcal{U}}$: $\mathcal{P}_{\mathcal{U}}(\mathcal{P}_{\mathcal{U}}(V)) = \mathcal{P}_{\mathcal{U}}(\mathcal{V})$, so applying $\mathcal{P}_\mathcal{U}$ again to a vector in its range will not remove that vector from the range.

Content from *Linear Algebra* by Meckes & Meckes. Theorem 4.16 on p. 255 contains many properties of $\mathcal{P}_\mathcal{U}$. Some are valid only under the assumption that $\mathcal{U}$ is finite-dimensional.

31.7 Example. If $\mathcal{U}$ is a finite-dimensional subspace of the inner product space $\mathcal{V}$, and if $(u_1, \ldots, u_n)$ is an orthonormal basis for $\mathcal{U}$, then

$$\mathcal{P}_\mathcal{U} v = \sum_{j=1}^{n} \langle v, u_j \rangle u_j.$$

This is the result of Theorem 30.10.

31.8 Problem (!). Let $\mathcal{V}$ be an inner product space, $w \in \mathcal{V} \setminus \{0_\mathcal{V}\}$, and $\mathcal{U} = \operatorname{span}(w)$. Prove that $\mathcal{P}_\mathcal{U} = \mathcal{P}_w$ with $\mathcal{P}_w$ defined in 29.1.

31.9 Problem ($\star$). Assuming the hypotheses and notation of Example 31.7, prove that $\mathcal{P}_\mathcal{U}$ is finite-rank and calculate $\operatorname{rank}(\mathcal{P}_\mathcal{U})$.

31.10 Problem (!). Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Suppose that for each $v \in \mathcal{V}$, there exists $u_0 \in \mathcal{U}$ such that $\langle u, u_0 \rangle = \langle u, v \rangle$ for all $u \in \mathcal{U}$. Prove that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$ and that $\mathcal{P}_\mathcal{U} v = u_0$. [Hint: *what do you know about $u_0 - v$?*]

We can now resolve part of the question of when a subspace $\mathcal{U}$ of an inner product space $\mathcal{V}$ satisfies $\mathcal{U} = \mathcal{U}^{\perp\perp}$: when $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$.

31.11 Theorem. *Let $\mathcal{V}$ be an inner product space and suppose that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$ for a subspace $\mathcal{U}$. Then $\mathcal{U} = \mathcal{U}^{\perp\perp}$.*

Proof. Lemma 30.7 gives $\mathcal{U} \subseteq \mathcal{U}^{\perp\perp}$, so we show $\mathcal{U}^{\perp\perp} \subseteq \mathcal{U}$. Let $v \in \mathcal{U}^{\perp\perp}$. Since $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$, we can write $v = u + u^\perp$ for some $u \in \mathcal{U}$ and $u^\perp \in \mathcal{U}^\perp$. If we can show that $u^\perp = 0_\mathcal{V}$, then we will have $v = u \in \mathcal{U}$.

We rewrite $u^\perp = v - u$. We have $u^\perp \in \mathcal{U}^\perp$, we are assuming $v \in \mathcal{U}^{\perp\perp} = (\mathcal{U}^\perp)^\perp$, and we know $u \in \mathcal{U} \subseteq \mathcal{U}^{\perp\perp} = (\mathcal{U}^\perp)^\perp$. So, $v - u \in (\mathcal{U}^\perp)^\perp$, and therefore $u^\perp \in \mathcal{U}^\perp \cap (\mathcal{U}^\perp)^\perp$. Recall that $\mathcal{W} \cap \mathcal{W}^\perp = \{0_\mathcal{V}\}$ for any subspace $\mathcal{W}$ of $\mathcal{V}$; here $\mathcal{W} = \mathcal{U}^\perp$. So, $u^\perp = 0_\mathcal{V}$, as desired. ■

31.12 Problem (!). In the proof of Theorem 31.11, show that $u^\perp = 0$ by computing $\langle v - u, u^\perp \rangle = 0$.

31.13 Problem ($\star$). Let $h \in \mathcal{C}([-1,1])$ have the property that $\int_{-1}^{1} g(x)h(x)\,dx = 0$ for all $g \in \mathcal{C}([-1,1])$ such that $\int_{-1}^{1} f(x)g(x)\,dx = 0$ for all even $f \in \mathcal{C}([-1,1])$. Prove that h is also even. [Hint: *use the hint, and result, of Problem 30.4 and Theorem 31.11.*]

31.14 Problem ($\star$). Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$.

(i) Suppose that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$. Show that $\mathcal{V}/\mathcal{U} = \{u^{\perp} + \mathcal{U} \mid u^{\perp} \in \mathcal{U}^{\perp}\}$.

(ii) Continue to assume that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$. Put

$$\mathcal{T} := \{(u^{\perp} + \mathcal{U}, u^{\perp}) \mid u^{\perp} \in \mathcal{U}^{\perp}\}.$$

Show that $\mathcal{T} \in \mathbf{L}(\mathcal{V}/\mathcal{U}, \mathcal{U}^{\perp})$ is an isomorphism.

(iii) Suppose that $\mathcal{V}$ is finite-dimensional. Use Problem 31.4 and Theorem 26.11 to give another argument that $\mathcal{V}/\mathcal{U}$ and $\mathcal{U}^{\perp}$ are isomorphic.

(iv) Conversely, assume that $\mathcal{T} \in \mathbf{L}(\mathcal{V}/\mathcal{U}, \mathcal{U}^{\perp})$ is an isomorphism with $\mathcal{T}(u^{\perp} + \mathcal{U}) = u^{\perp}$ for all $u^{\perp} \in \mathcal{U}^{\perp}$. Show that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$.

31.15 Problem ($\star$). **(i)** Here is a generalization of one step of part (ii) of Problem 31.4. Let $\mathcal{V}$ be a vector space with nontrivial subspaces $\mathcal{U}_1$ and $\mathcal{U}_2$ such that $\mathcal{U}_1 \cap \mathcal{U}_2 = \{0_{\mathcal{V}}\}$. Let $(v_1, \ldots, v_r)$ be an independent list in $\mathcal{U}_1$ and $(w_1, \ldots, w_s)$ be an independent list in $\mathcal{U}_2$. Prove that the list $(v_1, \ldots, v_r, w_1, \ldots, w_s)$ is independent.

(ii) Let $\mathcal{V}$ be a finite-dimensional vector space with $n := \dim(\mathcal{V}) \geq 1$. Suppose that $\mathcal{U}_1$ and $\mathcal{U}_2$ are subspaces of $\mathcal{V}$ such that $\dim(\mathcal{U}_1) + \dim(\mathcal{U}_2) > n$. Prove that $\mathcal{U}_1 \cap \mathcal{U}_2 \neq \{0_{\mathcal{V}}\}$.

Day 32: Friday, April 3.

Now we take up a question of approximation. Often it happens that a subspace $\mathcal{U}$ of an inner product space $\mathcal{V}$ has some particularly "nice" or tractable properties, and we would prefer to work with vectors in $\mathcal{U}$ over arbitrary vectors in $\mathcal{V}$. If $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$, then we can view $\mathcal{P}_{\mathcal{U}} v$ as giving the "best approximation" to $v \in \mathcal{V}$ by a vector in $\mathcal{U}$.

32.1 Problem (!). Let $\mathcal{V} = \mathbb{R}^2$ with the usual Euclidean inner product. Recall that $\mathcal{P}_{\mathrm{span}(w)} = \mathcal{P}_w$ for any $w \in \mathbb{R}^2$, with $\mathcal{P}_w$ defined in 29.1. Let $u = (2,3)$ and $w = (4,0)$. Draw a picture that shows why $\mathcal{P}_w u$ should be the "best approximation" to u in $\mathrm{span}(w)$. How does this resemble your picture from Problem 29.3?

32.2 Theorem. *Let $\mathcal{V}$ be an inner product space and suppose that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$ for a subspace $\mathcal{U}$. Then*

$$\|v - \mathcal{P}_{\mathcal{U}} v\| \leq \|v - u\| \tag{32.1}$$

for all $u \in \mathcal{U}$ with equality if and only if $u = \mathcal{P}_{\mathcal{U}}v$.

Proof. We begin with the classical trick of adding zero: for $u \in \mathcal{U}$ and $v \in \mathcal{V}$, we have

$$v - u = v - \mathcal{P}_{\mathcal{U}}v + \mathcal{P}_{\mathcal{U}}v - u.$$

Here $v - \mathcal{P}_{\mathcal{U}}v \in \mathcal{U}^{\perp}$ and $\mathcal{P}_{U}v - u \in \mathcal{U}$, so the Pythagorean identity gives

$$\|v-u\|^2 = \|(v - \mathcal{P}_{\mathcal{U}}v) + (\mathcal{P}_{\mathcal{U}}v - u)\|^2 = \|v - \mathcal{P}_{\mathcal{U}}v\|^2 + \|\mathcal{P}_{\mathcal{U}}v - u\|^2 \geq \|v - \mathcal{P}_{\mathcal{U}}v\|^2. \quad (32.2)$$

This proves (32.1).

Now we prove the if and only if statement. If $u = \mathcal{P}_{\mathcal{U}}v$, then certainly the inequality is an equality. Conversely, if the inequality is an equality and $\|v - \mathcal{P}_{\mathcal{U}}v\| = \|v - u\|$, then repeating the work with the Pythagorean identity gives

$$\|v - \mathcal{P}_{\mathcal{U}}v\|^2 + \|\mathcal{P}_{\mathcal{U}}v - u\|^2 = \|v-u\|^2 = \|v - \mathcal{P}_{\mathcal{U}}v\|^2,$$

thus $\|\mathcal{P}_{\mathcal{U}}v - u\|^2 = 0$, and so $u = \mathcal{P}_{\mathcal{U}}v$. ■

Content from *Linear Algebra* by Meckes & Meckes. This approximation result is part (ii) of Theorem 4.19 on p. 258. See Figure 4.6 and compare it to Figure 4.1 on p. 231.

32.3 Example. Let $\mathcal{V} = \mathcal{C}([-\pi,\pi])$ with the inner product $\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\ dx$, and let $\mathcal{U} = \operatorname{span}(1, \cos(\cdot), \sin(\cdot))$. Here we are working on $[-\pi,\pi]$, not $[0,1]$, for trigonometric convenience.

(i) It is easy to check that the list $(1, \cos(\cdot), \sin(\cdot))$ is orthogonal, and since

$$\int_{-\pi}^{\pi} 1\ dx = \frac{1}{2\pi}, \quad \text{and} \quad \int_{-\pi}^{\pi} \cos^2(x)\ dx = \int_{-\pi}^{\pi} \sin^2(x)\ dx = \pi,$$

the list

$$\left(\frac{1}{\sqrt{2\pi}}, \frac{\cos(\cdot)}{\sqrt{\pi}}, \frac{\sin(\cdot)}{\sqrt{\pi}}\right)$$

is orthonormal. Then

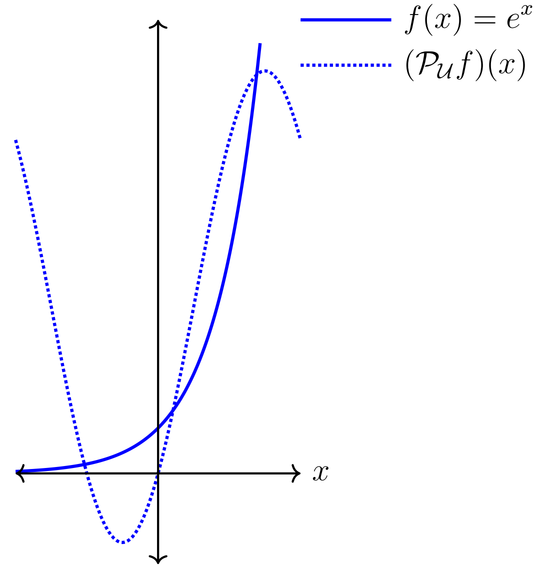
$$\begin{aligned}(\mathcal{P}_{\mathcal{U}}f)(x) &= \left\langle f, \frac{1}{\sqrt{2\pi}}\right\rangle \frac{1}{\sqrt{2\pi}} + \left\langle f, \frac{\cos(\cdot)}{\sqrt{\pi}}\right\rangle \frac{\cos(x)}{\sqrt{\pi}} + \left\langle f, \frac{f}{\sin(\cdot)}\sqrt{\pi}\right\rangle \frac{\sin(x)}{\sqrt{\pi}} \\ &= \frac{\langle f, 1\rangle}{2\pi} + \frac{\langle f, \cos(\cdot)\rangle}{\pi}\cos(x) + \frac{\langle f, \sin(\cdot)\rangle}{\pi}\sin(x)\end{aligned}$$

is the best approximation to any $f \in \mathcal{V}$ by a function in $\mathcal{U}$. Actually finding a formula for $\mathcal{P}_{\mathcal{U}}f$ then just boils down to computing some integrals.

(ii) In the particular case of $f(x) = e^x$, we have

$$(\mathcal{P}_{\mathcal{U}}f)(x) = \left(\frac{e^{\pi} - e^{-\pi}}{2\pi}\right)(1 - \cos(x) + \sin(x))$$

This is a low-order partial sum of the Fourier series for the exponential.



32.4 Remark. *Consider the situation of the previous example. The result is that* $\|f - \mathcal{P}_{\mathcal{U}} f\| \le \|f - g\|$ *for any* $g \in \mathcal{U}$. *This does not say that pointwise* $|f(x) - (\mathcal{P}_{\mathcal{U}} f)(x)| \le |f(x) - g(x)|$ *for all* $g \in \mathcal{U}$ *and* $x \in [-\pi, \pi]$. *Rather, this is a best approximation in a certain "average" sense over the interval* $[-\pi, \pi]$. *Remember that integrals measure average value of functions.*

This leads us to an approximation strategy for the fundamental problem $\mathcal{T} v = w$. Let $\mathcal{V}$ be a vector space, $\mathcal{W}$ be an inner product space, and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. If $w \in \mathcal{W} \setminus \mathcal{T}(\mathcal{V})$, then we cannot solve the problem $\mathcal{T} v = w$. However, suppose that $\mathcal{W} = \mathcal{T}(\mathcal{V}) \oplus \mathcal{T}(\mathcal{V})^{\perp}$. This is certainly possible when $\mathcal{T}(\mathcal{V})$ is finite-dimensional, i.e., for finite-rank $\mathcal{T}$.

Then $\widehat{w} := \mathcal{P}_{\mathcal{T}(\mathcal{V})} w$ satisfies

$$\|w - \widehat{w}\| \le \|w - \mathcal{T} v\|$$

for all $v \in \mathcal{T} v$. We therefore view $\mathcal{T} v = \widehat{w}$ as the best possible approximation to our unsolvable problem $\mathcal{T} v = w$. And we can certainly solve $\mathcal{T} v = \widehat{w}$ with $v = \widehat{v} \in \mathcal{V}$ for some $\widehat{v} \in \mathcal{V}$; we call $\widehat{v}$ a **LEAST SQUARES SOLUTION** to $\mathcal{T} v = w$. If $\mathcal{T}$ is not injective, then $\widehat{v}$ will not be unique, and an interesting question is how to choose (or even define) an *optimal* solution to $\mathcal{T} v = \widehat{w}$. One meaningful option is to ask for the **MINIMUM NORM LEAST SQUARES SOLUTION**: choose, if possible, $\widehat{v}_0 \in \mathcal{V}$ such that both $\mathcal{T} \widehat{v}_0 = \widehat{w}$ and $\|\widehat{v}\| \le \|v\|$ whenever $\mathcal{T} v = \widehat{w}$.

We will address these questions with some new technology later. In particular, we will return to the special case of $\mathcal{T} = \mathcal{M}_A$ when $A \in \mathbb{F}^{m \times n}$, and for the case $\operatorname{rank}(A) = n$ we will construct a transparent formula for $\mathcal{P}_{\mathcal{M}_A(\mathbb{F}^n)}$.

Part of this new technology will be the notion of adjoint, which we now take up rigorously, but which we met in our very first example.

32.5 Example. Let
$$A = \begin{bmatrix} 1 & 2 & 1 & 7 \\ 2 & 4 & 2 & 14 \\ 0 & 0 & 2 & 8 \end{bmatrix}.$$
Recall from Example 1.1 that $\mathbf{b} = (b_1, b_2, b_3) \in \mathbf{C}(A)$ if and only if $b_2 = 2b_1$, i.e., if and only if $b_2 - 2b_1 = 0$. We can interpret this "solvability condition" for the problem $A\mathbf{x} = \mathbf{b}$ in the language of inner products and orthogonality by recognizing
$$b_2 - 2b_1 = (-2)b_1 + (1 \cdot b_2) + (0 \cdot b_3) = \mathbf{b} \cdot \mathbf{z}, \qquad \mathbf{z} := \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$
Thus
$$\mathbf{C}(A) = \operatorname{span}(\mathbf{z})^\perp.$$

Can we "see" the vector $\mathbf{z} = (-2, 1, 0)$ somewhere in A? Yes: the second row of A is twice the first. While we perform elementary row operations on (obviously) the rows of a matrix in Gaussian elimination, proofs about matrices are usually more transparent in terms of columns. So, we flip rows to columns (and columns to rows) by taking the **TRANSPOSE**:
$$A^\mathsf{T} := \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 1 & 2 & 2 \\ 7 & 14 & 8 \end{bmatrix}.$$
It is then straightforward to calculate that $\mathbf{N}(A^\mathsf{T}) = \operatorname{span}(\mathbf{z})$, and so
$$\mathbf{C}(A) = \mathbf{N}(A^\mathsf{T})^\perp.$$

This "geometric" characterization of $\mathbf{C}(A)$ offers us a means of testing whether or not $\mathbf{b} \in \mathbb{R}^3$ is in $\mathbf{C}(A)$ *without* actually trying to solve $A\mathbf{x} = \mathbf{b}$: check if $\mathbf{b} \cdot \mathbf{z} = 0$ or not. And $\mathbf{z}$ arose by solving the "simpler" problem $A^\mathsf{T}\mathbf{z} = \mathbf{0}_4$.

32.6 Definition. *Let $A \in \mathbb{F}^{m \times n}$. The* ***TRANSPOSE*** *of A is the matrix $A^\mathsf{T} \in \mathbb{F}^{n \times m}$ whose (i, j)-entry is the (j, i)-entry of A. That is,*
$$A^\mathsf{T}_{ij} = A_{ji}.$$

This is a "static" definition of the transpose that arises from thinking about a matrix as a rectangular array of data (which is a perfectly valid way of thinking). However, it is often at least as fruitful, if not more fruitful, to think "dynamically"—what things do defines what things are. What does the transpose *do*, relative to the dot product?

Let $A \in \mathbb{F}^{m \times n}$ and let $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and $(\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_m)$ be the standard basis vectors for $\mathbb{F}^n$ and $\mathbb{F}^m$. Then
$$A^\mathsf{T}_{ij} = A^\mathsf{T}\widetilde{\mathbf{e}}_j \cdot \mathbf{e}_i$$

but also

$$A_{ij}^{\mathsf{T}} = A_{ji} = A\mathbf{e}_i \cdot \widetilde{\mathbf{e}}_j.$$

Thus

$$A\mathbf{e}_i \cdot \widetilde{\mathbf{e}}_j = A^{\mathsf{T}}\widetilde{\mathbf{e}}_j \cdot \mathbf{e}_i = \overline{\mathbf{e}_i \cdot A^{\mathsf{T}}\widetilde{\mathbf{e}}_j}.$$

So, at least when the vectors involved are standard basis vectors, the transpose almost "pops" across the dot product. (If $\mathbb{F} = \mathbb{R}$, there is no "almost" here.)

32.7 Definition. *Let $A \in \mathbb{F}^{m \times n}$. The* **CONJUGATE TRANSPOSE** *of A is the matrix $A^* \in \mathbb{F}^{n \times m}$ whose (i, j)-entry is the conjugate of (j, i)-entry of A, equivalently, the conjugate of the (i, j)-entry of A^{T}. That is,*

$$A_{ij}^* = \overline{A_{ji}} = \overline{A_{ij}^{\mathsf{T}}}.$$

We have

$$A_{ij}^* = A^*\widetilde{\mathbf{e}}_j \cdot \mathbf{e}_i$$

but also

$$A_{ij}^* = \overline{A_{ji}} = \overline{A\mathbf{e}_i \cdot \widetilde{\mathbf{e}}_j}.$$

Then

$$A\mathbf{e}_i \cdot \widetilde{\mathbf{e}}_j = \overline{A_{ij}^*} = \overline{A^*\widetilde{\mathbf{e}}_j \cdot \mathbf{e}_i} = \mathbf{e}_i \cdot A^*\widetilde{\mathbf{e}}_j.$$

So, at least when the vectors involved are standard basis vectors, the conjugate transpose (really) "pops" across the dot product.

32.8 Problem ($\star$). Let $A \in \mathbb{F}^{m \times n}$, $\mathbf{v} \in \mathbb{F}^n$, and $\mathbf{w} \in \mathbb{F}^m$. Prove that

$$A\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot A^*\mathbf{w}.$$

[Hint: *prove it first for $\mathbf{w} = \widetilde{\mathbf{e}}_j$ by expanding $\mathbf{v} \in \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and using the linearity of the dot product. Then prove it for an arbitrary $\mathbf{w} \in \operatorname{span}(\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_m)$, again by expanding $\mathbf{w}$ in this span and using linearity.*]

Content from *Linear Algebra* by Meckes & Meckes. Pages 96–97 discuss matrix transposes and conjugate transposes. We know $\mathcal{M}_A^* = \mathcal{M}_{A^*}$, but how is $\mathcal{M}_{A^{\mathsf{T}}}$ related to $\mathcal{M}_A$ when we are working with complex matrices and vectors? This is subtle, and it involves linear functionals. . .

Day 33: Monday, April 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."

Adjoint of a linear operator

We now abstract the "popping" behavior of the conjugate transpose. The proof of the following theorem is a great exercise in the slogan "What things do defines what things are."

33.1 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces with inner products $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{W}}$, respectively. Let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and suppose that $\mathcal{S} \in \mathcal{V}^{\mathcal{W}}$ with*

$$\langle \mathcal{T}v, w \rangle_{\mathcal{W}} = \langle v, \mathcal{S}(w) \rangle_{\mathcal{V}}, \ v \in \mathcal{V}, \ w \in \mathcal{W}. \tag{33.1}$$

Then $\mathcal{S}$ satisfies the following.

(i) $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$.

(ii) $\ker(\mathcal{S}) = \mathcal{T}(\mathcal{V})^{\perp}$ *and* $\mathcal{T}(\mathcal{V}) \subseteq \ker(\mathcal{S})^{\perp}$.

(iii) $\mathcal{S}$ *is the only operator in* $\mathbf{L}(\mathcal{V}, \mathcal{W})$ *to satisfy* (33.1).

Proof. **(i)** We want to show that $\mathcal{S}(w_1 + w_2) = \mathcal{S}(w_1) + \mathcal{S}(w_2)$ for all w_1, $w_2 \in \mathcal{W}$. We know

$$\begin{aligned} \langle \mathcal{T}v, w_1 \rangle_{\mathcal{W}} &= \langle v, \mathcal{S}(w_2) \rangle_{\mathcal{V}} \\ \langle \mathcal{T}v, w_2 \rangle_{\mathcal{W}} &= \langle v, \mathcal{S}(w_2) \rangle_{\mathcal{V}} \\ \langle \mathcal{T}v, w_1 + w_2 \rangle_{\mathcal{W}} &= \langle v, \mathcal{S}(w_1 + w_2) \rangle_{\mathcal{V}} \end{aligned}$$

for all $v \in \mathcal{V}$. We put this together to show that

$$\begin{aligned} \langle v, \mathcal{S}(w_1 + w_2) \rangle_{\mathcal{V}} &= \langle \mathcal{T}v, w_1 + w_2 \rangle_{\mathcal{W}} \\ &= \langle \mathcal{T}v, w_1 \rangle_{\mathcal{W}} + \langle \mathcal{T}v, w_2 \rangle_{\mathcal{W}} \\ &= \langle v, \mathcal{S}(w_1) \rangle_{\mathcal{V}} + \langle v, \mathcal{S}(w_2) \rangle_{\mathcal{V}} \\ &= \langle v, \mathcal{S}(w_1) + \mathcal{S}(w_2) \rangle_{\mathcal{V}}. \end{aligned}$$

Since

$$\langle v, \mathcal{S}(w_1 + w_2) \rangle_{\mathcal{V}} = \langle v, \mathcal{S}(w_1) + \mathcal{S}(w_2) \rangle_{\mathcal{V}}$$

for all $v \in \mathcal{V}$, we have

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

A similar argument shows $\mathcal{S}(\alpha w) = \alpha \mathcal{S}(w)$ for all $\alpha \in \mathbb{F}$ and $w \in \mathcal{W}$.

(ii) First we show that $\mathcal{T}(\mathcal{V}) \subseteq \ker(\mathcal{S})^\perp$. Let $w \in \mathcal{T}(\mathcal{V})$ and $z \in \ker(\mathcal{S})$. Take $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. Then

$$\langle w, z \rangle_{\mathcal{W}} = \langle \mathcal{T}v, z \rangle_{\mathcal{W}} = \langle v, \mathcal{S}z \rangle_{\mathcal{V}} = \langle v, 0_{\mathcal{V}} \rangle_{\mathcal{V}} = 0. \tag{33.2}$$

The first equality is the definition $w = \mathcal{T}v$, the second is the fundamental identity (33.1), the third is $z \in \ker(\mathcal{S})$, and the fourth is a property of inner products. We have shown that if $w \in \mathcal{T}(\mathcal{V})$, then $\langle w, z \rangle_{\mathcal{W}} = 0$ for all $z \in \ker(\mathcal{S})$, and so $w \in \ker(\mathcal{S})^\perp$. Thus $\mathcal{T}(\mathcal{V}) \subseteq \ker(\mathcal{S})^\perp$.

Now we prove that $\ker(\mathcal{S}) = \mathcal{T}(\mathcal{V})^\perp$. The calculation (33.2) also shows that if $z \in \ker(\mathcal{S})$, then $\langle w, z \rangle_{\mathcal{W}} = 0$ for all $w \in \mathcal{T}(\mathcal{V})$. Hence $z \in \mathcal{T}(\mathcal{V})^\perp$, and so $\ker(\mathcal{S}) \subseteq \mathcal{T}(\mathcal{V})^\perp$.

Last, suppose $u \in \mathcal{T}(\mathcal{V})^\perp$. Then $\langle w, u \rangle_{\mathcal{W}} = 0$ for all $w \in \mathcal{T}(\mathcal{V})$, so $\langle \mathcal{T}v, u \rangle_{\mathcal{W}} = 0$ for all $v \in \mathcal{V}$. The fundamental identity (33.1) gives

$$0 = \langle \mathcal{T}v, u \rangle_{\mathcal{W}} = \langle v, \mathcal{S}u \rangle_{\mathcal{V}}.$$

Since this is true for all $v \in \mathcal{V}$, Theorem 27.10 implies that $\mathcal{S}u = 0_{\mathcal{V}}$, thus $u \in \ker(\mathcal{S})$. We have shown that if $u \in \mathcal{T}(\mathcal{V})^\perp$, then $u \in \ker(\mathcal{S})$, and so $\mathcal{T}(\mathcal{V})^\perp \subseteq \ker(\mathcal{S})$.

(iii) Suppose that $\widetilde{\mathcal{S}} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ also satisfies

$$\langle \mathcal{T}v, w \rangle_{\mathcal{W}} = \left\langle v, \widetilde{\mathcal{S}}w \right\rangle_{\mathcal{V}}$$

for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$. We want to show $\mathcal{S}w = \widetilde{\mathcal{S}}w$ for all $w \in \mathcal{W}$, equivalently, $(\mathcal{S} - \widetilde{\mathcal{S}})w$, and this happens if

$$\left\langle v, (\mathcal{S} - \widetilde{\mathcal{S}})w \right\rangle_{\mathcal{V}} = 0$$

for all $v \in \mathcal{V}$. We compute

$$\left\langle v, (\mathcal{S} - \widetilde{\mathcal{S}})w \right\rangle_{\mathcal{V}} = \langle v, \mathcal{S}w \rangle_{\mathcal{V}} - \left\langle v, \widetilde{\mathcal{S}}w \right\rangle_{\mathcal{V}} = \langle \mathcal{T}v, w \rangle_{\mathcal{W}} - \langle \mathcal{T}v, w \rangle_{\mathcal{W}} = 0.$$ ■

33.2 Problem (!). Do that similar argument to show that $\mathcal{S}(\alpha w) = \alpha \mathcal{S}(w)$, as claimed in the proof of part (i) above.

Content from *Linear Algebra* by Meckes & Meckes. The uniqueness argument is Lemma 5.11 on p. 312.

33.3 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. The* ***ADJOINT*** *of $\mathcal{T}$ is the unique operator $\mathcal{T}^* \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that*

$$\langle \mathcal{T}v, w \rangle_{\mathcal{W}} = \langle v, \mathcal{T}^* w \rangle_{\mathcal{V}}, \ v \in \mathcal{V}, \ w \in \mathcal{W},$$

if such an operator $\mathcal{T}^*$ exists.

Content from *Linear Algebra* by Meckes & Meckes. The adjoint is defined on p. 311. Note the book's use of "an" adjoint at this point, not "the" adjoint.

Here is the fundamental utility of adjoints toward the operator equation $\mathcal{T}v = w$. Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has the adjoint $\mathcal{T}^*$. Part (ii) of Theorem 33.1 shows that

$$\mathcal{T}(\mathcal{V})^{\perp} = \ker(\mathcal{T}^*).$$

If we know that $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$, then we obtain the superb "range equals kernel perp" characterization of the range: $\mathcal{T}(\mathcal{V}) = \ker(\mathcal{T}^*)^{\perp}$. That is, given $w \in \mathcal{W}$, we can solve $\mathcal{T}v = w$ precisely when $\langle w, z \rangle_{\mathcal{W}} = 0$ for all $z \in \mathcal{W}$ such that $\mathcal{T}^* z = 0_{\mathcal{V}}$. In particular, if $\langle w, z \rangle_{\mathcal{W}} \neq 0$ for just one $z \in \mathcal{W}$ with $\mathcal{T}^* z = 0_{\mathcal{V}}$, then we cannot solve $\mathcal{T}v = w$. This geometric characterization of the range changes the dynamics of what could be a very challenging problem (solve $\mathcal{T}v = w$ or prove that no solution exists) to a possibly easier problem (solve the more specific problem $\mathcal{T}^* z = 0_{\mathcal{V}}$ and then check orthogonality to solutions). Depending on the nature of $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{T}$, it is possible to generate sufficient conditions for (1) $\mathcal{T}$ to have an adjoint and (2) $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$ such that these conditions are (3) natural and arise often in practice and (4) are (relatively) easily checked. In particular, if $\mathcal{T}$ is finite-rank, then $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$.

33.4 Problem (!). Why?

33.5 Example. Our motivating example for adjoints already gave us one. Let $A \in \mathbb{F}^{m \times n}$. Then the adjoint of the multiplication operator $\mathcal{M}_A$ is the multiplication operator $\mathcal{M}_{A^*}$, for given $\mathbf{v} \in \mathbb{F}^n$ and $\mathbf{w} \in \mathbb{F}^m$, we have

$$\mathcal{M}_A \mathbf{v} \cdot \mathbf{w} = A\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot A^* \mathbf{w} = \mathbf{v} \cdot \mathcal{M}_{A^*} \mathbf{w}.$$

Be careful in that the first two dot products are computed in $\mathbb{F}^m$, while the second two are in $\mathbb{F}^n$.

Finite-dimensionality guarantees the existence of many nice things in linear algebra, so it should be no surprise to find that adjoints exist for operators with finite-dimensional domains.

33.6 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces with $\mathcal{V}$ finite-dimensional. Then any operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has an adjoint. Specifically, if $(u_1, \ldots, u_n)$ is an orthonormal basis for $\mathcal{V}$, then*

$$\mathcal{T}^* w = \sum_{j=1}^{n} \langle w, \mathcal{T} u_j \rangle_{\mathcal{W}} u_j, \ \ w \in \mathcal{W}.$$

Proof. Finite-dimensionality means that we can use a basis, and since we are working with inner product spaces, we can use the best basis. Let $(u_1, \ldots, u_n)$ be an orthonormal basis for $\mathcal{V}$. We want to find an operator $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that $\langle \mathcal{T}v, w \rangle_{\mathcal{W}} = \langle v, \mathcal{S}w \rangle_{\mathcal{V}}$ for all $v \in \mathcal{V}$

and $w \in \mathcal{W}$. How should we define $\mathcal{S}w$? We manipulate

$$\langle \mathcal{T}v, w\rangle_{\mathcal{W}} = \left\langle \mathcal{T}\sum_{j=1}^{n} \langle v, u_j\rangle_{\mathcal{V}} u_j, w\right\rangle = \left\langle \sum_{j=1}^{n} \langle v, u_j\rangle_{\mathcal{V}} \mathcal{T}u_j, w\right\rangle_{\mathcal{W}} = \sum_{j=1}^{n} \langle v, u_j\rangle_{\mathcal{V}} \langle \mathcal{T}u_j, w\rangle_{\mathcal{W}}.$$

This is the linearity of the inner product in its first slot (with respect to both sums and scalar multiples). Now look at the terms of this sum:

$$\langle v, u_j\rangle_{\mathcal{V}} \langle \mathcal{T}u_j, w\rangle_{\mathcal{W}} = \left\langle v, \overline{\langle \mathcal{T}u_j, w\rangle_{\mathcal{W}}} u_j\right\rangle_{\mathcal{V}} = \left\langle v, \langle w, \mathcal{T}u_j\rangle_{\mathcal{W}} u_j\right\rangle_{\mathcal{V}}.$$

This is the conjugate linearity of the inner product in its second slot (with respect to scalar multiples). Then

$$\langle \mathcal{T}v, w\rangle_{\mathcal{W}} = \sum_{j=1}^{n} \left\langle v, \langle w, \mathcal{T}u_j\rangle_{\mathcal{W}} u_j\right\rangle_{\mathcal{V}} = \left\langle v, \sum_{j=1}^{n} \langle w, \mathcal{T}u_j\rangle_{\mathcal{W}} u_j\right\rangle_{\mathcal{V}}.$$

This is the linearity of the inner product in its second slot (with respect to sums). ■

Content from *Linear Algebra* by Meckes & Meckes. This result is Theorem 5.12 on p. 312.

33.7 Example. We are using the same notation for the adjoint of an operator and the conjugate transpose of a matrix. Of course, there is no conflict.

Let $\mathcal{T} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$. Then

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

so

$$\mathcal{T}^* = \mathcal{M}_{[\mathcal{T}]}^* = \mathcal{M}_{[\mathcal{T}]^*}.$$

But also

$$\mathcal{T}^* = \mathcal{M}_{[\mathcal{T}^*]}.$$

That is, $\mathcal{M}_{[\mathcal{T}]^*} = \mathcal{M}_{[\mathcal{T}^*]}$, and so $[\mathcal{T}^*] = [\mathcal{T}]^*$.

33.8 Problem ($\star$). Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces. Suppose that $\mathcal{V}$ is finite-dimensional with orthonormal bases $(u_1, \ldots, u_n)$ and $(\widetilde{u}_1, \ldots, \widetilde{u}_n)$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$.

(i) Use Theorem 33.6 to obtain the (possibly surprising) result that

$$\sum_{j=1}^{n} \langle w, \mathcal{T}u_j\rangle_{\mathcal{W}} u_j = \sum_{j=1}^{n} \langle t, \mathcal{T}\widetilde{u}_j\rangle_{\mathcal{W}} \widetilde{u}_j \tag{33.3}$$

for every $w \in \mathcal{W}$.

(ii) Show that (33.3) is true without invoking Theorem 33.6. [Hint: *do some intermediate* *calculations to show*

$$\left\langle \sum_{j=1}^{n} \langle w, \mathcal{T} u_j \rangle_{\mathcal{W}} u_j, \widetilde{u}_k \right\rangle_{\mathcal{V}} = \overline{\sum_{j=1}^{n} \langle \mathcal{T}(\langle \widetilde{u}_k, u_j \rangle_{\mathcal{V}} u_j, w \rangle_{\mathcal{W}}} = \langle w, \mathcal{T} \widetilde{u}_j \rangle_{\mathcal{W}}.$$

Then use Theorem 28.3.]

Day 34: Wednesday, April 8.

Vocabulary from today

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

Self-adjoint operator/matrix (N), symmetric matrix, skew-adjoint operator (N)

34.1 Example. Let $\mathcal{V} = \mathcal{C}([0,1])$ with the L^2-inner product, and let $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ be the multiplication operator $(\mathcal{T} f)(x) = x f(x)$. To determine if $\mathcal{T}$ has an adjoint, we compute

$$\langle \mathcal{T} f, g \rangle = \int_0^1 (\mathcal{T} f)(x) g(x) \, dx = \int_0^1 x f(x) g(x) \, dx = \int_0^1 f(x)(x g(x)) \, dx$$
$$= \int_0^1 f(x)(\mathcal{T} g)(x) \, dx = \langle f, \mathcal{T} g \rangle.$$

So, $\mathcal{T}$ does have an adjoint, and here $\mathcal{T}^* = \mathcal{T}$.

Operators that are their own adjoint are particularly nice (in the sense that they have many clear and useful properties) and deserve a special name.

34.2 Definition. **(i)** *Let $\mathcal{V}$ be an inner product space. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* ***SELF-ADJOINT*** *if $\mathcal{T}^* = \mathcal{T}$.*

(ii) *A matrix $A \in \mathbb{F}^{n \times n}$ is* ***SELF-ADJOINT*** *or* ***HERMITIAN*** *if $A^* = A$.*

(iii) *A matrix $A \in \mathbb{F}^{n \times n}$ is* ***SYMMETRIC*** *if $A^{\mathsf{T}} = A$.*

34.3 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces with $\mathcal{V} \neq \mathcal{W}$. Why does it not really make sense to talk about a self-adjoint operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$?

Our next examples require a slight variation on the L^2-inner product.

34.4 Lemma. *Let $p > 0$. The map $\langle f, g\rangle := \int_0^p f(x)g(x)\ dx$ is an inner product on the space*

$$\{f \in \mathcal{C}(\mathbb{R}) \mid f(x+p) = f(x) \text{ for all } x \in \mathbb{R}\}.$$

Proof. As is often the case, the only property that really needs checking is definiteness; all of the other properties follow from ordinary properties of integrals. So, suppose that $\langle f, f\rangle = 0$, which means $\int_0^p |f(x)|^2\ dx = 0$. Per Example 27.2, we have $f(x) = 0$ for all $0 \le x \le p$. Since f is p-periodic, for any $y \in \mathbb{R}$, there are $x \in [0, p]$ and $k \in \mathbb{Z}$ such that $f(y) = f(x + pk) = f(x) = 0$, thus $f = 0$. ∎

34.5 Problem (!). Why is $\langle f, g\rangle := \int_0^1 f(x)g(x)\ dx$ not an inner product on $\mathcal{C}^\infty(\mathbb{R})$? [Hint: *it is a fact, easily suggested by a picture and somewhat ticklish to prove, that for any interval $[a, b]$, there is an infinitely differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ such that $f(x) = 0$ for $a \le x \le b$ with $f(x) \ne 0$ for some $x \in \mathbb{R} \setminus [a, b]$.*]

34.6 Example. Let

$$\mathcal{V} = \mathcal{W} = \{f \in \mathcal{C}^\infty(\mathbb{R}) \mid f(x+1) = f(x) \text{ for all } x \in \mathbb{R}\}$$

and define

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

To determine if $\mathcal{T}$ has an adjoint, we compute

$$\begin{aligned}
\langle \mathcal{T}f, g\rangle &= \int_0^1 (\mathcal{T}f)(x)g(x)\ dx \\
&= \int_0^1 f'(x)g(x)\ dx \\
&= f(1)g(1) - f(0)g(0) - \int_0^1 f(x)g'(x)\ dx \\
&= -\int_0^1 f(x)g'(x)\ dx.
\end{aligned}$$

The third equality is integration by parts (with $u = g(x)$ and $dv = f'(x)\ dx$), and the fourth is the 1-periodicity of f and g. Thus

$$\langle \mathcal{T}f, g\rangle = \int_0^1 f(x)(-g'(x))\ dx,$$

and so the adjoint of $\mathcal{T}$ is

$$(\mathcal{T}^* g)(x) := -g(x).$$

That is, $\mathcal{T}^* = -\mathcal{T}$.

Operators $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ with $\mathcal{T}^* = -\mathcal{T}$ also have a special name.

34.7 Definition. *Let $\mathcal{V}$ be an inner product space. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* ***SKEW-ADJOINT*** *if $\mathcal{T}^* = -\mathcal{T}$.*

34.8 Example. Let

$$\mathcal{V} = \{f \in \mathcal{C}^{\infty}(\mathbb{R}) \mid f(x+2\pi) = f(x) \text{ for all } x \in \mathbb{R}\},$$

with inner product $\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\ dx$. Let $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ be the shift operator $(\mathcal{T}f)(x) := f(x+1)$. To find the adjoint of $\mathcal{T}$, we manipulate

$$\langle \mathcal{T}f, g\rangle = \int_{-\pi}^{\pi} (\mathcal{T}f)(x)g(x)\ dx = \int_{-\pi}^{\pi} f(x+1)g(x)\ dx.$$

We would like to turn this into an integral involving only a factor of $f(x)$ and "something" involving g, and the way to do that is to "remove" the $x+1$ by substituting $s = x+1$, $ds = dx$, $s(-\pi) = -\pi+1$, and $s(\pi) = \pi+1$ to find

$$\int_0^{2\pi} f(x+1)g(x)\ dx = \int_1^{1+2\pi} f(s)g(s-1)\ ds.$$

The problem now is that this integral is not over $[0, 2\pi]$, and so it is not really the original inner product.

But it is: if $h \in \mathcal{C}(\mathbb{R})$ is 2π-periodic, then $H(x) := \int_x^{x+2\pi} h(s)\ ds$ is 2π-periodic as well. We have used this fact several times throughout the course (possibly with periods other than 2π). The easiest way to prove it is to compute $H' = 0$, thus H is constant, so $H(x) = H(0) = \int_0^{2\pi} h(s)\ ds$.

In the particular case above, $h(s) = f(s)g(s-1)$ and so

$$\int_1^{1+2\pi} f(s)g(s-1)\ ds = \int_0^{2\pi} f(s)g(s-1)\ ds.$$

Thus with $(\mathcal{T}^*g)(s) := g(s-1)$, we have $\langle \mathcal{T}f, g\rangle = \langle f, \mathcal{T}^*g\rangle$.

34.9 Problem ($\star$). Define $\mathcal{T} \colon \ell^2 \to \ell^2$ by $(\mathcal{T}f)(k) = f(k+1)$, so f is the "forward shift" operator. That is, $\mathcal{T}(a_1, a_2, a_3, \ldots) = (a_2, a_3, a_4, \ldots)$. Show that

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

We can think of $\mathcal{T}^*$ as the "backward shift" operator (with slight tweaking to account for the absence of anything before $g(1)$ to shift back to). [Hint: *if $(a_k) \in \mathbb{F}^{\infty}$, m, $n \geq 1$, and the series $\sum_{k=1}^{\infty} a_k$ converges, then $\sum_{k=m}^{\infty} a_k = \sum_{k=m+n}^{\infty} a_{k-n}$. Apply this property to the series in $\langle \mathcal{T}f, g\rangle = f(2)\overline{g(1)} + \sum_{k=2}^{\infty} f(k+1)\overline{g(k)}$.*]

34.10 Problem ($\star$). Let $\mathcal{V} = \mathcal{C}([0,1])$ with the L^2-inner product and put

$$\mathcal{R} := \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; 0 \le x \le 1,\ 0 \le y \le 1\right\}.$$

Recall from multivariable calculus that if $h \colon \mathcal{R} \to \mathbb{R}$ is continuous, then

$$\int_0^1 \int_0^x h(x,y)\ dy\ dx = \int_0^1 \int_y^1 h(x,y)\ dx\ dy.$$

Use this fact to find the adjoint of the antiderivative operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{V}$ given by $(\mathcal{T}f)(x) = \int_0^x f(y)\ dy$.

An operator is not, however, guaranteed to have an adjoint.

34.11 Problem ($\star$). Let

$$\mathcal{U} = \left\{f \in \ell^2 \;\middle|\; f(k) = 0 \text{ for all but finitely many } k\right\}.$$

We worked with this space in Example 30.8, and we consider $\mathcal{U}$ here as an inner product space with the ℓ^2-inner product. Put $h(k) := 1/k$, so $h \in \ell^2 \setminus \mathcal{U}$. Define

$$\mathcal{T} \colon \mathcal{U} \to \ell^2 \colon f \mapsto \langle f, h \rangle\, h.$$

Show that $\mathcal{T}$ has no adjoint as follows. If there is $\mathcal{S} \in \mathbf{L}(\ell^2, \mathcal{U})$ such that $\langle \mathcal{T}f, g \rangle = \langle f, \mathcal{S}g \rangle$ for all $f \in \mathcal{U}$ and $g \in \ell^2$, take $f = e_j$ as defined in (30.2) and $g = h$ to compute $(\mathcal{S}h)(j)$. Since $\mathcal{S}h \in \mathcal{U}$, what contradiction results?

The situation in the previous problem that contributed to a lack of adjoint is somewhat baroque and contrived. The operators that one "naturally" encounters in practice tend to have adjoints, as their domains are suitably chosen to encode some symmetry properties that permit an adjoint's existence. Here is an abstraction of the situation of Problem 34.11 that further emphasizes its esoteric nature.

34.12 Problem ($\star$). Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}$ be a proper subspace of $\mathcal{V}$ (so $\mathcal{U} \ne \mathcal{V}$). Fix $v_0 \in \mathcal{V} \setminus \mathcal{U}$ and define

$$\mathcal{T} \colon \mathcal{U} \to \mathcal{V} \colon u \mapsto \langle u, v_0 \rangle\, v_0.$$

In particular, since $v_0 \in \mathcal{V} \setminus \mathcal{U}$, we have $v_0 \ne 0_{\mathcal{V}}$. We consider $\mathcal{U}$ as an inner product space with the same inner product as $\mathcal{V}$.

(i) Define

$$\mathcal{S} \colon \mathcal{V} \to \mathcal{V} \colon v \mapsto \langle v, v_0 \rangle\, v_0.$$

Then $\mathcal{T}u = \mathcal{S}u$ for all $u \in \mathcal{U}$, but $\mathcal{S}$ and $\mathcal{T}$ have different domains, and so they are not the same operator—rather, $\mathcal{S}$ is an extension of $\mathcal{T}$ to all of $\mathcal{V}$. Show that $\langle \mathcal{T}u, v \rangle = \langle u, \mathcal{S}v \rangle$ for all $u \in \mathcal{U}$ and $v \in \mathcal{V}$. Explain why $\mathcal{S}$ cannot be the adjoint of $\mathcal{T}$.

(ii) Suppose that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$. Prove that $\mathcal{T}$ has the adjoint given by $\mathcal{T}^* v = \langle v, v_0 \rangle \mathcal{P}_\mathcal{U} v_0$. [Hint: $\mathcal{P}_\mathcal{U}^* = \mathcal{P}_\mathcal{U}$ *and* $v_0 = \mathcal{P}_\mathcal{U} v_0 + u^\perp$ *for some* $u^\perp \in \mathcal{U}^\perp$.]

(iii) With $\mathcal{U}$ from Problem 34.11, check that $\ell^2 \neq \mathcal{U} \oplus \mathcal{U}^\perp$.

Day 35: Friday, April 10.

You took Exam 2.

Day 36: Monday, April 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."

Four fundamental subspaces associated with a linear operator, matrix representation of a linear operator (not just an operator between Euclidean spaces)

Now that we have more concrete experience with the adjoint, we return to viewing it in the abstract. It turns out that considering the adjoint as an operator-theoretic version of the complex conjugate can be very fruitful. Recall that if $z = x + iy \in \mathbb{C}$, then $\overline{z} = \overline{x + iy} := x - iy$. Then for z, $w \in \mathbb{C}$, we have identities like

$$\overline{z + w} = \overline{z} + \overline{w}, \qquad \overline{zw} = \overline{z}(\overline{w}), \qquad \overline{\overline{z}} = z, \qquad \text{and} \qquad \overline{z^{-1}} = (\overline{z})^{-1} \text{ if } z \neq 0.$$

The adjoint acts much the same.

36.1 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces.*

(i) *Suppose that $\mathcal{T}_1$, $\mathcal{T}_2 \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ have adjoints. Then $\mathcal{T}_1 + \mathcal{T}_2$ has the adjoint*

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

(ii) *Suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has an adjoint and $\alpha \in \mathbb{F}$. Then $\alpha\mathcal{T}$ has the adjoint*

$$(\alpha\mathcal{T})^* = \overline{\alpha}\mathcal{T}^*.$$

(iii) *Suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has an adjoint. Then $\mathcal{T}^*$ itself has the adjoint*

$$(\mathcal{T}^*)^* = \mathcal{T}.$$

(iv) *Suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is invertible has the adjoint $\mathcal{T}^*$. Then $\mathcal{T}^*$ is invertible and $\mathcal{T}^{-1}$ has the adjoint*

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

(v) *Suppose that $\mathcal{U}$ is also an inner product space and now $\mathcal{T}_1 \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{T}_2 \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ have adjoints. Then $\mathcal{T}_2\mathcal{T}_1 \in \mathbf{L}(\mathcal{U}, \mathcal{W})$ has the adjoint*

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

Proof. We prove only the third part. If $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has the adjoint $\mathcal{T}^* \in \mathbf{L}(\mathcal{W}, \mathcal{V})$, then we expect that $(\mathcal{T}^*)^* \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and

$$\langle \mathcal{T}^* w, v \rangle_{\mathcal{V}} = \langle w, (\mathcal{T}^*)^* v \rangle_{\mathcal{W}}. \tag{36.1}$$

Does $\mathcal{T}$ do this? Certainly $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, and we compute

$$\langle \mathcal{T}^* w, v \rangle_{\mathcal{V}} = \overline{\langle v, \mathcal{T}^* w \rangle_{\mathcal{V}}} = \overline{\langle \mathcal{T} v, w \rangle_{\mathcal{W}}} = \langle w, \mathcal{T} v \rangle_{\mathcal{W}}.$$

That is, $\mathcal{T}$ fulfills exactly the role of $(\mathcal{T}^*)^*$ in (36.1), and so by uniqueness of the adjoint, $(\mathcal{T}^*)^* = \mathcal{T}$. *What things do defines what things are.* ■

Content from *Linear Algebra* by Meckes & Meckes. These properties appear in Proposition 5.14 on pp. 313–314. Do Quick Exercise #9 on p. 312.

36.2 Problem ($\star$). We might wonder for what values of $\alpha \in \mathbb{F}$ can an operator $\mathcal{T}$ satisfy $\mathcal{T}^* = \alpha \mathcal{T}$. Self- and skew-adjoint operators tell us that $\alpha = \pm 1$ is an option, and so is $\alpha = 0$. (Why?) The parallels of adjoints and complex conjugates perhaps suggest that such nonzero α must satisfy $|\alpha| = 1$, and this is true.

(i) Prove it by computing

$$\langle \mathcal{T} v, w \rangle = |\alpha|^2 \langle \mathcal{T} v, w \rangle$$

and then taking $w = \mathcal{T} v$. [Hint: *pop $\mathcal{T}$ across the inner product twice with the adjoint.*]

(ii) By considering the matrices

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & -1 \\ i & 0 \end{bmatrix},$$

show that $\alpha \neq 1$ is possible.

36.3 Problem ($\star$). The following result will be useful multiple times later and right now just offers good practice with adjoints and orthogonality. Let $\mathcal{V}$ be an inner product space and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ has an adjoint. If $\mathcal{U}$ is a subspace of $\mathcal{V}$ such that $\mathcal{T}(\mathcal{U}) \subseteq \mathcal{U}$, show that $\mathcal{T}^*(\mathcal{U}^\perp) \subseteq \mathcal{U}^\perp$.

Our initial interest in the adjoint of $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ was the "range = kernel$^\perp$" characterization $\mathcal{T}(\mathcal{V}) = \ker(\mathcal{T}^*)^\perp$, which occurs when $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$. This meaningful geometric characterization of the range motivates awareness of four essential subspaces associated with an operator that has an adjoint—of course these are the foundational subspaces of kernel and range, just done twice.

36.4 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has an adjoint $\mathcal{T}^* \in \mathbf{L}(\mathcal{W}, \mathcal{V})$. The four* **FUNDAMENTAL SUBSPACES** *associated with $\mathcal{T}$ are the subspaces $\ker(\mathcal{T})$ and $\mathcal{T}^*(\mathcal{W})$ of $\mathcal{V}$ and $\mathcal{T}(\mathcal{V})$ and $\ker(\mathcal{T}^*)$ of $\mathcal{W}$.*

If $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional, then $\mathcal{T}(\mathcal{V}) = \mathcal{T}(\mathcal{V})^{\perp\perp} = \ker(\mathcal{T}^*)^{\perp}$, and likewise $\mathcal{T}^*(\mathcal{W}) = \mathcal{T}^*(\mathcal{W})^{\perp\perp} = \ker(\mathcal{T})^{\perp}$. We therefore have orthogonal direct sum decompositions

$$\mathcal{V} = \ker(\mathcal{T}) \oplus \ker(\mathcal{T})^{\perp} = \ker(\mathcal{T}) \oplus \mathcal{T}^*(\mathcal{W})^{\perp}$$

and

$$\mathcal{W} = \mathcal{T}(\mathcal{V}) \oplus \mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V}) \oplus \ker(\mathcal{T}^*).$$

Content from *Linear Algebra* by Meckes & Meckes. This is Proposition 5.16 on p. 315.

36.5 Problem (!). Let $\mathcal{V}$ be a finite-dimensional vector space, $\mathcal{W}$ be any vector space, and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Use Problem 31.4 and the identity $\mathcal{T}^*(\mathcal{W}) = \ker(\mathcal{T}^*)^{\perp}$ to obtain another proof of rank–nullity. In particular, conclude that $\mathcal{T}^*$ is also finite-rank and $\operatorname{rank}(\mathcal{T}) = \operatorname{rank}(\mathcal{T}^*)$.

36.6 Problem (!). Find bases for the four fundamental subspaces of $\mathcal{M}_A$, where

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

Use these bases to describe how to write $\mathbb{F}^5 = \mathbf{N}(A) \oplus \mathbf{C}(A^*)$ and $\mathbb{F}^3 = \mathbf{C}(A) \oplus \mathbf{N}(A^*)$.

To represent vectors in $\mathcal{V}$ or $\mathcal{W}$ in terms of these orthogonal decompositions, we could find bases for the fundamental subspaces, apply Gram–Schmidt to get orthonormal bases, and then construct projection operators. In the special case that $\mathcal{V} = \mathbb{F}^n$, there is an easier way of obtaining the orthogonal projection onto a subspace. This approach involves adjoints, which is why we did not discuss it earlier.

Let $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ be a linearly independent list in $\mathbb{F}^m$ and let $\mathcal{U} = \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. Then $\mathcal{U} = \mathbf{C}(A)$, where $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{F}^{m \times n}$. We will know the orthogonal projection $\mathcal{P}_{\mathcal{U}} \in \mathbf{L}(\mathbb{F}^m)$ onto $\mathcal{U}$ if we know its matrix representation $[\mathcal{P}_{\mathcal{U}}] \in \mathbb{F}^{m \times n}$, and it would be natural if we could express $[\mathcal{P}_{\mathcal{U}}]$ in terms of A.

Let $\mathbf{v} \in \mathbb{F}^m$. We will find a tame formula for $\mathcal{P}_{\mathcal{U}}\mathbf{v}$. The right idea is to exploit orthogonality: $\mathbf{v} - \mathcal{P}_{\mathcal{U}}\mathbf{v} \in \mathcal{U}^{\perp}$, so

$$(\mathbf{v} - \mathcal{P}_{\mathcal{U}}\mathbf{v}) \cdot \mathbf{a}_j = 0, \ j = 1, \ldots, n.$$

We have $\mathcal{P}_{\mathcal{U}}\mathbf{v} \in \mathbf{C}(A)$, so $\mathcal{P}_{\mathcal{U}}\mathbf{v} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{F}^n$. If we can figure out $\mathbf{x}$ in terms of A and $\mathbf{v}$, then we will know $\mathcal{P}_{\mathcal{U}}\mathbf{v}$. The trick is to rewrite $\mathbf{a}_j = A\mathbf{e}_j$ with $\mathbf{e}_j$ as the jth standard basis vector in $\mathbb{F}^n$. Thus

$$0 = (\mathbf{v} - \mathcal{P}_{\mathcal{U}}\mathbf{v}) \cdot \mathbf{a}_j = (\mathbf{v} - A\mathbf{x}) \cdot A\mathbf{e}_j = A^*(\mathbf{v} - A\mathbf{x}) \cdot \mathbf{e}_j, \ j = 1, \ldots, n.$$

Since this is true for each of the standard basis vectors, we have $A^*(\mathbf{v} - A\mathbf{x}) = \mathbf{0}_n$, and so $A^*A\mathbf{x} = A^*\mathbf{v}$. If we were lucky enough to have A^*A invertible, then $\mathbf{x} = (A^*A)^{-1}A^*\mathbf{v}$, and so, recalling that $\mathbf{x}$ satisfies $\mathcal{P}_{\mathcal{U}}\mathbf{v} = A\mathbf{x}$,

$$\mathcal{P}_{\mathcal{U}}\mathbf{v} = A(A^*A)^{-1}A^*\mathbf{v}.$$

We are indeed so lucky.

36.7 Lemma. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has an adjoint.*

(i) $\ker(\mathcal{T}) = \ker(\mathcal{T}^*\mathcal{T})$.

(ii) *$\mathcal{T}$ is injective if and only if $\mathcal{T}^*\mathcal{T}$ is injective.*

Proof. The second part follows immediately from the first, which establishes that $\ker(\mathcal{T})$ is trivial if and only if $\ker(\mathcal{T}^*\mathcal{T})$ is trivial. For the first part, if $\mathcal{T}v = 0_{\mathcal{W}}$, then $\mathcal{T}^*\mathcal{T}v = 0_{\mathcal{V}}$, so $\ker(\mathcal{T}) \subseteq \ker(\mathcal{T}^*\mathcal{T})$.

Now suppose that $\mathcal{T}^*\mathcal{T}v = 0_{\mathcal{V}}$. We might try to relate this to inner products by considering

$$0 = \|\mathcal{T}^*\mathcal{T}v\|_{\mathcal{V}}^2 = \langle \mathcal{T}^*\mathcal{T}v, \mathcal{T}^*\mathcal{T}v \rangle_{\mathcal{V}} = \langle \mathcal{T}v, \mathcal{T}\mathcal{T}^*\mathcal{T}v \rangle_{\mathcal{W}},$$

but that looks too complicated. However we could also compute

$$\|\mathcal{T}v\|_{\mathcal{W}}^2 = \langle \mathcal{T}v, \mathcal{T}v \rangle_{\mathcal{W}} = \langle v, \mathcal{T}^*\mathcal{T}v \rangle_{\mathcal{V}} = \langle v, 0_{\mathcal{V}} \rangle_{\mathcal{V}} = 0.$$

That forces $\mathcal{T}v = 0_{\mathcal{V}}$ and thus $v = 0_{\mathcal{V}}$, as desired. ■

Taking $\mathcal{T} = \mathcal{M}_A$ above, we have proved the following result.

36.8 Theorem. *Suppose that the columns of $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{F}^{m \times n}$ are independent. Then $[\mathcal{P}_{\mathbf{C}(A)}] = A(A^*A)^{-1}A^*$.*

Content from *Linear Algebra* by Meckes & Meckes. This is Proposition 4.18 on p. 257. Read the least squares examples on pp. 259–260.

While we have developed many results for inner product spaces without a hypothesis of finite-dimensionality, our best results do come from spans of orthonormal lists. Consider the following summary of uses of inner product spaces.

36.9 Remark. *Let $\mathcal{V}$ be an inner product space.*

(i) *Inner products compare vectors: $v = 0_{\mathcal{V}}$ if and only if $\langle v, w \rangle = 0$ for all $w \in \mathcal{V}$, and $v_1 = v_2$ if and only if $\langle v_1, w \rangle = \langle v_2, w \rangle$ for al $w \in \mathcal{V}$.*

(ii) *Inner products represent vectors: if $(u_1, \ldots, u_n)$ is an orthonormal list and $v \in \operatorname{span}(u_1, \ldots, u_n)$, then $v = \sum_{j=1}^n \langle v, u_j \rangle u_j$.*

(iii) *Inner products measure vectors: if $(u_1, \ldots, u_n)$ is an orthonormal list and $v \in$*

span$(u_1, \ldots, u_n)$, *then* $\|v\| = \left(\sum_{j=1}^n |\langle v, u_j\rangle|^2\right)^{1/2}$.

(iv) *Inner products identify how a subspace fits into the ambient space: if* $\mathcal{U}$ *is a subspace of* $\mathcal{V}$*, then* $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$ *with*

$$\|v\| = \left(\|\mathcal{P}_{\mathcal{U}}v\|^2 + \|v - \mathcal{P}_{\mathcal{U}}v\|^2\right)^{1/2} \quad \text{and} \quad \|v - \mathcal{P}_{\mathcal{U}}v\| = \min_{u \in \mathcal{U}} \|v - u\|$$

under suitable hypotheses on $\mathcal{U}$ *and* $\mathcal{V}$.

(v) *Inner products characterize the range of a linear operator: if* $\mathcal{W}$ *is another inner product space and* $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ *is a linear operator, then* $\mathcal{T}(\mathcal{V}) = \ker(\mathcal{T}^*)^\perp$ *under suitable hypotheses on* $\mathcal{V}$, $\mathcal{W}$, *and* $\mathcal{T}$.

Our fundamental goal has been to understand linear operators, and we have done so primarily through two major tools: finite-dimensionality and inner product spaces. Now we combine both of those tools to learn more about operators acting on *finite-dimensional inner product spaces*. We begin, however, by revisiting some aspects of operator theory on finite-dimensional vector spaces (no assumption of an inner product right now) to develop a deeper connection with matrices.

We know that any linear operator from $\mathbb{F}^n$ to $\mathbb{F}^m$ is just matrix multiplication. The same is essentially true for any operator between two finite-dimensional spaces.

Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces with bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$, respectively. Let $\mathcal{B}_{\mathcal{V}} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ and $\mathcal{B}_{\mathcal{W}} \in \mathbf{L}(\mathbb{F}^m, \mathcal{W})$ be the respective basis operators. That is, the maps

$$\mathcal{B}_{\mathcal{V}} \colon \mathbb{F}^n \to \mathcal{V} \colon (\alpha_1, \ldots, \alpha_n) \mapsto \sum_{j=1}^n \alpha_j v_j \quad \text{and} \quad \mathcal{B}_{\mathcal{W}} \colon \mathbb{F}^m \to \mathcal{W} \colon (\beta_1, \ldots, \beta_m) \mapsto \sum_{j=1}^m \beta_j w_j$$

are isomorphisms.

Let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. We multiply cleverly by the identity operators on $\mathcal{V}$ and $\mathcal{W}$ to get

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

Since $\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$, there is $A := [\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}] \in \mathbb{F}^{m \times n}$ such that $\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}} = \mathcal{M}_A$. Then

$$\mathcal{T} = \mathcal{B}_{\mathcal{W}}\mathcal{M}_A\mathcal{B}_{\mathcal{V}}^{-1}. \tag{36.2}$$

Up to "conjugation" by the basis operators, $\mathcal{T}$ is really a matrix multiplication operator!

36.10 Problem (!). Reread the (punishingly intricate) proof of Theorem 13.1. How does the identity (36.2) show up in that proof? [Hint: *in the notation of that theorem, use* $\mathcal{V}_1 = \mathcal{V}$, $\mathcal{V}_2 = \mathbb{F}^n$, $\mathcal{W}_1 = \mathcal{W}$, *and* $\mathcal{W}_2 = \mathbb{F}^m$.]

What exactly is $[\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}]$? Since $\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}} \in \mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$, of course

$$[\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}] = \begin{bmatrix} \mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}\mathbf{e}_1 & \cdots & \mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}\mathbf{e}_n \end{bmatrix},$$

but we can be more precise. Since

$$\mathcal{B}_{\mathcal{V}}(\alpha_1, \ldots, \alpha_n) = \sum_{j=1}^{n} \alpha_j v_j,$$

we have $\mathcal{B}_{\mathcal{V}}\mathbf{e}_j = v_j$, so

$$\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}\mathbf{e}_j = \mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}v_j = [\mathcal{T}v_j]_{\mathcal{B}_{\mathcal{W}}},$$

where we use the notation $[w]_{\mathcal{B}_{\mathcal{W}}} = \mathcal{B}_{\mathcal{W}}^{-1}w \in \mathbb{F}^m$ from Definition 14.4. In more words, the jth column of $[\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}]$ is the coordinates of $\mathcal{T}v_j$ with respect to the basis $(w_1, \ldots, w_m)$ of $\mathcal{W}$.

36.11 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces with bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$, respectively. Let $\mathcal{B}_{\mathcal{V}} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ and $\mathcal{B}_{\mathcal{W}} \in \mathbf{L}(\mathbb{F}^m, \mathcal{W})$ be the respective basis operators. The* ***MATRIX REPRESENTATION*** *of a linear operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ with respect to these bases is*

$$[\mathcal{T}]_{\mathcal{B}_{\mathcal{W}} \leftarrow \mathcal{B}_{\mathcal{V}}} := [\mathcal{B}_{\mathcal{W}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}] = \begin{bmatrix} [\mathcal{T}v_1]_{\mathcal{B}_{\mathcal{W}}} & \cdots & [\mathcal{T}v_n]_{\mathcal{B}_{\mathcal{W}}} \end{bmatrix} \in \mathbb{F}^{m \times n}.$$

If $\mathcal{V} = \mathcal{W}$, then $[\mathcal{T}]_{\mathcal{B}_{\mathcal{V}} \leftarrow \mathcal{B}_{\mathcal{V}}} := [\mathcal{B}_{\mathcal{V}}^{-1}\mathcal{T}\mathcal{B}_{\mathcal{V}}]$ is the matrix representation of $\mathcal{T}$ with respect to just the basis $(v_1, \ldots, v_n)$.

Content from *Linear Algebra* by Meckes & Meckes. The matrix representation of an operator is defined on p. 187 and uses coordinate notation from p. 185. Read Examples 1 and 2 on pp. 188–189.

Day 37: Wednesday, April 15.

Vocabulary from today

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

Diagonalizable operator (N), pair of similar operators

Computing the matrix representation of a linear operator with respect to given bases can be a fairly thankless task that involves a lot of manipulating linear systems of equations. We consider a hopefully straightforward example.

37.1 Example. **(i)** We study $\mathcal{T}\colon \mathbb{P}^2 \to \mathbb{P}^1\colon p \mapsto p'$. Let $\mathbb{P}^2$ have the basis (v_0, v_1, v_2) with $v_0(x) = 1$, $v_1(x) = x$, and $v_2(x) = x^2$. Let $\mathbb{P}^1$ have the basis (w_0, w_1) with $w_0(x) = 1$ and $w_1(x) = x$. Denote the associated basis operators by $\mathcal{B}_{\mathbb{P}^2}\colon \mathbb{F}^3 \to \mathbb{P}^2$ and $\mathcal{B}_{\mathbb{P}^1}\colon \mathbb{F}^2 \to \mathbb{P}^1$.

Then

$$(\mathcal{T}v_0)(x) = v_0'(x) = 0 = 0w_0(x) + 0w_1(x),$$

$$(\mathcal{T}v_1)(x) = v_1'(x) = 1 = 1w_0(x) + 0w_1(x),$$

and

$$(\mathcal{T}v_2)(x) = v_2'(x) = 2x = 0w_0(x) + 2w_1(x).$$

That is,

$$[\mathcal{T}v_0]_{\mathcal{B}_{\mathbb{P}^1}} = \begin{bmatrix}0\\0\end{bmatrix}, \qquad [\mathcal{T}v_1]_{\mathcal{B}_{\mathbb{P}^1}} = \begin{bmatrix}1\\0\end{bmatrix}, \quad \text{and} \quad [\mathcal{T}v_2]_{\mathcal{B}_{\mathbb{P}^1}} = \begin{bmatrix}0\\2\end{bmatrix}.$$

The matrix representation of $\mathcal{T}$ is therefore

$$[\mathcal{B}_{\mathbb{P}^1}^{-1}\mathcal{T}\mathcal{B}_{\mathbb{P}^2}]_{\mathcal{B}_{\mathbb{P}^1}\leftarrow\mathcal{B}_{\mathbb{P}^2}} = \begin{bmatrix}[\mathcal{T}v_0]_{\mathcal{B}_{\mathbb{P}^1}} & [\mathcal{T}v_1]_{\mathcal{B}_{\mathbb{P}^1}} & [\mathcal{T}v_1]_{\mathcal{B}_{\mathbb{P}^1}}\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\0 & 0 & 2\end{bmatrix}.$$

We observe that this matrix has dependent columns but full column rank; that is, its column space is $\mathbb{F}^2$ but its null space is nontrivial. This encodes the fact that $\mathcal{T}$ is surjective but not injective.

(ii) Keep the bases the same as in the previous part, and now define $\mathcal{S}\colon \mathbb{P}^1 \to \mathbb{P}^2$ by $(\mathcal{S}p)(x) := \int_0^x p(s)\ ds$. We have

$$(\mathcal{S}w_0)(x) = \int_0^x 1\ ds = x = 0v_0(x) + 1v_1(x) + 0v_2(x)$$

and

$$(\mathcal{S}w_1)(x) = \int_0^x s\ ds = \frac{x^2}{2} = 0v_0(x) + 0v_1(x) + \frac{v_2(x)}{2}.$$

Then

$$[\mathcal{S}w_0]_{\mathcal{B}_{\mathbb{P}^2}} = \begin{bmatrix}0\\1\\0\end{bmatrix} \quad \text{and} \quad [\mathcal{S}w_1]_{\mathcal{B}_{\mathbb{P}^1}} = \begin{bmatrix}0 & 0 & 1/2\end{bmatrix}.$$

The matrix representation of $\mathcal{S}$ is therefore

$$[\mathcal{B}_{\mathbb{P}^2}^{-1}\mathcal{S}\mathcal{B}_{\mathbb{P}^1}]_{\mathcal{B}_{\mathbb{P}^2}\leftarrow\mathcal{B}_{\mathbb{P}^1}} = \begin{bmatrix}[\mathcal{S}w_0]_{\mathcal{B}_{\mathbb{P}^2}} & [\mathcal{S}w_1]_{\mathcal{B}_{\mathbb{P}^2}}\end{bmatrix} = \begin{bmatrix}0 & 0\\1 & 0\\0 & 1/2\end{bmatrix}.$$

We observe that this matrix has independent columns but does not have full row rank. This encodes the fact that $\mathcal{S}$ is injective but not surjective.

37.2 Problem (!). Let $\mathbb{P}^2$ have the "standard" basis (v_0, v_1, v_2) with $v_j(x) = x^j$. Find the matrix representation of $\mathcal{T}p := p'$ with respect to this basis.

37.3 Problem (!). **(i)** Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces with bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$, respectively. Let $\mathcal{B}_{\mathcal{V}} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ and $\mathcal{B}_{\mathcal{W}} \in \mathbf{L}(\mathbb{F}^m, \mathcal{W})$ be the respective basis operators. For $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and $v \in \mathcal{V}$, show that

$$[\mathcal{T}v]_{\mathcal{B}_{\mathcal{W}}} = [\mathcal{T}]_{\mathcal{B}_{\mathcal{W}}\leftarrow\mathcal{B}_{\mathcal{V}}}[v]_{\mathcal{B}_{\mathcal{V}}}.$$

(ii) Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be finite-dimensional vector spaces with bases $(u_1, \ldots, u_n)$, $(v_1, \ldots, v_m)$, and $(w_1, \ldots, w_p)$, respectively. Let $\mathcal{B}_{\mathcal{U}} \in \mathbf{L}(\mathbb{F}^n, \mathcal{U})$, $\mathcal{B}_{\mathcal{V}} \in \mathbf{L}(\mathbb{F}^m, \mathcal{V})$, and $\mathcal{B}_{\mathcal{W}} \in \mathbf{L}(\mathbb{F}^p, \mathcal{W})$ be the respective basis operators. For $\mathcal{T} \in \mathbf{L}(\mathcal{U}, \mathcal{V})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, show that

$$[\mathcal{S}\mathcal{T}]_{\mathcal{B}_{\mathcal{W}} \leftarrow \mathcal{B}_{\mathcal{U}}} = [\mathcal{S}]_{\mathcal{B}_{\mathcal{W}} \leftarrow \mathcal{B}_{\mathcal{V}}} [\mathcal{T}]_{\mathcal{B}_{\mathcal{V}} \leftarrow \mathcal{B}_{\mathcal{U}}}$$

37.4 Problem (!). Let $\mathcal{V}$ be a finite-dimensional vector space with bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_n)$, and let $\mathcal{B}_{\mathcal{V}}$, $\mathcal{B}_{\mathcal{W}} \in \mathbf{L}(\mathcal{V})$ be the respective basis operators. For $v \in \mathcal{V}$, prove that

$$[v]_{\mathcal{B}_{\mathcal{W}}} = [\mathcal{I}_{\mathcal{V}}]_{\mathcal{B}_{\mathcal{W}} \leftarrow \mathcal{B}_{\mathcal{V}}} [v]_{\mathcal{B}_{\mathcal{V}}}.$$

Explain why $[\mathcal{I}_{\mathcal{V}}]_{\mathcal{B}_{\mathcal{W}} \leftarrow \mathcal{B}_{\mathcal{V}}}$ is often called the **CHANGE OF BASIS** matrix from the basis $(v_1, \ldots, v_n)$ to the basis $(w_1, \ldots, w_n)$.

37.5 Problem (!). **(i)** Compute the matrix products

$$[\mathcal{B}_{\mathbb{P}^1}^{-1}\mathcal{T}\mathcal{B}_{\mathbb{P}^2}]_{\mathcal{B}_{\mathbb{P}^1} \leftarrow \mathcal{B}_{\mathbb{P}^2}} [\mathcal{B}_{\mathbb{P}^2}^{-1}\mathcal{S}\mathcal{B}_{\mathbb{P}^1}]_{\mathcal{B}_{\mathbb{P}^2} \leftarrow \mathcal{B}_{\mathbb{P}^1}} \quad \text{and} \quad [\mathcal{B}_{\mathbb{P}^2}^{-1}\mathcal{S}\mathcal{B}_{\mathbb{P}^1}]_{\mathcal{B}_{\mathbb{P}^2} \leftarrow \mathcal{B}_{\mathbb{P}^1}} [\mathcal{B}_{\mathbb{P}^1}^{-1}\mathcal{T}\mathcal{B}_{\mathbb{P}^2}]_{\mathcal{B}_{\mathbb{P}^1} \leftarrow \mathcal{B}_{\mathbb{P}^2}}$$

from Example 37.1. Discuss your results in the context of part (ii) of Example 9.9 and part (ii) of Problem 37.3.

(ii) Let $v_0(x) = 1$, $v_1(x) = x$, $w_0(x) = x + 1$, $w_1(x) = x - 1$. Certainly (v_0, v_1) is a basis for $\mathbb{P}^1$; explain why (w_0, w_1) is also a basis for $\mathbb{P}^1$. Find the change of basis matrix from (v_0, v_1) to (w_0, w_1).

(iii) Let $p(x) = 1977x + 1138$. Find the coordinates of p with respect to the basis (w_0, w_1) for $\mathbb{P}^1$ from the previous part.

We have amassed a wealth of information about operators on finite-dimensional spaces, but we have not really considered how an operator interacts with a given basis; our approach has been that we are given an operator on a finite-dimensional space, and we are given a basis for the domain, and then we put the two together and see what they do. Suppose instead that we start with a finite-dimensional space $\mathcal{V}$ and an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$, and we consider what basis, or bases, of $\mathcal{V}$ interact "best" with $\mathcal{T}$. (We are taking the codomain of $\mathcal{T}$ to be $\mathcal{V}$ right now for simplicity; eventually we will ask this sort of question when $\mathcal{T}$ maps to a possibly different space.)

Often "best" means "simplest," and we know that the simplest action of an operator is on its eigenvectors. Suppose that $\mathcal{V}$ has a "basis of eigenvectors" of $\mathcal{T}$. That is, there is a basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$ such that each v_j is an eigenvector of $\mathcal{T}$. Then there is $\lambda_j \in \mathbb{F}$ such that $\mathcal{T}v_j = \lambda_j v_j$ for each j. We do *not* assume that the λ_j are distinct, or nonzero, or anything other than scalars in $\mathbb{F}$.

Now write any $v \in \mathcal{V}$ as $v = \sum_{j=1}^n \varphi_j(v) v_j$, where $(\varphi_1, \ldots, \varphi_n)$ is the dual basis associated with $(v_1, \ldots, v_n)$. Then

$$\mathcal{T}v = \sum_{j=1}^n \varphi_j(v)\mathcal{T}v_j = \sum_{j=1}^n \lambda_j \varphi_j(v) v_j \tag{37.1}$$

The action of $\mathcal{T}$ overall is therefore pretty simple, if we know the coordinates of a vector with respect to the basis $(v_1, \ldots, v_n)$.

We can push this a little further and factor $\mathcal{T}$. Let $\mathcal{B} \colon \mathbb{F}^n \to \mathcal{V}$ be the basis operator for $\mathcal{V}$ associated with this "eigenbasis." That is,

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

is an isomorphism, and

$$\mathcal{B}^{-1} v = (\varphi_1(v), \ldots, \varphi_n(v))$$

for all $v \in \mathcal{V}$. This, by the way, is the original formula for the dual basis functionals from (20.2). Then

$$\mathcal{T} v = \sum_{j=1}^{n} \lambda_j \varphi_j(v) v_j = \mathcal{B}(\lambda_1 \varphi_1(v), \ldots, \lambda_n \varphi_n(v)). \tag{37.2}$$

Put

$$\Lambda := \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{F}^{n \times n}.$$

That is, Λ is the $n \times n$ diagonal matrix whose jth diagonal entry is λ_j, and so

$$(\lambda_1 \varphi_1(v), \ldots, \lambda_n \varphi_n(v)) = \Lambda(\varphi_1(v), \ldots, \varphi_n(v)) = \Lambda \mathcal{B}^{-1} v. \tag{37.3}$$

We combine (37.1), (37.2), and (37.3) to conclude that $\mathcal{T}$ factors as

$$\mathcal{T} = \mathcal{B} \mathcal{M}_{\Lambda} \mathcal{B}^{-1}.$$

One upshot of this factorization of $\mathcal{T}$ is that computing powers is very easy:

$$\mathcal{T}^k = (\mathcal{B} \mathcal{M}_{\Lambda} \mathcal{B}^{-1})^k = \mathcal{B} \mathcal{M}_{\Lambda^k} \mathcal{B}^{-1},$$

and since $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal, we have

$$\Lambda^k = \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k).$$

Conversely, if we start with $\mathcal{T}$ factored as $\mathcal{B} \mathcal{M}_{\Lambda} \mathcal{B}^{-1}$, then there must be a basis of eigenvectors of $\mathcal{T}$.

37.6 Problem (!). **(i)** Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces. The operators $\mathcal{S} \in \mathbf{L}(\mathcal{V})$ and $\mathcal{T} \in \mathbf{L}(\mathcal{W})$ are **SIMILAR** if there is an invertible operator $\mathcal{B} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ such that $\mathcal{T} = \mathcal{B} \mathcal{S} \mathcal{B}^{-1}$. Prove that similar operators have the same eigenvalues.

(ii) Let $\mathcal{V}$ be a finite-dimensional vector space and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is diagonalizable with $\mathcal{T} = \mathcal{B} \mathcal{M}_{\Lambda} \mathcal{B}^{-1}$ for some invertible $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ and diagonal $\Lambda \in \mathbb{F}^{n \times n}$. Prove that the diagonal entries of Λ are the eigenvalues of $\mathcal{T}$ with eigenvectors $(\mathcal{B}\mathbf{e}_1, \ldots, \mathcal{B}\mathbf{e}_n)$, so in particular $(\mathcal{B}\mathbf{e}_1, \ldots, \mathcal{B}\mathbf{e}_n)$ is a basis for $\mathcal{V}$ consisting of eigenvectors of $\mathcal{T}$.

(iii) Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces. Suppose that $\mathcal{S} \in \mathbf{L}(\mathcal{V})$ and $\mathcal{T} \in \mathbf{L}(\mathcal{W})$ are similar.

Prove that $\mathcal{S}$ is finite-rank if and only if $\mathcal{T}$ is finite-rank, in which case $\operatorname{rank}(\mathcal{S}) = \operatorname{rank}(\mathcal{T})$.

Content from *Linear Algebra* by Meckes & Meckes. Pages 203–206 discuss similarity and diagonalization. Similarity is treated only at the level of matrices here. Do Quick Exercises #28 on p. 203 and #29 on p. 206.

We name this situation.

37.7 Definition. *Let $\mathcal{V}$ be a finite-dimensional vector space and let $n := \dim(\mathcal{V}) \geq 1$. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* ***DIAGONALIZABLE*** *if one of the following equivalent conditions holds:*

(i) *There exists a basis for $\mathcal{V}$ consisting of eigenvectors of $\mathcal{T}$.*

(ii) *There exist a diagonal matrix $\Lambda \in \mathbb{F}^{n \times n}$ and a basis for $\mathcal{V}$ such that if $\mathcal{B}$ is the associated basis operator, then*

$$\mathcal{T} = \mathcal{B}\mathcal{M}_{\Lambda}\mathcal{B}^{-1}.$$

That is, the matrix representation of $\mathcal{T}$ with respect to this basis is diagonal. In this case we say that $\mathcal{T}$ is ***DIAGONAL WITH RESPECT TO*** *this basis.*

(iii) *$\mathcal{T}$ is similar to matrix multiplication by a diagonal matrix.*

A matrix $A \in \mathbb{F}^{n \times n}$ is ***DIAGONALIZABLE*** *if the multiplication operator $\mathcal{M}_A \in \mathbf{L}(\mathbb{F}^{n \times n})$ is diagonalizable.*

37.8 Example. **(i)** Let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{F}^{n \times n}$ be diagonal. Then the multiplication operator $\mathcal{M}_{\Lambda} \in \mathbf{L}(\mathbb{F}^n)$ is diagonal with respect to the standard basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. Given $\mathbf{v} \in \mathbb{F}^n$, we have $\mathbf{v} = \sum_{j=1}^n (\mathbf{v} \cdot \mathbf{e}_j)\mathbf{e}_j$ and so

$$\mathcal{M}_{\Lambda}\mathbf{v} = \Lambda\mathbf{v} = \sum_{j=1}^n \lambda_j(\mathbf{v} \cdot \mathbf{e}_j)\mathbf{e}_j.$$

(ii) We consider $\mathbb{P}^n$ as a real vector space and define $\mathcal{T} \in \mathbf{L}(\mathbb{P}^n)$ by $\mathcal{T}p := p'$. If $\mathcal{T}p = \lambda p$, then $p'' = \lambda p$, thus $p(x) = p(0)e^{\lambda x}$. If $\lambda = 0$, then $p(x) = 1$ is an eigenvector of $\mathcal{T}$. However, for $\lambda \neq 0$, any function of the form $p(x) = p(0)e^{\lambda x}$ is not a polynomial (of any degree), and so no $\lambda \in \mathbb{R} \setminus \{0\}$ is an eigenvalue. That is, the only eigenvalue of $\mathcal{T}$ is 0. Moreover, the eigenspace corresponding to 0 is $\operatorname{span}(1)$, which is one-dimensional. So, there cannot exist a list of length $n+1$ consisting of eigenvectors of $\mathcal{T}$, and therefore $\mathcal{T}$ is not diagonalizable.

(iii) Consider the subspace $\mathcal{V}$ of $\mathcal{C}^{\infty}(\mathbb{R})$ given by

$$\mathcal{V} = \operatorname{span}(1, \sin(\cdot), \cos(\cdot)).$$

Define $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ by $\mathcal{T}f := f''$. Note that properties of trigonometric derivatives ensure $\mathcal{T}f \in \mathcal{V}$ for all $f \in \mathcal{V}$. The list $(1, \sin(\cdot), \cos(\cdot))$ is a basis for $\mathcal{V}$ since it is independent (there are multiple ways to check this, including the tried-and-true definition of independence, or an orthogonality argument in the style of part (i) of Example 32.3).

Since

$$\mathcal{T}1 = 0, \qquad \mathcal{T}\sin(\cdot) = -\sin(\cdot), \qquad \text{and} \qquad \mathcal{T}\cos(\cdot) = -\cos(\cdot),$$

the matrix representation of $\mathcal{T}$ with respect to this basis is

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

This is diagonal, so $\mathcal{T}$ is diagonalizable. Equivalently, for $f \in \mathcal{V}$, write

$$f(x) = \alpha_1 + \alpha_2 \sin(x) + \alpha_3 \cos(x)$$

to have

$$f''(x) = -\alpha_2 \sin(x) - \alpha_3 \cos(x),$$

which we recognize as the expansion (37.1) for $\mathcal{T}$ here.

Content from *Linear Algebra* by Meckes & Meckes. Pages 191–193 discuss diagonalization. Do Quick Exercise #25 on p. 192 and #26 on p. 194 (make sure you understand the book's notation for coordinates here).

37.9 Problem (!). **(i)** Let $\mathcal{V}$ be a finite-dimensional vector space and $\lambda \in \mathbb{F}$. Show that $\mathcal{T} := \lambda \mathcal{I}_{\mathcal{V}}$ is diagonal with respect to any basis of $\mathcal{V}$.

(ii) However, give an example of a diagonal matrix $\Lambda \in \mathbb{F}^{n \times n}$ and a basis $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ for $\mathbb{F}^n$ such that $\mathcal{M}_\Lambda$ is not diagonal with respect to this basis. (You can take $n = 2$ here.)

37.10 Problem (!). **(i)** Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim(\mathcal{V}) = n$. Prove that if $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ has n distinct eigenvalues, then $\mathcal{T}$ is diagonalizable.

(ii) Give an example of an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ for some finite-dimensional vector space $\mathcal{V}$ such that $\mathcal{T}$ has fewer than $\dim(\mathcal{V})$ distinct eigenvalues and yet is diagonalizable.

37.11 Problem ($\star$). Let $\mathcal{V}$ be a finite-dimensional vector space, let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ be diagonalizable, and let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$ consisting of eigenvectors of $\mathcal{T}$. Let $\lambda \in \mathbb{F}$ be an eigenvalue of $\mathcal{T}$, and suppose that $(v_{j_1}, \ldots, v_{j_k})$ is the sublist of $(v_1, \ldots, v_n)$ consisting of all eigenvectors in the list $(v_1, \ldots, v_n)$ corresponding to λ. That is,

$$\mathcal{T}v_j = \begin{cases} \lambda v_j, & j = j_1, \ldots, j_k \\ \lambda_j v_j, & j \neq j_1, \ldots, j_k, \lambda_j \neq \lambda. \end{cases}$$

Show that $(v_{j_1}, \ldots, v_{j_k})$ is a basis for the eigenspace $\mathcal{E}_\lambda(\mathcal{T})$ of $\mathcal{T}$ corresponding to λ. [Hint: *explain why we just have to show* $\mathcal{E}_\lambda(\mathcal{T}) = \operatorname{span}(v_{j_1}, \ldots, v_{j_k})$. *Let* $v \in \mathcal{E}_\lambda(\mathcal{T})$, *write*

$v = \sum_{j=1}^{n} \varphi_j(v) v_j$, where $(\varphi_1, \ldots, \varphi_n)$ is the dual basis for $\mathcal{V}$ corresponding to $(v_1, \ldots, v_n)$. Compute $\mathcal{T}v$ in two ways: $\mathcal{T}v = \sum_{j=1}^{n} \varphi_j(v)\mathcal{T}v_j$ and $\mathcal{T}v = \lambda v$. Equate the two expressions for $\mathcal{T}v$ and use independence to obtain $\varphi_j(v) = 0$ for $j \neq j_1, \ldots, j_k$.]

37.12 Problem ($\star$). Let $\mathcal{V}$ be an inner product space and suppose that $(u_1, \ldots, u_r)$ is an orthonormal list in $\mathcal{V}$. Let $\lambda_1, \ldots, \lambda_r \in \mathbb{F} \setminus \{0\}$ and define $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ by

$$\mathcal{T}v = \sum_{j=1}^{r} \lambda_j \langle v, u_j \rangle u_j.$$

Use Problem 24.10 to show that $\operatorname{rank}(\mathcal{T}) = r$.

Since not every operator is diagonalizable, we might ask what is the "next best thing." If $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is diagonalizable with $\mathcal{T} = \mathcal{B}\mathcal{M}_\Lambda\mathcal{B}^{-1}$ and $\Lambda \in \mathbb{F}^{n \times n}$ diagonal, we might call this factorization "eigenvalue-revealing" because we can read off the eigenvalues of $\mathcal{T}$ from the diagonal of Λ. So, given an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ with $\mathcal{V}$ finite-dimensional, is there an "eigenvalue-revealing" factorization of $\mathcal{T}$? That is, can we write $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}$ for some invertible $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ and some $A \in \mathbb{F}^{n \times n}$, where we can easily read off the eigenvalues of $\mathcal{T}$ from A? Yes, if we work on complex spaces (an unsurprising restriction, since eigenvalues are in play). It turns out that every operator on a complex finite-dimensional space is similar to matrix multiplication by an *upper-triangular* matrix, and experience teaches us that the eigenvalues of triangular matrices are their diagonal entries (though we have not proved this, outside of the special case of diagonal matrices in Problem 8.7).

Day 38: Friday, April 17.

Vocabulary from today

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

Upper-triangular matrix, operator that is upper-triangular with respect to a basis

38.1 Definition. *A matrix $A \in \mathbb{F}^{n \times n}$ is* **UPPER-TRIANGULAR** *if the (i, j)-entry of A is 0 for $i > j$. That is, the entries of A are zero below the diagonal.*

To prove that every operator on a finite-dimensional complex space is similar to matrix multiplication by an upper-triangular matrix, it will first be helpful to think dynamically: what does an upper-triangular matrix *do*, and how does that relate to what the similar operator does? Suppose that $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}$ with $A \in \mathbb{F}^{n \times n}$ upper-triangular and $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ invertible. Let $v_j = \mathcal{B}\mathbf{e}_j$. What is $\mathcal{T}$ doing to the basis $(v_1, \ldots, v_n)$? If A were diagonal, we would have $\mathcal{T}v_j = \lambda_j v_j$ for some scalars λ_j. What can we say now about $\mathcal{T}v_j$ relative to the other vectors in the basis $(v_1, \ldots, v_n)$?

The easiest case is when the operator is literally just matrix multiplication on $\mathbb{F}^n$ by an upper-triangular matrix. For simplicity, take $n = 3$ and say that $\mathcal{T} = \mathcal{M}_A$ with

$$A = \begin{bmatrix} a & b & d \\ 0 & c & e \\ 0 & 0 & f \end{bmatrix}.$$

What is $\mathcal{T}$ doing to the standard basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ for $\mathbb{F}^3$? We have

$$\mathcal{T}\mathbf{e}_1 = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} = a\mathbf{e}_1, \qquad \mathcal{T}\mathbf{e}_2 = \begin{bmatrix} b \\ c \\ 0 \end{bmatrix} = b\mathbf{e}_1 + c\mathbf{e}_2, \qquad \text{and} \qquad \mathcal{T}\mathbf{e}_3 = \begin{bmatrix} d \\ e \\ f \end{bmatrix} = d\mathbf{e}_1 + e\mathbf{e}_2 + f\mathbf{e}_3.$$

It looks like $\mathcal{T}\mathbf{e}_j \in \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_j)$ for each j; this is unsurprising at the end with $j = 3$ but noticeable for $j = 1, 2$.

This is true in general. There is a "static" way of thinking about upper-triangular matrices—the subdiagonal entries are zero—and a "dynamic" way—what upper-triangular matrices do to the standard basis vectors. (What things do defines what things are.)

38.2 Problem ($\star$). **(i)** Show that if $A \in \mathbb{F}^{n \times n}$ is upper-triangular, then

$$A\mathbf{e}_j \in \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_j) \text{ for } j = 1, \ldots, n. \tag{38.1}$$

(ii) Conversely, suppose (38.1) holds for some $A \in \mathbb{F}^{n \times n}$. Show that A is upper-triangular.

Now we can describe the behavior of an operator that is similar to matrix multiplication by an upper-triangular matrix. Suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ with $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}$, $A \in \mathbb{F}^{n \times n}$ upper-triangular, and $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ invertible. Let $v_j = \mathcal{B}\mathbf{e}_j$. Then $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$ and

$$\mathcal{T}v_j = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}v_j = \mathcal{B}\mathcal{M}_A\mathbf{e}_j = \mathcal{B}\sum_{k=1}^{j} \alpha_k \mathbf{e}_k$$

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

$$\mathcal{T}v_j = \sum_{k=1}^{j} \alpha_k \mathcal{B}\mathbf{e}_k = \sum_{k=1}^{j} \alpha_k v_k \in \operatorname{span}(v_1, \ldots, v_j).$$

38.3 Problem (!). Conversely, let $\mathcal{V}$ be a finite-dimensional vector space with basis $(v_1, \ldots, v_n)$; what if $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ satisfies $\mathcal{T}v_j \in \operatorname{span}(v_1, \ldots, v_j)$ for $j = 1, \ldots, n$? Show that the matrix representation of $\mathcal{T}$ with respect to this basis $(v_1, \ldots, v_n)$ is upper-triangular.

38.4 Definition. *Let $\mathcal{V}$ be a finite-dimensional vector space. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* ***UPPER-TRIANGULAR*** *with respect to a basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$ if one of the following equivalent conditions hold.*

(i) *There exists an upper-triangular matrix* $A \in \mathbb{F}^{n\times n}$ *such that if* $\mathcal{B}$ *is the basis operator associated with* $(v_1, \ldots, v_n)$*, then*

$$\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}.$$

That is, the matrix representation of $\mathcal{T}$ *with respect to this basis is upper-triangular.*

(ii) $\mathcal{T}v_j \in \operatorname{span}(v_1, \ldots, v_j)$ *for* $j = 1, \ldots, n$.

38.5 Example. Let $A \in \mathbb{F}^{n\times n}$ be upper-triangular: $A_{ij} = 0$ for $i \geq j+1$. Then the multiplication operator $\mathcal{M}_A$ is upper-triangular with respect to the standard basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$.

Before proving that every operator on a finite-dimensional complex space is similar to matrix multiplication by an upper-triangular matrix, we explore some properties of upper-triangular operators. Our experience from a first course in linear algebra might teach us that (1) an upper-triangular matrix is invertible if and only if all of its diagonal entries are nonzero, (2) the inverse of an invertible upper-triangular matrix is also upper-triangular, (3) the product of upper-triangular matrices is upper-triangular, and (4) the diagonal entries of a product of upper-triangular matrices are the products of the corresponding diagonal entries of the upper-triangular factors. We can prove all of these facts from Definition 38.4 and reinterpret them in an operator-theoretic context, not just a matrix context. We prove (2) and (3) in arbitrary finite-dimensional vector spaces, so we start there, but (1) and (4) are most naturally expressed in the framework of an inner product space.

38.6 Theorem. *Let* $\mathcal{V}$ *be a finite-dimensional vector space, and suppose that* $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ *is invertible and upper-triangular with respect to the basis* $(v_1, \ldots, v_n)$ *for* $\mathcal{V}$*. Then* $\mathcal{T}^{-1}$ *is also upper-triangular with respect to this basis.*

Proof. Our goal is that $\mathcal{T}^{-1}v_j \in \operatorname{span}(v_1, \ldots, v_j)$ for each j. That is, we want to write $\mathcal{T}^{-1}v_j = \sum_{k=1}^{j} \alpha_k v_k$. This is, of course, equivalent to $v_j = \sum_{k=1}^{j} \alpha_k \mathcal{T}v_k$. We know that $\mathcal{T}v_j \in \operatorname{span}(v_1, \ldots, v_j)$, and so what we want is that $v_j \in \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_j)$. Since $\mathcal{T}$ is invertible, this should feel plausible.

Indeed, when $\mathcal{T}$ is upper-triangular with respect to the basis $(v_1, \ldots, v_n)$ and also invertible, we do have $\operatorname{span}(v_1, \ldots, v_j) = \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_j)$ for each j. Here is why. Certainly $\mathcal{T}v_k \in \operatorname{span}(v_1, \ldots, v_k) \subseteq \operatorname{span}(v_1, \ldots, v_j)$ for $1 \leq k \leq j$. Then $(\mathcal{T}v_1, \ldots, \mathcal{T}v_j)$ is an independent list of length j in the j-dimensional subspace $\operatorname{span}(v_1, \ldots, v_j)$. This list is therefore a basis for $\operatorname{span}(v_1, \ldots, v_j)$, so $\operatorname{span}(v_1, \ldots, v_j) = \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_j)$. ■

38.7 Problem (⋆). Suppose that $\mathcal{V}$ is a finite-dimensional vector space and that $\mathcal{S}, \mathcal{T} \in \mathbf{L}(\mathcal{V})$ are both upper-triangular with respect to the *same* basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$. Prove that $\mathcal{S}\mathcal{T}$ is also upper-triangular with respect to this basis. [Hint: *if* $w_k \in \operatorname{span}(v_1, \ldots, v_k)$*, then* $\operatorname{span}(w_1, \ldots, w_j) \subseteq \operatorname{span}(v_1, \ldots, v_j)$.]

Content from *Linear Algebra* by Meckes & Meckes. Lemma 3.68 on p. 219 expresses a version of this notion of upper-triangular operator.

Upper-triangular operators on inner product spaces have particularly transparent behaviors, thanks to the representation powers of orthonormal bases. Let $\mathcal{V}$ be a finite-dimensional inner product space and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is upper-triangular with respect to an orthonormal basis $(u_1, \ldots, u_n)$ for $\mathcal{V}$. Since $\mathcal{T}$ is upper-triangular with respect to this basis, we know that $\mathcal{T}u_j \in \operatorname{span}(u_1, \ldots, u_j)$ for each j, and since the (sub)list $(u_1, \ldots, u_j)$ is orthonormal, we have

$$\mathcal{T}u_j = \sum_{k=1}^{j} \langle \mathcal{T}u_j, u_k \rangle \, u_k. \tag{38.2}$$

We will use (38.2) frequently in the following arguments.

38.8 Theorem. *Let $\mathcal{V}$ be a finite-dimensional inner product space and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is upper-triangular with respect to an orthonormal basis $(u_1, \ldots, u_n)$ for $\mathcal{V}$. Then $\mathcal{T}$ is invertible if and only if $\langle \mathcal{T}u_j, u_j \rangle \neq 0$ for each j.*

Proof. We prove the negation: $\mathcal{T}$ is not invertible if and only if $\langle \mathcal{T}u_j, u_j \rangle = 0$ for some j.

($\Longrightarrow$) If $\mathcal{T}$ is not invertible, there is $v \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$ such that $\mathcal{T}v = 0_{\mathcal{V}}$. Expand $v = \sum_{k=1}^{j} \langle v, u_k \rangle \, u_k$, where $\langle v, u_j \rangle \neq 0$ and $\langle v, u_k \rangle = 0$ for $k = j+1, \ldots, n$. We consider two cases on j.

1. $j = 1$. Then

$$0_{\mathcal{V}} = \mathcal{T}v = \mathcal{T}(\langle v, u_1 \rangle \, u_1) = \langle v, u_1 \rangle \, \mathcal{T}u_1,$$

so $\mathcal{T}u_1 = 0_{\mathcal{V}}$ since $\langle v, u_1 \rangle \neq 0$ here. Then

$$0_{\mathcal{V}} = \mathcal{T}u_1 = \langle \mathcal{T}u_1, u_1 \rangle \, u_1,$$

so $\langle \mathcal{T}u_1, u_1 \rangle = 0$.

2. $2 \leq j \leq n$. We have

$$0_{\mathcal{V}} = \mathcal{T}v = \mathcal{T}\sum_{k=1}^{j} \langle v, u_k \rangle \, u_k = \sum_{k=1}^{j} \langle v, u_k \rangle \, \mathcal{T}u_k = \sum_{k=1}^{j-1} \langle v, u_k \rangle \, \mathcal{T}u_k + \langle v, u_j \rangle \, \mathcal{T}u_j,$$

and so we can solve for $\mathcal{T}u_j$ as

$$\mathcal{T}u_j = \sum_{k=1}^{j-1} \left(-\frac{\langle v, u_k \rangle}{\langle v, u_j \rangle} \right) \mathcal{T}u_k.$$

For $k = 1, \ldots, j-1$, we have $\mathcal{T}u_k \in \operatorname{span}(u_1, \ldots, u_k) \subseteq \operatorname{span}(u_1, \ldots, u_{j-1})$, and so $\mathcal{T}u_j \in \operatorname{span}(u_1, \ldots, u_{j-1})$. Thus

$$\mathcal{T}u_j = \sum_{k=1}^{j-1} \langle \mathcal{T}u_k, u_k \rangle \, u_k.$$

Comparing this to (38.2), we must have $\langle \mathcal{T}u_j, u_j \rangle = 0$.

($\Longleftarrow$) Again we consider two cases on j.

1. $j = 1$. Here $\langle \mathcal{T}u_1, u_1 \rangle = 0$. Since $\mathcal{T}u_1 = \langle \mathcal{T}u_1, u_1 \rangle u_1 = 0_{\mathcal{V}}$, we immediately see that $\mathcal{T}$ is not invertible.

2. $2 \le j \le n$. For any k, let $\mathcal{V}_k := \text{span}(u_1, \ldots, u_k)$, so $\dim(\mathcal{V}_k) = k$. Then $\mathcal{T}(\mathcal{V}_k) \subseteq \mathcal{V}_k$, but for $k = j$, we have

$$\mathcal{T}u_j = \sum_{k=1}^{j} \langle \mathcal{T}u_j, u_k \rangle u_k = \sum_{k=1}^{j-1} \langle \mathcal{T}u_j, u_k \rangle u_k \in \mathcal{V}_{j-1}.$$

Thus $\mathcal{T}(\mathcal{V}_j) \subseteq \mathcal{T}(\mathcal{V}_{j-1})$. That is, the restriction

$$\mathcal{T}\big|_{\mathcal{V}_j} : \mathcal{V}_j \to \mathcal{V}_{j-1} : v \mapsto \mathcal{T}v$$

is a linear operator from $\mathcal{V}_j$ to $\mathcal{V}_{j-1}$).

Since $\dim(\mathcal{V}_{j-1}) = j - 1 < j = \dim(\mathcal{V}_j)$, the restriction $\mathcal{T}\big|_{\mathcal{V}_j}$ is not injective, so there is $v \in \mathcal{V}_j \setminus \{0_{\mathcal{V}}\}$ such that $\mathcal{T}v = \mathcal{T}\big|_{\mathcal{V}_j} v = 0_{\mathcal{V}}$. Then $\mathcal{T}$ is not injective, either. ■

38.9 Problem (!). Prove that the eigenvalues of an upper-triangular matrix are its diagonal entries.

38.10 Problem ($\star$). **(i)** Let $\mathcal{V}$ be a finite-dimensional inner product space and suppose that $\mathcal{S}$, $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ are upper-triangular with respect to the same orthonormal basis $(u_1, \ldots, u_n)$ for $\mathcal{V}$. Prove that

$$\langle \mathcal{S}\mathcal{T}u_j, u_j \rangle = \langle \mathcal{S}u_j, u_j \rangle \langle \mathcal{T}u_j, u_j \rangle$$

[Hint: *calculate* $\mathcal{S}\mathcal{T}u_j$ *by first expressing* $\mathcal{T}u_j$ *as a vector in* $\text{span}(u_1, \ldots, u_j)$ *and then, for* $k = 1, \ldots, j$, *expressing* $\mathcal{S}u_k$ *as a vector in* $\text{span}(u_1, \ldots, u_k)$.]

(ii) Let R_1, $R_2 \in \mathbb{F}^{n \times n}$ be upper-triangular. What are the diagonal entries of R_1R_2?

38.11 Problem ($\star$). We could introduce a notion of lower-triangular operators, but we will not really need that. However, the following will be useful. Let $\mathcal{V}$ be a finite-dimensional inner product space, and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is upper-triangular with respect to the orthonormal basis $(u_1, \ldots, u_n)$, $n \ge 2$. Prove that $\langle \mathcal{T}^*u_n, u_j \rangle = 0$ for $j = 1, \ldots, n-1$, and conclude that $\mathcal{T}^*u_n \in \text{span}(u_n)$.

Day 39: Monday, April 20.

Now we are ready to prove that every operator on a finite-dimensional complex vector space is similar to matrix multiplication by an upper-triangular matrix. We can do this in two ways: directly for an operator on an abstract space, or we can show that if $A \in \mathbb{C}^{n \times n}$, then there is an invertible $B \in \mathbb{C}^{n \times n}$ and an upper-triangular $T \in \mathbb{C}^{n \times n}$ such that $A = BTB^{-1}$. Applying the operator result to $\mathcal{M}_A \in \mathbf{L}(\mathbb{C}^{n \times n})$ gives the matrix result. Conversely, since $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ factors as $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}$, where $\mathcal{B}$ is any basis operator for $\mathcal{V}$ and A is the matrix representation for $\mathcal{T}$ with respect to that basis, the matrix result implies the operator result.

39.1 Problem (!). Fill in the details of that last sentence.

However, there is some value in proving each result separately in that while both proofs use induction, they stress different properties of matrices and operators. We begin with the matrix proof.

39.2 Theorem. *Let $A \in \mathbb{C}^{n \times n}$. There exist an invertible matrix $B \in \mathbb{C}^{n \times n}$ and an upper-triangular matrix $T \in \mathbb{C}^{n \times n}$ such that $A = BTB^{-1}$.*

Proof. We induct on n. This proof is largely an exercise in manipulating block matrices.

1. *The base case $n = 1$.* Any matrix in $\mathbb{C}^{1 \times 1}$ is already upper-triangular, since there are no subdiagonal entries in play here. More precisely, if $A \in \mathbb{C}^{1 \times 1}$, it is vacuously true that if $i > j$, then $A_{ij} = 0$ for $i = 1$ and $j = 1$.

2. *The illustrative case $n = 2$.* The logic of induction allows us to proceed directly from the base case of $n = 1$ to the induction hypothesis and step. Nonetheless, it may be helpful to see the less trivial, but still small, case of $n = 2$ first before working more generally.

So, suppose that $A \in \mathbb{C}^{2 \times 2}$ and let $\lambda \in \mathbb{C}$ be an eigenvalue for A with eigenvector $\mathbf{v} \in \mathbb{C}^2 \setminus \{\mathbf{0}_2\}$. Let $\mathbf{w} \in \mathbb{C}^2$ such that $(\mathbf{v}, \mathbf{w})$ is a basis for $\mathbb{C}^2$. Then $A\mathbf{v} = \lambda\mathbf{v}$, and since $A\mathbf{w} \in \mathbb{C}^2$, there are c_1, $c_2 \in \mathbb{C}$ such that $A\mathbf{w} = c_1\mathbf{v} + c_2\mathbf{w}$. Put $B := \begin{bmatrix}\mathbf{v} & \mathbf{w}\end{bmatrix} \in \mathbb{C}^{2 \times 2}$. Then B is invertible and

$$\begin{aligned} AB &= A\begin{bmatrix}\mathbf{v} & \mathbf{w}\end{bmatrix} \\ &= \begin{bmatrix}A\mathbf{v} & A\mathbf{w}\end{bmatrix} \\ &= \begin{bmatrix}\lambda\mathbf{v} & (c_1\mathbf{v} + c_2\mathbf{w})\end{bmatrix} \\ &= \begin{bmatrix}(\lambda\mathbf{v} + 0\mathbf{w}) & (c_1 + c_2\mathbf{w})\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{v} & \mathbf{w}\end{bmatrix}\begin{bmatrix}\lambda & c_1 \\ 0 & c_2\end{bmatrix}. \end{aligned}$$

Put

$$T := \begin{bmatrix}\lambda & c_1 \\ 0 & c_2\end{bmatrix}$$

to conclude that T is upper-triangular with $A = BTB^{-1}$.

3. *The induction hypothesis and step.* Suppose that for all $A \in \mathbb{C}^{n \times n}$, there are an invertible $B \in \mathbb{C}^{n \times n}$ and an upper-triangular $T \in \mathbb{C}^{n \times n}$ such that $A = BTB^{-1}$.

Now let $A \in \mathbb{C}^{(n+1) \times (n+1)}$. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A with eigenvector $\mathbf{v} \in \mathbb{C}^{n+1} \setminus \{\mathbf{0}_{n+1}\}$. Choose $\mathbf{w}_2, \ldots, \mathbf{w}_{n+1} \in \mathbb{C}^{n+1}$ so that $(\mathbf{v}, \mathbf{w}_2, \ldots, \mathbf{w}_{n+1})$ is a basis for $\mathbb{C}^{n+1}$. Put

$B := \begin{bmatrix} \mathbf{v} & \mathbf{w}_2 & \cdots & \mathbf{w}_{n+1} \end{bmatrix} \in \mathbb{C}^{(n+1)\times(n+1)}$. Then there are $\mathbf{x}_j \in \mathbb{C}^{n+1}$ such that $A\mathbf{w}_j = B\mathbf{x}_j$ for $j = 2, \ldots, n+1$, from which it follows (check this) that

$$AB = B\begin{bmatrix} \lambda\mathbf{e}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{n+1} \end{bmatrix}. \tag{39.1}$$

Now write

$$\begin{bmatrix} \lambda\mathbf{e}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{n+1} \end{bmatrix} = \begin{bmatrix} \lambda & \mathbf{v}^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{A} \end{bmatrix}, \tag{39.2}$$

where $\mathbf{v}^{\mathsf{T}} \in \mathbb{C}^{1\times n}$ and $\widetilde{A} \in \mathbb{C}^{n\times n}$. Apply the induction hypothesis to factor $\widetilde{A} = \widetilde{B}\widetilde{T}\widetilde{B}^{-1}$, where $\widetilde{B} \in \mathbb{C}^{n\times n}$ is invertible and $\widetilde{T} \in \mathbb{C}^{n\times n}$ is upper-triangular. Then

$$\begin{bmatrix} \lambda & \mathbf{v}^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{A} \end{bmatrix} = \begin{bmatrix} \lambda & \mathbf{v}^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{B}\widetilde{T}\widetilde{B}^{-1} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}_n^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{B} \end{bmatrix} \begin{bmatrix} \lambda & \mathbf{v}^{\mathsf{T}}\widetilde{B} \\ \mathbf{0}_n & \widetilde{T} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_n^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{B}^{-1} \end{bmatrix}. \tag{39.3}$$

We conclude

$$A = \left(B\begin{bmatrix} 1 & \mathbf{0}_n^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{B} \end{bmatrix}\right) \begin{bmatrix} \lambda & \mathbf{v}^{\mathsf{T}}\widetilde{B} \\ \mathbf{0}_n & \widetilde{T} \end{bmatrix} \left(\begin{bmatrix} 1 & \mathbf{0}_n^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{B}^{-1} \end{bmatrix} B^{-1}\right),$$

where

$$B\begin{bmatrix} 1 & \mathbf{0}_n^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{B} \end{bmatrix}$$

is invertible with inverse

$$\left(B\begin{bmatrix} 1 & \mathbf{0}_n^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{B} \end{bmatrix}\right)^{-1} = \begin{bmatrix} 1 & \mathbf{0}_n^{\mathsf{T}} \\ \mathbf{0}_n & \widetilde{B}^{-1} \end{bmatrix} B^{-1},$$

and

$$\begin{bmatrix} \lambda & \mathbf{v}^{\mathsf{T}}\widetilde{B} \\ \mathbf{0}_n & \widetilde{T} \end{bmatrix}$$

is upper-triangular. ■

39.3 Problem (+). Most of our results about upper-triangular operators and matrices require complex numbers. Here is a situation in which everything can be real. Let $A \in \mathbb{R}^{n\times n}$ and suppose that all of the eigenvalues of A are real. The following proves the existence of an invertible matrix $B \in \mathbb{R}^{n\times n}$ and an upper-triangular matrix $T \in \mathbb{R}^{n\times n}$ such that $A = BTB^{-1}$.

(i) Let $n \geq 2$ and suppose that $A \in \mathbb{F}^{n\times n}$ has the form

$$A = \begin{bmatrix} \lambda & \mathbf{v}^{\mathsf{T}} \\ \mathbf{0}_{n-1} & \widetilde{A} \end{bmatrix},$$

where $\lambda \in \mathbb{F}$, $\mathbf{v} \in \mathbb{F}^{n-1}$, and $\widetilde{A} \in \mathbb{F}^{(n-1)\times(n-1)}$. Prove that if $\widetilde{\lambda} \in \mathbb{F}$ is an eigenvalue of $\widetilde{A}$, then $\widetilde{\lambda}$ is also an eigenvalue of A. [Hint: *consider two cases:* $\widetilde{\lambda} = \lambda$ *and* $\widetilde{\lambda} \neq \lambda$. *In the* *latter, if* $\widetilde{\mathbf{v}}$ *is an eigenvector for* $\widetilde{A}$ *corresponding to* $\widetilde{\lambda}$*, what choice of* $v \in \mathbb{F}$ *guarantees that* $(v, \widetilde{\mathbf{v}})$ *is an eigenvector for* A *corresponding to* $\widetilde{\lambda}$*?*]

(ii) Suppose that all of the eigenvalues of $A \in \mathbb{R}^{n\times n}$ are real. Adapt the proof of Theorem 39.2 to prove the existence of an invertible matrix $B \in \mathbb{R}^{n\times n}$ and an upper-triangular matrix $T \in \mathbb{R}^{n\times n}$ such that $A = BTB^{-1}$. [Hint: *induct on* n *as in the proof of Theorem 39.2. At the induction step, obtain results like* (39.1) *and* (39.2). *Show that all of the eigenvalues of* $\widetilde{A}$ *here are real and apply the induction hypothesis as in* (39.3).]

Here is the operator proof.

39.4 Theorem. *Let* $\mathcal{V}$ *be a complex finite-dimensional vector space. Any operator* $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ *is upper-triangular with respect to some basis of* $\mathcal{V}$*.*

Proof. We induct on $n := \dim(\mathcal{V})$. This proof is largely an exercise in how operators interact with bases.

1. *The base case* $n = 1$. If $n = \dim(\mathcal{V}) = 1$, then any basis for $\mathcal{V}$ has the form (v_1), and any operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ satisfies $\mathcal{T}v_1 = \alpha v_1 \in \operatorname{span}(v_1)$ for some $\alpha \in \mathbb{F}$.

2. *The illustrative case* $n = 2$. Here $\dim(\mathcal{V}) = 2$. Since $\mathcal{V}$ is a finite-dimensional complex vector space, $\mathcal{T}$ has an eigenvalue $\lambda \in \mathbb{C}$. Let $v_1 \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$ be an eigenvector for $\mathcal{T}$ corresponding to λ. Let $v_2 \in \mathcal{V} \setminus \operatorname{span}(v_1)$ be any vector that is not a scalar multiple of v_1. The list (v_1, v_2) is therefore independent and, since $\dim(\mathcal{V}) = 2$, a basis for $\mathcal{V}$. Certainly $\mathcal{T}v_1 = \lambda v_1 \in \operatorname{span}(v_1)$, and $\mathcal{T}v_2 \in \operatorname{span}(v_1, v_2)$, since (v_1, v_2) is a basis for $\mathcal{V}$.

3. *The induction hypothesis and step.* Suppose that for some $n \geq 1$, if $\mathcal{V}$ is a finite-dimensional complex vector space with $1 \leq \dim(\mathcal{V}) \leq n$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$, then $\mathcal{T}$ is upper-triangular with respect to some basis of $\mathcal{V}$.

Now let $\mathcal{V}$ be a finite-dimensional complex vector space with $1 \leq \dim(\mathcal{V}) \leq n+1$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. If $\dim(\mathcal{V}) \leq n$, then the induction hypothesis gives the desired result, so assume here that $\dim(\mathcal{V}) = n+1$. As before, we can take an eigenvalue $\lambda \in \mathbb{C}$ of $\mathcal{T}$. If $\mathcal{T} = \lambda\mathcal{I}_{\mathcal{V}}$, then $\mathcal{T}$ is diagonalizable, so $\mathcal{T}$ is similar to matrix multiplication by a diagonal, thus upper-triangular, matrix. Otherwise, suppose that $\mathcal{T} \neq \lambda\mathcal{I}_{\mathcal{V}}$. Given an eigenvector v_1 for $\mathcal{T}$ corresponding to λ, we have $\mathcal{T}v_1 = \lambda v_1 \in \operatorname{span}(v_1)$, and our inclination might be to extend the list (v_1) to a basis $(v_1, v_2, \ldots, v_{n+1})$ for $\mathcal{V}$ such that $\mathcal{T}v_j \in \operatorname{span}(v_1, \ldots, v_j)$ for $j = 2, \ldots, n+1$. Putting $\mathcal{V}_n := \operatorname{span}(v_2, \ldots, v_{n+1})$, we would hope that $\mathcal{T}(\mathcal{V}_n) \subseteq \mathcal{V}_n$; then $\mathcal{V}_n$ is an n-dimensional space, so the restriction $\mathcal{T}\big|_{\mathcal{V}_n}$ of $\mathcal{T}$ to $\mathcal{V}_n$ is upper-triangular with respect to some basis $(w_2, \ldots, w_{n+1})$ of $\mathcal{V}_n$. Then we could check that $(v_1, w_2, \ldots, w_{n+1})$ is a basis for $\mathcal{V}$ (this needs some work) and that $\mathcal{T}$ is upper-triangular with respect to that basis.

The problem is that we have no guarantee that $\operatorname{span}(v_2, \ldots, v_{n+1})$ is invariant under $\mathcal{T}$ for an *arbitrary* extension of (v_1) to a basis $(v_1, v_2, \ldots, v_{n+1})$ for $\mathcal{V}$. Here is one workaround, which abandons the idea of starting the basis with the eigenvector v_1. Let $\mathcal{U} := (\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})(\mathcal{V})$. Since $\mathcal{T} \neq \lambda\mathcal{I}_{\mathcal{V}}$, the space $\mathcal{U}$ is nontrivial; since $\ker(\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})$ is nontrivial as λ is an eigenvalue of $\mathcal{T}$, the operator $\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}} \in \mathbf{L}(\mathcal{V})$ is not injective, hence not surjective, and thus $\mathcal{U} \neq \mathcal{V}$.

All together, $\mathcal{U}$ is a proper, nontrivial subspace of $\mathcal{V}$, so $1 \le \dim(\mathcal{U}) < \dim(\mathcal{V}) = n+1$. By Problem 12.12, $\mathcal{U}$ is invariant under $\mathcal{T}$. That is, $\mathcal{T}u \in \mathcal{U}$ for all $u \in \mathcal{U}$, so the restriction

$$\mathcal{T}\big|_{\mathcal{U}} \colon \mathcal{U} \to \mathcal{U} \colon u \mapsto \mathcal{T}u$$

is an operator in $\mathbf{L}(\mathcal{U})$. Since $1 \le \dim(\mathcal{U}) \le n$, by the induction hypothesis there is a basis $(u_1, \ldots, u_r)$ of $\mathcal{U}$ such that $\mathcal{T}\big|_{\mathcal{U}}$ is upper-triangular with respect to this basis.

Extend the basis $(u_1, \ldots, u_r)$ to a basis $(u_1, \ldots, u_{n+1})$ for $\mathcal{V}$. We check that $\mathcal{T}$ is upper-triangular with respect to this basis. Certainly for $j = 1, \ldots, r$, we have $\mathcal{T}u_j = \mathcal{T}\big|_{\mathcal{U}}u_j \in \operatorname{span}(u_1, \ldots, u_j)$, so let $r+1 \le j \le n+1$. Then

$$\mathcal{T}u_j = \mathcal{T}u_j - \lambda u_j + \lambda u_j = (\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})u_j + \lambda u_j.$$

We have $(\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})u_j \in \mathcal{U} = \operatorname{span}(u_1, \ldots, u_r) \subseteq \operatorname{span}(u_1, \ldots, u_r, u_{r+1}, \ldots, u_j)$, and so

$$\mathcal{T}u_j = (\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})u_j + \lambda u_j \in \operatorname{span}(u_1, \ldots, u_r, u_{r+1}, \ldots, u_j). \qquad \blacksquare$$

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.67 on p. 219, and this is the proof given on p. 220. This is a hard proof. It appears as well at least in Axler's *Linear Algebra Done Right*, and Axler credits the proof to Arveson, one of the great operator theorists (a branch of infinite-dimensional linear algebra + analysis) of the last century.

39.5 Problem ($\star$). Reread the proof of Theorem 39.4. This problem uses the notation of that proof. Put $\widetilde{\mathcal{U}} := \operatorname{span}(u_{r+1}, \ldots, u_n)$. Let $\mathcal{B}_{\mathcal{V}}$ be the basis operator for the basis $(v_1, \ldots, u_{n+1})$ of $\mathcal{V}$, $\mathcal{B}_{\mathcal{U}}$ be the basis operator for the basis $(u_1, \ldots, u_r)$ of $\mathcal{U}$, and $\mathcal{B}_{\widetilde{\mathcal{U}}}$ be the basis operator for the basis $(u_{r+1}, \ldots, u_n)$ of $\widetilde{\mathcal{U}}$. Show that

$$[\mathcal{T}]_{\mathcal{V}\leftarrow\mathcal{V}} = \begin{bmatrix} [\mathcal{T}\big|_{\mathcal{U}}]_{\mathcal{U}\leftarrow\mathcal{U}} & [(\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})\big|_{\widetilde{\mathcal{U}}}]_{\mathcal{U}\leftarrow\widetilde{\mathcal{U}}} \\ 0 & \lambda\mathcal{I}_{n-r} \end{bmatrix}.$$

Day 40: Wednesday, April 22.

Vocabulary from today

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

Unitary operator/matrix (N), unitarily diagonalizable operator/matrix, normal operator/matrix

We took up the study of triangularization by looking for a more broadly applicable alternative to diagonalization that was still eigenvalue-revealing. It turns out that the universal notion of triangularization has more specific applications to diagonalization in an inner product space. Previously we argued that, given a finite-dimensional vector space $\mathcal{V}$ and an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$, the "best" basis for $\mathcal{V}$ relative to $\mathcal{T}$ is a basis of eigenvectors of $\mathcal{T}$. (Of course, such a basis need not exist.) Suppose that we impose the additional structure of an inner product on $\mathcal{V}$. Since the best basis for an inner product space is an orthonormal basis, we might ask if $\mathcal{V}$ has an orthonormal basis of eigenvectors of $\mathcal{T}$. This would imply that $\mathcal{T}$ is diagonalizable, and a bit more, and so this is probably an even more restrictive situation than ordinary diagonalizability.

To understand this, it pays to ask a new question about an old object. What is special about the basis operator for an orthonormal basis? How does it behave?

40.1 Example. Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal basis $(u_1, \ldots, u_n)$ and let

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

be the basis operator. For $v \in \mathcal{V}$, since $v = \sum_{j=1}^{n} \langle v, u_j \rangle u_j$, we have $\mathcal{B}^{-1}v = (\langle v, u_1 \rangle_{\mathcal{V}}, \ldots, \langle v, u_n \rangle_{\mathcal{V}}) \in \mathbb{F}^n$. Conversely, fix $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$. Since $\mathcal{B}\boldsymbol{\alpha} = \sum_{j=1}^{n} \alpha_j u_j$, we have $\alpha_j = \langle \mathcal{B}\boldsymbol{\alpha}, u_j \rangle_{\mathcal{V}}$. Theorem 28.4 allows us to compute

$$\langle \mathcal{B}\alpha, v \rangle_{\mathcal{V}} = \sum_{j=1}^{n} \langle \mathcal{B}\boldsymbol{\alpha}, u_j \rangle_{\mathcal{V}} \overline{\langle v, u_j \rangle_{\mathcal{V}}} = \sum_{j=1}^{n} \alpha_j \overline{\langle v, u_j \rangle_{\mathcal{V}}} = \sum_{j=1}^{n} \alpha_j \overline{(\mathcal{B}^{-1}v)_j} = \boldsymbol{\alpha} \cdot (\mathcal{B}^{-1}v).$$

Thus $\mathcal{B}^* = \mathcal{B}^{-1}$.

40.2 Definition. **(i)** *Let $\mathcal{V}$ be an inner product space. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* ***UNITARY*** *if $\mathcal{T}$ is invertible and has an adjoint with $\mathcal{T}^* = \mathcal{T}^{-1}$.*

(ii) *A matrix $A \in \mathbb{F}^{n \times n}$ is* ***UNITARY*** *if A is invertible and $A^{-1} = A^*$.*

(iii) *A matrix $A \in \mathbb{R}^{n \times n}$ is* ***ORTHOGONAL*** *if A is invertible and $A^{-1} = A^{\mathsf{T}}$.*

40.3 Problem (!). Let $\mathcal{V}$ be an inner product space and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ has an adjoint. If $\mathcal{T}$ is both self-adjoint and unitary, what else do you know about $\mathcal{T}$?

The question that motivated our consideration of unitary operators was if if $\mathcal{V}$ is a finite-dimensional inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$, is there an orthonormal basis for $\mathcal{V}$ consisting of eigenvectors of $\mathcal{T}$? Now we see that this is equivalent to being able to factor $\mathcal{T} = \mathcal{B}\mathcal{M}_{\Lambda}\mathcal{B}^*$, where $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ is unitary and $\Lambda \in \mathbb{F}^{n \times n}$ is diagonal. If this is possible, note that here we are using $\mathcal{B}^* = \mathcal{B}^{-1}$ in the factorization. That is, we no longer have to invert the basis operator; this is particularly nice when $\mathcal{T}$ is matrix multiplication, as taking the adjoint of a matrix is much easier than computing a matrix inverse.

40.4 Definition. *Let $\mathcal{V}$ be a finite-dimensional vector space and let $n := \dim(\mathcal{V}) \geq 1$. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* ***UNITARILY DIAGONALIZABLE*** *if one of the following equivalent conditions holds:*

(i) *There exists an orthonormal basis for $\mathcal{V}$ consisting of eigenvectors of $\mathcal{T}$.*

(ii) *There exist a diagonal matrix $\Lambda \in \mathbb{F}^{n\times n}$ and an orthonormal basis for $\mathcal{V}$ such that if $\mathcal{B}$ is the associated basis operator, then*

$$\mathcal{T} = \mathcal{B}\mathcal{M}_\Lambda\mathcal{B}^*$$

(iii) *There exist a unitary operator $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ and a diagonal matrix $\Lambda \in \mathbb{F}^{n\times n}$ such that*

$$\mathcal{T} = \mathcal{B}\mathcal{M}_\Lambda\mathcal{B}^*.$$

A matrix $A \in \mathbb{F}^{n\times n}$ is ***UNITARILY DIAGONALIZABLE*** *if the multiplication operator $\mathcal{M}_A \in \mathbf{L}(\mathbb{F}^{n\times n})$ is unitarily diagonalizable. If $\mathbb{F} = \mathbb{R}$, a unitarily diagonalizable operator (matrix) is sometimes called* ***ORTHOGONALLY DIAGONALIZABLE****.*

A unitarily diagonalizable operator has perhaps the most "ideal" representation as a linear combination of rank-1 operators. We review two of our past results and then contrast them with what is now new. By Theorem 24.12, if $\mathcal{V}$ is a finite-dimensional vector space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ with $r := \operatorname{rank}(\mathcal{T}) \ge 1$, then there are independent lists $(\varphi_1, \ldots, \varphi_r)$ in $\mathcal{V}'$ and $(w_1, \ldots, w_r)$ in $\mathcal{T}(\mathcal{V})$ such that

$$\mathcal{T}v = \sum_{j=1}^{r} \varphi_j(v)w_j. \tag{40.1}$$

40.5 Problem (!). Go back to the proof of that theorem: what are the functionals φ_j, and why are they independent? [Hint: *we are overworking the symbol φ_j here.*]

This is very existential, and it does not say much about what φ_j and w_j are. With more hypotheses, we can be more precise. If $\mathcal{T}$ is diagonalizable and $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$ consisting of eigenvectors of $\mathcal{T}$ where $\mathcal{T}v_j = 0_{\mathcal{V}}$ for $j = r+1, \ldots, n$, then with $(\varphi_1, \ldots, \varphi_n)$ in $\mathcal{V}'$ as the associated dual basis, we have

$$\mathcal{T}v = \sum_{j=1}^{r} \lambda_j\varphi_j(v)v_j.$$

We established this in (37.1). This is more transparent than the otherwise unknown w_j in (40.1), but we still have to deal with the dual basis.

Here is what is new, and better: if $\mathcal{T}$ is unitarily diagonalizable and $(u_1, \ldots, u_n)$ is an orthonormal basis for $\mathcal{V}$ consisting of eigenvectors of $\mathcal{T}$, then since any $v \in \mathcal{V}$ expands as $v = \sum_{j=1}^{n} \langle v, u_j\rangle u_j$, and since $\mathcal{T}u_j = \lambda_j u_j$ for some $\lambda_j \in \mathbb{F}$, we have

$$\mathcal{T}v = \sum_{j=1}^{n} \lambda_j \langle v, u_j\rangle u_j.$$

In particular, if $\mathcal{T}v_j = 0_{\mathcal{V}}$ for $j = r+1, \ldots, n$, then this collapses to

$$\mathcal{T}v = \sum_{j=1}^{r} \lambda_j \langle v, u_j \rangle u_j.$$

This representation of $\mathcal{T}$ is as explicit as possible.

So, what operators have this nice property of unitary diagonalizability?

40.6 Problem (!). Here is a negative result, which we will revisit later with a concrete (and still negative) example: it is not immediately obvious that diagonalizable operators are unitarily diagonalizable. (They may not be!) Suppose that $\mathcal{V}$ is a finite-dimensional inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is diagonalizable, so there is a basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$ of eigenvectors of $\mathcal{T}$. Perform Gram–Schmidt on $(v_1, \ldots, v_n)$ to obtain an orthonormal basis $(u_1, \ldots, u_n)$ for $\mathcal{V}$ with $\operatorname{span}(v_1, \ldots, v_j) = \operatorname{span}(u_1, \ldots, u_j)$ for each j. Why are we not guaranteed that the u_j are also eigenvectors of $\mathcal{T}$?

There is a surprisingly simple characterization of unitarily diagonalizable operators: on a finite-dimensional real inner product space, they are the self-adjoint operators, and on a finite-dimensional complex inner product space, they are the *normal* operators. These results are the *spectral theorems*—so-called because their proofs and applications hinge on eigenvalues, and eigenvalues are part of the *spectrum* of a linear operator.

40.7 Definition. **(i)** *Let $\mathcal{V}$ be an inner product space. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* **NORMAL** *if $\mathcal{T}$ has an adjoint and if $\mathcal{T}^*\mathcal{T} = \mathcal{T}\mathcal{T}^*$.*

(ii) *A matrix $A \in \mathbb{F}^{n \times n}$ is* **NORMAL** *if $A^*A = AA^*$.*

The value of normality is that it is so easy to check: just compare $\mathcal{T}^*\mathcal{T}$ and $\mathcal{T}\mathcal{T}^*$. The spectral theorem then asserts that every normal operator (on a finite-dimensional complex space—and both the dimension and the field do matter here) is unitarily diagonalizable. No need to find eigenvalues, eigenvectors, or an orthonormal basis thereof.

40.8 Problem (!). Prove that self-adjoint, skew-adjoint, and unitary operators are normal. (In conjunction with Problem 40.6, we will shortly give examples of normal operators that are not self-adjoint, skew-adjoint, or unitary.)

Here is how normal operators arise in the context of unitary diagonalizability. We work backwards and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is unitarily diagonalizable with $\mathcal{T} = \mathcal{B}\mathcal{M}_\Lambda\mathcal{B}^*$, $\mathcal{B} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ unitary, and $\Lambda \in \mathbb{F}^{n \times n}$ diagonal. Since $\mathcal{T}^* = \mathcal{B}\mathcal{M}_\Lambda\mathcal{B}^*$ is so easy to compute, it might natural to try to compare $\mathcal{T}$ and $\mathcal{T}^*$ further. One way of comparing operators is to play them off each other: how different are $\mathcal{T}\mathcal{T}^*$ and $\mathcal{T}^*\mathcal{T}$? Not at all, if $\mathcal{T}$ is unitarily diagonalizable. Because Λ is diagonal, Λ and Λ^* commute ($\Lambda\Lambda^* = \Lambda^*\Lambda$), and since $\mathcal{T}^* = \mathcal{B}\mathcal{M}_{\Lambda^*}\mathcal{B}^*$, we can compute

$$\mathcal{T}\mathcal{T}^* = \mathcal{T}^*\mathcal{T}.$$

Thus every unitarily diagonalizable operator is normal. The actual computation here is quite simple, but thinking to do it in the first place takes some insight.

40.9 Problem (!). Let
$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$
Prove that $\mathcal{M}_A \in \mathbf{L}(\mathbb{R}^2)$ is normal but not (unitarily) diagonalizable. This bolsters our earlier remark that on a finite-dimensional real inner product space, an operator must be self-adjoint to be unitarily diagonalizable.

Content from *Linear Algebra* by Meckes & Meckes. However, check out the example on pp. 326–327 to see a unitary diagonalization of this matrix with complex numbers.

Content from *Linear Algebra* by Meckes & Meckes. The calculation that unitary diagonalizability implies normality (for matrices) appears on p. 324. The spectral theorems are proved throughout Section 5.4 using the singular value decomposition (SVD), which the book establishes earlier. We are going in the reverse order: spectral theorems, then SVD.

Here is the proof strategy for the spectral theorems. We will first improve Theorem 39.4 to show that every operator on a finite-dimensional complex inner product space $\mathcal{V}$ is upper-triangular with respect to an orthonormal basis for $\mathcal{V}$. Then we show that if a normal operator is upper-triangular with respect to an orthonormal basis, that operator is really diagonal with respect to that basis. This is the complex spectral theorem. After completing this program, we outline a proof of the real spectral theorem.

40.10 Theorem (Schur—operator version). *Let $\mathcal{V}$ be a finite-dimensional complex vector space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. There exists an orthonormal basis for $\mathcal{V}$ such that $\mathcal{T}$ is upper-triangular with respect to this basis.*

40.11 Problem (!). Let $\mathcal{V}$ be a finite-dimensional complex inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Prove Theorem 40.10 by using Theorem 39.4 to obtain a basis for $\mathcal{V}$ such that $\mathcal{T}$ is upper-triangular with respect to this basis. Apply the Gram–Schmidt procedure to this basis to produce an orthonormal basis for $\mathcal{V}$. Use the "span-preserving" property (28.9) of Gram–Schmidt to conclude that $\mathcal{T}$ is still upper-triangular with respect to this orthonormal basis.

40.12 Problem (+). Here is another proof of Schur's theorem that also uses induction but with a different flavor in the induction step. Let $\mathcal{V}$ be a complex finite-dimensional vector space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. We again induct on $n := \dim(\mathcal{V})$, and the $n = 1$ case and induction hypothesis are the same.

For the induction step, assume that $\dim(\mathcal{V}) = n + 1$. Since $\mathcal{V}$ is finite-dimensional, the adjoint $\mathcal{T}^* \in \mathbf{L}(\mathcal{V})$ exists; since $\mathcal{V}$ is complex and finite-dimensional, $\mathcal{T}^*$ has an eigenvalue $\lambda \in \mathbb{C}$ and an eigenvector $u \in \mathcal{V}$, which we assume is a unit vector: $\|u\| = 1$. Put $\mathcal{U} := \operatorname{span}(u)$.

(i) Use Problem 31.4 to show that $\dim(\mathcal{U}^\perp) = n$.

(ii) Use Problem 36.3 to show that $\mathcal{T}(\mathcal{U}^\perp) \subseteq \mathcal{U}^\perp$, so the restriction

$$\mathcal{T}\big|_{\mathcal{U}^\perp} : \mathcal{U}^\perp \to \mathcal{U}^\perp : u^\perp \mapsto u^\perp$$

is an operator in $\mathbf{L}(\mathcal{U}^\perp)$.

(iii) Use the induction hypothesis to show that $\mathcal{T}\big|_{\mathcal{U}^\perp}$ is upper-triangular with respect to some orthonormal basis $(u_1^\perp, \ldots, u_n^\perp)$ for $\mathcal{U}^\perp$.

(iv) Use Problem 31.4 to show that $(u_1^\perp, \ldots, u_n^\perp, u)$ is a basis for $\mathcal{V}$.

(v) Check that $\mathcal{T}$ is upper-triangular with respect to $(u_1^\perp, \ldots, u_n^\perp, u)$.

We state without proof a matrix version of Schur's theorem, which follows directly from the operator version.

40.13 Theorem (Schur—matrix version). *Let $A \in \mathbb{C}^{n \times n}$. There exist a unitary matrix $U \in \mathbb{C}^{n \times n}$ and an upper-triangular matrix $T \in \mathbb{C}^{n \times n}$ such that $A = UTU^*$.*

Content from *Linear Algebra* by Meckes & Meckes. Page 328 has an example of finding such a "Schur decomposition" of a matrix.

40.14 Problem ($\star$). Schur's theorem has a real matrix version, too. Suppose that all of the eigenvalues of $A \in \mathbb{R}^{n \times n}$ are real. Show that there exist an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ and an upper-triangular matrix $T \in \mathbb{R}^{n \times n}$ such that $A = UTU^\mathsf{T}$. [Hint: *use Problem 39.3 to triangularize $A = B\widetilde{T}B$ with B and $\widetilde{T}$ real and then adapt the argument from Problem 40.11.*]

Now we are ready to characterize unitary diagonalizability in terms of adjoints. This relies on combining Schur's theorem with the following preparatory results.

40.15 Problem. Let $\mathcal{V}$ be an inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ be normal. Show that $\|\mathcal{T}v\| = \|\mathcal{T}^*v\|$. [Hint: *compute $\|\mathcal{T}v\|^2$.*]

40.16 Lemma. *Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal basis $(u_1, \ldots, u_n)$. If $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is normal and upper-triangular with respect to this basis, then $\mathcal{T}$ is really diagonal with respect to this basis, thus unitarily diagonalizable.*

Proof. We induct on n.

1. *The base case* $n = 1$. There is really nothing to prove here, as if $\mathcal{V}$ is one-dimensional, then every operator on $\mathcal{V}$ is diagonal with respect to any basis of $\mathcal{V}$.

2. *The illustrative case* $n = 2$. The logic of induction allows us to proceed directly from the base case of $n = 1$ to the induction hypothesis and step. Nonetheless, it may be helpful to see the less trivial, but still small, case of $n = 2$ first before working more generally.

Suppose that (u_1, u_2) is an orthonormal basis for $\mathcal{V}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is normal and upper-triangular with respect to this basis. Then $\mathcal{T}u_1 \in \operatorname{span}(u_1)$, so $\mathcal{T}u_1 = \langle \mathcal{T}u_1, u_1\rangle u_1$, and $\|\mathcal{T}v\|^2 = \|\mathcal{T}^*v\|^2$ for all $v \in \mathcal{V}$. We want to find $\lambda_1, \lambda_2 \in \mathbb{F}$ such that

$$\mathcal{T}v = \lambda_1 \langle v, u_1\rangle u_1 + \lambda_2 \langle v, u_2\rangle u_2$$

for all $v \in \mathcal{V}$. We know

$$\begin{aligned}\mathcal{T}v = \mathcal{T}\big(\langle v, u_1\rangle u_1 + \langle v, u_2\rangle u_2\big) = \langle v, u_1\rangle \mathcal{T}u_1 + \langle v, u_2\rangle \mathcal{T}u_2 = \langle \mathcal{T}u_1, u_1\rangle \langle v, u_1\rangle u_1 \\ + \langle v, u_2\rangle \big(\langle \mathcal{T}u_2, u_1\rangle u_1 + \langle \mathcal{T}u_2u_2, u_2\rangle u_2\big).\end{aligned}$$

We will be done if we can show that $\langle \mathcal{T}u_2, u_1\rangle = 0$.

Here is how this works. We have $\|\mathcal{T}u_2\|^2 = \|\mathcal{T}^*u_2\|^2$ since $\mathcal{T}$ is normal, and, by orthonormality,

$$\|\mathcal{T}u_2\|^2 = \|\langle \mathcal{T}u_2, u_1\rangle u_1 + \langle \mathcal{T}u_2, u_2\rangle u_2\|^2 = |\langle \mathcal{T}u_2, u_1\rangle|^2 + |\langle \mathcal{T}u_2, u_2\rangle|^2$$

and

$$\begin{aligned}\|\mathcal{T}^*u_2\|^2 = \|\langle \mathcal{T}^*u_2, u_1\rangle u_1 + \langle \mathcal{T}^*u_2, u_2\rangle u_2\|^2 = |\langle \mathcal{T}^*u_2, u_1\rangle|^2 + |\langle \mathcal{T}^*u_2, u_2\rangle|^2 \\ = |\langle u_2, \mathcal{T}u_1\rangle|^2 + |\langle u_2, \mathcal{T}u_2\rangle|^2.\end{aligned}$$

Furthermore,

$$\langle u_2, \mathcal{T}u_1\rangle = \langle u_2, \langle \mathcal{T}u_1, u_1\rangle u_1\rangle = \langle u_1, \mathcal{T}u_1\rangle \langle u_2, u_1\rangle = 0$$

by (more) orthonormality.

Combining everything, we have

$$|\langle \mathcal{T}u_2, u_1\rangle|^2 + |\langle \mathcal{T}u_2, u_2\rangle|^2 = \|\mathcal{T}u_2\|^2 = \|\mathcal{T}^*u_2\|^2 = |\langle u_2, \mathcal{T}u_2\rangle|^2 = |\langle \mathcal{T}u_2, u_2\rangle|^2,$$

so $|\langle \mathcal{T}u_2, u_1\rangle| = 0$, as desired.

3. *The induction hypothesis and step.* Assume that for some $n \geq 1$, if $\dim(\mathcal{V}) = n$ and an operator on $\mathcal{V}$ is both normal and upper-triangular with respect to some orthonormal basis of $\mathcal{V}$, then the operator is diagonal with respect to that basis. Now suppose that $\dim(\mathcal{V}) = n+1$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is normal and upper-triangular with respect to the orthonormal basis $(u_1, \ldots, u_n, u_{n+1})$ of $\mathcal{V}$. Put $\mathcal{U}_n := \operatorname{span}(u_1, \ldots, u_n)$. Since $\mathcal{T}v_j \in \operatorname{span}(u_1, \ldots, u_j)$ for each j, we have $\mathcal{T}(\mathcal{U}_n) \subseteq \mathcal{U}_n$. Then $\mathcal{T}\big|_{\mathcal{U}_n}$ is normal and (still) upper-triangular with respect to the basis $(u_1, \ldots, u_n)$ for $\mathcal{U}_n$, thus diagonal with respect to this basis by the induction hypothesis. That is, for $j = 1, \ldots, n$, we have $\mathcal{T}\big|_{\mathcal{U}_n} u_j = \lambda_j u_j$ for some $\lambda_j \in \mathbb{F}$, and so $\mathcal{T} u_j = \lambda_j u_j$.

We want to show $\mathcal{T} u_{n+1} = \lambda_{n+1} u_{n+1}$ for some $\lambda_{n+1} \in \mathbb{F}$. This is equivalent to having $\mathcal{T} u_{n+1} \in \operatorname{span}(u_{n+1})$. We know $\mathcal{T} u_{n+1} = \sum_{j=1}^{n+1} \langle \mathcal{T} u_{n+1}, u_j \rangle u_j$, and we will show $\langle \mathcal{T} u_{n+1}, u_j \rangle = 0$ for $j = 1, \ldots, n$. We do this by calculating $\|\mathcal{T} u_{n+1}\|^2$ in two ways.

First,

$$\|\mathcal{T} u_{n+1}\|^2 = \sum_{j=1}^{n+1} |\langle \mathcal{T} u_{n+1}, u_j \rangle|^2 = \sum_{j=1}^{n} |\langle \mathcal{T} u_{n+1}, u_j \rangle|^2 + |\langle \mathcal{T} u_{n+1}, u_{n+1} \rangle|^2.$$

Second, because $\mathcal{T}$ is normal, we have

$$\begin{aligned} \|\mathcal{T} u_{n+1}\|^2 = \|\mathcal{T}^* u_{n+1}\|^2 = \sum_{j=1}^{n+1} |\langle \mathcal{T}^* u_{n+1}, u_j \rangle|^2 &= \sum_{j=1}^{n+1} |\langle u_{n+1}, \mathcal{T} u_j \rangle|^2 \\ &= \sum_{j=1}^{n} |\langle u_{n+1}, \mathcal{T} u_j \rangle|^2 + |\langle u_{n+1}, \mathcal{T} u_{n+1} \rangle|^2. \end{aligned}$$

For $1 \le j \le n$, we know

$$\langle u_{n+1}, \mathcal{T} u_j \rangle = \langle u_{n+1}, \lambda_j u_j \rangle = \overline{\lambda_j} \langle u_{n+1}, u_j \rangle = 0.$$

Thus

$$\sum_{j=1}^{n} |\langle \mathcal{T} u_{n+1}, u_j \rangle|^2 + |\langle \mathcal{T} u_{n+1}, u_{n+1} \rangle|^2 = \|\mathcal{T} u_{n+1}\|^2 = |\langle u_{n+1}, \mathcal{T} u_{n+1} \rangle|^2 = |\langle \mathcal{T} u_{n+1}, u_{n+1} \rangle|^2,$$

and so

$$\sum_{j=1}^{n} |\langle \mathcal{T} u_{n+1}, u_j \rangle|^2 = 0.$$

This is a sum of nonnegative terms, so each term is zero:

$$|\langle \mathcal{T} u_{n+1}, u_j \rangle|^2 = 0, \ j = 1, \ldots, n$$

thus

$$\langle \mathcal{T} u_{n+1}, u_j \rangle = 0, \ j = 1, \ldots, n,$$

as desired. ■

The following problem illustrates the matrix version of the preceding lemma in a simple case.

40.17 Problem (!). Let a, b, $c \in \mathbb{F}$ and

$$A := \begin{bmatrix} a & c \\ 0 & d \end{bmatrix}.$$

Suppose that A is normal. Prove that $c = 0$. [Hint: *compare the entries of* AA^* *and* A^*A.]

40.18 Lemma. *If* $A \in \mathbb{F}^{n \times n}$ *is both upper-triangular and normal, then* A *is diagonal.*

Proof. **1.** *The base case* $n = 1$. There is really nothing to prove here, as every 1×1 matrix is simultaneously upper-triangular, diagonal, and normal (although not necessarily Hermitian!).

2. *A reminder about the illustrative case* $n = 2$. Do Problem 40.17.

3. *The induction hypothesis and step.* Assume that for some $n \geq 1$, if $A \in \mathbb{F}^{n \times n}$ is both upper-triangular and normal, then A is diagonal. Suppose now that $A \in \mathbb{F}^{(n+1)\times(n+1)}$ is normal and upper-triangular, and write

$$A = \begin{bmatrix} \widetilde{A} & \mathbf{v} \\ \mathbf{0}_n^* & z \end{bmatrix}$$

for some $\mathbf{v} \in \mathbb{F}^n$ and $z \in \mathbb{F}$. To show that A is diagonal, it suffices to show that $\widetilde{A}$ is diagonal and $\mathbf{v} = \mathbf{0}_n$. We prove the former by showing that $\widetilde{A}$ is normal and applying the induction hypothesis; along the way, we obtain $\mathbf{v} = \mathbf{0}_n$.

We have

$$A^* = \begin{bmatrix} \widetilde{A}^* & \mathbf{0}_n \\ \mathbf{v}^* & \overline{z} \end{bmatrix}.$$

Then

$$AA^* = \begin{bmatrix} \widetilde{A} & \mathbf{v} \\ \mathbf{0}_n^* & z \end{bmatrix} \begin{bmatrix} \widetilde{A}^* & \mathbf{0}_n \\ \mathbf{v}^* & \overline{z} \end{bmatrix} = \begin{bmatrix} (\widetilde{A}\widetilde{A}^* + \mathbf{v}\mathbf{v}^*) & \overline{z}\mathbf{v} \\ z\mathbf{v}^* & z\overline{z} \end{bmatrix}$$

and

$$A^*A = \begin{bmatrix} \widetilde{A}^* & \mathbf{0}_n \\ \mathbf{v}^* & \overline{z} \end{bmatrix} \begin{bmatrix} \widetilde{A} & \mathbf{v} \\ \mathbf{0}_n^* & z \end{bmatrix} = \begin{bmatrix} \widetilde{A}^*\widetilde{A} & \widetilde{A}^*\mathbf{v} \\ \mathbf{v}^*\widetilde{A} & (\mathbf{v}^*\mathbf{v} + z\overline{z}) \end{bmatrix}.$$

Since $A^*A = AA^*$, we may equate the $(2,2)$-blocks to find

$$z\overline{z} = \mathbf{v}^*\mathbf{v} + z\overline{z}.$$

This gives $\mathbf{v}^*\mathbf{v} = 0$, thus $\mathbf{v} = \mathbf{0}_n$. With this we equate the $(1,1)$-blocks to find $\widetilde{A}^*\widetilde{A} = \widetilde{A}\widetilde{A}^*$. Thus $\widetilde{A}^*$ is both normal and upper-triangular, so by the induction hypothesis $\widetilde{A}$ is diagonal.

We have therefore shown that $\widetilde{A}$ is diagonal and $\mathbf{v} = \mathbf{0}_n$, as desired, so A is really diagonal. ■

40.19 Theorem (Complex spectral theorem—operator version). *Let* $\mathcal{V}$ *be a finite-dimensional complex inner product space. An operator* $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ *is unitarily diagonalizable if and only if* $\mathcal{T}$ *is normal.*

Proof. $(\Longrightarrow)$ This is the calculation that motivated our discussion of normal operators.

($\Longleftarrow$) By Schur's theorem, $\mathcal{T}$ is upper-triangular with respect to some orthonormal basis of $\mathcal{V}$. By Lemma 40.16, since $\mathcal{T}$ is normal and upper-triangular with respect to this basis, $\mathcal{T}$ is really diagonal with respect to this basis. Since the basis is orthonormal, $\mathcal{T}$ is unitarily diagonalizable. ■

Before finalizing the spectral theorems for real finite-dimensional inner product spaces, it will be helpful to have some unsurprising results about how unitary similarity preserves properties between operators and matrix representations.

40.20 Problem (!). Let $\mathcal{V}$ be a finite-dimensional inner product space. Suppose that $\mathcal{B}\colon \mathbb{F}^n \to \mathcal{V}$ is unitary and $A \in \mathbb{F}^{n \times n}$. Put $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^*$.

(i) Prove that $\mathcal{T}$ is unitary if and only if A is unitary.

(ii) Prove that $\mathcal{T}$ is self-adjoint if and only if A is self-adjoint.

(iii) Prove that $\mathcal{T}$ is normal if and only if A is normal.

40.21 Problem (!). Let $\mathcal{V}$ be a finite-dimensional complex inner product space and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ be self-adjoint. Prove that $\mathcal{T}$ is unitarily diagonalizable without using the spectral theorem. [Hint: *write* $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^*$, *where* A *is self-adjoint by Problem 40.20. Show that if* A *is self-adjoint and upper-triangular, then* A *is diagonal.*]

40.22 Theorem (Complex spectral theorem—matrix version). *A matrix* $A \in \mathbb{C}^{n \times n}$ *is unitarily diagonalizable if and only if* A *is normal.*

Proof. ($\Longrightarrow$) This is the calculation that motivated our discussion of normal operators.

($\Longleftarrow$) By Schur's theorem for matrices, we can factor $A = UTU^*$ for some unitary $U \in \mathbb{C}^{n \times n}$ and upper-triangular $T \in \mathbb{C}^{n \times n}$. By Problem 40.20, since A is normal, so is T. By Lemma 40.18, since T is normal and upper-triangular, T is really diagonal. ■

Day 41: Friday, April 24.

A version of the spectral theorem remains true for finite-dimensional real inner product spaces, but its proof requires some additional preparation. This is an indication of an occasional theme in our work that results on real vector spaces can be more challenging to obtain than results on complex vector spaces. While real vector spaces are somewhat "simpler" than complex ones (because the underlying field is "simpler"), they also lack some of the structure that complex vector spaces possess (e.g., $i^2 = -1$); this is morally similar to why some results in complex analysis are "nicer" than in real analysis.

We have previously (Theorem 36.1) compared the adjoint of an operator to the conjugate of a complex number. Recall that for $z \in \mathbb{C}$, we have $z \in \mathbb{R}$ if and only if $\overline{z} \in \mathbb{R}$. We might wonder, then, how real numbers show up when dealing with a self-adjoint operator. First, the eigenvalues of a self-adjoint operator (if they exist) are real.

41.1 Theorem. *Let $\mathcal{V}$ be an inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ be self-adjoint. If $\mathcal{T}$ has an eigenvalue $\lambda \in \mathbb{C}$, then $\lambda \in \mathbb{R}$.*

Proof. The proof is a classical trick well worth knowing. Let $v \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$ be an eigenvector for $\mathcal{T}$ corresponding to λ, so $\mathcal{T}v = \lambda v$. We want to show $\lambda = \overline{\lambda}$. We know that conjugates appear when moving scalar coefficients around in an inner product, and we know that $\mathcal{T}^*$ will show up when we involve $\mathcal{T}$ in the inner product. Since we want $\lambda - \overline{\lambda} = 0$, and we want λ to interact with the inner product, and since λ interacts best with v, and since $\langle v, v \rangle \neq 0$, a good idea is to try to show

$$(\lambda - \overline{\lambda}) \langle v, v \rangle = 0.$$

We compute this:

$$\begin{aligned}
(\lambda - \overline{\lambda}) \langle v, v \rangle &= \lambda \langle v, v \rangle - \overline{\lambda} \langle v, v \rangle \\
&= \langle \lambda v, v \rangle - \langle v, \lambda v \rangle \text{ by properties of inner products} \\
&= \langle \mathcal{T}v, v \rangle - \langle v, \mathcal{T}v \rangle \text{ since } \mathcal{T}v = \lambda v \\
&= \langle \mathcal{T}v, v \rangle - \langle \mathcal{T}^*v, v \rangle \\
&= \langle \mathcal{T}v, v \rangle - \langle \mathcal{T}v, v \rangle \text{ since } \mathcal{T} \text{ is self-adjoint} \\
&= 0.
\end{aligned}$$

■

Next, at least in finite dimensions, every self-adjoint operator has an eigenvalue. We have long known this to be true on a complex finite-dimensional vector space, so now we should pay attention to the real case. We do this by exploiting the versatility of matrices.

Let $\mathcal{V}$ be a finite-dimensional *real* vector space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Then $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}$ for some invertible operator $\mathcal{B} \in \mathbf{L}(\mathbb{R}^n, \mathcal{V})$ and matrix $A \in \mathbb{R}^{n \times n}$. In the factorization $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}$, we view $\mathcal{M}_A$ as an operator in $\mathbf{L}(\mathbb{R}^n)$, but we could also choose to view it as an operator in $\mathbf{L}(\mathbb{C}^n)$. In the latter case, $\mathcal{M}_A$ is guaranteed to have an eigenvalue: there exist $\lambda \in \mathbb{C}$ and $\mathbf{v} \in \mathbb{C}^n \setminus \{\mathbf{0}_n\}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. If, in fact $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$, then $\mathcal{B}\mathbf{v} \in \mathcal{V}$ is defined (and nonzero), and we can compute $\mathcal{T}(\mathcal{B}\mathbf{v}) = \lambda(\mathcal{B}\mathbf{v})$, thus $\mathcal{B}\mathbf{v}$ is an eigenvector of $\mathcal{T}$ corresponding to λ. However, if $\mathbf{v} \in \mathbb{C}^n \setminus \mathbb{R}^n$, none of this makes sense. Nonetheless, there is a situation in which we can use matrix representations to guarantee that an operator on a finite-dimensional real inner product space has an eigenvalue. Unsurprisingly, given the current context, this uses self-adjointness.

41.2 Lemma. *Let $\mathcal{V}$ be a finite-dimensional real inner product space and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ be self-adjoint. Then $\mathcal{T}$ has an eigenvalue (necessarily in $\mathbb{R}$).*

Proof. Take an orthonormal basis $(u_1, \ldots, u_n)$ for $\mathcal{V}$ and let $\mathcal{B} \colon \mathbb{F}^n \to \mathcal{V}$ be the associated basis operator, so $\mathcal{B}^* = \mathcal{B}$. Write $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^*$ for some $A \in \mathbb{R}^{n \times n}$. Here A is just the matrix representation of $\mathcal{T}$ with respect to this basis $(u_1, \ldots, u_n)$, and we are not claiming that the factorization $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^{-1}$ is a unitary diagonalization, or triangularization, of $\mathcal{T}$.

The matrix A has an eigenvalue $\lambda \in \mathbb{C}$, but because $\mathcal{T}$ is self-adjoint, so is A by Problem 40.20, and therefore $\lambda \in \mathbb{R}$. Let $\mathbf{v} \in \mathbb{C}^n$ be an eigenvector of A corresponding to λ. Since A

and λ are both real, we might expect that $\mathbf{v}$ should be real, and this is mostly true: $\mathbf{v}$ can be *taken to be* real, as we subsequently show. Then $\mathcal{B}\mathbf{v}$ is defined and the discussion above shows that $\mathcal{B}\mathbf{v}$ is an eigenvector of $\mathcal{T}$ corresponding to λ. ∎

41.3 Lemma. *Let $A \in \mathbb{R}^{n\times n}$ have a real eigenvalue $\lambda \in \mathbb{R}$. Then there is a real eigenvector corresponding to λ: there exists $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}_n\}$ such that $A\mathbf{v} = \lambda\mathbf{v}$.*

Proof. Write $\mathbf{v} = \mathbf{x} + i\mathbf{y}$ with $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$. Then

$$A\mathbf{x} + iA\mathbf{y} = A(\mathbf{x} + i\mathbf{y}) = A\mathbf{v} = \lambda\mathbf{v} = \lambda\mathbf{x} + i\lambda\mathbf{y}.$$

Since $A \in \mathbb{R}^{n\times n}$ and $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$, we have $A\mathbf{x}$, $A\mathbf{y} \in \mathbb{R}^n$. And since $\lambda \in \mathbb{R}$, we have $\lambda\mathbf{x}$, $\lambda\mathbf{y} \in \mathbb{R}^n$. Equating the real and imaginary parts of $A\mathbf{x} + iA\mathbf{y} = \lambda\mathbf{x} + i\lambda\mathbf{y}$ componentwise, we have $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \lambda\mathbf{y}$. Since $\mathbf{v} \neq \mathbf{0}_n$, at least one of $\mathbf{x}$ or $\mathbf{y}$ is nonzero and thus an eigenvector of A corresponding to λ. ∎

Now we can prove the real spectral theorem.

41.4 Theorem (Real spectral theorem—operator version). *Let $\mathcal{V}$ be a finite-dimensional real inner product space. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is unitarily diagonalizable if and only if $\mathcal{T}$ is self-adjoint.*

Proof. ($\Longrightarrow$) This is the calculation that motivated our discussion of normal operators.

($\Longleftarrow$) We prove this by induction on $n := \dim(\mathcal{V})$. For $n = 1$, there is, as before, nothing to prove, since every operator is diagonal with respect to any basis of a one-dimensional space.

Suppose that for some $n \geq 1$, if $\mathcal{V}$ is a real inner product space with $\dim(\mathcal{V}) = n$, then every self-adjoint operator on $\mathcal{V}$ is unitarily diagonalizable. Now let $\dim(\mathcal{V}) = n + 1$ and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is self-adjoint. By Lemma 41.2, $\mathcal{T}$ has an eigenvalue $\lambda \in \mathbb{R}$, and so we may find an eigenvector $u_1 \in \mathcal{V}$ for $\mathcal{T}$ corresponding to λ, and we assume $\|u_1\| = 1$. We will take u_1 to be the first vector in an orthonormal basis for $\mathcal{V}$ such that $\mathcal{T}$ is diagonal in this basis.

We can make an n-dimensional subspace of $\mathcal{V}$, to which the induction hypothesis might be applicable, show up by putting $\mathcal{U} = \operatorname{span}(u_1)$, so $\dim(\mathcal{U}^\perp) = n$ by Problem 31.4. Since u_1 is an eigenvector of $\mathcal{T}$, we have $\mathcal{T}u \in \mathcal{U}$ for all $u \in \mathcal{U}$, so $\mathcal{T}(\mathcal{U}) \subseteq \mathcal{U}$. Then $\mathcal{T}^*(\mathcal{U}^\perp) \subseteq \mathcal{U}^\perp$ by Problem 36.3. But $\mathcal{T}$ is self-adjoint, so $\mathcal{T}(\mathcal{U}^\perp) = \mathcal{T}^*(\mathcal{U}^\perp)$. Then the restriction

$$\mathcal{T}\big|_{\mathcal{U}^\perp} : \mathcal{U}^\perp \to \mathcal{U}^\perp$$

is linear.

We claim that $(\mathcal{T}\big|_{\mathcal{U}^\perp})^* = \mathcal{T}\big|_{\mathcal{U}^\perp}$ (thinking about what this equality says and parsing its notation is, by itself, a good exercise). Assuming this, the induction hypothesis gives an orthonormal basis $(u_2^\perp, \ldots, u_{n+1}^\perp)$ for $\mathcal{U}^\perp$ of eigenvectors of $\mathcal{T}$, so $(u_1, u_2^\perp, \ldots, u_{n+1}^\perp)$ is an orthonormal basis for $\mathcal{V}$ of eigenvectors of $\mathcal{T}$. ∎

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 5.19 on p. 321. Read the example on p. 322 to see a diagonalizable matrix that is not unitarily diagonalizable.

41.5 Problem (!).

41.6 Problem (!). This problem uses the notation and hypotheses of the proof of the real spectral theorem.

(i) Prove the claim that $\mathcal{T}\big|_{\mathcal{U}^\perp}$ is self-adjoint.

(ii) Reread the proof and explain why it works to prove that a self-adjoint operator on a finite-dimensional *complex* inner product space is unitarily diagonalizable.

41.7 Theorem (Real spectral theorem—matrix version). *A matrix $A \in \mathbb{R}^{n \times n}$ is orthogonally diagonalizable if and only if A is symmetric.*

41.8 Problem ($\star$). Prove this by applying the real Schur theorem for matrices from Problem 40.14 to produce an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ and an upper-triangular matrix $T \in \mathbb{R}^{n \times n}$ such that $A = UTU^\mathsf{T}$. Use Problem 40.20 to conclude that T is also self-adjoint, and argue that a self-adjoint upper-triangular matrix is really diagonal.

41.9 Problem ($\star$). We are working with three closely related (but distinct) types of operators (matrices): self-adjoint, unitary, and normal. We know that self-adjoint and unitary matrices are normal (Problem 40.8), but the other relations among the three classes are murkier. Give examples of each of the following kinds of matrices. It suffices to work with 2×2 matrices, and most examples can be diagonal.

(i) A normal matrix that is neither self-adjoint nor unitary.

(ii) A self-adjoint matrix that is not unitary.

(iii) A unitary matrix that is not self-adjoint.

(iv) A diagonalizable matrix that is not unitarily diagonalizable (equivalently, that is not normal). [Hint: *why will a diagonal matrix not work here?*]

We have interpreted the spectral theorems primarily as factorization results: under easily checked conditions for an operator $\mathcal{T}$ on a finite-dimensional space $\mathcal{V}$ (normality when $\mathcal{V}$ is complex, self-adjointness when $\mathcal{V}$ is real), $\mathcal{T}$ factors as $\mathcal{T} = \mathcal{B}\mathcal{M}_\Lambda\mathcal{B}^*$ for a diagonal matrix Λ and a unitary operator $\mathcal{B}$ from Euclidean space to $\mathcal{V}$. We can obtain another representation of a unitarily diagonalizable operator involving orthogonal projections. Suppose for simplicity that $\dim(\mathcal{V}) = 5$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is unitarily diagonalizable with two distinct eigenvalues λ_1 and λ_2. Suppose also that in an orthonormal eigenbasis $(u_1, \ldots, u_5)$ for $\mathcal{V}$, the vectors u_1,

u_2, and u_3 are the eigenvectors for λ_1 and u_4 and u_5 are the eigenvectors for λ_2. By Problem 37.11, (u_1, u_2, u_3) is an orthonormal basis for the eigenspace $\mathcal{E}_{\lambda_1}(\mathcal{T})$ of $\mathcal{T}$ corresponding to λ_1 and (u_4, u_5) is an orthonormal basis for $\mathcal{E}_{\lambda_2}(\mathcal{T})$. Then we have the straightforward representation formula

$$\mathcal{T}v = \lambda_1\big(\underbrace{\langle v, u_1\rangle\, u_1 + \langle v, u_2\rangle\, u_2 + \langle v, u_3\rangle\, u_3}_{\mathcal{P}_1 v}\big) + \lambda_2\big(\underbrace{\langle v, u_4\rangle\, u_4 + \langle v, u_5\rangle\, u_5}_{\mathcal{P}_2 v}\big),$$

where $\mathcal{P}_j := \mathcal{P}_{\mathcal{E}_{\lambda_j(\mathcal{T})}}$ is the orthogonal projection onto $\mathcal{E}_{\lambda_j(\mathcal{T})}$, and so

$$\mathcal{T} = \lambda_1\mathcal{P}_1 + \lambda_2\mathcal{P}_2.$$

Additionally, $\mathcal{P}_1$ and $\mathcal{P}_2$ are orthogonal in the sense that

$$\mathcal{P}_1\mathcal{P}_2 = \mathcal{P}_2\mathcal{P}_1 = 0_{\mathcal{V}\to\mathcal{V}}.$$

Finally, any $v \in \mathcal{V}$ has the form

$$v = \sum_{j=1}^{5} \langle v, u_j\rangle\, u_j = \left(\sum_{j=1}^{3} \langle v, u_j\rangle\, u_j\right) + \left(\sum_{j=4}^{5} \langle v, u_j\rangle\, u_j\right) = \mathcal{P}_1 v + \mathcal{P}_2 v,$$

and so

$$\mathcal{I}_{\mathcal{V}} = \mathcal{P}_1 + \mathcal{P}_2$$

With slightly more patient handling of the notation, the work above generalizes to establish the following.

41.10 Theorem (Spectral theorem—spectral decomposition/resolution version). *Suppose that $\mathcal{V}$ is a finite-dimensional inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is unitarily diagonalizable. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of $\mathcal{T}$, and let $\mathcal{P}_j = \mathcal{P}_{\mathcal{E}_{\lambda_j(\mathcal{T})}}$ be the orthogonal projection onto the eigenspace $\mathcal{E}_{\lambda_j(\mathcal{T})}$. Then*

$$\mathcal{T} = \sum_{j=1}^{m} \lambda_j\mathcal{P}_j \quad \text{and} \quad \mathcal{I}_{\mathcal{V}} = \sum_{j=1}^{m} \mathcal{P}_j.$$

The expression above for $\mathcal{T}$ is its ***SPECTRAL DECOMPOSITION*** *and the expression above for $\mathcal{I}_{\mathcal{V}}$ is its* ***SPECTRAL RESOLUTION*** *induced by $\mathcal{T}$.*

41.11 Problem ($\star$). Conversely, suppose that $\mathcal{V}$ is a finite-dimensional inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is a linear combination of mutually orthogonal nonzero projections that sum to the identity:

$$\mathcal{T} = \sum_{j=1}^{m} \lambda_j\mathcal{P}_j, \qquad \mathcal{P}_j\mathcal{P}_k = \begin{cases} \mathcal{P}_j, & j = k \\ 0_{\mathcal{V}\to\mathcal{V}}, & j \neq k, \end{cases} \qquad \mathcal{P}_j \neq 0_{\mathcal{V}\to\mathcal{V}}, \qquad \sum_{j=1}^{m} \mathcal{P}_j = \mathcal{I}_{\mathcal{V}}.$$

Show that the eigenvalues of $\mathcal{T}$ are $\lambda_1, \ldots, \lambda_m$. [Hint: *to show that any eigenvalue λ of $\mathcal{T}$*

must be one of these λ_k, compute $\lambda v = \mathcal{T} v$ and $\lambda v = \lambda \mathcal{I}_{\mathcal{V}} v$, and use the formulas for $\mathcal{T}$ and $\mathcal{I}_{\mathcal{V}}$ in terms of the projections above. Equate the results and apply $\mathcal{P}_k$ to both sides for k fixed.]

Day 42: Monday, April 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."

Singular value decomposition (SVD) of an operator (matrix)

When possible, unitary diagonalization provides a transparent and flexible factorization of an operator. What happens when an operator is not unitarily diagonalizable? If $\mathcal{V}$ is a finite-dimensional inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$, then we can always pick an orthonormal basis for $\mathcal{V}$, let $\mathcal{B} \colon \mathbb{F}^n \to \mathcal{V}$ be the associated basis operator, and factor $\mathcal{T} = \mathcal{B}\mathcal{M}_A\mathcal{B}^*$ for some $A \in \mathbb{F}^{n \times n}$. However, this gives no good control over the matrix representation A. If $\mathcal{V}$ is complex, then the orthonormal basis $(u_1, \ldots, u_n)$ can be chosen so that A is upper-triangular (A is not guaranteed to be upper-triangular for an *arbitrary* orthonormal basis). Nonetheless, the "pointwise" expansion of $\mathcal{T}$ in this basis is

$$\mathcal{T} v = \sum_{j=1}^{n} \langle v, u_j \rangle \mathcal{T} u_j = \sum_{j=1}^{n} \langle v, u_j \rangle \sum_{k=1}^{n} \langle \mathcal{T} u_j, u_k \rangle u_k,$$

which is not nearly as transparent as the expansion

$$\mathcal{T} v = \sum_{j=1}^{n} \lambda_j \langle v, u_j \rangle u_j$$

when $\mathcal{T}$ is diagonal in this basis.

And what about an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, where $\mathcal{V}$ and $\mathcal{W}$ are *different* inner product spaces? Is there any hope for a slick "almost diagonal" representation of $\mathcal{T}$ with the right choice of *bases* for $\mathcal{V}$ and $\mathcal{W}$? The answer is yes, and this points to a new approach for factoring nondiagonalizable operators whose codomain is the same as their domain: choose different bases for domain and codomain.

It turns out that if $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional inner product spaces with $\dim(\mathcal{V}) = n$ and $\dim(\mathcal{W}) = m$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has rank $r \geq 1$, then there exist orthonormal bases $(v_1, \ldots, v_n)$ for $\mathcal{V}$ and $(u_1, \ldots, u_m)$ for $\mathcal{W}$ (the use of letters u_j, not w_j, here is traditional) and positive real numbers $\sigma_1 \geq \cdots \geq \sigma_r > 0$ such that

$$\mathcal{T} v = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j. \tag{42.1}$$

We can obtain a factored form for $\mathcal{T}$ out of this expansion by taking $\mathcal{B}_{\mathcal{V}} \colon \mathbb{F}^n \to \mathcal{V}$ and $\mathcal{B}_{\mathcal{W}} \colon \mathbb{F}^m \to \mathcal{W}$ to be the associated basis operators for the bases $(v_1, \ldots, v_n)$ for $\mathcal{V}$ and

$(u_1, \ldots, u_m)$ for $\mathcal{W}$ and defining $\Sigma \in \mathbb{R}^{m \times n}$ by

$$\Sigma_{ij} = \begin{cases} 0, \ i \neq j \text{ and } i = j > r \\ \sigma_j, \ 1 \le i = j \le r. \end{cases} \tag{42.2}$$

Then (42.1) becomes

$$\mathcal{T} = \mathcal{B}_{\mathcal{W}} \mathcal{M}_{\Sigma} \mathcal{B}_{\mathcal{V}}^*.$$

42.1 Problem (!). Check that.

For example, if $\mathcal{T} \in \mathbf{L}(\mathbb{F}^4, \mathbb{F}^3)$ has rank 2, then we expect

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

while if $\mathcal{T} \in \mathbf{L}(\mathbb{F}^3, \mathbb{F}^4)$ has rank 2, then we expect

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Such a factorization of $\mathcal{T}$ is called a **SINGULAR VALUE DECOMPOSITION (SVD)**. We will see that the basis vectors $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$ are not unique, but the **SINGULAR VALUES** $\sigma_1, \ldots, \sigma_r$ are. We will also see that even if $\mathcal{V} = \mathcal{W}$ and $\mathcal{T}$ is diagonalizable, the singular values need not be the eigenvalues of $\mathcal{T}$; in some sense, the SVD has simpler coefficients than a diagonal factorization and pushes complexities further into the basis vectors, not the coefficients. We take the approach of proving the existence of the SVD via the spectral theorem, although one could start with different proofs of the SVD and obtain the spectral theorem as a consequence of that.

Content from *Linear Algebra* by Meckes & Meckes. This is the book's approach. The SVD is used to prove that every self-adjoint operator has an eigenvalue; this is Lemma 5.18 on p. 321. Then the spectral theorem for self-adjoint operators is proved in Theorem 5.19, which has the same approach as our proof of Theorem 41.4. After some other auxiliary results (Lemmas 5.21 and 5.22 on p. 325) are developed, the spectral theorem for normal operators is proved in Theorem 5.23 on p. 326; neither of these auxiliary results nor Theorem 5.23 require the SVD. Read the proof of Lemma 5.21, do Quick Exercise #12 on p. 326, and then read the proof of Theorem 5.23 to see this strategy of proving the spectral theorem effectively without Schur's theorem. Indeed, here Schur's theorem appears as a consequence of the spectral theorem and Gram–Schmidt; see Corollary 5.24 on p. 327.

Before giving the proof of the SVD, we assume the existence of the SVD and deduce from that how we should be able to obtain the vectors v_j and w_j and the singular values.

So, assume that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional inner product spaces with $n = \dim(\mathcal{V})$ and $m = \dim(\mathcal{W})$ and that $\mathcal{T}$ factors as

$$\mathcal{T} = \mathcal{B}_{\mathcal{W}} \mathcal{M}_{\Sigma} \mathcal{B}_{\mathcal{V}}^*,$$

where $\mathcal{B}_{\mathcal{V}} \in \mathbf{L}(\mathbb{F}^n, \mathcal{V})$ and $\mathcal{B}_{\mathcal{W}} \in \mathbf{L}(\mathbb{F}^m, \mathcal{W})$ are unitary and $\Sigma \in \mathbb{R}^{m \times n}$ is "almost diagonal" in the sense that (42.2) holds. Inspired, perhaps, by how we developed the condition of normality as a consequence of unitary diagonalizability, we compute

$$\mathcal{T}^*\mathcal{T} = \mathcal{B}_{\mathcal{V}} \mathcal{M}_{\Sigma^*\Sigma} \mathcal{B}_{\mathcal{V}}^*$$

to see what we can learn.

42.2 Problem (!). Show that $\Sigma^*\Sigma \in \mathbb{F}^{n \times n}$ is diagonal with

$$\Sigma^*\Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0).$$

If $r = n$, just eliminate the 0 entries.

So, $\mathcal{T}^*\mathcal{T}$ is unitarily diagonalizable, the list $(v_1, \ldots, v_n)$ from the SVD forms an orthonormal basis for $\mathcal{V}$ consisting of eigenvectors of $\mathcal{T}^*\mathcal{T}$, and the eigenvalues of $\mathcal{T}^*\mathcal{T}$ are nonnegative with

$$\mathcal{T}^*\mathcal{T}v = \sum_{j=1}^{r} \sigma_j^2 \langle v, v_j \rangle_{\mathcal{V}} v_j. \tag{42.3}$$

None of this should really be surprising. First, the eigenvalues of $\mathcal{T}^*\mathcal{T}$ are necessarily nonnegative.

42.3 Problem ($\star$). Prove that if $\mathcal{V}$ and $\mathcal{W}$ are inner product spaces (not necessarily finite-dimensional) and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has an adjoint, then any eigenvalue of $\mathcal{T}^*\mathcal{T}$ is nonnegative. [Hint: *suppose that* $\mathcal{T}^*\mathcal{T}v = \lambda v$ *and compute* $\langle \mathcal{T}^*\mathcal{T}v, v \rangle_{\mathcal{V}}$.]

Second, we should expect that the eigenexpansion for $\mathcal{T}^*\mathcal{T}$ in (42.3) contains only r terms. Suppose that we diagonalize

$$\mathcal{T}^*\mathcal{T}v = \sum_{j=1}^{n} \lambda_j \langle v, v_j \rangle_{\mathcal{V}} v_j.$$

Since $\mathcal{T}^*\mathcal{T} \neq 0_{\mathcal{V} \to \mathcal{V}}$, at least one λ_j is positive; without loss of generality, order the basis $(v_1, \ldots, v_n)$ so that $\lambda_j > 0$ for $j = 1, \ldots, p$ with $p \geq 1$ and $\lambda_j = 0$ for $j \geq p + 1$. Then

$$\mathcal{T}^*\mathcal{T}v = \sum_{j=1}^{p} \lambda_j \langle v, v_j \rangle_{\mathcal{V}} v_j.$$

Problem 37.12 tells us that $\operatorname{rank}(\mathcal{T}^*\mathcal{T}) = p$. But by Lemma 36.7, we have $\ker(\mathcal{T}) = \ker(\mathcal{T}^*\mathcal{T})$, and so (since we are in a finite-dimensional setting)

$$r = \operatorname{rank}(\mathcal{T}) = \dim(\mathcal{V}) - \dim(\ker(\mathcal{T})) = \dim(\mathcal{V}) - \dim(\ker(\mathcal{T}^*\mathcal{T})) = \operatorname{rank}(\mathcal{T}^*\mathcal{T}) = p.$$

This largely exhausts what we can learn about the basis $(v_1, \ldots, v_n)$ and the singular values σ_j without considering the other basis $(u_1, \ldots, u_m)$. Still working backward from the SVD's presumed existence, we can figure out *some* of the vectors $(u_1, \ldots, u_m)$ by computing, for $1 \le k \le r$,

$$\mathcal{T}v_k = \sum_{j=1}^{r} \sigma_j \langle v_j, v_j \rangle_{\mathcal{V}} u_j = \sigma_k u_k,$$

thus

$$u_k = \frac{\mathcal{T}v_k}{\sigma_k}_{\mathcal{W}}, \ 1 \le k \le r.$$

We should check that defining $(u_1, \ldots, u_r)$ in this way does give an orthonormal list, so we compute

$$\langle u_j, u_k \rangle_{\mathcal{W}} = \frac{1}{\sigma_j \sigma_k} \langle \mathcal{T}v_j, \mathcal{T}v_k \rangle_{\mathcal{W}} = \frac{1}{\sigma_j \sigma_k} \langle \mathcal{T}^*\mathcal{T}v_j, v_k \rangle_{\mathcal{V}} = \frac{1}{\sigma_j \sigma_k} \langle \lambda_j v_j, v_k \rangle_{\mathcal{V}} = \frac{\lambda_j}{\sigma_j \sigma_k} \langle v_j, v_k \rangle_{\mathcal{V}}. \tag{42.4}$$

For the $j = k$ case, we used the definition $\sigma_j^2 = \lambda_j$ and the orthonormality $\langle v_j, v_j \rangle = 1$ to obtain

$$\frac{\lambda_j}{\sigma_j^2} \langle v_j, v_j \rangle_{\mathcal{V}} = 1.$$

For $j \ne k$, we use the orthogonality $\langle v_j, v_k \rangle = 0$ to obtain

$$\frac{\lambda_j}{\sigma_j \sigma_k} \langle v_j, v_k \rangle_{\mathcal{V}} = 0.$$

We might then try to extend $(u_1, \ldots, u_r)$ to an orthonormal basis for $\mathcal{W}$ and hope that somehow everything will work out.

It does, and at this point we have done at least half of the work of the proof, going backwards, so now we state and prove the theorem.

42.4 Theorem (Singular value decomposition). *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional inner product spaces and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ with $r := \operatorname{rank}(\mathcal{T}) \ge 1$.*

(i) *There exist orthonormal bases $(v_1, \ldots, v_n)$ for $\mathcal{V}$ and $(u_1, \ldots, u_m)$ for $\mathcal{W}$ and unique numbers $\sigma_1, \ldots, \sigma_r \in \mathbb{R}$, called the* ***SINGULAR VALUES*** *of $\mathcal{T}$, such that*

$$\mathcal{T}v = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle u_j, \tag{42.5}$$

where $\sigma_1 \ge \cdots \ge \sigma_r > 0$. In particular, the singular values of $\mathcal{T}$ are the positive eigenvalues of $\mathcal{T}^*\mathcal{T}$.

(ii) *Equivalently, if $\mathcal{B}_{\mathcal{V}} \colon \mathbb{F}^n \to \mathcal{V}$ and $\mathcal{B}_{\mathcal{W}} \colon \mathbb{F}^m \to \mathcal{W}$ are the basis operators for $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_m)$, respectively, and $\Sigma \in \mathbb{R}^{m \times n}$ satisfies*

$$\Sigma_{ij} = \begin{cases} \sigma_j, \ 1 \le i = j \le r \\ 0, \ i \ne j \text{ or } i = j > r, \end{cases}$$

then

$$\mathcal{T} = \mathcal{B}_{\mathcal{W}} \mathcal{M}_{\Sigma} \mathcal{B}_{\mathcal{V}}.$$

(iii) *The four fundamental subspaces for $\mathcal{T}$ have the following bases:*

- $(v_{r+1}, \ldots, v_n)$ *is an orthonormal basis for* $\ker(\mathcal{T})$.
- $(u_1, \ldots, u_r)$ *is an orthonormal basis for* $\mathcal{T}(\mathcal{V})$.
- $(u_{r+1}, \ldots, u_m)$ *is an orthonormal basis for* $\ker(\mathcal{T}^*)$.
- $(v_1, \ldots, v_r)$ *is an orthonormal basis for* $\mathcal{T}^*(\mathcal{W})$.

Proof. The operator $\mathcal{T}^*\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is self-adjoint, thus unitarily diagonalizable by the spectral theorem. Let $(v_1, \ldots, v_n)$ be an orthonormal basis for $\mathcal{V}$ such that $\mathcal{T}^*\mathcal{T}$ is diagonal in this basis:

$$\mathcal{T}^*\mathcal{T}v = \sum_{j=1}^{n} \lambda_j \langle v, v_j \rangle_{\mathcal{V}} v_j$$

for some $\lambda_j \in \mathbb{R}$. By Problem 42.3, $\lambda_j \geq 0$ for all j. Since $\operatorname{rank}(\mathcal{T}^*\mathcal{T}) = \operatorname{rank}(\mathcal{T}) = r \geq 1$ by Problem 37.12, at least one λ_j is positive, as otherwise $\mathcal{T}^*\mathcal{T} = 0_{\mathcal{V} \to \mathcal{V}}$. Reordering the vectors in this basis if needed, we may write

$$\mathcal{T}^*\mathcal{T}v = \sum_{j=1}^{p} \lambda_j \langle v, v_j \rangle_{\mathcal{V}} v_j,$$

where $\lambda_1 \geq \cdots \geq \lambda_p > 0$ with $p \leq n$. By Problem 36.7, we obtain $r = \operatorname{rank}(\mathcal{T}^*\mathcal{T}) = p$.

Let $\sigma_j := \sqrt{\lambda_j}$, so $\sigma_1 \geq \cdots \geq \sigma_r > 0$. For $1 \leq k \leq r$, put

$$u_k := \frac{\mathcal{T}v_k}{\sigma_k}.$$

We check that the list $(u_1, \ldots, u_r)$ is orthonormal as in (42.4).

We also have $u_k \in \mathcal{T}(\mathcal{V})$ for each k, so, by orthonormality, the list $(u_1, \ldots, u_r)$ is an independent list in $\mathcal{T}(\mathcal{V})$, and it has length $r = \operatorname{rank}(\mathcal{T}) = \dim(\mathcal{T}(\mathcal{V}))$. Hence this list $(u_1, \ldots, u_r)$ is a basis for $\mathcal{T}(\mathcal{V})$. Now extend the list $(u_1, \ldots, u_r)$ to an orthonormal basis $(u_1, \ldots, u_m)$ for $\mathcal{W}$. In particular, by part (ii) of Problem 31.4, the list $(u_{r+1}, \ldots, u_m)$ is an orthonormal basis for

$$\operatorname{span}(u_1, \ldots, u_r)^{\perp} = \mathcal{T}(\mathcal{V})^{\perp} = \ker(\mathcal{T}^*).$$

Finally, we expand $\mathcal{T}v$ in the orthonormal basis $(u_1, \ldots, u_m)$ for $\mathcal{W}$ as

$$\mathcal{T}v = \sum_{j=1}^{m} \langle \mathcal{T}v, u_j \rangle_{\mathcal{W}} u_j,$$

where we expand v in the orthonormal basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$ as

$$\langle \mathcal{T}v, u_j \rangle_{\mathcal{W}} = \left\langle \sum_{k=1}^{n} \langle v, v_k \rangle_{\mathcal{V}} \mathcal{T}v_k, u_j \right\rangle_{\mathcal{W}} = \sum_{k=1}^{n} \langle v, v_k \rangle_{\mathcal{V}} \langle \mathcal{T}v_k, u_j \rangle_{\mathcal{W}}.$$

If $1 \le j \le r$, then

$$\langle \mathcal{T}v_k, u_j \rangle_{\mathcal{W}} = \left\langle \mathcal{T}v_k, \frac{\mathcal{T}v_j}{\sigma_j} \right\rangle_{\mathcal{W}} = \frac{1}{\sigma_j} \langle \mathcal{T}^*\mathcal{T}v_k, v_j \rangle_{\mathcal{V}} = \frac{1}{\sigma_j} \langle \lambda_k v_k, v_j \rangle_{\mathcal{V}} = \frac{\lambda_k}{\sigma_j} \langle v_k, v_j \rangle_{\mathcal{V}} = \begin{cases} \lambda_j/\sigma_j, \ j = k \\ 0, \ k \ne j. \end{cases}$$

Since $\lambda_j/\sigma_j = \sigma_j^2/\sigma_j = 1$, we conclude

$$\langle \mathcal{T}v_k, u_j \rangle_{\mathcal{W}} = \begin{cases} \sigma_j, \ j = k \\ 0, \ j \ne k, \end{cases} \qquad 1 \le j \le r.$$

If $j > r$, then since $(u_{r+1}, \ldots, u_m)$ is an orthonormal basis for $\mathcal{T}(\mathcal{V})^\perp$, we have

$$\langle \mathcal{T}v_k, u_j \rangle_{\mathcal{W}} = 0$$

for all k. Thus

$$\langle \mathcal{T}v, u_j \rangle_{\mathcal{W}} = 0, \ r+1 \le j \le m.$$

We combine the two cases on j to conclude

$$\mathcal{T}v = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j.$$

■

42.5 Problem (!). Finish the proof of the SVD by developing the bases in $\mathcal{V}$ for $\ker(\mathcal{T})$ and $\mathcal{T}^*(\mathcal{W})$.

(i) Use the formula (42.5) to check that $\mathcal{T}v_j = 0_{\mathcal{W}}$ for $j = r+1, \ldots, n$. Conclude that $(v_{r+1}, \ldots, v_n)$ is a list of length $n - r$ in $\ker(\mathcal{T})$ and therefore an orthonormal basis for $\ker(\mathcal{T})$.

(ii) Give another proof that $(v_{r+1}, \ldots, v_n)$ is a basis for $\ker(\mathcal{T})$ by using the expansion $\mathcal{T}^*\mathcal{T}v = \sum_{j=1}^{r} \lambda_j \langle v, v_j \rangle_{\mathcal{V}} v_j$ with $\lambda_j > 0$ for all j to conclude that $(v_{r+1}, \ldots, v_n)$ is a basis for the eigenspace of $\mathcal{T}^*\mathcal{T}$ corresponding to 0, and then using $\ker(\mathcal{T}) = \ker(\mathcal{T}^*\mathcal{T})$.

(iii) Use part (ii) of Problem 31.4 to conclude that the list $(v_1, \ldots, v_r)$ is an orthonormal basis for $\mathcal{T}^*(\mathcal{V})$.

(iv) Use the SVD to give yet another proof of rank–nullity.

42.6 Problem (+). Suppose that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional inner product spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ has the SVD $\mathcal{T} = \mathcal{B}_{\mathcal{W}} \mathcal{M}_{\Sigma} \mathcal{B}_{\mathcal{V}}^*$. Work backwards to learn about $\mathcal{T}\mathcal{T}^*$. How do the bases and singular values show up in connection with $\mathcal{T}\mathcal{T}^*$? Then redevelop the proof of the SVD starting with a spectral decomposition of $\mathcal{T}\mathcal{T}^*$, not $\mathcal{T}^*\mathcal{T}$. What is similar and what is different now?

Content from *Linear Algebra* by Meckes & Meckes. All of Section 5.1 is devoted to a proof of the SVD. This proof does not rely on the spectral theorem but rather properties of the operator norm developed in Section 4.4. Section 5.2 develops the SVD for matrices from this operator-theoretic proof. Examples of actual SVDs appear on pp. 289–290, 298–299, and 300–302. Pages 301–302 give an important, meaningful geometric interpretation of the SVD. Pages 316–318 give a matrix-focused algorithm for computing the SVD that follows our spectral theorem-based proof.

Day 43: Wednesday, April 29.

The SVD has many, many applications. One of its chief virtues is approximation. Recall the framework of least squares: if $\mathcal{V}$ and $\mathcal{W}$ are inner product spaces (not necessarily finite-dimensional), if $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, and if $w \in \mathcal{W} \setminus \mathcal{T}(\mathcal{V})$, then if $\mathcal{W} = \mathcal{T}(\mathcal{V}) \oplus \mathcal{T}(\mathcal{V})^\perp$, we solve the approximate problem $\mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w$ instead of the unsolvable problem $\mathcal{T}v = w$. Leaving aside the question of what conditions on $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{T}$ guarantee $\mathcal{W} = \mathcal{T}(\mathcal{V}) \oplus \mathcal{T}(\mathcal{V})^\perp$, there is still a uniqueness concern here. Perhaps $\ker(\mathcal{T}) \neq \{0_{\mathcal{V}}\}$ and $\mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w$ has many solutions. Is there a "best approximate solution" to the unsolvable problem $\mathcal{T}v = w$?

While "best" is quite subjective, one interpretation is "smallest," if perhaps "smallest" means "least complicated." Does there exist a smallest solution $\widehat{v}_0$ to $\mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{W})}w$ in the sense that

$$\mathcal{T}\widehat{v}_0 = \mathcal{P}_{\mathcal{T}(\mathcal{W})}w \quad \text{and} \quad \mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{W})}w \Longrightarrow \|\widehat{v}_0\| \le \|\widehat{v}\|?$$

By the way, in this discussion $\|u\| = \sqrt{\langle u, u\rangle_{\mathcal{V}}}$; that is, the only norm under consideration here is the one induced by the inner product in $\mathcal{V}$.

Yes, if we also have $\mathcal{V} = \ker(\mathcal{T}) \oplus \ker(\mathcal{T})^\perp$. (There are conditions on $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{T}$ under which this decomposition holds, and they are nicely related to the conditions for the decomposition above in $\mathcal{W}$.) For then we could pick one solution $\widehat{v}$ to $\mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{W})}w$ and set $\widehat{v}_0 := \widehat{v} - \mathcal{P}_{\ker(\mathcal{T})}\widehat{v}$, so $\widehat{v}_0 \in \ker(\mathcal{T})^\perp$ and

$$\mathcal{T}\widehat{v}_0 = \mathcal{T}\widehat{v} - \mathcal{T}\mathcal{P}_{\ker(\mathcal{T})}\widehat{v} = \mathcal{T}\widehat{v} = w$$

since $\mathcal{P}_{\ker(\mathcal{T})}\widehat{v} \in \ker(\mathcal{T})$, thus $\mathcal{T}\mathcal{P}_{\ker(\mathcal{T})}\widehat{v} = 0_{\mathcal{W}}$. And then if $\mathcal{T}u = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w$ for some other $u \in \mathcal{V}$, we know that we can write $u = \widehat{v}_0 + z$ for some $z \in \ker(\mathcal{T})$. Since $z \in \ker(\mathcal{T})$ and $\widehat{v}_0 \in \ker(\mathcal{T})^\perp$, the Pythagorean theorem gives

$$\|u\|^2 = \|\widehat{v}_0 + z\|^2 = \|\widehat{v}_0\|^2 + \|z\|^2 \ge \|\widehat{v}_0\|^2,$$

thus $\|\widehat{v}_0\| \le \|u\|$.

This discussion is, at best, annoying. It is highly existential and hinges on the existence of orthogonal decompositions of both the domain and codomain of the operator under consideration. What we probably would prefer is an operator $\mathcal{T}^\dagger \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that

$$\mathcal{T}(\mathcal{T}^\dagger w) = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w \quad \text{and} \quad \mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w \Longrightarrow \left\|\mathcal{T}^\dagger w\right\| \le \|\widehat{v}\|.$$

We can construct such an operator $\mathcal{T}^\dagger$ when $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional via the SVD.

Day 44: Friday, May 1.

Vocabulary from today

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

Pseudoinverse of an operator, trace of a matrix, trace of an operator (on a finite-dimensional inner product space), trace/Frobenius inner product, trace/Frobenius norm

Here is how this works. Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional inner product spaces, and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ have the SVD

$$\mathcal{T}v = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j. \tag{44.1}$$

Then $(u_1, \ldots, u_r)$ is an orthonormal basis for $\mathcal{T}(\mathcal{V})$, so

$$\mathcal{P}_{\mathcal{T}(\mathcal{V})}w = \sum_{j=1}^{r} \langle w, u_j \rangle_{\mathcal{W}} u_j. \tag{44.2}$$

We want to find $\widehat{v} \in \mathcal{V}$ such that $\mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w$.

Using the formula for $\mathcal{T}$ from its SVD (44.1) and for $\mathcal{P}_{\mathcal{T}(\mathcal{V})}$ from (44.2), we want

$$\sum_{j=1}^{r} \sigma_j \langle \widehat{v}, v_j \rangle_{\mathcal{V}} u_j = \sum_{j=1}^{r} \langle w, u_j \rangle_{\mathcal{W}} u_j.$$

Because $(u_1, \ldots, u_r)$ is independent, we should match coefficients in these linear combinations: we want

$$\sigma_j \langle \widehat{v}, v_j \rangle_{\mathcal{V}} = \langle w, u_j \rangle_{\mathcal{W}}.$$

And since $\sigma_j > 0$ for $j = 1, \ldots, r$, we can solve for $\langle \widehat{v}, v_j \rangle_{\mathcal{V}}$ as

$$\langle \widehat{v}, v_j \rangle_{\mathcal{V}} = \frac{\langle w, u_j \rangle_{\mathcal{W}}}{\sigma_j}.$$

While this does not completely determine v, as the Fourier coefficients $\langle \widehat{v}, v_j \rangle_{\mathcal{V}}$ are unspecified for $j > r$, it does suggest putting

$$\mathcal{T}^{\dagger}w := \sum_{j=1}^{r} \frac{\langle w, u_j \rangle_{\mathcal{W}}}{\sigma_j} v_j. \tag{44.3}$$

A direct calculation using (44.1) and (44.3) then shows that $\mathcal{T}(\mathcal{T}^{\dagger}w) = w$.

44.1 Problem (!). Do it.

Now suppose that $\mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{W})}w$ for some $\widehat{v} \in \mathcal{V}$ and $w \in \mathcal{W}$. The work above shows that $\langle \widehat{v}, v_j \rangle_{\mathcal{V}} = \langle w, u_j \rangle_{\mathcal{W}} / \sigma_j$ for $j = 1, \ldots, r$, so

$$\widehat{v} = \sum_{j=1}^{n} \langle \widehat{v}, v_j \rangle_{\mathcal{V}} v_j = \sum_{j=1}^{r} \langle \widehat{v}, v_j \rangle_{\mathcal{V}} v_j + \underbrace{\sum_{j=r+1}^{n} \langle \widehat{v}, v_j \rangle_{\mathcal{V}} v_j}_{z} = \mathcal{T}^{\dagger}w + z.$$

Since $(v_1, \ldots, v_n)$ is orthonormal and $\mathcal{T}^{\dagger}w \in \operatorname{span}(v_1, \ldots, v_r)$ and $z \in \operatorname{span}(v_{r+1}, \ldots, v_n)$, we have $\langle \mathcal{T}^{\dagger}w, z \rangle_{\mathcal{V}} = 0$, from which the Pythagorean theorem gives $\|\mathcal{T}^{\dagger}w\| \le \|\widehat{v}\|$.

44.2 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional inner product spaces with $\dim(\mathcal{V}) = n$ and $\dim(\mathcal{W}) = m$. Let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ with $r = \operatorname{rank}(\mathcal{T}) \ge 1$ and suppose that $\mathcal{T}$ has the SVD*

$$\mathcal{T}v = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j,$$

equivalently, $\mathcal{T} = \mathcal{B}_{\mathcal{W}}\mathcal{M}_{\Sigma}\mathcal{B}_{\mathcal{V}}^*$. Define

$$\mathcal{T}^{\dagger} \colon \mathcal{W} \to \mathcal{V} \colon w \mapsto \sum_{j=1}^{r} \frac{\langle w, u_j \rangle_{\mathcal{W}}}{\sigma_j} v_j.$$

Then

$$\mathcal{T}(\mathcal{T}^{\dagger}w) = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w \quad \text{and} \quad \mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w \Longrightarrow \|\mathcal{T}^{\dagger}w\| \le \|\widehat{v}\|.$$

This operator $\mathcal{T}^{\dagger}$ is the ***PSEUDOINVERSE*** *of $\mathcal{T}$, and it factors as*

$$\mathcal{T}^{\dagger} = \mathcal{B}_{\mathcal{V}}\mathcal{M}_{\Sigma^{\dagger}}\mathcal{B}_{\mathcal{W}}^*,$$

where $\Sigma^{\dagger} \in \mathbb{R}^{n \times m}$ has entries

$$\Sigma_{ij}^{\dagger} := \begin{cases} 1/\sigma_j, & 1 \le i = j \le r \\ 0, & i \ne j \text{ or } i = j > r. \end{cases}$$

For example, if

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

then

$$\Sigma^{\dagger} = \begin{bmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The SVD helps us approximate more than solutions to unsolvable problems. Consider the representation $\mathcal{T}v = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j$. This gives $\mathcal{T}$ as a sum of rank-1 operators—since $\sigma_j \ne 0$, each operator $\sigma_j \langle \cdot, v_j \rangle_{\mathcal{V}} u_j$ is definitely rank-1. If r is large, we may wonder if truncating this (already finite) sum for $\mathcal{T}$ to, say,

$$\mathcal{T}_k v := \sum_{j=1}^{k} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j$$

will somehow approximate $\mathcal{T}$ itself.

But what does that even mean? How can, or should, we measure the size of an operator? Right now, our only notion of size for a *vector* is the norm induced by the inner product.

Here it is helpful to start small. Consider

$$A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \in \mathbb{F}^{2\times 2}.$$

It is natural to associate A with the vector $(a, b, c, d) \in \mathbb{F}^4$, and so we might want to measure the size of A as

$$\|A\| = \sqrt{|a|^2 + |b|^2 + |c|^2 + |d|^2}.$$

Is there an inner product on $\mathbb{F}^{2\times 2}$ such that the norm induced by the inner product is this formula above?

Yes, although finding it takes some insight, or luck. It so happens that

$$A^*A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} \overline{a} & \overline{b} \\ \overline{c} & \overline{d} \end{bmatrix} = \begin{bmatrix} (a\overline{a} + c\overline{c}) & (a\overline{b} + c\overline{d}) \\ (b\overline{a} + d\overline{c}) & (b\overline{b} + d\overline{d}) \end{bmatrix} = \begin{bmatrix} (|a|^2 + |c|^2) & (a\overline{b} + c\overline{d}) \\ (b\overline{a} + d\overline{c}) & (|b|^2 + |d|^2) \end{bmatrix},$$

and so the sum of the diagonal entries of A^*A is $\|A\|^2$ as (tentatively) defined above.

44.3 Definition. *Let $B \in \mathbb{F}^{n\times n}$. The* **TRACE** *of B, denoted $\mathrm{tr}(B)$, is the sum of the diagonal entries of B:*

$$\mathrm{tr}(B) = \sum_{j=1}^{n} B_{jj}.$$

Here is the generalization of our 2×2 calculation above to arbitrary, nonsquare matrices.

44.4 Lemma. *Let $A \in \mathbb{F}^{m\times n}$.*

$$\mathrm{tr}(A^*A) = \sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2.$$

We might then hope that putting $\|A\| = \sqrt{\mathrm{tr}(A^*A)}$ gives a norm induced by an inner product. But what is that inner product? Since we have thought of the adjoint as an analogue of the complex conjugate, and since the dot product in $\mathbb{F}$ boils down to $z \cdot w = z\overline{w}$, and since A^*A resembles $z\overline{z} = |z|^2$, perhaps we should conjecture that this inner product is

$$\langle A, B \rangle = \mathrm{tr}(B^*A).$$

We can indeed prove that this is an inner product on $\mathbb{F}^{m\times n}$, but we can obtain some rather more general results at the cost of very little additional work. The first step is to notice how the trace interacts with the standard basis vectors: for $B \in \mathbb{F}^{n\times n}$, we have

$$\mathrm{tr}(B) = \sum_{j=1}^{n} B_{jj} = \sum_{j=1}^{n} B\mathbf{e}_j \cdot \mathbf{e}_j.$$

This suggests that if $\mathcal{V}$ is a finite-dimensional inner product space with orthonormal basis $(u_1, \ldots, u_n)$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$, we should define

$$\mathrm{tr}(\mathcal{T}) = \sum_{j=1}^{n} \langle \mathcal{T}u_j, u_j \rangle .$$

However, this definition appears to depend on the choice of orthonormal basis.

44.5 Problem (!). Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal bases $(u_1, \ldots, u_n)$ and $(\widetilde{u}_1, \ldots, \widetilde{u}_n)$. Define $\mathcal{S} \in \mathbf{L}(\mathcal{V})$ by setting $\mathcal{S}u_j := \widetilde{u}_j$ and extending $\mathcal{S}$ by linearity (Theorem 18.13) to $\mathcal{V}$.

(i) Show that $\mathcal{S}^*\widetilde{u}_k = u_k$ for each k. [Hint: *Theorem 33.6 gives a formula for $\mathcal{S}^*$ in this setting.*]

(ii) Prove that $\mathcal{S}$ is unitary. [Hint: *compute $\mathcal{S}^*\mathcal{S}u_j$. Then expand $v \in \mathcal{V}$ as $v = \sum_{j=1}^{n} \langle v, u_j \rangle u_j$ and compute $\mathcal{S}^*\mathcal{S}v$.*]

The following fact is the key to showing that the definition of the trace is independent of the choice of orthonormal basis.

44.6 Lemma. *Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal basis $(u_1, \ldots, u_n)$, and let $\mathcal{S}$, $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Then*

$$\sum_{j=1}^{n} \langle \mathcal{S}\mathcal{T}u_j, u_j \rangle = \sum_{j=1}^{n} \langle \mathcal{T}\mathcal{S}u_j, u_j \rangle .$$

Proof. For each j, we have

$$\begin{aligned}
\langle \mathcal{S}\mathcal{T}u_j, u_j \rangle &= \langle \mathcal{T}u_j, \mathcal{S}^*u_j \rangle \\
&= \left\langle \sum_{k=1}^{n} \langle \mathcal{T}u_j, u_k \rangle u_k, \mathcal{S}^*u_j \right\rangle \\
&= \sum_{k=1}^{n} \langle \mathcal{T}u_j, u_k \rangle u_k \mathcal{S}^*u_j \\
&= \sum_{k=1}^{n} \langle \mathcal{T}u_j, u_k \rangle \langle \mathcal{S}u_k, u_j \rangle .
\end{aligned}$$

Interchanging the roles of $\mathcal{S}$ and $\mathcal{T}$ here, and (for future convenience) making j the index of summation but fixing k, we also have

$$\langle \mathcal{T}\mathcal{S}u_k, u_k\rangle = \sum_{j=1}^{n} \langle \mathcal{S}u_k, u_j\rangle \langle \mathcal{T}u_j, u_k\rangle .$$

Then, since we can interchange the order of finite sums,

$$\begin{aligned}\sum_{j=1}^{n} \langle \mathcal{S}\mathcal{T}u_j, u_j\rangle &= \sum_{j=1}^{n}\sum_{k=1}^{n} \langle \mathcal{T}u_j, u_k\rangle \langle \mathcal{S}u_k, u_j\rangle \\ &= \sum_{k=1}^{n}\sum_{j=1}^{n} \langle \mathcal{T}u_j, u_k\rangle \langle \mathcal{S}u_k, u_j\rangle \\ &= \sum_{k=1}^{n} \langle \mathcal{T}\mathcal{S}u_k, u_k\rangle .\end{aligned}$$

■

44.7 Theorem. *Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal bases $(u_1, \ldots, u_n)$ and $(\widetilde{u}_1, \ldots, \widetilde{u}_n)$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Then*

$$\sum_{j=1}^{n} \langle \mathcal{T}u_j, u_j\rangle = \sum_{j=1}^{n} \langle \mathcal{T}\widetilde{u}_j, \widetilde{u}_j\rangle .$$

Proof. Let $\mathcal{S} \in \mathbf{L}(\mathcal{V})$ be the unitary operator from Problem 44.5 such that $\mathcal{S}u_j = \widetilde{u}_j$. Then

$$\begin{aligned}\sum_{j=1}^{n} \langle \mathcal{T}\widetilde{u}_j, \widetilde{u}_j\rangle &= \sum_{j=1}^{n} \langle \mathcal{T}\mathcal{S}u_j, \mathcal{S}u_j\rangle \\ &= \sum_{j=1}^{n} \langle \mathcal{S}^*(\mathcal{T}\mathcal{S})u_j, u_j\rangle \\ &= \sum_{j=1}^{n} \langle (\mathcal{T}\mathcal{S})\mathcal{S}^*u_j, u_j\rangle \\ &= \sum_{j=1}^{n} \langle \mathcal{T}(\mathcal{S}\mathcal{S}^*)u_j, u_j\rangle \text{ by Lemma 44.6} \\ &= \sum_{j=1}^{n} \langle \mathcal{T}u_j, u_j\rangle \text{ since } \mathcal{S} \text{ is unitary.}\end{aligned}$$

■

Now we are assured that our idea for the trace of an operator in general is unambiguous.

44.8 Definition. *Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal basis*

$(u_1, \ldots, u_n)$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. The **TRACE** *of $\mathcal{T}$ is*

$$\operatorname{tr}(\mathcal{T}) := \sum_{j=1}^{n} \langle \mathcal{T}u_j, u_j \rangle .$$

Content from *Linear Algebra* by Meckes & Meckes. Pages 208–209 discuss the trace first for matrices and then for operators on arbitrary finite-dimensional spaces via matrix representations.

A nice consequence of Lemma 44.6 is the identity

$$\operatorname{tr}(\mathcal{S}\mathcal{T}) = \operatorname{tr}(\mathcal{T}\mathcal{S}), \ \mathcal{S}, \mathcal{T} \in \mathbf{L}(\mathcal{V}),$$

sometimes called the "cyclicity of the trace."

44.9 Problem (!). **(i)** Let $A \in \mathbb{F}^{n \times n}$. Prove that $\operatorname{tr}(\mathcal{M}_A) = \operatorname{tr}(A)$, with $\operatorname{tr}(A)$ defined in Definition 44.3.

(ii) Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal basis $(u_1, \ldots, u_n)$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Let $A \in \mathbb{F}^{n \times n}$ be the matrix representation of $\mathcal{T}$ with respect to this orthonormal basis. Use the cyclicity of the trace to prove that $\operatorname{tr}(\mathcal{T}) = \operatorname{tr}(A)$, with, again, $\operatorname{tr}(A)$ defined in Definition 44.3.

More broadly, now that we are equipped with the trace, we can define a meaningful inner product on an operator space.

44.10 Problem. Let $\mathcal{V}$ be a finite-dimensional inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$.

(i) If $\langle \mathcal{T}v, v \rangle \geq 0$ for all $v \in \mathcal{V}$, show that $\operatorname{tr}(\mathcal{T}) \geq 0$ and, moreover, that $\operatorname{tr}(\mathcal{T}) = 0$ if and only if $\mathcal{T} = 0_{\mathcal{V} \to \mathcal{W}}$.

(ii) Show that $\operatorname{tr}(\mathcal{T}^*) = \overline{\operatorname{tr}(\mathcal{T})}$.

44.11 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional inner product spaces and for $\mathcal{S}$, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, put*

$$\langle \mathcal{S}, \mathcal{T} \rangle_{\mathrm{tr}} := \operatorname{tr}(\mathcal{T}^*\mathcal{S}).$$

Then $\langle \cdot, \cdot \rangle_{\mathrm{tr}}$ is an inner product on $\mathbf{L}(\mathcal{V}, \mathcal{W})$, sometimes called the **TRACE INNER PRODUCT** *or the* **FROBENIUS INNER PRODUCT**.

Proof. Before doing anything, we note that for $\mathcal{S}$, $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, we have $\mathcal{T}^* \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ and so $\mathcal{T}^*\mathcal{S} \in \mathbf{L}(\mathcal{V})$. We check conjugacy, nonnegativity, and definiteness. For conjugacy, we want to show that $\langle \mathcal{S}, \mathcal{T} \rangle_{\mathrm{tr}} = \overline{\langle \mathcal{T}, \mathcal{S} \rangle_{\mathrm{tr}}}$. That is, we want $\operatorname{tr}(\mathcal{T}^*\mathcal{S}) = \overline{\operatorname{tr}(\mathcal{S}^*\mathcal{T})}$. Since $\mathcal{T}^*\mathcal{S} = (\mathcal{S}^*\mathcal{T})^*$, this follow from Problem 44.10.

For nonnegativity, we want $\langle \mathcal{T}, \mathcal{T} \rangle_{\text{tr}} \geq 0$. Since $\langle \mathcal{T}, \mathcal{T} \rangle_{\text{tr}} = \text{tr}\, \mathcal{T}^*\mathcal{T}$, and since

$$\langle \mathcal{T}^*\mathcal{T}v, v \rangle = \langle \mathcal{T}v, \mathcal{T}v \rangle = \|\mathcal{T}v\|^2 \geq 0$$

for all $v \in \mathcal{V}$, this too follows from Problem 44.10. Last, that problem also implies that $\langle \mathcal{T}, \mathcal{T} \rangle_{\text{tr}} = 0$ if and only if $\mathcal{T}^*\mathcal{T} = 0_{\mathcal{V}\to\mathcal{V}}$. Then, since $\ker(\mathcal{T}) = \ker(\mathcal{T}^*\mathcal{T})$, we have $\ker(\mathcal{T}) = \ker(0_{\mathcal{V}\to\mathcal{V}}) = \mathcal{V}$, in which case $\mathcal{T} = 0_{\mathcal{V}\to\mathcal{W}}$. ■

44.12 Corollary. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional inner product spaces. Then*

$$\|\mathcal{T}\|_{\text{tr}} := \sqrt{\langle \mathcal{T}, \mathcal{T} \rangle_{\text{tr}}} = \sqrt{\text{tr}(\mathcal{T}^*\mathcal{T})}$$

is a norm on $\mathbf{L}(\mathcal{V}, \mathcal{W})$, sometimes called the ***TRACE NORM*** *or the* ***FROBENIUS NORM****.*

Content from *Linear Algebra* by Meckes & Meckes. Example #3 on p. 234 presents the trace inner product and norm for matrices.

44.13 Example. Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional inner product spaces and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ have the SVD

$$\mathcal{T}v = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j.$$

Factor $\mathcal{T} = \mathcal{B}_{\mathcal{W}}\mathcal{M}_{\Sigma}\mathcal{B}_{\mathcal{V}}^*$. Then

$$\|\mathcal{T}\|_{\text{tr}}^2 = \text{tr}(\mathcal{T}^*\mathcal{T}) = \text{tr}(\mathcal{B}_{\mathcal{V}}\mathcal{M}_{\Sigma^*}\mathcal{B}_{\mathcal{W}}^*\mathcal{B}_{\mathcal{W}}\mathcal{M}_{\Sigma}\mathcal{B}_{\mathcal{V}}^*) = \text{tr}(\mathcal{B}_{\mathcal{V}}\mathcal{M}_{\Sigma^*\Sigma}\mathcal{B}_{\mathcal{V}}^*)$$

since $\mathcal{B}_{\mathcal{W}}$ is unitary. The cyclicity of the trace then gives

$$\text{tr}(\mathcal{B}_{\mathcal{V}}\mathcal{M}_{\Sigma^*\Sigma}\mathcal{B}_{\mathcal{V}}^*) = \text{tr}(\mathcal{B}_{\mathcal{V}}(\mathcal{M}_{\Sigma^*\Sigma}\mathcal{B}_{\mathcal{V}}^*)) = \text{tr}((\mathcal{M}_{\Sigma^*\Sigma}\mathcal{B}_{\mathcal{V}}^*)\mathcal{B}_{\mathcal{V}}) = \text{tr}(\mathcal{M}_{\Sigma^*\Sigma})$$

since $\mathcal{B}_{\mathcal{V}}$ is unitary. Last, since

$$\Sigma^*\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0)$$

by Problem 42.2, we conclude

$$\text{tr}(\mathcal{M}_{\Sigma^*\Sigma}) = \text{tr}(\Sigma^*\Sigma) = \sum_{j=1}^{r} \sigma_j^2$$

by Problem 44.9.

Now, at last, we can state how the "partial sums" of the SVD are a good approximation to an operator.

44.14 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional inner product spaces and let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ have the SVD*

$$\mathcal{T}v = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j.$$

Then, for $k = 1, \ldots, r$, the kth partial sum

$$\mathcal{T}_k v := \sum_{j=1}^{k} \sigma_j \langle v, v_j \rangle_{\mathcal{V}} u_j$$

is the "best rank-k approximation" to $\mathcal{T}$ in the sense that

$$\|\mathcal{T} - \mathcal{T}_k\|_{\mathrm{tr}} = \min_{\substack{\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W}) \\ \mathrm{rank}(\mathcal{S}) \leq k}} \|\mathcal{T} - \mathcal{S}\|_{\mathrm{tr}}.$$

In particular,

$$\|\mathcal{T} - \mathcal{T}_k\|_{\mathrm{tr}} = \left(\sum_{j=k+1}^{r} \sigma_j^2 \right)^{1/2}.$$

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 5.9 on p. 306, which is discussed in detail on pp. 306–308.

SELECTED SOLUTIONS FOR THE DAILY LOG
MATH 4260: LINEAR ALGEBRA II

Timothy E. Faver
May 5, 2026

Only solutions to problems assigned on problem sets appear here.

1.2. (i) This is probably done most efficiently using dot products of rows of A with the column vector $\mathbf{x}$.

(ii) The second row of $A\mathbf{x}$ is 2 times the first row, so b_2 must equal $2b_1$.

(iii) We want the vector $\mathbf{b} = (b_1, b_2, b_3)$ to satisfy $b_2 = 2b_1$, equivalently,

$$2b_1 - b_2 + 0b_3 = 0.$$

That is,

$$\begin{bmatrix} b_1 \\ b_2 \; b_3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = 0,$$

so we can take $\mathbf{z} = (2, -1, 0)$ here.

1.3. (i) First we subtract 2 times row 1 from row 2:

$$\begin{bmatrix} 1 & 2 & 1 & 7 & b_1 \\ 2 & 4 & 2 & 14 & b_2 \\ 0 & 0 & 2 & 8 & b_3 \end{bmatrix} \xrightarrow{\text{R2} \mapsto \text{R2}-2\times\text{R1}} \begin{bmatrix} 1 & 2 & 1 & 7 & b_1 \\ 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 2 & 8 & b_3 \end{bmatrix}.$$

Then we multiply row 3 by 1/2:

$$\begin{bmatrix} 1 & 2 & 1 & 7 & b_1 \\ 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 2 & 8 & b_3 \end{bmatrix} \xrightarrow{\text{R3} \mapsto (1/2)\times\text{R3}} \begin{bmatrix} 1 & 2 & 1 & 7 & b_1 \\ 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 1 & 4 & b_3/2 \end{bmatrix}.$$

Then we subtract row 3 from row 1:

$$\begin{bmatrix} 1 & 2 & 1 & 7 & b_1 \\ 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 1 & 4 & b_3/2 \end{bmatrix} \xrightarrow{\text{R1} \mapsto \text{R1}-\text{R3}} \begin{bmatrix} 1 & 2 & 0 & 3 & b_1 - b_3/2 \\ 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 1 & 4 & b_3/2 \end{bmatrix}.$$

Last, we interchange rows 2 and 3:

$$\begin{bmatrix} 1 & 2 & 0 & 3 & b_1 - b_3/2 \\ 0 & 0 & 0 & 0 & b_2 - 2b_1 \\ 0 & 0 & 1 & 4 & b_3/2 \end{bmatrix} \xrightarrow{\text{R2} \mapsto \text{R3},\ \text{R3}\mapsto\text{R2}} \begin{bmatrix} 1 & 2 & 0 & 3 & b_1 - b_3/2 \\ 0 & 0 & 1 & 4 & b_3/2 \\ 0 & 0 & 0 & 0 & b_2 - 2b_1 \end{bmatrix}.$$

(ii) The problem $R\mathbf{x} = \mathbf{c}$ reads

$$\begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 - b_3/2 \\ b_3/2 \\ b_2 - 2b_1 \end{bmatrix}.$$

Doing the matrix-vector multiplication on the left and equating componentwise leads to

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

(iii) We solve for x_1 and x_3 as

$$x_1 = b_1 - \frac{b_3}{2} - 2x_2 - 3x_4 \quad \text{and} \quad x_3 = \frac{b_3}{2} - 4x_4.$$

Then the solution (assuming $b_2 - 2b_1 = 0$) is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 - b_3/2 - 2x_2 - 3x_4 \\ x_2 \\ b_3/2 - 4x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 - b_3/2 \\ 0 \\ b_3/2 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

1.6. (i) We want to show that with $f(x) := \int_0^x g(s)\ ds$, we have $f(x+1) = f(x)$ for all x. That is, we want

$$\int_0^{x+1} g(s)\ ds = \int_0^x g(s)\ ds$$

for all x. By integral property ($\int 2$), this is equivalent to

$$\int_0^x g(s)\ ds + \int_x^{x+1} g(s)\ ds = \int_0^x g(s)\ ds,$$

and that is the same as having $G(x) = 0$ for all x.

Now we check that $G = 0$. Differentiating and using the fundamental theorem of calculus, we find

$$G'(x) = g(x+1) - g(x) = 0$$

since g is 1-periodic. So, G is constant, and therefore $G(x) = G(0)$ for all x. In particular, $G(0) = \int_0^1 g(s)\ ds = 0$.

(ii) Since $\int_0^1 f(s)\ ds = 0$ and $f(s) = f(0) + \int_0^s g(x)\ dx$, we have

$$0 = \int_0^1 f(s)\ ds = \int_0^1 \left(f(0) + \int_0^s g(x)\ dx \right)\ ds = \int_0^1 f(0)\ ds + \int_0^1 \int_0^s g(x)\ dx\ ds$$
$$= f(0) + \int_0^1 \int_0^s g(x)\ dx\ ds.$$

Interchanging the order of integration gives

$$\int_0^1 \int_0^s g(x)\ dx\ ds = \int_0^1 \int_x^1 g(x)\ ds\ dx = \int_0^1 g(x) \int_x^1 1\ ds\ dx = \int_0^1 (1-x)g(x)\ dx$$
$$= \int_0^1 g(x)\ dx - \int_0^1 xg(x)\ dx = -\int_0^1 xg(x)\ dx$$

since $\int_0^1 g(x)\ dx = 0$. Thus

$$0 = f(0) - \int_0^1 xg(x)\ dx.$$

2.8. (i) The ordered pairs $(1,-1)$ and $(1,1)$ are both in $f := \{(1,-1),(1,1),(2,1),(3,-1),(4,1)\}$. This violates the requirement that for each $x \in \{1,2,3,4\}$, there is a unique $y \in \{1,-1\}$ such that $(x,y) \in f$. Specifically, it violates the uniqueness of y.

(ii) Let $y \in f(I)$. Then $y = f(x)$ for some $x \in I$, so $y = x^2$ for some $x \geq 0$. Since $x^2 \geq 0$ for all $x \in \mathbb{R}$, we have $y = x^2 \in [0,\infty) = I$.

Now let $y \in I$, so $y \geq 0$. Then $x := \sqrt{y} \in \mathbb{R}$ satisfies $x^2 = y$, so $f(x) = y$. Thus $y \in f(I)$.

(iii) Since $(\pm 1)^2 = 1$, both $(1,1)$ and $(1,-1)$ belong to this set. This violates the uniqueness part of a function's definition. However, since $y^2 \geq 0$ for any $y \in \mathbb{R}$, if (x,y) is in this set, then $x = y^2 \geq 0$. Thus $(-1,1)$ is not in this set, so it cannot be a function with domain equal to $\mathbb{R}$.

2.21. Since $\mathbb{F}^{1\times n} = \mathbb{F}^{\{1\}\times\{1,\ldots,n\}}$, the domain of a function in $\mathbb{F}^{1\times n}$ is the set of ordered pairs $\{(1,k)\}_{k=1}^n$. But $\mathbb{F}^n = \mathbb{F}^{\{1,\ldots,n\}}$, so the domain of a function in $\mathbb{F}^n$ is just the set of integers $\{1,\ldots,n\}$.

3.7. Since $(f +_{\mathbb{F}^X} g)(x) = f(x) + g(x)$, we have

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

Since $(\alpha \cdot f)(x) = \alpha f(x)$, we have

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

And since $(-f)(x) = -f(x) = (-1)f(x)$, we have

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

4.8. Denote by 0 the function $0\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto 0$. Then $0 \in \mathcal{V}$, since $\mathcal{V}$ is a subspace of $\mathbb{R}^{\mathbb{R}}$. Also, $0(-x) = 0 = 0(x)$, so 0 is even. Thus $0 \in \mathcal{U}$.

Next, let $f,\ g \in \mathcal{U}$. Then $f,\ g \in \mathcal{V}$, so $f+g \in \mathcal{V}$ since $\mathcal{V}$ is a subspace of $\mathbb{R}^{\mathbb{R}}$. Also, $(f+g)(-x) = f(-x)+g(-x) = f(x)+g(x) = (f+g)(x)$, so $f+g$ is even. Hence $f+g \in \mathcal{U}$.

Finally, let $\alpha \in \mathbb{R}$ and $f \in \mathcal{U}$. Then $f \in \mathcal{V}$, so $\alpha f \in \mathcal{V}$, since $\mathcal{V}$ is a subspace of $\mathbb{R}^{\mathbb{R}}$. Also, $(\alpha f)(-x) = \alpha f(-x) = \alpha f(x) = (\alpha f)(x)$, so αf is even. Hence $\alpha f \in \mathcal{U}$.

4.10. (i) If $(a_k),\ (b_k) \in c_0$, then since $a_k \to 0$ and $b_k \to 0$, properties of limits imply $a_k+b_k \to 0+0 = 0$, thus $(a_k)+(b_k) = (a_k+b_k) \in c_0$. If $\alpha \in \mathbb{F}$ and $(a_k) \in c_0$, more properties of limits limply $\alpha a_k \to \alpha \cdot 0 = 0$, thus $\alpha(a_k) = (\alpha a_k) \in c_0$. Finally, if $z_k := 0$ for all k, then $z_k \to 0$, so $(0) \in c_0$.

(ii) Assume $\alpha \neq 0$. If $(a_k),\ (b_k) \in \mathcal{U}_\alpha$, then $a_k + b_k \to 2\alpha$, and, since $\alpha \neq 0$, $2\alpha \neq \alpha$. Thus $(a_k) + (b_k) \notin \mathcal{U}_\alpha$. If $\beta \in \mathbb{F}$ and $(a_k) \in \mathcal{U}_\alpha$, then $\beta a_k \to \beta\alpha$. If $\beta \neq 1$, then $\beta\alpha \neq \alpha$. So, $\beta(a_k) \notin \mathcal{U}_\alpha$ for $\beta \neq 1$. Finally, if $z_k := 0$ for all k, then $z_k \to 0 \neq \alpha$, so $(0) \notin \mathcal{U}_\alpha$.

5.6. One way to prove this is to compute, as suggested,

$$\mathcal{T}0_{\mathcal{V}} = \mathcal{T}(0_{\mathcal{V}} + 0_{\mathcal{V}}) = \mathcal{T}0_{\mathcal{V}} + \mathcal{T}0_{\mathcal{V}}$$

by linearity of $\mathcal{T}$ and the identity $0_{\mathcal{V}} = 0_{\mathcal{V}} + 0_{\mathcal{V}}$. Subtracting, we then find $\mathcal{T}0_{\mathcal{V}} = 0_{\mathcal{W}}$.

Another way to prove this is to compute, also as suggested,

$$\mathcal{T}0_{\mathcal{V}} = \mathcal{T}(00_{\mathcal{V}}) = 0\mathcal{T}0_{\mathcal{V}} = 0_{\mathcal{W}}.$$

Here we have used the linearity of $\mathcal{T}$ (in the second equality) and the identities $00_{\mathcal{V}} = 0_{\mathcal{V}}$ and $00_{\mathcal{W}} = 0_{\mathcal{W}}$.

5.7. (i) We have

$$0_{\mathcal{V}\to\mathcal{W}}(v_1 + v_2) = 0_{\mathcal{W}} = 0_{\mathcal{W}} + 0_{\mathcal{W}} = 0_{\mathcal{V}\to\mathcal{W}}v_1 + 0_{\mathcal{V}\to\mathcal{W}}v_2$$

and

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

(ii) Yes, when $w_0 = 0_{\mathcal{W}}$, as shown above. For $w_0 \neq 0_{\mathcal{W}}$, we have

$$\mathcal{T}(v_1 + v_2) = w_0 \neq 2w_0 = w_0 + w_0 = \mathcal{T}v_1 + \mathcal{T}v_2$$

and, for $\alpha \neq 1$,

$$\mathcal{T}(\alpha v) = w_0 \neq \alpha w_0 = \alpha\mathcal{T}v.$$

So, $\mathcal{T}$ is not linear when $w_0 \neq 0_{\mathcal{W}}$.

7.8. Recall that $\mathbb{F}^{m\times n} = \mathbb{F}^{\{1,\ldots,m\}\times\{1,\ldots,n\}}$. The domain of $A \in \mathbb{F}^{m\times n}$ is therefore $\{1,\ldots,m\} \times \{1,\ldots,n\}$, while the domain of $\mathcal{M}_A$ is $\mathbb{F}^n = \mathbb{F}^{\{1,\ldots,n\}}$, and so $\{1,\ldots,m\} \times \{1,\ldots,n\} \neq \mathbb{F}^n$. For two functions to be equal, they must have the same domain; here A and $\mathcal{M}_A$ do not. The codomain of A is $\mathbb{F}$, whereas the codomain of $\mathcal{M}_A$ is $\mathbb{F}^m$. If $m \neq 1$, then $\mathbb{F} \neq \mathbb{F}^m$.

7.11. Working backwards,

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

Take

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

The action of $\mathcal{M}_A$ on $\mathbf{v}$ is to subtract 2 times the first row of $\mathbf{v}$ from the second row.

7.20. We have $\lambda_1 v = \mathcal{T}v = \lambda_2 v$, thus $(\lambda_1 - \lambda_2)v = 0_{\mathcal{V}}$. Since v is an eigenvector, $v \neq 0_{\mathcal{V}}$, thus $\lambda_1 - \lambda_2 = 0$, and so $\lambda_1 = \lambda_2$.

8.2. **(i)** If $\int_0^x f(s)\,ds = 0$ for all $x \in [0,1]$, then differentiating both sides yields $f(x) = 0$, so f cannot be an eigenvector. That is, no $f \neq 0$ satisfies $\mathcal{T}f = 0$, so 0 cannot be an eigenvalue.

(ii) Since $\mathcal{T}f$ is differentiable and $f = \lambda^{-1}\mathcal{T}f$, f is also differentiable. Differentiating both sides yields $f'(x) = \lambda^{-1}(\mathcal{T}f)'(x) = \lambda^{-1}f(x)$. That is, f solves the ODE $f' = \lambda^{-1}f$, and so $f(x) = f(0)e^{x/\lambda}$. But $\mathcal{T}f = \lambda f$, and so

$$\lambda f(0)e^{x/\lambda} = \lambda f(x) = (\mathcal{T}f)(x) = \int_0^x f(s)\,ds = \int_0^x f(0)e^{s/\lambda}\,ds = \frac{f(0)e^{x/\lambda} - f(0)}{1/\lambda}.$$

That is,

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

and so $\lambda f(0) = 0$. Since $\lambda \neq 0$, we must have $f(0) = 0$. But then $f = 0$, so no $f \neq 0$ satisfies $\mathcal{T}f = \lambda f$, and therefore λ is not an eigenvalue.

8.4. If $\lambda \in \mathbb{C} \setminus \mathbb{R}$ is an eigenvalue, then, as in Example 8.3, there is $f \in \mathcal{V}$ such that f is nonzero and $(x - \lambda)f(x) = 0$ for all $x \in [0,1]$. Since $x \in [0,1] \subseteq \mathbb{R}$ and $\lambda \in \mathbb{C} \setminus \mathbb{R}$, we have $x - \lambda \neq 0$, thus $f(x) = 0$ for all $x \in [0,1]$. Hence f cannot be an eigenvector of $\mathcal{T}$, and so λ is not an eigenvalue.

8.6. **(i)** If $\mathcal{T}(a_k) = \lambda(a_k)$, then $(0, a_1, a_2, a_3, \ldots) = (\lambda a_1, \lambda a_2, \lambda a_3, \lambda a_4, \ldots)$, and so $0 = \lambda a_1$, $a_1 = \lambda a_2$, $a_2 = \lambda a_3$, and, in general, for $k \geq 1$, $a_k = \lambda a_{k+1}$.

(ii) We have $a_k = 0a_{k+1} = 0$ for all k, so $(a_k) = (0)$. So, no nonzero sequence can be an eigenvector for 0, and 0 is not an eigenvalue.

(iii) Since $a_k = \lambda a_{k+1}$ and $\lambda \neq 0$, we have $a_{k+1} = \lambda^{-1}a_k$. And since $a_1 = 0$, we have $a_2 = \lambda^{-1}a_1 = 0$, $a_3 = \lambda^{-1}a_2 = 0$, and, in general, $a_k = 0$. This could be proved by induction to show first that $a_{k+1} = \lambda^{-k}a_1$. Again, no nonzero sequence can be an eigenvector for λ, and so λ is not an eigenvalue.

9.5. Since $\alpha = \alpha \cdot 1$, we have

$$\mathcal{T}\alpha = \mathcal{T}(\alpha \cdot 1) = \alpha\mathcal{T}1 = \alpha w_1.$$

9.10. We have

$$0_{\mathcal{V}\to\mathcal{W}}\mathcal{T}u = 0_{\mathcal{W}} \quad \text{and} \quad \mathcal{S}0_{\mathcal{U}\to\mathcal{V}}u = \mathcal{S}0_{\mathcal{V}} = 0_{\mathcal{W}},$$

so

$$0_{\mathcal{V}\to\mathcal{W}}\mathcal{T} = \mathcal{S}0_{\mathcal{U}\to\mathcal{V}} = 0_{\mathcal{U}\to\mathcal{W}}.$$

9.11. We have

$$\mathcal{I}_{\mathcal{W}}\mathcal{T}v = \mathcal{T}v \quad \text{and} \quad \mathcal{T}\mathcal{I}_{\mathcal{V}}v = \mathcal{T}v,$$

so

$$\mathcal{I}_{\mathcal{W}}\mathcal{T} = \mathcal{T}\mathcal{I}_{\mathcal{V}} = \mathcal{T}.$$

10.5. Take $\mathcal{U} = \mathcal{V} = \mathcal{W} = \mathbb{R}^2$, $\mathcal{T} = \mathcal{M}_A$, and $\mathcal{S} = \mathcal{M}_B$, where

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

11.3. ($\Longrightarrow$) Let x_1, $x_2 \in \mathcal{X}$ and put $y_1 = \mathcal{T}x_1$, $y_2 = \mathcal{T}x_2$. Then (x_1, y_1), $(x_2, y_2) \in \mathcal{T}$, and so $(x_1 + x_2, y_1 + y_2) \in \mathcal{T}$. That is, $\mathcal{T}(x_1 + x_2) = y_1 + y_2 = \mathcal{T}x_1 + \mathcal{T}x_2$.

Next, let $\alpha \in \mathbb{F}$ and $x \in \mathcal{X}$. Put $y = \mathcal{T}x$. Then $(x, y) \in \mathcal{T}$, so $(\alpha x, \alpha y) \in \mathcal{T}$. That is, $\mathcal{T}(\alpha x) = \alpha y = \alpha\mathcal{T}x$.

($\Longleftarrow$) Let (x_1, y_1), $(x_2, y_2) \in \mathcal{T}$. Then $y_1 = \mathcal{T}x_1$ and $y_2 = \mathcal{T}x_2$, so $y_1 + y_2 = \mathcal{T}x_1 + \mathcal{T}x_2 = \mathcal{T}(x_1 + x_2)$ since $\mathcal{T}$ is linear. So $(x_1 + y_1, x_2 + y_2) \in \mathcal{T}$.

Next, let $\alpha \in \mathbb{F}$ and $(x, y) \in \mathcal{T}$. Then $y = \mathcal{T}x$, so $\alpha y = \alpha\mathcal{T}x = \mathcal{T}(\alpha x)$ since $\mathcal{T}$ is linear. So $(\alpha x, \alpha y) \in \mathcal{T}$.

11.11. (i) For surjectivity, let $g \in \mathcal{C}([0, 1])$ and put $f(x) := \int_0^x g(s)\ ds$, so $f' = g$. For the lack of injectivity, let $f_1(x) = 0$ and $f_2(x) = 1$, so $f_1' = f_2'(= 0)$, but $f_1 \neq f_2$.

(ii) For injectivity, if $\mathcal{T}f = \mathcal{T}g$, then $(\mathcal{T}f)' = (\mathcal{T}g)'$, thus $f = g$. For the lack of surjectivity, observe that $(\mathcal{T}f)(0) = 0$. Consequently, there is no f such that $\mathcal{T}f = 1$.

11.12. (i) If $\mathcal{T}v = 0_{\mathcal{W}}$, then $\mathcal{L}\mathcal{T}v = \mathcal{L}0_{\mathcal{W}} = 0_{\mathcal{V}}$. But $\mathcal{L}\mathcal{T}v = v$ as well, so $v = 0_{\mathcal{V}}$. Thus $\ker(\mathcal{T}) = \{0_{\mathcal{V}}\}$ and therefore $\mathcal{T}$ is injective.

(ii) Let $w \in \mathcal{V}$. Then $\mathcal{T}\mathcal{R}w = w$, so $w \in \mathcal{T}(\mathcal{V})$. Thus $\mathcal{T}(\mathcal{V}) = \mathcal{W}$, so $\mathcal{T}$ is surjective.

(iii) The previous two parts establish that if $\mathcal{T}$ has a left inverse and a right inverse, then $\mathcal{T}$ is both injective and surjective, thus bijective, thus invertible. We adapt the calculation in (10.5) to compute, for $w \in \mathcal{W}$,

$$\begin{aligned} \mathcal{L}w &= \mathcal{L}(\mathcal{T}\mathcal{R}w) \text{ since } \mathcal{T}\mathcal{R}w = w \text{ for all } w \in \mathcal{W} \\ &= (\mathcal{L}\mathcal{T})(\mathcal{R}w) \text{ by associativity of operator composition} \\ &= \mathcal{R}w \text{ since } \mathcal{L}\mathcal{T}v = v \text{ for all } v \in \mathcal{V}. \end{aligned}$$

(iv) That $\mathcal{T}\mathcal{G}(f) = f$ is the fundamental theorem of calculus (for the first term in $\mathcal{G}$) combined with the derivative of a constant function (for the second term). Certainly $\mathcal{G}$ is not linear:

$$(\mathcal{G}(2f))(x) = 2\int_0^x f(s)\ ds + 4(f(0))^2.$$

If, say, $f(x) = 1$ for all x, then

$$(\mathcal{G}(2f))(x) = 2x + 4,$$

but

$$(\mathcal{G}f)(x) = x + 1,$$

so $\mathcal{G}(2f) \neq 2\mathcal{G}f$.

12.11. Suppose that $(\mathcal{T} - \lambda\mathcal{I})f = g$. Then $g(x) = (\mathcal{T}f)(x) - \lambda f(x) = xf(x) - \lambda f(x) = (x - \lambda)f(x)$, and so $g(\lambda) = (\lambda - \lambda)f(\lambda) = 0$. The function $g(x) = 1$ therefore cannot be in the range of $\mathcal{T} - \lambda\mathcal{I}$, and so this operator is not surjective.

12.12. Let $v \in \ker(\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})$, so $\mathcal{T}v - \lambda v = 0_{\mathcal{V}}$. We want to show that $\mathcal{T}v \in \ker(\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})$, too. That is, we want $(\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})(\mathcal{T}v) = 0_{\mathcal{V}}$. We compute

$$(\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})(\mathcal{T}v) = \mathcal{T}(\mathcal{T}v) - \lambda\mathcal{I}_{\mathcal{V}}(\mathcal{T}v) = \mathcal{T}^2v - \lambda\mathcal{T}v = \mathcal{T}(\mathcal{T}v - \lambda v) = \mathcal{T}0_{\mathcal{V}} = 0_{\mathcal{V}}.$$

Now let $w \in (\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})(\mathcal{V})$, so $w = \mathcal{T}v - \lambda v$ for some $v \in \mathcal{V}$. We want to show that $\mathcal{T}w \in (\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})(\mathcal{V})$, too. That is, we want $\mathcal{T}w = (\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})u$ for some $u \in \mathcal{V}$. We compute

$$\mathcal{T}w = \mathcal{T}(\mathcal{T}v - \lambda v) = \mathcal{T}^2v - \lambda\mathcal{T}v = \mathcal{T}(\mathcal{T}v) - \lambda\mathcal{I}_{\mathcal{V}}(\mathcal{T}v) = (\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})(\mathcal{T}v).$$

Take $u = \mathcal{T}v$ to conclude $\mathcal{T}w = (\mathcal{T} - \lambda\mathcal{I}_{\mathcal{V}})u$.

14.3. First suppose that $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$. We check injectivity of $\mathcal{B}$. If $\mathcal{B}(\alpha_1, \ldots, \alpha_n) = 0_{\mathcal{V}}$, then $\sum_{j=1}^n \alpha_j v_j = 0_{\mathcal{V}}$. Since $(v_1, \ldots, v_n)$ is linearly independent, we have $\alpha_j = 0$ for all j. That is, $\ker(\mathcal{B}) = \{\mathbf{0}_n\}$. For surjectivity, if $v \in \mathcal{V}$, then since $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$, there is $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$ such that $v = \sum_{j=1}^n \alpha_j v_j = \mathcal{B}(\alpha_1, \ldots, \alpha_n)$.

Now suppose that $\mathcal{B}$ as defined in the problem statement is an isomorphism. We check that $(v_1, \ldots, v_n)$ is independent. If $\sum_{j=1}^n \alpha_j v_j = 0_{\mathcal{V}}$, then $\mathcal{B}(\alpha_1, \ldots, \alpha_n) = 0_{\mathcal{V}}$. Since $\mathcal{B}$ is injective, $(\alpha_1, \ldots, \alpha_n) = \mathbf{0}_n$. To show that $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$, we take $v \in \mathcal{V}$ and use the surjectivity of $\mathcal{B}$ to find $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$ such that $v = \mathcal{B}(\alpha_1, \ldots, \alpha_n) = \sum_{j=1}^n \alpha_j v_j \in \operatorname{span}(v_1, \ldots, v_n)$.

Finally, by definition of $\mathcal{B}$, we have $\mathcal{B}\mathbf{e}_j = v_j$.

14.20. Suppose that $\sum_{j=1}^n \alpha_j \mathcal{T}_j = 0_{\mathcal{V}\to\mathcal{V}}$. Fix k and compute

$$0_{\mathcal{V}\to\mathcal{V}} = \mathcal{T}_k 0_{\mathcal{V}\to\mathcal{V}} = \mathcal{T}_k \sum_{j=1}^n \alpha_j \mathcal{T}_j = \sum_{j=1}^n \alpha_j \mathcal{T}_k \mathcal{T}_j = \alpha_k \mathcal{T}_k^2.$$

Since $\mathcal{T}_k^2 \neq 0_{\mathcal{V}\to\mathcal{V}}$, we must have $\alpha_k = 0$.

15.5. (i) Suppose that $\sum_{j=1}^n \alpha_j \mathcal{T}v_j = 0_{\mathcal{W}}$. Since $\sum_{j=1}^n \alpha_j \mathcal{T}v_j = \mathcal{T}\sum_{j=1}^n \alpha_j v_j$ by linearity, this says $\mathcal{T}\sum_{j=1}^n \alpha_j v_j = 0_{\mathcal{W}}$. Injectivity then forces $\sum_{j=1}^n \alpha_j v_j = 0_{\mathcal{V}}$, and the independence of $(v_1, \ldots, v_n)$ implies $\alpha_j = 0$ for all j.

(ii) Let $\mathcal{T} = 0_{\mathcal{V}\to\mathcal{W}}$. Then the image of any list under $\mathcal{T}$ has $0_{\mathcal{W}}$ as a term and therefore must be dependent.

(iii) If $\sum_{j=1}^n \alpha_j v_j = 0_{\mathcal{V}}$, then

$$0_{\mathcal{W}} = \mathcal{T}\sum_{j=1}^n \alpha_j v_j = \sum_{j=1}^n \alpha_j \mathcal{T}v_j = \sum_{j=1}^n \alpha_j w_j,$$

thus $\alpha_j = 0$ for all j.

15.6. (i) By surjectivity, for each $w \in \mathcal{W}$, there is $v \in \mathcal{V}$ such that $w = \mathcal{T}v$. Write $v = \sum_{j=1}^{n} \alpha_j v_j$. Then $w = \mathcal{T}v = \sum_{j=1}^{n} \alpha_j \mathcal{T}v_j$.

(ii) Suppose that $\mathcal{W}$ is not the zero vector space and let $\mathcal{T} = 0_{\mathcal{V}\to\mathcal{W}}$. Then the image of any list under $\mathcal{T}$ is just the list whose entries are all $0_{\mathcal{W}}$, and that list cannot span $\mathcal{W}$.

15.7.

18.6. Let $v_0 \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$. For each $\alpha_1, \alpha_2 \in \mathbb{F}$, we have $\alpha_1 v_0 \neq \alpha_2 v_0$; otherwise, $(\alpha_1 - \alpha_2)v_0 = 0_{\mathcal{V}}$, and then since $v_0 \neq 0_{\mathcal{V}}$, it must be the case that $\alpha_1 - \alpha_2 = 0$. So, $\{\alpha v_0 \mid \alpha \in \mathbb{F}\} \subseteq \mathcal{V}$ contains infinitely many vectors.

18.7. Let $\mathcal{V} = \text{span}(v_1)$ for some $v_1 \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$. For any $v \in \mathcal{V}$, we have $v = \alpha_1 v_1$ for some $\alpha_1 \in \mathbb{F}$, so $\mathcal{T}v = \mathcal{T}(\alpha_1 v_1) = \alpha_1 \mathcal{T}v_1$. That is, $\mathcal{T}$ effectively acts as scalar multiplication on the vector $\mathcal{T}v_1 \in \mathcal{W}$.

19.7. (i) Let $n \geq 1$. Since $\mathcal{V}$ is infinite-dimensional, there is an independent list $(v_1, \ldots, v_n)$ of length n in $\mathcal{V}$. Since $\mathcal{T}$ is injective, the list $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ is independent in $\mathcal{W}$ by Problem 15.5. So, for each $n \geq 1$, there is an independent list of length n in $\mathcal{W}$. Hence $\mathcal{W}$ is infinite-dimensional.

(ii) Suppose instead that $\mathcal{V}$ is finite-dimensional. If $\mathcal{V} = \{0_{\mathcal{V}}\}$, then $\mathcal{W} = \mathcal{T}(\mathcal{V}) = \{\mathcal{T}0_{\mathcal{V}}\} = \{0_{\mathcal{W}}\}$, and so $\mathcal{W}$ is finite-dimensional. If $\mathcal{V} \neq \{0_{\mathcal{V}}\}$, let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$. Then $\mathcal{V} = \text{span}(v_1, \ldots, v_n)$, so by Problem 15.6, since $\mathcal{T}$ is surjective, $\mathcal{W} = \text{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$. Thus $\mathcal{W}$ is finite-dimensional, too. Let $w \in \mathcal{W}$. Since $\mathcal{T}$ is surjective, there is $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. Write $v = \sum_{j=1}^{n} \alpha_j v_j$. Then $w = \sum_{j=1}^{n} \alpha_j \mathcal{T}v_j \in \text{span}(v_1, \ldots, v_n)$.

19.9. If $\mathcal{U} = \{0_{\mathcal{V}}\}$, let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$ and let $(w_1, \ldots, w_n)$ be any list in $\mathcal{W}$. Put $\widetilde{\mathcal{T}}v_j = w_j$ and extend $\widetilde{\mathcal{T}}$ by linearity to all of $\mathcal{V}$. Of course $\widetilde{\mathcal{T}}0_{\mathcal{V}} = 0_{\mathcal{W}} = \mathcal{T}0_{\mathcal{V}}$, so $\widetilde{\mathcal{T}}v = \mathcal{T}v$ for all $v \in \mathcal{U}$.

If $\mathcal{U} = \mathcal{V}$, we can take $\widetilde{\mathcal{T}} = \mathcal{T}$. Then $\widetilde{\mathcal{T}}v = \mathcal{T}v$ for all $v \in \mathcal{U} = \mathcal{V}$, so $\widetilde{\mathcal{T}} = \mathcal{T}$ and the extension is unique.

Otherwise, suppose that $\mathcal{U} \neq \mathcal{V}$ and $\mathcal{U} \neq \{0_{\mathcal{V}}\}$ and let $(u_1, \ldots, u_r)$ be a basis for $\mathcal{U}$. (That $\mathcal{U} \neq \{0_{\mathcal{V}}\}$ implies that $\mathcal{U}$ has a basis; that $\mathcal{U} \neq \mathcal{V}$ implies $r < n$.) Extend $(u_1, \ldots, u_r)$ to a basis $(u_1, \ldots, u_r, v_{r+1}, \ldots, v_n)$ for $\mathcal{V}$. Put $\widetilde{\mathcal{T}}u_j = \mathcal{T}u_j$ for $j = 1, \ldots, r$, and let $\widetilde{\mathcal{T}}v_j$ be any desired vector in $\mathcal{W}$ for $j = r+1, \ldots, n$. Extend $\widetilde{\mathcal{T}}$ by linearity to all of $\mathcal{V}$. For $u = \sum_{j=1}^{r} \alpha_j u_j \in \mathcal{U}$, we have

$$\widetilde{\mathcal{T}}u = \sum_{j=1}^{r} \alpha_j \widetilde{\mathcal{T}}u_j = \sum_{j=1}^{r} \alpha_j \mathcal{T}u_j = \mathcal{T}u.$$

This extension is not unique because we can choose the values of $\widetilde{\mathcal{T}}v_j$ arbitrarily.

20.5. 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 $\mathcal{V}$, and let $(\varphi_1, \ldots, \varphi_n)$ be the dual basis for $\mathcal{V}'$ relative to $(v_1, \ldots, v_n)$. Suppose that $\sum_{j=1}^n \alpha_j \varphi_j = 0$; we want to show $\alpha_j = 0$ for all j. The assumption $\sum_{j=1}^n \alpha_j \varphi_j = 0$ means $\sum_{j=1}^n \alpha_j \varphi_j(v) = 0$ for all $v \in \mathcal{V}$. Taking $v = v_k$ for $k = 1, \ldots, n$ gives

$$0 = \sum_{j=1}^n \alpha_j \varphi_j(v_k) = 0 = \alpha_k,$$

since $\varphi_j(v_k) = 0$ for $j \neq k$ and $\varphi_j(v_k) = 1$ for $j = k$. That is, $\alpha_k = 0$ for all k, as desired.

20.6. (i) Assume that $\alpha_1\varphi_1 + \alpha_2\varphi_2 = 0$; we want to show that $\alpha_1 = \alpha_2 = 0$. The assumption means that $\alpha_1\varphi_1(p) + \alpha_2\varphi_2(p) = 0$ for all $p \in \mathbb{P}^1$. We therefore try $p = p_0$, $p = p_1$, where $p_j(x) = x^j$, to learn more. We have

$$0 = \alpha_1\varphi_1(p_0) + \alpha_2\varphi_2(p_0) = \alpha_1 p_0(0) + \alpha_2 \int_0^1 p_0(x)\ dx = \alpha_1 + \alpha_2 \int_0^1 1\ dx = \alpha_1 + \alpha_2$$

and

$$0 = \alpha_1\varphi_1(p_1) + \alpha_2\varphi_2(p_1) = \alpha_1 p_1(0) + \alpha_2 \int_0^1 p_1(x)\ dx = 0 + \alpha_2 \int_0^1 x\ dx = \frac{\alpha_2}{2}.$$

This immediately gives $\alpha_2 = 0$ and then $\alpha_1 = 0$, as desired, thus the independence of (φ_1, φ_2). Since $\dim(\mathcal{V}') = \dim(\mathcal{V}) = \dim(\mathbb{P}^1) = 2$, the list (φ_1, φ_2) is therefore a basis for $\mathcal{V}'$.

(ii) Per the hint, we want the basis (p_1, p_2) to satisfy

$$\varphi_1(p_1) = 1, \qquad \varphi_1(p_2) = 0, \qquad \varphi_2(p_1) = 0, \quad \text{and} \quad \varphi_2(p_2) = 1.$$

Since p_1, $p_2 \in \mathbb{P}^1$, we can write $p_1(x) = a_0 + a_1 x$ and $p_2(x) = b_0 + b_1 x$. We want to use the four equations above to find the four unknowns a_0, a_1, b_0, and b_1. That is, we want

$$1 = \varphi_1(p_1) = p_1(0) = a_0,$$

$$0 = \varphi_1(p_2) = p_2(0) = b_0,$$

$$0 = \varphi_2(p_1) = \int_0^1 p_1(x)\ dx = \int_0^1 (a_0 + a_1 x)\ dx = \int_0^1 (1 + a_1 x)\ dx = 1 + \frac{a_1}{2},$$

and

$$1 = \varphi_2(p_2) = \int_0^1 p_2(x)\ dx = \int_0^1 (b_0 + b_1 x)\ dx = \int_0^1 (1 + b_1 x)\ dx = 1 + \frac{b_1}{2},$$

thus $a_1 = -2$ and $b_1 = 0$. The original basis, therefore, should be given by $p_1(x) = -2x + 1$ and $p_2(x) = x$.

20.7. By ??, the dual basis satisfies $v = \sum_{j=1}^{n} \varphi_j(v)v_j$ for all $v \in \mathcal{V}$. Since $\mathcal{B}$ is an isomorphism, we also have $v = \mathcal{B}\mathcal{B}^{-1}v$, where

$$\mathcal{B}^{-1}v = ((\mathcal{B}^{-1}v \cdot \mathbf{e}_1), \ldots, (\mathcal{B}^{-1}v \cdot \mathbf{e}_n)) = \sum_{j=1}^{n} (\mathcal{B}^{-1}v \cdot \mathbf{e}_j)\mathbf{e}_j.$$

Then

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

In particular, $\mathcal{B}\mathbf{e}_j = v_j$, and so we have two representations for v:

$$v = \sum_{j=1}^{n} (\mathcal{B}^{-1}v \cdot \mathbf{e}_j)v_j \quad \text{and} \quad v = \sum_{j=1}^{n} \varphi_j(v)v_j.$$

Equating these representations and using the independence of the list $(v_1, \ldots, v_n)$, we conclude

$$\varphi_j(v) = \mathcal{B}^{-1}v \cdot \mathbf{e}_j.$$

21.4. Following the hint, we start with a basis $(\varphi_1', \ldots, \varphi_n')$ for $\mathcal{V}'$, construct the dual basis $(\varphi_1'', \ldots, \varphi_n'')$ for $\mathcal{V}''$ relative to $(\varphi_1', \ldots, \varphi_n')$, and then write $\varphi_k'' = \mathcal{J}v_k$ for some $v_k \in \mathcal{V}$, where $\mathcal{J} \colon \mathcal{V} \to \mathcal{V}''$ is the canonical isomorphism. First we show that $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$, and then we check that $(\varphi_1', \ldots, \varphi_n')$ is the dual basis for $\mathcal{V}'$ relative to $(v_1, \ldots, v_n)$.

Since $\dim(\mathcal{V}) = n$, we only need to check independence of $(v_1, \ldots, v_n)$. Suppose that $\sum_{j=1}^{n} \alpha_j v_j = 0_{\mathcal{V}}$; we want $\alpha_j = 0$ for all j. Then

$$0_{\mathcal{V}' \to \mathbb{F}} = 0_{\mathcal{V}''} = \mathcal{J}0_{\mathcal{V}} = \mathcal{J}\sum_{j=1}^{n} \alpha_j v_j = \sum_{j=1}^{n} \alpha_j \mathcal{J}v_j = \sum_{j=1}^{n} \alpha_j \varphi_j''.$$

Since $(\varphi_1'', \ldots, \varphi_n'')$ is a basis for $\mathcal{V}''$, we have $\alpha_j = 0$ for all j, as desired.

Last, to check that $(\varphi_1', \ldots, \varphi_n')$ is the dual basis for $\mathcal{V}'$ relative to $(v_1, \ldots, v_n)$, we want to show that $\varphi_j'(v_k) = 0$ for $j \neq k$ and $\varphi_j'(v_j) = 1$. Since $\varphi_j'' = \mathcal{J}v_j$, we have $\varphi_j''(\varphi) = \varphi(v_j)$ for all j. In particular, $\varphi_j''(\varphi_k') = \varphi_k'(v_j)$. And since $(\varphi_1'', \ldots, \varphi_n'')$ is the dual basis for $\mathcal{V}''$ relative to $(\varphi_1', \ldots, \varphi_n')$, we have $\varphi_j''(\varphi_k') = 0$ for $j \neq k$ and $\varphi_j''(\varphi_j') = 1$. This proves the desired identities for $\varphi_j'(v_k)$.

22.3. (i) We have $\mathcal{T}v = \mathbf{0}_n$ if and only if $\sum_{j=1}^{n} \varphi_j(v)\mathbf{e}_j$. By independence, this happens if and only if $\varphi_j(v) = 0$ for all j, which is the same as $v \in \cap_{j=1}^{n} \ker(\varphi_j)$.

(ii) First we show that $\psi = \{(\mathcal{T}v, \varphi(v)) \mid v \in \mathcal{V}\}$ is a function. Assume that (w, α_1), $(w, \alpha_2) \in \psi$. We need to show that $\alpha_1 = \alpha_2$. Since $(w, \alpha_1) \in \psi$, there is $v_1 \in \mathcal{V}$ such that $w = \mathcal{T}v_1$ and $\alpha_1 = \varphi(v_1)$, Likewise, since $(w, \alpha_2) \in \psi$, there is $v_2 \in \mathcal{V}$ such that $w = \mathcal{T}v_2$ and $\alpha_2 = \varphi(v_2)$. Here the definition of ψ does not allow us to assume that $v_1 = v_2$. However, we do have $\mathcal{T}v_1 = w = Tv_2$, thus $\mathcal{T}(v_1 - v_2) = \mathbf{0}_n$, and so $v_1 - v_2 \in \ker(\mathcal{T})$. Since $\ker(\mathcal{T}) \subseteq \ker(\varphi)$, we have $\varphi(v_1 - v_2) = 0$, thus $\alpha_1 = \varphi(v_1) = \varphi(v_2) = \alpha_2$.

(iii) Let $w_1,\ w_2 \in \mathcal{T}(\mathcal{V})$ and write $w_1 = \mathcal{T}v_1$ and $w_2 = \mathcal{T}v_2$ for some $v_1,\ v_2 \in \mathcal{V}$. Then $w_1 + w_2 = \mathcal{T}(v_1 + v_2)$, so

$$\psi(w_1+w_2) = \psi\mathcal{T}(v_1+v_2) = \varphi(v_1+v_2) = \varphi(v_1)+\varphi(v_2) = \psi(\mathcal{T}v_1)+\psi(\mathcal{T}v_2) = \psi(w_1)+\psi(w_2).$$

Next, let $w \in \mathcal{T}(\mathcal{V})$ with $w = \mathcal{T}v$ for some $v \in \mathcal{V}$, and let $\alpha \in \mathbb{F}$. Then $\alpha w = \mathcal{T}(\alpha v)$, so

$$\psi(\alpha w) = \psi(\mathcal{T}(\alpha v)) = \varphi(\alpha v) = \alpha\varphi(v) = \alpha\psi(\mathcal{T}v) = \alpha\psi(w).$$

23.3. Let $n \geq 1$. Since $\mathcal{V}'$ is infinite-dimensional, there is an independent list $(\varphi_1, \ldots, \varphi_n)$ in $\mathcal{V}'$. Let $(v_1, \ldots, v_n)$ be the list in $\mathcal{V}$ constructed by Theorem 23.1. By Problem 23.2, this list $(v_1, \ldots, v_n)$ is independent. Since $n \geq 1$ was arbitrary, $\mathcal{V}$ is infinite-dimensional.

23.14. (i) We have

$$(\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 = A\mathbf{e}_1 - i\mathbf{e}_1 = \mathbf{e}_2 - i\mathbf{e}_1 \neq \mathbf{0}_2$$

and

$$(\mathcal{M}_A+i\mathcal{I}_{\mathbb{C}^2})(\mathcal{M}_A-i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 = (A+iI_2)(\mathbf{e}_2-i\mathbf{e}_1) = A\mathbf{e}_2+i\mathbf{e}_2-iA\mathbf{e}_1-i^2\mathbf{e}_1 = -\mathbf{e}_1+i\mathbf{e}_2-i\mathbf{e}_2+\mathbf{e}_1 = \mathbf{0}_2.$$

This shows that $(\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 = \mathbf{e}_2 - i\mathbf{e}_1 = (-i, 1)$ is an eigenvector of A (of $\mathcal{M}_A$) corresponding to the eigenvalue $-i$.

(ii) We redo the work of Example 23.13. The list $(\mathbf{e}_2, A\mathbf{e}_2, A^2\mathbf{e}_2)$ is dependent in $\mathbb{C}^2$. Specifically, $A\mathbf{e}_2 = -\mathbf{e}_1$, so $A^2\mathbf{e}_2 = -A\mathbf{e}_1 = -\mathbf{e}_2$. Then we have

$$\mathbf{0}_2 = -\mathbf{e}_2 + \mathbf{e}_2 = A^2\mathbf{e}_2 + \mathbf{e}_2 = (A^2 + I_2)\mathbf{e}_2 = (\mathcal{M}_A^2 + \mathcal{I}_{\mathbb{C}^2})\mathbf{e}_2.$$

Again, put $p(z) = z^2 + 1 = (z - i)(z + i)$, where we have intentionally written the factors in the opposite order to their appearance in Example 23.13. Then

$$\mathbf{0}_2 = p(\mathcal{M}_A)\mathbf{e}_2 = (A - iI_2)(A + iI_2)\mathbf{e}_2.$$

We check

$$(A + iI_2)\mathbf{e}_2 = A\mathbf{e}_2 + i\mathbf{e}_2 = -\mathbf{e}_1 + i\mathbf{e}_2 \neq \mathbf{0}_2$$

and

$$(A-iI_2)(A+iI_2)\mathbf{e}_2 = (A-iI_2)(-\mathbf{e}_1+i\mathbf{e}_2) = -A\mathbf{e}_1+i\mathbf{e}_1+iA\mathbf{e}_2-i^2\mathbf{e}_2 = -\mathbf{e}_2+i\mathbf{e}_1-i\mathbf{e}_1+\mathbf{e}_2 = \mathbf{0}_2.$$

This shows that $(A + iI_2)\mathbf{e}_2 = -\mathbf{e}_1 + i\mathbf{e}_2 = (-1, i)$ is an eigenvector of A corresponding to the eigenvalue i.

24.2. (i) Suppose that $\mathcal{T}v = \lambda v$ for some $\lambda \in \mathbb{F}$ and $v \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$. We want to show $\lambda = \lambda_j$ for some j. Write $v = \sum_{j=1}^{n}\alpha_j v_j$ for some $\alpha_j \in \mathbb{F}$. Then

$$\sum_{j=1}^{n} \lambda\alpha_j v_j = \lambda v = \mathcal{T}v = \mathcal{T}\sum_{j=1}^{n} \alpha_j v_j = \sum_{j=1}^{n} \alpha_j \mathcal{T}v_j = \sum_{j=1}^{n} \alpha_j \lambda_j v_j,$$

so

$$\sum_{j=1}^{n} \alpha_j(\lambda - \lambda_j)v_j = 0_{\mathcal{V}}.$$

Since $(v_1, \ldots, v_n)$ is independent, $\alpha_j(\lambda - \lambda_j) = 0$ for all j. Since $v \neq 0$, there is at least one k such that $\alpha_k \neq 0$. Then $\lambda - \lambda_k = 0$, so $\lambda = \lambda_k$.

(ii) Let v be a vector in the eigenspace corresponding to λ_k, so $\mathcal{T}v = \lambda_k v$. If $v = 0_{\mathcal{V}}$, then $v \in \operatorname{span}(v_k)$ already, so suppose $v \neq 0_{\mathcal{V}}$. Write $v = \sum_{j=1}^n \alpha_j v_j$, where since $v \neq 0_{\mathcal{V}}$, we have $\alpha_\ell \neq 0$ for at least one ℓ. Since $\mathcal{T}v = \lambda_k v$, we can repeat the calculations from the first part now with $\lambda = \lambda_k$ to obtain

$$\sum_{j=1}^n \alpha_j(\lambda_k - \lambda_j)v_j = 0_{\mathcal{V}}.$$

By the independence of $(v_1, \ldots, v_n)$, we have $\alpha_j(\lambda_k - \lambda_j) = 0$ for all j. For $j \neq k$, we have $\lambda_j \neq \lambda_k$, so $\alpha_j = 0$ for $j \neq k$. Thus $v = \alpha_k v_k \in \operatorname{span}(v_k)$.

24.3. Let $\dim(\mathcal{V}) = n$. If $\mathcal{V} = \{0_{\mathcal{V}}\}$, then $\mathbf{L}(\mathcal{V}) = \{0_{\mathcal{V}\to\mathcal{V}}\}$, and $0_{\mathcal{V}\to\mathcal{V}} = 0\mathcal{I}_{\mathcal{V}}$. So, suppose $n \geq 1$, and let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$. For $j = 1, \ldots, n$, there is $\lambda_j \in \mathbb{F}$ such that $\mathcal{T}v_j = \lambda_j v_j$, and likewise there is $\lambda \in \mathbb{F}$ such that $\mathcal{T}\sum_{j=1}^n v_j = \lambda \sum_{j=1}^n v_j$. (By the way, since $(v_1, \ldots, v_n)$ is independent, $0 \neq \sum_{j=1}^n 1v_j = \sum_{j=1}^n v_j$, so $\sum_{j=1}^n v_j$ is an eigenvector for $\mathcal{T}$ under these hypotheses.)

That is,

$$\sum_{j=1}^n \lambda v_j = \mathcal{T}\sum_{j=1}^n v_j = \sum_{j=1}^n \mathcal{T}v_j = \sum_{j=1}^n \lambda_j v_j,$$

and so

$$\sum_{j=1}^n (\lambda - \lambda_j)v_j = 0_{\mathcal{V}}.$$

By independence, $\lambda - \lambda_j = 0$ for all j, so $\lambda_j = \lambda$ for all j.

Now let $v \in \mathcal{V}$ and write $v = \sum_{j=1}^n \alpha_j v_j$. Then

$$\mathcal{T}v = \mathcal{T}\sum_{j=1}^n \alpha_j v_j = \sum_{j=1}^n \alpha_j \mathcal{T}v_j = \sum_{j=1}^n \alpha_j \lambda v_j = \lambda \sum_{j=1}^n \alpha_j v_j = \lambda v,$$

so $\mathcal{T} = \lambda \mathcal{I}_{\mathcal{V}}$.

24.5. (i) If $\operatorname{rank}(\mathcal{T}) = 0$, then $\dim(\mathcal{T}(\mathcal{V})) = 0$, so $\mathcal{T}(\mathcal{V}) = \{0_{\mathcal{W}}\}$. That is, $\mathcal{T}v = 0_{\mathcal{W}}$ for all $v \in \mathcal{V}$, so $\mathcal{T} = 0_{\mathcal{V}\to\mathcal{W}}$. Conversely, $0_{\mathcal{V}\to\mathcal{W}}(\mathcal{V}) = \{0_{\mathcal{W}}\}$, so $\dim(0_{\mathcal{V}\to\mathcal{W}}(\mathcal{V})) = 0$, and therefore $0_{\mathcal{V}\to\mathcal{W}}$ is finite-rank with $\operatorname{rank}(0_{\mathcal{V}\to\mathcal{W}}(\mathcal{V})) = 0$.

(ii) Since $\mathcal{V}$ is finite-dimensional, $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$ for some $v_j \in \mathcal{V}$, thus $\mathcal{T}(\mathcal{V}) = \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$, so $\mathcal{T}(\mathcal{V})$ is finite-dimensional, and therefore $\mathcal{T}$ is finite-rank.

(iii) Since $\mathcal{I}_{\mathcal{V}}(\mathcal{V}) = \mathcal{V}$, if $\mathcal{V}$ is infinite-dimensional, then so is $\mathcal{I}_{\mathcal{V}}(\mathcal{V})$, and therefore $\mathcal{I}_{\mathcal{V}}$ cannot be finite-rank.

24.6. Since $\mathcal{T}$ is finite-rank, the image $\mathcal{T}(\mathcal{U})$ is finite-dimensional. Say that $\operatorname{rank}(\mathcal{T}) = r$ and $(v_1, \ldots, v_r)$ is a basis for $\mathcal{T}(\mathcal{U})$. Let $w \in (\mathcal{S}\mathcal{T})(\mathcal{U})$. Then $w = \mathcal{S}(\mathcal{T}u)$ for some $u \in \mathcal{U}$. Since $\mathcal{T}u \in \mathcal{T}(\mathcal{U})$, we can write $\mathcal{T}u = \sum_{j=1}^r \alpha_j v_j$ for some $\alpha_j \in \mathbb{F}$, and then

$$w = \mathcal{S}\sum_{j=1}^r \alpha_j v_j = \sum_{j=1}^r \alpha_j \mathcal{S}v_j \in \operatorname{span}(\mathcal{S}v_1, \ldots, \mathcal{S}v_r).$$

That is, $(\mathcal{S}\mathcal{T})(\mathcal{U}) \subseteq \operatorname{span}(\mathcal{S}v_1, \ldots, \mathcal{S}v_r)$. Hence $(\mathcal{S}\mathcal{T})(\mathcal{U})$ is finite-dimensional, in particular with $\dim((\mathcal{S}\mathcal{T})(\mathcal{U})) \le r$, and so $\mathcal{S}\mathcal{T}$ is finite-rank. In particular, $\operatorname{rank}(\mathcal{S}\mathcal{T}) \le r$.

This inequality may be strict; for example, if $\mathcal{S} = 0_{\mathcal{V}\to\mathcal{W}}$, then $\mathcal{S}(\mathcal{T}(\mathcal{U})) = \{0_{\mathcal{W}}\}$. However, suppose that $\mathcal{S}$ is invertible. Since $(v_1, \ldots, v_r)$ is independent in $\mathcal{V}$ and $\mathcal{S}\colon \mathcal{V} \to \mathcal{W}$ is injective, the list $(\mathcal{S}v_1, \ldots, \mathcal{S}v_r)$ is also independent in $\mathcal{W}$. In particular, $(\mathcal{S}v_1, \ldots, \mathcal{S}v_r)$ is independent in $\mathcal{S}(\mathcal{T}(\mathcal{U}))$, and since $\dim(\mathcal{S}(\mathcal{T}(\mathcal{U}))) \le r$, it must be the case that $\dim(\mathcal{S}(\mathcal{T}(\mathcal{U}))) = r$. That is, when $\mathcal{S}$ is invertible, $\operatorname{rank}(\mathcal{S}\mathcal{T}) = \operatorname{rank}(\mathcal{T})$. Another case of equality is when $r = 0$, for then $\mathcal{T}(\mathcal{U}) = \{0_{\mathcal{V}}\}$, and so $\mathcal{S}(\mathcal{T}(\mathcal{U})) = \{0_{\mathcal{W}}\}$, in which case $\operatorname{rank}(\mathcal{S}\mathcal{T}) = 0$ as well.

24.9. (i) The list (φ_1, φ_2) is independent: if $\alpha_1\varphi_1 + \alpha_2\varphi_2 = 0_{\mathbb{F}^2\to\mathbb{F}}$, evaluate this linear combination at $\mathbf{e}_1$ and $\mathbf{e}_2$ to conclude $\alpha_1 = \alpha_2 = 0$. The list $(\mathbf{w}_1, \mathbf{w}_2)$ is obviously dependent because it contains a repeated vector. The list $(\mathcal{T}_1, \mathcal{T}_2)$ is independent: assume $\mathcal{S} := \alpha_1\mathcal{T}_1 + \alpha_2\mathcal{T}_2 = 0_{\mathbb{F}^2\to\mathbb{F}^2}$ and compute

$$\mathbf{0}_2 = \mathcal{S}\mathbf{e}_1 = \alpha_1(\mathbf{e}_1 \cdot \mathbf{e}_1)\mathbf{e}_1 + \alpha_2(\mathbf{e}_1 \cdot \mathbf{e}_2)\mathbf{e}_1 = \alpha_1\mathbf{e}_1$$

and

$$\mathbf{0}_2 = \mathcal{S}\mathbf{e}_2 = \alpha_1(\mathbf{e}_2 \cdot \mathbf{e}_1)\mathbf{e}_1 + \alpha_2(\mathbf{e}_2 \cdot \mathbf{e}_2)\mathbf{e}_1 = \alpha_2\mathbf{e}_1$$

to conclude $\alpha_1 = \alpha_2 = 0$.

(ii) The list (φ_1, φ_2) is obviously dependent because it contains a repeated functional. The list $(\mathbf{w}_1, \mathbf{w}_2)$ is independent by properties of the standard basis vectors. The list $(\mathcal{T}_1, \mathcal{T}_2)$ is independent: assume $\mathcal{S} := \alpha_1\mathcal{T}_1 + \alpha_2\mathcal{T}_2 = 0_{\mathbb{F}^2\to\mathbb{F}^2}$ and compute

$$\mathbf{0}_2 = \mathcal{S}\mathbf{e}_1 = \alpha_1(\mathbf{e}_1 \cdot \mathbf{e}_1)\mathbf{e}_1 + \alpha_2(\mathbf{e}_1 \cdot \mathbf{e}_1)\mathbf{e}_2 = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2,$$

thus $\alpha_1 = \alpha_2 = 0$.

(iii) The list $(\varphi_1, \varphi_2, \varphi_3)$ is obviously dependent because it contains a repeated functional. The list $(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$ is dependent because $\mathbf{w}_3$ is a linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$. The list $(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)$ is independent: assume $\mathcal{S} := \alpha_1\mathcal{T}_1 + \alpha_2\mathcal{T}_2 + \alpha_3\mathcal{T}_3 = 0_{\mathbb{F}^2\to\mathbb{F}^2}$ and compute

$$\mathbf{0}_2 = \mathcal{S}\mathbf{e}_1 = \alpha_1(\mathbf{e}_1 \cdot \mathbf{e}_1)\mathbf{e}_1 + \alpha_2(\mathbf{e}_1 \cdot \mathbf{e}_2)\mathbf{e}_2 + \alpha_3((\mathbf{e}_1 \cdot \mathbf{e}_1) + (\mathbf{e}_1 \cdot \mathbf{e}_2))(\mathbf{e}_1 + \mathbf{e}_2) = \alpha_1\mathbf{e}_1 + \alpha_3(\mathbf{e}_1 + \mathbf{e}_2)$$

and

$$\mathbf{0}_2 = \mathcal{S}\mathbf{e}_2 = \alpha_1(\mathbf{e}_2 \cdot \mathbf{e}_1)\mathbf{e}_1 + \alpha_2(\mathbf{e}_2 \cdot \mathbf{e}_2)\mathbf{e}_2 + \alpha_3((\mathbf{e}_2 \cdot \mathbf{e}_1) + (\mathbf{e}_2 \cdot \mathbf{e}_2))(\mathbf{e}_1 + \mathbf{e}_2) = \alpha_2\mathbf{e}_2 + \alpha_3(\mathbf{e}_1 + \mathbf{e}_2).$$

This turns into the upper-triangular systems

$$A\begin{bmatrix}\alpha_1\\ \alpha_3\end{bmatrix} = A\begin{bmatrix}\alpha_2\\ \alpha_3\end{bmatrix} = \mathbf{0}_2, \qquad A = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix},$$

thus $\alpha_1 = \alpha_2 = \alpha_3 = 0$

25.6. (i) If $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is surjective, then $\mathcal{T}(\mathcal{V}) = \mathcal{W}$, so $\dim(\mathcal{T}(\mathcal{V})) = \dim(\mathcal{W})$. Then $\dim(\mathcal{V}) = \dim(\ker(\mathcal{T})) + \dim(\mathcal{T}(\mathcal{V})) \ge \dim(\mathcal{T}(\mathcal{V})) = \dim(\mathcal{W})$.

(ii) If $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is injective, then $\ker(\mathcal{T}) = \{0_\mathcal{V}\}$, so $\dim(\ker(\mathcal{T})) = 0$. Since $\mathcal{T}(\mathcal{V})$ is a subspace of $\mathcal{W}$, we have $\dim(\mathcal{T}(\mathcal{V})) \le \dim(\mathcal{W})$. Then $\dim(\mathcal{V}) = \dim(\ker(\mathcal{T})) + \dim(\mathcal{T}(\mathcal{V})) = \dim(\mathcal{T}(\mathcal{V})) \le \dim(\mathcal{W}) < \dim(\mathcal{V})$.

25.7. Let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. If $\mathcal{T}$ is injective, then $\ker(\mathcal{T}) = \{0_\mathcal{V}\}$, and so $\dim(\ker(\mathcal{T})) = 0$. Then $\dim(\mathcal{T}(\mathcal{V})) = \dim(\mathcal{V}) - \dim(\ker(\mathcal{T})) = \dim(\mathcal{V}) = \dim(\mathcal{W})$, so $\mathcal{T}(\mathcal{V}) = \mathcal{W}$, and therefore $\mathcal{T}$ is surjective. If $\mathcal{T}$ is surjective, then $\mathcal{T}(\mathcal{V}) = \mathcal{W}$, so $\dim(\mathcal{T}(\mathcal{V})) = \dim(\mathcal{W}) = \dim(\mathcal{V})$. Then $\dim(\ker(\mathcal{T})) = \dim(\mathcal{V}) - \dim(\mathcal{T}(\mathcal{V})) = 0$, so $\ker(\mathcal{T}) = \{0_\mathcal{V}\}$, and therefore $\mathcal{T}$ is injective.

25.16. We have

$$\mathcal{V}/\{0_\mathcal{V}\} = \{v + \{0_\mathcal{V}\} \mid v \in \mathcal{V}\},$$

and

$$v + \{0_\mathcal{V}\} = \{v + u \mid u \in \{0_\mathcal{V}\}\} = \{v\}.$$

So,

$$\mathcal{V}/\{0_\mathcal{V}\} = \{\{v\} \mid v \in \mathcal{V}\},$$

and we should think of this as effectively $\mathcal{V}$.

Next, we have

$$\mathcal{V}/\mathcal{V} = \{v + \mathcal{V} \mid v \in \mathcal{V}\},$$

and

$$v + \mathcal{V} = \{v + u \mid u \in \mathcal{V}\} \subseteq \mathcal{V}.$$

But also, given $w \in \mathcal{V}$, we have $w = v + (w - v) \in v + \mathcal{V}$, so $\mathcal{V} \subseteq v + \mathcal{V}$ for each $v \in \mathcal{V}$. Thus $v + \mathcal{V} = \mathcal{V}$ for all $v \in \mathcal{V}$, and so

$$\mathcal{V}/\mathcal{V} = \{\mathcal{V}\}.$$

Since $0_{\mathcal{V}/\mathcal{V}} = \mathcal{V}$, we should think of $\mathcal{V}/\mathcal{V}$ as the trivial space.

25.17. Suppose that $v_1 + \mathcal{U} \neq v_2 + \mathcal{U}$; we will show $(v_1 + \mathcal{U}) \cap (v_2 + \mathcal{U}) = \varnothing$. Let $v \in (v_1 + \mathcal{U}) \cap (v_2 + \mathcal{U})$. Then $v \in v_j + \mathcal{U}$ for $j = 1$, 2, so $v = v_1 + u_1 = v_2 + u_2$ for some u_1, $u_2 \in \mathcal{U}$. Then $v_1 - v_2 = u_2 - u_1 \in \mathcal{U}$, so $v_1 = v_2 + (v_1 - v_2) \in v_2 + \mathcal{U}$. For $u \in \mathcal{U}$, we therefore have $v_1 + u = v_2 + (v_1 - v_2 + u) \in v_2 + \mathcal{U}$, so $v_1 + \mathcal{U} \subseteq v_2 + \mathcal{U}$. Reversing the roles of v_1 and v_2 here gives $v_2 + \mathcal{U} \subseteq v_1 + \mathcal{U}$, and so $v_1 + \mathcal{U} = v_2 + \mathcal{U}$ after all, a contradiction.

Now we characterize when these situations hold. First suppose that $v_1 + \mathcal{U} = v_2 + \mathcal{U}$. Then $v_1 = v_1 + 0_\mathcal{V} \in v_1 + \mathcal{U} = v_2 + \mathcal{U}$, so there is $u \in \mathcal{U}$ such that $v_1 = v_2 + u$. Then $v_1 - v_2 = u \in \mathcal{U}$. So, if $v_1 + \mathcal{U} = v_2 + \mathcal{U}$, then $v_1 - v_2 \in \mathcal{U}$. Conversely, the previous paragraph shows that if $v_1 - v_2 \in \mathcal{U}$, then $v_1 + \mathcal{U} = v_2 + \mathcal{U}$.

So, $v_1 + \mathcal{U} = v_2 + \mathcal{U}$ if and only if $v_1 - v_2 \in \mathcal{U}$. If $(v_1 + \mathcal{U}) \cap (v_2 + \mathcal{U}) = \varnothing$, then by the prior work, it must be the case that $v_1 + \mathcal{U} \neq v_2 + \mathcal{U}$, and so $v_1 - v_2 \notin \mathcal{U}$. Conversely, if $v_1 - v_2 \notin \mathcal{U}$, then $v_1 + \mathcal{U} \neq v_2 + \mathcal{U}$, and so, again by the prior work, it must be the case that $(v_1 + \mathcal{U}) \cap (v_2 + \mathcal{U}) = \varnothing$. So, $(v_1 + \mathcal{U}) \cap (v_2 + \mathcal{U}) = \varnothing$ if and only if $v_1 - v_2 \notin \mathcal{U}$.

26.3. (i) We know that

$$(a_k) + \mathcal{U} = \{(c_k) \in \mathcal{V} \mid (a_k) - (c_k) \in \mathcal{U}\},$$

and we have $(a_k) - (c_k) = (a_k - c_k) \in \mathcal{U}$ if and only if

$$\lim_{k\to\infty} (a_k - c_k) = 0.$$

Since $\lim_{k\to\infty} a_k$ and $\lim_{k\to\infty} c_k$ exist, we have

$$\lim_{k\to\infty} (a_k - c_k) = 0$$

if and only if

$$\lim_{k\to\infty} a_k = \lim_{k\to\infty} c_k.$$

(ii) The sequences $(c_k) \in (1/k+1) + \mathcal{U}$ are precisely those that satisfy

$$\lim_{k\to\infty} c_k = \lim_{k\to\infty} \left(\frac{1}{k} + 1\right) = 1.$$

For example, (1), $(1/k^2+1) \in (1/k+1) + \mathcal{U}$.

26.15. (i) The first isomorphism theorem tells us that $\mathcal{V}/\ker(\mathcal{T})$ and $\mathcal{T}(\mathcal{V})$ are isomorphic, and since $\mathcal{V}$ is finite-dimensional, $\mathcal{T}(\mathcal{V})$ is also finite-dimensional, thus $\mathcal{V}/\ker(\mathcal{T})$ is finite-dimensional. (We could also use Theorem 26.11 to conclude that $\mathcal{V}/\ker(\mathcal{T})$ is finite-dimensional.) The dimensions of $\mathcal{V}/\ker(\mathcal{T})$ and $\mathcal{T}(\mathcal{V})$ are therefore defined and, by that isomorphism, $\dim(\mathcal{V}/\ker(\mathcal{T})) = \dim(\mathcal{T}(\mathcal{V}))$. Now we (must) use Theorem 26.11 to obtain

$$\dim(\mathcal{V}/\ker(\mathcal{T})) = \dim(\mathcal{V}) - \dim(\ker(\mathcal{T})),$$

thus

$$\dim(\mathcal{V}) = \dim(\ker(\mathcal{T})) + \dim(\mathcal{T}(\mathcal{V})),$$

which is rank–nullity.

(ii) Rank–nullity tells us that $\dim(\mathcal{V}) - \dim(\ker(\mathcal{T})) = \dim(\mathcal{T}(\mathcal{V}))$, and Theorem 26.11 tells us that $\dim(\mathcal{V}/\ker(\mathcal{T})) = \dim(\mathcal{V}) - \dim(\ker(\mathcal{T}))$. So, $\dim(\mathcal{V}/\ker(\mathcal{T})) = \dim(\mathcal{T}(\mathcal{V}))$, and therefore $\mathcal{V}/\ker(\mathcal{T})$ and $\mathcal{T}(\mathcal{V})$ are isomorphic. This does not quite prove the first isomorphism theorem for two reasons. First, this argument only works when $\mathcal{V}$ is finite-dimensional, which the first isomorphism theorem does not assume. Second, the first isomorphism theorem establishes that $\mathcal{V}/\ker(\mathcal{T})$ and $\mathcal{T}(\mathcal{V})$ are isomorphic via a specific (and "natural") isomorphism, and this argument does not prove that that operator is an isomorphism.

26.16. In linear algebraic language, the coset $f + \ker(\mathcal{T})$ consists of all functions $f + g$ with $g \in \ker(\mathcal{T})$. Since the functions in $\ker(\mathcal{T})$ are the constant functions, in calculus language, the coset $f + \ker(\mathcal{T})$ consists of all functions that differ from f by a constant, equivalently, all functions that have the same derivative as f. In further calculus language, the operator $\mathcal{T}$ just pairs all functions that have the same derivative as f with f'.

27.6. Take $f(x) = 1$. Then $f \neq 0$ but $f' = 0$, so $\langle f, f \rangle = \int_0^1 0 \, dx = 0$.

28.13. Put $\varphi_j(v) := \langle v, u_j \rangle$. We want to show that $\varphi_j(u_j) = 1$ and $\varphi_j(u_k) = 0$ for $j \neq k$. This follows immediately from the definition of orthonormality. (We are not explicitly using the assumption that $(u_1, \ldots, u_n)$ is an orthonormal *basis* for $\mathcal{V}$. Rather, this assumption gives $\dim(\mathcal{V}) = \dim(\mathcal{V}') = n$, so the dual basis for $\mathcal{V}'$ will have length n.)

28.14. (i) Linearity in the first slot is a consequence of the linearity of the φ_j:

$$\langle u+v, w \rangle = \sum_{j=1}^{n} \varphi_j(u+v)\overline{\varphi_j(w)} = \sum_{j=1}^{n} (\varphi_j(u)+\varphi_j(v))\overline{\varphi_j(w)} = \sum_{j=1}^{n} \varphi_j(u)\overline{\varphi_j(w)} + \sum_{j=1}^{n} \varphi_j(v)\overline{\varphi_j(w)} = \langle u, w \rangle + \langle v, w \rangle$$

and

$$\langle \alpha v, w \rangle = \sum_{j=1}^{n} \varphi_j(\alpha v)\overline{\varphi_j(w)} = \sum_{j=1}^{n} \alpha \varphi_j(v)\overline{\varphi_j(w)} = \alpha \sum_{j=1}^{n} \varphi_j(v)\overline{\varphi_j(w)} = \alpha \langle v, w \rangle .$$

Conjugate linearity is a consequence of algebraic properties of the complex conjugate:

$$\langle w, v \rangle = \sum_{j=1}^{n} \varphi_j(w)\overline{\varphi_j(v)} = \sum_{j=1}^{n} \overline{\overline{\varphi_j(w)}\varphi_j(v)} = \overline{\sum_{j=1}^{n} \overline{\varphi_j(w)}\varphi_j(v)} = \overline{\langle v, w \rangle}.$$

Nonnegativity is also a consequence of properties of the conjugate:

$$\langle v, v \rangle = \sum_{j=1}^{n} \varphi_j(v)\overline{\varphi_j(v)} = \sum_{j=1}^{n} |\varphi_j(v)|^2 \geq 0.$$

Finally, for definiteness, assume $\langle v, v \rangle = 0$, so $\sum_{j=1}^{n} |\varphi_j(v)|^2 = 0$. This is a sum of nonnegative terms, so $|\varphi_j(v)|^2 = 0$ for all j, thus $\varphi_j(v) = 0$ for all j, and so $v = \sum_{j=1}^{n} \varphi_j(v) v_j = 0_{\mathcal{V}}$.

(ii) We have

$$\langle v_j, v_k \rangle = \sum_{\ell=1}^{n} \varphi_\ell(v_j)\overline{\varphi_\ell(v_k)},$$

where $\varphi_\ell(v_j) = 0$ for $j \neq \ell$, and $\varphi_\ell(v_k) = 0$ for $\ell \neq k$. If $j \neq k$, at least one of the factors $\varphi_\ell(v_j)$ or $\varphi_\ell(v_k)$ will always be 0, regardless of the value of ℓ, and so in this case the sum is 0. For $j = k$, the product $\varphi_\ell(v_j)\overline{\varphi_\ell(v_k)} = \varphi_\ell(v_j)\overline{\varphi_\ell(v_j)}$ will be 0 for $j \neq \ell$ and 1 for $j = \ell$, and so in this case the sum is 1.

29.10. Here we are assuming $w \neq 0_{\mathcal{V}}$. The inequality appears in estimating

$$\|v\|^2 = \|\mathcal{P}_w v\|^2 + \|v - \mathcal{P}_w v\|^2 \geq \|\mathcal{P}_w v\|^2 .$$

The inequality is an equality if and only if $\|v - \mathcal{P}_w v\|^2 = 0$, which is equivalent to $v = \mathcal{P}_w v$.

Per the hint, if $v = \mathcal{P}_w v$, then $v \in \operatorname{span}(w)$, since $\mathcal{P}_w v = (\langle v, w \rangle / \|w\|^2) w \in \operatorname{span}(w)$. Conversely, if $v \in \operatorname{span}(w)$, then $v = \alpha w$ for some $\alpha \in \mathbb{F}$. Then

$$\mathcal{P}_w v = \frac{\langle v, w \rangle}{\|w\|^2} w = \frac{\langle \alpha w, w \rangle}{\|w\|^2} w = \alpha \frac{\langle w, w \rangle}{\|w\|^2} w = \alpha w = v.$$

We conclude that equality holds in the Cauchy–Schwarz inequality $|\langle v, w\rangle| \le \|v\|\,\|w\|$ when $w \ne 0_{\mathcal{V}}$ if and only if $v \in \text{span}(w)$. This is equivalent to the list (v, w) being linearly dependent with $w \ne 0_{\mathcal{V}}$. Since equality also holds when $w = 0_{\mathcal{V}}$, we can summarize this work as follows: equality holds in the Cauchy–Schwarz inequality $|\langle v, w\rangle| \le \|v\|\,\|w\|$ if and only if the list (v, w) is dependent.

30.4. By definition, $g \in \mathcal{U}^\perp$ if and only if $\langle g, f\rangle = 0$ for all $f \in \mathcal{U}$. That is, $g \in \mathcal{U}^\perp$ if and only if $\int_{-1}^1 f(x)g(x)\,dx = 0$ for all $f \in \mathcal{U}$. If $f \in \mathcal{U}$, then f is even, so if $g \in \mathcal{C}([-1,1])$ is odd, then so is fg, and therefore $\int_{-1}^1 f(x)g(x)\,dx = 0$. Thus all odd functions in $\mathcal{C}([-1,1])$ are in $\mathcal{U}^\perp$.

Is every function in $\mathcal{U}^\perp$ odd? Write $g \in \mathcal{U}^\perp$ as $g = g_e + g_o$, where g_e is even and g_o is odd. For all $f \in \mathcal{U}$, we have

$$0 = \langle g, f\rangle = \langle g_e + g_o, f\rangle = \langle g_e, f\rangle + \langle g_o, f\rangle = \langle g_e, f\rangle,$$

since $\langle g_o, f\rangle = 0$ because its integrand $g_o f$ is odd. Take $f = g_e$ to obtain $\langle g_e, g_e\rangle = 0$, so $g_e = 0$, and therefore $g = g_o$ is odd.

30.5. (i) First, $\langle 0_{\mathcal{V}}, u\rangle = 0$ for all $u \in \mathcal{U}$, so $0_{\mathcal{V}} \in \mathcal{U}^\perp$. Second, let $u_1^\perp$, $u_2^\perp \in \mathcal{U}^\perp$, so $\langle u_1^\perp, u\rangle = \langle u_2^\perp, u\rangle = 0$ for all $u \in \mathcal{U}$. Then

$$\langle u_1^\perp + u_2^\perp, u\rangle = \langle u_1^\perp, u\rangle + \langle u_2^\perp, u\rangle = 0$$

for all $u \in \mathcal{U}$, so $u_1^\perp + u_2^\perp \in \mathcal{U}$. Finally, let $\alpha \in \mathbb{F}$ and $u^\perp \in \mathcal{U}^\perp$, so $\langle u^\perp, u\rangle = 0$ for all $u \in \mathcal{U}$. Then

$$\langle \alpha u^\perp, u\rangle = \alpha \langle u^\perp, u\rangle = 0$$

for all $u \in \mathcal{U}$, so $\alpha u^\perp \in \mathcal{U}^\perp$.

(ii) Let $v \in \mathcal{U} \cap \mathcal{U}^\perp$. Then $v \in \mathcal{U}$ and, since $v \in \mathcal{U}^\perp$, we have $\langle v, u\rangle = 0$ for all $u \in \mathcal{U}$. Taking $v = u$, we obtain $\langle v, v\rangle = 0$, thus $v = 0_{\mathcal{V}}$.

(iii) If $v \in \mathcal{V}^\perp$, then $\langle v, w\rangle = 0$ for all $w \in \mathcal{V}$; taking $w = v$, we have $\langle v, v\rangle = 0$, thus $v = 0_{\mathcal{V}}$. Since we already know $0_{\mathcal{V}} \in \mathcal{V}^\perp$, we have $\mathcal{V}^\perp = \{0_{\mathcal{V}}\}$.

Next, $\langle v, 0_{\mathcal{V}}\rangle = 0$ for all $v \in \mathcal{V}$, so $\mathcal{V} \subseteq \{0_{\mathcal{V}}\}^\perp \subseteq \mathcal{V}$, thus $\{0_{\mathcal{V}}\}^\perp = \mathcal{V}$.

(iv) First let $v \in (\text{span}(v_1, \ldots, v_n))^\perp$. Then $\langle v, w\rangle = 0$ for all $w \in \text{span}(v_1, \ldots, v_n)$. Taking $w = v_j$, we have $\langle v, v_j\rangle = 0$ for each j.

Conversely, suppose $\langle w, v_j\rangle = 0$ for all j. Let $v \in \text{span}(v_1, \ldots, v_n)$. We show $\langle w, v\rangle = 0$. By definition of span, $v = \sum_{j=1}^n \alpha_j v_j$ for some $\alpha_j \in \mathbb{F}$. Then

$$\langle w, v\rangle = \left\langle w, \sum_{j=1}^n \alpha_j v_j \right\rangle = \sum_{j=1}^n \overline{\alpha_j}\,\langle w, v_j\rangle = 0.$$

31.13. Let $\mathcal{U}$ be the space of all even functions in $\mathcal{C}([-1,1])$. Problem 30.4 showed that $\mathcal{U}^\perp$ is the space of all odd functions in $\mathcal{C}([-1,1])$. The hint to that problem showed that any $f \in \mathcal{C}([-1,1])$ can be written as $f = f_e + f_o$, with $f_e \in \mathcal{U}$ and $f_o \in \mathcal{U}^\perp$. This gives $\mathcal{C}([-1,1]) = \mathcal{U} \oplus \mathcal{U}^\perp$, and Theorem 31.11 then implies $\mathcal{U}^{\perp\perp} = \mathcal{U}$.

If $h \in \mathcal{C}([-1,1])$ has the property that $\int_{-1}^1 g(x)h(x)\ dx = 0$ for all $g \in \mathcal{C}([-1,1])$ such that $\int_{-1}^1 f(x)g(x)\ dx = 0$ for all even $f \in \mathcal{C}([-1,1])$, then $\langle h, g\rangle = 0$ for all $g \in \mathcal{C}([-1,1])$ such that $\langle f, g\rangle = 0$ for all even functions in $\mathcal{C}([-1,1])$. That is, $\langle h, g\rangle = 0$ for all $g \in \mathcal{U}^\perp$, and so $h \in \mathcal{U}^{\perp\perp}$. Thus h is even.

34.9. We have

$$\langle \mathcal{T}f, g\rangle = \sum_{k=1}^\infty (\mathcal{T}f)(k)\overline{g(k)} = \sum_{k=1}^\infty f(k+1)\overline{g(k)} = f(2)\overline{g(1)} + \sum_{k=2}^\infty f(k+1)\overline{g(k)}.$$

Per the hint (with $m = 2$ and $n = 1$), we rewrite

$$\sum_{k=2}^\infty f(k+1)\overline{g(k)} = \sum_{k=3}^\infty f(k)\overline{g(k-1)}.$$

Then

$$\langle \mathcal{T}f, g\rangle = (f(1)\cdot\overline{0}) + f(2)\overline{g(1)} + \sum_{k=3}^\infty f(k)\overline{g(k-1)} = \sum_{k=1}^\infty f(k)\overline{(\mathcal{T}^*g)(k)}$$

with $\mathcal{T}^*g$ as defined in the problem statement.

34.11. We have $\langle \mathcal{T}e_j, h\rangle = \langle e_j, \mathcal{S}h\rangle$, where

$$\langle \mathcal{T}e_j, h\rangle = \langle \langle e_j, h\rangle h, h\rangle = \langle e_j, h\rangle \langle h, h\rangle = h(j)\,\|h\|^2 = \frac{\|h\|^2}{j}$$

and

$$\langle e_j, \mathcal{S}h\rangle = (\mathcal{S}h)(j).$$

Thus

$$(\mathcal{S}h)(j) = \frac{\|h\|^2}{j} \neq 0$$

for all j, and so $\mathcal{S}h \notin \mathcal{U}$, a contradiction.

34.12. (i) We have

$$\langle \mathcal{T}u, v\rangle = \langle \langle u, v_0\rangle v_0, v\rangle = \langle u, v_0\rangle \langle v_0, v\rangle = \left\langle u, \overline{\langle v_0, v\rangle} v_0\right\rangle = \langle u, \langle v, v_0\rangle v_0\rangle = \langle u, \mathcal{S}v\rangle.$$

If $\mathcal{S}$ is the adjoint of $\mathcal{T}$, then $\mathcal{S}v \in \mathcal{U}$ for all $v \in \mathcal{V}$. However,

$$\mathcal{S}v_0 = \langle v_0, v_0\rangle v_0 = \|v_0\|^2 v_0.$$

If $\mathcal{S}v_0 \in \mathcal{U}$, then $v_0 = \|v_0\|^{-2}\mathcal{S}v_0 \in \mathcal{U}$ as well. (Here we may divide by $\|v_0\|$ since $v_0 \neq 0_{\mathcal{V}}$.) This is impossible.

(ii) As calculated above, we know that

$$\langle \mathcal{T}u, v\rangle = \langle \langle u, v_0\rangle v_0, v\rangle = \langle u, v_0\rangle \langle v_0, v\rangle = \left\langle u, \overline{\langle v_0, v\rangle} v_0 \right\rangle = \langle u, \langle v, v_0\rangle v_0\rangle$$

for all $u \in \mathcal{U}$ and $v \in \mathcal{V}$. Since $v_0 = \mathcal{P}_{\mathcal{U}} v_0 + u^\perp$, where $u^\perp := v_0 - \mathcal{P}_{\mathcal{U}} v_0 \in \mathcal{U}^\perp$, we have

$$\langle u, \langle v, v_0\rangle v_0\rangle = \langle u, \langle v, v_0\rangle \mathcal{P}_{\mathcal{U}} v_0\rangle + \left\langle u, \langle v, v_0\rangle u^\perp \right\rangle,$$

with

$$\left\langle u, \langle v, v_0\rangle u^\perp \right\rangle = \langle v, v_0\rangle \left\langle u, u^\perp \right\rangle = 0$$

since $u \in \mathcal{U}$ and $u^\perp \in \mathcal{U}^\perp$. We conclude that

$$\langle \mathcal{T}u, v\rangle = \langle u, \langle v, v_0\rangle \mathcal{P}_{\mathcal{U}} v_0\rangle$$

and that $\langle \cdot, v_0\rangle \mathcal{P}_{\mathcal{U}} v_0 \in \mathbf{L}(\mathcal{V}, \mathcal{U})$, thus $\mathcal{T}^* = \langle \cdot, v_0\rangle \mathcal{P}_{\mathcal{U}} v_0$.

(iii) If $\ell^2 = \mathcal{U} \oplus \mathcal{U}^\perp$, then $\mathcal{U}^{\perp\perp} = \mathcal{U}$ by Theorem 31.11. But in Example 30.8, we checked that $\mathcal{U}^{\perp\perp} = \ell^2 \neq \mathcal{U}$.

36.2. (i) We have

$$\langle \mathcal{T}v, w\rangle = \langle v, \mathcal{T}^* w\rangle = \langle v, \alpha \mathcal{T} w\rangle = \overline{\alpha} \langle v, \mathcal{T} w\rangle = \overline{\alpha} ip\mathcal{T}^* vw = \overline{\alpha} \langle \alpha \mathcal{T} v, w\rangle = \alpha\overline{\alpha} \langle \mathcal{T} v, w\rangle = |\alpha|^2 \langle \mathcal{T} v, w\rangle.$$

With $w = \mathcal{T}v$, we have

$$\|\mathcal{T}v\|^2 = \langle \mathcal{T}v, \mathcal{T}v\rangle = |\alpha|^2 \langle \mathcal{T}v, \mathcal{T}v\rangle = |\alpha|^2 \|\mathcal{T}v\|^2.$$

Since $\mathcal{T}$ is nonzero, $\mathcal{T}v \neq 0_{\mathcal{V}}$ for at least one v; thus $1 = |\alpha|^2$, and so $|\alpha| = 1$.

(ii) The conjugate transpose of

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

is

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

and the conjugate transpose of

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

is

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

36.3. Let $u^\perp \in \mathcal{U}^\perp$. Our goal is to show that $\left\langle \mathcal{T}^* u^\perp, u \right\rangle = 0$ for all $u \in \mathcal{U}$. We have $\left\langle \mathcal{T}^* u^\perp, u \right\rangle = \left\langle u^\perp, \mathcal{T}u \right\rangle$, and we are assuming $\mathcal{T}u \in \mathcal{U}$. Thus $\left\langle u^\perp, \mathcal{T}u \right\rangle = 0$.

37.5. (i) We have

$$[\mathcal{B}_{\mathbb{P}^1}^{-1}\mathcal{T}\mathcal{B}_{\mathbb{P}^2}]_{\mathcal{B}_{\mathbb{P}^2}\leftarrow\mathcal{B}_{\mathbb{P}^1}}[\mathcal{B}_{\mathbb{P}^2}^{-1}\mathcal{S}\mathcal{B}_{\mathbb{P}^1}]_{\mathcal{B}_{\mathbb{P}^1}\leftarrow\mathcal{B}_{\mathbb{P}^2}} = \begin{bmatrix}0&1&0\\0&0&2\end{bmatrix}\begin{bmatrix}0&0\\1&0\\0&1/2\end{bmatrix} = \begin{bmatrix}1&0\\0&1\end{bmatrix} = I_2,$$

which encodes the fact that $\mathcal{T}\mathcal{S}p = p$, and

$$[\mathcal{B}_{\mathbb{P}^2}^{-1}\mathcal{S}\mathcal{B}_{\mathbb{P}^1}]_{\mathcal{B}_{\mathbb{P}^1}\leftarrow\mathcal{B}_{\mathbb{P}^2}}[\mathcal{B}_{\mathbb{P}^1}^{-1}\mathcal{T}\mathcal{B}_{\mathbb{P}^2}]_{\mathcal{B}_{\mathbb{P}^2}\leftarrow\mathcal{B}_{\mathbb{P}^1}} = \begin{bmatrix}0&0\\1&0\\0&1/2\end{bmatrix}\begin{bmatrix}0&1&0\\0&0&2\end{bmatrix} = \begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix} \neq I_3,$$

which encodes the reality that $\mathcal{S}\mathcal{T}p = p - p(0) \neq p$ unless $p(0) = 0$.

(ii) Let $p \in \mathbb{P}^1$ and write $p(x) = a_1x + a_0$. If we can find α, $\beta \in \mathbb{F}$ such that

$$a_1x + a_0 = \alpha(x+1) + \beta(x-1),$$

then we will have $p \in \operatorname{span}(w_0, w_1)$, and thus $\operatorname{span}(w_0, w_1) = \mathbb{P}^1$. Then (w_0, w_1) will be a basis for $\mathbb{P}^1$ because it is a spanning list of length 2. So, we want

$$a_0 = \alpha - \beta \quad \text{and} \quad a_1 = \alpha + \beta.$$

This is the linear system

$$\begin{bmatrix}1&-1\\1&1\end{bmatrix}\begin{bmatrix}\alpha\\\beta\end{bmatrix} = \begin{bmatrix}a_0\\a_1\end{bmatrix},$$

and this is invertible. Specifically,

$$\begin{bmatrix}\alpha\\\beta\end{bmatrix} = \begin{bmatrix}1/2&1/2\\-1/2&1/2\end{bmatrix}\begin{bmatrix}a_0\\a_1\end{bmatrix}.$$

This matrix inverse here *is* the change of basis matrix, since

$$v_0(x) = 1 = \frac{x+1}{2} - \frac{x-1}{2} = \frac{1}{2}w_0(x) + \left(-\frac{1}{2}w_1(x)\right)$$

and

$$v_1(x) = x = \frac{x+1}{2} + \frac{x-1}{2} = \frac{1}{2}w_0(x) + \frac{1}{2}w_1(x).$$

(iii) The coordinates of p with respect to the basis (p_0, p_1) are $(1138, 1977)$, so the coordinates of p with respect to the basis (w_0, w_1) are

$$\begin{bmatrix}1/2&1/2\\-1/2&1/2\end{bmatrix}\begin{bmatrix}1138\\1977\end{bmatrix} = \begin{bmatrix}1,557.5\\419.5\end{bmatrix}.$$

38.10. (i) We have

$$\mathcal{S}\mathcal{T}u_j = \mathcal{S}\sum_{k=1}^{j} \langle \mathcal{T}u_j, u_k\rangle u_k = \sum_{k=1}^{j} \langle \mathcal{T}u_j, u_k\rangle \mathcal{S}u_k = \sum_{k=1}^{j} \langle \mathcal{T}u_j, u_k\rangle \sum_{\ell=1}^{k} \langle \mathcal{S}u_k, u_\ell\rangle u_\ell,$$

and so

$$\langle \mathcal{S}\mathcal{T}u_j, u_j\rangle = \left\langle \sum_{k=1}^{j} \langle \mathcal{T}u_j, u_k\rangle \sum_{\ell=1}^{k} \langle \mathcal{S}u_k, u_\ell\rangle u_\ell, u_j \right\rangle = \sum_{k=1}^{j} \langle \mathcal{T}u_j, u_k\rangle \sum_{\ell=1}^{k} \langle \mathcal{S}u_k, u_\ell\rangle \langle u_\ell, u_j\rangle .$$

For $1 \le \ell \le k < j$, we have $\langle u_\ell, u_j\rangle = 0$. So,

$$\langle \mathcal{S}\mathcal{T}u_j, u_j\rangle = \langle \mathcal{T}u_j, u_j\rangle \langle \mathcal{S}u_j, u_j\rangle .$$

39.5. First suppose $1 \le j \le r$. The jth column of $[\mathcal{T}]_{\mathcal{V}\leftarrow\mathcal{V}}$ contains the coordinates of $\mathcal{T}u_j$ with respect to the basis $(u_1, \ldots, u_n)$. Because $\mathcal{T}u_j \in \operatorname{span}(u_1, \ldots, u_j$

40.3. Since $\mathcal{T}$ has an adjoint and is self-adjoint, we know that $\mathcal{T}^* = \mathcal{T}$, and since $\mathcal{T}$ is unitary, we know that $\mathcal{T}^* = \mathcal{T}^{-1}$. Thus $\mathcal{T} = \mathcal{T}^* = \mathcal{T}^{-1}$, so $\mathcal{T}^2 = \mathcal{I}_{\mathcal{V}}$. This does not guarantee that $\mathcal{T} = \mathcal{I}_{\mathcal{V}}$; take, for example, $\mathcal{T} = -\mathcal{I}_{\mathcal{V}}$, or $\mathcal{T} = -i\mathcal{I}_{\mathcal{V}}$ if $\mathcal{V}$ is complex.

40.6. Since $u_1 = v_1/\|v_1\|$, the first vector u_1 will still be an eigenvector. Beyond that, however, $u_j \in \operatorname{span}(v_1, \ldots, v_j)$, and there is no guarantee that a linear combination of eigenvectors possibly corresponding to different eigenvalues is still an eigenvector.

40.15. We compute

$$\|\mathcal{T}v\|^2 = \langle \mathcal{T}v, \mathcal{T}v\rangle = \langle \mathcal{T}^*\mathcal{T}v, v\rangle = \langle \mathcal{T}\mathcal{T}^*v, v\rangle = \langle \mathcal{T}^*v, \mathcal{T}^*v\rangle = \|\mathcal{T}^*v\|^2 .$$

40.20. (i) First suppose that $\mathcal{T}$ is unitary. Then $\mathcal{T}\mathcal{T}^* = \mathcal{I}_{\mathcal{V}}$. We compute

$$\begin{aligned}
\mathcal{M}_{AA^*} &= \mathcal{M}_A\mathcal{M}_{A^*} \\
&= \mathcal{M}_A\mathcal{M}_A^* \\
&= \mathcal{I}_{\mathbb{F}^n}\mathcal{M}_A\mathcal{M}_A^*\mathcal{I}_{\mathbb{F}^n} \\
&= (\mathcal{B}^*\mathcal{B})\mathcal{M}_A(\mathcal{B}^*\mathcal{B})\mathcal{M}_A^*(\mathcal{B}^*\mathcal{B}) \\
&= \mathcal{B}^*(\mathcal{B}\mathcal{M}_A\mathcal{B}^*)(\mathcal{B}\mathcal{M}_A^*\mathcal{B}^*)\mathcal{B} \\
&= \mathcal{B}^*\mathcal{T}\mathcal{T}^*\mathcal{B} \\
&= \mathcal{B}^*\mathcal{I}_{\mathcal{V}}\mathcal{B} \\
&= \mathcal{B}^*\mathcal{B} \\
&= \mathcal{I}_{\mathbb{F}^n}.
\end{aligned}$$

Hence $\mathcal{M}_{AA^*} = \mathcal{I}_{\mathbb{F}^n}$, so $AA^* = I_n$, and therefore A is invertible with $A^{-1} = A^*$.

Conversely, suppose that A is unitary. Then

$$\mathcal{T}\mathcal{T}^* = (\mathcal{B}\mathcal{M}_A\mathcal{B}^*)(\mathcal{B}\mathcal{M}_A^*\mathcal{B}^*)$$

$$\begin{aligned}
&= \mathcal{B}\mathcal{M}_A(\mathcal{B}^*\mathcal{B})\mathcal{M}_{A^*}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_A\mathcal{I}_{\mathbb{F}^n}\mathcal{M}_{A^*}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_A\mathcal{M}_{A^*}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_{AA^*}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_{I_n}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{I}_{\mathbb{F}^n}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{B}^* \\
&= \mathcal{I}_{\mathbb{F}^n}.
\end{aligned}$$

Hence $\mathcal{T}\mathcal{T}^* = \mathcal{I}_{\mathcal{V}}$, so $\mathcal{T}$ is invertible with $\mathcal{T}^{-1} = \mathcal{T}^*$.

By the way, since we are working in a finite-dimensional space here, we only checked that $\mathcal{T}\mathcal{T}^* = \mathcal{I}_{\mathcal{V}}$, and then (by dimension-counting, rank–nullity, etc.), we get $\mathcal{T}^*\mathcal{T} = \mathcal{I}_{\mathcal{V}}$ immediately.

(ii) First suppose that $\mathcal{T} = \mathcal{T}^*$. Then $\mathcal{B}\mathcal{M}_A\mathcal{B}^* = \mathcal{B}\mathcal{M}_A^*\mathcal{B}^*$. Since $\mathcal{B}$ and $\mathcal{B}^*$ are invertible, this gives $\mathcal{M}_A = \mathcal{M}_A^* = \mathcal{M}_{A^*}$, thus $A = A^*$.

Conversely, suppose that $A = A^*$. Then $\mathcal{M}_A = \mathcal{M}_A^*$, so $\mathcal{M}_A = \mathcal{M}_{A^*}$, and so $\mathcal{B}\mathcal{M}_A\mathcal{B}^* = \mathcal{B}\mathcal{M}_{A^*}\mathcal{B}^*$, thus $\mathcal{T} = \mathcal{T}^*$.

(iii) First suppose that $\mathcal{T}\mathcal{T}^* = \mathcal{T}^*\mathcal{T}$. Then

$$\begin{aligned}
\mathcal{M}_{AA^*} &= \mathcal{M}_A\mathcal{M}_{A^*} \\
&= M_A\mathcal{M}_A^* \\
&= \mathcal{I}_{\mathbb{F}^n}\mathcal{M}_A\mathcal{I}_{\mathbb{F}^n}\mathcal{M}_{A^*}\mathcal{I}_{\mathbb{F}^n} \\
&= (\mathcal{B}^*\mathcal{B})\mathcal{M}_A(\mathcal{B}^*\mathcal{B})\mathcal{M}_{A^*}(\mathcal{B}^*\mathcal{B}) \\
&= \mathcal{B}^*(\mathcal{B}\mathcal{M}_A\mathcal{B}^*)(\mathcal{B}\mathcal{M}_{A^*}\mathcal{B}^*)\mathcal{B} \\
&= \mathcal{B}^*\mathcal{T}\mathcal{T}^*\mathcal{B} \\
&= \mathcal{B}^*\mathcal{T}^*\mathcal{T}\mathcal{B} \\
&= \mathcal{B}^*(\mathcal{B}\mathcal{M}_{A^*}\mathcal{B}^*)(\mathcal{B}\mathcal{M}_A\mathcal{B}^*)\mathcal{B} \\
&= (\mathcal{B}^*\mathcal{B})\mathcal{M}_{A^*}(\mathcal{B}^*\mathcal{B})\mathcal{M}_A(\mathcal{B}^*\mathcal{B}) \\
&= \mathcal{M}_{A^*}\mathcal{M}_A \\
&= \mathcal{M}_{A^*A},
\end{aligned}$$

and so $AA^* = A^*A$.

Conversely, suppose that $AA^* = A^*A$. Then

$$\begin{aligned}
\mathcal{T}\mathcal{T}^* &= (\mathcal{B}\mathcal{M}_A\mathcal{B}^*)(\mathcal{B}\mathcal{M}_{A^*}\mathcal{B}^*) \\
&= \mathcal{B}\mathcal{M}_A(\mathcal{B}^*\mathcal{B})\mathcal{M}_{A^*}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_A\mathcal{I}_{\mathbb{F}^n}\mathcal{M}_{A^*}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_A\mathcal{M}_{A^*}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_{AA^*}\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_{A^*A}\mathcal{B}^*
\end{aligned}$$

$$
\begin{aligned}
&= \mathcal{B}\mathcal{M}_{A^*}\mathcal{M}_A\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_A^*\mathcal{I}_{\mathbb{F}^n}\mathcal{M}_A\mathcal{B}^* \\
&= \mathcal{B}\mathcal{M}_A^*(\mathcal{B}^*\mathcal{B})\mathcal{M}_A\mathcal{B}^* \\
&= (\mathcal{B}\mathcal{M}_A^*\mathcal{B}^*)(\mathcal{B}\mathcal{M}_A\mathcal{B}^*) \\
&= \mathcal{T}^*\mathcal{T}.
\end{aligned}
$$

41.8. If A is orthogonally diagonalizable, then $A = U\Lambda U^\top$ for some orthogonal $U \in \mathbb{R}^{n\times n}$. Then

$$A^\top = (U\Lambda U^\top)^\top = U\Lambda^\top U^\top = U\Lambda U^\top = A.$$

Conversely, suppose that $A = A^\top$. Then the eigenvalues of A are real, so the real Schur's theorem for matrices applies to allow us to factor $A = UTU^\top$ for some orthogonal $U \in \mathbb{R}^{n\times n}$ and upper-triangular $T \in \mathbb{R}^{n\times n}$. Since $T = U^\top AU$, we have

$$T^\top = (U^\top AU)^\top = U^\top A^\top U = U^\top AU = T.$$

Hence T is really diagonal.

41.9. (i) The matrix

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

is diagonal, thus unitarily diagonalizable, thus normal. However, it is diagonal and nonreal, so it is not self-adjoint, and while its conjugate tranpose is

$$\begin{bmatrix} -i & 0 \\ 0 & 2 \end{bmatrix},$$

its inverse is

$$\begin{bmatrix} 1/i & 0 \\ 0 & 1/2 \end{bmatrix} = \begin{bmatrix} i/i^2 & 0 \\ 0 & 1/2 \end{bmatrix} = \begin{bmatrix} -i & 0 \\ 0 & 2 \end{bmatrix},$$

so it is not unitary.

(ii) The matrix

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

is real and diagonal, thus Hermitian, but its inverse is

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

and so it (more precisely, its conjugate transpose) is not its own inverse, thus not unitary.

(iii) The matrix

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

is unitary, because its conjugate transpose is

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

and its inverse is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1/i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & i/i^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix},$$

so its inverse and conjugate transpose are equal. However, this matrix is diagonal and not real, so it cannot be self-adjoint.

(iv) The matrix

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

is 2×2 and upper-triangular with diagonal entries equal to 1 and 2, so it has two distinct eigenvalues (1 and 2) and therefore is diagonalizable. However, this matrix is upper-triangular but not diagonal, so it is not normal, thus not unitarily diagonalizable.