MATH 4260: LINEAR ALGEBRA II

Daily Log for Lectures and Readings

Timothy E. Faver

December 9, 2025

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, August 18.

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).

The second problem comes from differential equations, about which we presume no prior knowledge beyond the basics of a standard course in single-variable calculus.

1.4 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?

1.5 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

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

1.6 Problem ($\star$). **(i)** 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.]

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

(iii) 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 the following facts, which you do not need to prove, to show that $\mathcal{S}g$ will be 2π-periodic if $\mathcal{I}(x) := \int_0^{2\pi} \sin(x-y)g(y)\ dy = 0$. (1) The integral property $\int_a^b h(y)\ dy = \int_a^c h(y)\ dy + \int_c^b h(y)\ dy$, valid for any continuous* *function h on $\mathbb{R}$. (2) $\int_x^{x+2\pi} h(y)\ dy = \int_0^{2\pi} h(y)\ dy$ for all $x \in \mathbb{R}$, whenever h is 2π-periodic and continuous on $\mathbb{R}$. Then use a trigonometric addition formula to rewrite $\mathcal{I}(x)$ in such a way that the solvability conditions from* (1.7) *appear. This last step is morally similar to the proof of* (1.9).]

(iv) 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, August 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."

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)

The problems of Examples 1.1 and 1.4 should look cosmetically different, but they are really asking many of the same questions. The first problem is a linear system of the form $A\mathbf{x} = \mathbf{b}$ for $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$ with, ideally, a solution $\mathbf{x} \in \mathbb{R}^4$. The second problem is the ODE $\mathcal{A}f = g$, where $\mathcal{A}f := f'' + f$ with $g\colon \mathbb{R} \to \mathbb{R}$ continuous (and possibly more than that) and $f\colon \mathbb{R} \to \mathbb{R}$ twice-continuously differentiable (and possibly more than that). Both problems are **LINEAR** in the sense that

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2,\ \mathbf{x}_1,\ \mathbf{x}_2 \in \mathbb{R}^n \quad \text{and} \quad A(\alpha\mathbf{x}) = \alpha(A\mathbf{x}),\ \alpha \in \mathbb{R},\ \mathbf{x} \in \mathbb{R}^n \tag{2.1}$$

and

$$\mathcal{A}(f_1 + f_2) = \mathcal{A}f_1 + \mathcal{A}f_2 \quad \text{and} \quad \mathcal{A}(\alpha f) = \alpha(\mathcal{A}f) \tag{2.2}$$

for f_1, f_2, $f\colon \mathbb{R} \to \mathbb{R}$ twice-continuously differentiable. The identities (2.1) are a consequence of the definition of matrix-vector multiplication; we presume familiarity with this from prior exposure to linear algebra.

2.1 Problem (!). Use the linearity of the derivative to prove (2.2).

These kinds of problems are the central object of study in this course. Specifically, we will consider vector spaces $\mathcal{V}$ and $\mathcal{W}$ and a linear operator $\mathcal{T}\colon \mathcal{V} \to \mathcal{W}$. (Many first courses in linear algebra consider this already, and many do not. We will review notions of vector spaces and linear operators from scratch.) Given a vector $w \in \mathcal{W}$, we ask if we can find $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. If we can, we then ask if v is unique; if it is, then we write

$v = \mathcal{T}^{-1}w$. If v is not unique, we ask if there is a natural, meaningful way to choose v to be unique (what is "natural" and "meaningful" will depend on the precise context of what $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{T}$ are). If we cannot solve $\mathcal{T}v = w$, then we ask if there is a natural, meaningful way to "approximate" w by some other $\widehat{w} \in \mathcal{W}$ such that the problem $\mathcal{T}v = \widehat{w}$ *does* have a solution. In short, *how do we characterize and understand the range of a linear operator?*

Addressing these questions will involve three 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. 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?*

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.

2.2 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.

2.3 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.

2.4 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*

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

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:*

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

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

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

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

(i) $B \times A$

(ii) $\varnothing \times A$

2.7 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.*

2.8 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.

2.9 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 $\mathbb{R}$?

2.10 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?

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

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. 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

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

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.

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.

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

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

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

Day 3: Friday, August 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."

List of length n in a set, $\mathbb{R}^{m \times n}$ (as a set of functions), vector space, additive identity for a vector space, additive inverse for a vector space, zero vector for a vector space

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

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

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

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

3.1 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).*

Here is a slight generalization of this notion of $\mathbb{R}^n$ that will be useful when we speak precisely about bases.

3.2 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,*

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

So, vectors in $\mathbb{R}^n$ are lists of length n in $\mathbb{R}$. But we will also think about lists of length n in $\mathbb{R}^m$ from time to time. This arises naturally when we talk about matrices.

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.

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

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

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

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

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

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.

3.5 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**.

3.6 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.

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 $f + (-f) = 0$, where $(-f)(x) = -f(x)$.

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.

3.7 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.

3.8 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* α, $\beta \in \mathbb{F}$ *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* α, $\beta \in \mathbb{F}$ *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.

3.9 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* $v+(-v) = 0_{\mathcal{V}}$. *It is also possible to show that* $-v = (-1) \cdot v$; *that is, there is an intimate, and expected, connection between the additive inverse in* $\mathcal{V}$ *and scalar multiplication by the additive inverse of the multiplicative identity in* $\mathbb{F}$.

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

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

(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. We may or may not write 0 in lieu of $0_{\mathcal{V}}$. (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.)*

3.10 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 $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$ are vector spaces over $\mathbb{R}$, and more generally $\mathbb{F}^n$ 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.

3.11 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.

Day 4: Monday, August 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."

Subspace of a vector space (N)

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.

4.1 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: that is,

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

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:

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

4.2 Problem (!). Complete the proof of the uniqueness of the additive inverse begun in part (v) of Example 4.1. [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.

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.

4.3 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.

4.4 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.

4.5 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 $\alpha v \in \mathcal{U}$.*

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 $\mathbb{F}^X$.

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

4.7 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 $\mathbf{v} \in \mathcal{W}$.

4.8 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.

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 $M > 0$ such that $|\alpha a_k| \le M$ for all k. Since $(a_k) \in \ell^\infty$, we know there is $N > 0$ such that $|a_k| \le N$ for all k. Since

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

we have $|\alpha a_k| = |\alpha||a_k| \le |\alpha|N$. Taking $M = |\alpha|N$ 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 $\mathcal{C}^k([a,b])$.

Day 5: Wednesday, August 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."

Polynomial, linear combination, span, linear operator (N), identity operator

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

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

$$p(x) = \sum_{k=0}^{n} a_k x^k$$

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

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

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

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

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

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

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

5.3 Problem (!). Why is $\mathcal{U} := \left\{p \in \mathbb{P}^2 \;\middle|\; \deg(p) = 2\right\}$ 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.

5.4 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*

5.5 Problem (+). Let X be a nonempty set and let $\mathcal{V}$ be a vector space over $\mathbb{F}$. What is a "natural" way of defining vector addition and scalar multiplication in $\mathcal{V}^X$ so that $\mathcal{V}^X$ is a vector space over $\mathbb{F}$?

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

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

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

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

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

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

5.6 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 $v = \sum_{j=1}^{n} v_j$. 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 $\mathcal{B} \subseteq \mathcal{V}$ is nonempty, the* **SPAN** *of $\mathcal{B}$ is the set of all (necessarily finite) linear combinations of vectors in $\mathcal{B}$:*

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

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.

5.7 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."

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

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

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

5.9 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 $\mathbb{R}$.*

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

(i) Define

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

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

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

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

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

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

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

5.11 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and let $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ 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 $v \in \mathcal{V}$?*]

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.

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

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

That is, $\mathcal{T}f = mf$. 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

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

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

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

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

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

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?

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

5.14 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

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

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

Day 6: Friday, August 29.

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, matrix-vector product, eigenvalue of a linear operator (N), eigenvector of a linear operator (N), eigenvalue of a matrix (N), eigenvector of a matrix (N)

6.1 Example. Let $\mathcal{V} = \mathcal{C}([0,1])$ and $\mathcal{W} = \mathcal{C}^1([0,1])$. Calculus teaches us the following about integrals: for each $f \in \mathcal{V}$ and each a, $b \in [0,1]$, 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 $c \in [0,1]$.

($\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 f, $g \in \mathcal{V}$ 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

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

is differentiable with $F' = f$. Also, if $f \in \mathcal{W}$, then

$$\int_a^b f'(s) \ ds = f(b) - f(a).$$

This is effectively all of the information about integrals that we will need. For $f \in \mathcal{V}$, define

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

Then $\mathcal{T}f \in \mathcal{W}$, since $\mathcal{T}f$ is differentiable with $(\mathcal{T}f)' = f \in \mathcal{V}$. Linearity of $\mathcal{T}$ follows immediately from the linearity of the integral. (This is one of those times where showing the linearity of the operator is not so hard, but showing that the operator maps to the claimed codomain requires more technology.)

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 $\mathbb{F}$.

6.2 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.

6.3 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

$$\begin{aligned}
\left\{\begin{array}{rcrcr} x_1 & - & 2x_2 & = & 1 \\ 3x_1 & + & 2x_2 & = & 11 \end{array}\right.
&\overset{(1)}{\iff} \begin{bmatrix} x_1 - 2x_2 \\ 3x_1 + 2x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix} \\
&\overset{(2)}{\iff} \begin{bmatrix} x_1 \\ 3x_1 \end{bmatrix} + \begin{bmatrix} -2x_2 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix} \\
&\overset{(3)}{\iff} x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix} \\
&\overset{(4)}{\iff} \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}
\end{aligned}$$

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.

6.4 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

$$A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w} \quad \text{and} \quad A(\alpha\mathbf{v}) = \alpha A\mathbf{v} \tag{6.1}$$

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

6.5 Problem ($\star$). Show that (6.1) is true.

6.6 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.*

6.7 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.

6.8 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 6.6 does not indicate any of this; if it matters, context will make it clear.*

6.9 Problem ($\star$). The definition of matrix-vector multiplication from Definition 6.4 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 6.4 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.

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.

Perhaps the simplest operator between spaces $\mathcal{V}$ and $\mathcal{W}$ is the one that maps all vectors to $0_{\mathcal{W}}$:

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

6.10 Problem (!). Check that this is a linear operator. Then fix $w_0 \in \mathcal{W}$. Is the map

$$\mathcal{T}\colon \mathcal{V} \to \mathcal{W}\colon v \mapsto w_0$$

ever linear?

We saw some very simple operators in Example 5.10. 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?

6.11 Definition. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{T}\colon \mathcal{V} \to \mathcal{V}$ be linear. A vector $v \in \mathcal{V} \setminus \{0\}$ 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 $0 \in \mathcal{V}$ as an eigenvector because then $\mathcal{T}0 = \lambda 0$ 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 $w = 0$.

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

6.12 Example. Let

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

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.

6.13 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 $\mathbf{v} \in \mathbb{R}^2 \setminus \{\mathbf{0}_2\}$.*

6.14 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) $A \in \mathbb{F}^{m\times n}$.

Eigenvalues and eigenvectors of matrices are basically the same as those of the multiplication operator $\mathcal{M}_A$. We make this (im)precise.

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

6.16 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.

6.17 Remark. *We will only talk about eigenvalues of matrices with complex, possibly nonreal entries, and so we will always allow matrix eigenvalues to be complex. However, we will point out some situations in which a real matrix has only nonreal eigenvalues. The point is that a linear operator on a vector space over $\mathbb{R}$ may have no real eigenvalues. We* *will see how taking the underlying field to be $\mathbb{C}$ guarantees, in broad circumstances, the existence of eigenvalues. 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.*

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

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

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 $\mathcal{T}$.

6.19 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?

6.20 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 $\lambda_1 = \lambda_2$.

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."

Day 7: Wednesday, September 3.

Here is an operator that has no eigenvalues.

7.1 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

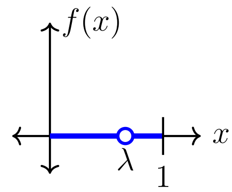
$$x - \lambda = 0 \quad \text{or} \quad f(x) = 0, \tag{7.1}$$

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 (7.1), 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 (7.1) 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.

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

7.2 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

$$a_{k+1} = \lambda a_k \tag{7.2}$$

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,

$$a_k = \lambda^{k-1}a_1 \tag{7.3}$$

for all k. We could prove this by induction on k from the relation (7.2).

This is the classical mathematical technique of working backwards: if $\mathcal{T}(a_k) = \lambda(a_k)$, then λ and (a_k) must satisfy (7.3). Does the logic go the other way? If λ and (a_k) satisfy (7.3), 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 (7.3), 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 (7.3), 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.

7.3 Problem ($\star$). We saw in Example 5.14 that the map

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

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

(i) Show that

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

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

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

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

Here is an eigenvalue example in which the choice of field—context again!—matters.

7.4 Example. Let

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

(i) We consider $\mathcal{M}_A\mathbf{v} = A\mathbf{v}$ 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 6.11 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.

(iii) Because we choose to consider a scalar as an eigenvalue of a matrix whether or not that scalar is real, the matrix A always has eigenvalues. But A has only real entries and these eigenvalues are purely imaginary. It can happen.

7.5 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.

Day 8: Friday, September 5.

Vocabulary from today

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

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 5.5 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.

8.1 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:

$$\begin{aligned}(\mathcal{T} + \mathcal{S})(v_1 + v_2) &= \mathcal{T}(v_1 + v_2) + \mathcal{S}(v_1 + v_2) \text{ by definition of } \mathcal{T} + \mathcal{S} \\ &= \mathcal{T}v_1 + \mathcal{T}v_2 + \mathcal{S}v_1 + \mathcal{S}v_2 \text{ by the linearity of } \mathcal{T} \text{ and } \mathcal{S} \\ &= (\mathcal{T}v_1 + \mathcal{S}v_1) + (\mathcal{T}v_2 + \mathcal{S}v_2) \text{ by commutativity of addition in } \mathcal{W} \\ &= (\mathcal{T} + \mathcal{S})v_1 + (\mathcal{T} + \mathcal{S})v_2 \text{ by definition, again, of } \mathcal{T} + \mathcal{S}.\end{aligned} \tag{8.1}$$

The rest of the proof follows from similar manipulations. ∎

8.2 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.

8.3 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 8.1 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 $0_{\mathcal{V}\to\mathcal{W}} \colon \mathcal{V} \to \mathcal{W} \colon v \mapsto 0_{\mathcal{W}}$, and it is straightforward to check that $0_{\mathcal{V}\to\mathcal{W}}$ is linear, thus $0_{\mathcal{V}\to\mathcal{W}} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. ■

8.4 Problem (!). **(i)** Prove that the zero map $0_{\mathcal{V}\to\mathcal{W}}$ from $\mathcal{V}$ to $\mathcal{W}$ is linear.

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

8.5 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 $\mathbf{0}_n \in \mathbb{F}^n$.*

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

8.6 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 $\mathbf{L}(\mathcal{V}, \mathbb{F})$.

8.7 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.

8.8 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*

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

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

$$\begin{tikzcd} \mathcal{U} \arrow[r, "\mathcal{T}"] \arrow[dr, dashed, "\mathcal{S}\mathcal{T}"'] & \mathcal{V} \arrow[d, "\mathcal{S}"] \\ & \mathcal{W} \end{tikzcd}$$

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

$$\begin{aligned} (\mathcal{S}\mathcal{T})(u_1 + u_2) &= \mathcal{S}(\mathcal{T}(u_1 + u_2)) \\ &= \mathcal{S}(\mathcal{T}u_1 + \mathcal{T}u_2) \\ &= \mathcal{S}(\mathcal{T}u_1) + \mathcal{S}(\mathcal{T}u_2) \\ &= (\mathcal{S}\mathcal{T})u_1 + (\mathcal{S}\mathcal{T})u_2. \end{aligned} \tag{8.2}$$

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

8.9 Problem (!). **(i)** Justify each equality in (8.2). Compare your justifications to the ones given for (8.1) in the proof of Theorem 8.1.

(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.

8.10 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 $v_1 = 0$.

8.11 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: $(\mathcal{S}f)' \neq \mathcal{S}f'$.

8.12 Problem (!). Here we use the notation for the zero operator from the proof of Corollary 8.3. 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 $\mathcal{S}0_{\mathcal{U}\to\mathcal{V}}$?

8.13 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.

8.14 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*

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

$$\begin{tikzcd} \mathcal{U} \arrow[r, "\mathcal{T}_j"] \arrow[dr, dashed, "\mathcal{S}\mathcal{T}_j"'] & \mathcal{V} \arrow[d, "\mathcal{S}"] \\ & \mathcal{W} \end{tikzcd}$$

(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*

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

$$\begin{tikzcd} \mathcal{U} \arrow[r, "\mathcal{T}"] \arrow[dr, dashed, "\mathcal{S}_j\mathcal{T}"'] & \mathcal{V} \arrow[d, "\mathcal{S}_j"] \\ & \mathcal{W} \end{tikzcd}$$

(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*

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

$$\begin{tikzcd} \mathcal{U} \arrow[r, "\mathcal{T}_1"] \arrow[rrr, dashed, bend left, "\mathcal{T}_3\mathcal{T}_2\mathcal{T}_1"] & \mathcal{V} \arrow[r, "\mathcal{T}_2"] & \mathcal{W} \arrow[r, "\mathcal{T}_3"] & \mathcal{X} \end{tikzcd}$$

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

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

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

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

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

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

8.15 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 3.8.

8.16 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.

8.17 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.

8.18 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 9: Monday, September 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."

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

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.

9.1 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.

9.2 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*

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

9.3 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*

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

9.4 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.

9.5 Problem ($\star$). Give examples of matrices A and B for which AB is the "zero matrix," whose entries are all 0, yet at least one entry in each of A and B is nonzero (so neither A nor B is the zero matrix). [Hint: *work at the 2×2 level with upper-triangular matrices. These are matrices of the form*

$$\begin{bmatrix} a & c \\ 0 & d \end{bmatrix}.$$

Along the way, convince yourself that diagonal matrices, which have the form

$$\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix},$$

intrinsically cannot work here.]

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

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

9.6 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*

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

9.7 Problem (!). Let $\mathcal{T}$ be the operator from Problem 6.14. How did that problem show how to find $[\mathcal{T}]$?

9.8 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

$$\mathcal{T}\mathbf{v} = \mathcal{T}\left(\sum_{j=1}^n v_j\mathbf{e}_j\right) = \sum_{j=1}^n v_j\mathcal{T}\mathbf{e}_j = \begin{bmatrix} \mathcal{T}\mathbf{e}_1 & \cdots & \mathcal{T}\mathbf{e}_n \end{bmatrix}\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = [\mathcal{T}]\mathbf{v} = \mathcal{M}_{[\mathcal{T}]}\mathbf{v}. \tag{9.1}$$

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

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

9.9 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

$$\mathcal{S}_2\mathcal{S}_1\mathcal{T} = \mathcal{S}_2[\mathcal{T}] = \mathcal{M}_{[\mathcal{T}]} = \mathcal{T}. \tag{9.2}$$

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

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

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

Thus

$$\mathcal{S}_1\mathcal{S}_2 A = \mathcal{S}_1\mathcal{M}_A = [\mathcal{M}_A] = A. \tag{9.3}$$

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 8.11, except the "undoing" did not quite work out because of the constant of integration.

9.11 Example. Let

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

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 8.11, 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

$$\mathcal{S}\mathcal{T}v = v \text{ for all } v \in \mathcal{V} \quad \text{and} \quad \mathcal{T}\mathcal{S}w = w \text{ for all } w \in \mathcal{W}. \tag{9.4}$$

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 (9.4)? Suppose that $\mathcal{S}_1$, $\mathcal{S}_2 \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ both meet (9.4). Then

$$\begin{aligned} \mathcal{S}_1 w &= \mathcal{S}_1(\mathcal{T}\mathcal{S}_2 w) \text{ because } \mathcal{T}\mathcal{S}_2 w = w \ \forall w \in \mathcal{W} \\ &= \mathcal{S}_1(\mathcal{T}(\mathcal{S}_2 w)) \text{ by definition of } \mathcal{T}\mathcal{S}_2 \\ &= (\mathcal{S}_1\mathcal{T})(\mathcal{S}_2 w) \text{ by associativity of operator composition} \\ &= \mathcal{S}_2 w \text{ because } \mathcal{S}\mathcal{T}_1 v = v \ \forall v \in \mathcal{V}. \end{aligned} \tag{9.5}$$

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

9.12 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 $\mathcal{T}^{-1} := \mathcal{S}$.*

9.13 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.*]

9.14 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

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

9.15 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.

9.16 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 10: Wednesday, September 10.

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)

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 $\mathcal{T} v = w$.

10.1 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, $\mathcal{S}(\mathcal{T} v) = v$.

10.2 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.

10.3 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 $(w_1 + w_2, v_1 + v_2) \in \mathcal{S}$.

10.4 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.

10.5 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.

10.6 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.

10.7 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 $B = A^{-1}$.*

10.8 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.

10.9 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 10.5. 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.

10.10 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 $\mathcal{T}(1, 0, 0, \ldots) = \mathcal{T}(0, 0, 0, \ldots) = (0)$.

10.11 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.*]

10.12 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* (9.5).]

(iv) In proving Theorem 10.5, 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 11: Friday, September 12.

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."

Isomorphism (N), isomorphic vector spaces (N)

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 2.7:

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

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

11.1 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 4.5). 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).

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 $v = 0$.

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

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

11.3 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 4.5). 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 11.2. ■

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

If $\ker(\mathcal{T}) \neq \{0\}$, 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 $z \neq 0$. Then $\mathcal{T}(v + \alpha z) = w$

for all $\alpha \in \mathbb{F}$, and since $z \neq 0$, when $\alpha_1 \neq \alpha_2$, we have $v + \alpha_1 z \neq v + \alpha_2 z$, thus every α gives a different solution.

11.4 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.

11.5 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.

11.6 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.

11.7 Problem ($\star$). **(i)** Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ be invertible. Prove that $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathcal{T}$ if and only if λ^{-1} is an eigenvalue of $\mathcal{T}^{-1}$. (By Problem 11.6, 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 6.18 showed that every $\lambda \in \mathbb{R}$ is an eigenvalue of the differentiation operator on $\mathcal{V}$. Problem 6.19 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 9.11 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.*]

11.8 Problem ($\star$). Recall from Example 7.1 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 11.6. Show that if $0 \le \lambda \le 1$, 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.

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

11.9 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."

11.10 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.

11.11 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 (9.2) and (9.3), 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 (9.1) shows $\mathcal{T} = \mathcal{M}_{[\mathcal{T}]} = \mathcal{S}[\mathcal{T}]$.

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

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

Day 12: Monday, September 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."

Finite-dimensional vector space, infinite-dimensional vector space, basis for a vector space

Here is an exercise in diagram-chasing.

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

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

$$\begin{tikzcd} \mathcal{V}_1 \arrow[r, "\mathcal{A}"] \arrow[d, "\mathcal{T}"'] & \mathcal{W}_1 \arrow[d, "\mathcal{S}"] \\ \mathcal{V}_2 \arrow[r, dashed, "?"'] & \mathcal{W}_2 \end{tikzcd}$$

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

$$\begin{tikzcd} \mathcal{V}_1 \arrow[r, "\mathcal{A}"] & \mathcal{W}_1 \arrow[d, "\mathcal{S}"] \\ \mathcal{V}_2 \arrow[u, "\mathcal{T}^{-1}"] \arrow[r, dashed, "\mathcal{S}\mathcal{A}\mathcal{T}^{-1}"'] & \mathcal{W}_2 \end{tikzcd}$$

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

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

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

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

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

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

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

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

$$\begin{tikzcd} \mathcal{V}_1 \arrow[r, dashed, "\mathcal{A}?"] \arrow[d, "\mathcal{T}"'] & \mathcal{W}_1 \arrow[d, "\mathcal{S}"] \\ \mathcal{V}_2 \arrow[r, "\widetilde{\mathcal{A}}"'] & \mathcal{W}_2 \end{tikzcd}$$

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

$$\begin{tikzcd} \mathcal{V}_1 \arrow[r, dashed, "\mathcal{S}^{-1}\widetilde{\mathcal{A}}\mathcal{T}"] \arrow[d, "\mathcal{T}"'] & \mathcal{W}_1 \\ \mathcal{V}_2 \arrow[r, "\widetilde{\mathcal{A}}"'] & \mathcal{W}_2 \arrow[u, "\mathcal{S}^{-1}"'] \end{tikzcd}$$

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

12.2 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 9.13, 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.

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

We summarize this formally.

12.4 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*

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

12.5 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:

$$\mathbb{F}^n = \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_n) \quad \text{and} \quad \mathbb{P}^n = \operatorname{span}(p_0, p_1, \ldots, p_n), \tag{12.1}$$

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.

12.6 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.

12.7 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.

12.8 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 $j = 1\ldots, m-1$.)*

12.9 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.

12.10 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.

12.11 Example. The spaces $\mathbb{F}^n$ and $\mathbb{P}^{n+1}$ are finite-dimensional by (12.1). We will prove, with some nontrivial technology, that $\mathbb{F}^\infty$ and $\mathcal{C}^r([0,1])$ are infinite-dimensional.

Definition 12.10 allows some unfortunate inefficiency in writing a vector space as a span.

12.12 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.

12.13 Problem (!). Prove all of the claims in Example 12.12.

12.14 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 $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n, w_1, \ldots, w_m)$.

We should try to avoid the "linear redundancy" that Definition 12.10 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.

12.15 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 $v = \sum_{j=1}^n \alpha_j v_j$.*

12.16 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.

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

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

is an isomorphism. (You do not have to prove that $\mathcal{T}$ is linear.)

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 12.16.

12.18 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 12.15 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.

Day 13: Wednesday, September 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."

Linearly independent list (N), linearly dependent list (N)

Condition (ii) in the above characterization of basis (possibly the surprising one—spans should not be surprising) is hugely important by itself.

13.1 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,*

$$\exists(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n : \sum_{j=1}^n \alpha_j v_j = 0_{\mathcal{V}} \text{ and } \exists j : \alpha_j \neq 0.$$

13.2 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

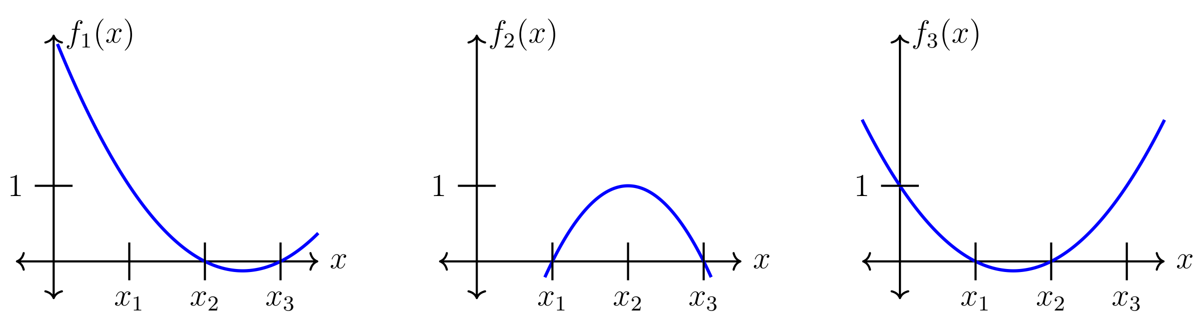
$$f_j(x_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

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.

13.3 Problem (!). Show that the list $(1, \sin(\cdot), \cos(\cdot))$ is independent in $\mathcal{C}([0,1])$.

Here is an easier situation.

13.4 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. ■

13.5 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?)

13.6 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 $\mathbb{F} = \mathbb{C}$.

13.7 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 $\sum_{j=1}^n \alpha_j v_j \neq 0_{\mathcal{V}}$.

Day 14: Friday, September 19.

Here is an important characterization of a dependent list.

14.1 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*

$$v_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} \alpha_j v_j \tag{14.1}$$

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 $\sum_{j=1}^{k} \alpha_j v_j = 0$; 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, (14.1) 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.

14.2 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.

14.3 Lemma. *If a list of length $n \ge 2$ 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:

$$v_k = \sum_{\substack{j=1 \\ j \ne k}} \alpha_j v_j.$$

Then any linear combination of the original list $(v_1, \ldots, v_n)$ has the form

$$\sum_{j=1}^n \beta_j v_j = \sum_{\substack{j=1 \\ j \ne k}} \beta_j v_j + \beta_k \sum_{\substack{j=1 \\ j \ne k}} \alpha_j v_j = \sum_{\substack{j=1 \\ j \ne k}} (\beta_j + \beta_k \alpha_j) v_j.$$

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:

$$\sum_{\substack{j=1 \\ j \ne k}}^n \gamma_j v_j = \sum_{j=1}^n \mu_j v_j, \qquad \mu_j = \begin{cases} \gamma_j, & j \ne k \\ 0, & j = k. \end{cases}$$

■

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.

14.4 Lemma (Linear (in)dependence lemma). *Let $(v_1, \ldots, v_n)$ be a list of length $n \ge 2$ with $v_1 \ne 0_{\mathcal{V}}$. Then $(v_1, \ldots, v_n)$ is dependent if and only if there exists $k \ge 2$ such that $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$. Equivalently, when $n \ge 2$ and $v_1 \ne 0_{\mathcal{V}}$, 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 $j \ge 2$.*

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. ■

14.5 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 $(0, v_2, \ldots, v_n)$ 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.

Here is a classical deployment of the linear dependence lemma, which in its fullest form uses induction.

14.6 Example. A list of eigenvectors corresponding to distinct eigenvalues is independent. That is, assume $n \geq 2$ and let $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ have the distinct eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ with corresponding eigenvectors $v_1, \ldots, v_n \in \mathcal{V}$. So, $\lambda_j \neq \lambda_k$ for $j \neq k$, and $\mathcal{T} v_j = \lambda_j v_j$. Then the list $(v_1, \ldots, v_n)$ is independent.

1. We first show this for the simple case of $n = 3$. What if the list is dependent? Since the entries of the list are eigenvectors, none is $0_{\mathcal{V}}$, so in particular $v_1 \neq 0_{\mathcal{V}}$. So, it must be the case that either $v_2 \in \operatorname{span}(v_1)$ or $v_3 \in \operatorname{span}(v_1, v_2)$.

(i) In the former, we have $v_2 = \alpha_1 v_1$. In particular, $\alpha_1 \neq 0$, as otherwise $v_2 = 0_{\mathcal{V}}$. The only other thing that we know about v_1 and v_2 is how they talk with $\mathcal{T}$, so we apply $\mathcal{T}$ to both sides to get $\mathcal{T} v_2 = \alpha_1 \mathcal{T} v_1$, and thus $\lambda_2 v_2 = \alpha_1 \lambda_1 v_1$. Substitute $v_2 = \alpha_1 v_1$ on the left to find $\lambda_2 \alpha_1 v_1 = \alpha_1 \lambda_1 v_1$. Since $\alpha_1 \neq 0$, we may divide to find $\lambda_2 v_1 = \lambda_1 v_1$, thus $(\lambda_1 - \lambda_2) v_1 = 0_{\mathcal{V}}$. Since $\lambda_1 \neq \lambda_2$, we have $v_1 = 0_{\mathcal{V}}$, a contradiction. So, $v_2 \notin \operatorname{span}(v_1)$.

(ii) What if $v_3 \in \operatorname{span}(v_1, v_2)$? Then $v_3 = \alpha_1 v_1 + \alpha_2 v_2$. Apply $\mathcal{T}$ to both sides to obtain

$$\lambda_3 v_3 = \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2.$$

Substitute $v_3 = \alpha_1 v_1 + \alpha_2 v_2$ to obtain

$$\lambda_3\alpha_1 v_1 + \lambda_3\alpha_2 v_2 = \alpha_1\lambda_1 v_1 + \alpha_2\lambda_2 v_2.$$

Rearrange to get

$$\alpha_1(\lambda_1 - \lambda_3)v_1 + \alpha_2(\lambda_2 - \lambda_3)v_2 = 0_{\mathcal{V}}.$$

We know $v_2 \notin \operatorname{span}(v_1)$ and $v_1 \neq 0_{\mathcal{V}}$, so the list (v_1, v_2) is independent. Thus

$$\alpha_1(\lambda_1 - \lambda_3) = \alpha_2(\lambda_2 - \lambda_3) = 0,$$

and since $\lambda_1 \neq \lambda_3$ and $\lambda_2 \neq \lambda_3$, we must have $\alpha_1 = \alpha_2 = 0$. Thus $v_3 = 0_{\mathcal{V}}$, another contradiction. So, $v_3 \notin \operatorname{span}(v_1, v_2)$.

2. Here is how the argument works in general. We induct on n.

(i) *The base case.* If $n = 1$, then the list (v_1) has only one entry, which is nonzero, since v_1 is an eigenvector, and so this list is independent.

(ii) *The induction hypothesis and step.* 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 $v_1 \neq 0$ (again, because v_1 is an eigenvector), the linear (in)dependent lemma says that $v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$ for some $j \geq 2$. But $(v_1, \ldots, v_n)$ is a list of n eigenvectors of $\mathcal{T}$ corresponding to distinct eigenvalues, so by the induction hypothesis it is independent, and therefore $v_j \notin \operatorname{span}(v_1, \ldots, v_{j-1})$ for $2 \leq j \leq n$. The only possibility is that $v_{n+1} \in \operatorname{span}(v_1, \ldots, v_n)$.

Write $v_{n+1} = \sum_{j=1}^{n} \alpha_j v_j$, thus

$$\lambda_{n+1} v_{n+1} = \mathcal{T} v_{n+1} = \mathcal{T} \sum_{j=1}^{n} \alpha_j v_j = \sum_{j=1}^{n} \alpha_j \mathcal{T} v_j = \sum_{j=1}^{n} \alpha_j \lambda_j v_j.$$

Substitute $v_{n+1} = \sum_{j=1}^{n} \alpha_j v_j$ to find

$$\lambda_{n+1} \sum_{j=1}^{n} \alpha_j v_j = \sum_{j=1}^{n} \alpha_j \lambda_j v_j.$$

Rearrange to find

$$\sum_{j=1}^{n} \alpha_j(\lambda_j - \lambda_{n+1})v_j = 0.$$

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.

Day 15: Monday, September 22.

15.1 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.

Finally, while we worked in $\mathcal{C}^\infty([0,1])$ to prove independence of $(p_0, \ldots, p_n)$, this list is also independent in $\mathcal{C}^r([0,1])$ for any $r \geq 0$. After all, independence in any of those spaces just means that if $\sum_{j=0}^n \alpha_j p_j = 0$, then $\alpha_j = 0$ for all j. The underlying space really does not matter!

Content from *Linear Algebra* by Meckes & Meckes. For another perspective on this example, see the example on p. 146.

15.2 Problem ($\star$). Prove (and also state precisely) the following.

(i) Any sublist of an independent list is independent.

(ii) A list with a dependent sublist is dependent.

(iii) A dependent list may have an independent sublist.

Now we prove that every finite-dimensional vector space has a basis.

15.3 Lemma. *Any spanning list can be reduced to a basis. That is, if $\mathcal{V}$ 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.

1. *The case* $n = 1$. For $n = 1$, $\mathcal{V} = \text{span}(v_1)$ and since $\mathcal{V}$ is nonzero, $v_1 \neq 0$, so the list (v_1) is independent. Since it spans $\mathcal{V}$ already, it is a basis.

2. *The induction hypothesis and step.* Assume that the result is true for some $n \geq 1$ and suppose now that $\mathcal{V} = \text{span}(v_1, \ldots, v_n, v_{n+1})$. Let $\mathcal{V}_n = \text{span}(v_1, \ldots, v_n)$. By the induction hypothesis, $\mathcal{V}_n = \text{span}(v_{j_1}, \ldots, v_{j_r})$ for an independent sublist of $(v_{j_1}, \ldots, v_{j_r})$, and so

$$\text{span}(v_{j_1}, \ldots, v_{j_r}, v_{n+1}) = \text{span}(v_1, \ldots, v_n, v_{n+1}) = \mathcal{V}.$$

If $v_{n+1} \in \mathcal{V}_n$, then $v_{n+1} \in \text{span}(v_{j_1}, \ldots, v_{j_r})$. The "removal" process from Lemma 14.3 then says that

$$\mathcal{V} = \text{span}(v_{j_1}, \ldots, v_{j_r}, v_{n+1}) = \text{span}(v_{j_1}, \ldots, v_{j_r}),$$

and so we are done.

If $v_{n+1} \notin \text{span}(v_{j_1}, \ldots, v_{j_r})$, then the list $(v_{j_1}, \ldots, v_{j_r}, v_{n+1})$ is independent. This is because $v_{j_1} \neq 0_{\mathcal{V}}$, since the list $(v_{j_1}, \ldots, v_{j_r})$ is independent, and also $v_{j_k} \notin \text{span}(v_{j_1}, \ldots, v_{j_{k-1}})$ for $k \geq 2$, again by independence of the list $(v_{j_1}, \ldots, v_{j_r})$. 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 14.3 and 14.4).

15.4 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 Lemma 15.3 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.

Day 16: Wednesday, September 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."

Dimension of a vector space

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. There are multiple ways to prove that all bases have the same length; here are some motivating considerations and then the full argument.

16.1 Example. **(i)** *A contradiction with systems.* Suppose that the list $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is independent in $\mathbb{F}^m$ but the list $(\mathbf{w}_1, \mathbf{w}_2)$ spans $\mathbb{F}^m$. Write

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

We can recognize this as saying

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

Since the matrix

$$A := \begin{bmatrix}x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2\end{bmatrix}$$

has more columns than rows, a first course in linear algebra leads us to expect that it has a nontrivial null space. That is, there is $\mathbf{c} \in \mathbb{F}^3 \setminus \{\mathbf{0}_3\}$ such that $A\mathbf{c} = \mathbf{0}_2$. But then

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2\end{bmatrix} A\mathbf{c} = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2\end{bmatrix}\mathbf{0}_2 = \mathbf{0}_m,$$

and so $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is dependent.

(ii) *A contradiction with "swapping."* Say that (v_1, v_2, v_3) is independent in $\mathcal{V}$ but (w_1, w_2) spans $\mathcal{V}$. The list (v_1, w_1, w_2) still spans $\mathcal{V}$ and now is dependent since v_1 is in the span of (w_1, w_2). Since (v_1, v_2, v_3) is independent, $v_1 \neq 0_{\mathcal{V}}$, so the linear dependence lemma means that either

$$w_1 \in \operatorname{span}(v_1) \qquad \text{or} \qquad w_2 \in \operatorname{span}(w_1). \tag{16.1}$$

Suppose the latter. Then

$$\operatorname{span}(v_1, w_1) = \operatorname{span}(v_1, w_1, w_2) = \mathcal{V}.$$

Next, the list (v_1, v_2, w_1) also spans $\mathcal{V}$ and i dependent, since $v_2 \in \mathcal{V} = \operatorname{span}(v_1, w_1)$. Since $v_1 \neq 0_{\mathcal{V}}$, the linear dependence lemma implies that either $v_2 \in \operatorname{span}(v_1)$ or $w_1 \in \operatorname{span}(v_1, v_2)$. The former cannot happen, since (v_1, v_2) is a sublist of the independent list (v_1, v_2) and therefore independent itself. So, $w_1 \in \operatorname{span}(v_1, v_2)$, and then

$$\operatorname{span}(v_1, v_2) = \operatorname{span}(v_1, v_2, w_1) = \mathcal{V}.$$

But then $v_2 \in \mathcal{V} = \operatorname{span}(v_1, v_2)$, so (v_1, v_2, v_3) is dependent, after all.

16.2 Problem (!). Redo the work of the previous problem starting now with $w_2 \in \operatorname{span}(w_1)$, as in (16.1).

Content from *Linear Algebra* by Meckes & Meckes. Lemma 3.17 on p. 163 generalizes the "contradiction with systems" approach.

Here is the general result.

16.3 Theorem (Steinitz exchange). *Any independent list is no longer than any spanning list. That is, if $\mathcal{V} = \text{span}(w_1, \ldots, w_m)$ and $(v_1, \ldots, v_n)$ is independent in $\mathcal{V}$, then $n \le m$. Additionally, the following hold.*

(i) *If $n = m$, then*
$$\mathcal{V} = \text{span}(v_1, \ldots, v_n).$$

(ii) *If $n < m$, then there is a sublist $(w_{j_1}, \ldots, w_{j_{m-n}})$ of $(w_1, \ldots, w_m)$ such that*
$$\mathcal{V} = \text{span}(v_1, \ldots, v_n, w_{j_1}, \ldots, w_{j_{m-n}}).$$

Proof. We induct on n.

1. *The base case.* If $n = 1$, then necessarily $1 \le m$.

(i) *The case $1 = m$.* If $m = 1$, then $v_1 = \alpha_1 w_1$ for some $\alpha_1 \in \mathbb{F}$. Since (v_1) is independent, $v_1 \ne 0$, so $\alpha_1 \ne 0$, and therefore $w_1 = \alpha_1^{-1} v_1$. Then $\mathcal{V} = \text{span}(w_1) = \text{span}(\alpha_1^{-1} v_1) = \text{span}(v_1)$.

(ii) *The case $1 < m$.* If $m \ge 2$, consider the list $(v_1, w_1, \ldots, w_m)$, which also spans $\mathcal{V}$. (This is the sort of "helpful redundancy" foretold by Problem 12.14, and we will exploit this several more times in the proof.) Since $v_1 \in \text{span}(w_1, \ldots, w_m)$, the list $(v_1, w_1, \ldots, w_m)$ is dependent. Since $v_1 \ne 0$, by the linear (in)dependence lemma, either
$$w_1 \in \text{span}(v_1) \quad \text{or} \quad w_j \in \text{span}(v_1, w_1, \ldots, w_{j-1}) \text{ for some } j \ge 2.$$

In the first case,
$$\text{span}(v_1, w_1, \ldots, w_m) = \text{span}(v_1, w_2, \ldots, w_m).$$

In the second case, let $(w_{j_1}, \ldots, w_{j_{m-1}})$ be the list formed by removing w_j from the list $(w_1, \ldots, w_m)$. Then
$$\text{span}(v_1, w_1, \ldots, w_m) = \text{span}(v_1, w_{j_1}, \ldots, w_{j_{m-1}}).$$

2. *The induction hypothesis and step.* Assume that the statement is true for some $n \ge 1$. Now let $(v_1, \ldots, v_n, v_{n+1})$ be independent in $\mathcal{V}$. Then $(v_1, \ldots, v_n)$ is also independent, so $n \le m$. If $n = m$, then $\mathcal{V} = \text{span}(v_1, \ldots, v_n)$, so $v_{n+1} \in \text{span}(v_1, \ldots, v_n)$. This contradicts the independence of $(v_1, \ldots, v_n, v_{n+1})$, so $n < m$, and therefore $n + 1 \le m$.

(i) *The case $n + 1 = m$.* If $n + 1 = m$, we have $\mathcal{V} = \text{span}(v_1, \ldots, v_n, w_{j_1})$ for some j_1. The list $(v_1, \ldots, v_n, v_{n+1}, w_{j_1})$ is dependent, and by the linear dependence lemma, the only possibility is that $w_{j_1} \in \text{span}(v_1, \ldots, v_n, v_{n+1})$. Thus $\mathcal{V} = \text{span}(v_1, \ldots, v_n, v_{n+1})$.

(ii) *The case $n + 1 < m$.* If $n + 1 < m$, we have $\mathcal{V} = \text{span}(v_1, \ldots, v_n, w_{j_1}, \ldots, w_{j_{n-m}})$ for some sublist $(w_{j_1}, \ldots, w_{j_{n-m}})$. Here $n - m > 1$, thus $n - m \ge 2$. Again, the list $(v_1, \ldots, v_n, v_{n+1}, w_{j_1}, \ldots, w_{j_{n-m}})$ is dependent, so either
$$w_{j_1} \in \text{span}(v_1, \ldots, v_n, v_{n+1})$$

or

$$w_{j_k} \in \operatorname{span}(v_1, \ldots, v_n, v_{n+1}, w_{j_1}, \ldots, w_{j_{k-1}}) \text{ for some } k \text{ with } 2 \le k \le n-m.$$

In the first case, we have

$$\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n, v_{n+1}, w_{j_2}, \ldots, w_{j_{n-m}}).$$

In the second case, let $(\widetilde{w}_{\ell_1}, \ldots, \widetilde{w}_{\ell_{n-m-1}})$ be the list formed by removing w_{j_k} from the list $(w_{j_1}, \ldots, w_{j_{n-m}})$. Then

$$\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n, v_{n+1}, \widetilde{w}_{\ell_1}, \ldots, \widetilde{w}_{\ell_{n-m-1}}).$$ ■

As is often the case in mathematics, what we really want is the corollary.

16.4 Corollary. *All bases for a finite-dimensional space have the same length. That is, if $(v_1, \ldots, v_n)$ and $(\widetilde{v}_1, \ldots, \widetilde{v}_p)$ are bases for $\mathcal{V}$, then $n = p$.*

Proof. The list $(v_1, \ldots, v_n)$ is a basis and therefore independent and the list $(\widetilde{v}_1, \ldots, \widetilde{v}_p)$ spans $\mathcal{V}$ (because it, too, is a basis), and so $n \le p$. Reverse the roles of the lists to conclude $p \le n$, thus $p = n$. ■

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.

16.5 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.

16.6 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 $\dim(\mathbb{F}^{m\times n}) = mn$.

16.7 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.

16.8 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 8.7.

The following two problems contain extremely useful results about independence, basis, dimension, and subspaces that we will use often.

16.9 Problem ($\star$). Prove (and state precisely) each of the following. Let $\mathcal{V}$ be a finite-dimensional vector space.

(i) Any independent list is no longer than any basis.

(ii) Any independent list of length $\dim(\mathcal{V})$ is a basis for $\mathcal{V}$.

(iii) Any spanning list of length $\dim(\mathcal{V})$ is a basis for $\mathcal{V}$. [Hint: *if it is not a basis, apply Lemma 15.3. What contradiction results?*]

(iv) Any independent list of length less than $\dim(\mathcal{V})$ can be extended to a basis for $\mathcal{V}$.

Content from *Linear Algebra* by Meckes & Meckes. Some of these results appear as Propositions 3.20 and 3.21 on p. 165 and Theorems 3.26 and 3.27 on p. 167 and Theorem 3.28 on p. 168. Read the example on p. 167.

16.10 Problem ($\star$). Let $\mathcal{V}$ be a vector space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Prove the following.

(i) If $\mathcal{V}$ is finite-dimensional, then $\mathcal{U}$ is also finite-dimensional and $\dim(\mathcal{U}) \le \dim(\mathcal{V})$.

(ii) If $\mathcal{V}$ is finite-dimensional and $\dim(\mathcal{U}) = \dim(\mathcal{V})$, then $\mathcal{U} = \mathcal{V}$.

Content from *Linear Algebra* by Meckes & Meckes. This problem is Theorem 3.29 on p. 168.

16.11 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 11.4) that if $\mathcal{T}$ is not injective and $\mathcal{T}v = w$ has *a* solution, then the problem has infinitely many solutions.

Day 17: Friday, September 26.

You took Exam 1.

Day 18: Monday, September 29.

Our first application of the results of Problem 16.9 will be to show that infinite-dimensional vector spaces exist.

18.1 Theorem. *If a vector space contains an independent list of any arbitrary length, the space is infinite-dimensional. That is, let $\mathcal{V}$ be a vector space such that for each integer $n \ge 1$, there is an independent list $(v_1, \ldots, v_n)$ in $\mathcal{V}$. Then $\mathcal{V}$ is infinite-dimensional.*

Proof. Suppose that $\mathcal{V}$ is finite-dimensional with $n = \dim(\mathcal{V})$. The hypotheses imply that $\mathcal{V}$ has an independent list of length $n + 1$. This is impossible in a finite-dimensional vector space. ■

Content from *Linear Algebra* by Meckes & Meckes. This is Theorem 3.18 on p. 163.

18.2 Example. For each $r \ge 1$, the space $\mathcal{C}^r([0,1])$ is infinite-dimensional. Here is why. Example 15.1 shows that the list $(p_0, \ldots, p_n)$ is independent in $\mathcal{C}^\infty([0,1])$ for any $n \ge 1$, where $p_j(x) = x^j$. This is also an independent list in $\mathcal{C}^r([0,1])$ for any r. Since $\mathcal{C}^r([0,1])$ contains an independent list of arbitrary (finite) length, it must be infinite-dimensional.

The Steinitz exchange lemma is very powerful because it relates the lengths of independent lists and spanning lists. Here is a slightly weaker result that more directly relates the lengths of bases and spanning lists. We discuss it to reinforce the proof technique of induction and to show an interesting role of linear operators in the proof. Our primary tool* is the following.

* The proofs of Lemma 18.3 and Theorem 18.5 are due to Jochen Gluck, available at https://mathoverflow.net/questions/499774/alternative-proofs-that-two-bases-of-a-vector-space-have-

18.3 Lemma. *Suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is not injective and $\mathcal{V} = \operatorname{span}(v_1, \ldots, v_n)$. Then the list $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ is dependent.*

Proof. Since $\mathcal{T}$ is not injective, there is $v \in \mathcal{V}$ such that $\mathcal{T}v = 0_{\mathcal{W}}$ and $v \neq 0_{\mathcal{V}}$. Write $v = \sum_{j=1}^{n} \alpha_j v_j$, where $\alpha_j \neq 0$ for at least one j (otherwise, $v = 0_{\mathcal{V}}$). Then $0_{\mathcal{W}} = \mathcal{T}v = \sum_{j=1}^{n} \alpha_j \mathcal{T}v_j$ with, again, $\alpha_j \neq 0$ for at least one j. ■

We also need the following result.

18.4 Problem ($\star$). **(i)** Show that surjections preserve spans. That is, if $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ 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 $\mathcal{W} \neq \operatorname{span}(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$.

18.5 Theorem. *No spanning list is shorter than any basis. We give two versions of this result.*

(i) *Let $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ be a list in $\mathbb{F}^n$ such that $\mathbb{F}^n = \operatorname{span}(\mathbf{w}_1, \ldots, \mathbf{w}_m)$. Then $n \leq m$.*

(ii) *Let $\mathcal{V}$ be a finite-dimensional vector space with $\mathcal{V} = \operatorname{span}(w_1, \ldots, w_m)$. Suppose that $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$. Then $n \leq m$.*

Proof. **(i)** We induct on n.

1. *The case $n = 1$.* Any list has length at least 1, so $1 \leq m$.

2. *The induction hypothesis and step.* Suppose that the result is true for some $n \geq 1$ and consider a spanning list $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ in $\mathbb{F}^{n+1}$. The goal is to show that $n + 1 \leq m$. Equivalently, we want $n \leq m-1$. This, combined with the nature of the induction hypothesis, suggests that we try to relate $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ to a spanning list for $\mathbb{F}^n$ of length $m - 1$. Then we will have $n \leq m - 1$ by the induction hypothesis.

We accomplish this in the most natural way of relating $\mathbb{F}^{n+1}$ to $\mathbb{F}^n$: just take the first n components of a vector in $\mathbb{F}^{n+1}$. Define

$$\mathcal{T} \colon \mathbb{F}^{n+1} \to \mathbb{F}^n \colon \mathbf{v} \mapsto (v_1, \ldots, v_n).$$

Then $\mathcal{T}$ is surjective, as $\mathcal{T}(v_1, \ldots, v_n, 0) = (v_1, \ldots, v_n)$ for any $(v_1, \ldots, v_n) \in \mathbb{F}^n$, and not injective, as $\mathcal{T}\mathbf{e}_{n+1} = \mathbf{0}_n$. By surjectivity, the list $(\mathcal{T}\mathbf{w}_1, \ldots, \mathcal{T}\mathbf{w}_m)$ spans $\mathbb{F}^n$. By the lack of injectivity, this list is also dependent, and so there exists a sublist of length $m - 1$ that also spans $\mathbb{F}^n$. By the induction hypothesis, $n \leq m - 1$, so $n + 1 \leq m$.

(ii) Define $\mathcal{T} \colon \mathbb{F}^n \to \mathcal{V} \colon (\alpha_1, \ldots, \alpha_n) \mapsto \sum_{j=1}^{n} \alpha_j v_j$. Then $\mathcal{T}$ is an isomorphism. In particular, $\mathcal{T}^{-1}$ is surjective, so $(\mathcal{T}^{-1}w_1, \ldots, \mathcal{T}^{-1}w_m)$ spans $\mathbb{F}^n$. Then $n \leq m$ by the previous part. ■

---

the-same-size.

We have previously seen that an operator on a vector space over $\mathbb{R}$ need not have eigenvalues (Example 7.4), and likewise an operator on an infinite-dimensional space also need not have eigenvalues (Example 7.1, Problems 6.19 and 7.3), but also that an operator on an infinite-dimensional space can have infinitely many eigenvalues (Examples 6.18 and 7.2). None of this can cannot happen on a finite-dimensional vector space over $\mathbb{C}$.

18.6 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

$$p(\mathcal{T}) := \sum_{k=1}^{n} a_k \mathcal{T}^k.$$

18.7 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.

18.8 Theorem (Fundamental theorem of algebra). *Let $p(z) = \sum_{k=0}^{n} a_k z^k$ be a polynomial of degree n with coefficients in $\mathbb{C}$: $a_k \in \mathbb{C}$, $a_n \neq 0$. There is a list $(z_1, \ldots, z_n) \in \mathbb{C}^n$*

such that

$$p(z) = a_n \prod_{j=1}^{n} (z - z_j).$$

Content from *Linear Algebra* by Meckes & Meckes. Page 217 discusses the FTA.

For example, $z^2+1 = (z+i)(z-i)$. Thus every polynomial p has (at least) two expressions: the Taylor expansion $p(z) = \sum_{k=0}^{n} a_k z^k$ and the factored form above. The key difference is that even though all of the coefficients a_k may be real, some or all of the roots z_j may be complex. Just consider $p(z) = z^2 + 1$.

We can also consider arbitrary operator products. If $\mathcal{V}$ is a vector space and $(\mathcal{S}_1, \ldots, \mathcal{S}_n)$ is a list in $\mathbf{L}(\mathcal{V})$, we put

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

Now let $\mathcal{V}$ be a vector space over $\mathbb{C}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. If a polynomial $p(z) = \sum_{k=0}^{n} a_k z^k$ with coefficients in $\mathbb{C}$ factors as

$$p(z) = a_n \prod_{j=1}^{n} (z - z_j),$$

do we have

$$p(\mathcal{T}) = a_n \prod_{j=1}^{n} (\mathcal{T} - z_j I)$$

as well?

18.9 Lemma. *Yes.*

Proof. We induct on n.

1. *The base case.* 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

$$a_1 \mathcal{T} + a_0 \mathcal{I}_{\mathcal{V}} = a_1 \left(\mathcal{T} - \left(\frac{a_0}{a_1} \right) \mathcal{I}_{\mathcal{V}} \right).$$

2. *The induction hypothesis and step.* 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

$$p(\mathcal{T}) = q(\mathcal{T})(\mathcal{T} - z_{n+1}\mathcal{I}_{\mathcal{V}}), \tag{18.1}$$

then we will be done.

We do this in two passes. First, we rewrite

$$\begin{aligned}
p(z) &= q(z)(z - z_{n+1}) \\
&= \sum_{k=0}^{n} b_k z^k (z - z_{n+1}) \\
&= \sum_{k=0}^{n} (b_k z^{k+1} - b_k z_{n+1} z^k) \\
&= \sum_{k=0}^{n} b_k z^{k+1} - \sum_{k=0}^{n} b_k z_{n+1} z^k \\
&= \sum_{k=1}^{n+1} b_{k-1} z^k - \sum_{k=0}^{n} b_k z_{n+1} z^k \\
&= -b_0 z_{n+1} + \sum_{k=1}^{n} (b_{k-1} - b_k z_{n+1}) z^k + b_n z^{n+1}.
\end{aligned} \tag{18.2}$$

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 (18.2) with z replaced by $\mathcal{T}$ shows

$$\begin{aligned} q(\mathcal{T})(\mathcal{T} - z_{n+1}\mathcal{I}_{\mathcal{V}}) &= \sum_{k=0}^{n} b_k \mathcal{T}^k (\mathcal{T} - z_{n+1}\mathcal{I}_{\mathcal{V}}) \\ &= -b_0 z_{n+1} I + \sum_{k=1}^{n} (b_{k-1} - b_k z_{n+1}) \mathcal{T}^k + b_n \mathcal{T}^{n+1} \text{ (this is the fruit of (18.2))} \\ &= \sum_{k=0}^{n+1} c_k \mathcal{T}^k = \sum_{k=0}^{n+1} a_k \mathcal{T}^k = p(\mathcal{T}). \end{aligned}$$

This is the desired equality (18.1). ■

18.10 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)* $\mathcal{T} - z_j I$ *is invertible for each* $1 \le j \le n$.]

Day 19: Wednesday, October 1.

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.

19.1 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,

$$A^2\mathbf{e}_1 + I_2\mathbf{e}_1 = \mathbf{0}_2,$$

and so

$$(\mathcal{M}_A^2 + \mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 = \mathbf{0}_2.$$

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

$$(\mathcal{M}_A + i\mathcal{I}_{\mathbb{C}^2})(\mathcal{M}_A - i\mathcal{I}_{\mathbb{C}^2})\mathbf{e}_1 = \mathbf{0}_2. \tag{19.1}$$

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 (19.1) 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 $-i$.

19.2 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.

The trick of Example 19.1 generalizes substantially.

19.3 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 $v \in \mathcal{V} \setminus \{0\}$ such that $\mathcal{T}v = \lambda v$.*

Proof. Let $v \in \mathcal{V} \setminus \{0_{\mathcal{V}}\}$ and let $n = \dim(\mathcal{V})$. The list $(v, \mathcal{T}v, \mathcal{T}^2v, \ldots, \mathcal{T}^nv)$ of length $n+1$ whose jth entry is $\mathcal{T}^jv$ 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

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

If $\alpha_{n+1} = 0$, let $m \le n$ be such that $\alpha_m \neq 0$ and $\alpha_j = 0$ for $j \ge m+1$. Then

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

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 $z_j \in \mathbb{C}$. Then, ignoring $\alpha_m \neq 0$,

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

If $(\mathcal{T} - z_j \mathcal{I}_{\mathcal{V}})v = 0_{\mathcal{V}}$, then $\mathcal{T}v = z_j v$, and since $v \neq 0_{\mathcal{V}}$, this shows that v is an eigenvector of $\mathcal{T}$ with eigenvalue λ. Otherwise, let $2 \le r < m$ be such that

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

Such an r must exist, as if it does not, we would have $\prod_{j=1}^{m} (\mathcal{T} - z_j \mathcal{I}_{\mathcal{V}}) v \neq 0_{\mathcal{V}}$. Then

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

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.

It has taken a fair amount of machinery, but we have a powerful result: if a vector space can be expressed as the span of a finite list of its vectors, then it can be expressed as a span of a finite list of its vectors with superb efficiency. This is the double efficiency of a basis: every vector has a unique representation as a linear combination of vectors in the basis, and every basis has the same length. The greater service of these results is not, however, to vector space structure but to *operator behavior* (an example of which we just saw regarding eigenvalues). In particular, dimension counting gives us powerful control over the existence and uniqueness of solutions to our fundamental problem.

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

(i) *Suppose that $\mathcal{W}$ is finite-dimensional and $\mathcal{T}$ is surjective. Then either $\mathcal{V}$ is infinite-dimensional or* $\dim(\mathcal{W}) \le \dim(\mathcal{V})$.

(ii) *Suppose that $\mathcal{V}$ is finite-dimensional and $\mathcal{T}$ is injective. Then either $\mathcal{W}$ is infinite-dimensional or* $\dim(\mathcal{V}) \le \dim(\mathcal{W})$.

(iii) *Suppose that $\mathcal{T}$ 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. We leave the proofs of the first two parts as exercises. The third part is a mostly immediate consequence of the first two. If $\mathcal{V}$ is finite-dimensional with basis $(v_1, \ldots, v_n)$, then since $\mathcal{T}$ is surjective, Problem 18.4 shows that the list $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ spans $\mathcal{W}$, so $\mathcal{W}$ is finite-dimensional as well. The second part then implies $\dim(\mathcal{V}) \le \dim(\mathcal{W})$, and the first part implies $\dim(\mathcal{W}) \le \dim(\mathcal{V})$. If we only know that $\mathcal{W}$ is finite-dimensional, reverse the preceding argument with $\mathcal{T}^{-1}$ in place of $\mathcal{T}$. ■

The contrapositives are quite useful, in a negative sense. The first 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.)

The second 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.)

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.

Content from *Linear Algebra* by Meckes & Meckes. Proposition 3.39 presents the contrapositives (based on a different proof). Read the paragraph on p.181 that begins "Proposition 3.39 also has important geometric consequences…" for an enlightening discussion of those geometric consequences.

The following problems outline approaches to proving the first two parts of Theorem 19.4.

19.5 Problem ($\star$). This problem completes the proof of part (i) of Theorem 19.4. You should reread (and maybe redo) Problem 18.4 along with this problem, too. Let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ be surjective and suppose that $\dim(\mathcal{W}) = m$. Show that either $\mathcal{V}$ is infinite-dimensional or $\mathcal{V}$ is finite-dimensional with $m \le \dim(\mathcal{V})$. [Hint: *given a basis* $(w_1, \ldots, w_m)$ *for $\mathcal{W}$, let $v_j \in \mathcal{V}$ satisfy $\mathcal{T}v_j = w_j$. Argue that $(v_1, \ldots, v_m)$ is independent.*]

19.6 Problem ($\star$). This problem studies part (ii) of Theorem 19.4.

(i) Show that the image of an independent list under an injection is still independent. That is, prove that if $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is injective and $(v_1, \ldots, v_n)$ is independent, then $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ is also independent.

(ii) Give an example to show how this result may fail if $\mathcal{T}$ is not injective.

(iii) Let $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ be injective and $\dim(\mathcal{V}) = n$. Show that $\dim(\mathcal{T}(\mathcal{V})) = n$ and thus $n \le \dim(\mathcal{W})$.

We can prove a stronger result than part (iii) of Theorem 19.4 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. The tool that we use encourages us to consider how bases *discretize* problems: they reduce consideration from an entire vector space to just a (very special) list of vectors. Here is one way that this discretization shows up in encounters with operators.

19.7 Theorem. *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*

$$\mathcal{T}\sum_{j=1}^n \alpha_j v_j = \sum_{j=1}^n \alpha_j w_j. \tag{19.2}$$

This method of defining $\mathcal{T}$ is EXTENSION BY LINEARITY.

Proof. **1.** *Uniqueness.* First we should see why (19.2) is the “right” definition of $\mathcal{T}$. If $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ satisfies $\mathcal{T}v_j = w_j$, and if $v \in \mathcal{V}$ has the expansion $v = \sum_{j=1}^n \alpha_j v_j$, then

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

This also basically leads to a proof of uniqueness for $\mathcal{T}$. Suppose that $\mathcal{S} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ also satisfies $\mathcal{S}v_j = w_j$ for each j. We want to show that $\mathcal{T} = \mathcal{S}$, equivalently, that $\mathcal{T}v = \mathcal{S}v$ for each $v \in \mathcal{V}$. So, fix $v \in \mathcal{V}$ and write $v = \sum_{j=1}^n \alpha_j v_j$. Then

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

$$= \mathcal{S}\sum_{j=1}^{n} \alpha_j v_j \text{ by the linearity of } \mathcal{S}$$
$$= \mathcal{S}v \text{ by definition of } v.$$

This calculation did not use the linearity of $\mathcal{T}$ as defined in (19.2) 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 (19.2) actually gives a function in $\mathcal{W}^{\mathcal{V}}$. We are saying that

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

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 $\mathcal{T}v$. If we could write $v = \sum_{j=1}^{n} \alpha_j v_j$ and $v = \sum_{j=1}^{n} \beta_j v_j$ with perhaps $\alpha_k \neq \beta_k$ for some k, then our definition of $\mathcal{T}$ could yield two different values for $\mathcal{T}v$! But the representation of v is unique, since $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$.

One (possibly excessive) way to view this set-theoretically is to put

$$\mathcal{T}_1 := \left\{ (v, (\alpha_1, \ldots, \alpha_n)) \in \mathcal{V} \times \mathbb{F}^n \;\middle|\; v = \sum_{j=1}^{n} \alpha_j v_j \right\}$$

and

$$\mathcal{T}_2 := \left\{ \left((\alpha_1, \ldots, \alpha_n), \sum_{j=1}^{n} \alpha_j w_j \right) \;\middle|\; (\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n \right\}.$$

The unique representation property of a basis ensures $\mathcal{T}_1 \in (\mathbb{F}^n)^{\mathcal{V}}$. In fact, $\mathcal{T}_1$ is the inverse of the operator from Problem 12.17. And certainly $\mathcal{T}_2 \in \mathcal{W}^{\mathbb{F}^n}$. Then we can define $\mathcal{T}$ as the composition $\mathcal{T} := \mathcal{T}_2 \circ \mathcal{T}_1$, or $\mathcal{T}(v) = \mathcal{T}_2(\mathcal{T}_1(v))$. Problem 12.17 and the linearity of the inverse of a linear operator ensure that $\mathcal{T}_1$ is linear, and we could also check the linearity of $\mathcal{T}_2$. We do that, effectively, in the next step.

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. ■

19.8 Problem (!). Finish the proof.

Content from *Linear Algebra* by Meckes & Meckes. "Extension by linearity" is Theorem 3.14 on p. 155.

19.9 Problem (!). 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.

19.10 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 11.5) 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?*]

19.11 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* $\sum_{j=1}^n v_j$.]

Day 20: Friday, October 3.

Vocabulary from today

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

Dual basis, double algebraic dual space

Now here is the improvement of part (iii) of Theorem 19.4.

20.1 Theorem. *Let* $\mathcal{V}$ *and* $\mathcal{W}$ *be 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 $n = \dim(\mathcal{V}) = \dim(\mathcal{W})$.

1. *Construction of the isomorphism.* Let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$ and $(w_1, \ldots, w_n)$ be a basis for $\mathcal{W}$. We claim that

$$\mathcal{T} \colon \mathcal{V} \to \mathcal{W} \colon \sum_{j=1}^n \alpha_j v_j \mapsto \sum_{j=1}^n \alpha_j w_j$$

is an isomorphism. Theorem 19.7 assures us that $\mathcal{T} \in \mathcal{W}^{\mathcal{V}}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, so we just have to check that $\mathcal{T}$ is bijective.

2. *Checking surjectivity.* Let $w \in \mathcal{W}$. We want to find $v \in \mathcal{V}$ such that $\mathcal{T}v = w$. Since $\mathcal{W} = \operatorname{span}(w_1, \ldots, w_n)$, we may write $w = \sum_{j=1}^n \alpha_j w_j$ for some $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$. With $v := \sum_{j=1}^n \alpha_j v_j$, the definition of $\mathcal{T}$ gives $\mathcal{T}v = w$.

3. *Checking injectivity.* Suppose $\mathcal{T}v = 0_{\mathcal{W}}$. We want to show that $v = 0_{\mathcal{V}}$. With $v = \sum_{j=1}^n \alpha_j v_j$, so

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

Since $(w_1, \ldots, w_n)$ is independent, $\alpha_j = 0$ for all j, and therefore $v = 0_{\mathcal{V}}$.

($\Longleftarrow$) This is part (iii) of Theorem 19.4. ■

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.

20.2 Example. **(i)** Part (i) of Example 11.11 showed that $\mathbb{P}^n$ and $\mathbb{F}^{n+1}$ are isomorphic, so $\dim(\mathbb{P}^n) = \dim(\mathbb{F}^{n+1}) = n + 1$.

(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 12.1, $\mathbf{L}(\mathcal{V}, \mathcal{W})$ and $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ are isomorphic. And by part (ii) of Example 11.11, $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ 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 12.5, and so

$$\dim(\mathbf{L}(\mathcal{V}, \mathcal{W})) = \dim(\mathbb{F}^{m \times n}) = mn$$

by part (iii) of Example 16.6.

The following gives us practice with linear functionals and dual spaces via dimension counting.

20.3 Example. Let $\mathcal{V}$ be a finite-dimensional vector space. Recall that the dual space of $\mathcal{V}$ is $\mathcal{V}' := \mathbf{L}(\mathcal{V}, \mathbb{F})$. Then $\dim(\mathcal{V}') = \dim(\mathbf{L}(\mathcal{V}, \mathbb{F})) = \dim(\mathcal{V}) \dim(\mathbb{F}) = \dim(\mathcal{V})$, so $\mathcal{V}$ and $\mathcal{V}'$ are isomorphic. We analyze this isomorphism in the following steps.

1. If $\dim(\mathcal{V}) = n$, then there are bases $(v_1, \ldots, v_n)$ for $\mathcal{V}$ and $(\varphi_1, \ldots, \varphi_n)$ for $\mathcal{V}'$ of the same length. Any $v \in \mathcal{V}$ can be written as $v = \sum_{j=1}^n \alpha_j v_j$, while any $\varphi \in \mathcal{V}'$ has the form $\varphi = \sum_{k=1}^n \beta_k \varphi_k$. Then

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

This is a pretty awful expression; uncharacteristically, bases have not made things simpler.

The problem is that we have not chosen the *right* bases here, or more precisely the right basis for $\mathcal{V}'$. Things *would* be much simpler if we had better control over $\varphi_k(v_j)$.

2. Start again with the basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$. We can define a linear operator $\varphi \in \mathbf{L}(\mathcal{V}, \mathbb{F}) = \mathcal{V}'$ by specifying its values on this basis. Specifically, for $k = 1, \ldots, n$, let

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

Then if $v = \sum_{j=1}^{n} \alpha_j v_j$, we have $\varphi_k(v) = \alpha_k$. We can therefore think of φ_k as the kth "coordinate functional" on $\mathcal{V}$, and the list $(\varphi_1, \ldots, \varphi_n)$ turns out to be a basis for $\mathcal{V}'$, which we call the **DUAL BASIS FOR $\mathcal{V}'$ RELATIVE TO THE BASIS $(v_1, \ldots, v_n)$ FOR $\mathcal{V}$.**

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 extract from this example two useful identities: if $\mathcal{V}$ is a finite-dimensional space with $\dim(\mathcal{V}) = n \geq 1$, basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$, and corresponding dual basis $(\varphi_1, \ldots, \varphi_n)$ for $\mathcal{V}'$, then

$$v = \sum_{j=1}^{n} \varphi_j(v) v_j, \ v \in \mathcal{V} \quad \text{and} \quad \varphi = \sum_{j=1}^{n} \varphi(v_j) \varphi_j, \ \varphi \in \mathcal{V}'.$$

20.4 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* $k = 1, \ldots, n$.]

20.5 Problem ($\star$). 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. Let $\mathcal{V}$ be a finite-dimensional vector space and $v \in \mathcal{V}$. Prove that $v = 0_{\mathcal{V}}$ if and only if $\varphi(v) = 0$ for all $\varphi \in \mathcal{V}'$. [Hint: *use the generosity of the universal quantifier to evaluate the dual basis vectors at* v.]

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 $\varphi_1,\ \varphi_2 \in \mathcal{V}'$ (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 Definition. *Let* $\mathcal{V}$ *be a vector space. The* **DOUBLE (ALGEBRAIC) DUAL SPACE** *of* $\mathcal{V}$ *is* $\mathcal{V}'' := (\mathcal{V}')' = \mathbf{L}(\mathcal{V}', \mathbb{F})$.

By Example 20.3, 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.

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

$$\varphi_k'(v_j) = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \qquad \text{and} \qquad \varphi_\ell''(\varphi_k) = \begin{cases} 1, & \ell = k \\ 0, & \ell \neq k \end{cases} \tag{20.2}$$

as well as the representations

$$v = \sum_{j=1}^n \varphi_j'(v)v_j,\ v \in \mathcal{V} \qquad \text{and} \qquad \varphi' = \sum_{k=1}^n \varphi_k''(\varphi)\varphi_k',\ \varphi' \in \mathcal{V}'. \tag{20.3}$$

If it looks as though we are defining v and φ' in terms of themselves, 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."

For the latter sum in (20.3), this means

$$\varphi'(v) = \sum_{k=1}^n \varphi_k''(\varphi)\varphi_k'(v),\ v \in \mathcal{V},$$

so in particular taking $v = v_j$, we get

$$\varphi'(v_j) = \sum_{k=1}^n \varphi_k''(\varphi)\varphi_k'(v_j) = \varphi_j''(\varphi).$$

We record this for future use.

20.8 Lemma. *Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim(\mathcal{V}) = n \ge 1$. 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." This sort of representation turns out to be true for all $\varphi'' \in \mathcal{V}''$ when $\mathcal{V}$ is finite-dimensional. First we check that the "evaluate at" functional is indeed a linear functional on $\mathcal{V}'$.

20.9 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

$$\mathcal{J} \colon \mathcal{V} \to \mathcal{V}'' \colon v \mapsto \varphi_v'' \tag{20.4}$$

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

Day 21: Monday, October 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."

Finite-rank operator (N), rank of a linear operator

Now we prove that "evaluate at" 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.

21.1 Theorem. *Let $\mathcal{V}$ be a finite-dimensional vector space. The map $\mathcal{J} \in \mathbf{L}(\mathcal{V}, \mathcal{V}'')$ defined in (20.4) 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 20.5, 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 $(\varphi_1', \ldots, \varphi_n')$. This allows us to use (20.2) and (20.3).

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 $\varphi' \in \mathcal{V}'$. Then in particular $\varphi''(\varphi_k') = \varphi_k'(v)$, and so we are led to take $v = \sum_{k=1}^n \varphi''(\varphi_k')v_k$.

3. Here is the actual proof of surjectivity. Let $\varphi'' \in \mathcal{V}''$ and put $v = \sum_{k=1}^n \varphi''(\varphi_k')v_k$. Let $\varphi' \in \mathcal{V}'$. We show that $\varphi''(\varphi') = \varphi'(v)$.

The representation $\varphi' = \sum_{\ell=1}^n \varphi_\ell''(\varphi')\varphi_\ell'$ gives

$$\varphi''(\varphi') = \varphi''\left(\sum_{\ell=1}^n \varphi_\ell''(\varphi')\varphi_\ell'\right) = \sum_{\ell=1}^n \varphi_\ell''(\varphi')\varphi''(\varphi_\ell').$$

The representation $v = \sum_{k=1}^n \varphi_k'(v)v_k$ gives

$$\varphi'(v) = \varphi'\left(\sum_{k=1}^n \varphi''(\varphi_k')v_k\right) = \sum_{k=1}^n \varphi''(\varphi_k')\varphi'(v_k).$$

Comparing these calculations (and adjusting the dummy index of summation), we just need to show that $\varphi_k''(\varphi') = \varphi'(v_k)$. And this is the result of Lemma 20.8, which motivated our entire study of "evaluate at" in the first place. ■

21.2 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 $\mathcal{V}$.*]

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.

21.3 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces. 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 $\operatorname{rank}(\mathcal{T}) := \dim(\mathcal{T}(\mathcal{V}))$.*

21.4 Problem (!). **(i)** Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Prove that $\mathcal{T}$ is finite-rank with $\operatorname{rank}(\mathcal{T}) = 0$ if and only if $\mathcal{T} = 0_{\mathcal{V} \to \mathcal{W}}$, i.e., the zero operator.

(ii) Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces with $\mathcal{V}$ finite-dimensional. Prove that every operator in $\mathbf{L}(\mathcal{V}, \mathcal{W})$ is finite-rank.

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

21.5 Problem ($\star$). Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces with $\mathcal{U}$ finite-dimensional. 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?

The "nullity" of a linear operator is a somewhat old-fashioned way of referring to the dimension of $\ker(\mathcal{T})$; "rank" is rather more in vogue.

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.

21.6 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*

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

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."

Day 22: Wednesday, October 8.

22.1 Example. We can use rank–nullity to prove a curious result about linear functionals. Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim(\mathcal{V}) = n \geq 1$. 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 $v_0 \in \mathcal{V}$.

Now we prove rank–nullity.

Proof (of Theorem 21.6). 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 $\operatorname{span}(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n) = \mathcal{T}(\mathcal{V})$. For $v \in \mathcal{V}$, write $v = \sum_{j=1}^n \alpha_j v_j$, so $\mathcal{T}v = \sum_{j=1}^n \alpha_j \mathcal{T}v_j = \sum_{j=p+1}^n \alpha_j \mathcal{T}v_j$. Thus $\mathcal{T}v \in \operatorname{span}(\mathcal{T}v_{p+1}, \ldots, \mathcal{T}v_n)$, and certainly this span is contained in $\mathcal{T}(\mathcal{V})$.

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 $\sum_{j=p+1}^n \alpha_j v_j = \sum_{j=1}^p \beta_j v_j$. Rearrange this to read $\sum_{j=1}^n \gamma_j v_j = 0_{\mathcal{V}}$, where

$$\gamma_j := \begin{cases} \alpha_j, & 1 \le j \le p \\ -\beta_j, & p+1 \le j \le n. \end{cases}$$

Since $(v_1, \ldots, v_n)$ is a basis for $\mathcal{V}$, we have $\gamma_j = 0$, so in. particular $\alpha_j = 0$. ■

22.2 Problem (!). Finish the proof of rank–nullity by treating the cases $\dim(\ker(\mathcal{T})) = 0$ and $\dim(\ker(\mathcal{T})) = n = \dim(\mathcal{V})$.

22.3 Problem ($\star$). Try proving the rank–nullity theorem in a different way. Let $\dim(\mathcal{V}) = n$ and $\dim(\mathcal{T}(\mathcal{V})) = r$. If $r = 0$, then $\mathcal{T}(\mathcal{V}) = \{0_{\mathcal{W}}\}$ and $\mathcal{T} = 0_{\mathcal{V}\to\mathcal{W}}$, thus $\ker(\mathcal{T}) = \mathcal{V}$, and there is nothing to prove; if $r = n$, then given a basis $(v_1, \ldots, v_n)$ for $\mathcal{V}$, the list $(\mathcal{T}v_1, \ldots, \mathcal{T}v_n)$ spans $\mathcal{T}(\mathcal{V})$ and therefore is a basis for $\mathcal{V}$, so $\mathcal{T}v_j \ne 0_{\mathcal{W}}$, and therefore $\ker(\mathcal{T}) = \{0_{\mathcal{V}}\}$. So, assume $1 \le r < n$ and let $(w_1, \ldots, w_r)$ be a basis for $\mathcal{T}(\mathcal{V})$. Try to show then that $\dim(\ker(\mathcal{T})) = n - r$. Describe as precisely as possible where you get stuck.

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

22.4 Problem (!). Use the rank–nullity theorem to explain why we could have stopped the proof of Theorem 21.1 after proving the injectivity of $\mathcal{J}$ (which was probably easier than proving surjectivity).

22.5 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces. Use rank–nullity to prove the contrapositives of parts (i) and (ii) of Theorem 19.4. (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* $\dim(\mathcal{V}) \ge \dim(\mathcal{W})$.]

(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.*

22.6 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.

Day 23: Friday, October 10.

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."

RREF of a matrix (N—be able to give an example of a matrix not in RREF)

A first course in linear algebra typically develops many results about basis and dimension from the point of view of matrices and subspaces of Euclidean space. There is a fairly transparent way to prove the rank–nullity theorem for matrix multiplication operators from the reduced row echelon form (RREF) of a matrix. We briefly sketch this approach here, starting with an existential construction of the RREF that does not rely on Gaussian elimination at all.

23.1 Definition. *The* RANK *of* $A \in \mathbb{F}^{m \times n}$ *is the dimension of its column space:*

$$\operatorname{rank}(A) := \dim(\mathbf{C}(A)) = \dim(\mathcal{M}_A(\mathbb{F}^n)).$$

23.2 Problem (!). Here is a version of Problem 22.5 for matrices. Let $A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix} \in \mathbb{F}^{m \times n}$. First explain why $\operatorname{rank}(A) \le \min\{m, n\}$. Then prove the following. (In all arguments, you may use Problem 22.5 and rank–nullity, but do not refer to any standard results about the RREF or pivots. The logically correct, if pedagogically unsound, goal here is to develop properties of matrices from the more abstract operator results.)

(i) If $m < n$, then the list $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ in $\mathbb{F}^m$ is dependent and $\ker(A) \neq \{\mathbf{0}_n\}$.

(ii) If $n < m$, then $\mathbf{C}(A) \neq \mathbb{R}^m$.

(iii) If $m = n$, then there is a matrix $B \in \mathbb{F}^{n \times n}$ such that $AB = I_n$ if and only if there is a matrix $C \in \mathbb{F}^{m \times m}$ such that $CA = I_m$.

A major result of a first course in linear algebra should be that for any $A \in \mathbb{F}^{m \times n}$, there is an invertible matrix $E \in \mathbb{F}^{m \times m}$ such that EA is in reduced row echelon form, sometimes "row reduced echelon form," and either way RREF. We refresh our memory of the RREF with the following six examples, which effectively encompass all possible cases.

23.3 Example. The matrices

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

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

all have something in common (beyond their rank, which is 2). Actually, they have the following four properties in common.

- Any row whose entries are all 0 is below any row with nonzero entries.

- If a row has nonzero entries, then the first nonzero entry in that row is 1.

- Such a "leading 1" is the only nonzero entry in its column.

- If $i < j$ and row j has a leading 1, then the leading 1 of row j appears in a column after the column of the leading 1 in row i. (This is possibly the hardest condition of the RREF to state in language that is both technically precise and intuitively evocative.)

We should also note that by "shuffling" the columns of the fourth and sixth matrices, we can make the 2×2 identity matrix appear in a block. Put

$$P_{23} := \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_3 & \mathbf{e}_2 & \mathbf{e}_4 \end{bmatrix} \in \mathbb{F}^{4 \times 4}.$$

Then

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

It turns out that the four (or six, with shuffling) matrix structures in this example are "canonical": every matrix can be converted into one of these forms. First we name this form.

23.4 Definition. *A matrix $R \in \mathbb{F}^{m \times n}$ is in* REDUCED ROW ECHELON FORM *or* ROW REDUCED ECHELON FORM, *abbreviated either way as* **RREF**, *if it has the following four properties.*

Row Property 1. *Any row of R whose entries are all zero is below any row with some nonzero entries. That is, if the entries of row i_2 of R are all zero, and row i_1 has at least one nonzero entry, then $i_1 < i_2$.*

Row Property 2. *If a row contains nonzero entries, the first nonzero entry of that row is 1. This entry is called the* LEADING 1 *or the* PIVOT *for that row.*

Column Property 1. *The other entries of any column containing a leading one are 0. That is, a column containing a leading 1 is a standard basis vector for $\mathbb{F}^m$.*

Column Property 2. *Suppose that $1 \le i_1 < i_2 \le m$ and rows i_1 and i_2 both contain nonzero entries. Let the leading 1 of row i_1 be an entry in column j_{i_1} and the leading 1 of row i_2 be an entry in column j_{i_2}. Then $j_{i_1} < j_{i_2}$. That is, the leading 1 of a given row is "to the left" of the leading 1's in the rows below.*

23.5 Problem (!). Explain *all* of the reasons why

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

is not in RREF.

Now we state precisely the major result on reducing (nonzero) every matrix to RREF.

23.6 Theorem (RREF). *Let $A \in \mathbb{F}^{m \times n}$ be nonzero. There exists an invertible matrix $E \in \mathbb{F}^{m \times m}$ and a unique matrix $R_0 \in \mathbb{F}^{m \times n}$ such that $EA = R_0$ and R_0 is in RREF. The matrix R_0 is the* ROW REDUCED ECHELON FORM, *or* REDUCED ROW ECHELON FORM, *of A, and we write $\mathsf{rref}(A) := R_0$. In particular, R_0 has one of the following four possible forms, depending on the value of $r := \operatorname{rank}(A)$.*

(i) $r = m = n$. *Then $R_0 = I_n$.*

(ii) $r = n < m$. *Then*

$$R_0 = \begin{bmatrix} I_n \\ 0 \end{bmatrix}.$$

(iii) $r = m < n$. *Then*

$$R_0 = \begin{bmatrix} I_m & F \end{bmatrix} P$$

for some permutation matrix $P \in \mathbb{F}^{n \times n}$ *and* $F \in \mathbb{F}^{m \times (n-m)}$.

(iv) $1 \le \operatorname{rank}(A) < \min\{m, n\}$. *Then*

$$R_0 = \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P$$

for some permutation matrix $P \in \mathbb{F}^{n \times n}$ *and* $F \in \mathbb{F}^{r \times (n-m)}$.

We will prove this theorem later. For now, we use it to prove rank–nullity for matrices.

Content from *Linear Algebra* by Meckes & Meckes. Page 15 defines the RREF and Theorem 1.1 on that page outlines the standard existence proof using Gaussian elimination and elementary row operations. Do Quick Exercise #8 on p. 16. Section 2.4 reviews encoding those elementary row operations via matrix multiplication. This is Theorem 2.21 on p. 103, and the factorization $EA = \mathsf{rref}(A)$ is Theorem 2.24 on p. 105. Do Quick Exercises #16 on p. 104 and #17 on p. 105.

23.7 Theorem (Rank–nullity for matrices). *Let* $A \in \mathbb{F}^{m \times n}$. *Then*

$$\dim(\mathbf{N}(A)) + \operatorname{rank}(A) = n.$$

Proof. If A is the zero matrix, then $\mathbf{N}(A) = \mathbb{F}^n$ and $\mathbf{C}(A) = \{\mathbf{0}_m\}$, so $\dim(\mathbf{N}(A)) = n$, $\dim(\mathbf{C}(A)) = 0$, and the result is immediate. Now assume $r := \operatorname{rank}(A) \ge 1$. Then $\mathsf{rref}(A)$ has one of the four forms from Theorem 23.6. We assume that it is it the last: there is an invertible matrix $E \in \mathbb{F}^{m \times m}$, a matrix $F \in \mathbb{F}^{r \times (n-r)}$, and a permutation matrix $P \in \mathbb{F}^{n \times n}$ such that

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

Here we adopt the sometimes uncomfortable custom that the "row" of zero blocks, or the block of F, could be missing, which really would allow $\mathsf{rref}(A)$ to have one of the other three forms in Theorem 23.6. With this convention, it will not be necessary to treat separately the cases $r = m = n$, or $r = m < n$, or $r = n < m$. (Alternatively, one could prove this theorem for each of those cases individually.)

We want to show that $\dim(\mathbf{N}(A)) = n - r$. First, we have $A\mathbf{x} = \mathbf{0}_m$ if and only if

$$E^{-1} \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P\mathbf{x} = \mathbf{0}_m.$$

Put $\mathbf{y} = P\mathbf{x}$, so this is equivalent to

$$\begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} \mathbf{y} = \mathbf{0}_m.$$

Now compress $\mathbf{y} = (\mathbf{y}^{(r)}, \mathbf{y}_{(n-r)})$, with possibly $\mathbf{y}_{(n-r)}$ absent if $n = r$. We therefore have $A\mathbf{x} = \mathbf{0}_m$ if and only if

$$\mathbf{0}_r = I_r \mathbf{y}^{(r)} + F\mathbf{y}_{(n-r)} = \mathbf{y}^{(r)} + F\mathbf{y}_{(n-r)}.$$

If $n = r$ and $\mathbf{y}_{(n-r)}$ is absent, then this says $\mathbf{y} = \mathbf{y}^{(r)} = \mathbf{0}_r = \mathbf{0}_n$, thus $P\mathbf{x} = \mathbf{0}_n$. Since P is a permutation matrix and therefore invertible, we get $\mathbf{x} = \mathbf{0}_n$, in which case $\mathbf{N}(A)$ is trivial (as expected).

If $r < n$, then we have $\mathbf{y}^{(r)} = -F\mathbf{y}^{(n-r)}$. Then

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}^{(r)} \\ \mathbf{y}^{(n-r)} \end{bmatrix} = \begin{bmatrix} -F\mathbf{y}^{(n-r)} \\ \mathbf{y}^{(n-r)} \end{bmatrix} = \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{y}^{(n-r)}.$$

Since $\mathbf{x} = P^{-1}\mathbf{y}$, we conclude

$$\mathbf{x} = P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \mathbf{y}^{(n-r)}.$$

All together, we have shown

$$\mathbf{N}(A) \subseteq \mathbf{C}\left(P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right).$$

The reverse inclusion is also true, and so we have

$$\mathbf{N}(A) = \mathbf{C}\left(P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right), \tag{23.1}$$

Thus

$$\dim(\mathbf{N}(A)) = \operatorname{rank}\left(P^{-1} \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix}\right) = n - r. \tag{23.2}$$

■

23.8 Problem (!). Prove (23.1) and (23.2).

Content from *Linear Algebra* by Meckes & Meckes. This version of rank–nullity is (still) Theorem 3.35 on p. 175 with its matrix-oriented proof.

23.9 Problem ($\star$). Let

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

Express $\mathbf{N}(A)$ as the column space of some matrix, and show that this other matrix has the form

$$P^{-1} \begin{bmatrix} -F \\ I_2 \end{bmatrix}$$

for some permutation matrix P and some block F. Is F what you expect it to be? [Hint: *feel free to cite Example 1.1 throughout your work.*]

Now we develop the factorization of a matrix into its RREF. We work only with nonzero matrices. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{F}^{m \times n}$ be a nonzero matrix, so A has at least one nonzero entry. Let $\operatorname{rank}(A) = r$, so $1 \le r \le \min\{m, n\}$. We construct a basis for $\mathbf{C}(A)$ as follows.

Let j_1 be the smallest integer in $\{1, \ldots, n\}$ such that $a_{j_1} \neq \mathbf{0}_m$. That is, $1 \leq j_1 \leq n$, and if $j_1 \geq 2$, then $\mathbf{a}_j = \mathbf{0}_m$ for $j = 1, \ldots, j_1 - 1$. If $j_1 = n$, stop here. Otherwise, proceed recursively and, for $2 \leq i \leq r$, let j_i be the least integer greater than j_{i-1} such that $\mathbf{a}_{j_i} \notin \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$. That is, $j_{i-1} + 1 \leq j_i \leq n$, and if $j_{i-1} + 1 < j_i$, then $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$ for $j = j_{i-1} + 1, \ldots, j_i - 1$.

Each list $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_s})$ is independent by the linear (in)dependence lemma, so we cannot continue this procedure *past* column $k = r$. Otherwise, we would have an independent list of length $r + 1$ in $\mathbf{C}(A)$. But we can continue this process *through* column $k = r$. Otherwise, for some $1 \leq s < r$, every column of A would be in $\operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_s})$, and then $\dim(\mathbf{C}(A)) \leq s < \operatorname{rank}(A)$.

We call these columns $\mathbf{a}_{j_i}$ of A selected in this manner the **PIVOT COLUMNS** of A. It is not possible to choose the pivot columns in more than one way: the first pivot column is the first nonzero column of A, while the ith pivot column ($i \geq 2$) is the first column of A not in the span of the previous $i - 1$ pivot columns. That is, if $\operatorname{rank}(A) = r$, then there is a unique list $(j_1, \ldots, j_r)$ in $\{1, \ldots, n\}$ such that $j_i < j_{i+1}$ for $i = 1, \ldots, r - 1$ and that each entry in $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$ is a pivot column of A.

Here is a summary of what we have established.

23.10 Lemma. *Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{F}^{m \times n}$ with $\operatorname{rank}(A) = r \geq 1$. There exists a unique list $(j_1, \ldots, j_r)$ in $\{1, \ldots, n\}$ such that the following hold.*

(i) *If $j < j_1$, then $\mathbf{a}_j = \mathbf{0}_m$.*

(ii) $\mathbf{a}_{j_1} \neq \mathbf{0}_m$.

(iii) *If $r \geq 2$ and $i \geq 2$, then $\mathbf{a}_{j_i} \notin \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$.*

(iv) *If $r \geq 2$, $i \geq 2$, and $j < j_i$, then $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$. If $j > j_r$, then $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$.*

Proof. We did not explicitly prove the last part, so we do so now. If $j < j_1$, then $\mathbf{a}_j = \mathbf{0}_m$, as otherwise the first nonzero column would occur before column j_1, which, by definition, *is* the first nonzero column. And if $j_i < j < j_{i+1}$, then column j is not a pivot column. But if $\mathbf{a}_j \notin \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_i})$, then column j_{i+1} would not be the first column after column j_i not to be in $\operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_i})$. So, $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_i})$. ∎

Parts (i) and (ii) of Lemma 23.10 together encode the idea that the first pivot column of a matrix is the first nonzero column of that matrix. Parts (iii) and (iv) together encode the idea that the ith pivot column of a matrix ($i \geq 2$) is the first column of that matrix not in the span of the previous $i - 1$ pivot columns.

This notation is getting burdensome, so we should pause for an example.

23.11 Example. Let
$$A = \begin{bmatrix} 0 & 1 & 2 & 1 & 7 \\ 0 & 2 & 4 & 2 & 14 \\ 0 & 0 & 0 & 2 & 8 \end{bmatrix}.$$
We should recognize this matrix as a variant of our very first Example 1.1. Here $\mathbf{a}_1 = \mathbf{0}_3$, $\mathbf{a}_2$ is the first nonzero column, $\mathbf{a}_3 = 2\mathbf{a}_1$, $\mathbf{a}_4$ is not a multiple of $\mathbf{a}_2$, and $\mathbf{a}_5 = 3\mathbf{a}_2 + 4\mathbf{a}_4$. So, the list $(\mathbf{a}_2, \mathbf{a}_4)$ is independent and spans $\mathbf{C}(A)$, thus $r = \operatorname{rank}(A) = 2$, $j_1 = 2$, and $j_2 = 4$.

Day 24: Monday, October 13.

We pause to check the consistency of our pivot terminology.

24.1 Lemma. *Let $R \in \mathbb{F}^{m \times n}$ be a nonzero matrix in RREF. Suppose that columns $j_1, \ldots, j_r$ contain leading $1'$s, with $j_i < j_{i+1}$ for $i = 1, \ldots, r-1$. Then $\mathbf{r}_{j_i} = \mathbf{e}_i$, the list of pivot columns of R is $(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r})$, and* $\operatorname{rank}(R) = r$.

Proof. **1.** *For each $i = 1, \ldots, r$, there is k with $1 \le k \le r$ such that $\mathbf{r}_{j_i} = \mathbf{e}_k$.* Any column with a leading 1 is a standard basis vector $\mathbf{e}_k \in \mathbb{F}^m$ by Column Property 1. We need to show $1 \le k \le r$. Suppose that for some i, there is $k \ge r+1$ such that $\mathbf{r}_{j_i} = \mathbf{e}_k$. In particular, row $k \ge r+1$ has a nonzero entry.

However, rows 1 through r now contain at most $r-1$ leading 1's, and so at least one row between rows 1 and r has no leading 1. Then that row must have all zero entries (as otherwise it has a nonzero entry, thus a leading nonzero entry, thus a leading 1). But then a row between rows 1 and r has all zero entries, whereas a row between rows $r+1$ and m has a nonzero entry. This contradicts Row Property 1.

So, for each $i = 1, \ldots, r$, there exists k_i with $1 \le k_i \le r$ such that $\mathbf{r}_{j_i} = \mathbf{e}_{k_i}$. We want to show $k_i = i$. To be clear, saying $\mathbf{r}_{j_1} = \mathbf{e}_{k_i}$ means that the leading 1 in row k_i occurs in column j_i. Moreover, each row from row 1 to row r has a leading 1; otherwise, as demonstrated in the paragraph above, that row would have all zero entries, whereas a row below it would have nonzero entries.

2. *If $j < j_1$, then $\mathbf{r}_j = \mathbf{0}_m$.* Suppose instead that $\mathbf{r}_j \ne \mathbf{0}_m$ for some $j < j_1$. Then row i of $\mathbf{r}_j$ is nonzero for some i, and so row i has a leading nonzero entry, thus a leading 1, in some column $k \le j$. But column j_1 is the first column with a leading 1.

3. $\mathbf{r}_{j_1} = \mathbf{e}_1$. Suppose instead that $\mathbf{r}_{j_1} = \mathbf{e}_k$ with $k \ge 2$. Since R is a nonzero matrix, row 1 cannot have all zero entries, for then there would be a nonzero entry in row 2 or below, and that would contradict Row Property 1. So, row 1 has a nonzero entry, thus a leading nonzero entry, thus a leading 1. This entry cannot occur in column j with $j < j_1$, as column j_1 is the first column with a leading 1, and it cannot occur in column j_1, since row 1 of column j is zero. The leading 1 in row 1 therefore occurs in column $j > j_1$.

Now take $i_1 = 1$, $j_{i_1} = j$, $i_2 = k$, and $j_{i_2} = j_1$. Then $i_1 < i_2$ but $j_{i_2} < j_{i_1}$. The leading

1 of row i_1 is an entry in column j_{i_1}; the leading 1 of row i_2 is an entry in column j_{i_2}. This contradicts Column Property 2. We must therefore have $\mathbf{r}_{j_1} = \mathbf{e}_1$.

4. $\mathbf{r}_{j_i} = \mathbf{e}_i$ *for* $2 \le i \le r$. Suppose that this is not true and that i is the first column for which it fails. That is, $i \ge 2$, $\mathbf{r}_{j_i} \ne \mathbf{e}_i$, but

$$\mathbf{r}_{j_\ell} = \mathbf{e}_\ell, \ 1 \le \ell \le i-1. \tag{24.1}$$

We must have $\mathbf{r}_{j_i} = \mathbf{e}_k$ for some $1 \le k \le r$, so now $k \ne i$. Then the leading 1 of row k is in column j_i. Let the leading 1 of row i be in column j.

Case 1: $i < k$. Take $i_1 = i$, $i_2 = k$, $j_{i_1} = j$, and $j_{i_2} = j_i$ to find, from Column Property 2, that $j < j_i$. Since j is the index of a column with a leading 1, and $j < j_i$, it must be the case that $j = j_\ell$ for some $1 \le \ell \le i-1$. But then $\mathbf{r}_j = \mathbf{r}_{j_\ell} = \mathbf{e}_\ell$ by (24.1), and so the leading 1 in column j is in row $\ell < i$, not in row i.

Case 2: $k < i$. Then $\mathbf{r}_{j_k} = \mathbf{e}_k$ by (24.1), but also $\mathbf{r}_{j_i} = \mathbf{e}_k$ by the assumption above. This says that columns j_k and j_i both contain leading 1's for row k, which is impossible, because a leading 1 for a row can only appear in one column.

5. *If* $j_i < j < j_{i+1}$, *then* $\mathbf{r}_j \in \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_i})$. Since $\operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_i}) = \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_i)$, this is true if $\mathbf{r}_j$ has zero entries in rows $i+1$ through m. Suppose instead that $\mathbf{r}_j$ has a nonzero entry in one of those rows $i+1$ through m. Suppose that this happens in row $\ell \ge i+1$, so row ℓ has a nonzero entry in column j and so it has a leading nonzero entry in column j or before, thus a leading 1 in column j or before. As column j is not a pivot column, it does not have a leading nonzero entry in any of its rows, and so the leading 1 in row $\ell \ge i+1$ occurs in a column before column j. If this column is column $j_k < j$, then $\mathbf{r}_{j_k} = \mathbf{e}_\ell$. We now know that $\mathbf{r}_{j_k} = \mathbf{e}_k$, so $k = \ell \ge i+1$. Then $j_{i+1} = j_k < j < j_{i+1}$, a contradiction.

6. *If* $j > j_r$, *then* $\mathbf{r}_j \in \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r})$. First, the entries in rows $r+1$ through m of R are all zero. Otherwise, a row in rows $r+1$ through m would have a nonzero entry, thus a leading nonzero entry, thus a leading 1. But R already has leading 1's in rows 1 through r, and so R would have at least $r+1$ leading 1's, a contradiction. So, every column in R has zero entries in rows $r+1$ through m, so every column of R is in $\operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_r) = \operatorname{span}(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r})$.

7. $\operatorname{rank}(R) = r$. Since $\mathbf{r}_{j_i} = \mathbf{e}_i$, the list $(\mathbf{r}_{j_1}, \ldots, \mathbf{r}_{j_r})$ is independent. We have shown that any column $\mathbf{r}_j$ with $j \ne j_i$ for all i is in the span of this list, so this list spans $\mathbf{C}(R)$ and therefore is a basis for $\mathbf{C}(R)$. Thus $\operatorname{rank}(R) = r$. ■

We will use these properties of pivot columns to factor a matrix into its RREF—without doing any Gaussian elimination at all. Let $A \in \mathbb{F}^{m \times n}$ with $\operatorname{rank}(A) = r \ge 1$, and let $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$ be the pivot columns of A. The map

$$\mathcal{T} \colon \mathbb{F}^r \to \mathbf{C}(A) \colon (\alpha_1, \ldots, \alpha_r) \mapsto \sum_{i=1}^r \alpha_k \mathbf{a}_{j_i} \tag{24.2}$$

is an isomorphism. In particular,

$$\mathcal{T}^{-1}\mathbf{a}_{j_i} = \mathbf{e}_i. \tag{24.3}$$

24.2 Example. With

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

Example 23.11 tells us that $(\mathbf{a}_2, \mathbf{a}_4)$ is the list of pivot columns. Here $j_1 = 2$ and $j_2 = 4$. So, the isomorphism $\mathcal{T}$ from (24.2) is

$$\mathcal{T}\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

We have

$$\mathcal{T}^{-1}\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \mathcal{T}^{-1}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathcal{T}^{-1}\begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \qquad \mathcal{T}^{-1}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$\text{and} \quad \mathcal{T}^{-1}\begin{bmatrix} 7 \\ 14 \\ 8 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}. \tag{24.4}$$

In particular, $\mathbf{a}_1 = \mathbf{0}_3$ since $1 < 2 = j_2$, $\mathcal{T}^{-1}\mathbf{a}_2 = \mathbf{e}_1$ and $\mathcal{T}^{-1}\mathbf{a}_4 = \mathbf{e}_2$, as expected, and since the nonpivot column $\mathbf{a}_3$ satisfies $j_1 < 3 < j_2$, we have $\mathbf{a}_3 \in \operatorname{span}(\mathbf{a}_{j_1}) = \operatorname{span}(\mathbf{a}_2)$.

Returning to the more general case of (24.2), put

$$C := \begin{bmatrix} \mathbf{a}_{j_1} & \cdots & \mathbf{a}_{j_r} \end{bmatrix},$$

so

$$\mathcal{T}\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{bmatrix} = C\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{bmatrix} \qquad \text{and} \qquad \mathbf{a}_j = C(\mathcal{T}^{-1}\mathbf{a}_j).$$

In particular, C is the matrix representation of $\mathcal{T}$ with respect to the standard bases for $\mathbb{F}^r$ and $\mathbb{F}^m$: $C = [\mathcal{T}]$. Then with

$$R := \begin{bmatrix} \mathcal{T}^{-1}\mathbf{a}_1 & \cdots & \mathcal{T}^{-1}\mathbf{a}_n \end{bmatrix},$$

we have

$$A = CR.$$

We have (mostly) established the following factorization.

24.3 Theorem (CR-factorization). *Let $A \in \mathbb{F}^{m\times n}$ with $r := \operatorname{rank}(A) \geq 1$. Denote the list of pivot columns of A by $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$, where $1 \leq j_i < j_{i+1} \leq n$ for $i = 1, \ldots, r$, and let $C = \begin{bmatrix} \mathbf{a}_{j_1} & \cdots & \mathbf{a}_{j_r} \end{bmatrix} \in \mathbb{F}^{m\times r}$. There exists a unique $R \in \mathbb{F}^{r\times n}$ such that $A = CR$. In particular, R is in RREF and $\mathbf{r}_{j_i} = \mathbf{e}_i \in \mathbb{F}^r$.*

Proof. The work above proves existence of such a factorization.

1. For uniqueness of R, if $CR = C\widetilde{R}$, then $C(R - \widetilde{R})\mathbf{x} = \mathbf{0}_m$ for all $\mathbf{x} \in \mathbb{F}^m$. Hence $(R - \widetilde{R})\mathbf{x} \in \mathbf{N}(C) = \{\mathbf{0}_r\}$ for all $\mathbf{x} \in \mathbb{F}^m$, thus $R = \widetilde{R}$. That $\mathbf{N}(C) = \{\mathbf{0}_r\}$ follows because the columns of C are independent.

So, the matrix $R = \begin{bmatrix} \mathbf{r}_1 & \cdots \mathbf{r}_n \end{bmatrix} \in \mathbb{F}^{r\times n}$ is the one that we constructed above. In particular, from (24.3), we have $\mathbf{r}_{j_i} = \mathbf{e}_i$. (Alternatively, since $C\mathbf{r}_{j_i} = \mathbf{a}_{j_i} = \mathbf{c}_{j_i} = C\mathbf{e}_i$, the independence of the columns of C forces $\mathbf{r}_{j_i} = \mathbf{e}_i$).

2. We check that R is in RREF. First, no row of R has all zero entries because R has r rows and the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r \in \mathbb{F}^r$ are columns of r. So, R trivially satisfies Row Property 1.

3. For the other properties, we treat the first row as a special case.

Case 1: $j_1 = 1$. Then $\mathbf{r}_1 = \mathbf{e}_1 \in \mathbb{F}^r$, so the $(1,1)$-entry of R is 1. This proves Row Property 2 for row 1 and Column Property 1 for the leading 1 in row 1 in the case that $j_1 = 1$.

Case 2: $j_1 \geq 2$. Then the first $j_1 - 1$ columns of A are $\mathbf{0}_m$, so $C\mathbf{r}_j = \mathbf{0}_m$ for $j = 1, \ldots, j_1 - 1$. By independence, $\mathbf{r}_j = \mathbf{0}_r$ for $j = 1, \ldots, j_1 - 1$. The entries of row 1 in columns 1 through $j_1 - 1$ are therefore 0, but row 1 now has the entry of 1 in column j_1 since $\mathbf{r}_{j_1} = \mathbf{e}_1$. This proves Row Property 2 for row 1 and Column Property 1 for the first leading 1 in the case that $j_1 \geq 2$.

Finally, this also proves Column Property 2 for row 1 (i.e., for $i_1 = 1$ in Column Property 2), since the first nonzero column of R is $\mathbf{e}_1$ in column $j_1 \geq 1$; thus all of the entries in rows 2 and below are 0 for columns 1 through j_1. That is, the first nonzero entry in rows 2 and below must appear in columns $j_1 + 1$ or later.

4. Now let $i \geq 2$. Since $\mathbf{r}_{j_i} = \mathbf{e}_i$, we know that row i has a nonzero entry in column j_i. If we can show that the entries in row i of columns 1 through $j_i - 1$ are zero, then the first nonzero entry in row i will be the 1 from $\mathbf{e}_i$. Additionally, the other entries of the column in which that 1 is located (column j_i) will be zero, since that column is $\mathbf{e}_i$.

So, suppose that the first nonzero entry in row i occurs in some column j with $j < j_i$. We know that $\mathbf{a}_j$ is a linear combination of the columns of C weighted by the entries of column j of R; since the (i,j)-entry of R is nonzero, that linear combination *nontrivially* includes $\mathbf{a}_{j_i}$. But since $j < j_i$ and $i \geq 2$, we have $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$. It follows that $\mathbf{a}_{j_i} \in \operatorname{span}(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_{i-1}})$ as well, a contradiction.

We have therefore established that the first nonzero entry in row i is 1, that this entry occurs in column j_i, and that column j_i is $\mathbf{e}_i$. This proves Row Property 2 and Column Property 1.

5. Last, if $1 \le i_1 < i_2 \le m$, we now know that the leading 1 in row i_1 appears in column j_{i_1}, and the leading 1 in column i_2 appears in column j_{i_2}. By construction, $j_{i_1} < j_{i_2}$ since $i_1 < i_2$. This proves Column Property 2. ■

24.4 Example. Continuing Examples 23.11 and 24.2 with

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

we take

$$C = \begin{bmatrix} \mathbf{a}_{j_1} & \mathbf{a}_{j_2} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 2 \end{bmatrix}$$

and, per (24.4),

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

to find $A = CR$.

We are finally ready to prove the existence and uniqueness of the RREF from Theorem 23.6.

Proof (of Theorem 23.6—existence). Let $A \in \mathbb{F}^{m \times n}$ with $\operatorname{rank}(A) = r$ and write $A = CR$ as in Theorem 24.3. We consider some cases.

1. $r = m = n$. Here A is square and every column of A is a pivot column, so $C = A \in \mathbb{F}^{n \times n}$ and $R = I_n$. We reinterpret, somewhat elaborately, $A = CR$ as $A = CI_n$, so $C^{-1}A = I_n$. (In particular, A is invertible.)

2. $r = n < m$. Again, every column of A is a pivot column and $R = I_n$. Now extend the list $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ to a basis $(\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{v}_1, \ldots, \mathbf{v}_{m-n})$ for $\mathbb{F}^m$. Write $V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_{m-n} \end{bmatrix}$, so the block matrix $\begin{bmatrix} A & V \end{bmatrix} \in \mathbb{F}^{m \times m}$ is invertible. Then

$$A = AI_n = \begin{bmatrix} A & V \end{bmatrix} \begin{bmatrix} I_n \\ 0 \end{bmatrix}.$$

We conclude

$$EA = \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \qquad E^{-1} = \begin{bmatrix} A & V \end{bmatrix}.$$

3. $r = m < n$. Here $C \in \mathbb{F}^{m \times m}$ has independent columns and is invertible. Let $P \in \mathbb{F}^{n \times n}$ be a permutation matrix that interchanges columns i and j_i for $i = 1, \ldots, m$. That is, if $B \in \mathbb{F}^{m \times n}$, then column i of BP is column j_i of B for $i = 1, \ldots, m$. Then $R = \begin{bmatrix} I_m & F \end{bmatrix} P$ for some $F \in \mathbb{F}^{m \times (n-m)}$, so

$$A = C \begin{bmatrix} I_m & F \end{bmatrix} P,$$

and therefore

$$C^{-1}A = R = \begin{bmatrix} I_m & F \end{bmatrix} P.$$

4. $1 \le r \le \min\{m, n\}$. This is a combination of the previous two cases. First, extend the list $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$ to a basis $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}, \mathbf{v}_1, \ldots, \mathbf{v}_{m-r})$ for $\mathbb{F}^m$ and put $V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_{m-r} \end{bmatrix}$. Then $\begin{bmatrix} C & V \end{bmatrix}$ is invertible. Next, let $P \in \mathbb{F}^{n \times n}$ be a permutation matrix that interchanges columns i and j_i for $i = 1, \ldots, r$. That is, if $B \in \mathbb{F}^{m \times n}$, then column i of BP is column j_i of B for $i = 1, \ldots, r$. Then $R = \begin{bmatrix} I_r & F \end{bmatrix} P$ for some $F \in \mathbb{F}^{m \times (n-r)}$, and so

$$A = CR = C \begin{bmatrix} I_r & F \end{bmatrix} P = \begin{bmatrix} C & V \end{bmatrix} \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P,$$

thus

$$EA = \begin{bmatrix} I_r & F \\ 0 & 0 \end{bmatrix} P, \qquad E^{-1} = \begin{bmatrix} C & V \end{bmatrix}. \qquad \blacksquare$$

Our running example has been the fourth situation.

24.5 Example. This is the last installment begun in Examples 23.11 and 24.2. The pivot columns of

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

are the entries of the list $(\mathbf{a}_2, \mathbf{a}_4)$, and we have

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

Let $\mathbf{v} \in \mathbb{F}^3$ be any vector such that the list $(\mathbf{a}_2, \mathbf{a}_4, \mathbf{v})$ is a basis for $\mathbb{F}^3$. Then

$$A = \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_4 & \mathbf{v} \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 2 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

so

$$EA = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 2 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad E^{-1} = \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_4 & \mathbf{v} \end{bmatrix}.$$

Here is the last part of the proof of Theorem 23.6.

Proof (of Theorem 23.6—uniqueness). We just need to establish uniqueness. So, suppose that $EA = R_0$, where E is invertible and R_0 is in RREF. Write

$$R_0 = \begin{bmatrix} R \\ 0 \end{bmatrix},$$

where $R = \begin{bmatrix} \mathbf{r}_1 & \cdots & \mathbf{r}_n \end{bmatrix} \in \mathbb{F}^{r \times n}$ has nonzero rows. (This is the motivation for our use of the notation R_0 for the original RREF. If all of the rows of R_0 are nonzero, then just take $R = R_0$.) Then R is still in RREF, except that Row Property 1 is trivially satisfied.

Now let $E^{-1} = \begin{bmatrix} C & V \end{bmatrix}$, where $C \in \mathbb{F}^{m \times r}$. Then

$$A = E^{-1}R_0 = \begin{bmatrix} C & V \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} = CR.$$

Since R is in RREF, if we can show that C contains the pivot columns of A, then R must have the form from Theorem 24.3. This in turn dictates what R_0 is. Suppose that the pivot columns of R (equivalently, of R_0) are columns $j_1, \ldots, j_r$. Then $\mathbf{r}_{j_i} = \mathbf{e}_i \in \mathbb{F}^r$, and so $\mathbf{a}_{j_i} = C\mathbf{r}_{j_i} = C\mathbf{e}_i = \mathbf{c}_i$. Certainly, then, the columns of C are columns of A.

If we can show that the list of pivot columns of A is $(\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r})$, then we will be done. Suppose that this list is $(\mathbf{a}_{k_1}, \ldots, \mathbf{a}_{k_r})$. We want to show $j_i = k_i$ for each i. Throughout, we use the identity

$$E\mathbf{a}_j = \begin{bmatrix} \mathbf{r}_j \\ \mathbf{0}_{m-r} \end{bmatrix}. \tag{24.5}$$

If $k_1 = 1$, then since E is invertible,

$$\mathbf{0}_m \neq E\mathbf{a}_1 = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{0}_{m-r} \end{bmatrix},$$

thus $\mathbf{r}_1 \neq \mathbf{0}_r$, and so the first column of R is nonzero. Then $j_1 = 1$, so $j_1 = k_1$. If $k_1 > 1$, then $\mathbf{a}_j = \mathbf{0}_m$ for $j = 1, \ldots, k_1 - 1$ and $\mathbf{a}_{k_1} \neq \mathbf{0}_m$. Then (24.5) shows that $\mathbf{r}_j = \mathbf{0}_r$ for $j = 1, \ldots, k_1 - 1$ and $\mathbf{r}_{k_1} \neq \mathbf{0}_r$, so $j_1 = k_1$.

For $i \geq 2$, we know that $\mathbf{a}_{k_i}$ is the first column of A not in $\operatorname{span}(\mathbf{a}_{k_1}, \ldots, \mathbf{a}_{k_{i-1}})$. That is, $\mathbf{a}_j \in \operatorname{span}(\mathbf{a}_{k_1}, \ldots, \mathbf{a}_{k_{i-1}})$ for $j = 1, \ldots, k_i - 1$ and $\mathbf{a}_{k_i} \notin \operatorname{span}(\mathbf{a}_{k_1}, \ldots, \mathbf{a}_{k_{i-1}})$. Then $E\mathbf{a}_j \in \operatorname{span}(E\mathbf{a}_{k_1}, \ldots, E\mathbf{a}_{k_{i-1}})$ for $j = 1, \ldots, k_i - 1$ and, since E is invertible, $E\mathbf{a}_{k_i} \notin \operatorname{span}(E\mathbf{a}_{k_1}, \ldots, E\mathbf{a}_{k_{i-1}})$. Using (24.5), this reads

$$\begin{bmatrix} \mathbf{r}_j \\ \mathbf{0}_{m-r} \end{bmatrix} \in \operatorname{span}\left(\begin{bmatrix} \mathbf{r}_{k_1} \\ \mathbf{0}_{m-r} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{r}_{k_i-1} \\ \mathbf{0}_{m-r} \end{bmatrix} \right), \ j = 1, \ldots, k_i - 1$$

and

$$\begin{bmatrix} \mathbf{r}_{k_i} \\ \mathbf{0}_{m-r} \end{bmatrix} \notin \operatorname{span}\left(\begin{bmatrix} \mathbf{r}_{k_1} \\ \mathbf{0}_{m-r} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{r}_{k_i-1} \\ \mathbf{0}_{m-r} \end{bmatrix} \right)$$

Thus the list $(\mathbf{r}_{k_1}, \ldots, \mathbf{r}_{k_r})$ is also the list of pivot columns of R, and so by the unique construction of the pivot columns, $(j_1, \ldots, j_r) = (k_1, \ldots, k_r)$. ■

24.6 Problem (!). Why does Theorem 23.6 not have a uniqueness claim about the matrix E?

Day 25: Wednesday, October 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."

Basis (in an arbitrary vector space), inner product (be able to check the axioms), inner product space

We have devoted a lot of effort to understanding finite-dimensional vector spaces via the tool of bases. Recall that fundamentally a basis is a unique coordinate system for a vector space: each vector in the space can be expressed as a (by definition *finite*) linear combination of vectors in the basis in a unique way. We made this more precise with quantifiers and then we decoupled the notion of basis into the more easily checked conditions of linear independence and span.

What happens in an infinite-dimensional vector space? Suppose that $\mathcal{V}$ is a vector space such that $\mathcal{V} \neq \operatorname{span}(v_1, \ldots, v_n)$ for any list $(v_1, \ldots, v_n)$ in $\mathcal{V}$? Does $\mathcal{V}$ still have a coordinate system?

Yes. To express this, we need some new terminology that repackages familiar ideas. In finite dimensions, a basis is a list of some length n in the vector space. This is just a function from $\{1, \ldots, n\}$ to the space. We will now think about lists whose lengths may not be finite (or, indeed, easily defined).

25.1 Definition. *Let J and X be sets. A* LIST IN X INDEXED BY J *is a function in X^J. If $f \in X^J$ with $f(j) = x_j$, then we write $f = (x_j)_{j \in J}$. That is,*

$$(x_j)_{j \in J} = \{(j, x_j) \mid j \in J\}.$$

All that this notion of list is doing is making more explicit the outputs of $f \in X^J$. Now let $\mathcal{V}$ be a vector space and let $(w_j)_{j \in J}$ be a list in $\mathcal{V}$. We want to assign meaning to the expression

$$\sum_{j \in J} w_j,$$

which intuitively should mean "add up all of the w_j in the list." This is fine if J is a finite set, dicey if $J = \mathbb{N}$ (think series), and confounding if, say, $J = \mathbb{R}$. However, there is a natural meaning for this expression if w_j is the zero vector "most of the time."

25.2 Definition. *Let $\mathcal{V}$ be a vector space and $(w_j)_{j \in J}$ be a list in $\mathcal{V}$ such that $w_j = 0_{\mathcal{V}}$ for all but finitely many $j \in J$. If $w_j = 0_{\mathcal{V}}$ for all $j \in J$, we define*

$$\sum_{j \in J} w_j := 0_{\mathcal{V}}.$$

Otherwise, suppose that $\{j \in J \mid w_j \neq 0_{\mathcal{V}}\} = \{j_k\}_{k=1}^n$ with $j_{k_1} \neq j_{k_2}$ for $k_1 \neq k_2$ and $n \geq 1$.

Then
$$\sum_{j \in J} w_j := \sum_{k=1}^{n} w_{j_k}.$$

Now we can really talk about bases.

25.3 Definition. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. A* **BASIS** *for $\mathcal{V}$ is a list $(v_j)_{j \in J}$ in $\mathcal{V}$ such that for each $v \in \mathcal{V}$, there is a unique list $(\alpha_j)_{j \in J}$ in $\mathbb{F}$ such that $\alpha_j = 0$ for all but finitely many $j \in J$ and $v = \sum_{j \in J} \alpha_j v_j$.*

The uniqueness of the list $(\alpha_j)_{j \in J}$ encodes the idea of uniqueness of the coordinates of v with respect to this basis; that $\alpha_j = 0$ for all but finitely many j encodes the idea that each vector is really a *finite* linear combination of vectors in the basis. (To be fair, we have only talked about finite linear combinations, never infinite series.) This is really the independence of the list $(v_j)_{j \in J}$.

25.4 Definition. *A list $(v_j)_{j \in J}$ in the vector space $\mathcal{V}$ over $\mathbb{F}$ is* **(LINEARLY) INDEPENDENT** *if whenever $(\alpha_j)_{j \in J}$ is a list in $\mathbb{F}$ such that $\alpha_j = 0$ for all but finitely many $j \in J$ and $\sum_{j \in J} \alpha_j v_j = 0_{\mathcal{V}}$, then $\alpha_j = 0$ for all $j \in J$. If a list is not independent, then it is* **(LINEARLY) DEPENDENT**.

25.5 Problem ($\star$). Define the list $(e_j)_{j \in \mathbb{N}}$ in $\mathbb{F}^\infty$ by setting $e_j(k) = 1$ for $j = k$ and 0 for $j \neq k$. Prove that this list is independent but not a basis for $\mathbb{F}^\infty$.

It can be shown that every vector space has a basis in the sense of Definition 25.3; we did that already under the assumption that the space was spanned by a finite list. Such a proof typically uses Zorn's lemma. Moreover, once the existence of a basis is established, it can also be shown that any two bases have the same length. That is, if $(v_j)_{j \in J}$ and $(w_k)_{k \in K}$ are bases for the same space, then J and K have the same cardinality, i.e., there is a bijection between J and K. Naturally, we want to define the dimension of the space as the cardinality of J (or K).

Bases in infinite-dimensional spaces certainly have their uses, but they are not quite as useful as bases in finite dimensions. For one reason, some of the most interesting infinite-dimensional spaces are so large that their dimension is at least the cardinality of $\mathbb{R}$. For another, infinite-dimensional spaces that arise in applications typically have additional structure, and that structure leads to representations that are often more useful than a plain old basis. We now turn to such a structure: geometry and the inner product.

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 boil down to 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, which is an instrument for extracting data about vectors and for representing vectors. 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.

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.

25.6 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 $1 \le k \le n$. Then

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

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.

25.7 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

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

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.

25.8 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.

25.9 Example. **(i)** The map from Example 25.6 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

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{q}} := \sum_{j=1}^{n} (q_j v_j)\overline{w_j} \tag{25.1}$$

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 25.7 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. (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.

25.10 Problem (!). Give an example of $\mathbf{q} \in \mathbb{R}^n$ such that the weighted inner product $\langle \cdot, \cdot \rangle_{\mathbf{q}}$ from (25.1) is not an inner product on $\mathbb{F}^n$. Discuss exactly what goes wrong.

25.11 Problem (!). Why does defining

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

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

25.12 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.

25.13 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,

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

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.")

25.14 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

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

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.

25.15 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.

25.16 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 19.7 that it suffices to know the values of $\mathcal{T} v_j$; use Theorem 25.15 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).

Day 26: Friday, October 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."

Orthogonal list (N), orthonormal list (N), norm induced by the inner product, orthonormal basis

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

26.1 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

$$\mathbf{v} = \sum_{j=1}^{n} v_j \mathbf{e}_j = \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{e}_j \rangle \mathbf{e}_j. \tag{26.1}$$

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:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{j=1}^{n} v_j \overline{w_j} = \sum_{k=1}^{n} \langle \mathbf{v}, \mathbf{e}_j \rangle \overline{\langle \mathbf{w}, \mathbf{e}_j \rangle}. \tag{26.2}$$

(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":

$$\langle \mathbf{e}_j, \mathbf{e}_k \rangle = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases} \tag{26.3}$$

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

26.2 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.

26.3 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 $u = \cos(x)$.

26.4 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).

26.5 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.

26.6 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*

$$v = \sum_{j=1}^{n} \frac{\langle v, v_j \rangle}{\langle v_j, v_j \rangle} v_j. \tag{26.4}$$

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:

$$\langle v, v_k \rangle = \left\langle \sum_{j=1}^{n} \alpha_j v_j, v_k \right\rangle = \sum_{j=1}^{n} \langle \alpha_j v_j, v_k \rangle = \sum_{j=1}^{n} \alpha_j \langle v_j, v_k \rangle = \alpha_k \langle v_k, v_k \rangle . \tag{26.5}$$

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.

26.7 Remark. *The calculation in* (26.5) *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.

26.8 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 $\alpha_j = \langle v_j, 0_{\mathcal{V}} \rangle / \langle v_j, v_j \rangle = 0$. ■

26.9 Problem (!). Give an example of a linearly independent list in an inner product space that is not an orthogonal list.

Orthogonality by itself does not account for the excellent behavior of the standard basis vectors in (26.3). In particular, orthogonality says nothing about the behavior of the inner products $\langle v, v \rangle$ when both vectors are the same.

26.10 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 $\|u\| = 1$ 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.

We can adapt the calculation in (26.5) to generalize the expansion (26.1). We can paraphrase this via the slogan "Inner products extract coefficients from linear combinations."

26.11 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*

$$v = \sum_{j=1}^{n} \langle v, u_j \rangle \, u_j. \tag{26.6}$$

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 (26.2), mostly as an exercise in manipulating sums and indices.

26.12 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*

$$\langle v, w \rangle = \sum_{j=1}^{n} \langle v, u_j \rangle \overline{\langle w, u_j \rangle}. \tag{26.7}$$

Proof. We use Theorem 26.6 to expand

$$\langle v, w \rangle = \left\langle \sum_{j=1}^{n} \langle v, u_j \rangle, \sum_{j=1}^{n} \langle w, u_j \rangle \right\rangle. \tag{26.8}$$

Here we are continuing with the good grammar of Remark 26.7 by not overworking the indices of summation and using k and j separately. Linearity of the inner product in the first slot gives

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

$$= \sum_{j=1}^{n} \langle v, u_j \rangle \left\langle u_j, \sum_{k=1}^{n} \langle w, u_k \rangle u_k \right\rangle. \quad (26.9)$$

Then we use antilinearity in the second slot to compute, for each j,

$$\left\langle u_j, \sum_{k=1}^{n} \langle w, u_k \rangle u_k \right\rangle = \sum_{k=1}^{n} \langle u_j, \langle w, u_k \rangle u_k \rangle = \sum_{k=1}^{n} \overline{\langle w, u_k \rangle} \langle u_j, u_k \rangle = \langle w, u_j \rangle, \quad (26.10)$$

where the last equality is orthonormality with $\langle u_j, u_j \rangle = 1$ and $\langle u_j, u_k \rangle = 0$ for $j \neq k$. Combining (26.8), (26.9), and (26.10) yields the desired identity for $\langle v, w \rangle$. ■

The identities (26.4), (26.6), and (26.7) 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.

26.13 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*

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

26.14 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.

26.15 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* $v = 0_{\mathcal{V}}$.

26.16 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 in Example 20.3.) It would be nice if we could guarantee that any finite-dimensional inner product space has a basis consisting of orthonormal vectors.

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

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

Day 27: Monday, October 20.

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

$$\operatorname{span}(v_1, \ldots, v_j) = \operatorname{span}(u_1, \ldots, u_j). \tag{27.1}$$

This is what we want for $j = n$, but our procedure gives (27.1) 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 want (u_1) to be orthonormal, so $\|u_1\| = 1$, and $\operatorname{span}(u_1) = \operatorname{span}(v_1)$. Then we want $u_1 = \alpha_1 v_1$ for some $\alpha_1 \in \mathbb{F}$. How should we choose α_1? All that we know is

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

Since (v_1) is independent, $v_1 \neq 0_{\mathcal{V}}$, so $\|v_1\| \neq 0$. Then we may solve for

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

This still leaves many (if $\mathbb{F} = \mathbb{C}$) or some (if $\mathbb{F} = \mathbb{R}$) options for α_1; the simplest is to take

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

2. $n = 2$. We want (u_1, u_2) to be orthonormal and $\operatorname{span}(v_1, v_2) = \operatorname{span}(u_1, u_2)$. Our success with $n = 1$ suggests taking $u_1 = v_1 / \|v_1\|$. Since we want $\operatorname{span}(v_1, v_2) = \operatorname{span}(u_1, u_2)$, and since we know $\operatorname{span}(v_1) = \operatorname{span}(u_1)$, it should suffice to have $u_2 \in \operatorname{span}(u_1, v_1)$. (Why should it?) So, we must be able to write $u_2 = \alpha_1 u_1 + \alpha_2 v_2$, and so to have $\langle u_2, u_1 \rangle = 0$, we need

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

since $\langle u_1, u_1 \rangle = \|u_1\|^2 = 1$. Then

$$\alpha_1 = -\alpha_2 \langle v_2, u_1 \rangle$$

and so

$$u_2 = -\alpha_2 \langle v_2, u_1 \rangle u_1 + \alpha_2 v_2 = \alpha_2 (v_2 - \langle v_2, u_1 \rangle u_1) \tag{27.2}$$

Put

$$w_2 := v_2 - \langle v_2, u_1 \rangle u_1,$$

so we want $u_2 = \alpha_2 w_2$. Since (v_1, v_2) is independent and u_1 is a nonzero scalar multiple of v_1, the list (u_1, v_2) is also independent (think about it...), and therefore $w_2 \neq 0_{\mathcal{V}}$. To have $\|u_2\| = 1$, we need $|\alpha_2| \|w_2\| = 1$; this suggests taking $\alpha_2 = 1 / \|w_2\|$, which is permissible.

To summarize, we put

$$\begin{cases} u_1 := v_1 / \|v_1\| \\ w_2 := v_2 - (\langle v_2, u_1 \rangle u_1 \\ u_2 := w_2 / \|w_2\| \end{cases} \tag{27.3}$$

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 $\operatorname{span}(u_1) = \operatorname{span}(u_1)$.

3. $n = 3$. We want (u_1, u_2, u_3) to be orthonormal and $\operatorname{span}(v_1, v_2, v_3) = \operatorname{span}(u_1, u_2, u_3)$. Based on our prior success, we might use (27.3) to define u_1 and u_2. Then (check this) we would have $\operatorname{span}(v_1, v_2, v_3) = \operatorname{span}(u_1, u_2, v_3)$. We therefore want $u_3 \in \operatorname{span}(u_1, u_2, v_3)$, 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

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

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 $v_3 \in \operatorname{span}(u_1, u_2) = \operatorname{span}(v_1, v_2)$, 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 $\operatorname{span}(v_1, v_2, v_3) = \operatorname{span}(u_1, u_2, u_3)$, and the added bonus that $\operatorname{span}(v_1) = \operatorname{span}(u_1)$ and $\operatorname{span}(v_1, v_2) = \operatorname{span}(u_1, u_2)$.

Here is how this works in general.

27.1 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*

$$u_j := \begin{cases} \dfrac{v_1}{\|v_1\|}, & j = 1 \\[2ex] \dfrac{w_j}{\|w_j\|}, & j \geq 2, \end{cases} \qquad w_j := v_j - \sum_{k=1}^{j-1} \langle v_j, u_k \rangle u_k, \ j \geq 2. \tag{27.4}$$

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

$$\operatorname{span}(v_1, \ldots, v_j) = \operatorname{span}(u_1, \ldots, u_j), \ j = 1, \ldots, n. \tag{27.5}$$

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 $\operatorname{span}(v_1) = \operatorname{span}(u_1)$, 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 (27.4). Then $(u_1, \ldots, u_n)$ is orthonormal and preserves spans in the sense of (27.5).

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 26.5).

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

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

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.

27.2 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 (26.6). In that sense, orthonormal bases are closely related to the coordinate functionals from the corresponding dual basis (Example 20.3).

27.3 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 $(\langle \cdot, u_1 \rangle, \ldots, \langle \cdot, u_n \rangle)$.

Day 28: Wednesday, October 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."

Projection onto a vector in a inner product space, Cauchy–Schwarz inequality, Pythagorean identity

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.

28.1 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*

$$\mathcal{P}_w : \mathcal{V} \to \mathcal{V} : v \mapsto \frac{\langle v, w \rangle}{\|w\|^2} w. \tag{28.1}$$

If we revisit (27.2) in the Gram–Schmidt 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 $\mathcal{P}_{v_1} v_1$ is a scalar multiple of v_1, this also says that $v_2 - \mathcal{P}_{v_1} v_2$ and $\mathcal{P}_{v_1} v_1$ are orthogonal.

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

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

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

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

28.3 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 $aw = \mathcal{P}_w v$.

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

28.4 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

$$\|v+w\| \leq \|v\| + \|w\|.$$

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

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

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

$$\operatorname{Re}[\langle v, w\rangle] \leq \|v\|\,\|w\|. \tag{28.2}$$

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

So, do we have this peculiar inequality (28.2)? Yes. In fact, we get something a bit stronger.

28.6 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 $2ab \leq a^2+b^2$, valid for all a, $b \in \mathbb{R}$ (since $0 \leq (a-b)^2$), to find

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

Factor the right side as

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

28.7 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

$$\|v+w\|^2 = \|v\|^2 + \|w\|^2. \tag{28.3}$$

[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 (28.3) 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: $\sqrt{a^2+b^2} \neq |a| + |b|$ *in general.*]

(iv) Generalize (28.3) to the case where $(v_1, \ldots, v_n)$ is an orthogonal list:

$$\left\|\sum_{j=1}^{n} v_j\right\|^2 = \sum_{j=1}^{n} \|v_j\|^2. \tag{28.4}$$

28.8 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 (26.7) and then with (28.4).

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 (28.2), and much more.

28.9 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 $0 \le \|v\| \|w\|$, and that is the inequality. If v and w are not orthogonal, then experience teaches us that we can separate them into "orthogonal components."

Suppose $w \neq 0_{\mathcal{V}}$. (The case $w = 0_{\mathcal{V}}$ is handled by the previous remarks about $\langle v, w\rangle = 0$.) With $\mathcal{P}_w$ from (28.1), we have

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

and

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

by Problem 28.2. Then the Pythagorean identity (28.3) implies

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

The vector $\mathcal{P}_w v$ contains factors that resemble what we want out of the Cauchy–Schwarz inequality, whereas $v - \mathcal{P}_w v$ may be more opaque. Fortunately, the latter is irrelevant here: since $\|v - P_w v\|^2 \ge 0$, we have

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

Rewritten, this says

$$\left\| \frac{\langle v, w \rangle}{\|w\|^2} w \right\| \le \|v\| ,$$

and the left side of this inequality is

$$\left\| \frac{\langle v, w \rangle}{\|w\|^2} w \right\| = \frac{|\langle v, w \rangle|}{\|w\|^2} \|w\| = \frac{|\langle v, w \rangle|}{\|w\|}.$$

Thus

$$\frac{|\langle v, w \rangle|}{\|w\|} \le \|v\| ,$$

and this turns into the Cauchy–Schwarz inequality. ■

28.10 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: $\mathcal{P}_w v = v$ *if and only if* $v \in \operatorname{span}(w)$.]

28.11 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

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

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

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

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

28.12 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?

28.13 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 28.5 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.

28.14 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. Suppose that (d_k) is a sequence and there is $C > 0$ such that $\sum_{k=1}^{n}|d_k| \le C$ for all $n \ge 1$. Then $\sum_{k=1}^{\infty}|d_k|$ converges. 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 25.6.

28.15 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.

We have previously seen (Problem 25.5) that the list $(e_j)_{j \in \mathbb{N}}$, where $e_j(k) = 1$ for $j = k$ and 0 for $j \neq k$, is not a basis for ℓ^2. Nonetheless, it is an orthonormal list in the sense that $\langle e_j, e_k \rangle = 1$ for $j = k$ and 0 for $j \neq k$. (We have never talked about orthonormality of any kind of list other than a finite list, but what else could orthonormality mean here?) Moreover, given $f \in \ell^2$, we have $f(j) = \langle f, e_j \rangle$, and if we put $s_n := \sum_{j=1}^n \langle f, e_j \rangle e_j$, then we have $s_n(k) = f(k)$ for $k = 1, \ldots, n$. Consequently,

$$\|f - s_n\|^2 = \sum_{k=n+1}^{\infty} |f(k)|^2 \to 0 \tag{28.5}$$

as $n \to \infty$. Our calculus instincts perhaps make us want to say $f = \lim_{n \to \infty} s_n$.

We briefly formalize this notion of convergence.

28.16 Definition. *Let (v_n) be a sequence in the inner product space $\mathcal{V}$. We say that (v_n)* **CONVERGES** *to $v \in \mathcal{V}$ and write $\lim_{n \to \infty} v_n = v$ or $v_n \to v$ if*

$$\lim_{n \to \infty} \|v_n - v\| = 0, \tag{28.6}$$

with the limit (28.6) being the usual limit in $\mathbb{R}$.

All of the usual properties of limits hold in inner product spaces, starting with uniqueness.

28.17 Theorem. *Let (v_n) be a sequence in the inner product space $\mathcal{V}$ and $v, w \in \mathcal{V}$. If both*

$$\lim_{n \to \infty} \|v_n - v\| = 0 \quad \text{and} \quad \lim_{n \to \infty} \|v_n - w\| = 0,$$

then $v = w$.

This follows from the triangle inequality and the squeeze theorem:

$$\|v - w\| = \|v - v_n + v_n - w\| \leq \|v - v_n\| + \|v_n - w\| \to 0.$$

Limit arithmetic also works as we expect.

28.18 Theorem. *Let (v_n) and (w_n) be sequences in the inner product space $\mathcal{V}$ with $v_n \to v$ and $w_n \to w$. Then the following hold.*

(i) *(v_n) is bounded in the sense that there is $C > 0$ such that $\|v_n\| \leq C$ for all n.*

(ii) *$v_n + w_n \to v + w$.*

(iii) *If $\alpha \in \mathbb{F}$, then $\alpha v_n \to \alpha v$.*

(iv) $\langle v_n, w_n \rangle \to \langle v, w \rangle$.

Proof. We prove only the fourth part, as the others are direct analogues of properties of real sequences. Rewrite

$$\langle v_n, w_n \rangle - \langle v, w \rangle = \langle v_n, w_n \rangle - \langle v, w_n \rangle + \langle v, w_n \rangle - \langle v, w \rangle = \langle v_n - v, w_n \rangle + \langle v, w_n - w \rangle .$$

Use the Cauchy–Schwarz inequality to estimate

$$| \langle v, w_n - w \rangle | \le \|v\| \, \|w_n - w\| \to 0.$$

Use Cauchy–Schwarz again and this time the boundedness of (w_n), with $\|w_n\| \le C$ for all n, to estimate

$$| \langle v_n - v, w_n \rangle | \le \|v_n - v\| \, \|w_n\| \le C \, \|v_n - v\| \to 0.$$

Thus

$$| \langle v_n, w_n \rangle - \langle v, w \rangle | \le | \langle v_n - v, w_n \rangle | + | \langle v, w_n - w \rangle | \to 0. \qquad \blacksquare$$

Sequences allow us to define series.

28.19 Definition. *Let (v_j) be a sequence in the inner product space $\mathcal{V}$.*

(i) *We put*

$$\sum_{j=1}^{\infty} v_j := \left(\sum_{j=1}^{n} v_j \right).$$

That is, the symbol $\sum_{j=1}^{\infty} v_j$ denotes the sequence of partial sums $(\sum_{j=1}^{n} v_j)$, and we call the sequence $\sum_{j=1}^{\infty} v_j$ the **SERIES** *whose nth partial sum is $\sum_{j=1}^{n} v_j$.*

(ii) *If $\lim_{n\to\infty} \sum_{j=1}^{n} v_j$ exists, then we call this limit the* **SUM** *of the series $\sum_{j=1}^{\infty} v_j$, and we also write*

$$\sum_{j=1}^{\infty} v_j := \lim_{n\to\infty} \sum_{j=1}^{n} v_j.$$

In this case the symbol $\sum_{j=1}^{\infty} v_j$ plays the dual role of denoting both a sequence in $\mathcal{V}$ and an element in $\mathcal{V}$.

28.20 Theorem. *Let $\mathcal{V}$ be an inner product space, let (v_j) be a sequence in $\mathcal{V}$, and suppose that the series $\sum_{j=1}^{\infty} v_j$ converges in $\mathcal{V}$. Then for any $w \in \mathcal{V}$, the series $\left\langle \sum_{j=1}^{\infty} v_j, w \right\rangle$ converges in $\mathbb{F}$, and*

$$\left\langle \sum_{j=1}^{\infty} v_j, w \right\rangle = \sum_{j=1}^{\infty} \langle v_j, w \rangle . \tag{28.7}$$

Proof. Abbreviate $s := \sum_{j=1}^{\infty} v_j$, so

$$\lim_{n\to\infty} \left\| \sum_{j=1}^{n} v_j - s \right\| = 0.$$

The desired identity (28.7) is equivalent to showing

$$\lim_{n\to\infty}\sum_{j=1}^{n}\langle v_j, w\rangle = \langle s, w\rangle\,.$$

This is straightforward: algebraic properties of the inner product and the Cauchy–Schwarz inequality give

$$\left|\sum_{j=1}^{n}\langle v_j, w\rangle - \langle s, w\rangle\right| = \left|\left\langle \sum_{j=1}^{n} v_j - s, w\right\rangle\right| \le \left\|\sum_{j-1}^{n} v_j - s\right\| \|w\| \to 0. \qquad \blacksquare$$

Day 29: Friday, October 24.

Vocabulary from today

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

Orthonormal basis for an infinite-dimensional vector space (the infinite-dimensional part is important!)

Now we can talk about the situation in ℓ^2 with that special list $(e_j)_{j\in\mathbb{N}}$ more precisely, and also more generally.

29.1 Definition. *Let $\mathcal{V}$ be an inner product space. A list $(u_j)_{j\in J}$ is* **ORTHONORMAL** *if*

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

29.2 Problem (!). Prove that any inner product space with an orthonormal list $(u_j)_{j\in\mathbb{N}}$ is infinite-dimensional. [Hint: *for each integer $n \ge 1$, find a linearly independent list with length n.*]

Suppose that $\mathcal{V}$ is an inner product space and $(u_j)_{j\in\mathbb{N}}$ is an orthonormal list in $\mathcal{V}$. If $v \in \mathcal{V}$ and $(\alpha_j)_{j\in\mathbb{N}}$ is a list in $\mathbb{F}$ such that $v = \sum_{j=1}^{\infty}\alpha_j u_j$, then we are really saying that

$$\lim_{n\to\infty}\left\|v - \sum_{j=1}^{n}\alpha_j v_j\right\| = 0.$$

For a given k, Theorem 28.20 allows

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

This is exactly what we expect from the tamer situation (26.6) with a *finite* orthonormal list.

29.3 Definition. *Let $\mathcal{V}$ be an infinite-dimensional inner product space and $(u_j)_{j\in\mathbb{N}}$ be an orthonormal list in $\mathcal{V}$. The* **Fourier series** *of $v \in \mathcal{V}$ with respect to $(u_j)_{j\in\mathbb{N}}$ is the series*

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

This series may or may not converge; in the latter case, we just think of it as the sequence of partial sums $(\sum_{j=1}^{n} \langle v, u_j \rangle \, u_j)$. With the notion of Fourier series, we can adapt our prior version of an orthonormal basis to the infinite-dimensional case.

29.4 Definition. *Let $\mathcal{V}$ be an infinite-dimensional inner product space. An orthonormal list $(u_j)_{j\in\mathbb{N}}$ is an* **orthonormal basis** *for $\mathcal{V}$ if*

$$v = \sum_{j=1}^{\infty} \langle v, u_j \rangle \, u_j$$

for all $v \in \mathcal{V}$, i.e., if every $v \in \mathcal{V}$ is the sum of its Fourier series with respect to this list. If $\mathcal{V}$ has an orthonormal basis, we say that $\mathcal{V}$ is **separable**.

Our use of the term orthonormal basis here is quite different from Definition 26.17 in finite dimensions. In the infinite-dimensional setting, we do not require that a vector be written as a *finite* linear combination of entries from the orthonormal basis—which surely goes against all of our instincts regarding bases up to now. To be clear, and annoying, the list from Definition 29.4 is probably not a spanning list! (It is a linearly independent list.) Rather, the power of the orthonormal basis in infinite dimensions is that it allows an arbitrarily close *approximation* of a vector by finite linear combinations of entries from the list. For most applications, the orthonormal basis (if it exists) turns out to be much more preferable than a classical basis (the latter is sometimes called a **Hamel basis** for distinction).

29.5 Example. **(i)** The inner product space ℓ^2 is separable because the list $(e_j)_{j\in\mathbb{N}}$ with $e_j(k) = 1$ for $j = k$ and 0 otherwise is an orthonormal basis. This was the calculation (28.5).

(ii) The notion of Fourier series is probably most naturally associated with trigonometric functions. For notational simplicity, let $\mathcal{V} = \{f \in \mathcal{C}([-\pi,\pi]) \mid f \text{ is odd}\}$ with the L^2-inner product. For integers $j \geq 1$, put $u_j(x) := \sin(jx)/\left\|\sin(j\cdot)\right\|$. It can be shown that $(u_j)_{j\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{V}$. The orthonormality is not the hard part; orthogonality of the u_j is one of the best (and perhaps only?) reasons to study trigonometric integrals in calculus.

However, showing that if $f\colon [-\pi, \pi] \to \mathbb{R}$ is continuous and odd, then

$$\lim_{n\to\infty} \left\| f - \sum_{j=1}^{n} \langle f, u_j \rangle u_j \right\| = 0$$

is quite the feat of hard analysis. Moreover, this "L^2-convergence" does not address the ticklish issue of *pointwise* convergence of the Fourier series back to f, i.e., whether the series $\sum_{j=0}^{\infty} \langle f, u_j \rangle u_j(x)$ converges and equals $f(x)$ for all continuous odd functions on $[-\pi, \pi]$. (It does not.)

By the way, we can remove the restriction of oddness by including cosines in the list, or we could allow complex-valued functions and use complex exponentials. The notation in either case is just distracting enough that we do not address it here.

The definition of orthonormal list here did not require that the indexing set J be finite or countable (i.e., there need not be a bijection between J and $\{1, \ldots, n\}$ for some n or between J and $\mathbb{N}$). Consequently, we could talk about an orthonormal basis indexed by an *arbitrary* set J . . . if we could give meaning to a sum of the form $\sum_{j\in J} v_j$ where J is not finite or countable. This is possible! It involves a topological generalization of sequences called *nets*. Limits of nets, by the way, give the right interpretation of the Riemann integral as a "limit" of Riemann sums (a notion of limit that, in calculus, looks nothing like the limit definition of the derivative). However, because most interesting and useful inner product spaces turn out to be separable, summing over arbitrary indexing sets is not high among our priorities.

Day 30: Monday, October 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."

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 $\mathcal{U}_1 := \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $\mathcal{U}_2 := \text{span}(\mathbf{e}_3)$. 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*

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

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 ($\star$). 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 $\mathcal{U} \cap \mathcal{U}^{\perp} = \{0_{\mathcal{V}}\}$.

(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 $\mathcal{U} = \operatorname{span}(\mathbf{e}_1)$. Draw a picture illustrating what this result is saying about $\mathcal{U}^{\perp} = \operatorname{span}(\mathbf{e}_2)$.

(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 28.11, for $\alpha \in \mathbb{F}$, put $p(\alpha) := \|v - \alpha u\|^2$. Conclude that

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

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 $\mathrm{Re}[\langle v, u\rangle] = 0$. Then take $\alpha = i\beta$ with $\beta \in \mathbb{R}$, consider the limit as $\beta \to 0^{\pm}$, and conclude $\mathrm{Im}[\langle v, u\rangle] = 0$, 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 28.14 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 $e_j(k) = 1$ for $j = k$ and 0 for $j \ne k$, 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 $\mathcal{U} \ne \ell^2$.

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, October 29.

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, orthogonal projection

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

31.1 Definition. *Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Then $\mathcal{V}$ is the* ORTHOGONAL DIRECT SUM *of $\mathcal{U}$ and $\mathcal{U}^\perp$, abbreviated $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$, if for all $v \in \mathcal{V}$, there exist (necessarily) unique $u \in \mathcal{U}$ and $u^\perp \in \mathcal{U}^\perp$ such that $v = u + u^\perp$.*

31.2 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.

31.3 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 $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$. [Hint: *what do you know about $u_0 - v$?*]

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, $\beta_k \in \mathbb{F}$.*]

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 $v = \mathcal{P}_\mathcal{U} v$.*

We call $\mathcal{P}_\mathcal{U}$ 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 28.1.

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 $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$.

31.9 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. ■

31.10 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 31.9.*]

Day 32: Friday, October 31.

You took Exam 2.

Day 33: Monday, November 3.

Vocabulary from today

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

Least squares solution to $\mathcal{T}v = w$, bounded linear functional on an inner product space

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 $\mathcal{U}$.

33.1 Problem (!). Let $\mathcal{V} = \mathbb{R}^2$ with the usual Euclidean inner product. Recall that $\mathcal{P}_{\operatorname{span}(w)} = \mathcal{P}_w$ for any $w \in \mathbb{R}^2$, with $\mathcal{P}_w$ defined in 28.1. 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 $\text{span}(w)$. How does this resemble your picture from Problem 28.3?

33.2 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*

$$\|v - \mathcal{P}_{\mathcal{U}} v\| \le \|v - u\| \tag{33.1}$$

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

$$\|v-u\|^2 = \|(v - \mathcal{P}_{\mathcal{U}} v) + (\mathcal{P}_{\mathcal{U}} v - u)\|^2 = \|v - \mathcal{P}_{\mathcal{U}} v\|^2 + \|\mathcal{P}_{\mathcal{U}} v - u\|^2 \ge \|v - \mathcal{P}_{\mathcal{U}} v\|^2. \tag{33.2}$$

This proves (33.1).

Now we prove the if and only if statement. If $u = \mathcal{P}_{\mathcal{U}} v$, then certainly the inequality is an equality. If $u \ne \mathcal{P}_{\mathcal{U}} v$, then $\|\mathcal{P}_{\mathcal{U}} v - u\|^2 > 0$, so the inequality in (33.2) is strict. ■

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.

33.3 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 $\mathcal{U} = \text{span}(1, \cos(\cdot), \sin(\cdot))$. 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

$$\int_{-\pi}^{\pi} 1\, dx = \frac{1}{2\pi}, \qquad \text{and} \qquad \int_{-\pi}^{\pi} \cos^2(x)\, dx = \int_{-\pi}^{\pi} \sin^2(x)\, dx = \pi,$$

the list

$$\left(\frac{1}{\sqrt{2\pi}}, \frac{\cos(\cdot)}{\sqrt{\pi}}, \frac{\sin(\cdot)}{\sqrt{\pi}}\right)$$

is orthonormal. Then

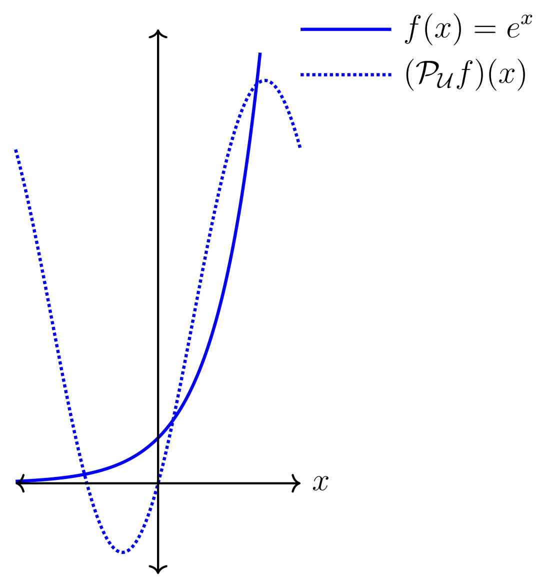
$$\begin{aligned}(\mathcal{P}_{\mathcal{U}} f)(x) = \left\langle f, \frac{1}{\sqrt{2\pi}} \right\rangle \frac{1}{\sqrt{2\pi}} + \left\langle f, \frac{\cos(\cdot)}{\sqrt{\pi}} \right\rangle \frac{\cos(x)}{\sqrt{\pi}} + \left\langle f, \frac{f}{\sin(\cdot)}\sqrt{\pi} \right\rangle \frac{\sin(x)}{\sqrt{\pi}} \\ = \frac{\langle f, 1\rangle}{2\pi} + \frac{\langle f, \cos(\cdot)\rangle}{\pi}\cos(x) + \frac{\langle f, \sin(\cdot)\rangle}{\pi}\sin(x)\end{aligned}$$

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.



33.4 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 $\mathcal{P}_{\mathcal{M}_A(\mathbb{F}^n)}$.

Now that we have an appreciation of some of the results that follow from an orthogonal decomposition $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$ of an inner product space $\mathcal{V}$ by a subspace $\mathcal{U}$, we study conditions that facilitate this representation. Certainly this is possible if $\mathcal{U}$ is finite-dimensional, and even when $\mathcal{V}$ is infinite-dimensional, many of the interesting subspaces in applications do turn out to be fiinite-dimensional. But we can push a bit beyond the restriction that $\mathcal{U}$ is finite-dimensional. Perhaps surprisingly, one possible replacement for the finite-dimensionality of $\mathcal{U}$ hinges on the behavior of certain functionals on $\mathcal{V}$.

33.5 Definition. *Let $\mathcal{V}$ be an inner product space. A functional $\varphi \in \mathcal{V}'$ is* **REPRESENTED** *by the inner product if there is $r \in \mathcal{V}$ such that if $\varphi = \langle \cdot, r \rangle$, i.e., if $\varphi(v) = \langle v, r \rangle$ for all $v \in \mathcal{V}$.*

33.6 Problem (!). Let $\mathcal{V}$ be an inner product space and r_1, $r_2 \in \mathcal{V}$. Prove that if $\langle \cdot, r_1 \rangle = \langle \cdot, r_2 \rangle$, then $r_1 = r_2$. [Hint: *if $\langle v, r_1 - r_2 \rangle = 0$ for all $v \in \mathcal{V}$, what do you know about r_1 and r_2?*] If $\varphi \in \mathcal{V}'$ is represented by the inner product with $\varphi = \langle \cdot, r \rangle$ for some $r \in \mathcal{V}$, we therefore call r *the* **REPRESENTING VECTOR** for φ.

Suppose that $\mathcal{V}$ is an inner product space and $\varphi \in \mathcal{V}'$ is represented by the inner product, and write $\varphi = \langle \cdot, r \rangle$ for some $r \in \mathcal{V}$. Then the Cauchy–Schwarz inequality implies that

$$|\varphi(v)| = |\langle v, r \rangle| \le \|v\| \, \|r\| \tag{33.3}$$

for all $v \in \mathcal{V}$. Functionals satisfying this sort of inequality have a special name.

33.7 Definition. *Let $\mathcal{V}$ be an inner product space. A functional $\varphi \in \mathcal{V}'$ is* **BOUNDED** *if there exists $C > 0$ such that*

$$|\varphi(v)| \le C \|v\| .$$

We denote the set of all bounded functionals on $\mathcal{V}$ by $\mathcal{V}^*$.

Day 34: Wednesday, November 5.

Vocabulary from today

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

Riesz representation property (for an inner product space *or* a subspace thereof)

34.1 Theorem. *Let $\mathcal{V}$ be an inner product space. The set of bounded functionals $\mathcal{V}^*$ on $\mathcal{V}$ is a subspace of $\mathcal{V}'$.*

Proof. Certainly $0_{\mathcal{V}\to\mathbb{F}} = 0_{\mathcal{V}\to\mathbb{F}}$ is bounded, since $|0_{\mathcal{V}\to\mathbb{F}}(v)| = 0 \le \|v\|$ for all $v \in \mathcal{V}$.

Now let φ_1, $\varphi_2 \in \mathcal{V}^*$; we will show $\varphi_1 + \varphi_2 \in \mathcal{V}^*$. We want to find $C > 0$ such that $|(\varphi_1 + \varphi_2)(v)| \le C \|v\|$ for all $v \in \mathcal{V}$. Since φ_1, $\varphi_2 \in \mathcal{V}^*$, we know that there are C_1, $C_2 > 0$ such that $|\varphi_1(v)| \le C_1 \|v\|$ and $|\varphi_2(v)| \le C_2 \|v\|$ for all $v \in \mathcal{V}$. The triangle inequality in $\mathbb{F}$ then gives

$$|(\varphi_1 + \varphi_2)(v)| = |\varphi_1(v) + \varphi_2(v)| \le |\varphi_1(v)| + |\varphi_2(v)| \le C_1 \|v\| + C_2 \|v\| = (C_1 + C_2) \|v\| .$$

Take $C = C_1 + C_2$. ■

34.2 Problem (!). Complete the proof by showing that $\alpha\varphi \in \mathcal{V}^*$ for all $\alpha \in \mathbb{F}$ and $\varphi \in \mathcal{V}^*$.

Here is a nice property of bounded linear functionals: they are continuous.

34.3 Example. Let $\mathcal{V}$ be an inner product space, $\varphi \in \mathcal{V}^*$, and (v_n) be a sequence in $\mathcal{V}$ with $v_n \to v \in \mathcal{V}$ (in the sense of Definition 28.16). Take $C > 0$ such that $|\varphi(w)| \le C \|w\|$ for all $w \in \mathcal{V}$. Then

$$|\varphi(v_n) - \varphi(v)| = |\varphi(v_n - v)| \le C \|v_n - v\| \to 0,$$

and so $\lim_{n\to\infty} \varphi(v_n) = \varphi(v)$ in $\mathbb{F}$.

Every functional on an inner product space that is represented by the inner product is a bounded functional by (33.3). A functional that is *not* bounded therefore *cannot* be represented by the inner product. In finite dimensions, it turns out that every functional is represented by the inner product and therefore bounded.

34.4 Example. Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal basis $(u_1, \ldots, u_n)$, and let $\varphi \in \mathcal{V}'$. Then

$$\varphi(v) = \varphi\left(\sum_{j=1}^n \langle v, u_j\rangle u_j\right) = \sum_{j=1}^n \langle v, u_j\rangle \varphi(u_j) = \sum_{j=1}^n \left\langle v, \overline{\varphi(u_j)} u_j\right\rangle = \left\langle v, \sum_{j=1}^n \overline{\varphi(u_j)} u_j\right\rangle .$$

In particular, every functional on a finite-dimensional vector space is represented by the inner product and therefore bounded.

34.5 Problem ($\star$). **(i)** Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal bases $(u_1, \ldots, u_n)$ and $(w_1, \ldots, w_n)$ and let $\varphi \in \mathcal{V}'$. Use Example 34.4 and Theorem 25.15 to deduce the following possibly surprising relationship:

$$\sum_{j=1}^n \overline{\varphi(u_j)} u_j = \sum_{j=1}^n \overline{\varphi(w_j)} w_j. \tag{34.1}$$

(ii) Continue to assume that $\mathcal{V}$ is a finite-dimensional inner product space with orthonormal bases $(u_1, \ldots, u_n)$ and $(w_1, \ldots, w_n)$ and let $\varphi \in \mathcal{V}'$. Here is why (34.1) is not such a surprise. Let $v := \sum_{j=1}^n \overline{\varphi(u_j)} u_j$. Explain why (34.1) says that

$$\langle v, w_j \rangle = \overline{\varphi(w_j)}. \tag{34.2}$$

Now calculate $\varphi(w_j)$ directly using the representation $w_j = \sum_{k=1}^n \langle w_j, u_k \rangle u_k$ (note the different index of summation k here) to obtain

$$\varphi(w_j) = \overline{\left\langle \sum_{k=1}^n \overline{\varphi(u_k)} u_k, w_j \right\rangle},$$

and this is (34.2).

34.6 Problem ($\star$). Here is another proof that every functional on a finite-dimensional inner product space is bounded. Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal basis $(u_1, \ldots, u_n)$ and let $\varphi \in \mathcal{V}'$. Let $v \in \mathcal{V}$.

(i) Use the triangle inequality to estimate

$$|\varphi(v)| \le \sum_{j=1}^n |\langle v, u_j \rangle| |\varphi(u_j)|.$$

(ii) By considering the n-tuples $(|\langle v, u_1 \rangle|, \ldots, |\langle v, u_n \rangle|)$ and $(|\varphi(u_1)|, \ldots, |\varphi(u_n)|)$ as vectors in $\mathbb{R}^n$, use the Cauchy–Schwarz inequality for the Euclidean inner product to estimate

$$\sum_{j=1}^n |\langle v, u_j \rangle| |\varphi(u_j)| \le C \|v\|, \qquad C := \left(\sum_{j=1}^n |\varphi(u_j)|^2 \right)^{1/2}.$$

(iii) Consider the representation of $\varphi(v)$ given in Example 34.4. Use the Cauchy–Schwarz inequality (for the inner product in $\mathcal{V}$) on that representation to estimate $|\varphi(v)| \le \widetilde{C} \|v\|$. Explain why $C = \widetilde{C}$.

34.7 Example. Consider the inner product space $\mathcal{C}([0,1])$ with the L^2-inner product. Define $\varphi(f) := \int_0^1 f(x) \, dx$. Then φ is a functional on $\mathcal{C}([0,1])$, and if we look closely at it, we will see that

$$\varphi(f) = \int_0^1 f(x) \, dx = \int_0^1 (f(x) \cdot 1) \, dx = \langle f, 1 \rangle,$$

so φ is represented by the inner product and therefore bounded. Specifically,

$$|\varphi(f)| = |\langle f, 1 \rangle| \le \|f\| \|1\| = \|f\|.$$

However, not every functional on an inner product space is bounded.

34.8 Example. Consider the inner product space $\mathcal{C}([0,1])$ with the L^2-inner product. Let $\varphi(f) = f(0)$, so φ is "evaluate at 0." We show that there cannot exist a constant $C > 0$ such that $|\varphi(f)| \le C\|f\|$ for all $f \in \mathcal{C}([0,1])$. That is, there is no $C > 0$ such that

$$|f(0)| \le C\left(\int_0^1 |f(x)|^2\, dx\right)^{1/2}$$

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

This is perhaps not hugely surprising. The integral on the right is a measure of the total area under the graph of f. Does a function's value at 0 really control the area under its graph? We might think about functions whose value at $= 0$ is "pinned" at a certain level, say, $f(0) = 1$, but whose areas become quite small.

A good class of such functions is $f_n(x) := (1-x)^n$. Since $0 < 1-x < 1$ for $0 < x < 1$, we find that the area under f_n becomes progressively smaller as n becomes larger. However, $f_n(0) = 1$ for all n. Specifically, we compute

$$\int_0^1 |f_n(x)|^2\, dx = \int_0^1 (1-x)^{2n}\, dx = -\int_1^0 u^{2n}\, du = \int_0^1 u^{2n}\, du = \frac{u^{2n+1}}{2n+1}\bigg|_{u=0}^{u=1} = \frac{1}{2n+1}.$$

Now suppose there is $C > 0$ such that $|\varphi(f)| \le C\|f\|$ for all $f \in \mathcal{C}([0,1])$. Taking $f = f_n$, we have

$$1 = |f_n(0)|^2 = |\varphi(f)|^2 \le C^2\|f\|^2 = \frac{C^2}{2n+1} \to 0$$

as $n \to \infty$. This is a contradiction, and so φ cannot be bounded.

And not every bounded functional on an inner product space needs to be represented by the inner product.

34.9 Example. Consider the inner product space $\mathcal{C}([0,1])$ with the L^2-inner product. Let $\varphi(f) = \int_0^{1/2} f(x)\, dx$, so φ is a functional on $\mathcal{C}([0,1])$. We first check that φ is indeed bounded:

$$|\varphi(f)| \le \int_0^{1/2} |f(x)|\, dx \le \int_0^1 |f(x)|\, dx = |\langle |f|, 1\rangle| \le \||f|\|_{L^2}\|1\| = \|f\|.$$

We would probably like to say that $\varphi = \langle \cdot, r_0\rangle$, where

$$r_0(x) = \begin{cases} 1, & 0 \le x < 1/2 \\ 0, & 1/2 \le x \le 1. \end{cases}$$

It is absolutely true that $\varphi(f) = \int_0^1 f(x)r_0(x)\, dx$, assuming that we have a notion of integral for piecewise continuous functions, but $r_0 \notin \mathcal{C}([0,1])$.

Can there possibly be $r \in \mathcal{C}([0,1])$ such that $\varphi = \langle \cdot, r\rangle$? No. It is possible to show that such an r would need to satisfy

$$r(x) = \begin{cases} 1, & 0 \le x \le 1/2 \\ 0, & 1/2 \le x \le 1. \end{cases}$$

This both means that r is discontinuous and not a function!

Here is what goes wrong. Our work relies on the fundamental lemma of the calculus of variations, which says that if $\phi \in \mathcal{C}([a,b])$ satisfies $\int_a^b \phi(x)\psi(x)\ dx = 0$ for all $\psi \in \mathcal{C}([a,b])$ with $\psi(a) = \psi(b) = 0$, then $\phi = 0$. This lemma is not terribly difficult to prove using effectively only the basic properties of integrals that we have already employed, but we will not pursue that proof here.

Instead, we deploy that fundamental lemma in two passes. First, let $g \in \mathcal{C}([1/2,1])$ with $g(1/2) = g(1) = 0$ and let

$$\widetilde{g}(x) = \begin{cases} 0, & 0 \le x < 1/2 \\ g(x), & 1/2 \le x \le 1. \end{cases}$$

Then $\widetilde{g} \in \mathcal{C}([0,1])$, so

$$0 = \int_0^{1/2} \widetilde{g}(x)\ dx = \varphi(\widetilde{g}) = \int_0^1 \widetilde{g}(x)\overline{r(x)}\ dx = \int_{1/2}^1 g(x)\overline{r(x)}\ dx.$$

The fundamental lemma of the calculus of variations implies that $r(x) = 0$ for $1/2 \le x \le 1$.

Thus

$$\int_0^{1/2} f(x)\ dx = \varphi(f) = \int_0^1 f(x)\overline{r(x)}\ dx = \int_0^{1/2} f(x)\overline{r(x)}\ dx$$

for any $f \in \mathcal{C}([0,1])$, so also

$$\int_0^{1/2} f(x)(1 - \overline{r(x)})\ dx = 0$$

for all $f \in \mathcal{C}([0,1])$. Now let $h \in \mathcal{C}([0,1/2])$ with $h(0) = h(1/2) = 0$. Put

$$\widetilde{h}(x) = \begin{cases} h(x), & 0 \le x < 1/2 \\ 0, & 1/2 \le x \le 1. \end{cases}$$

Then $\widetilde{h} \in \mathcal{C}([0,1])$ and

$$0 = \int_0^{1/2} \widetilde{h}(x)(1 - \overline{r(x)})\ dx = \int_0^{1/2} h(x)(1 - \overline{r(x)})\ dx.$$

Another application of the fundamental lemma of the calculus of variations gives $1 - \overline{r(x)} = 0$ for $0 \le x \le 1/2$. In particular, $0 = r(1/2) = 1$, a contradiction.

The problem with the previous example was that the vector that we wanted to represent our functional did not belong to the domain of the functional. The same situation arises in the next example.

34.10 Example. Let

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

with the ℓ^2-inner product. We will think of $\mathcal{V}$ as the inner product space in this example and not of $\mathcal{V}$ as a subspace of ℓ^2 (as we previously did in Example 30.8).

Let $g(k) = 1/k$, so $g \in \ell^2 \setminus \mathcal{V}$. Define $\varphi \in \mathcal{V}^*$ by $\varphi(f) = \langle f, g \rangle$. (Since f, $g \in \ell^2$, this inner product is defined.) While φ is defined via the inner product, we cannot strictly say that φ is represented by the inner product on $\mathcal{V}$, since $g \notin \mathcal{V}$. However, $\varphi \in \mathcal{V}^*$ by the Cauchy–Schwarz inequality for the ℓ^2-inner product.

Suppose that we can represent φ by the inner product on $\mathcal{V}$: assume there is $r \in \mathcal{V}$ such that $\varphi(f) = \langle f, r \rangle$ for all $f \in \mathcal{V}$. With $e_j(k) = 1$ for $j = k$ and 0 otherwise, take $f = e_j$ to find

$$\frac{1}{j} = \overline{g(j)} = \langle e_j, g \rangle = \varphi(e_j) = \langle e_j, r \rangle = r(j),$$

thus $r(j) \neq 0$ for all j, so $r \notin \mathcal{V}$. So, we cannot represent φ by the inner product on $\mathcal{V}$.

We might, somewhat euphemistically, think that every bounded functional on an inner product space "should" be represented by the inner product, and that an inner product space for which this representation does not always hold is somehow "incomplete." The best inner product spaces are the ones that *do* have this representation property.

34.11 Definition. **(i)** *An inner product space $\mathcal{V}$ has the* **Riesz representation property (RRP)** *if every bounded functional on $\mathcal{V}$ can be represented by the inner product: for all $\varphi \in \mathcal{V}^*$, there is a (necessarily unique) $r \in \mathcal{V}$ such that $\varphi = \langle \cdot, r \rangle$. That is, for all $\varphi \in \mathcal{V}^*$, there is $r \in \mathcal{V}$ such that $\varphi(v) = \langle v, r \rangle$ for all $v \in \mathcal{V}$.*

(ii) *A subspace $\mathcal{U}$ of the inner product space $\mathcal{V}$ has the Riesz representation property (RRP) if $\mathcal{U}$ has the RRP when considered as an inner product space with the inner product of $\mathcal{V}$. That is, if $\langle \cdot, \cdot \rangle$ is the inner product on $\mathcal{V}$, then for each $\varphi \in \mathcal{U}^*$, there is $r \in \mathcal{U}$ such that $\varphi(u) = \langle u, r \rangle$ for all $u \in \mathcal{U}$.*

Day 35: Friday, November 7.

35.1 Example. **(i)** Every finite-dimensional inner product space has the RRP by Example 34.4.

(ii) The space $\mathcal{C}([0, 1])$ with the L^2-inner product does not have the RRP by Example 34.9. This is rather disappointing: $\mathcal{C}([0, 1])$ has been absolutely central to our blending of calculus and linear algebra. Part of the journey from mathematical innocence to experience is realizing that some of the familiar, comfortable settings of calculus are inadequate for deeper analysis, and new structures are needed.

(iii) While the space $\left\{f \in \ell^2 \;\middle|\; f(k) = 0 \text{ for all but finitely many } f\right\}$ does not have the RRP, by Example 34.10, the larger space ℓ^2 does, as we now show. Let $\varphi \in (\ell^2)^*$ and take $C > 0$ such that $|\varphi(f)| \leq C\,\|f\|$ for all $f \in \ell^2$.

First we determine a *candidate* for the representing vector for φ, then we show that this candidate is in ℓ^2, and last we show that this candidate *does* represent φ. If there is $r \in \ell^2$ such that $\varphi(f) = \langle f, r\rangle_{\ell^2}$ for all $f \in \ell^2$, then in particular $\varphi(e_j) = \langle e_j, r\rangle_{\ell^2} = \overline{r(j)}$, thus $r(j) = \overline{\varphi(e_j)}$. Here e_j is, as usual, the sequence with $e_j(k) = 1$ for $j = k$ and 0 for $j \neq k$.

We first show that r as defined by $r(k) = \overline{\varphi(e_k)}$ is in ℓ^2. That is, we want to show that the series $\sum_{k=1}^{\infty}|\varphi(e_k)|^2$ converges; here we are using $|\varphi(e_k)| = |\overline{\varphi(e_k)}|$. We use an auxiliary fact from calculus: if (a_k) is a sequence in $\mathbb{F}$ and there exists $c > 0$ such that for all n, the estimate $\sum_{k=1}^{n}|a_k| \leq c$ holds, then the series $\sum_{k=1}^{\infty}|a_k|$ converges.

So, we want to find $c > 0$ such that $s_n := \sum_{k=1}^{n}|\varphi(e_k)|^2 \leq c$ for all n; this will prove convergence of the series $\sum_{k=1}^{\infty}|\varphi(e_k)|^2$. We have

$$s_n = \sum_{k=1}^{n}|\varphi(e_k)|^2 = \sum_{k=1}^{n}\overline{\varphi(e_k)}\varphi(e_k) = \varphi\left(\sum_{k=1}^{n}\overline{\varphi(e_k)}e_k\right),$$

and so

$$s_n = |s_n| = \left|\varphi\left(\sum_{k=1}^{n}\overline{\varphi(e_k)}e_k\right)\right| \leq C\left\|\sum_{k=1}^{n}\overline{\varphi(e_k)}e_k\right\|.$$

The last inequality is true because φ is bounded.

The sequence $(e_1, \ldots, e_n)$ is orthonormal in ℓ^2, so

$$\left\|\sum_{k=1}^{n}\overline{\varphi(e_k)}e_k\right\|^2 = \sum_{k=1}^{n}|\overline{\varphi(e_k)}|^2 = \sum_{k=1}^{n}|\varphi(e_k)|^2 = s_n.$$

And so we have shown

$$s_n^2 \leq C\left\|\sum_{k=1}^{n}\overline{\varphi(e_k)}e_k\right\|^2 = Cs_n.$$

If $s_n \neq 0$ for some n, then we may divide to find $s_n \leq C^2$. If $s_n = 0$ for some n, then of course we have $s_n \leq C^2$. So, for all n, we have $s_n \leq C^2 =: c$, as desired; thus $r \in \ell^2$.

Now we check that $\varphi(f) = \langle f, r\rangle$ for all $f \in \ell^2$. Since $(e_j)_{j\in\mathbb{N}}$ is an orthonormal basis for ℓ^2, we have $f = \sum_{j=1}^{\infty}\langle f, e_j\rangle\, e_j$, thus

$$\varphi(f) = \varphi\left(\sum_{j=1}^{\infty}\langle f, e_j\rangle\, e_j\right) = \sum_{j=1}^{\infty}\langle f, e_j\rangle\,\varphi(e_j) = \sum_{j=1}^{\infty}f(j)\overline{r(j)} = \langle f, r\rangle_{\ell^2}.$$

The second equality follows from Example 34.3.

(iv) Let $\mathcal{V}$ be a separable inner product space with orthonormal basis $(u_j)_{j\in\mathbb{N}}$. Does this imply that $\mathcal{V}$ has the RRP? Given $\varphi \in \mathcal{V}^*$, suppose that $\varphi = \langle \cdot, r \rangle$ for some $r \in \mathcal{V}$. Then $\langle r, u_j \rangle = \overline{\varphi(u_j)}$, so $r = \sum_{j=1}^{\infty} \overline{\varphi(u_j)} u_j$.

So, to prove that $\mathcal{V}$ has the RRP, we would start with $\varphi \in \mathcal{V}^*$ and try to show that

$$\varphi(v) = \left\langle v, \sum_{j=1}^{\infty} \overline{\varphi(u_j)} u_j \right\rangle.$$

This would largely proceed along the lines of the calculations in Example 34.4, but there is one questionable point: what guarantees that the series $\sum_{j=1}^{\infty} \overline{\varphi(u_j)} u_j$ converges in the first place? We could mimic the work in part (iii) above to show that the series $\sum_{j=1}^{\infty} |\varphi(u_j)|^2$ converges in $\mathbb{F}$, but now, outside the tangible setting of ℓ^2, it is not clear—and in fact not guaranteed—that the convergence of a series of the form $\sum_{j=1}^{\infty} |\alpha_j|^2$ in $\mathbb{F}$ guarantees convergence of the series $\sum_{j=1}^{\infty} \alpha_j u_j$ in $\mathcal{V}$.

35.2 Problem (!). Suppose that the inner product space $\mathcal{V}$ has the RRP. Define

$$\mathcal{T} : \mathcal{V} \to \mathcal{V}^* : r \mapsto \langle \cdot, r \rangle.$$

Prove that $\mathcal{T}$ is bijective, that $\mathcal{T}(r_1 + r_2) = \mathcal{T} r_1 + \mathcal{T} r_2$ for all $r_1, r_2 \in \mathcal{V}$, and that $\mathcal{T}(\alpha r) = \overline{\alpha} \mathcal{T} r$ for all $\alpha \in \mathbb{F}$, $r \in \mathcal{V}$. Because of this last identity, $\mathcal{T}$ is not linear, and instead we call $\mathcal{T}$ **ANTILINEAR**.

We now show how the RRP addresses our desire to generalize those inner product spaces $\mathcal{V}$ and subspaces $\mathcal{U}$ that have the orthogonal direct sum decomposition $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$. First, we need some additional properties of the orthogonal projection operator from Theorem 31.5.

35.3 Lemma. *Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Suppose that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$. The orthogonal projection $\mathcal{P}_{\mathcal{U}} : \mathcal{V} \to \mathcal{U}$ satisfies the following.*

(i) $\langle \mathcal{P}_{\mathcal{U}} v, w \rangle = \langle v, \mathcal{P}_{\mathcal{U}} w \rangle$ *for all $v, w \in \mathcal{V}$.*

(ii) $\|\mathcal{P}_{\mathcal{U}} v\| \le \|v\|$ *for all $v \in \mathcal{V}$.*

Proof. **(i)** Fix $v, w \in \mathcal{V}$ and write $w = \mathcal{P}_{\mathcal{U}} w + u^\perp$ for some $u^\perp \in \mathcal{U}^\perp$. Then

$$\langle \mathcal{P}_{\mathcal{U}} v, w \rangle = \langle \mathcal{P}_{\mathcal{U}} v, \mathcal{P}_{\mathcal{U}} w + u^\perp \rangle = \langle \mathcal{P}_{\mathcal{U}} v, \mathcal{P}_{\mathcal{U}} w \rangle + \langle \mathcal{P}_{\mathcal{U}} v, u^\perp \rangle.$$

Since $\mathcal{P}_{\mathcal{U}} v \in \mathcal{U}$ and $u^\perp \in \mathcal{U}^\perp$, we have

$$\langle \mathcal{P}_{\mathcal{U}} v, u^\perp \rangle = 0,$$

thus

$$\langle \mathcal{P}_{\mathcal{U}} v, w \rangle = \langle \mathcal{P}_{\mathcal{U}} v, \mathcal{P}_{\mathcal{U}} w \rangle.$$

Interchanging the roles of v and w therefore gives

$$\langle \mathcal{P}_{\mathcal{U}} w, v \rangle = \langle \mathcal{P}_{\mathcal{U}} w, \mathcal{P}_{\mathcal{U}} v \rangle.$$

Then

$$\langle \mathcal{P}_{\mathcal{U}} v, w\rangle = \langle \mathcal{P}_{\mathcal{U}} v, \mathcal{P}_{\mathcal{U}} w\rangle = \overline{\langle \mathcal{P}_{\mathcal{U}} w, \mathcal{P}_{\mathcal{U}} v\rangle} = \overline{\langle \mathcal{P}_{\mathcal{U}} w, v\rangle} = \langle v, \mathcal{P}_{\mathcal{U}} w\rangle .$$

(ii) We work backwards: let $v \in \mathcal{V}$ and write $v = \mathcal{P}_{\mathcal{U}} v + u^{\perp}$. Since $\langle \mathcal{P}_{\mathcal{U}} v, u^{\perp}\rangle = 0$, the Pythagorean identity gives

$$\|v\|^2 = \left\|\mathcal{P}_{\mathcal{U}} v + u^{\perp}\right\|^2 = \|\mathcal{P}_{\mathcal{U}} v\|^2 + \left\|u^{\perp}\right\|^2 \geq \|\mathcal{P}_{\mathcal{U}} v\|^2 .$$ ■

35.4 Theorem. *Let $\mathcal{V}$ be an inner product space with the RRP and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Then $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$ if and only if $\mathcal{U}$ has the RRP.*

Proof. ($\Longrightarrow$) Let $\varphi \in \mathcal{U}$. Take $C > 0$ such that $|\varphi(u)| \leq C\|u\|$ for all $u \in \mathcal{U}$. The goal is to find $r \in \mathcal{U}$ such that $\varphi(u) = \langle u, r\rangle$ for all $u \in \mathcal{U}$. Since $\mathcal{V}$ has the RRP, a natural first step is to relate φ to a functional defined on all of $\mathcal{V}$, invoke the RRP for $\mathcal{V}$ to represent *that* functional by a vector in $\mathcal{V}$, and then show that this representing vector really is in $\mathcal{U}$. And since $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$, the most natural way to get a functional on $\mathcal{V}$ out of a functional on $\mathcal{U}$ is to involve the orthogonal projection $\mathcal{P}_{\mathcal{U}} \colon \mathcal{V} \to \mathcal{U}$.

So, define

$$\psi \colon \mathcal{V} \to \mathbb{F} \colon v \mapsto \varphi(\mathcal{P}_{\mathcal{U}} v).$$

Then $\psi \in \mathcal{V}'$ since $\varphi \in \mathcal{U}^* \subseteq \mathcal{U}'$ and $\mathcal{P}_{\mathcal{U}} \in \mathbf{L}(\mathcal{V}, \mathcal{U})$. Next, we estimate

$$|\psi(v)| \leq C\|\mathcal{P}_{\mathcal{U}} v\| \leq C\|v\| .$$

This uses the boundedness of φ and the estimate on $\mathcal{P}_{\mathcal{U}}$ from part (ii) of Lemma 35.3. Thus $\psi \in \mathcal{V}^*$, and so there is $r \in \mathcal{V}$ such that $\psi = \langle \cdot, r\rangle$.

We claim that $r \in \mathcal{U}$. If so, then we have

$$\varphi(u) = \varphi(\mathcal{P}_{\mathcal{U}} u) = \psi(u) = \langle u, r\rangle .$$

This shows that $\mathcal{U}$ has the RRP.

We show $r \in \mathcal{U}$ by proving that $\langle v, r\rangle = \langle v, \mathcal{P}_{\mathcal{U}} r\rangle$ for all $v \in \mathcal{V}$. It then follows from Theorem 25.15 that $r = \mathcal{P}_{\mathcal{U}} r \in \mathcal{U}$. So, we compute

$$\begin{aligned}
\langle v, r\rangle &= \psi(v) \text{ by definition of } r \\
&= \varphi(\mathcal{P}_{\mathcal{U}} v) \text{ by definition of } \psi \\
&= \varphi(\mathcal{P}_{\mathcal{U}}^2 v) \text{ since } \mathcal{P}_{\mathcal{U}} = \mathcal{P}_{\mathcal{U}}^2 \text{ for orthogonal projections} \\
&= \varphi(\mathcal{P}_{\mathcal{U}}(\mathcal{P}_{\mathcal{U}} v)) \text{ by definition of } \mathcal{P}_{\mathcal{U}}^2 \\
&= \psi(\mathcal{P}_{\mathcal{U}} v) \text{ by definition of } \psi \text{ again} \\
&= \langle \mathcal{P}_{\mathcal{U}} v, r\rangle \text{ by definition of } r \text{ again} \\
&= \langle v, \mathcal{P}_{\mathcal{U}} r\rangle \text{ by part (i) of Lemma 35.3.}
\end{aligned}$$ ■

35.5 Problem ($\star$). The reverse implication in Theorem 35.4 can be proved by applying an idea from Problem 31.3. Specifically, for $v \in \mathcal{V}$, define $\varphi \in \mathcal{U}^*$ by

$$\varphi \colon \mathcal{U} \to \mathbb{F} \colon u \mapsto \langle u, v\rangle .$$

Apply the RRP for $\mathcal{U}$ to get $\varphi = \langle \cdot, r\rangle$ for some $r \in \mathcal{U}$, and show that $v - r \in \mathcal{U}^{\perp}$. Why does this show that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$?

Day 36: Monday, November 10.

We know that if an inner product space $\mathcal{V}$ satisfies $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$ for a subspace $\mathcal{U}$, then $\mathcal{U}^{\perp\perp} = \mathcal{U}$. This was Theorem 31.9, and that theorem did *not* require $\mathcal{V}$ to have the RRP. We might wonder if having $\mathcal{U}^{\perp\perp} = \mathcal{U}$ is enough to guarantee the orthogonal direct sum decomposition $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$. The answer is yes if $\mathcal{U}^{\perp}$ has the RRP, for then $\mathcal{V} = \mathcal{U}^{\perp} \oplus \mathcal{U}^{\perp\perp}$. But since $\mathcal{U}^{\perp\perp} = \mathcal{U}$ here, that just says $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$.

It turns out that if $\mathcal{V}$ has the RRP and $\mathcal{U}$ is any subspace of $\mathcal{V}$, then $\mathcal{U}^{\perp}$ *always* has the RRP (regardless of whether or not $\mathcal{U}$ does). Proving this[†] requires a deep theorem, whose proof in turn hinges on Zorn's lemma, like the proof that every vector space has a basis.

36.1 Theorem (Hahn–Banach extension theorem). *Let $\mathcal{V}$ be an inner product space and let $\mathcal{W}$ be a subspace of $\mathcal{V}$. Suppose that $\varphi \in \mathcal{W}^*$.Then there is $\widetilde{\varphi} \in \mathcal{V}^*$ such that $\widetilde{\varphi}(w) = \varphi(w)$ for all $w \in \mathcal{W}$ and $|\widetilde{\varphi}(v)| \le C \|v\|$ for all $v \in \mathcal{V}$.*

36.2 Lemma. *Let $\mathcal{V}$ be an inner product space with the RRP and let $\mathcal{U}$ be a subspace of $\mathcal{X}$. Then $\mathcal{U}^{\perp}$ has the RRP.*

Proof. Let $\varphi \in (\mathcal{U}^{\perp})^*$. Our goal is to find $r \in \mathcal{U}^{\perp}$ such that $\varphi(u^{\perp}) = \langle u^{\perp}, r\rangle$ for all $u^{\perp} \in \mathcal{U}^{\perp}$. We do this by passing through an "intermediate" subspace of $\mathcal{V}$.

Put

$$\mathcal{W} := \left\{u + u^{\perp} \;\middle|\; u \in \mathcal{U},\ u^{\perp} \in \mathcal{U}^{\perp}\right\}.$$

Then $\mathcal{W}$ is a subspace of $\mathcal{V}$. Each $w \in \mathcal{W}$ can be written uniquely in the form $w = u + u^{\perp}$ for some $u \in \mathcal{U}$ and $u^{\perp} \in \mathcal{U}^{\perp}$; existence of this decomposition is in the definition of $\mathcal{W}$, and uniqueness is Lemma 30.9. Define

$$\psi \colon \mathcal{W} \to \mathbb{F} \colon u + u^{\perp} \mapsto \varphi(u^{\perp}).$$

Then $\psi \in \mathcal{W}^*$ and if $u \in \mathcal{U}$, then

$$\psi(u) = \psi(u + 0_{\mathcal{V}}) = \varphi(0_{\mathcal{V}}) = 0. \tag{36.1}$$

By the Hahn–Banach theorem, there is $\widetilde{\psi} \in \mathcal{V}^*$ such that $\widetilde{\psi}(w) = \psi(w)$ for all $w \in \mathcal{W}$. Since $\mathcal{V}$ has the RRP, we can write $\widetilde{\psi} = \langle \cdot, r\rangle$ for some $r \in \mathcal{V}$. Then for $u^{\perp} \in \mathcal{U}^{\perp}$, we have

$$\left\langle u^{\perp}, r\right\rangle = \widetilde{\psi}(u^{\perp}) = \widetilde{\psi}(0_{\mathcal{V}} + u^{\perp}) = \psi(0_{\mathcal{V}} + u^{\perp}) = \varphi(u^{\perp}).$$

We will show $r \in \mathcal{U}^{\perp}$ by checking $\langle u, r\rangle = 0$ for all $u \in \mathcal{U}$, and then we will be done. For $u \in \mathcal{U} \subseteq \mathcal{W}$, we compute

$$\langle u, r\rangle = \widetilde{\psi}(u) = \psi(u) = 0$$

[†] The proof of Lemma 36.2 is due to `Tuvasbien`'s post https://math.stackexchange.com/questions/5082676/riesz-representation-for-orthogonal-complements-without-using-completeness.

by (36.1). ■

So, per the discussion above, if $\mathcal{U}^{\perp\perp} = \mathcal{U}$ for a subspace $\mathcal{U}$ of the inner product space $\mathcal{V}$ with the RRP, then $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$.

There is also a more geometric characterization of when a subspace has the RRP. We know that if an inner product space $\mathcal{V}$ and subspace $\mathcal{U}$ satisfy $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$, then the orthogonal projection $\mathcal{P}_{\mathcal{U}} \colon \mathcal{V} \to \mathcal{U}$ gives the "best approximation" of any $v \in \mathcal{V}$ by a vector in $\mathcal{U}$ via the inequality $\|\mathcal{P}_{\mathcal{U}}v - v\| \le \|u - v\|$ for all $u \in \mathcal{U}$.

36.3 Definition. *Let $\mathcal{V}$ be an inner product space. A subspace $\mathcal{U}$ of $\mathcal{V}$ has the* BEST APPROXIMATION PROPERTY *if for each $x \in \mathcal{U}$, there exists a unique $\mathcal{A}(v) \in \mathcal{U}$ such that $\|v - \mathcal{A}(v)\| \le \|v - u\|$ for all $u \in \mathcal{U}$.*

36.4 Theorem. *Let $\mathcal{V}$ be an inner product space and $\mathcal{U}$ be a subspace of $\mathcal{V}$. Then $\mathcal{U}$ has the RRP if and only if $\mathcal{U}$ has the best approximation property.*

Proof. ($\Longrightarrow$) If $\mathcal{U}$ has the RRP, then $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$, and we can take $\mathcal{A}(v) = \mathcal{P}_{\mathcal{U}}v$ to get the best approximation of v by a vector in $\mathcal{U}$. ■

36.5 Problem ($\star$). Let $\mathcal{V}$ be an inner product space and $\mathcal{U}$ be a subspace of $\mathcal{V}$. Suppose that $\mathcal{U}$ has the best approximation property; in the notation of Definition 36.3, write any $v \in \mathcal{V}$ as $v = \mathcal{A}(v) + (v - \mathcal{A}(v))$, so $\mathcal{A}(v) \in \mathcal{U}$. Show that $v - \mathcal{A}(v) \in \mathcal{U}^{\perp}$ by using the characterization (30.1) of $\mathcal{U}^{\perp}$. [Hint: *$\mathcal{A}(v) + u \in \mathcal{U}$ for all $v \in \mathcal{V}$ and $u \in \mathcal{U}$.*]

We can now augment Theorem 35.4 with two more equivalent conditions.

36.6 Theorem. *Let $\mathcal{V}$ be an inner product space with the RRP and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. The following are equivalent:*

(i) *$\mathcal{U}$ has the RRP.*

(ii) *$\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$.*

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

(iv) *$\mathcal{U}$ has the best approximation property.*

What guarantees that an inner product space has the RRP? There are multiple equivalent conditions; with sufficient background in analysis and topology, one can show that the RRP is equivalent to the COMPLETENESS of $\mathcal{V}$ with respect to the norm induced by the inner product (and so $\mathcal{V}$ is a HILBERT SPACE). We develop a characterization based on the kernels of bounded linear functionals.

36.7 Lemma. *Let $\mathcal{V}$ be a vector space and let $\varphi_1, \varphi_2 \in \mathcal{V}'$. Suppose that $\ker(\varphi_1) \subseteq \ker(\varphi_2)$. Then there is $\alpha \in \mathbb{F}$ such that $\varphi_2 = \alpha\varphi_1$.*

Proof. 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 $\alpha = \varphi_2(v_0)/\varphi_1(v_0)$. ■

36.8 Problem ($\star$). Let $\mathcal{V}$ be an inner product space and $r \in \mathcal{V}$. Show that

$$\ker(\langle\cdot, r\rangle) = \operatorname{span}(r)^\perp \quad \text{and} \quad \ker(\langle\cdot, r\rangle)^\perp = \operatorname{span}(r).$$

[Hint: *for the first equality, use definitions; for the second, what do you know about $\mathcal{U}^{\perp\perp}$ when $\mathcal{U}$ is a finite-dimensional subspace of $\mathcal{V}$?*]

Day 37: Wednesday, November 12.

Now we are ready to characterize spaces with the RRP.

37.1 Theorem. *Let $\mathcal{V}$ be an inner product space. The following are equivalent.*

(i) *$\mathcal{V}$ has the RRP.*

(ii) $\mathcal{V} = \ker(\varphi) \oplus \ker(\varphi)^{\perp}$ *for all* $\varphi \in \mathcal{V}^*$.

(iii) $\ker(\varphi)^{\perp\perp} = \ker(\varphi)$ *for all* $\varphi \in \mathcal{V}^*$.

Proof. (i) $\Longrightarrow$ (ii) Let $\varphi \in \mathcal{V}^*$. Since $\mathcal{V}$ has the RRP, we can write $\varphi = \langle \cdot, r\rangle$ for some $r \in \mathcal{V}$. Problem 36.8 tells us that $\ker(\varphi) = \operatorname{span}(r)^{\perp}$ and $\ker(\varphi)^{\perp} = \operatorname{span}(r)$. Since $\operatorname{span}(r)$ is finite-dimensional, $\mathcal{V} = \operatorname{span}(r) \oplus \operatorname{span}(r)^{\perp}$ by Theorem 30.10, thus $\mathcal{V} = \ker(\varphi) \oplus \ker(\varphi)^{\perp}$.

(ii) $\Longrightarrow$ (iii) This is Theorem 31.9.

(iii) $\Longrightarrow$ (i) Let $\varphi \in \mathcal{V}^*$. We want to find $r \in \mathcal{V}$ such that $\varphi = \langle \cdot, r\rangle$. This is easy in the case that $\varphi = 0_{\mathcal{V}\to\mathbb{F}}$, for then we can just take $r = 0_{\mathcal{V}}$.

Assume from now on that $\varphi \neq 0_{\mathcal{V}\to\mathbb{F}}$. We work backwards: if we have such an r with $\varphi = \langle \cdot, r\rangle$, what do we know about r? Problem 36.8 tells us that r satisfies $\ker(\varphi) = \operatorname{span}(r)^{\perp}$ and $\ker(\varphi)^{\perp} = \operatorname{span}(r)$. In particular, we want $r \in \ker(\varphi)^{\perp}$.

So, we might take some $w \in \ker(\varphi)^{\perp}$ and use w to obtain the representing vector r. (Since there can be only one r with $\varphi = \langle \cdot, r\rangle$, we are not guaranteed that the $w \in \ker(\varphi)^{\perp}$ that we pick will automatically satisfy $\varphi = \langle \cdot, w\rangle$. Perhaps there will be some intermediate work.) There is one w that we may want to avoid: if we take $w = 0_{\mathcal{V}}$, then $\langle \cdot, w\rangle = 0_{\mathcal{V}\to\mathbb{F}}$. But now we are assuming $\varphi \neq 0_{\mathcal{V}\to\mathbb{F}}$. Is it possible to choose $w \in \ker(\varphi)^{\perp} \setminus \{0_{\mathcal{V}}\}$?

Yes: for if $\ker(\varphi)^{\perp} = \{0_{\mathcal{V}}\}$, then

$$\ker(\varphi) = \ker(\varphi)^{\perp\perp} = \{0_{\mathcal{V}}\}^{\perp} = \mathcal{V},$$

and so $\varphi = 0_{\mathcal{V}\to\mathbb{F}}$ after all. That is, if $\varphi \neq 0_{\mathcal{V}\to\mathbb{F}}$, then there is $w \in \ker(\varphi)^{\perp} \setminus \{0_{\mathcal{V}}\}$.

Here is what any such w does: $\langle v, w\rangle = 0$ for all $v \in \ker(\varphi)$ and $w \neq 0_{\mathcal{V}}$. That is, $\ker(\varphi) \subseteq \ker(\langle \cdot, w\rangle)$. By Lemma 36.7, there is $\alpha \in \mathbb{F}$ such that $\langle \cdot, w\rangle = \alpha\varphi$. If $\alpha \neq 0$, then we can divide to find $\varphi = \langle \cdot, w/\overline{\alpha}\rangle$, and so we take $r = w/\overline{\alpha}$. If $\alpha = 0$, then $\langle \cdot, w\rangle = 0_{\mathcal{V}\to\mathbb{F}}$. That is, $\langle v, w\rangle = 0$ for all $v \in \mathcal{V}$, and then $w = 0_{\mathcal{V}}$ by Theorem 25.15, a contradiction. ■

We promised two major uses of inner product spaces in understanding the fundamental problem $\mathcal{T}v = w$ for an operator $\mathcal{T}$ between the spaces $\mathcal{V}$ and $\mathcal{W}$. The first was least squares approximation: when we cannot solve $\mathcal{T}v = w$, what is the best problem to solve instead? If $\mathcal{W}$ and $\mathcal{T}(\mathcal{V})$ have the RRP, write $\mathcal{W} = \mathcal{T}(\mathcal{V}) \oplus \mathcal{T}(\mathcal{V})^{\perp}$ and solve $\mathcal{T}\widehat{v} = \mathcal{P}_{\mathcal{T}(\mathcal{V})}w$.

The next application is a geometric characterization of the range. What does it mean to have $w \in \mathcal{T}(\mathcal{V})$ beyond the definition? We begin with a familiar example. Let

$$\mathcal{V} = \{f \in \mathcal{C}^{\infty}(\mathbb{R}) \mid f(x) = f(x + 2\pi) \text{ for all } x \in \mathbb{R}\}$$

and

$$\langle f, g\rangle := \int_0^{2\pi} f(x)g(x)\, dx, \ f, \ g \in \mathcal{V}.$$

37.2 Lemma. *The map* $\langle f, g\rangle := \int_0^{2\pi} f(x)g(x)\, dx$ *is an inner product on* $\mathcal{V}$ *as defined above.*

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 $\int_0^{2\pi} |f(x)|^2 \, dx = 0$. Per Example 25.7, we have $f(x) = 0$ for all $0 \le x \le 2\pi$. Since f is 2π-periodic, for any $y \in \mathbb{R}$, there are $x \in [0, 2\pi]$ and $k \in \mathbb{Z}$ such that $f(y) = f(x + 2\pi k) = f(x) = 0$, thus $f = 0$. ■

37.3 Problem (!). Why is $\langle f, g\rangle := \int_0^{2\pi} f(x)g(x) \, dx$ 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* $x \in \mathbb{R} \setminus [a, b]$.]

Now let $g \in \mathcal{V}$ and consider the problem $\mathcal{T} f = g$. By the fundamental theorem of calculus, this is equivalent to

$$f(x) = f(0) + \int_0^x g(s) \, ds. \tag{37.1}$$

Does defining f in this way actually yield a function in $\mathcal{V}$? What is new here is that we require $f(x + 2\pi) = f(x)$. Taking $x = 0$, we see that this definition of f and 2π-periodicity require

$$\int_0^{2\pi} g(s) \, ds = 0. \tag{37.2}$$

Conversely, if g satisfies this "mean-zero" condition (37.2), then

$$\int_0^{x+2\pi} g(s) \, ds = \int_0^x g(s) \, ds + \int_x^{x+2\pi} g(s) \, ds.$$

Put $G(x) := \int_x^{x+2\pi} g(s) \, ds$, so $G'(x) = g(x + 2\pi) - g(x) = 0$ since g is 2π-periodic. Thus G is constant and $G(0) = \int_0^{2\pi} g(s) \, ds = 0$, hence $G = 0$. We conclude $\int_0^{x+2\pi} g(s) \, ds = \int_0^x g(s) \, ds$ for all x, which gives the 2π-periodicity of f as defined in (37.1).

Here is what we have shown:

$$\mathcal{T}(\mathcal{V}) = \left\{ g \in \mathcal{V} \;\middle|\; \int_0^{2\pi} g(s) \, ds = 0 \right\}. \tag{37.3}$$

We did all of this with calculus and did not use the inner product at all. What does this characterization of the range mean in terms of the inner product? How does the inner product show up in the mean-zero condition $\int_0^{2\pi} g(s) \, ds = 0$?

Day 38: Friday, November 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."

Adjoint of a linear operator

The trick is multiplying by 1. Put $u_1(x) := 1$ for all x. Then

$$\int_0^{2\pi} g(x)\ dx = \int_0^{2\pi} g(x)u_1(x)\ dx = \langle g, u_1 \rangle .$$

The identity (37.3) then reads

$$\mathcal{T}(\mathcal{V}) = \{g \in \mathcal{V} \mid \langle g, u_1 \rangle = 0\} .$$

Here is a (somewhat) different way of obtaining $\langle g, u_1 \rangle = 0$ for $g \in \mathcal{T}(\mathcal{V})$. If $g = \mathcal{T}f$ for some $f \in \mathcal{V}$, then

$$\langle g, u_1 \rangle = \langle \mathcal{T}f, u_1 \rangle = \langle f', u_1 \rangle = \int_0^{2\pi} f'(x)u_1(x)\ dx.$$

Now we integrate by parts:

$$\int_0^{2\pi} f'(x)u_1(x)\ dx = f(2\pi)u_1(2\pi) - f(0)u_1(0) - \int_0^{2\pi} f(x)u_1'(x)\ dx.$$

Since f is 2π-periodic and $u_1 = 1$, the "boundary terms" vanish:

$$f(2\pi)u_1(2\pi) - f(0)u_1(0) = 0.$$

And since u_1 is constant, $u_1' = 0$, thus

$$\int_0^{2\pi} f(x)u_1'(x)\ dx = 0$$

as well. This gives $\langle g, u_1 \rangle = 0$ for all $g \in \mathcal{T}(\mathcal{V})$.

However, this calculation still relied heavily on properties of integrals, just not the ones that we used before. In the language of inner products, integration by parts and the vanishing of the boundary terms showed

$$\langle \mathcal{T}f, u_1 \rangle = \langle f, -\mathcal{T}u_1 \rangle .$$

This is actually valid for more than just u_1.

38.1 Problem (!). Let f, $g \in \mathcal{C}^1([a,b])$ with $f(a) = f(b)$ and $g(a) = g(b)$. Show that

$$\int_a^b f'(x)g(x)\ dx = -\int_a^b f(x)g'(x)\ dx.$$

In the concrete situation above, we have

$$\langle \mathcal{T}f, g\rangle = \langle f, \mathcal{S}g\rangle$$

for all f, $g \in \mathcal{V}$, where $\mathcal{S} := -\mathcal{T}$. The function u_1 satisfies $\ker(\mathcal{T}) = \operatorname{span}(u_1)$, and so also $\ker(\mathcal{S}) = \operatorname{span}(u_1)$. Then if $g \in \mathcal{T}(\mathcal{V})$ with $g = \mathcal{T}f$ for some $f \in \mathcal{V}$, we have

$$\langle g, u_1\rangle = \langle \mathcal{T}f, u_1\rangle = \langle f, \mathcal{S}u_1\rangle = 0.$$

Any $h \in \ker(\mathcal{S})$ has the form $h = \alpha u_1$ for some $\alpha \in \mathbb{F}$, thus if $g \in \mathcal{T}(\mathcal{V})$, then

$$\langle g, h\rangle = \langle g, \alpha u_1\rangle = \alpha \langle g, u_1\rangle = 0.$$

This is true for any $g \in \mathcal{T}(\mathcal{V})$ and $h \in \ker(\mathcal{S})$, and so such g and h also satisfy $g \in \ker(\mathcal{S})^{\perp}$ and $h \in \mathcal{T}(\mathcal{V})^{\perp}$. That is,

$$\mathcal{T}(\mathcal{V}) \subseteq \ker(\mathcal{S})^{\perp} \quad \text{and} \quad \ker(\mathcal{S}) \subseteq \mathcal{T}(\mathcal{V})^{\perp}.$$

Now let $h \in \mathcal{T}(\mathcal{V})^{\perp}$ and $f \in \mathcal{V}$. Then

$$0 = \langle \mathcal{T}f, h\rangle = \langle f, \mathcal{S}h\rangle,$$

and since this is true for any $f \in \mathcal{V}$, we have $\mathcal{S}h = 0_{\mathcal{V}}$. Hence $\mathcal{T}(\mathcal{V})^{\perp} \subseteq \ker(\mathcal{S})$. And so here

$$\mathcal{T}(\mathcal{V})^{\perp} = \ker(\mathcal{S}),$$

thus $\mathcal{T}(\mathcal{V})^{\perp\perp} = \ker(\mathcal{S})^{\perp}$.

This, frankly, is about as far as the inner product approach will take us, but it is pretty far. If we knew $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$, then we would have $\mathcal{T}(\mathcal{V}) = \ker(\mathcal{S})^{\perp}$. Of course, we know this already, since we worked that out with calculus.

38.2 Problem ($\star$). We also know that $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$ if we have $\mathcal{V} = \mathcal{T}(\mathcal{V}) \oplus \mathcal{T}(\mathcal{V})^{\perp}$. This turns out to be true in the concrete situation above. Start with $f \in \mathcal{V}$, so $f \in \mathcal{C}^{\infty}(\mathbb{R})$ and $f(x+2\pi) = f(x)$ for all x. We want to write $f = g' + \alpha u_1$ for some 2π-periodic $g \in \mathcal{C}^{\infty}(\mathbb{R})$ and some $c \in \mathbb{R}$. The natural idea may be to take $g(x) = \int_0^x f(s)\ ds =: (\mathcal{A}f)(x)$ and $c = 0$, but we are not guaranteed that such a g is 2π-periodic. Indeed, we have established that if $h \in \mathcal{C}(\mathbb{R})$ is 2π-periodic, then $\mathcal{A}h$ is 2π-periodic if and only if $(\mathcal{A}h)(2\pi) = 0$.

The right idea is to adjust the integrand defining g so that g is genuinely 2π-periodic and then adjust c to compensate for that new integrand. Show that taking

$$g = \mathcal{A}\left(f - \frac{(\mathcal{A}f)(2\pi)}{2\pi}\right) \quad \text{and} \quad c = \frac{(\mathcal{A}f)(2\pi)}{2\pi}$$

yields $f = \mathcal{T}g + cu_1$ with $g \in \mathcal{V}$.

Here is what we are going to show in general: under suitable hypotheses on $\mathcal{T}$ (and its domain and codomain), we can write $\mathcal{T}(\mathcal{V})^{\perp\perp} = \ker(\mathcal{S})^{\perp}$ for an operator $\mathcal{S}$. With more hypotheses on $\mathcal{T}$ (and its domain and codomain), we can get $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$, and then we have $\mathcal{T}(\mathcal{V}) = \ker(\mathcal{S})^{\perp}$. This is our long-desired geometric characterization of the range in terms of a simpler object (a kernel).

38.3 Problem (!). Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and suppose that $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is finite-rank. Prove that $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$.

38.4 Remark. *Does all of this feel excessive, gratuitous, unnecessary? It should. We do not need a single piece of linear algebra to solve the problem $f' = g$ for f and g 2π-periodic. Or to understand the solution. However, as always in this course, the point is to use familiar objects from calculus as teaching tools: what is the deeper linear algebra underlying a (hopefully) familiar and accessible problem? What does this concrete situation teach us about abstraction?*

Here is the essential abstraction. Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ and $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that

$$\langle \mathcal{T}v, w \rangle_{\mathcal{W}} = \langle v, \mathcal{S}(w) \rangle_{\mathcal{V}}, \ v \in \mathcal{V}, \ w \in \mathcal{W}. \tag{38.1}$$

Take $w \in \mathcal{T}(\mathcal{V})$, so $w = \mathcal{T}v$ for some $v \in \mathcal{V}$, and let $z \in \ker(\mathcal{S})$.

1. We have

$$\langle w, z \rangle_{\mathcal{W}} = \langle \mathcal{T}v, z \rangle_{\mathcal{W}} = \langle v, \mathcal{S}z \rangle_{\mathcal{V}} = \langle v, 0_{\mathcal{V}} \rangle_{\mathcal{V}} = 0. \tag{38.2}$$

The first equality is the definition $w = \mathcal{T}v$, the second is the fundamental identity (38.1), 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 $\mathcal{T}(\mathcal{V}) \subseteq \ker(\mathcal{S})^{\perp}$.

2. The calculation (38.2) 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 $\ker(\mathcal{S}) \subseteq \mathcal{T}(\mathcal{V})^{\perp}$.

3. 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 (38.1) 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 25.15 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 $\mathcal{T}(\mathcal{V})^{\perp} \subseteq \ker(\mathcal{S})$.

Combining the second and third results give

$$\mathcal{T}(\mathcal{V})^{\perp} = \ker(\mathcal{S}).$$

If we also know that $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$, perhaps because $\mathcal{W}$ and $\mathcal{T}(\mathcal{V})$ have the RRP, then we obtain

$$\mathcal{T}(\mathcal{V}) = \ker(\mathcal{S})^{\perp}. \tag{38.3}$$

This is a geometric characterization of $\mathcal{T}(\mathcal{V})$ in terms of a simpler object—"geometric" because (38.3) involves an orthogonal complement, simpler because $\ker(\mathcal{S})$ requires us to solve $\mathcal{S}w = 0_{\mathcal{V}}$, which is arguably simpler than trying to solve $\mathcal{T}v = w$ for every $w \in \mathcal{W}$, or at least every $w \in \mathcal{T}(\mathcal{V})$.

We now have two questions to consider. First, under what conditions on $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is there an operator $\mathcal{S} \in \mathbf{L}(\mathcal{W}, \mathcal{V})$ such that (38.1) holds? Second, what conditions on $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ guarantee $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$?

We begin by studying $\mathcal{S}$ in much more detail. The proof of the following theorem is a great exercise in the slogan "What things do defines what things are."

38.5 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*

$$\langle \mathcal{T}v, w \rangle_{\mathcal{W}} = \langle v, \mathcal{S}(w) \rangle_{\mathcal{V}}, \ v \in \mathcal{V}, \ w \in \mathcal{W}. \tag{38.4}$$

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 linear operator from $\mathcal{W}$ to $\mathcal{V}$ to satisfy* (38.1).

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) We proved this just before the statement of this theorem.

(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

$$\left\langle v, (\mathcal{S} - \widetilde{\mathcal{S}})w\right\rangle_{\mathcal{V}} = \langle v, \mathcal{S}w\rangle_{\mathcal{V}} - \left\langle v, \widetilde{\mathcal{S}}w\right\rangle_{\mathcal{V}} = \langle \mathcal{T}v, w\rangle_{\mathcal{W}} - \langle \mathcal{T}v, w\rangle_{\mathcal{W}} = 0. \qquad \blacksquare$$

38.6 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.

38.7 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.

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 38.5 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: $\mathcal{T}(\mathcal{V}) = \ker(\mathcal{T}^*)^{\perp}$.

What else can we learn about the adjoint? It turns out that viewing 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

$$\overline{z + w} = \overline{z} + \overline{w}, \qquad \overline{zw} = \overline{z}(\overline{w}), \qquad \overline{\overline{z}} = z, \qquad \text{and} \qquad \overline{z^{-1}} = (\overline{z})^{-1} \text{ if } z \neq 0.$$

The adjoint acts much the same.

38.8 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

$$\langle(\mathcal{T}^*)^*v, w\rangle_{\mathcal{W}} = \langle v, \mathcal{T}^*w\rangle_{\mathcal{W}}, \ v \in \mathcal{V}, \ w \in \mathcal{W}. \tag{38.5}$$

Does $\mathcal{T}$ do this? Certainly $\mathcal{T} \in \mathbf{L}(\mathcal{V},\mathcal{W})$, and

$$\langle\mathcal{T}v, w\rangle_{\mathcal{W}} = \langle v, \mathcal{T}^*w\rangle_{\mathcal{W}}.$$

That is, $\mathcal{T}$ fulfills exactly the role of $(\mathcal{T}^*)^*$ in (38.5), and so by uniqueness of the adjoint, $(\mathcal{T}^*)^* = \mathcal{T}$. ■

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.

Day 39: Monday, November 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."

Self-adjoint operator, skew-adjoint operator

39.1 Example. Our motivating example for adjoints involved the periodic function space $\mathcal{V} = \{f \in \mathcal{C}^\infty \mid f(x+2\pi) = f(x) \text{ for all } x \in \mathbb{R}\}$ with the inner product $\langle f, g\rangle := \int_0^{2\pi} f(s)g(s)\, ds$ and the operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{V} \colon f \mapsto f'$. There we showed that $\langle \mathcal{T}f, g\rangle = \langle f, -\mathcal{T}g\rangle$ for all f, $g \in \mathcal{V}$, so $\mathcal{T}$ has the adjoint $\mathcal{T}^* = -\mathcal{T}$.

Operators $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ with $\mathcal{T}^* = -\mathcal{T}$ also have a special name.

39.2 Definition. *Let $\mathcal{V}$ be an inner product space. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* **SKEW-ADJOINT** *if $\mathcal{T}^* = -\mathcal{T}$.*

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.

39.3 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

$$\langle \mathcal{T}v, w\rangle_{\mathcal{W}} = \sum_{j=1}^n \langle v, \langle w, \mathcal{T}u_j\rangle_{\mathcal{W}} u_j\rangle_{\mathcal{V}} = \left\langle v, \sum_{j=1}^n \langle w, \mathcal{T}u_j\rangle_{\mathcal{W}} u_j\right\rangle_{\mathcal{V}}.$$

This is the linearity of the inner product in its second slot (with respect to sums). ■

39.4 Problem (!). Does this proof remind you of the calculation in Example 34.4?

Content from *Linear Algebra* by Meckes & Meckes. This result is Theorem 5.12 on p. 312.

39.5 Example. Let $A \in \mathbb{F}^{m \times n}$, and let $\mathcal{M}_A \colon \mathbb{F}^n \to \mathbb{F}^m \colon \mathbf{v} \mapsto A\mathbf{v}$. We know what the adjoint $\mathcal{M}_A^* \colon \mathbb{F}^m \to \mathbb{F}^n$ *does*:

$$\langle \mathcal{M}_A \mathbf{v}, \mathbf{w} \rangle_m = \langle \mathbf{v}, \mathcal{M}_A^* \mathbf{w} \rangle_n ,$$

where $\langle \cdot, \cdot \rangle_p$ is the dot product on $\mathbb{F}^p$. And we know what $\mathcal{M}_A^*$ *is*:

$$\mathcal{M}_A^* \mathbf{w} = \sum_{j=1}^n \langle \mathbf{w}, \mathcal{M}_A \mathbf{e}_j \rangle_m \mathbf{e}_j = \sum_{j=1}^n \langle \mathbf{w}, A\mathbf{e}_j \rangle_m \mathbf{e}_j$$

with $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ as the standard basis for $\mathbb{F}^n$.

So what is the matrix representation $[\mathcal{M}_A^*] \in \mathbb{F}^{n \times m}$, and how is it related to A? Let $(\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_m)$ be the standard basis for $\mathbb{F}^m$. Then the jth column of $[\mathcal{M}_A^*]$ is $[\mathcal{M}_A^*]\widetilde{\mathbf{e}}_j = \mathcal{M}_A^* \widetilde{\mathbf{e}}_j \in \mathbb{F}^n$, and so the (i,j)-entry of $[\mathcal{M}_A^*]$ is $\langle \mathcal{M}_A^* \widetilde{\mathbf{e}}_j, \mathbf{e}_i \rangle_n$. We compute this more precisely as

$$\langle \mathcal{M}_A^* \widetilde{\mathbf{e}}_j, \mathbf{e}_i \rangle_n = \langle \widetilde{\mathbf{e}}_j, \mathcal{M}_A \mathbf{e}_i \rangle_m = \langle \widetilde{\mathbf{e}}_j, A\mathbf{e}_i \rangle_m = \overline{\langle A\mathbf{e}_i, \widetilde{\mathbf{e}}_j \rangle_m}.$$

This is the conjugate of the (j,i)-entry of A: the ith column of A is $A\mathbf{e}_i \in \mathbb{F}^m$, and its jth entry is $\langle A\mathbf{e}_i, \widetilde{\mathbf{e}}_j \rangle_m$.

We name the matrix that appeared in the previous example.

39.6 Definition. *Let $A \in \mathbb{F}^{m \times n}$. The* CONJUGATE TRANSPOSE *or the* ADJOINT *of A is the matrix $A^* \in \mathbb{F}^{n \times m}$ whose (i,j)-entry is the conjugate of the (j,i)-entry of A.*

So, for $A \in \mathbb{F}^{m \times n}$, the matrix representation of the adjoint of the multiplication-by-A operator is the adjoint of A:

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

39.7 Example. If

$$A := \begin{bmatrix} 1 & i \\ -2i & 3 \\ 4 & 5 \end{bmatrix},$$

then

$$A^* = \begin{bmatrix} \overline{1} & \overline{-2i} & \overline{4} \\ \overline{i} & \overline{3} & \overline{5} \end{bmatrix} = \begin{bmatrix} 1 & 2i & 4 \\ -i & 3 & 5 \end{bmatrix}.$$

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 $\mathcal{M}_{A^\intercal}$ related to $\mathcal{M}_A$ when we are working with complex matrices and vectors? This is subtle, and it involves linear functionals...

39.8 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 $(\mathcal{T}f)(x) = xf(x)$. To determine if $\mathcal{T}$ has an adjoint, we compute

$$\begin{aligned}\langle \mathcal{T}f, g\rangle = \int_0^1 (\mathcal{T}f)(x)g(x)\ dx = \int_0^1 xf(x)g(x)\ dx &= \int_0^1 f(x)(xg(x))\ dx \\ &= \int_0^1 f(x)(\mathcal{T}g)(x)\ dx = \langle f, \mathcal{T}g\rangle\,.\end{aligned}$$

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.

39.9 Definition. *Let $\mathcal{V}$ be an inner product space. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is* **SELF-ADJOINT** *if $\mathcal{T}^* = \mathcal{T}$.*

39.10 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 $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$?

39.11 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 $\langle \mathcal{T}f, g\rangle = \langle f, \mathcal{T}^*g\rangle$.

39.12 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. 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* $\langle \mathcal{T}f, g\rangle = f(2)\overline{g(1)} + \sum_{k=2}^\infty f(k+1)\overline{g(k)}$.]

39.13 Problem ($\star$). Let $\mathcal{V} = \mathcal{C}([0,1])$ with the L^2-inner product. Recall from multivariable calculus that if h is continuous on the unit square $\{(x,y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1,\ 0 \leq y \leq 1\}$, 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.

39.14 Problem ($\star$). Suppose that we remove the periodic boundary conditions from Example 39.1 and work on the simpler interval $[0,1]$. Let $\mathcal{V} = \mathcal{C}^\infty([0,1])$ with the L^2-inner product. Define $\mathcal{T} \in \mathbf{L}(\mathcal{V})$ by $\mathcal{T}f := f'$. We show that $\mathcal{T}$ does not have an adjoint. Proving this directly involves a complicated negative: for all $\mathcal{S} \in \mathbf{L}(\mathcal{V})$, there exist f, $g \in \mathcal{V}$ such that $\langle \mathcal{T}f, g\rangle \neq \langle f, \mathcal{S}g\rangle$. Instead, we proceed with contradiction: assume there is $\mathcal{S} \in \mathbf{L}(\mathcal{V})$ such that $\langle \mathcal{T}f, g\rangle = \langle f, \mathcal{S}g\rangle$ for all f, $g \in \mathcal{V}$.

(i) Integrate by parts to get

$$\langle f, \mathcal{S}g\rangle = \langle \mathcal{T}f, g\rangle = f(1)g(1) - f(0)g(0) - \langle f, \mathcal{T}g\rangle\,.$$

The absence of boundary conditions (i.e., conditions on the values of functions in $\mathcal{V}$ at $x = 0$, 1) means that we cannot easily get rid of the terms $f(1)g(1) - f(0)g(0)$ and end up with just the inner product term $\langle f, -\mathcal{T}g\rangle$. This should warn us that an adjoint of $\mathcal{T}$ may not exist.

(ii) Obtain

$$\langle f, (\mathcal{T}+\mathcal{S})g\rangle = f(1)g(1) - f(0)g(0).$$

(iii) Put $g(x) = x - 1$ to get

$$\langle f, (\mathcal{T}+\mathcal{S})g\rangle = f(0).$$

Conclude that the functional $\varphi\colon \mathcal{V} \to \mathbb{F}\colon f \mapsto f(0)$ is bounded, and obtain a contradiction from (the methodology of) Example 34.8. (That example involved the larger inner product space $\mathcal{C}([0,1])$, but the functions f_n in that example are infinitely differentiable.)

Day 40: Wednesday, November 19.

Vocabulary from today

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

Bounded linear operator

What conditions, then, can we put on inner product spaces $\mathcal{V}$ and $\mathcal{W}$ and an operator $\mathcal{T} \in \mathbf{L}(\mathcal{V},\mathcal{W})$ to ensure that $\mathcal{T}$ has an adjoint $\mathcal{S} \in \mathbf{L}(\mathcal{W},\mathcal{V})$? Consider the fundamental property of the adjoint: we want

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

for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$. Fix $w \in \mathcal{W}$; the goal is to define what $\mathcal{S}w$ should be. Consider the map

$$\varphi_w\colon \mathcal{V} \to \mathbb{F}\colon v \mapsto \langle \mathcal{T}v, w\rangle_{\mathcal{W}}.$$

This is certainly a linear functional: $\varphi_w \in \mathcal{V}'$. If φ_w is also a *bounded* linear functional ($\varphi_w \in \mathcal{V}^*$), and if $\mathcal{V}$ has the RRP, then there is a unique $r_w \in \mathcal{V}$ such that $\varphi_w = \langle \cdot, r_w\rangle_{\mathcal{V}}$. That is,

$$\langle \mathcal{T}v, w\rangle_{\mathcal{W}} = \varphi_w(v) = \langle v, r_w\rangle_{\mathcal{V}}.$$

If we define $\mathcal{S}w := r_w$, then $\mathcal{S}$ meets the fundamental property (38.4) of the adjoint.

What guarantees that φ_w is bounded? We can estimate

$$|\varphi_w(v)| = |\langle \mathcal{T}v, w\rangle_{\mathcal{W}}| \le \|\mathcal{T}v\|_{\mathcal{W}} \|w\|_{\mathcal{W}}$$

by the Cauchy–Schwarz inequality. Here

$$\|v\|_{\mathcal{V}} = \sqrt{\langle v, v\rangle} \quad \text{and} \quad \|w\|_{\mathcal{W}} = \sqrt{\langle w, w\rangle}.$$

However, we would like an estimate of the form

$$|\varphi_w(v)| \le C\|v\|.$$

The right idea is to demand that $\mathcal{T}$ is bounded in the following sense.

40.1 Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces. An operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is* BOUNDED *if there exists $C > 0$ such that*

$$\|\mathcal{T}v\|_{\mathcal{W}} \leq C\,\|v\|_{\mathcal{V}}$$

for all $v \in \mathcal{V}$. We denote the set of all bounded linear operators from $\mathcal{V}$ to $\mathcal{W}$ by $\mathbf{B}(\mathcal{V}, \mathcal{W})$. An operator is UNBOUNDED *if it is not bounded.*

In particular, $\mathcal{V}^* = \mathbf{B}(\mathcal{V}, \mathbb{F})$.

40.2 Problem ($\star$). Adapt the proof of Theorem 34.1 to prove that $\mathbf{B}(\mathcal{V}, \mathcal{W})$ is a subspace of $\mathbf{L}(\mathcal{V}, \mathcal{W})$.

40.3 Problem ($\star$). Since almost everything seems to work in finite-dimensional spaces, it probably will not be surprising to know that all operators on a finite-dimensional space are bounded. Let $\mathcal{V}$ be a finite-dimensional inner product space with orthonormal basis $(u_1, \ldots, u_n)$ and let $\mathcal{W}$ be any inner product space (perhaps infinite-dimensional). Adapt the method of (the first two parts of) Problem 34.6 to prove that any $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is bounded. [Hint: *replace φ with $\mathcal{T}$ and $|\varphi(u_j)|$ with $\|\mathcal{T}u_j\|_{\mathcal{W}}$.*]

40.4 Problem ($\star$). Not every operator with an infinite-dimensional domain is bounded. Let $\mathcal{V} = \mathcal{C}^1([0, 2\pi])$ and $\mathcal{W} = \mathcal{C}([0, 2\pi])$, both with the L^2-inner product $\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\,dx$. Define $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ by $\mathcal{T}f = f'$. Suppose that $\mathcal{T}$ is bounded and obtain a contradiction by considering the functions $f_n(x) := \sin(nx)$. [Hint: *if f is 2π-periodic and $n \geq 1$ is an integer, then $\int_0^{2n\pi} f(x)\,dx = \int_0^{2\pi} f(x)\,dx$.*]

40.5 Problem (!). A finite-dimensional codomain does not guarantee the boundedness of a linear operator. Give an example of an infinite-dimensional inner product space $\mathcal{V}$, a finite-dimensional inner product space $\mathcal{W}$, and a linear operator $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ such that $\mathcal{T}$ is not bounded. [Hint: *work with* $\dim(\mathcal{W}) = 1$.]

We combine our old ally, the RRP, with the new notion of bounded linear operator to spell out a situation that guarantees the existence of the adjoint.

40.6 Theorem. *Let $\mathcal{V}$ be an inner product space. Then $\mathcal{V}$ has the RRP if and only if for every inner product space $\mathcal{W}$, every bounded operator $\mathcal{T} \colon \mathcal{V} \to \mathcal{W}$ has an adjoint.*

Proof. ($\Longrightarrow$) This is the argument above. Since $\mathcal{T}$ is bounded, the functional φ_w defined above is bounded, so the RRP applies to produce the representing vector r_w. Since r_w is unique, the map $\mathcal{S} \colon \mathcal{W} \to \mathcal{V} \colon w \mapsto r_w$ is defined, and by construction it satisfies the fundamental property of the adjoint, which in turn implies its linearity.

($\Longleftarrow$) Take $\mathcal{W} = \mathcal{V}$. Let $\varphi \in \mathcal{V}^*$ and $v_0 \in \mathcal{V}$ with $v_0 \neq 0_{\mathcal{V}}$. Then $\mathcal{T}v := \varphi(v)v_0$ is a bounded operator on $\mathcal{V}$ and so has an adjoint $\mathcal{T}^*\colon \mathcal{V} \to \mathcal{V}$. That is, $\langle \mathcal{T}v, w\rangle = \langle v, \mathcal{T}^*w\rangle$ for all v, $w \in \mathcal{V}$. Taking $w = v_0$, we obtain

$$\varphi(v)\left\|v_0\right\|^2 = \varphi(v)\left\langle v_0, v_0\right\rangle = \left\langle \varphi(v)v_0, v_0\right\rangle = \left\langle \mathcal{T}v, v_0\right\rangle = \left\langle v, \mathcal{T}^*v_0\right\rangle,$$

thus

$$\varphi(v) = \left\langle v, \frac{\mathcal{T}^*v_0}{\left\|v_0\right\|^2}\right\rangle. \qquad \blacksquare$$

We emphasize that from this theorem only the *domain* of an operator needs to have the RRP for the operator to have an adjoint, not the codomain.

40.7 Problem ($\star$). Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and let $\mathcal{T} \in \mathbf{B}(\mathcal{V}, \mathcal{W})$ have the adjoint $\mathcal{T}^* \in \mathbf{L}(\mathcal{W}, \mathcal{V})$. Prove that $\mathcal{T}^* \in \mathbf{B}(\mathcal{W}, \mathcal{V})$. More precisely, show that if $C > 0$ satisfies $\left\|\mathcal{T}v\right\|_{\mathcal{W}} \le C\left\|v\right\|_{\mathcal{V}}$ for all $v \in \mathcal{V}$, then this same C satisfies $\left\|\mathcal{T}^*w\right\|_{\mathcal{V}} \le C\left\|w\right\|_{\mathcal{W}}$. [Hint: *if $w \in \ker(\mathcal{T}^*)$, why is there nothing really to prove? So, assume $w \notin \ker(\mathcal{T}^*)$; as is often the case with the norm induced by the inner product, it is easier to square. Rewrite $\left\|\mathcal{T}^*w\right\|^2$ as an inner product and use the Cauchy–Schwarz inequality to eke out an estimate involving $\mathcal{T}$.*]

40.8 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

$$\mathcal{T}\colon \mathcal{U} \to \mathcal{V}\colon u \mapsto \left\langle u, v_0\right\rangle v_0.$$

We consider $\mathcal{U}$ as an inner product space with the same inner product as $\mathcal{V}$.

(i) For $v \in \mathcal{V}$, put $\mathcal{S}v := \left\langle v, v_0\right\rangle v_0$ (so $\mathcal{S} \in \mathbf{L}(\mathcal{V})$ and $\mathcal{T}u = \mathcal{S}u$ for all $u \in \mathcal{U}$; note, however, that $\mathcal{S}$ and $\mathcal{T}$ have different domains, and so they are not the same operator). 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 $\mathcal{T}^*v = \left\langle v, v_0\right\rangle \mathcal{P}_{\mathcal{U}}v_0$. [Hint: $\mathcal{P}_{\mathcal{U}}^* = \mathcal{P}_{\mathcal{U}}$.]

We have most of the tools that we need to wrap up definitively our geometric characterization of the range of an operator. Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces, and suppose that $\mathcal{V}$ has the RRP. Let $\mathcal{T} \in \mathbf{B}(\mathcal{V}, \mathcal{W})$. Then $\mathcal{T}$ has an adjoint $\mathcal{T}^*\colon \mathcal{W} \to \mathcal{V}$, and so $\mathcal{T}(\mathcal{V})^{\perp\perp} = \ker(\mathcal{T}^*)$. The question remains: what guarantees $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$?

This is a hard question. Here is an answer that often turns out to be straightforward to verify in practice but whose proof in the abstract requires tools that we do not really possess.

40.9 Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and suppose that both $\mathcal{V}$ and $\mathcal{W}$ have the RRP. Let $\mathcal{T} \in \mathbf{B}(\mathcal{V}, \mathcal{W})$. The following are equivalent.*

(i) $\mathcal{T}(\mathcal{V})^{\perp\perp} = \mathcal{T}(\mathcal{V})$.

(ii) *There is $C > 0$ such that $\|v\|_{\mathcal{V}} \le C \|\mathcal{T}v\|_{\mathcal{W}}$ for all $v \in \ker(\mathcal{T})^{\perp}$.*

Specifically, the proof of this result can be achieved using some deep properties of *normed* spaces in conjunction with results about inner product spaces.

40.10 Problem (!). Suppose that $\mathcal{V}$ and $\mathcal{W}$ are inner product spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$.

(i) Suppose that there exists $C > 0$ such that $\|v\|_{\mathcal{V}} \le C \|\mathcal{T}v\|_{\mathcal{W}}$ for all $v \in \mathcal{V}$. Such an estimate is sometimes called a COERCIVE ESTIMATE for $\mathcal{T}$. Prove that $\mathcal{T}$ is injective. [Hint: *what happens if $\mathcal{T}v = 0_{\mathcal{W}}$?*]

(ii) Suppose that $\mathcal{V}$ is nonzero and $\mathcal{T}$ is not injective. Explain why there exists no $C > 0$ such that $\|v\|_{\mathcal{V}} \le C \|\mathcal{T}v\|_{\mathcal{W}}$ for all $v \in \mathcal{V}$.

40.11 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 $\mathcal{T}(\mathcal{V}) = \ker(\mathcal{T}^*)^{\perp}$, and likewise $\mathcal{T}^*(\mathcal{W}) = \ker(\mathcal{T})^{\perp}$. 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.

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 $\mathcal{P}_{\mathcal{U}}$ if we know its matrix representation $[\mathcal{P}_{\mathcal{U}}]$, and it would be natural if we could express $[\mathcal{P}_{\mathcal{U}}]$ in terms of A.

Let $\mathbf{v} \in \mathbb{F}^m$. Of course, $\mathbf{v} - \mathcal{P}_{\mathcal{U}}\mathbf{v} \in \mathcal{U}^{\perp}$, so

$$\langle \mathbf{v} - \mathcal{P}_{\mathcal{U}}\mathbf{v}, \mathbf{a}_j \rangle_m = 0$$

for all j. Here $\langle \cdot, \cdot \rangle_p$ is the dot product in $\mathbb{F}^p$.

Now we introduce A. We have $\mathcal{P}_{\mathcal{U}}\mathbf{v} \in \mathbf{C}(A)$, so $\mathcal{P}_{\mathcal{U}}\mathbf{v} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{F}^n$. And also $\mathbf{a}_j = A\mathbf{e}_j$ with $\mathbf{e}_j$ as the jth standard basis vector in $\mathbb{F}^n$. Thus

$$0 = \langle \mathbf{v} - \mathcal{P}_{\mathcal{U}}\mathbf{v}, \mathbf{a}_j \rangle_m = \langle \mathbf{v} - A\mathbf{x}, A\mathbf{e}_j \rangle_m = \langle A^*(\mathbf{v} - A\mathbf{x}), \mathbf{e}_j \rangle_n .$$

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.

40.12 Lemma. *Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces. If $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$ is injective and has an adjoint $\mathcal{T}^*$, then $\mathcal{T}^*\mathcal{T} \in \mathbf{L}(\mathcal{V})$ is also injective.*

Proof. We want to show that if $\mathcal{T}^*\mathcal{T}v = 0_{\mathcal{V}}$, then $v = 0_{\mathcal{V}}$, and we know that if $\mathcal{T}v = 0_{\mathcal{V}}$, then $v = 0_{\mathcal{V}}$. So, it would be enough to show that if $\mathcal{T}^*\mathcal{T}v = 0_{\mathcal{V}}$, then $\mathcal{T}v = 0_{\mathcal{V}}$ (i.e., that $\ker(\mathcal{T}^*\mathcal{T}) \subseteq \ker(\mathcal{T})$). We might try to relate all this to inner products by assuming $\mathcal{T}^*\mathcal{T}v = 0_{\mathcal{V}}$ and 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 (possibly inspired by a solution to Problem 40.7), 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.

40.13 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 $[\mathcal{P}_{\mathsf{C}(A)}] = A(A^*A)^{-1}A^*$.*

Content from *Linear Algebra* by Meckes & Meckes. This is Proposition 4.18 on p. 257. Read the least squares examples on pp. 259–260.

Day 41: Friday, November 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."

Norm on a vector space, equivalent norms

We have been using the norm induced by the inner product almost as much as the inner product itself, and there is more that we can learn about it. Let $\mathcal{V}$ be an inner product space. The great *utility* of the norm $\|v\| = \sqrt{\langle v, v \rangle}$ induced by the inner product is that it allows us to quantify approximations and gives rigorous meaning to the "closest" element in a subspace $\mathcal{U}$ of $\mathcal{V}$ to some $v \in \mathcal{V}$. The great *comfort* of this norm is that it respects the fundamental properties that we associate with a measurement of length: nonnegativity, which is what lengths should be ($\|v\| \geq 0$); definiteness, which says that the only object of length 0 is the zero object ($\|v\| = 0$ if and only if $v = 0_{\mathcal{V}}$); homogeneity, which says that scaling an object scales its length ($\|\alpha v\| = |\alpha| \, \|v\|$); and the triangle inequality, which says that going from point A to point C directly is shorter than stopping at point B along the way ($\|v + w\| \leq \|v\| + \|w\|$).

Norms induced by inner products are not the only valid, meaningful way of defining length or size of vectors. Even on inner product spaces, we can define maps that retain the above properties of the norm induced by the inner product, and yet which *cannot* be the norm induced by *any* inner product.

41.1 Example. Probably the most natural way of measuring the length of a vector $\mathbf{v} \in \mathbb{R}^2$ is its Euclidean norm, which we write as

$$\|\mathbf{v}\|_2 = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2}.$$

This gives the length of $\mathbf{v}$ as a line segment from $(0,0)$ to (v_1, v_2) in $\mathbb{R}^2$.

But there are other meaningful ways of measuring the "length" of $\mathbf{v}$. Suppose that we are constrained in the plane to move only horizontally or vertically (say, as we would be on any reasonably designed grid of city streets). Then

$$\|\mathbf{v}\|_1 := |v_1| + |v_2|$$

also measures the length of $\mathbf{v}$ if we think of length as "distance from $(0,0)$ to (v_1, v_2)."

Continuing the notion that we move only horizontally or vertically, perhaps we are interested in which direction is longer. Then

$$\|\mathbf{v}\|_\infty := \max\{|v_1|, |v_2|\}$$

captures that longer direction.

More generally, for $1 \leq p < \infty$, we could put

$$\|\mathbf{v}\|_p := \left(|v_1|^p + |v_2|^p\right)^{1/p},$$

and then the following are true (though none are quite trivial to prove).

(i) $\|\cdot\|_p : \mathbb{R}^2 \to \mathbb{R}$ satisfies the same properties of nonnegativity, definiteness, homogeneity, and the triangle inequality as does the norm induced by an inner product.

(ii) $\lim_{p \to \infty} \|\mathbf{v}\|_p = \|\mathbf{v}\|_\infty$.

(iii) The triangle inequality fails for $\|\cdot\|_p$ when $0 < p < 1$, so we will not consider that. (Also, why does taking $p = 0$ give meaningless length?)

(iv) There exists *no* inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^2$ such that $\|\mathbf{v}\|_p = \sqrt{\langle\mathbf{v},\mathbf{v}\rangle}$ for *any* p except $p = 2$.

We sometimes call these maps the ℓ^p-norms ("little ell-p norms"), although these are defined on $\mathbb{R}^2$, not the sequence (sub)space ℓ^2. Of course, we could also extend these maps to $\mathbb{F}^n$ just by keeping track of more components in the sums or maxima (note that each ℓ^p-norm is defined on each space $\mathbb{F}^n$), and we could also define analogues on subspaces of $\mathbb{F}^\infty$, e.g., putting $\ell^p := \{(a_k) \in \mathbb{R}^\infty \mid \sum_{k=1}^\infty |a_k|^p < \infty\}$. (Verifying that this is a subspace for $p \neq 1,\ 2,\ \infty$ is tricky.)

41.2 Example. For $f \in \mathcal{C}([0,1])$ and $1 \le p < \infty$, we could put

$$\|f\|_p = \left(\int_0^1 |f(x)|^p\ dx\right)^{1/p}$$

and

$$\|f\|_\infty := \max_{0 \le x \le 1} |f(x)|$$

to obtain the four same results as the previous example. For historical and cultural reasons, we call these the L^p-norms (pronounced "big ell-p" if we need to say this at the same time as ℓ^p). Hopefully the L^∞-norm feels like a valid measurement of function size—it is, after all, the largest value that the function takes—and perhaps the L^1-norm feels even more natural than the L^2-norm for quantifying size. The L^1-norm $\|f\|_1 = \int_0^1 |f(x)|\ dx$ just captures the total area "under" the graph of f, whereas the area interpretation of the L^2-norm needs to be considered with the effect of squaring the integrand (which makes "small" values smaller and large values "larger").

Of course, there is nothing special about the underlying interval $[0,1]$ here, as has been the case in most of our examples with $\mathcal{C}([0,1])$. We could replace $[0,1]$ with any closed, bounded subinterval $[a,b] \subseteq \mathbb{R}$, and everything above would be true. More challenging is moving to a possibly open and/or unbounded interval $I \subseteq \mathbb{R}$, in which case we would have to work with improper integrals for $1 \le p < \infty$ and a *supremum*, not a maximum, for $p = \infty$.

Here is what the ℓ^p- and L^p-norms have in common with the norm induced by an inner product.

41.3 Definition. *Let $\mathcal{V}$ be a vector space. A* NORM *on $\mathcal{V}$ is a map $\|\cdot\| : \mathcal{V} \to \mathbb{R}$ such that the following hold.*

(i) [Nonnegativity] $\|v\| \ge 0$ *for all* $v \in \mathcal{V}$.

(ii) [Definiteness] $\|v\| = 0$ *if and only if* $v = 0$.

(iii) [Homogeneity] $\|\alpha v\| = |\alpha|\ \|v\|$ *for all* $\alpha \in \mathbb{F}$ *and* $v \in \mathcal{V}$.

(iv) [Triangle inequality] $\|v+w\| \le \|v\| + \|w\|$ *for all* v, $w \in \mathcal{V}$.

A NORMED SPACE *is a vector space on which a norm is defined; strictly speaking (recall the similar language in Definition 25.8), we might declare a normed space to be an ordered list* $(\mathcal{V}, \mathbb{F}, +, \cdot, \|\cdot\|)$, *where* $(\mathcal{V}, \mathbb{F}, +, \cdot)$ *is a vector space over* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and* $\|\cdot\|$ *is a norm on* $\mathcal{V}$.

41.4 Example. Let $1 \le p \le \infty$. Then the ℓ^p-norm is a norm on $\mathbb{F}^n$ and the L^p-norm is a norm on $\mathcal{C}([0,1])$. For $p = 2$, this follows because these norms are induced by the Euclidean or L^2-inner product. For $p = 1$, ∞, this is mostly straightforward by familiar properties of absolute value and integrals; perhaps the hardest step is proving definiteness for the L^1-norm, but that uses the strategy for the L^2-inner product in Example 25.7. For $1 < p < \infty$, proving the triangle inequality is, at the least, involved, and it relies on classical results from analysis called Hölder's inequality and Minkowski's inequality. In practice, the $p = 1$, 2, and ∞ cases arise most frequently.

Content from *Linear Algebra* by Meckes & Meckes. Norms are defined on p. 267. Do Quick Exercise #14 on that page (so the if and only if in our definition is not really necessary). Read the examples on pp. 267–268.

We should not need further convincing that the norm induced by an inner product is special. And yet there is still more to learn: not every norm on a vector space can be induced by an inner product. The following result first gives a necessary condition for a norm to be induced by an inner product: the norm must satisfy an identity called the PARALLELOGRAM IDENTITY. A norm that fails to satisfy this identity cannot be induced by an inner product. Then we state two formulas for inner products strictly in terms of the norms that they induce. These two results show that the "diagonal" values $\langle v, v\rangle$ $(= \|v\|^2)$ completely determine the other values ($\langle v, w\rangle$, $v \ne w$) of the inner product.

41.5 Theorem. *Let* $\mathcal{V}$ *be an inner product space and* v, $w \in \mathcal{V}$.

(i) [PARALLELOGRAM IDENTITY] $\dfrac{\|v+w\|^2 + \|v-w\|^2}{2} = \|v\|^2 + \|w\|^2$.

(ii) *If* $\mathcal{V}$ *is a real inner product space, then*

$$\langle v, w\rangle = \frac{\|v+w\|^2 - \|v-w\|^2}{4}.$$

(iii) *If* $\mathcal{V}$ *is a complex inner product space, then*

$$\langle v, w\rangle = \frac{\|v+w\|^2 - \|v-w\|^2 + i\|v+iw\|^2 - i\|v-iw\|^2}{4}.$$

41.6 Problem ($\star$). Prove it.

In fact, if a norm satisfies the parallelogram identity, then the two subsequent formulas can be taken as the *definition* of the inner product that induces the norm (and there can only be one such inner product by the "diagonal determination" remark above). Proving that these formulas do define inner products is a nontrivial endeavor.

Content from *Linear Algebra* by Meckes & Meckes. Figure 4.9 on p. 269 illustrates the parallelogram identity.

We have seen throughout the course that results for finite-dimensional spaces can be particularly transparent and straightforward. . .relatively speaking, of course. (Now is a good time to pause and enumerate some reasons why.) So, we might ask what is special about a *finite-dimensional* normed space. It turns out that all possible norms on a finite-dimensional space talk to each other very nicely.

41.7 Definition. *Let $\mathcal{V}$ be a vector space. The norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on $\mathcal{V}$ are* EQUIVALENT *if there exist constants C_1, $C_2 > 0$ such that*

$$\|v\|_1 \le C_1 \|v\|_2 \qquad \text{and} \qquad \|v\|_2 \le C_2 \|v\|_1 \tag{41.1}$$

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

Combining the inequalities in (41.1) yields

$$\frac{1}{C_1} \|v\|_1 \le \|v\|_2 \le C_2 \|v\|_1$$

for all $v \in \mathcal{V}$, and so the values of $\|\cdot\|_2$ are "trapped" (one might say "squozen") between two multiples of the values of $\|\cdot\|_1$. Essentially, if we know everything about one norm, then, up to some global constants, we know effectively everything about any other norm. Proving this precisely requires a lemma.

Norms measure length and also distance: if $\|v\|$ is the length of v, i.e., the distance from $0_{\mathcal{V}}$ to v, then $\|v - w\|$ should be the distance between v and w. We think that v and w are close if $\|v - w\|$ is small; this idea appeared in least squares solutions. It unsurprisingly turns out that if v and w are close, then their lengths $\|v\|$ and $\|w\|$ are also close. That is, if $\|v - w\|$ is small, then $\big|\, \|v\| - \|w\| \,\big|$ is also small. In the language of analysis, the norm is *continuous*.

41.8 Lemma (Reverse triangle inequality). *Let $\mathcal{V}$ be a normed space and v, $w \in \mathcal{V}$. Then*

$$\big|\, \|v\| - \|w\| \,\big| \le \|v - w\| . \tag{41.2}$$

Proof. The trick is to add zero:

$$\|v\| = \|v - w + w\| = \|(v - w) + w\| \le \|v - w\| + \|w\| ,$$

so

$$\|v\| - \|w\| \leq \|v - w\|. \tag{41.3}$$

And likewise

$$\|w\| = \|w - v + v\| = \|(w - v) + v\| \leq \|w - v\| + \|v\| = \|v - w\| + \|v\|,$$

so

$$\|w\| - \|v\| \leq \|v - w\|. \tag{41.4}$$

Since $|\,\|v\| - \|w\|\,|$ is either $\|v\| - \|w\|$ or $\|w\| - \|v\|$, the two estimates (41.3) and (41.4) prove (41.2). ■

Content from *Linear Algebra* by Meckes & Meckes. This is Proposition 4.21 on p. 269.

41.9 Theorem. *Let $\mathcal{V}$ be a finite-dimensional vector space. Any two norms on $\mathcal{V}$ are equivalent.*

Proof. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on $\mathcal{V}$. We find a constant $C_1 > 0$ such that $\|v\|_1 \leq C_1 \|v\|_2$ for all $v \in \mathcal{V}$. Since our method will not care which norm we label as which, it will also produce the second inequality in (41.1).

1. *The proof for $\mathcal{V} = \mathbb{F}^n$.* We effectively show that these norms are equivalent to the very familiar and tractable ℓ^2-norm. For $\mathbf{v} = \sum_{j=1}^n \alpha_j \mathbf{e}_j \in \mathbb{F}^n$, put

$$\|\mathbf{v}\|_{\ell^2} := \left(\sum_{j=1}^n |\alpha_j|^2\right)^{1/2}.$$

That is, $\|\cdot\|_{\ell^2}$ is just the usual ℓ^2-norm on $\mathbb{F}^n$, i.e., the norm induced by the dot product. (Do not confuse $\|\cdot\|_{\ell^2}$ with the given norm $\|\cdot\|_2$.) Then

$$\|\mathbf{v}\|_1 = \left\|\sum_{j=1}^n \alpha_j \mathbf{e}_j\right\| \leq \sum_{j=1}^n |\alpha_j| \|\mathbf{e}_j\|_1.$$

By the way, we have no guarantee that $\|\mathbf{e}_j\|_1 = 1$ for any j here.

The Cauchy–Schwarz inequality implies that

$$\sum_{j=1}^n |\alpha_j| \|\mathbf{e}_j\|_1 \leq \left(\sum_{j=1}^n |\alpha_j|^2\right)^{1/2} \left(\sum_{j=1}^n \|\mathbf{e}_j\|_1^2\right).$$

Put $C := (\sum_{j=1}^n \|\mathbf{e}_j\|_1^2)^{1/2}$ to obtain

$$\|\mathbf{v}\|_1 \leq C \|\mathbf{v}\|_{\ell^2}.$$

If we can find $\widetilde{C} > 0$ such that

$$\|\mathbf{v}\|_{\ell^2} \leq \widetilde{C}\,\|\mathbf{v}\|_2 \tag{41.5}$$

for all $\mathbf{v}$, we will be done.

It suffices to show that

$$1 \leq \widetilde{C}\,\|\mathbf{v}\|_2 \tag{41.6}$$

for all $\mathbf{v} \in \mathbb{F}^n$ with $\|\mathbf{v}\|_{\ell^2} = 1$. For if we know this, then we can take an arbitrary $\mathbf{v} \in \mathbb{F}^n \setminus \{\mathbf{0}_n\}$ (the case $\mathbf{v} = \mathbf{0}_n$ holds with equality) to find

$$1 \leq \widetilde{C}\left\|\frac{\mathbf{v}}{\|\mathbf{v}\|_{\ell^2}}\right\|_2,$$

and this yields the desired inequality (41.5).

Here we need to appeal to a result from multivariable calculus. Put

$$\mathcal{D} := \{\mathbf{v} \in \mathbb{F}^n \mid \|\mathbf{v}\|_{\ell^2} = 1\}.$$

Then $\mathcal{D}$ is closed and bounded. By the reverse triangle inequality, the function

$$f\colon \mathcal{D} \to \mathbb{R}\colon \mathbf{v} \mapsto \|\mathbf{v}\|_2$$

is continuous on $\mathcal{D}$ and therefore has a minimum on $\mathcal{D}$ by the extreme value theorem: there exists $\mathbf{v}_0 \in \mathcal{D}$ such that

$$\|\mathbf{v}_0\|_2 = f(\mathbf{v}_0) \leq f(\mathbf{v}) = \|\mathbf{v}\|_2$$

for all $\mathbf{v} \in \mathcal{D}$. In particular, since $\mathbf{v}_0 \in \mathcal{D}$, $\mathbf{v}_0 \neq \mathbf{0}_n$, thus $\|\mathbf{v}_0\|_2 > 0$. Take $\widetilde{C} := 1/\|\mathbf{v}_0\|_2$ to conclude (41.6).

2. *The proof for arbitrary $\mathcal{V}$.* Let $(v_1, \ldots, v_n)$ be a basis for $\mathcal{V}$. Define

$$N_1\colon \mathbb{F}^n \to \mathbb{R}\colon (\alpha_1, \ldots, \alpha_n) \mapsto \left\|\sum_{j=1}^n \alpha_j v_j\right\|_1$$

and

$$N_2\colon \mathbb{F}^n \to \mathbb{R}\colon (\alpha_1, \ldots, \alpha_n) \mapsto \left\|\sum_{j=1}^n \alpha_j v_j\right\|_2.$$

Then N_1 and N_2 are norms on $\mathbb{F}^n$; this is mostly straightforward to verify, and definiteness is a consequence of independence. So, there is $C_1 > 0$ such that

$$N_1(\alpha_1, \ldots, \alpha_n) \leq C_1 N_2(\alpha_1, \ldots, \alpha_n)$$

for all $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$.

Any $v \in \mathcal{V}$ has the unique form $v = \sum_{j=1}^n \alpha_j v_j$, and so

$$\|v\|_1 = \left\|\sum_{j=1}^n \alpha_j v_j\right\|_1 = N_1(\alpha_1, \ldots, \alpha_n) \leq C_1 N_2(\alpha_1, \ldots, \alpha_n) = \left\|\sum_{j=1}^n \alpha_j v_j\right\|_2 = \|v\|_2. \qquad \blacksquare$$

41.10 Problem ($\star$). Abbreviate $\|\cdot\|_p := \|\cdot\|_{\ell^p}$ on $\mathbb{F}^n$. For $p = 1,\ 2,\ \infty$ and $q = 1,\ 2,\ \infty$, find constants $C_{p,q} > 0$ such that

$$\|\mathbf{v}\|_p \le C_{p,q} \|\mathbf{v}\|_q$$

for all $\mathbf{v} \in \mathbb{F}^n$ when $p \neq q$. Prove that your constants are "optimal" by finding $\mathbf{v}_{p,q} \in \mathbb{F}^n$ such that

$$\|\mathbf{v}_{p,q}\|_p = C_{p,q} \|\mathbf{v}_{p,q}\|_q .$$

[Hint: *you need to find six constants here; for example, you do not need to find* $C_{1,1}$ *such that* $\|\mathbf{v}\|_1 \le C_{1,1} \|\mathbf{v}\|_1$ *for all* $\mathbf{v} \in \mathbb{F}^n$. *For optimality, try working with the standard basis vectors and the vector* $(1, \ldots, 1)$ *whose entries are all* 1. *Save the case* $p = 2$ *for last; square both sides and use some prior estimates relating* $\|\cdot\|_\infty$ *and* $\|\cdot\|_1$.]

41.11 Problem ($+$). Redo Problem 41.10 for arbitrary p and q satisfying $1 \le p \le \infty$, $1 \le q \le \infty$, and $p \neq q$.

Norms on infinite-dimensional spaces may not be equivalent.

41.12 Problem ($\star$). Show that one, but not both, of the inequalities in (41.1) hold for the L^2- and L^∞-norms on $\mathcal{C}([0,1])$. [Hint: *one inequality holds because integrals are increasing: if* $f(x) \le g(x)$ *for all* $0 \le x \le 1$, *then* $\int_0^1 f(x)\ dx \le \int_0^1 g(x)\ dx$. *To show that the remaining inequality is false, assume that it is true and use the information in Example 34.8 to get a contradiction.*]