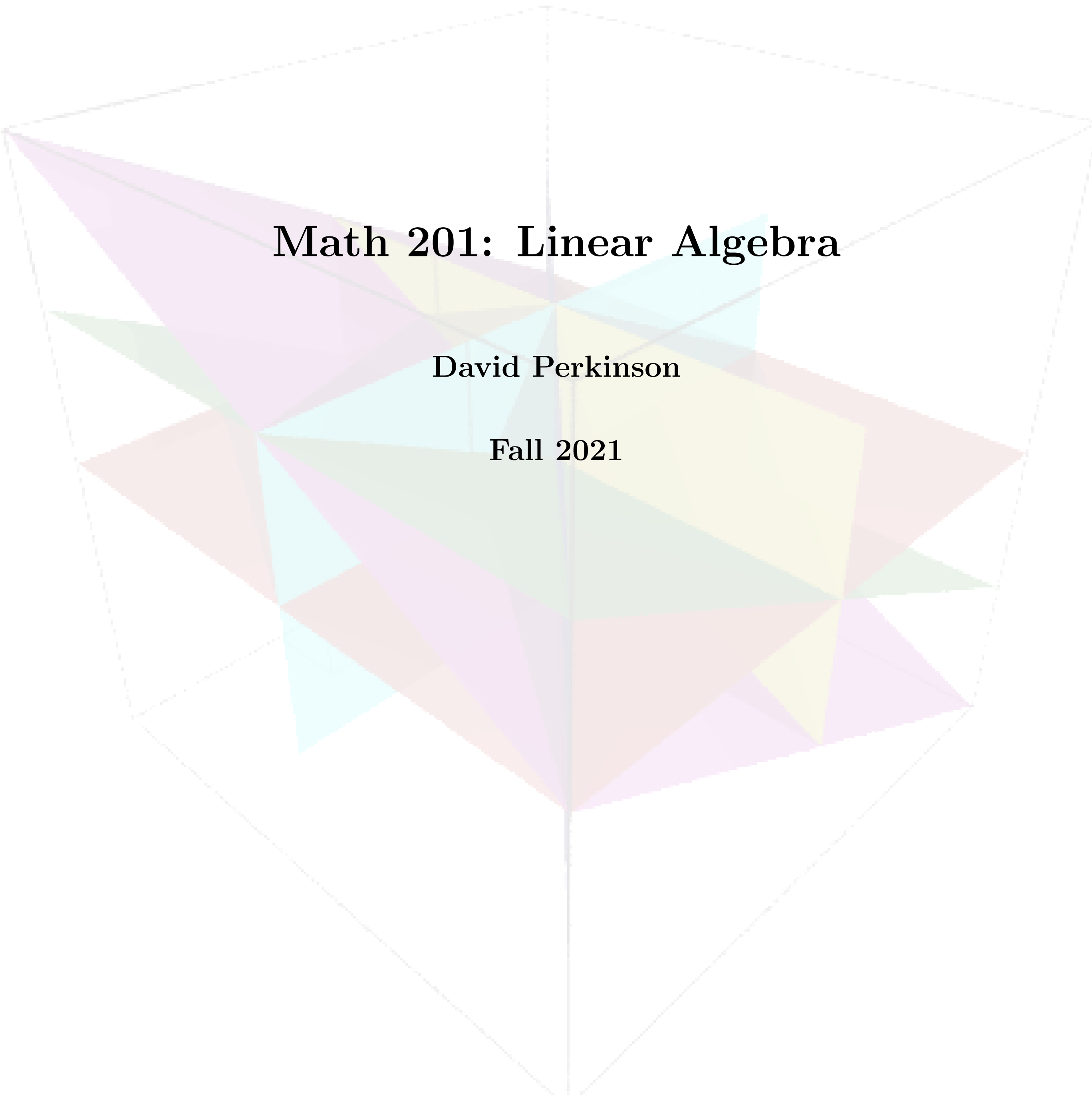


Math 201: Linear Algebra

David Perkinson

Fall 2021

Contents

Monday, Week 1: Solving systems of linear equations

First goal: solve systems of linear equations using Gaussian elimination.

Note: In all of the examples today, we will work over the real numbers.

Example 1. Solve the following system of two linear equations:

$$\begin{aligned} 3x + 2y &= 5 \\ 2x - y &= 1. \end{aligned}$$

SOLUTION: We find a solution by eliminating variables. To get rid of y, multiply the second equation through by 2 (which does not change the set of solutions), and add the equations:

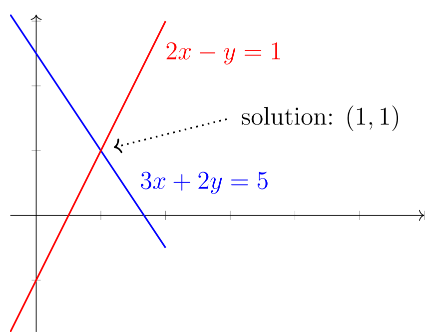
$$\begin{array}{ll} 3x + 2y & = 5 \\ 4x - 2y & = 2 \\ \hline 7x & = 7 \end{array} \quad \Rightarrow \quad x = 1.$$

Now substitute $x = 1$ back into either of the equations, and solve for y:

$$x = 1 \text{ and } 3x + 2y = 5 \quad \Rightarrow \quad y = 1.$$

So there is a unique solution: $x = y = 1$. $\square$

Here is the geometric picture:



Remark. Note that the line given by $3x+2y=5$ is perpendicular to the vector $(3,2)$ and the line given by $2x-y=1$ is perpendicular to $(2,-1)$. Do you see the relationship between these vectors and the coefficients of the equations? We will get back to this latter in the course.

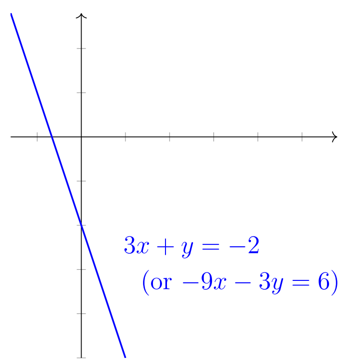
Example 2. System:

$$\begin{aligned} -9x - 3y &= 6 \\ 3x + y &= -2. \end{aligned}$$

Since the first equation is a scalar multiple of the second, they have the same solution set. In this case, the solution set is infinite:

$$\{(x,y) : y = -3x - 2\}.$$

Geometry:



Example 3. System:

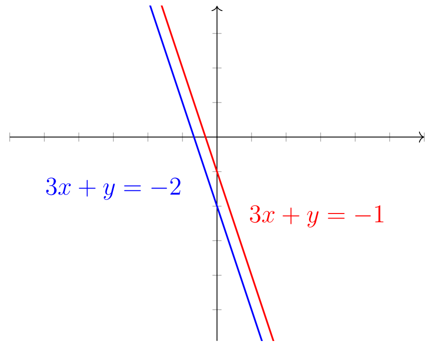
$$\begin{aligned} -9x - 3y &= 6 \\ 3x + y &= -1. \end{aligned}$$

Dividing through by -3, we see that the set of solutions to the first equation is the same as the set of solutions to the equation

$$3x + y = -2.$$

It is clear, though, that if (x,y) satisfies $3x+y=-2$, it cannot also satisfy the second equation in the system, $3x+y=1$. So the two equations in the system are incompatible, and the solution set is empty.

The lines defined by the two equations are parallel:



Example 4. System:

$$\begin{aligned} x + 2y + z &= 0 \\ x \phantom{{}+2y} + z &= 4 \\ x + y + 2z &= 1. \end{aligned}$$

We will use this example to illustrate the general method (called *Gaussian elimination.*

General idea: replace the set of equations with an equivalent set of equations (i.e., having the same solutions set) but from which the set of solutions is evident. We find equivalent sets of equations by using the following:

Row operations.

(a) multiply an equation by a nonzero scalar

(b) swap two equations

(c) add a multiple of one equation to another.

The reader should stop now and convince themselves that the solution set is invariant under these operations.

The good news is that these operations are all we need to solve any system of linear equations. We will illustrate with the system of three equations displayed above. We first introduce a convenient way of notating our system:

$$\begin{aligned} x + 2y + z &= 0 \\ x \phantom{{}+2y} + z &= 4 \\ x + y + 2z &= 1. \end{aligned} \quad \rightsquigarrow \quad \left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 1 & 0 & 1 & 4 \\ 1 & 1 & 2 & 1 \end{array}\right).$$

The matrix on the right is called the *augmented matrix* for our system of linear equations. At any point in the following string of calculations, one could convert back from an augmented matrix to its corresponding system of linear equations, and that system would be equivalent to our orignal system.

In the following calculation, r_i denotes the i-th row of the augmented matrix, and we introducing some notation for describing the row operations leading from one matrix to the next.

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 1 & 0 & 1 & 4 \\ 1 & 1 & 2 & 1 \end{array}\right) \xrightarrow[r_3 \to r_3 - r_1]{r_2 \to r_2 - r_1} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & -2 & 0 & 4 \\ 0 & -1 & 1 & 1 \end{array}\right) \xrightarrow{r_2 \to -\frac{1}{2} r_2} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & -1 & 1 & 1 \end{array}\right) \xrightarrow{r_3 \to r_3 + r_2}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right) \xrightarrow{r_1 \to -2r_2 + r_1} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right) \xrightarrow{r_1 \to -r_3 + r_1} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right).$$

The last augmented matrix is in *reduced echelon form.* We will define this term carefully later. Translate the lasted augmented matrix back into a system of equations to get a system that is equivalent to the original system, but from which the set of solutions is evident:

$$\begin{aligned} x &= 5 \\ y &= -2 \\ z &= -1. \end{aligned}$$

So there is a unique solution in this case. Now for the most **important step**: check your solution works for the original system:

$$\begin{aligned} 5 + 2(-2) + (-1) &= 0 \\ 5 \qquad\quad + (-1) &= 4 \\ 5 + (-2) + 2(-1) &= 1. \end{aligned}$$

That works. (We will suppress checking solutions throughout the result of this lecture, but in practice, you should always check your solutions.)

Example 5. The following system is a slight modification of the previous one:

$$\begin{aligned} x + 2y + z &= 0 \\ x \qquad + z &= 4 \\ x + \ y + z &= 1. \end{aligned}$$

Converting to the corresponding augmented matrix and performing a sequence of row operations similar to those in the previous example gives

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 1 & 0 & 1 & 4 \\ 1 & 1 & 1 & 1 \end{array}\right) \quad \rightsquigarrow \quad \left(\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & -1 \end{array}\right).$$

Converting back to equations gives the equivalent system:

$$\begin{aligned} x + z &= 4 \\ y &= -2 \\ 0 &= -1 \end{aligned}$$

which clearly has no solutions. Thus, our original system has no solutions.

Example 6. System:

$$\begin{aligned} x + 2y + z &= 0 \\ x \qquad + z &= 4 \\ x + \ \ y + z &= 2. \end{aligned}$$

Converting to the corresponding augmented matrix and performing a sequence of row operations similar to those in the previous example gives

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 1 & 0 & 1 & 4 \\ 1 & 1 & 1 & 2 \end{array}\right) \quad \rightsquigarrow \quad \left(\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Equivalent system:

$$\begin{aligned} x + z &= 4 \\ y &= -2 \\ 0 &= 0 \end{aligned}$$

We now get an infinite set of solutions:

$$\{(x, y, z) : x + z = 4 \text{ and } y = -2\} = \{(x, -2, 4 - x) : x \in \mathbb{R}\}.$$

This is a line in 3-space.

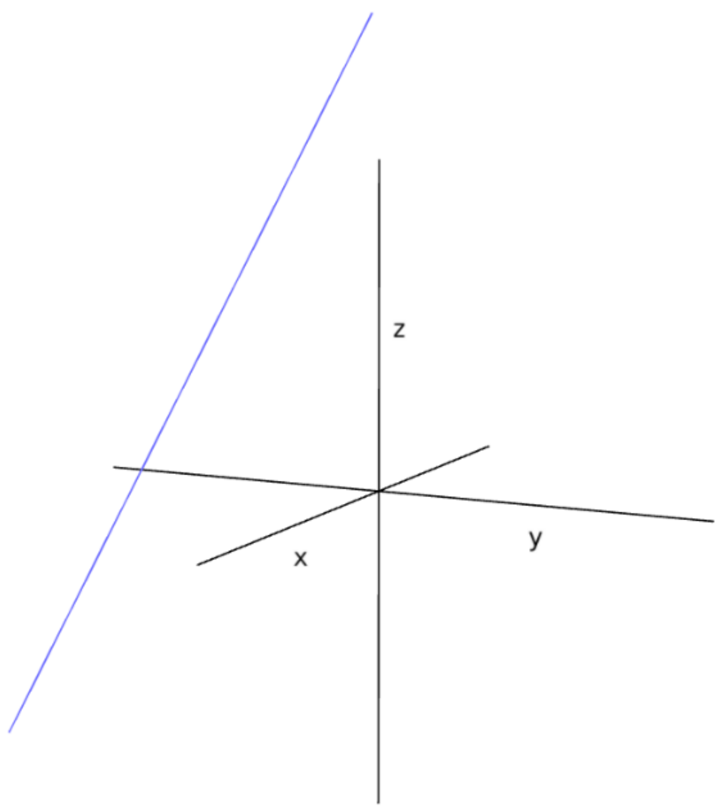


Figure 1.1: The line $\{(x, -2, 4-x) : x \in \mathbb{R}\}$.

Wednesday, Week 1: Reduced row echelon form

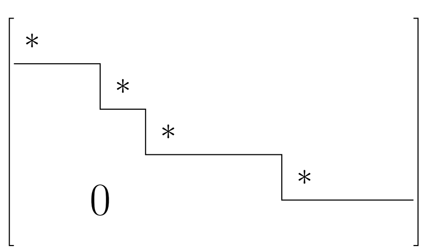
Goals for today:

- Discuss the computation of the reduced row echelon form of a system (or matrix).
- If there is an infinite number of solutions to a system, know how to describe the solution set in two ways (which we will call "parametric form" and "vector form").
- Vocabulary: reduced row echelon form, pivot variables (or pivot columns or pivot entries), free variables.

Procedure:

(a) Convert the linear system into an augmented matrix.
(b) Compute the reduced row echelon form of the matrix.
(c) Convert the reduced matrix back into a system of equations.
(d) Solve for the pivot variables.
(e) Express your solution in one of two forms, as described in the examples below.

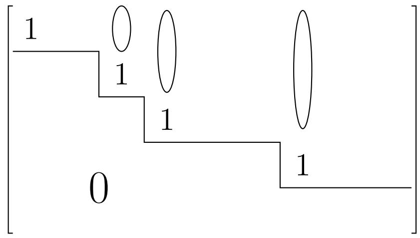
Echelon forms See our text for the precise definitions of *echelon form* (Chapter I, Definition 1.10) and *reduced row echelon form* (Chapter III, Definition 1.3). The general structure of a matrix in row echelon form is:



$*$ = nonzero *pivot* entry

zeros below "staircase"

The general structure of a matrix in *reduced* row echelon form is:



pivot entries = 1

zeros below "staircase"
and above pivot entries

Example 1. Find the solutions to the following linear system over the real numbers:

$$\begin{aligned} 2x_3 + 6x_4 \qquad &= 0 \\ x_1 + 2x_2 + x_3 + 3x_4 \qquad &= 1 \\ 2x_1 + 4x_2 + 3x_3 + 9x_4 + x_5 &= 5. \end{aligned}$$

Solution: The associated augmented matrix is

$$\left(\begin{array}{ccccc|c} 0 & 0 & 2 & 6 & 0 & 0 \\ 1 & 2 & 1 & 3 & 0 & 1 \\ 2 & 4 & 3 & 9 & 1 & 5 \end{array}\right).$$

We now use row operations to compute the reduced row echelon form of this matrix:

$$\left(\begin{array}{ccccc|c} 0 & 0 & 2 & 6 & 0 & 0 \\ 1 & 2 & 1 & 3 & 0 & 1 \\ 2 & 4 & 3 & 9 & 1 & 5 \end{array}\right) \xrightarrow{r_1 \leftrightarrow r_2} \left(\begin{array}{ccccc|c} 1 & 2 & 1 & 3 & 0 & 1 \\ 0 & 0 & 2 & 6 & 0 & 0 \\ 2 & 4 & 3 & 9 & 1 & 5 \end{array}\right) \xrightarrow{r_3 \to r_3 - 2r_1}$$

$$\left(\begin{array}{ccccc|c} 1 & 2 & 1 & 3 & 0 & 1 \\ 0 & 0 & 2 & 6 & 0 & 0 \\ 0 & 0 & 1 & 3 & 1 & 3 \end{array}\right) \xrightarrow{r_2 \to \frac{1}{2} r_2} \left(\begin{array}{ccccc|c} 1 & 2 & 1 & 3 & 0 & 1 \\ 0 & 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 3 & 1 & 3 \end{array}\right) \xrightarrow[r_3 \to r_3 - r_2]{r_1 \to r_1 - r_2}$$

$$\left(\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{array}\right).$$

The final matrix is the *reduced row echelon form* of the original matrix. (See our text for the precise definition.) The first, third, and fifth column are the *pivot columns*, and the leading 1s in these column are called *pivots*.

The system of linear equations represented by the reduced echelon form has the same set of solutions as the original system and is in a much simpler form:

$$\begin{aligned} x_1 + 2x_2 \qquad\qquad &= 1 \\ x_3 + 3x_4 \quad &= 0 \\ x_5 &= 3. \end{aligned}$$

The variables corresponding to the pivots—in this case, x_1, x_3, and x_5—are called the *pivot variables*. The others—in this case, x_2 and x_4—are the *free variables*. The free variables can take on any values, and once we assign values to the free variables, they determine the values of the pivot variables. (Since we have two free variables, the solution set is *two-dimensional* in a sense that will be precisely defined later in the course.) The solution set is

$$\left\{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 : x_1 + 2x_2 = 1, x_3 + 3x_4 = 0, x_5 = 3\right\}.$$

Solving for the pivot variables, we have

$$\begin{aligned} x_1 &= 1 - 2x_2 \\ x_3 &= -3x_4 \\ x_5 &= 3. \end{aligned}$$

You should know how to express this solution in the following two ways. The first involves writing the pivot variables in terms of the free variables:

$$\{(1 - 2x_2, x_2, -3x_4, x_4, 3) : x_2, x_4 \in \mathbb{R}\}.$$

We will call this the *parametric* version of the solution set. The second way of writing the solution set we will call the *vector* version. It looks like this:

$$\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix} : x_2, x_4 \in \mathbb{R}\right\}.$$

Here we are using *column vectors*, and for instance,

$$x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -2x_2 \\ x_2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and we can add column vectors component-wise. Therefore,

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 - 2x_2 \\ x_2 \\ -3x_4 \\ x_4 \\ 3 \end{pmatrix}.$$

Example 2. Suppose the reduced row echelon form for an augmented matrix has the form

$$\left(\begin{array}{ccccccc|c} 0 & 1 & 0 & 1 & 2 & 0 & 1 & 7 \\ 0 & 0 & 1 & 4 & -1 & 0 & 3 & -2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Write the solution set in parametric and vector form.

Solution: The first step is to convert the augmented matrix as a system of equations:

$$\begin{aligned} x_2 \quad + \quad x_4 + 2x_5 \quad + \quad x_7 &= 7 \\ x_3 + 4x_4 - \quad x_5 \quad + 3x_7 &= -2 \\ x_6 + \quad x_7 &= 3. \end{aligned}$$

(The last row corresponds to the equation $0 = 0$, which we can discard.) The pivot variables are x_2, x_3, and x_6. The rest are free variables. Solving for the pivot variables:

$$\begin{aligned} x_2 &= 7 - x_4 - 2x_5 - x_7 \\ x_3 &= -2 - 4x_4 + x_5 - 3x_7 \\ x_6 &= 3 - x_7. \end{aligned}$$

The parametric form for the solutions is

$$\{(x_1,\ 7 - x_4 - 2x_5 - x_7,\ -2 - 4x_4 + x_5 - 3x_7,\ x_4,\ x_5,\ 3 - x_7,\ x_7) : x_4, x_5, x_7 \in \mathbb{R}\}, \tag{2.1}$$

and the vector form is

$$\left\{ \begin{pmatrix} 0 \\ 7 \\ -2 \\ 0 \\ 0 \\ 3 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ -1 \\ -4 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_7 \begin{pmatrix} 0 \\ -1 \\ -3 \\ 0 \\ 0 \\ -1 \\ 1 \end{pmatrix} : x_4, x_5, x_7 \in \mathbb{R} \right\}.$$

Important: Note how these two forms of the solutions are related. For instance, looking at just the constant terms in (2.1) gives the first column vector displayed above. Looking at just the coefficients of x_4 gives the second column vector, and so on. We get one column for the constants and one column for each of the free variables.

The next example illustrates how linear algebra can be used to say something about non-linear objects.

Example 3. Find all parabolas $f(x) = ax^2 + bx + c$ passing through the points $(1,4)$ and $(3,6)$ (determine a, b, and c).

Solution: To pass through $(1,4)$, we need $f(1) = 4$, i.e.,

$$4 = a \cdot 1^2 + b \cdot 1 + c = a + b + c,$$

and to pass through $(3,6)$, we need $f(3) = 6$, i.e.,

$$6 = a \cdot 3^2 + b \cdot 3 + c = 9a + 3b + c.$$

So we need to solve the system

$$\begin{aligned} a + b + c &= 1 \\ 9a + 3b + c &= 6. \end{aligned}$$

Apply our algorithm:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 9 & 3 & 1 & 6 \end{array}\right) \xrightarrow{r_2 \to r_2 - 9r_1} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & -6 & -8 & -30 \end{array}\right) \xrightarrow{r_2 \to -\frac{1}{6} r_2}$$

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 4/3 & 5 \end{array}\right) \xrightarrow{r_1 \to r_1 - r_2} \left(\begin{array}{ccc|c} 1 & 0 & -1/3 & -1 \\ 0 & 1 & 4/3 & 5 \end{array}\right).$$

The corresponding system is

$$\begin{aligned} a \quad - \tfrac{1}{3}c &= -1 \\ b + \tfrac{4}{3}c &= 5. \end{aligned}$$

The solution set is

$$\left\{(-1 + \tfrac{1}{3}c, 5 - \tfrac{4}{3}c, c) : c \in \mathbb{R}\right\}$$

or

$$\left\{ \begin{pmatrix} -1 \\ 5 \\ 0 \end{pmatrix} + c \begin{pmatrix} 1/3 \\ -4/3 \\ 1 \end{pmatrix} : c \in \mathbb{R} \right\}.$$

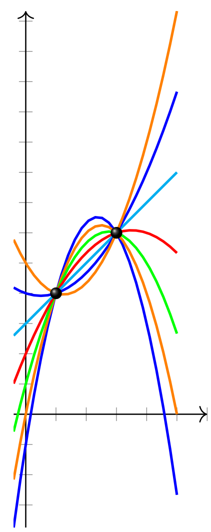
In this way, the set of parabolas passing through the two given points is parametrized by a line in 3-space. For each c, we get a corresponding parabola—the graph of the function

$$f(x) = \left(-1 + \tfrac{1}{3}c\right) x^2 + \left(5 - \tfrac{4}{3}c\right) x + c.$$

Here is a check that all of these parabolas pass through $(1,4)$ and $(3,6)$:

$$f(1) = \left(-1 + \tfrac{1}{3}c\right) + \left(5 - \tfrac{4}{3}c\right) + c = 4$$
$$f(3) = \left(-1 + \tfrac{1}{3}c\right)3^2 + \left(5 - \tfrac{4}{3}c\right)3 + c = 6.$$

Here are graphs of a few of these ($c = -1, 0, 1, 2, 3, 4, 5$):



Friday, Week 1: Introduction to $\mathbb{R}$

We will start the study of abstract vector spaces in the next class. Today, we will introduce a particular vector space, $\mathbb{R}^n$, and informally discuss some of its subspaces (lines, planes, etc.)

Definition. *Real n-space* is the set

$$\mathbb{R}^n := \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n\text{-factors}} := \{(x_1, \ldots, x_n) : x_i \in \mathbb{R} \text{ for } i = 1, \ldots, n\}.$$

with two operations: *addition* $+\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and *scalar multiplication* $\cdot\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ defined, respectively, as follows:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) := (x_1 + y_1, \ldots, x_n + y_n),$$

and

$$\lambda(x_1, \ldots, x_n) := (\lambda x_1, \ldots, \lambda x_n)$$

for all $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

Example 1. In $\mathbb{R}^4$,

$$(4, 0, -2, 1) + (3, 1, 2, -4) = (7, 1, 0, -3)$$

and

$$2(0, 3, 3, 7) = (0, 6, 6, 14).$$

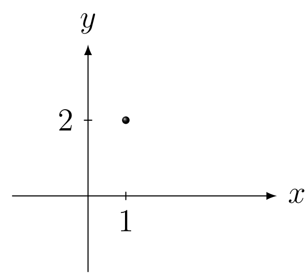
We will often think of points in $(x_1, \ldots, x_n) \in \mathbb{R}^n$ as column vectors or column matrices:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

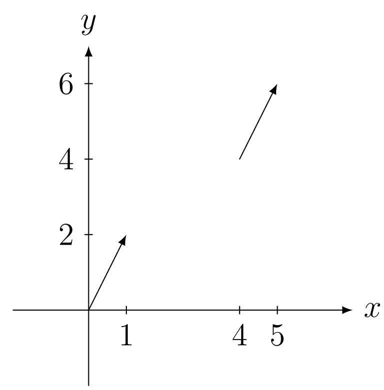
Example 2.

$$2\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \\ -4 \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \\ 12 \end{pmatrix}.$$

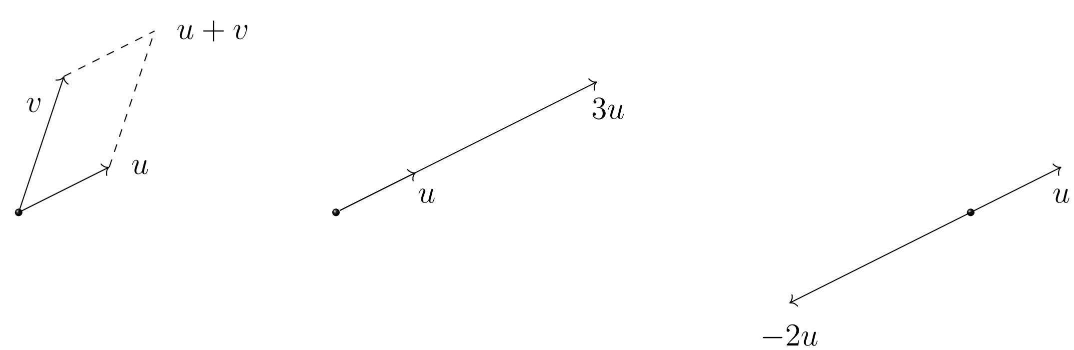
There are two standard ways to interpret an element $p \in \mathbb{R}^n$: either as *point* in space or as a *vector* (thought of as an arrow or direction[1]) For instance, we can think of $(1, 2) \in \mathbb{R}^2$ as point:



or as an arrow/direction:



Geometrically, addition of vectors is given by the "parallelogram rule" and scalar multiplication amounts to scaling the length of a vector but not its directions (except that scaling by a negative number reverses direction):



[1]The word "vector" will soon have a technical definition: an element of a vector space.

Definition (line in $\mathbb{R}^n$). Let $p, v \in \mathbb{R}^n$, with $v \neq (0, \ldots, 0)$. Then

$$\{p + \lambda v : \lambda \in \mathbb{R}\}$$

is the *line in $\mathbb{R}^n$ in the direction of v and passing through the point p.*

Remarks.

(a) In the above definition, the set of all scalar multiples of v, i.e., $\{\lambda v : \lambda \in \mathbb{R}\}$, gives a line *through the origin* in the direction of v. We then translate that line through the origin by p. In that translation, the origin is translated to the point p (and, thus, the resulting line contains the point p).

(b) We say the function

$$\begin{aligned} \mathbb{R} &\to \mathbb{R}^n \\ t &\mapsto p + tv \end{aligned}$$

is a *parametrization* of the line passing through p in the direction of v. Its image is the line, itself. We could use λ or any symbol here instead of t, of course. We are using t to connote time. We think of the parametrization as giving the position at each time t of a point traveling along the line.)

(c) **Exercise.** Try to show that if q is any point on the line $\{p + \lambda v : \lambda \in \mathbb{R}\}$ and w is any nonzero scalar multiple of v, then

$$\{q + \lambda w : \lambda \in \mathbb{R}\} = \{p + \lambda v : \lambda \in \mathbb{R}\}$$

So the same line may be described in many different ways. (This is not hard to do: start with an arbitrary element of the set on the left-hand side, and then show that is contained in the set on the right-hand side.)

Example 3. Give a parametrization of the line through the points $(1, 2, 0)$ and $(0, 1, 1)$ in $\mathbb{R}^3$.

Solution. For the direction, we may choose $v = (0, 1, 1) - (1, 2, 0) = (-1, -1, 1)$. (We think of v as the vector with head $(0, 1, 1)$ and tail $(1, 2, 0)$). We then pick any point on the line, say $p = (1, 2, 0)$. Then our line has the parametrization

$$t \mapsto p + tv = (1, 2, 0) + t\left((0, 1, 1) - (1, 2, 0)\right) = (1, 2, 0) + t(-1, -1, 1).$$

Setting t equal to 0 and 1, respectively, shows that this line really does pass throught the points $(1, 2, 0)$ and $(0, 1, 1)$. Here is an alternative way to write this parametrization:

$$t \mapsto (1 - t, 2 - t, t).$$

Example 4. *A line in $\mathbb{R}^2$ can always be expressed as the solution set to a single linear equation.* For example, consider the line

$$L := \{(3,2) + \lambda(1,5) : \lambda \in \mathbb{R}\}.$$

A point (x, y) lies on this line L if and only if

$$(x, y) = (3, 2) + \lambda(1, 5) = (3 + \lambda, 2 + 5\lambda)$$

for some $\lambda \in \mathbb{R}$. It follows that

$$\begin{aligned}
(x, y) \in L &\iff x = 3 + \lambda \quad \text{and} \quad y = 2 + 5\lambda \text{ for some } \lambda \\
&\iff x - 3 = \lambda \quad \text{and} \quad \frac{1}{5}(y - 2) = \lambda \text{ for some } \lambda \\
&\iff x - 3 = \frac{1}{5}(y - 2) \\
&\iff 5x - 15 = (y - 2) \\
&\iff 5x - y = 13.
\end{aligned}$$

So the line L is the set of solutions to $5x - y = 13$.

Example 5. *A line in $\mathbb{R}^3$ is always the solution set to a system of two linear equations.* For example, consider the line through the points $(1, 0, 2)$ and $(3, 1, -1)$, parametrized by

$$t \mapsto (1, 0, 2) + t\left((3, 1, -1) - (1, 0, 2)\right) = (1, 0, 2) + t(2, 1, -3).$$

We would like to find a system of two linear equations whose solution set is this line. A linear equation in $\mathbb{R}^3$ has the form

$$ax + by + cz = d.$$

We would like to find a, b, c, d so that the solution set for this equation contains the points $(1, 0, 2)$ and $(3, 1, -1)$. It turns out that this will force the whole line to be contained in the solution set. To contain these points, we need

$$\begin{aligned}
a \phantom{{}+b} + 2c &= d \\
3a + b - c &= d.
\end{aligned}$$

To solve this system, we first put it in reduced row echelon form by substracting 3 times the first equation from the second to get

$$\begin{aligned}
a \phantom{{}+b} + 2c &= d \\
b - 7c &= -2d.
\end{aligned}$$

Solve for the pivot variables:

$$\begin{aligned} a &= -2c + d \\ b &= 7c - 2d. \end{aligned}$$

To find two *independent* solutions[2], we first set $(c, d) = (1, 0)$, in which case

$$a = -2, b = 7, c = 1 \text{ and } d = 0.$$

We stick these values into $ax + by + cz = d$ to get our first linear equation

$$-2x + 7y + z = 0.$$

Next we set $(c, d) = (0, 1)$ to get

$$a = 1, b = -2, c = 0 \text{ and } d = 1$$

with corresponding linear equation

$$x - 2y = 1.$$

In sum, our line is the solution set to the system

$$\begin{aligned} -2x + 7y + z &= 0 \\ x - 2y \qquad &= 1. \end{aligned}$$

The reader should check that our original points, $(1, 0, 2)$ and $(3, 1, -1)$ are both solutions to this system.

Definition (plane in $\mathbb{R}^n$). Let $p, v, w \in \mathbb{R}^n$. Suppose that v and w are nonzero and that neither is a scalar multiple of the other. Then

$$\left\{p + \lambda v + \mu w : (\lambda, \mu) \in \mathbb{R}^2\right\}$$

is the plane in $\mathbb{R}^n$ containing p and with directions[3] v and w.

The plane in the above definition has the *parametrization*

$$\begin{aligned} \mathbb{R}^2 &\to \mathbb{R}^n \\ (s, t) &\mapsto p + sv + tw. \end{aligned}$$

Example 6. Find the plane P through the points $(0, 2, -1)$, $(4, 2, 1)$, and $(1, 0, 1)$. Describe both parametrically and as the solution set to a single linear equation.

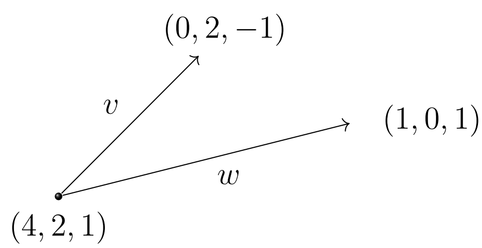
[2] The precise meaning of *independence* is left for later.

[3] The word “directions” here is not quite standard but will do for now

Solution. To find the directions, we fix any of the three points, say $(4, 2, 1)$, and we consider the two arrows having this point as their tail and $(0, 2, -1)$ and $(1, 0, 1)$ as their heads:

$$\begin{aligned} v &= ((0, 2, -1) - (4, 2, 1)) = (-4, 0, -2) \\ w &= ((1, 0, 1) - (4, 2, 1)) = (-3, -2, 0). \end{aligned}$$

Geometrically, we have the following picture for these directions:



So a parametric description of the line is

$$\{(4, 2, 1) + \lambda(-4, 0, 2) + \mu(-3, -2, 0) : \lambda, \mu \in \mathbb{R}\} .$$

or

$$\{(4 - 4\lambda - 3\mu, 2 - 2\mu, 1 + 2\lambda) : \lambda, \mu \in \mathbb{R}\} .$$

As an aside: this parametric description is great for drawing the plane using a computer. For instance, the Sage code for plotting this plane could be:

```
s,t = var("s,t")
parametric_plot3d((4-4*s-3*t,2-2*t,1+2*s),(s,-1,1),(t,-1,1))
```

We would now like to describe the plane P as the solution set of a single linear equation. That equation will have the form

$$ax + by + cz = d$$

for some $a, b, c, d \in \mathbb{R}$. Our job is to find a, b, c, d. (Scaling an equation does not change its solution set, so our solution will only be unique up to such scaling.) For P to contain $(0, 2, -1)$, $(4, 2, 1)$, and $(1, 0, 1)$, we need

$$\begin{aligned} 2b - c &= d \\ 4a + 2b + c &= d \\ a + c &= d. \end{aligned}$$

Reducing to row echelon form and solving for pivot variables, we find

$$\begin{aligned} a &= -d \\ b &= \frac{3}{2}d \\ c &= 2d. \end{aligned}$$

We are free to choose any nonzero value for d (again: our solution is only unique up to scaling). We choose $d = 2$ (to get rid of the denominator of 2). Therefore, the plane P is the solution set to the equation

$$-2x + 3y + 4z = 2.$$

Week 2, Wednesday: Vector spaces

Let F be a field, e.g., $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}/2\mathbb{Z}$ (but not $\mathbb{Z}$).

Definition. A *vector space over* F is a set V with two operations

$$\text{vector addition:} \qquad \begin{aligned} +\colon V \times V &\to V \\ (v, w) &\mapsto v + w \end{aligned}$$

$$\text{scalar multiplication:} \qquad \begin{aligned} +\colon F \times V &\to V \\ (a, v) &\mapsto av \end{aligned}$$

such that the following hold for all $x, y, z \in V$ and $a, b \in F$:

(a) $x + y = y + x$ (commutativity of addition).

(b) $(x + y) + z = (x + y) + z$ (associativity of addition).

(c) There exists $0 \in V$ such that $0 + w = w$ for all $w \in V$.

(d) There exists $-x \in V$ such that $x + (-x) = 0$.

(e) For $1 \in F$, we have $1 \cdot x = x$.

(f) $(ab)x = a(bx)$ (associativity of scalar multiplication).

(g) $a(x + y) = ax + ay$ (distributivity).

(h) $(a + b)x = ax + bx$ (distributivity).

Remark. Rules 1–4 provide the *additive structure* and say that under addition V forms an *abelian group*. Rules 5–8 deal with the second operation, scalar multiplication. Together, they provide a *linear structure* for the set V.

Exercise. Let v be an element of a vector space. Prove that $(-1)v = -v$.

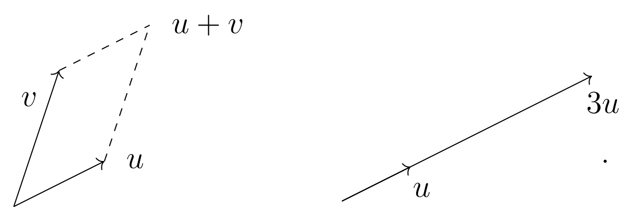
Example. Let $F^n = \underbrace{F \times \cdots \times F}_{n \text{ times}} = \{(a_1, \ldots, a_n) : a_i \in F \text{ for } i = 1, \ldots, n\}$ with the operations

$$\begin{aligned}(a_1, \ldots, a_n) + (b_1, \ldots, b_n) &:= (a_1 + b_1, \ldots, a_n + b_n)\\ c(a_1, \ldots, a_n) &:= (ca_1, \ldots, ca_n)\end{aligned}$$

for all $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in F^n$ and $c \in F$. Then F^n is a vector space.

Special cases:

(a) $F = \mathbb{R}$ and $n = 2$. This gives $\mathbb{R}^2$ with its usual linear structure. Addition is given by the "parallelogram rule" and scalar multiplication scales length:



Here are examples of the vector space axioms in the special case $V = \mathbb{R}^2$:

1. commutativity of +:

$$(6,3) + (-2,4) = (4,7) = (-2,4) + (6,3);$$

2. associativity of +:

$$\begin{aligned}((6,3) + (-2,4)) + (0,2) &= (4,7) + (0,2)\\ &= (4,9)\\ &= (6,3) + (-2,6)\\ &= (6,3) + ((-2,4) + (0,2));\end{aligned}$$

3. zero vector:

$$(0,0) + (6,3) = (6,3);$$

4. additive inverses:

$$(6,3) + (-6,-3) = (0,0);$$

5. scaling by 1:

$$1 \cdot (6,3) = (1 \cdot 6, 1 \cdot 3) = (6,3);$$

6. associativity of scalar multiplication:
$$(3 \cdot 2)(6,3) = 6(6,3) = (36,18) = 3(12,6) = 3\,(2\,(6,3))\,;$$

7. distributivity:
$$3((6,3)+(-2,4)) = 3(4,7) = (12,21)$$
and
$$3(6,3)+3(-2,4) = (18,9)+(-6,12) = (12,21);$$

8. distributivity:
$$(3+2)(6,3) = 5(6,3) = (30,15)$$
and
$$3(6,3)+2(6,3) = (18,9)+(12,6) = (30,15).$$

(b) $F = \mathbb{Z}/3\mathbb{Z}$ and $n = 4$. For example, $(0,1,0,0), (1,1,0,2) \in (\mathbb{Z}/3\mathbb{Z})^4$, and
$$(0,1,0,0)+2(1,1,0,2) = (0,1,0,0)+(2,2,0,1) = (2,0,0,1).$$

(c) The field F is a vector space over itself (this is the case of F^n with $n = 1$).

More examples of vector spaces.

(i) The field $\mathbb{C}$ is a vector space over $\mathbb{R}$. For all $a, b, c, d, t \in \mathbb{R}$, we have
$$\begin{aligned}&(a+bi)+(c+di) = (a+c)+(b+d)i\\&t(a+bi) = ta+(tb)i.\end{aligned}$$

(ii) The field $\mathbb{R}$ is a vector space over $\mathbb{Q}$.

(iii) The set of $m \times n$ matrices with entries in F:
$$M_{m\times n} := \left\{ \begin{pmatrix} a_{11} & \dots & a_{1n} \\ & \vdots & \\ a_{m1} & \dots & a_{mn} \end{pmatrix} : a_{ij} \in F \text{ for all } i,j \right\}.$$
has a standard vector space structure. Given $A \in M_{m\times n}$, denote the entry in its i-the row and j-th column by A_{ij}. Define the vector space operations on $M_{m\times n}$ as follows:

addition: $(A+B)_{ij} := A_{ij} + B_{ij}$ for all $A, B \in M_{m\times n}$;

scalar multiplication: $(cA)_{ij} := cA_{ij}$ for all $A \in M_{m\times n}$ and $c \in F$.

For example, let $F = \mathbb{Q}$, $m = 2$, and $n = 3$. We have

$$2\begin{pmatrix} 1 & 0 & 3 \\ -1 & 2 & 0 \end{pmatrix} + 5\begin{pmatrix} 0 & 2 & -1 \\ 1 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 10 & 1 \\ 3 & 4 & 20 \end{pmatrix}.$$

Calling this last matrix A, we have $A_{1,1} = 2$, $A_{1,2} = 10, \ldots, A_{2,3} = 20$.

(iv) (**Important.**) If S is any set, let F^S be the set of functions $f\colon S \to F$. This function space is naturally an F-vector space (i.e., a vector space with scalar field F) with the following operations: for $f, g \in F^S$ and $t \in F$ define $f + g$ and tf by

addition: $$(f + g)(s) := f(s) + g(s)$$

scalar multiplication: $$(tf)(s) := t(f(s)).$$

Special cases:

- If $S = \{1, \ldots, n\}$, then F^S is essentially F^n. For example, we can think of $(3, 2) \in \mathbb{R}$ as the function

$$\begin{aligned} f\colon \{1, 2\} &\to \mathbb{R} \\ 1 &\mapsto 3 \\ 2 &\mapsto 2. \end{aligned}$$

  In general, $(a_1, \ldots, a_n) \in F^n$ can be thought of as the function

$$\begin{aligned} f\colon \{1, \ldots, n\} &\to F \\ i \quad &\mapsto a_i. \end{aligned}$$

- Similarly, if $S = \{(i, j) : i = 1, \ldots, m \text{ and } j = 1, \ldots, n\}$, then F^S may be identified with $M_{m\times n}$ with $f \in F^S$ corresponding to the matrix A where $A_{ij} = f(i, j)$.
- If $S = \{1, 2, 3, \ldots\}$, then F^S is the vector space of infinite sequences in F. For example, the sequence $1, 1/2, 1/4, 1/8, \ldots$ in $\mathbb{Q}$ can be identified with the function $f\colon S \to \mathbb{Q}$ defined by $f(i) = 1/2^i$.

Definition. A subset $W \subseteq V$ of a vector space V is a *subspace* of V is it is a vector space with the operations of addition and scalar multiplication inherited from V.

We will talk about subspaces in the next class.

Week 2, Friday: Subspaces and spanning sets I

Note: Unless otherwise stated, from now on V will denote a vector space over a field F.

The over-arching goal of the next several classes is to define the notions of *dimension* and *isomorphism* and show that every finite-dimensional vector space over F is isomorphic to the vector space of d-tuples, F^d, where d is the dimension. Today's class lays some of the groundwork for reaching that goal.

Definition. Let S be a nonempty subset of V. Then $v \in V$ is a *linear combination* of vectors in S if there exist $u_1, \ldots, u_n \in S$ and $a_1, \ldots, a_n \in F$ (for some n) such that

$$v = \sum_{i=1}^{n} a_i u_i = a_1 u_1 + \cdots + a_n u_n.$$

Example. Let $S = \{(3,2), (2,-1)\} \subset \mathbb{R}$. Is $(-1,4)$ a linear combination of vectors in S? In other words, do there exist $a, b \in \mathbb{R}$ such that

$$a(3,2) + b(2,-1) = (-1,4)?$$

Since $a(3,2) + b(2,-1) = (3a+2b, 2a-b)$, the above requirement is equivalent to the existence of $a, b \in \mathbb{R}$ such that

$$\begin{aligned} 3a + 2b &= -1 \\ 2a - b &= 4, \end{aligned}$$

a system of linear equations! Apply our algorithm to look for solutions:

$$\left(\begin{array}{cc|c} 3 & 2 & -1 \\ 2 & -1 & 4 \end{array}\right) \xrightarrow{r_1 \to r_1 - r_2} \left(\begin{array}{cc|c} 1 & 3 & -5 \\ 2 & -1 & 4 \end{array}\right) \xrightarrow{r_2 \to r_2 - 2r_1} \left(\begin{array}{cc|c} 1 & 3 & -5 \\ 0 & -7 & 14 \end{array}\right) \xrightarrow{r_2 \to -\frac{1}{7} r_2}$$

$$\left(\begin{array}{cc|c} 1 & 3 & -5 \\ 0 & 1 & -2 \end{array}\right) \xrightarrow{r_1 \to r_1 - 3r_2} \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -2 \end{array}\right).$$

Thus, $a = 1$ and $b = -2$. Check:

$$1 \cdot (3, 2) - 2(2, -1) = (-1, 4). \quad \checkmark$$

So $(-1, 4)$ is a linear combination of the two given vectors. (If it were not, we would have had an inconsistent system, i.e., a system with no solutions.)

Definition. Let S be a nonempty subset of V. The *span* of S, denoted $\operatorname{Span}(S)$, is the set of all linear combinations of elements of S. By convention $\operatorname{Span} \emptyset := \{0\}$, and we say that 0 is the *empty linear combination*.

Example. In $\mathbb{R}^2$,

$$\operatorname{Span}\{(1, 1)\} = \{(a, a) : a \in \mathbb{R}\}.$$

In $\mathbb{R}^3$,

$$\operatorname{Span}\{(1, 0, 0), (0, 1, 0)\} = \{a(1, 0, 0) + b(0, 1, 0) : a, b \in \mathbb{R}\} = \{(a, b, 0) : a, b \in \mathbb{R}\}.$$

Note that the same set can be spanned by different sets of vectors, for instance,

$$\begin{aligned}\operatorname{Span}\{(1, 0, 0), (0, 1, 0)\} &= \operatorname{Span}\{(1, 0, 0), (0, 2, 0)\} \\ &= \operatorname{Span}\{(1, 0, 0), (0, 1, 0), (2, 3, 0)\}.\end{aligned}$$

A point in $\mathbb{R}^3$ is in any of these sets if and only if its third component is 0.

Definition. A subset $W \subseteq V$ is a *subspace* of V if W is a vector space itself with the operations of addition and scalar multiplication inherited from V.

Proposition. $W \subseteq V$ is a subspace of V if and only if

(a) $0 \in W$

(b) W is closed under addition $(x, y \in W \Rightarrow x + y \in W)$

(c) W is closed under scalar multiplication $(c \in F$ and $w \in W \Rightarrow cw \in W)$.

Proof. Exercise. Part 1 is there to ensure that W is nonempty. (Note that Part 2 and Part 3 are vacuously true for the empty set, and yet the empty set is not a subspace because of Part 1.) $\square$

Examples.

(a) $W = \{(a, 0) : a \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.

Proof. Letting $a = 0$, we see $(0, 0) \in W$. If $(a, 0), (b, 0) \in W$, then $(a, 0) + (b, 0) = (a + b, 0) \in W$. If $c \in \mathbb{R}$ and $(a, 0) \in W$, then $c(a, 0) = (ca, 0) \in W$. Thus, W is a subspace of $\mathbb{R}^2$. $\square$

(b) Let

$$V = \{f\colon \mathbb{R} \to \mathbb{R} : f \text{ is continuous}\},$$
$$W = \{f\colon \mathbb{R} \to \mathbb{R} : f \text{ is differentiable}\}.$$

Both V and W are subspaces of the vector space $\mathbb{R}^{\mathbb{R}}$ of functions from $\mathbb{R}$ to $\mathbb{R}$ (recall our earlier notation F^S for functions from a set S to a field F), and W is a subspace of V.

(c) Let $W = \{(a, b) \in \mathbb{R}^2 : ab = 0\}$. So W is the union of the two coordinate axes in $\mathbb{R}^2$. Each of these coordinate axes is a subspace of $\mathbb{R}^2$, but W is not. For instance, $(1, 0), (0, 1) \in W$, but $(1, 0) + (0, 1) = (1, 1) \notin W$. So W is not closed under addition.

(d) $\{0\}$ and V are always subspaces of V. The empty set $\emptyset$ is not a subspace (since it does not contain 0).

Proposition. If W_1 and W_2 are subspaces of V, so is $W_1 \cap W_2$.

Proof. Since W_1 and W_2 are subspaces, we have $0 \in W_i$ for $i = 1, 2$. Hence, $0 \in W_1 \cap W_2$. If $u, v \in W_1 \cap W_2$, then $u, v \in W_i$ for $i = 1, 2$. Hence, $u + v \in W_i$ for $i = 1, 2$. Similarly, for each $\lambda \in F$,

$$\begin{aligned} u \in W_1 \cap W_2 \quad &\Rightarrow \quad u \in W_1 \text{ and } u \in W_2 \\ &\Rightarrow \quad \lambda u \in W_1 \text{ and } \lambda u \in W_2 \\ &\Rightarrow \quad \lambda u \in W_1 \cap W_2. \end{aligned}$$

□

Proposition. Let S be a subset of V. Then:

(a) $\operatorname{Span}(S)$ is a subspace of V.

(b) If $W \subseteq V$ is a subspace and $S \subseteq W$, then $\operatorname{Span}(S) \subseteq W$. (In other words: a subspace is closed under the process of taking linear combinations of its elements.)

(c) Every subspace of V is the span of some subset of V.

Proof. 1. If $S = \emptyset$, then $\operatorname{Span}(S) = \{0\}$, which is a subspace of V. Otherwise, we will show $0 \in \operatorname{Span}(S)$ and $\operatorname{Span}(S)$ is closed under addition and scalar multiplication. Since $S \neq \emptyset$, there exists some $u \in S$. Then $0 \cdot u$ is a linear combination of elements in S, and $0 \cdot u = 0$ (the first 0 in this equation is in F, and the second is in V). Hence, $0 \in \text{Span}(S)$. Now let $x, y \in \text{Span}(S)$ so that

$$x = a_1u_1 + \cdots + a_mu_m$$
$$y = b_1v_1 + \cdots + b_nv_n$$

for some $a_i, b_i \in F$ and $u_i, v_i \in S$. Then

$$x + y = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n \in \text{Span}(S)$$

and for each $\lambda \in F$,

$$\lambda x = \lambda(a_1u_1 + \cdots + a_mu_m) = (\lambda a_1)u_1 + \cdots + (\lambda_m a_m)u_m \in \text{Span}(S).$$

2. Take $x \in \text{Span}(S)$. Then $x = a_1u_1 + \cdots + a_mu_m$ for some $a_i \in F$ and $u_i \in S$. Since $S \subseteq W$, we have $u_i \in W$ for all i, and since W is a subspace, it is closed under vector addition and scalar multiplication. Therefore, $x \in W$.

3. $\text{Span}(W) = W$. □

Definition. A subset $S \subseteq V$ *generates* a subspace W if $\text{Span}(S) = W$.

Examples.

(a) $\{1, x, x^2, \ldots, \}$ generates $P(F)$, the vector space of polynomials in one variable over F. More commonly, this vector space is denoted $F[x]$.

(b) $\{(1, 0), (0, 1)\}$ generates $\mathbb{R}^2$. So do $\{(1, 0), (0, 1), (3, -2)\}$ and $\{(1, 1), (0, 1)\}$.

(c) The *i-the standard basis vector* for F^n is $e_i := (0, \ldots, 0, 1, 0, \ldots, 0)$, the vector whose only nonzero entry is in the i-th component. We have that $\{e_1, \ldots, e_n\}$ generates F^n.

Week 3, Monday: Subspaces and spanning sets II

In today's lecture, we start by proving a simple (but useful) results about spanning sets. We then present several examples of subspaces and spanning sets.

Recall the definitions presented last time:

Definition. Let S be a nonempty subset of V. Then $v \in V$ is a *linear combination* of vectors in S if there exist $u_1, \ldots, u_n \in S$ and $a_1, \ldots, a_n \in F$ (for some n) such that

$$v = \sum_{i=1}^{n} a_i u_i = a_1 u_1 + \cdots + a_n u_n.$$

Definition. Let S be a nonempty subset of V. The *span* of S, denoted $\operatorname{Span}(S)$, is the set of all linear combinations of elements of S. By convention $\operatorname{Span} \emptyset := \{0\}$, and we say that 0 is the *empty linear combination*.

Lemma. Let V be a vector space over F, let $S \subseteq V$, and let $v \in V$. Then

$$\operatorname{Span}(S \cup \{v\}) = \operatorname{Span}(S) \quad \Leftrightarrow \quad v \in \operatorname{Span}(S).$$

Proof. ($\Rightarrow$) If $\operatorname{Span}(S \cup \{v\}) = \operatorname{Span}(S)$, then since $v \in \operatorname{Span}(S \cup \{v\})$, it follows that $v \in \operatorname{Span}(S)$.

($\Leftarrow$) Suppose that $v \in \operatorname{Span}(S)$. We wish to show that $\operatorname{Span}(S \cup \{v\}) = \operatorname{Span}(S)$. Suppose that $w \in \operatorname{Span}(S \cup \{v\})$. Then we can write

$$w = a_1 s_1 + \cdots + a_k s_k + bv$$

for some $s_1, \ldots, s_k \in S$ and some $a_1, \ldots, a_k, b \in F$. We are given that $v \in \operatorname{Span}(S)$. Hence,

$$v = c_1 t_1 + \cdots + c_\ell t_\ell$$

for some $t_1, \ldots, t_\ell \in S$ and some $c_1, \ldots, c_\ell \in F$. Substituting into the previous equation, we see

$$w = a_1 s_1 + \cdots + a_k s_k + b(c_1 t_1 + \cdots + c_\ell t_\ell)$$

$$\begin{aligned}&= a_1 s_1 + \cdots + a_k s_k + bc_1 t_1 + \cdots + bc_\ell t_\ell \\ &\in \operatorname{Span}(S).\end{aligned}$$

We have shown that $\operatorname{Span}(S \cup \{v\}) \subseteq \operatorname{Span}(S)$. The opposite inclusion also holds since one is easily sees that

$$S \subseteq S \cup \{v\} \Rightarrow \operatorname{Span}(S) \subseteq \operatorname{Span}(S \cup \{v\}).$$

$\square$

We now move on to examples of subspaces and spanning sets.

Example. Recall from the reading that $P_k(F)$ is the vector space of polynomials of degree at most k with coefficients in F. Another, more standard, notation for this vector space is $F[x]_{\leq k}$. We have that

$$P_k(F) = F[x]_{\leq 2} = \operatorname{Span}\{1, x, \ldots, x^k\}.$$

Now let

$$S = \left\{x^2 + 3x - 2, 2x^2 + 5x - 3\right\} \subset \mathbb{R}[x]_{\leq 2}.$$

Is $-x^2 - 4x + 4 \in \operatorname{Span}(S)$?

Solution. We are looking for $a, b \in \mathbb{R}$ such that

$$-x^2 - 4x + 4 = a(x^2 + 3x - 2) + b(2x^2 + 5x - 3),$$

in other words, such that

$$-x^2 - 4x + 4 = (a + 2b)x^2 + (3a + 5b)x + (-2a - 3b).$$

So we need to see if the following system of linear equations has a solution:

$$\begin{aligned} a + 2b &= -1 \\ 3a + 5b &= -4 \\ -2a - 3b &= 4. \end{aligned}$$

Applying Gaussian elimination, we find

$$\left(\begin{array}{rr|r} 1 & 2 & -1 \\ 3 & 5 & -4 \\ -2 & -3 & 4 \end{array}\right) \quad \rightsquigarrow \quad \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right).$$

We see that the system is inconsistent, i.e., it has no solutions. So $-x^2 - 4x + 4 \notin \operatorname{Span}(S)$.

Definition. Let S be any set, and consider the function space $F^S := \{f\colon S \to F\}$. For each $s \in S$, define the *characteristic function* $\chi_s \in F^S$ for s by

$$\begin{aligned}\chi_s\colon S &\to F\\ t &\mapsto \begin{cases} 1 & \text{if } t = s\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Example. Let $S = \{1, 2, 3\}$, and consider the function $f\colon S \to \mathbb{R}$ given by $f(1) = -1$, $f(2) = \pi$, and $f(3) = 16$. Then we can write f as a linear combination of characteristic functions:

$$f = -\chi_1 + \pi\chi_2 + 16\chi_3.$$

For instance,

$$\begin{aligned}f(2) &= (-\chi_1 + \pi\chi_2 + 16\chi_3)(2)\\ &= -\chi_1(2) + \pi\chi_2(2) + 16\chi_3(2)\\ &= -0 + \pi \cdot 1 + 16 \cdot 0 = \pi.\end{aligned}$$

In this way, if S is finite, then $\{\chi_s : s \in S\}$ generates F^S. On the other hand, if S is infinite, things are more complicated. For instance, consider the case where $S = \mathbb{N} = \{0, 1, 2, \dots\}$. Then $\mathbb{R}^{\mathbb{N}}$ is the vector space of infinite real sequences. For instance, the sequence $1, 1/2, 1/4, /8, \dots$ is the function $f \in \mathbb{R}^{\mathbb{N}}$ given by $f(i) := 1/2^i$. If we try to write f as a linear combination of characteristic functions, we would have

$$f = \chi_0 + \frac{1}{2}\chi_1 + \frac{1}{4}\chi_2 + \frac{1}{8}\chi_3 + \cdots,$$

an infinite sum. However, by definition, the span of a set is the collection of all *finite* linear combinations of elements in the set. Infinite linear combinations like those above involve questions of convergence, and we are not concerned with those issues at the moment.

Definition. A linear equation of the form $a_1x_1 + \cdots + a_nx_n = 0$ where $a_i \in F$ is called *homogeneous*.

Proposition. The solution set to a system of homogeneous linear equations in n unknowns and with coefficients in F is a subspace of F^n.

Proof. First note that the zero vector satisfies any homogeneous linear equation. So the solution set is nonempty. Next, let $u = (u_1, \dots, u_n)$ and $v = (v_1, \dots, v_n)$ be solutions to a system of homogeneous linear equations, and let $a_1x_1 + \cdots + a_nx_n = 0$ be any equation in the system. Thus,

$$\begin{aligned} a_1u_1 + \cdots + a_nu_n &= 0 \\ a_1v_1 + \cdots + a_nv_n &= 0. \end{aligned}$$

Now let $\lambda \in F$ and consider

$$u + \lambda v = (u_1 + \lambda v_1, \ldots, u_n + \lambda v_n).$$

The following calculation shows that $u + \lambda v$ is also a solution to the equation

$$\begin{aligned} a_1(u_1 + \lambda v_1) + \cdots + a_n(u_n + \lambda v_n) &= a_1u_1 + \cdots + a_nu_n + \lambda(a_1v_1 + \cdots + a_nv_n) \\ &= 0 + \lambda \cdot 0 = 0. \end{aligned}$$

$\square$

Terminology. Since the solution set to a system of homogeneous linear equations is a subspace, we usually refer to the solution set as the *solution space* for the system.

Example. Writing a solution to a system of homogeneous linear equations in vector form yields a set of generators for the solution space. For example, consider the system

$$\begin{aligned} x \qquad + z + w &= 0 \\ 2x + y \qquad - w &= 0 \\ 3x + y + z \qquad &= 0. \end{aligned}$$

We solve the system by performing Gaussian elimination (intermediary steps omitted):

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 0 \\ 2 & 1 & 0 & -1 & 0 \\ 3 & 1 & 1 & 0 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -2 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Converting back into equations and solving for the leading (pivot) variables gives

$$\begin{aligned} x &= -z - w \\ y &= 2z + 3w. \end{aligned}$$

So the set of solutions (in parametric form) is

$$\{(-z - w, 2z + 3w, z, w) : z, w \in \mathbb{R}\},$$

or, written in vector form,

$$\left\{ z \begin{pmatrix} -1 \\ 2 \\ 1 \\ 0 \end{pmatrix} + w \begin{pmatrix} -1 \\ 3 \\ 0 \\ 1 \end{pmatrix} : z, w \in \mathbb{R} \right\} = \operatorname{Span} \left\{ \begin{pmatrix} -1 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

The solution space is generated by two vectors.

Week 3, Wednesday: Linear Independence

Definition. A set $S \subset V$ is *linearly dependent* if there exist distinct[1] $u_1, \ldots, u_n \in S$, for some $n \geq 1$, and scalars $a_1, \ldots, a_n$, not all zero, such that

$$a_1 u_1 + \cdots + a_n u_n = 0.$$

We call the above expression a *non-trivial dependence relation* among the u_i.

Example. The empty set is not linearly dependent.

Example. If $0 \in S$, then S is linearly dependent. For instance, $1 \cdot 0 = 0$ is a non-trivial dependence relation.

Example. Let $S = \{(1, -1, 0), (-1, 0, 2), (-5, 3, 4)\} \subset \mathbb{R}^3$. Is S linearly dependent? We look for $a_1, a_2, a_3 \in \mathbb{R}$ such that

$$a_1(1, -1, 0) + a_2(-1, 0, 2) + a_3(-5, 3, 4) = (0, 0, 0),$$

i.e., such that

$$(a_1 - a_2 - 5a_3, -a_1 + 3a_3, 2a_2 + 4a_3) = (0, 0, 0).$$

So we are looking for a solution to the system of linear equations

$$\begin{aligned} a_1 - a_2 - 5a_3 &= 0 \\ -a_1 + 3a_3 &= 0 \\ 2a_2 + 4a_3 &= 0. \end{aligned}$$

Apply our algorithm:

$$\left(\begin{array}{rrr|r} 1 & -1 & -5 & 0 \\ -1 & 0 & 3 & 0 \\ 0 & 2 & 4 & 0 \end{array}\right) \xrightarrow{r_2 \to r_2 + r_1} \left(\begin{array}{rrr|r} 1 & -1 & -5 & 0 \\ 0 & -1 & -2 & 0 \\ 0 & 2 & 4 & 0 \end{array}\right) \xrightarrow{r_2 \to -r_2}$$

[1] Note the easily forgotten but necessary word "distinct", here.

$$\left(\begin{array}{ccc|c} 1 & -1 & -5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 2 & 4 & 0 \end{array}\right) \xrightarrow[r_3 \to r_3 - 2r_2]{r_1 \to r_1 + r_2} \left(\begin{array}{ccc|c} 1 & 0 & -3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Converting back to a system of equations and solving for the pivot variables gives

$$a_1 = 3a_3, \quad a_2 = -2a_3,$$

and a_3 is arbitrary. Take $a_3 = 1$ to get the solution $a_1 = 3$, $a_2 = -2$, and $a_3 = 1$:

$$3(1,-1,0) - 2(-1,0,2) + (-5,3,4) = (0,0,0).$$

Therefore, these vectors are linearly dependent.

Proposition 1. Let $S \subseteq V$. Then S is linearly dependent if and only if there exists $v \in S$ such that v is a linear combination of vectors in $S \setminus \{v\}$, i.e., if and only if $v \in \operatorname{Span}(S \setminus \{v\})$.

Proof. First note that we may assume $S \neq \emptyset$ since the empty set is not linearly dependent.

($\Rightarrow$) Suppose $a_1u_1 + \cdots + a_nu_n = 0$ for distinct $u_i \in S$ and $a_i \in F$, not all zero. Without loss of generality, we may assume that $a_1 \neq 0$. In that case, we have

$$u_1 = -\frac{a_2}{a_1}u_2 - \frac{a_3}{a_1}u_3 - \cdots - \frac{a_n}{a_1}u_n,$$

expressing u_1 as a linear combination of elements in $S \setminus \{u_1\}$. Note the special case where $S = \{0\}$. The result still holds in that case since $\{0\} = \operatorname{Span}(\emptyset)$. By definition, the empty linear combination is 0.

($\Leftarrow$) Say $v = a_1u_1 + \cdots + a_nu_n$ with distinct $u_i \in S \setminus \{v\}$ and $v \in S$. Then

$$a_1u_1 + \cdots + a_nu_n - v = 0$$

shows that S is linearly dependent. □

Definition. A set $S \subset V$ is *linearly independent* if it is not linearly dependent. This means that for all $n \geq 1$ and distinct $u_1, \ldots, u_n \in S$, if $a_1u_1 + \cdots + a_nu_n = 0$ for some $a_i \in F$, then $a_1 = \cdots = a_n = 0$. (In particular, the empty set is linearly independent.)

Remark. We say there is a *linear relation* among vectors $u_1, \ldots, u_n$ if there exist $a_i \in F$ such that $a_1u_1 + \cdots + a_nu_n = 0$. The linear relation is *trivial* if all $a_i = 0$.

Thus, a subset S of V is linearly independent if every linear relation distinct elements of S is trivial.

IMPORTANT. To prove that a set of (distinct) vectors $S = \{v_1, \dots, v_k\}$ is linearly independent start by writing the following:

> Suppose that
> $$a_1 v_1 + \cdots + a_k v_k = 0$$
> for some $a_1, \dots, a_k \in F$.

The goal is then to use some knowledge you are given about the vectors $v_1, \dots, v_k$ to show that the relation is trivial, i.e., $a_i = 0$ for all i.

AVOID. Another way to prove that a set of vectors $S = \{v_1, \dots, v_k\}$ is linearly independent is to suppose that some v_i is a linear combination of the vectors in $S \setminus \{v_i\}$ or to suppose that there is some nontrivial linear combination of elements in S, and then show a contradiction arises. Whenever tempted to give such a proof, check to see if the standard proof, described just above, would be clearer (as it almost always will).

Examples.

- The set $\{u\}$ is linearly independent for any nonzero $u \in V$: if $\lambda u = 0$ for some $\lambda \neq 0$, then scaling by $1/\lambda$ would give $u = 0$. But we are supposing $u \neq 0$. (Here is a case where the indirect proof of independence seems warranted.)

- $S = \{(1,-1,0), (-1,0,2), (0,1,1)\} \subset \mathbb{R}$ is linearly independent. To see this, we follow the standard proof. Suppose that
$$a(1,-1,0) + b(-1,0,2) + c(0,1,1) = 0,$$
which means
$$\begin{aligned} a - \quad b \quad\quad &= 0 \\ -a \quad\quad + c &= 0 \\ 2b + c &= 0. \end{aligned}$$
Apply our algorithm (I'll just show the result of row reduction):
$$\left(\begin{array}{rrr|r} 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{rrr|r} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right).$$
Thus, the only solution is $a = b = c = 0$.

- The set $S = \{1 + x, 1 + x + x^2\} \subset P_2(\mathbb{R}) = \mathbb{R}[x]_{\leq 2}$ is linearly independent. To see this, suppose that

$$a(1 + x) + b(1 + x + x^2) = 0$$

for some $a, b \in \mathbb{R}$. It follows that

$$(a + b) + (a + b)x + bx^2 = 0,$$

and, therefore, $a + b = 0$ (the coefficient the constant term or of the x-term) and $b = 0$ (the coefficient of the x^2-term). It then follows that $a = b = 0$.

Problem (leading to an important algorithm). Let

$$S = ((2, 0, 0), (0, 1, 0), (2, 2, 0), (0, 3, 1), (3, 0, 1)) .$$

Find a linearly independent subset of S and write the remaining vectors as linear combinations of vectors in that subset.

Solution. Look for linear relations

$$c_1(2, 0, 0) + c_2(0, 1, 0) + c_3(2, 2, 0) + c_4(0, 3, 1) + c_5(3, 0, 1) = (0, 0, 0).$$

Convert the above relation to as system of three homogeneous linear equations in c_1, c_2, c_3, c_4, c_5 and solve:

$$\left(\begin{array}{ccccc|c} 2 & 0 & 2 & 0 & 3 & 0 \\ 0 & 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & \frac{3}{2} & 0 \\ 0 & 1 & 2 & 0 & -3 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right).$$

(Note that the first matrix has the vectors in S as columns.) So the solution space is

$$\left(-c_3 - \frac{3}{2}c_5, -2c_3 + 3_5, c_3, -c_5, c_5 : c_3, c_5 \in \mathbb{R}\right),$$

or, in parametric form

$$\left\{ c_3 \begin{pmatrix} -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + c_5 \begin{pmatrix} -\frac{3}{2} \\ 3 \\ 0 \\ -1 \\ 1 \end{pmatrix} : c_3, c_5 \in \mathbb{R} \right\}.$$

Let T be the set of columns in our original matrix with the same indices as those for the non-free (i.e., pivot or leading) variables in the row-reduced matrix. In other words,

$$T = \left\{ \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} \right\}.$$

We claim that T is linearly independent. Suppose there is a linear relation (switching to row notation for convenience):

$$a(2,0,0) + b(0,1,0) + c(0,3,1) = 0.$$

To show that $a = b = c = 0$ is the only solution, we convert to a matrix and row-reduce as usual:

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Therefore, we must have $a = b = c = 0$, as claimed. **Important:** In fact, there was no need to do that last computation since we have already done it. To see that, go back to our original row-reduction

$$\left(\begin{array}{ccccc|c} 2 & 0 & 2 & 0 & 3 & 0 \\ 0 & 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & \frac{3}{2} & 0 \\ 0 & 1 & 2 & 0 & -3 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right)$$

and only pay attention to the first, second, and fourth columns. So the verification that T is linearly independent was secretly guaranteed by its construction.

It remains to be shown that the remaining columns (those corresponding to the free variables), i.e., $(2,2,0)$ and $(3,0,1)$, in row notation) are in the span of T. We have found all solutions to

$$c_1(2,0,0) + c_2(0,1,0) + c_3(2,2,0) + c_4(0,3,1) + c_5(3,0,1) = (0,0,0) \tag{7.1}$$

and found that c_3 and c_5 are free variables. To see that $(2,2,0)$ is in the span of T, find the solution to our system for which $(c_3, c_5) = (1,0)$, then solve for $(2,2,0)$ in (7.1). The solution in this case is

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

Therefore,

$$-(2,0,0) - 2(0,1,0) + 1 \cdot (2,2,0) + 0 \cdot (0,3,1) + 0 \cdot (3,0,1) = (0,0,0),$$

and, thus,

$$(2,2,0) = (2,0,0) + 2(0,1,0).$$

Similarly, to show $(3,0,1)$ is in the span of T, we set $(c_3, c_5) = (0,1)$. The corresponding solution is

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} -\frac{3}{2} \\ 3 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

Therefore,

$$-\frac{3}{2}(2,0,0) + 3(0,1,0) + 0 \cdot (2,2,0) - 1 \cdot (0,3,1) + 1 \cdot (3,0,1) = (0,0,0),$$

Solving for $(3,0,1)$ gives

$$(3,0,1) = \frac{3}{2}(2,0,0) - 3(0,1,0) + (0,3,1).$$

We summarize the underlying important algorithm: Let $S = \{v_1, \ldots, v_k\} \in F^n$. To find a linearly independent subset T of S such that $\operatorname{Span}(T) = \operatorname{Span}(S)$:

- Let M be the matrix with *columns* $v_1, \ldots, v_k$.
- Compute M', the row-reduced form of M.
- Let $j_1, \ldots, j_d$ be the indices of the pivot columns of M' (the ones containing the leading 1s).
- Set $T = \{v_{j_1}, \ldots, v_{j_d}\}$.

Note: The set T is a subset of the columns of M *not* of M'!

The elements of $S \backslash T$ correspond to the free variables, and we can write these elements as linear combinations of the elements of T by setting each free variable in turn equal to 1 and setting the remaining free variables equal to 0.

We end with a result of fundamental importance:

Theorem. Let $S \subseteq V$ be linearly independent, and let $v \in \operatorname{Span}(S)$. Then v has a unique expression as a linear combination of elements of S. In other words, if $v = \sum_{i=1}^{k} a_i u_i$ and $v = \sum_{i=1}^{\ell} b_i w_i$ for some nonzero $a_i, b_i \in F$ and some distinct $u_i \in S$ and distinct $w_i \in S$, then up to re-indexing, we have $k = \ell$, $u_i = w_i$, and $a_i = b_i$ for all i.

Proof. Say $v = \sum_{i=1}^{n} a_i u_i$ and $v = \sum_{i=1}^{n} b_i u_i$ for some $a_i, b_i \in F$ and $u_i \in S$. (By letting some a_i and b_i equal zero, these expressions represent two arbitrary representations of v as linear combinations of elements of S, i.e., we can use the same u_i and n for both expressions.) It follows that

$$0 = v - v = \sum_{i=1}^{n} a_i u_i - \sum_{i=1}^{n} b_i u_i = \sum_{i=1}^{n} (a_i - b_i) u_i.$$

Since S is linearly independent, it follows that $a_i - b_i = 0$ for all i. The result follows. $\square$

Example. The previous result does not hold if S is linearly dependent. For instance, consider the set $S = \{(1,1),(2,2)\} \subset \mathbb{R}$. Then

$$(3,3) = (1,1) + (2,2) = 2(1,1) + \frac{1}{2}(2,2) = 3(1,1) + 0(2,2) = \text{ etc.}$$

Exercise. Prove that the converse of the previous proposition holds: if each element of $\operatorname{Span} S$ can be expressed uniquely as a linear combination of elements of S, then S is linearly independent.

Week 3, Friday: Bases

Definition. A subset $B \subset V$ is a *basis* if it is linearly independent and spans V. An *ordered basis* is a basis whose elements have been listed as a sequence: $B = \langle b_1, b_2, \dots \rangle$.[1]

Warning: Our book defines a *basis* to be what we are calling an *ordered basis*. That's not standard, and there are problems with that idea when talking about infinite-dimensional vector spaces, which we will not go into here. We will, however, use the book's notation of "$\langle$" and "$\rangle$" to denote an ordered basis. Thus, for us, the word *basis* will mean "unordered basis", and we will try to be careful to say "ordered basis" when relevant (but will sometimes forget).

Examples.

(a) The *standard ordered basis* for F^n is $\langle e_1, \dots, e_n \rangle$ where the i-th standard basis vector is $e_i = (0, \dots, 0, 1, 0, \dots, 0)$, the vector with i-th component 1 and all other components 0. For instance, the standard ordered basis for F^3 is

$$\langle (1,0,0), (0,1,0), (0,0,1) \rangle.$$

Here is another possible ordered basis for F^3:

$$\langle (1,0,0), (0,1,0), (1,1,1) \rangle.$$

Exercise: check that the above vectors are linearly independent and span F^3.

(b) One ordered basis for the vector space $P_3(F) = F[x]_{\leq 3}$ of polynomials of degree most three is

$$\langle 1, x, x^2, x^3 \rangle.$$

[1]Every vector space has a basis—we will prove this in the finite-dimensional case. An infinite-dimensional vector space may not have a countable basis, i.e., one that can be indexed by the natural numbers. There is a link to a supplemental article at our course homepage, if you would like to know more.

(c) One ordered basis for, $M_{2\times 3}(F)$, the vector space of 2×3 matrices over a field F, is

$$M_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ M_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ M_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

$$M_4 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},\ M_5 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},\ M_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

These matrices span $M_{2\times 3}(F)$:

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = aM_1 + bM_2 + cM_3 + dM_4 + eM_5 + fM_6.$$

To see they are linearly independent, suppose the above sum is 0, i.e., the zero matrix. Then we must have $a = b = c = d = e = f = 0$.

Last time, we showed the following proposition:

Proposition 1 from previous lecture. Let $S \subseteq V$. Then S is linearly dependent if and only if there exists $v \in S$ such that v is a linear combination of vectors in $S \setminus \{v\}$, i.e., if and only if $v \in \operatorname{Span}(S \setminus \{v\})$.

We use this result to prove the following:

Proposition 1. Any finite subset S of V has a linearly independent subset with the same span. In other words, if S is a finite set, then there is a subset of S that is a basis for $\operatorname{Span}(S)$.

Proof. If S is linearly independent, we are done. If not, then by Proposition 1 from the previous lecture, there exists $v \in S$ such that $v \in \operatorname{Span}(S \setminus \{v\})$. It follows that $\operatorname{Span}(S) = \operatorname{Span}(S \setminus \{v\})$. If $S \setminus \{v\}$ is linearly independent, we are done. If not, repeat the above step. The process will end eventually since S is finite. We are OK even if the process ends at the empty set since the empty set is linearly independent. (For instance, if $S = \{0\}$, our process would end at $\emptyset$.) □

In the above, we create a basis for $\operatorname{Span}(S)$ by discarding elements of S. Another possibility is to start at the empty set and start adding elements S that are linearly independent of those we have so far. This follows from:

Proposition 2. If $T \subset V$ is linearly independent and $v \in V \setminus T$, then $T \cup \{v\}$ is linearly dependent if and only if $v \in \operatorname{Span}(T)$.

Proof. ($\Rightarrow$) Suppose that $v \in V \setminus \{v\}$ and that $T \cup \{v\}$ is linearly dependent. Then we may write

$$av + a_1u_1 + \cdots + a_nu_n = 0 \qquad (\star)$$

for some $a, a_1, \ldots, a_n \in F$, *not all zero*, and distinct $u_i \in T$. We can always assume that v appears in this expression by taking $a = 0$, if necessary. But, in fact, $a \neq 0$ since otherwise $(\star)$ would be a linear relation among distinct elements of T. Since T is linearly independent, this would mean that all the $a_i = 0$, in addition to $a = 0$. However, we know that at least one of these scalars in nonzero.

Thus, it must be that $a \neq 0$. We can then solve for v in $(\star)$:

$$v = -\frac{a_i}{a}u_1 - \cdots - \frac{a_n}{a}u_n \in \operatorname{Span}(T).$$

($\Leftarrow$) Suppose that $v \in \operatorname{Span}(T)$. Then

$$v = a_1u_1 + \cdots + a_nu_n$$

for some $a_i \in F$ and $u_i \in T$. Since $v \notin T$, it follows that

$$a_1u_1 + \cdots + a_nu_n + (-1) \cdot v = 0$$

is a nontrivial relation among elements of $T \cup \{v\}$. So $T \cup \{v\}$ is linearly dependent. $\square$

Alternate proof of Proposition 1. We are starting with a finite set S and looking for a subset T of S that is linearly independent and generates $V = \operatorname{Span}(S)$. If $S = \emptyset$ or $S = \{0\}$, we take $T = \emptyset$ and are done. If not, there exists a nonzero element $u_1 \in S$, and we set $T = \{u_1\}$. If $\operatorname{Span}(T) = \operatorname{Span}(S)$, we are done. If not, then there exists $u_2 \in S$ such that $u_2 \notin \operatorname{Span}(T)$. We then append u_2 to T. So now $T = \{u_1, u_2\}$, and by Proposition 2, the set T is linearly independent. If $\operatorname{Span}(T) \neq \operatorname{Span}(S)$, repeat to find $u_3 \in S$ linearly independent of u_1 and u_2. Etc. Since S is finite, the process eventually stops. $\square$

Example. Let $V = (\mathbb{Z}/3\mathbb{Z})^3$, a vector space over $\mathbb{Z}/3\mathbb{Z}$.

How many elements are in V? A point in V has the form (x_1, x_2, x_3), and there are 3 choices for each x_i. Hence, the number of elements in V is $|V| = 3^3 = 27$.

As an exercise, check that the following is a subspace of V:

$$W = \{(x_1, x_2, x_3) \in V : x_1 + x_2 + x_3 = 0\}.$$

How many elements are in W? We have,

$$W = \{(-x_2 - x_3, x_2, x_3) : x_2, x_3 \in \mathbb{Z}/3\mathbb{Z}\}.$$

As we let x_2 and x_3 vary, we get 9 elements:

$$\{(0,0,0),(2,1,0),(1,2,0),(2,0,1),(1,1,1),(0,2,1),(1,0,2),(0,1,2),(2,2,2)\}.$$

Let's try to find a linearly independent generating set for W. Start with $v_1 := (2,1,0)$. The span of $\{v_1\}$ has three elements:

$$\begin{aligned}0\cdot(2,1,0) &= (0,0,0)\\ 1\cdot(2,1,0) &= (2,1,0)\\ 2\cdot(2,1,0) &= (1,2,0).\end{aligned}$$

Next, note that $v_2 = (2,0,1)$ is not in $\operatorname{Span}(\{v_1\})$. By Proposition 2, we see that $S := \{v_1, v_2\}$ is linearly independent. We claim $\operatorname{Span}(S) = W$. First, since $v_1, v_2 \in W$, we see $\operatorname{Span}(S) \subseteq W$. Next, by Theorem 1, every element of $\operatorname{Span}(S)$ has a unique expression of the form

$$a_1v_1 + a_2v_2$$

where $a_1, a_2 \in \mathbb{Z}/3\mathbb{Z}$. Hence, $|\operatorname{Span}(S)| = 3^2 = 9$. Since $\operatorname{Span}(S) \subseteq W$ and $|\operatorname{Span}(S)| = |W| = 9$, it follows that $\operatorname{Span}(S) = W$.

Proposition 3. If B is a basis for V, then every element of V can be expressed uniquely as a linear combination of elements of B.

Proof. Since B is linearly independent, we've already seen that every element in $\operatorname{Span}(B)$ can be written uniquely as a linear combination of elements of B. Since B is a basis, $\operatorname{Span}(B) = V$. $\square$

Definition. Let $B = \langle v_1, \dots, v_n\rangle$ be an ordered basis for V. Given $v \in V$, there are unique $a_1, \dots, a_n \in F$ such that

$$v = a_1v_1 + \cdots + a_nv_n.$$

The *coordinates of v with respect to the basis B* are the components of the vector $(a_1, \dots, a_n) \in F^n$. We write

$$[v]_B = (a_1, \dots, a_n).$$

Examples.

(a) Let $v = (x,y,z) \in F^3$. The coordinates of v with respect to the standard ordered basis $B = \langle e_1, e_2, e_3\rangle$ are (x,y,z) since

$$(x,y,z) = x(1,0,0) + y(0,1,0) + z(0,0,1) = xe_1 + ye_2 + ze_3.$$

Now consider $B' = \langle (1,0,0), (1,1,0), (1,1,1) \rangle$. Then the coordinates of v with respect to B' are $(x-y, y-z, z)$ since

$$(x,y,z) = (x-y)(1,0,0) + (y-z)(1,1,0) + z(1,1,1).$$

(b) Recall the ordered basis $\langle M_1, \ldots, M_6 \rangle$ for $M_{2\times 3}(F)$ defined earlier. Then the coordinates of the matrix

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

are $(a,b,c,d,e,f) \in F^6$.

Week 4, Monday: Dimension I

Recall the following from last time:

- A set B is a *basis* for V if it
  - is linearly independent, and
  - spans V.
- If B is a basis for V, each element of V can be expressed uniquely as a linear combination of vectors in B.
- If $B = \langle v_1, \ldots, v_n \rangle$ is an ordered basis for V, then the *coordinates of* $v \in V$ *with respect to* B are $(a_1, \ldots, a_n)$ where

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Example. Find the coordinates of $(7, -6) \in \mathbb{R}^2$ with respect to the ordered basis $B = \langle (5, 3), (1, 4) \rangle$.

Solution. We need to find $a, b \in \mathbb{R}$ such that

$$(7, -6) = a(5, 3) + b(1, 4).$$

Therefore, we solve the system of equations

$$\begin{aligned} 5a + b &= 7 \\ 3a + 4b &= -6. \end{aligned}$$

Applying our algorithm yields $a = 2$ and $b = -3$. So the coordinates of $(7, -6)$ with respect to B are given by $(2, -3)$. We write

$$[(7, -6)]_B = (2, -3).$$

Figure 9.1 gives the geometry. The basis vectors are in blue, and the red vectors indicate how $(7, -6)$ is a linear combination of the basis vectors.

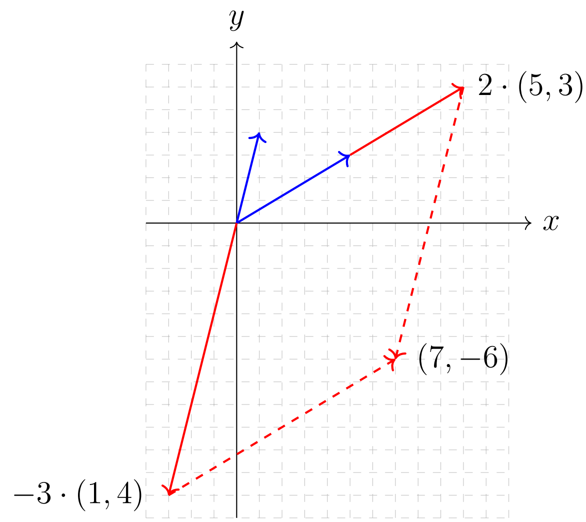


Figure 9.1: The coordinates of $(7,-6)$ with respect to the ordered basis $\langle (5,3),(1,4)\rangle$.

Remark. Let $B = \langle v_1, \dots, v_n\rangle$ be an ordered basis for a vector space V. Then taking coordinates defines a bijective (why?) function

$$\begin{aligned} \phi\colon V &\to F^n \\ v &\mapsto [v]_B. \end{aligned}$$

This function has an important property: it *preserves linear structure*. By this, we mean the following: let $u, v \in V$ and let $\lambda \in F$, then we claim that

$$\phi(u + \lambda v) = \phi(u) + \lambda\phi(v). \tag{9.1}$$

Note that addition and scalar multiplication happens in V on the left-hand side of this equation, and they happen in F^n on the right-hand side. The fact that ϕ is bijective and preserves linear structure means that as vector spaces V and F^n are "essentially the same". We can be more precise when we introduce linear transformations next week. For now, let us prove that equation (9.1) holds. We express u and v in terms of the basis:

$$\begin{aligned} u &= a_1v_1 + \cdots + a_nv_n \\ v &= b_1v_1 + \cdots + b_nv_n. \end{aligned}$$

It follows that

$$u + \lambda v = (a_1 + \lambda b_1)v_1 + \cdots + (a_n + \lambda b_n)v_n.$$

Then

$$\begin{aligned}
\phi(u + \lambda v) &= [u + \lambda v]_B \\
&= (a_1 + \lambda b_1, \ldots, a_n + \lambda b_n) \\
&= (a_1, \ldots, a_n) + \lambda(b_1, \ldots, b_n) \\
&= [u]_B + \lambda[v]_B \\
&= \phi(u) + \lambda\phi(v).
\end{aligned}$$

Definition. A vector space is *finite-dimensional* if it has a basis with a finite number of elements. If a vector space is not finite-dimensional, it is *infinite-dimensional*.

Examples. The following vector spaces are finite-dimensional:

- F^n (has a basis with n elements)
- $\mathcal{P}_d(F) = F[x]_{\leq d}$ (has a basis with $d + 1$ elements)
- $M_{m\times n}$ (has a basis with $m \times n$ elements)
- $\mathbb{C}$ as a vector space over $\mathbb{R}$ (basis $\{1, i\}$).

The following are infinite-dimensional:

- $\mathcal{P}(F) = F[x]$
- $\mathbb{R}^{\mathbb{R}} = \{f : \mathbb{R} \to \mathbb{R}\}$
- $\{f : \mathbb{R} \to \mathbb{R} : f \text{ is continuous}\}$
- $\{f : \mathbb{R} \to \mathbb{R} : f \text{ is differentiable}\}$
- $\mathbb{R}$ as a vector space over $\mathbb{Q}$
- $\mathbb{C}$ as a vector space over $\mathbb{Q}$.

Our goal today is to show that if V is a finite-dimensional vector space, then every basis for V has the same number of elements. Thus, the following definition makes sense:

Definition. If V is a finite-dimensional vector space, then the *dimension* of V, denoted $\dim V$ or $\dim_F V$, if we want to make the scalar field explicit, is the number of elements in any of its bases.

Exchange Lemma. Suppose $B = \{v_1, \ldots, v_n\}$ is a basis for a vector space V over a field F. Further, suppose that

$$w = a_1 v_1 + \cdots + a_n v_n \in V \qquad (\star)$$

with $a_i \in F$, and such that $a_\ell \neq 0$ for some $\ell \in \{1, \ldots, n\}$. Let B' be the set of vectors obtained from B by exchanging w for v_ℓ, i.e., $B' := (B \setminus \{v_\ell\}) \cup \{w\}$. Then B' is also a basis for V.

Proof. We first show that B' is linearly independent. For ease of notation, we may assume that $\ell = 1$, i.e., that $a_1 \neq 0$. Suppose we have a linear relation among the elements of B':

$$bw + b_2 v_2 + \cdots + b_n v_n = 0$$

Substituting for w:

$$0 = b(a_1 v_1 + \cdots + a_n v_n) + b_2 v_2 + \cdots + b_n v_n = ba_1 v_1 + (ba_2 + b_2)v_2 + \cdots + (ba_3 + b_n)v_n.$$

Since the v_i are linearly independent,

$$ba_1 = ba_2 + b_2 = \cdots = ba_n + b_n = 0.$$

Since $a_1 \neq 0$, it follows that $b = 0$ and then that $b_2 = \cdots = b_n = 0$, as well. Therefore, B' is linearly independent.

We now show that B' spans V. First, solve for v_1 in $(\star)$:

$$v_1 = \frac{1}{a_1} w - \frac{a_2}{a_1} v_2 - \cdots - \frac{a_n}{a\ _n}.$$

To see that B' spans, take $v \in V$. Since B is a basis, v can be written as a linear combination of $B = \{v_1, \ldots, v_n\}$, but then substituting the above expression for v_1 will express v as a linear combination of $B' = \{w, v_2, \ldots, v_n\}$, as required:

$$\begin{aligned}
v &= c_1 v_1 + \cdots + c_n v_n \\
&= \left(\frac{1}{a_1} w - \frac{a_2}{a_1} v_2 - \cdots - \frac{a_n}{a\ _n} v_n \right) + c_2 v_2 + \cdots + c_n v_n \\
&= \frac{1}{a_1} w + \left(-\frac{a_2}{a_1} + c_2 \right) v_2 + \cdots + \left(-\frac{a_n}{a_1} + c_n \right) v_n.
\end{aligned}$$

$\square$

Corollary. Suppose $B = \{v_1, \ldots, v_n\}$ is a basis for a vector space V over a field F. Further, suppose that $w \in V$ is nonzero. Then there exists $\ell \in \{1, \ldots, n\}$ such that $B' := (B \setminus \{v_\ell\}) \cup \{w\}$ is also a basis for V.

Theorem. In a finite-dimensional vector space, every basis has the same number of elements.

Proof. Let V be a finite-dimensional vector space. Among all the bases for V, let $B = \{u_1, \ldots, u_n\}$ be one of minimal size. Since B has minimal size, we know that $n = |B| \leq |C|$. Therefore C contains at least n distinct vectors $w_1, \ldots, w_n$ and possibly more. (Our goal is to show that, in fact, C contains no others.)

To take care of a trivial case, suppose $B = \emptyset$ (the case $n = 0$). In that case, we have

$$V = \operatorname{Span}(C) = \operatorname{Span}(B) = \operatorname{Span}(\emptyset) = \left\{\vec{0}\right\}.$$

The only linearly independent set whose span is $\left\{\vec{0}\right\}$ is $\emptyset$. So in this case, $0 = |C| = |B|$, as desired.

Now suppose that $n \geq 1$. We would again like to show that C has the same number of elements as B. The idea is to start with B, then use the exchange lemma to swap in the n elements $w_1, \ldots, w_n$ from C, one at a time, maintaining a basis at each step. To that end, let $B_0 = B$ and consider $w_1 \in C$. By the exchange lemma, we get a new basis B_1 by swapping w_1 with some $u_\ell \in B_0$. For ease of notation, let's suppose that $\ell = 1$. Therefore, $B_1 = \{w_1, u_2, \ldots, u_n\}$. Since B_1 is a basis for V, it is linearly independent and $V = \operatorname{Span}(B_1) = \operatorname{Span}(B) = \operatorname{Span}(C)$.

Next, consider $w_2 \in C$. Since B_1 is a basis, we know $w_2 \in \operatorname{Span}(B_1)$, hence, we can write

$$w_2 = a_1 w_1 + a_2 u_2 + \ldots a_n u_n$$

for some $a_i \in F$. Since w_1 and w_2 are linearly independent, at least one of $a_2, \ldots, a_n$ is nonzero. Without loss of generality, suppose $a_2 \neq 0$. Then by the exchange algorithm, $B_3 := \{w_1, w_2, u_3, \ldots, u_n\}$ is a basis. Continuing in this way, we eventually reach the basis $B_n = \{w_1, \ldots, w_n\} \subseteq C$. In fact, we must have $B_n = C$. Otherwise, there is a $w \in C \setminus B_n$. Since B_n is a basis, $w \in \operatorname{Span}(B_n)$, in other words, $w = \sum_{i=1}^n d_i w_i$ for some $d_i \in F$. But that can't happen since C is a basis: it's elements are linearly independent. So, in fact, C also has n elements. $\square$

Remark. If V is infinite-dimensional, it turns out that any two bases have the same cardinality. The above proof does not work to prove that, though.

Week 4, Wednesday: Dimension II

Last time, we showed that if V is finite-dimensional, then all of its bases have the same number of elements. Then, by definition, the number of elements in any basis for V is the *dimension* of V.

Examples.

- $\dim F^n = n$ (for instance, $\{e_1, \ldots, e_n\}$ is a basis).
- $\dim \mathcal{P}_d(F) = \dim F[x]_{\leq d} = d + 1$ (for instance, $\{1, x, \ldots, x^d\}$ is a basis).
- $\dim\{(x, y, z) \in F^3 : x + y + z = 0\} = 2$ (for instance, $\{(1, 0, -1), (0, 1, -1)\}$ is a basis).
- $\dim_{\mathbb{R}} \mathbb{C} = 2$ (for instance, $\{1, i\}$ is a basis).
- $\dim_{\mathbb{C}} \mathbb{C} = 1$ (for instance, $\{1\}$ is a basis).
- $\dim\{\vec{0}\} = 0$ (the basis is $\emptyset$, which has 0 elements).

Corollary. Let V be a vector space of dimension n. Then

(a) If $S \subseteq V$ is linearly independent, then S has at most n elements.

(b) If $S \subseteq V$ is linearly independent, then S can be completed to a basis for V, i.e., there exists a basis containing S as a subset.

(c) If S has n elements, then S is linearly independent if and only if it spans V.

(d) If S spans V, then S has at least n elements.

(e) A basis is a minimal spanning set for V. (Here, "minimal" can mean the set has no strict subsets that also span V, or it can mean minimal in number of elements.)

(f) A basis is a maximal linearly independent subset of V. (Here, "maximal" can mean there is no strict superset that is also linearly independent, or it can mean maximal in number.)

Example. Before proving the Corollary, here is an example of its use. Prove that $\{(5,3),(1,4)\}$ is a basis for $\mathbb{R}^2$. Since $\dim \mathbb{R} = 2$, we just need to check that these two vectors are linearly independent (by part (c)). Since neither is a scalar multiple of the other, we are done.

Proof. (a) Here we repeat the key idea of the proof from last time showing that all bases have the same number of elements. If $S = \emptyset$, we are done. Otherwise, say $S = \{s_1, \ldots, s_k\}$ for some $k \geq 1$. We know that V has some basis $C = \{v_1, \ldots, v_n\}$. Since $V = \operatorname{Span}(C)$, we can write

$$s_1 = a_1 v_1 + \cdots + a_n v_n.$$

Since S is linearly independent, $s_1 \neq \vec{0}$, and hence, some $a_i \neq 0$. Without loss of generality, say $a_1 \neq 0$. By the exchange lemma, we can swap s_1 for v_1 in C to get a new basis $C' = \{s_1, v_2, v_3, \ldots, v_n\}$.

If $k \geq 2$, since C' is a basis, we can write

$$s_2 = b_1 s_1 + b_2 v_2 + \cdots + b_n v_n.$$

Since s_1 and s_2 are linearly independent, at least one of $b_2, \ldots, b_n$ is nonzero. For convenience, say $b_2 \neq 0$. By the exchange lemma, the set $C'' = \{s_1, s_2, v_3, \ldots, v_n\}$ is a basis. We can repeat this process until all elements of S have been swapped into C, thus showing that $k \leq n$, as required.

(b) If $V = \operatorname{Span}(S)$, we are done. If not, take $v \in V \setminus \operatorname{Span}(S)$. By an earlier result, $S \cup \{v\}$ is linearly independent. We can repeat this process, but once we reach n elements, the process stops by part (a).

(c) ($\Rightarrow$) Suppose that S is linearly independent. By part (b), we can complete S to a basis B. Since $\dim V = n$, we know that $|B| = n$. So we have $S \subseteq B$ and $|S| = |B| = n$. It follows that $S = B$ is a basis, and hence, it spans V.

($\Leftarrow$) Suppose $V = \operatorname{Span}(S)$. We saw in an earlier lecture that there is a linearly subset of S' of S with the same span as S. Since S' is linearly independent and $V = \operatorname{Span}(S) = \operatorname{Span}(S')$, it follows that S' is a basis, and hence $|S'| = \dim V = n$. Since $S' \subseteq S$ and $n = |S'| = |S|$, it follows that $S' = S$, and therefore, S is linearly independent.

(d) If S is infinite, there is nothing to prove. Otherwise, by removing elements from S we can find a linearly independent subset $S' \subseteq S$ with the same span. Then S' is a basis for V and hence has n elements. Since $S' \subseteq S$, we have $n = |S'| \leq |S|$.

(e) HW.

(f) HW.

Example. Prove that $\{(3,1,2),(1,0,-1),(-1,2,4),(1,3,0)\} \subset \mathbb{R}^3$ is linearly dependent.

Solution. Since $\dim \mathbb{R}^3 = 3$, a linearly independent set has at most 3 elements. $\square$

Extra time activity. Let $F = \mathbb{Z}/3\mathbb{Z}$, and consider the following twelve points in F^4:

$$\begin{array}{ccc} (1,1,2,1) & (1,1,2,0) & (2,1,2,1) \\ (1,1,0,1) & (2,0,1,0) & (1,0,1,1) \\ (2,1,1,0) & (1,2,0,0) & (1,2,2,1) \\ (1,2,0,1) & (2,0,1,1) & (0,0,2,2) \end{array}$$

Goal: find subsets of size three of this array that sum to $(0,0,0,0)$.

All solutions:

- $(1,1,2,1),(1,0,1,1),(1,2,0,1)$
- $(1,1,0,1),(1,0,1,1),(1,2,2,1)$
- $(2,1,1,0),(1,2,0,1),(0,0,2,2)$

Observations:

- Three vectors sum to zero if and only if in each component, the entries are either all the same or all different. For example, in the solution $(2,0,0,1),(2,0,1,2),(2,0,2,0)$, the entries in the first component are all 2, the entries in the second component are all 0, the entries in the third and fourth components are $0,1,2$—all different.
- If u,v,w is a solution so that $u+v+w=0$, consider the value of

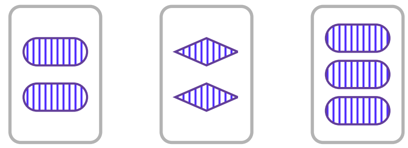
$$u + t(v-u)$$

as t varies among the element of F. When $t = 0$, we get u. When $t = 1$, we get v, and when $t = 2$, we get

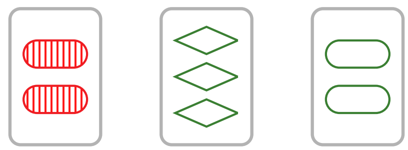
$$u + 2(v - u) = -u + 2v = -u - v = w,$$

recalling that $2 = -1$ in $F = \mathbb{Z}/3\mathbb{Z}$. We may think of $t(v - u)$ as determining a line through the origin as t varies. So then $u + t(v - u)$ is that line translated by the vector u. So finding these triples of points whose sum is zero is the same as finding lines in F^4 containing the three points.

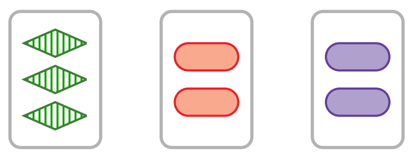
Relation to the game Set (number-1, shading, color, shape):



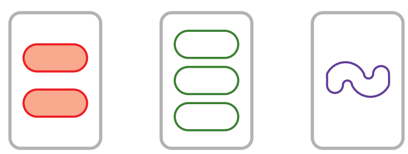
$(1,1,2,1)$ $(1,1,2,0)$ $(2,1,2,1)$



$(1,1,0,1)$ $(2,0,1,0)$ $(1,0,1,1)$



$(2,1,1,0)$ $(1,2,0,0)$ $(1,2,2,1)$



$(1,2,0,1)$ $(2,0,1,1)$ $(0,0,2,2)$

Week 4, Friday: Row and column spaces

Row rank and column rank.

Definition. Let A be an $m \times n$ matrix over F. The *row space* of A is the subspace of F^n spanned by its rows, and the *column space* of A is the subspace of F^m spanned by its columns. The *row rank* of A is the dimension of its row space, and the *column rank* of A is the dimension of its column space.

Since row operations are reversible, any matrix obtained from a matrix A by performing row operations has the same row space. In particular, the row space of A is the same as the row space of its reduced echelon form. From the structure of the reduced echelon form, it's clear that its nonzero rows form a basis for its row space. To summarize:

> **The nonzero rows of the reduced echelon form of A form a basis for the row space of A.**

This gives an algorithm for computing a basis for the row space of a matrix.

Algorithm for computing a basis for the row space and the row rank. Given an $m \times n$ matrix A, compute its reduced echelon form E. Then the rows of E are a basis for the row space of A. The number of nonzero rows in E is the row rank of A.

Example. Let

$$A = \begin{pmatrix} 1 & 2 & 0 & 4 \\ 3 & 3 & 1 & 0 \\ 7 & 8 & 2 & 4 \end{pmatrix}.$$

To compute a basis for the row space of A, compute its reduced echelon form:

$$A = \begin{pmatrix} 1 & 2 & 0 & 4 \\ 3 & 3 & 1 & 0 \\ 7 & 8 & 2 & 4 \end{pmatrix} \longrightarrow E = \begin{pmatrix} 1 & 0 & \frac{2}{3} & -4 \\ 0 & 1 & -\frac{1}{3} & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

So a basis for the row space of A is:

$$\left\{\left(1, 0, \tfrac{2}{3}, -4\right), \left(0, 1, -\tfrac{1}{3}, 4\right)\right\}.$$

Proposition. Let A be an $m \times n$ matrix with columns $A_1, \dots, A_n \in F^m$. Let $\tilde{A}$ be any matrix formed from A by performing row operations, and let $\tilde{A}_1, \dots, \tilde{A}_n \in F^m$ be its columns. Let $x_1, \dots, x_n \in F$ be any scalars. Then

$$x_1 A_1 + \cdots + x_n A_n = 0 \quad \text{if and only if} \quad x_1 \tilde{A}_1 + \cdots + x_n \tilde{A}_n = 0.$$

Proof. Write out the relation $x_1 A_1 + \cdots + x_n A_n = 0$ longhand:

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = 0.$$

Adding up the left-hand side, we see the relation is equivalent to a solution $(x_1, \dots, x_n)$ to the linear system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0. \end{aligned}$$

Or result follows since row operations do not change the set of solutions to a system of equations. $\square$

Corollary. Let E be the reduced row echelon form of a matrix A, and suppose the basic (pivot) columns have indices $j_1, \dots, j_r$. Then the columns of A indexed by $j_1, \dots, j_r$ form a basis for the column space of A.

Proof. For ease of notation, assume $j_1 = 1, j_2 = 2, \dots, j_r = r$, i.e., the first r columns of E are the pivot columns. For instance, in the case $m = 5$, $n = 7$, and $r = 3$, the matrix E would have the form

$$\begin{pmatrix} 1 & 0 & 0 & * & * & * & * \\ 0 & 1 & 0 & * & * & * & * \\ 0 & 0 & 1 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where the $*$s are arbitrary scalars.

Let $E_1, \dots, E_n$ denote the columns of E, and let $A_1, \dots, A_n$ denote the columns of A. It is clear that $E_1, \dots, E_r$ form a basis for the columns space of E. We need to show that $A_1, \dots, A_r$ form a basis for the columns space of A. So we need to show $A_1, \dots, A_r$

are linearly independent and span the column space of A. For linear independence, suppose that

$$x_1A_1 + \cdots + x_rA_r = 0.$$

for some $x_i \in F$. Then, by the Proposition,

$$x_1E_1 + \cdots + x_rE_r = 0.$$

Since $E_1, \ldots, E_r$ are linearly independent, it follows that $x_1 = \cdots = x_r = 0$, as desired. Next, to show $A_1, \ldots, A_r$ span the column space of A, it suffices to show that every other column of A is in the span. So consider a column A_j with $j > r$. Since $E_1, \ldots, E_r$ form a basis for the column space of E, we can find scalars $c_1, \ldots, c_r$ such that

$$E_j = c_1E_1 + \cdots + c_rE_r.$$

Rewriting this equation, we get

$$c_1E_1 + \cdots + c_rE_r - E_j = 0.$$

It then follows from the Proposition that

$$c_1A_1 + \cdots + c_rA_r - A_j = 0,$$

which implies

$$A_j = c_1A_1 + \cdots + c_rA_r - A_j.$$

So A_j is in the span of $A_1, \ldots, A_r$. □

We turn the Corollary into an algorithm:

Algorithm for computing a basis for the column space and the column rank. Given a matrix A, compute its reduced echelon from E. Say that columns $j_1, \ldots, j_r$ are the basic columns of E (those corresponding to the non-free variables—the one that have a single non-zero entry and that entry is equal to 1. Then columns $j_1, \ldots, j_r$ are a basis for the columns space of A. The column rank of A is r, the number of basic columns of its reduced echelon form.

WARNING: Be sure to take columns $j_1, \ldots, j_k$ of the orginal matrix, A, not of the echelon form, E. (So computing a basis for the row space is little easier, since it does not require this last step.)

Example: In the previous example, we computed the reduced echelon form of a matrix:

$$A = \begin{pmatrix} 1 & 2 & 0 & 4 \\ 3 & 3 & 1 & 0 \\ 7 & 8 & 2 & 4 \end{pmatrix} \longrightarrow E = \begin{pmatrix} 1 & 0 & \frac{2}{3} & -4 \\ 0 & 1 & -\frac{1}{3} & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The first two columns of E are its basic columns. Therefore, the first two columns of A form a basis for its column space:

$$\begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 8 \end{pmatrix}.$$

NOTE: The first two columns of E in this case are the first two standard basis vectors, which clearly don't have the same span as the above two vectors.

A consequence of our discussion above is the following, rather surprising, result:

Theorem. The row rank of a matrix A is equal to its column rank.

Proof. Let E be the reduced echelon form of A. Then the number of its nonzero rows is equal to the number of its basic columns. □

Definition. The *rank* of a matrix A, denoted $\text{rank}(A)$ is the dimension of its row space or column space.

Suppose we have a homogeneous system of linear equations

$$\begin{aligned} a_{11}x_{11} + \cdots + a_{1n}x_n &= 0 \\ \vdots \quad \vdots \quad &\vdots \\ a_{m1}x_m + \cdots + a_{mn}x_n &= 0. \end{aligned}$$

Let $A = (a_{ij})$ be the matrix of coefficients. To solve the system, we compute the reduced echelon form of the matrix A. The number of free parameters for the solution space is then the number of non-basic columns, i.e., $n - \text{rank}(A)$. There is a unique solution $\vec{0}$ exactly when the reduced echelon form is the matrix with 1s along its diagonal and 0s, otherwise, i.e., exactly when there are no non-basic columns. Hence, there is only the trivial solution if and only if $\text{rank}(A) = n$.

For a non-homogeneous system

$$\begin{aligned} a_{11}x_{11} + \cdots + a_{1n}x_n &= b_1 \qquad (11.1) \\ \vdots \quad \vdots \quad &\vdots \\ a_{m1}x_m + \cdots + a_{mn}x_n &= b_n. \end{aligned}$$

we would compute the echelon of the augmented matrix $[A|b]$ where b is the column with entries $b_1, \ldots, b_n$. If the system is consistent, we have seen that the set of solutions consists of any particular solution plus any vector in the span of $n - \text{rank}(A)$

vectors that are solutions to the corresponding homogeneous system. So if the system is consistent, there is a unique solution if and only if $\operatorname{rank}(A) = n$.

Summary. The system (11.1), above, has a unique solution if and only if it is consistent and $\operatorname{rank}(A) = n$. In the case $b_1 = \cdots = b_n = 0$, the system is homogeneous and, thus, consistent ($x_1 = \cdots = x_n = 0$ is a solution). So in the homogeneous case, there is a unique solution if and only if $\operatorname{rank}(A) = n$.

Week 5, Monday: Linear transformations

Linear transformations. We have now defined the objects of study—vector spaces. Next, we need to consider the appropriate mappings between those objects—those that preserve the linear structure.

Definition. Let V and W be vector spaces over a field F. A *linear transformation* from V to W is a function

$$f\colon V \to W$$

satisfying, for all $v, v' \in V$ and $\lambda \in F$,

$$f(v + v') = f(v) + f(v') \quad \text{and} \quad f(\lambda v) = \lambda f(v).$$

Remarks. Using the notation from the definition:

- If $f(v+v') = f(v) + f(v')$, we say f *preserves addition*. Note that the addition on the left side is in V and the addition on the right side is in W. Thus, if $V \neq W$, they are two different operations (with the same name). Similarly, if $f(\lambda v) = \lambda f(v)$, we say f *preserves scalar multiplication*.

- One may combine the two conditions, above, for linearity into one: for f to be linear, we require
$$f(v + \lambda v') = f(v) + \lambda f(v')$$
for all $v, v' \in V$ and $\lambda \in F$.

- Synonyms for "linear transformation" are: "linear mapping" and "linear homomorphism", often with the word "linear" dropped when clear from context (and it will be since this is a course in linear algebra!).

- Our book restricts "linear transformation" to mean a linear transformation of the form $f\colon V \to V$, where the domain and codomain are equal. That is non-standard, and we won't use that terminology. Linear mappings from a vector space to itself are called linear *endomorphisms* or linear *self-mappings*.

Template for a proof that a mapping is linear. Consider the function

$$\begin{aligned} f\colon \mathbb{R}^3 &\to \mathbb{R}^2 \\ (x, y, z) &\mapsto (2x + 3y, x + y - 3z). \end{aligned}$$

Claim: f is linear.

Proof. Let $(x, y, z), (x', y', z') \in \mathbb{R}^3$ and $\lambda \in R$.

$$\begin{aligned} f((x, y, z) + (x', y', z')) &= f(x + x', y + y', z + z') \\ &= (2(x + x') + 3(y + y'), (x + x') + (y + y') - 3(z + z')) \\ &= ((2x + 3y) + (2x' + 3y'), (x + y - 3z) + (x' + y' - 3z')) \\ &= (2x + 3y, x + y - 3z) + (2x' + 3y', x' + y' - 3z') \\ &= f(x, y, z) + f(x', y', z'). \end{aligned}$$

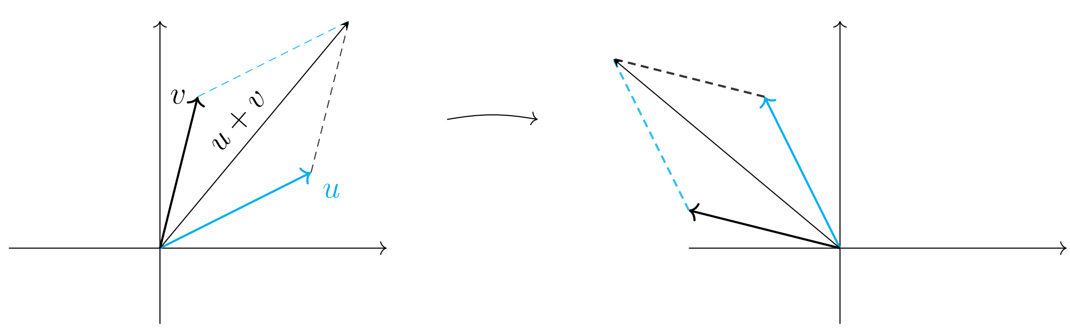
Thus, f preserves addition. Next,

$$\begin{aligned} f(\lambda(x, y, z)) &= f(\lambda x, \lambda y, \lambda z) \\ &= (2(\lambda x) + 3(\lambda y), (\lambda x + \lambda y - (3\lambda z)))) \\ &= (\lambda(2x + 3y), \lambda(x + y - 3z)) \\ &= \lambda(2x + 3y, x + y - 3z) \\ &= \lambda f(x, y, z). \end{aligned}$$

Thus, f preserves scalar multiplication. □

Note: People sometimes confuse proofs that subsets are subspaces with proofs that mappings are linear. To prove that $W \subseteq V$ is a subspace, we show that W is *closed under* addition and scalar multiplication by taking $u, v \in W$ and $\lambda \in F$ and showing $u + \lambda v \in W$. To prove $f\colon V \to W$ is linear, we show that f *preserves* addition and scalar multiplication. Be careful not to confuse the words "closed under" with "preserves".

Example. Rotation about the origin in the plane $\mathbb{R}^2$ is a linear transformation:



Exercise. Show that $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not linear.

Proof. We have $f(1+1) = f(2) = 4 \neq f(1) + f(1) = 1 + 1 = 2$. $\square$

The following proposition is often useful for showing a function is not linear.

Proposition 1. If $f\colon V \to W$ is linear, then $f(\vec{0}_V) = \vec{0}_W$.

Proof. Since f is linear,

$$f(\vec{0}_V) = f(0 \cdot \vec{0}_V) = 0 \cdot f(\vec{0}_V) = \vec{0}_W.$$

Thus, for instance,

$$\begin{aligned} f\colon \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto x + 2y + 5 \end{aligned}$$

is not linear since $f(0,0) = 5 \neq 0$.

Proposition 2. (A linear mapping is determined by its action on a basis.) Let V and W be vector spaces over F, and let B be a basis for V. For each $b \in B$, let $w_b \in W$. Then there exists a unique linear function $f\colon V \to W$ such that $f(b) = w_b$.

Proof. We define f as follows: Given $v \in V$, since B is a basis, we can write $v = \alpha_1 b_1 + \cdots + \alpha_k b_k$ for some $\alpha_i \in F$, $b_i \in B$, and $k \in \mathbb{Z}_{\geq 0}$. Define

$$f(v) := \alpha_1 f(b_1) + \ldots \alpha_k f(b_k) = \alpha_1 w_{b_1} + \cdots + \alpha_k w_{b_k}.$$

Since B is a basis, the expression for v as a linear combination of elements in B is unique. Hence, f is well-defined. Further, linearity of f forces us to define $f(v)$ as we have. To see that f is linear, let $v, w \in V$ and $\lambda \in \mathbb{R}$. Write v and w as linear combinations of the basis vectors:

$$\begin{aligned} v &= \alpha_1 b_1 + \cdots + \alpha_k b_k \\ w &= \beta_1 b_1 + \cdots + \beta_k b_k \end{aligned}$$

for some scalars α_i and β_i. It follows that

$$v + \lambda w = (\alpha_1 + \lambda\beta_1) b_1 + \cdots + (\alpha_k + \lambda\beta_k) b_k.$$

Using the definition of f, we see

$$\begin{aligned} f(v + \lambda w) &= (\alpha_1 + \lambda\beta_1) w_{b_1} + \cdots + (\alpha_k + \lambda\beta_k) w_{b_k} \\ &= (\alpha_1 w_{b_1} + \cdots + \alpha_k w_{b_k}) + \lambda(\beta_1 w_{b_1} + \cdots + \beta_k w_{b_k}) \end{aligned}$$

$$= f(v) + \lambda f(w).$$

□

Terminology. We say the function f as in Proposition 2 has been defined on B then *extended linearly* to all of V.

Example. Define a linear function $f\colon \mathbb{R}^2 \to M_{2\times 3}(\mathbb{R})$ by

$$f(1,0) = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 2 \end{pmatrix} \quad \text{and} \quad f(0,1) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}.$$

What is $f(2,-1)$?

Solution. In general, we have

$$\begin{aligned} f(x,y) &= f\left(x(1,0) + y(0,1)\right) \\ &= xf(1,0) + yf(0,1) \\ &= x\begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 2 \end{pmatrix} + y\begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix} \\ &= \begin{pmatrix} x+2y & y & 2x \\ 3x & -x+3y & 2x+y \end{pmatrix}. \end{aligned}$$

In particular,

$$f(2,-1) = 2\begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 2 \end{pmatrix} - \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 4 \\ 6 & -5 & 3 \end{pmatrix}.$$

Question. What goes wrong if we try to define a linear function by specifying its values on a non-basis? For instance, what happens if we try to define a linear function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ by specifying the values for the non-basis $\{(1,0),(2,0)\}$ as follows:

$$f(1,0) = (3,2) \quad \text{and} \quad f(2,0) = (1,1).$$

Note. Let V and W be vector spaces over F, and let X be a linearly subset of V. For each $x \in X$, let $w_x \in W$. Then there exists a linear function $f\colon V \to W$ such that $f(x) = w_x$ for all $x \in W$. To see this, let B be any completion of X to a basis for V, and apply Proposition 2. The map created this way is not unique: we are free to choose any values for elements of $B \setminus X$ (the value $\vec{0}$ might be a natural choice).

Here is something interesting that we will talk more about later:

Definition. Let V and W be vector spaces over F. The collection of all linear functions from V to W is denoted $\mathrm{Hom}(V,W)$ or $\mathcal{L}(V,W)$. It is a vector space over F under addition and scalar multiplication of functions: for linear $f,g\colon V\to W$,

$$\begin{aligned} f+\lambda g\colon V &\to W \\ v &\mapsto f(v)+\lambda g(v). \end{aligned}$$

Week 5, Wednesday: Range and nullspace

Recall the definition of a linear function from last time: a function $f\colon V \to W$ between vectors spaces V and W over the (same) field F is a function $f\colon V \to W$ that preserves addition and scalar multiplication. In detail, this means that for all $u, v \in V$ and $\lambda \in F$,

$$f(u+v) = f(u) + f(v) \quad \text{and} \quad f(\lambda v) = \lambda f(v).$$

Definition/Proposition 1. Suppose $f\colon V \to W$ is linear and $U \subseteq V$ is a subspace of V. Then the *image of U under f* is the set

$$f(U) := \{f(u) : u \in U\} \subseteq W.$$

The image of U under f is a subspace of W.

Proof. Since U is a subspace of V, it follows that $0_V \in U$, and hence, $f(0_V) = 0_W \in f(U)$. Thus, $f(U)$ is nonempty. Next, let $x, y \in f(U)$, and let $\lambda \in F$. By definition of $f(U)$, there are vectors $u, v \in U$ such that $f(u) = x$ and $f(v) = y$. Then since f is linear, is preserves addition and scalar multiplication. Therefore,

$$\begin{aligned} x + \lambda y &= f(u) + \lambda f(v) \\ &= f(u) + f(\lambda v) \\ &= f(u + \lambda v). \end{aligned}$$

Since U is a subspace, it is closed under addition and scalar multiplication. Therefore, $u + \lambda v \in U$. It follows that $x + \lambda y = f(u + \lambda v) \in f(U)$, as required. $\square$

In particular, since V is a subspace of itself, its image under a linear function is a subspace of the codomain of the function.

Definition. The *image* or *range* of a linear function $f\colon V \to W$ is the subspace

$$\mathrm{im}(f) := \mathcal{R}(f) := f(V) := \{f(v) : v \in V\} \subseteq W.$$

The dimension of the image of f is the *rank* of f (provided it is finite-dimensional) and is denoted $\operatorname{rank}(f)$ or $\operatorname{rk}(f)$.

Example. Define a linear function $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ by letting $f(1,0) = (2,1,0)$ and $f(0,1) = (0,-1,1)$ and extending linearly. Thus, for all $x, y \in \mathbb{R}$,

$$\begin{aligned} f(x,y) &= f(x(1,0) + y(0,1)) \\ &= xf(1,0) + yf(0,1) \\ &= x(2,1,0) + y(0,-1,1). \end{aligned}$$

We have

$$\operatorname{im}(f) = \mathcal{R}(f) = \operatorname{Span}\{(2,1,0),(0,-1,1)\}.$$

Since $(2,1,0)$ and $(0,-1,1)$ are linearly independent and span the image, they are a basis for the image of f, and thus, $\operatorname{rank}(f) = 2$.

Remark. If $f\colon V \to W$ is a linear function, and B is a basis for V, then

$$\operatorname{im}(f) = \operatorname{Span}(f(B)).$$

To see this, let $w \in \operatorname{im}(f)$. Then there exists $v \in V$ such that $w = f(v)$. Since B is a basis, there exists $b_1, \ldots, b_k \in B$ and $a_1, \ldots, a_k \in F$ such that $v = \sum_{i=1}^k a_i b_i$. Then since f is linear,

$$w = f(v) = f\left(\sum_{i=1}^k a_i b_i\right) = \sum_{i=1}^k a_i f(b_i) \in \operatorname{Span}(B).$$

Note, however, that $f(B)$ is not necessarily a basis for $\operatorname{im}(f)$.

Example. Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x,y) = (x,0)$, and let $B = \{(1,0),(0,1)\}$ be the standard basis for $\mathbb{R}^2$. Then

$$f(B) = \{f(1,0), f(0,1)\} = \{(1,0),(0,0)\}.$$

Although $f(B)$ spans $\operatorname{im}(f)$, it is not linearly independent and is thus not a basis for $\operatorname{im}(f)$.

Definition/Proposition 2. Let $f\colon V \to W$ be a linear mapping, and let U be a subspace of W. Then the *inverse image of U under f* is the set

$$f^{-1}(U) := \{v \in V : f(v) \in U\} \subseteq V.$$

The inverse image of U under f is a subspace of V.

Proof. Since U is a subspace of W, we know $0_W \in U$. Then, since $f(0_V) = 0_W$, it follows that $0_V \in f^{-1}(U)$. So $f^{-1}(U)$ is nonempty. Next, let $v, v' \in f^{-1}(U)$, and let $\lambda \in F$. It follows that $f(v) \in U$ and $f(v') \in U$. Since U is a subspace, it follows that $f(v) + \lambda f(v') \in U$. Since f is linear,

$$f(v + \lambda v') = f(v) + \lambda f(v') \in U.$$

It follows that $v + \lambda v' \in f^{-1}(U)$. $\square$

Definition. Let $f\colon V \to W$ be a linear mapping. The *kernel* or *nullspace* of f, denoted $\ker f$ or $\mathcal{N}(f)$, respectively, is the inverse image of $\{0_W\}$:

$$\ker(f) := \mathcal{N}(f) := f^{-1}(\{0_W\}) := \{v \in V : f(v) = 0\}\,.$$

It is a subspace of V (by Proposition 2). The dimension of the kernel is called the *nullity* of f (provided it is finite-dimensional) and is denoted $\operatorname{nullity}(f)$.

Example. Consider the linear mapping

$$\begin{aligned} f\colon \mathbb{R}^2 &\to \mathbb{R}^3 \\ (x, y) &\mapsto (2x, x - y, y). \end{aligned}$$

To find the kernel of f, we look for vectors (x, y) such that

$$f(x, y) = (2x, x - y, y) = (0, 0, 0).$$

Comparing vector components, we see that $x = y = 0$ is the only possibility. Therefore,

$$\ker(f) = \{(0, 0)\}\,,$$

and $\operatorname{nullity}(f) = 0$.

Example. Let $\mathbb{R}[x]_{\leq 2}$ denote polynomials in x of degree at most two and with real coefficients. Consider the linear mapping

$$\begin{aligned} f\colon \mathbb{R}[x]_{\leq 2} &\longrightarrow \mathbb{R}^2 \\ a + bx + cx^2 &\mapsto (a + b, a + c). \end{aligned}$$

To find the kernel of f, we need to find a, b, c such that $f(a + bx + cx^2) = (0, 0)$. This amounts to solving the system of equations

$$\begin{aligned} a + b &= 0 \\ a + c &= 0. \end{aligned}$$

Apply our algorithm:

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right). \tag{$\star$}$$

Solving for the pivot variables, we get

$$\begin{aligned} a &= -c \\ b &= c. \end{aligned}$$

Therefore,

$$\ker(f) = \left\{-c + cx + cx^2 : c \in \mathbb{R}\right\} = \operatorname{Span}\left\{-1 + x + x^2\right\}.$$

Therefore, the nullity of f is $\dim(\ker(f) = 1$. A basis for $\mathbb{R}[x]_{\leq 2}$ is the set $\{1, x, x^2\}$, and the image of these vectors forms a basis for the image of f:

$$f(1) = (1,1), \quad f(x) = (1,0), \quad f(x^2) = (0,1).$$

So the image of f is the column space of the matrix for the linear system we solved to find the kernel (cf. Equation ($\star$)). Using our algorithm for finding the basis of the column space, we get the basis $\{(1,1),(1,0)\}$. Another basis is $\{(1,0),(0,1)\}$. Therefore, the rank of f is $\operatorname{rank}(f) = 2$.

Our main goal next time will to prove the following:

Theorem. (Rank-nullity theorem) Let $f\colon V \to W$ be a linear mapping, and suppose that V is finite-dimensional. Then

$$\operatorname{rank}(f) + \operatorname{nullity}(f) = \dim V.$$

In other words, $\dim(\operatorname{im}(f)) + \dim(\ker(f)) = \dim V$.

Example. In the previous example, we found

$$\operatorname{rank}(f) + \operatorname{nullity}(f) = 2 + 1 = 3 = \dim \mathbb{R}[x]_{\leq 2}.$$

Week 5, Friday: Rank-nullity theorem; Isomorphisms

Let $f\colon V \to W$ be a linear mapping between vectors spaces V and W over a field F. Recall the definitions from last time:

Definition. The *kernel* or *null space* of f is

$$\mathcal{N}(f) := \ker(f) := f^{-1}(\{0_W\}) := \{v \in V : f(v) = 0\}.$$

The *nullity*[1] of f is the dimension of the kernel.

The *image* or *range* of f is

$$\mathcal{R}(f) = \operatorname{im}(f) = f(V) = \{f(v) \in W : v \in V\}.$$

The *rank* of f is the dimension of the image.

Theorem. (Rank-nullity theorem) Let $f\colon V \to W$ be a linear mapping, and suppose that V is finite-dimensional. Then

$$\operatorname{rank}(f) + \operatorname{nullity}(f) = \dim V.$$

In other words, the $\dim(\operatorname{im}(f)) + \dim(\ker(f)) = \dim V$.

Proof. Let $K = \{v_1, \dots, v_k\}$ be a basis for $\ker(f)$ (and therefore, $\operatorname{nullity}(f) = k$). Complete K to a basis for V:

$$B = \{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}.$$

To prove the theorem, it suffices to show that $\{f(v_{k+1}), \dots, f(v_n)\}$ is a basis for image(f). We first show linear independence. Suppose that

$$a_{k+1} f(v_{k+1}) + \cdots + a_n f(v_n) = 0_W.$$

[1]Don't confuse this concept with the *mullity* of f, defined as follows: $\operatorname{mullity}(f) = p(f) + b(f)$ where $p(f)$ is the amount of party of f in the back and $b(f)$ is the amount of business of f in the front.

Since f is linear, it follows that

$$f(a_{k+1}v_{k+1} + \cdots + a_n v_n) = a_{k+1}f(v_{k+1}) + \cdots + a_n f(v_n) = 0_W.$$

Therefore, $a_{k+1}v_{k+1} + \cdots + a_n v_n \in \ker(f)$. Since $K = \{v_1, \ldots, v_k\}$ is a basis for $\ker(f)$, there are scalars $a_1, \ldots, a_k$ such that

$$a_{k+1}v_{k+1} + \cdots + a_n v_n = a_1 v_1 + \cdots + a_k v_k,$$

i.e.,

$$a_1 v_1 + \cdots + a_k v_k - a_{k+1}v_{k+1} - \cdots - a_n v_n = 0_V.$$

This is a linear relation among the vectors of B, the basis we constructed for V. Since B is a linearly independent set, all of the a_i must be 0. In particular, $a_{k+1} = \cdots = a_n = 0$, as we were trying to show.

Next, we show that $\{f(v_{k+1}), \ldots, f(v_n)\}$ spans $\operatorname{im}(f)$. We know that since $B = \{v_1, \ldots, v_n\}$ is a basis for V that

$$\{f(v_1), \ldots, f(v_n)\}$$

spans the image of f. However, $v_1, \ldots, v_k$ are in $\ker(f)$, so

$$\begin{aligned}
\operatorname{im}(f) &= \operatorname{Span}\{f(v_1), \ldots, f(v_k), f(v_{k+1}), \ldots, f(v_n)\} \\
&= \operatorname{Span}\{0_W, \ldots, 0_W, f(v_{k+1}), \ldots, f(v_n)\} \\
&= \operatorname{Span}\{f(v_{k+1}), \ldots, f(v_n)\}.
\end{aligned}$$

□

Proposition 1. The linear mapping $f\colon V \to W$ is injective (i.e., one-to-one) if and only if $\ker(f) = \{0_V\}$.

Proof. ($\Rightarrow$) First suppose that f is injective, and let $v \in \ker(f)$. Therefore, $f(v) = 0_W$. We also know that since f is linear, $f(0_V) = 0_W$. So $f(v) = 0_W = f(0_V)$. Since f is injective and $f(v) = f(0_V)$, it follows that $v = 0_V$. We have shown that $\ker(f) = \{0_V\}$.

($\Leftarrow$) For the converse, now suppose that $\ker(f) = \{0_V\}$, and let $u, v \in V$ with $f(u) = f(v)$. It follows that $f(u - v) = f(u) - f(v) = 0_W$. Hence, $u - v \in \ker(f)$. However, we are assuming $\ker(f) = \{0_V\}$. So $u - v = 0_V$, which means $u = v$. Therefore, f is injective. □

Proposition 2. Let $S \subseteq V$.

(a) If S is linearly dependent, then $f(S) := \{f(s) : s \in S\} \subseteq W$ is linearly dependent. (The image of a dependent set is dependent.)

(b) If f is injective and S is linearly independent, then $f(S) \subseteq W$ is linearly independent. (The image of an independent set is independent *provided* f is injective.)

Proof. Suppose that $\sum_{i=1}^{k} a_i s_i = 0_V$ for some $a_i \in F$ and $s_i \in S$. Since f is linear, we have

$$0_W = f(0_V) = f(\textstyle\sum_{i=1}^{k} a_i s_i) = \sum_{i=1}^{k} a_i f(s_i).$$

Thus, f preserves linear dependencies, as claimed in part (a).

Suppose now that f is injective and S is linearly independent. If $\sum_{i=1}^{k} a_i f(s_i) = 0_W$ for some $a_i \in F$ and $s_i \in S$, then since f is linear,

$$0_W = \textstyle\sum_{i=1}^{k} a_i f(s_i) = f(\sum_{i=1}^{k} a_i s_i).$$

Therefore, $\sum_{i=1}^{k} a_i s_i$ is in the kernel of f. Since, f is injective, $\ker(f) = \{0_V\}$ by Proposition 1. It follows that $\sum_{i=1}^{k} a_i s_i = 0_V$. Then, since S is linearly independent, it follows that $a_i = 0$ for all i. This shows that $f(S)$ is linearly independent. $\square$

Definition. The linear function $f : V \to W$ is an *isomorphism* if there exists a linear function $g : W \to V$ such that $g \circ f = \mathrm{id}_V$ and $f \circ g = \mathrm{id}_W$. The function g is called the *inverse of* f.

Remark. Suppose that $f : V \to W$ is an isomorphism. Then, just as proved in Math 112 for mappings of sets, it follows that f is bijective, i.e., both injective and surjective. For mappings of sets, being bijective is equivalent to having an inverse. The same is true for mappings of vector spaces: A linear function $f : V \to W$ is an isomorphism if and only if it is bijective. It turns out that if a linear function is bijective, then its inverse mapping (as a mapping of sets) is automatically linear. (Check this for yourself.)

Example. The space of 2×2 matrices over F is isomorphic to F^4. One isomorphism is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto (a, b, c, d).$$

Exercise. Write $V \sim W$ if there is an isomophism $V \to W$. Check that $\sim$ is an equivalence relation.

Proposition 3. A linear mapping $f : V \to W$ is an isomorphism if and only if $\ker(f) = \{0_V\}$ and $\mathrm{im}(f) = W$, (i.e., if and only if its kernel is trivial and it is surjective).

Proof. We have just seen that $\ker f = \{0_V\}$ if and only if f is injective, and by definition of surjectivity, f is surjective if and only if $\mathrm{im}(f) = W$. Thus, the condition that $\ker(f)$ is trivial and $\mathrm{im}(f) = W$ is equivalent to the bijectivity of f. $\square$

Theorem 4. Let V be a vector space over F. Then V is isomorphic to F^n if and only if $\dim V = n$.

Proof. ($\Rightarrow$) Suppose that $f\colon V \to F^n$ is an isomorphism with inverse $g\colon F^n \to V$, and let $e_1, \ldots, e_n$ be the standard basis for F^n. Define $v_i = g(e_i) \in V$ for $i = 1, \ldots, n$. We claim that $B := \{v_1, \ldots, v_n\}$ is a basis for V (and hence, $\dim V = n$). First note that B is linearly independent by Proposition 2 (a) since $\{e_1, \ldots, e_n\}$ is linearly independent. Next, to see that B spans, let $v \in V$, and write

$$f(v) = \sum_{i=1}^{n} a_i e_i$$

for some $a_i \in F$. It follows that

$$v = g(f(v)) = g\left(\textstyle\sum_{i=1}^{n} a_i e_i\right) = \textstyle\sum_{i=1}^{n} a_i g(e_i) = \sum_{i=1}^{n} a_i v_i \in \mathrm{Span}(B).$$

($\Leftarrow$) Now suppose $\dim V = n$. Choose a basis $\{b_1, \ldots, b_n\}$ for V, and let $\{e_1, \ldots, e_n\}$ be the standard basis for F^n. Define $f\colon V \to F^n$ by $f(b_i) = e_i$ for $i = 1, \ldots, n$ and extending linearly. Recall what this means: given $v \in V$, there are unique $\alpha_i \in F$ such that $v = \sum_{i=1}^{n} \alpha_i b_i$. Then by definition of "extend linearly",

$$f(v) = \sum_{i=1}^{n} \alpha_i f(b_i) = \sum_{i=1}^{n} \alpha_i e_i = (\alpha_1, \ldots, \alpha_n) \in F^n.$$

Earlier, we called $(\alpha_1, \ldots, \alpha_n)$ the *coordinates* of v with respect to the ordered basis $\langle b_1, \ldots, b_n \rangle$.

Suppose $v \in \ker(f)$, and write $v = \sum_{i=1}^{n} \alpha_i b_i$. Then $0_W = f(v) = \sum_{i=1}^{n} \alpha_i e_i$ implies $\alpha_i = 0$ for all i since the e_i are linearly independent. So $v = 0_V$. This shows that the kernel of f is trivial, and hence, f is injective. For surjectivity, note that the image contains all linear combinations of the standard basis vectors, $e_1, \ldots, e_n$ for F^n. $\square$

Remarks: Theorem 4 says that for each $n = 0, 1, 2, \ldots,$ there is essentially only one vector space over F of dimension n. More precisely, under the equivalence relation $V \sim W$ defined earlier, there is one equivalence class for each natural number n.

Theorem 4 and its proof say that the difference between a vector space V of dimension n and F^n is the choice of a basis. Once a basis B is chosen, we get an isomorphism $V \to F^n$ by sending each vector to its coordinates with respect to B:

$$\begin{aligned} V &\to F^n \\ v &\mapsto [v]_B. \end{aligned}$$

The practical importance of this result is that if we have a problem involving vectors in V, we can use the isomorphism to translate problem into one about n-tuples in F^n. We apply our algorithms, e.g., Gaussian elimination, to solve the problem in F^n and then use the inverse of the isomorphism to translate the solution back to V.

Corollary 5. Let V and W be finite-dimensional vectors spaces. Then V and W are isomorphic if and only if they have the same dimension.

Proof. First, suppose that $f\colon V \to W$ is an isomorphism, and let $b_1, \dots, b_n$ be a basis for V. By Proposition 2, $f(b_1), \dots, f(b_n)$ are linearly independent, and since f is surjective, they span W. So $\{f(b_1), \dots, f(b_n)\}$ is a basis for W. Thus, the number of elements in a basis for V is the same as the number of elements in a basis for W, which says that $\dim V = \dim W$.

Conversely, suppose that $\dim V = \dim W = n$. By Theorem 4, we have isomorphisms $f_V\colon V \to F^n$ and $f_W\colon W \to F^n$. Let $f_W^{-1}\colon F^n \to W$ be the inverse of f_W. It follows that the composition,

$$V \xrightarrow{f_V} F^n \xrightarrow{f_W^{-1}} W$$

is an isomorphism. (From Math 112, you know that a composition of bijections of sets is a bijection of sets, and you should do the easy check that a composition of linear functions is linear.) □

Proposition 6. Let $f\colon V \to W$ be a linear function, and let $\dim V = \dim W < \infty$. (An important special case is $f\colon V \to V$ when $\dim V < \infty$.) Then the following are equivalent:

(a) f is injective (1-1),

(b) f is surjective (onto),

(c) f is an isomorphism.

Proof. The proof is left as an exercise. The central idea is to use the rank-nullity theorem to relate injectivity and surjectivity. □

Note: Proposition 6 is not true if the dimensions of V and W are not finite. For instance, consider the infinite-dimensional vector space $\mathcal{P}(F) = F[x]$ and the mapping

$$\begin{aligned} F[x] &\to F[x] \\ f &\mapsto xf, \end{aligned}$$

given by multiplication by x. For instance, under this mapping, $1+x+x^2 \mapsto x+x^2+x^3$. This mapping is linear and injective, but not surjective. For instance, 1 is not in the image (nor is any other constant besides 0).

Week 6, Monday: Linear transformations and matrices I

Our next goal is to encode linear functions by matrices. We first treat the special case of linear functions of the form $F^n \to F^m$. Next, we consider linear functions $V \to W$ between general finite-dimensional vector spaces. If $\dim V = n$ and $\dim W = m$ we saw last time that a choices of bases give isomorphisms $V \simeq F^n$ and $W \simeq F^m$, which reduces the problem to the special case.

Matrices for linear functions $F^n \to F^m$. The *dot product* of vectors $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ in F^n is defined by

$$(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) := \sum_{i=1}^{n} a_i b_i = a_1 b_1 + \cdots + a_n b_n.$$

From now on we make adopt the convention of identifying vectors $(a_1, \ldots, a_n) \in F^n$ with $n \times 1$ matrices, also called *column vectors*:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

If $A \in M_{m\times n}(F)$ and $x = (x_1, \ldots, x_n) \in F^n$, we define $Ax \in F^m$ to be the element of F^m whose i-th component $(Ax)_i$ is the dot product of the i-th row of A with x:

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

$$= (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \ldots, a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n).$$

The latter equals sign is just making the identification of column vectors with elements of F^m. Equivalently,

$$Ax := x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

We could similarly, convert the above notation into a statement about a linear combination of m-tuples in F^m instead of using column vectors.

Definition. Let $A \in M_{m\times n}(F)$. The linear associated with A is

$$\begin{aligned} L_A\colon F^n &\to F^m \\ x &\mapsto Ax. \end{aligned}$$

Exercise. The reader should perform the routine check that L_A is a linear function: $L_A(x + \lambda y) = L_A(x) + \lambda L_A(y)$.

Examples.

(1) The matrix

$$A = \begin{pmatrix} 2 & -5 & 4 \\ 3 & 0 & 2 \end{pmatrix}$$

has corresponding linear mapping

$$\begin{aligned} L_A\colon F^3 &\to F^2 \\ (x, y, z) &\mapsto \begin{pmatrix} 2x - 5y + 4z \\ 3x + 2z \end{pmatrix}. \end{aligned}$$

Recall that we are identifying To save space, we could we will write this as

$$\begin{aligned} L_A\colon F^3 &\to F^2 \\ (x, y, z) &\mapsto (2x - 5y + 4z, 3x + 2z). \end{aligned}$$

(2) Note that if you were given the linear function L_A, you could easily recover the matrix: just read off the coefficients of each component of $L_A(x)$ to find the rows of A. (We will see another way of recovering A below.) For example, find the matrix corresponding to the linear function $\phi\colon F^3 \to F^2$ defined by $\phi(u, v) = (4u - 3v, 6u + 2v, 3v)$.

Solution. Reading off the coefficients of each component of ϕ gives our matrix. Defining

$$A := \begin{pmatrix} 4 & -3 \\ 6 & 2 \\ 0 & 3 \end{pmatrix},$$

it is easy to check that $\phi = L_A$.

(3) Here are some important special cases of this correspondence between linear functions and matrices:

$$L_A(x) = (2x, 5x, 7x) \quad \leftrightsquigarrow \quad A = \begin{pmatrix} 2 \\ 5 \\ 7 \end{pmatrix}$$

$$L_B(w, x, y, z) = w + 2x - 4y + z \quad \leftrightsquigarrow \quad B = \begin{pmatrix} 1 & 2 & -4 & 1 \end{pmatrix}$$

$$L_C(t) = 8t \quad \leftrightsquigarrow \quad C = \begin{pmatrix} 8 \end{pmatrix}.$$

We have formally defined the linear mapping L_A associated with a matrix A, and from the examples above, it may be clear how to go in the other direction to find the matrix of a given linear function. Here is the formal definition:

Definition. The matrix associated with the linear function $L\colon F^n \to F^m$ is the element $A \in M_{m\times n}(F)$ whose j-th column is $L(e_j)$ where e_j is the j-th standard basis vector for F^n.

Examples. Consider the first two examples given above.

(1) Consider the linear function $L\colon F^3 \to F^2$ given by $L(x, y, z) = (2x-5y+4z, 3x+2z)$. Evaluate L at the three standard basis vectors for F^3:

$$\begin{aligned} L(e_1) &= L(1, 0, 0) = (2, 3) \\ L(e_2) &= L(0, 1, 0) = (-5, 0) \\ L(e_3) &= L(0, 0, 1) = (4, 2). \end{aligned}$$

Use these three vectors to form a matrix:

$$A = \begin{pmatrix} 2 & -5 & 4 \\ 3 & 0 & 2 \end{pmatrix}.$$

Thus, $L = L_A$.

(2) Consider the linear function $\phi\colon F^3 \to F^2$ given by $\phi(u,v) = (4u-3v, 6u+2v, 3v)$. Then,

$$\phi(1,0) = (4,6,0) \quad \text{and} \quad \phi(0,1) = (-3,2,3).$$

Place these vectors as columns to get the matrix

$$\begin{pmatrix} 4 & -3 \\ 6 & 2 \\ 0 & 3 \end{pmatrix}.$$

We have, thus, created a bijective correspondence between linear function $F^n \to F^m$ and matrices in $M_{m\times n}(F)$.

Matrices for linear functions $V \to W$.

Let V and W be vector spaces with ordered bases $\mathcal{B} = \langle v_1, \dots, v_n\rangle$ and $\mathcal{D} = \langle w_1, \dots, w_m\rangle$, respectively. Taking coordinates with respect to these bases yields isomorphisms $\phi_{\mathcal{B}} : V \to F^n$ and $\phi_{\mathcal{D}}\colon W \to F^m$. For instance, if $v \in V$, we write $v = \sum_{i=1}^n a_i v_i$, and then $\phi_{\mathcal{B}}(v) := (a_1, \dots, a_n)$. Now suppose we have a linear function $f\colon V \to W$. So up to now we have three mappings we are considering:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle \phi_{\mathcal{B}}}\downarrow\wr & & \wr\downarrow{\scriptstyle \phi_{\mathcal{D}}} \\ F^n & & F^m \end{array}$$

We now describe how to use this diagram to create a linear function $L\colon F^n \to F^m$. Since $\phi_{\mathcal{B}}$ is an isomorphism, we can invert it and then define L by starting at F^n, applying $\phi_{\mathcal{B}}^{-1}$ to go up the left-hand side of the diagram arriving at V, then applying f to go to W, and finally using $\phi_{\mathcal{D}}$ to go from W to F^m. More succinctly, define:

$$L := \phi_{\mathcal{D}} \circ f \circ \phi_{\mathcal{B}}^{-1}.$$

In sum, we have the following important *commutative diagram*:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle \phi_{\mathcal{B}}}\downarrow\wr & & \wr\downarrow{\scriptstyle \phi_{\mathcal{D}}} \\ F^n & \xrightarrow{L} & F^m \end{array}$$

Saying the diagram is *commutative* means the no matter which path we take from V to F^m, we arrive at the same place, i.e.,

$$L \circ \phi_{\mathcal{B}} = \phi_{\mathcal{D}} \circ f.$$

Now L is a mapping between tuples and, thus, has a matrix, as discussed at the beginning of this lecture. To keep track of all of the input data, we use the following, necessarily complicated, notation for this matrix:

$$[f]_{\mathcal{B}}^{\mathcal{D}} := \text{ matrix corresponding to } L.$$

How do we compute this matrix? The algorithm for computing $[f]_\alpha^\beta$ is summarized in the diagram below:

$$\begin{tikzcd}
v_j \arrow[rrr, bend left=10] \arrow[ddd, dashed] & & & f(v_j) \arrow[ddd, "\text{take coords. wrt. } \mathcal{D}"] \\
& V \arrow[r, "f"] \arrow[d, "\phi_{\mathcal{B}}"', "\wr"] & W \arrow[d, "\phi_{\mathcal{D}}", "\wr"'] & \\
& F^n \arrow[r, "{[f]_{\mathcal{B}}^{\mathcal{D}}}"] & F^m. & \\
e_j \arrow[rrr, dashed] & & & j\text{-th col. of } [f]_{\mathcal{B}}^{\mathcal{D}}
\end{tikzcd}$$

In words: since $[f]_{\mathcal{B}}^{\mathcal{D}}$ is a matrix, its j-th column is given by $[f]_{\mathcal{B}}^{\mathcal{D}}(e_j)$. By definition,

$$[f]_{\mathcal{B}}^{\mathcal{D}}(e_j) = \phi_{\mathcal{D}} \circ f \circ \phi_{\mathcal{B}}^{-1}(e_j) = \phi_{\mathcal{D}}\left(f\left(\phi_{\mathcal{B}}^{-1}(e_j)\right)\right).$$

We have $\phi_{\mathcal{B}}(v_j) = e_j$. Hence, $\phi_{\mathcal{B}}^{-1}(e_j) = v_j \in V$. So,

$$[f]_{\mathcal{B}}^{\mathcal{D}}(e_j) = \phi_{\mathcal{D}}\left(f\left(\phi_{\mathcal{B}}^{-1}(e_j)\right)\right) = \phi_{\mathcal{D}}\left(f(v_j)\right).$$

So here is the algorithm for computing $[f]_{\mathcal{B}}^{\mathcal{D}}$:

> To find the j-th column of $[f]_{\mathcal{B}}^{\mathcal{D}}$ compute the coordinates of $f(v_j)$ with respect to $\mathcal{D} = \langle w_1, \ldots, w_m \rangle$ for each $v_j \in \mathcal{B} = \langle v_1, \ldots, v_n \rangle$.

Example. Consider linear function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix

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

Thus, $f(x, y) = (x + 4y, 2x + 3y$. Using the notation above, we are letting $V = W = \mathbb{R}^2$. Take the same ordered basis for both V and W given by

$$\mathcal{B} = \mathcal{D} = \langle (1, 1), (-2, 1) \rangle.$$

Find the matrix representing f with respect to this choice of bases for domain and codomain.

Solution. To conform with our earlier notation, we take $v_1 = (1, 1)$ and $v_2 = (-2, 1)$. First apply f to each of the basis vectors for V:

$$\begin{aligned} f(v_1) &= (5, 5) \\ f(v_2) &= (2, -1). \end{aligned}$$

Next, take the coordinates of these vectors with respect to the basis $\mathcal{B}$ for W:

$$\begin{aligned} (5, 5) &= 5v_1 + 0 \cdot v_2 \\ (2, -1) &= 0 \cdot v_1 - v_2. \end{aligned}$$

Hence,

$$\begin{aligned} \phi_{\mathcal{B}}(v_1) &= (5, 0) \\ \phi_{\mathcal{B}}(v_2) &= (0, -1). \end{aligned}$$

These are the columns for our matrix:

$$[f]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 5 & 0 \\ 0 & -1 \end{pmatrix}.$$

We arrive at the commutative diagram:

$$\begin{tikzcd} \mathbb{R}^2 \arrow[r, "f"] \arrow[d, "\phi_{\mathcal{B}}"', "\wr"] & \mathbb{R}^2 \arrow[d, "\phi_{\mathcal{B}}", "\wr"'] \\ \mathbb{R}^2 \arrow[r, "L"] & \mathbb{R}^2 \end{tikzcd}$$

Where L is the linear function corresponding to $[f]_{\mathcal{B}}^{\mathcal{B}}$, i.e.,

$$L(x, y) = (5x, -y).$$

Example. Consider the linear mapping

$$\begin{aligned} f \colon \mathbb{R}[x]_{\leq 2} &\to \mathbb{R}[x]_{\leq 3} \\ p &\mapsto xp. \end{aligned}$$

Thus, f consists of multiplying a polynomial by x. Choose bases $\mathcal{B} = \langle 1, x, x^2 \rangle$ for the domain and $\mathcal{D} = \langle 1, x, x^2, x^3 \rangle$ for the codomain. Thus, $\phi_{\mathcal{B}}(a + bx + cx^2) = (a, b, c)$ and $\phi_{\mathcal{D}}(a + bx + cx^2 + dx^3) = (a, b, c, d)$. To find $[f]_{\mathcal{B}}^{\mathcal{D}}$, compute the images of the elements in $\mathcal{B}$ and express them as linear combinations of elements of $\mathcal{D}$:

$$f(1) = x = 0 \cdot 1 + 1 \cdot x + 0 \cdot x^2 + 0 \cdot x^3$$

$$
\begin{aligned}
f(x) &= x^2 = 0 \cdot 1 + 0 \cdot x + 1 \cdot x^2 + 0 \cdot x^3 \\
f(x^2) &= x^3 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 1 \cdot x^3.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
[f(1)]_{\mathcal{D}} &= (0, 1, 0, 0) \\
[f(x)]_{\mathcal{D}} &= (0, 0, 1, 0) \\
[f(x^2)]_{\mathcal{D}} &= (0, 0, 0, 1).
\end{aligned}
$$

These vectors are the columns for our matrix:

$$
[f]_{\mathcal{B}}^{\mathcal{D}} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
$$

Week 6, Wednesday: Linear transformations and matrices II

Recall from last time that given a linear mapping $f\colon V \to W$ and ordered bases $\mathcal{B} = \langle v_1, \ldots, v_n \rangle$ and $\mathcal{D} = \langle w_1, \ldots, w_m \rangle$ for V and W, respectively, we have a commutative diagram

$$
\begin{tikzcd}
V \arrow[r, "f"] \arrow[d, "\phi_{\mathcal{B}}"', "\wr"] & W \arrow[d, "\phi_{\mathcal{D}}", "\wr"'] \\
F^n \arrow[r, "L"] & F^m
\end{tikzcd}
$$

where

$$L := \phi_{\mathcal{D}} \circ f \circ \phi_{\mathcal{B}}^{-1}.$$

The matrix representing L is denoted $[f]_{\mathcal{B}}^{\mathcal{D}}$ and its calculation is displayed in the following diagram:

$$
\begin{tikzcd}
v_j \arrow[rrr, bend left] \arrow[dddd, dashed] & & & f(v_j) \arrow[dddd, "\text{take coords. wrt. } \mathcal{D}"] \\
& V \arrow[r, "f"] \arrow[d, "\phi_{\mathcal{B}}"', "\wr"] & W \arrow[d, "\phi_{\mathcal{D}}", "\wr"'] & \\
& F^n \arrow[r, "{[f]_{\mathcal{B}}^{\mathcal{D}}}"] & F^m. & \\
& & & \\
e_j \arrow[rrr, dashed] & & & j\text{-th col. of } [f]_{\mathcal{B}}^{\mathcal{D}}
\end{tikzcd}
$$

We compute $[f]_{\mathcal{B}}^{\mathcal{D}}$ by finding each of its columns: To find the j-th column of $[f]_{\mathcal{B}}^{\mathcal{D}}$ compute the coordinates of $f(v_j)$ with respect to $\mathcal{D} = \langle w_1, \ldots, w_m \rangle$ for each $v_j \in \mathcal{B} = \langle v_1, \ldots, v_n \rangle$.

Commutativity of the diagram says that for each $v \in V$

$$[f]_{\mathcal{B}}^{\mathcal{D}}(\phi_{\mathcal{B}}(v)) = \phi_{\mathcal{D}}(f(v)).$$

Recall our notation for the coordinates of a vector with respect to a ordered basis, we can rewrite that above as

$$[f]_{\mathcal{B}}^{\mathcal{D}}[v]_{\mathcal{B}} = [f(v)]_{\mathcal{D}}.$$

Example. Consider the linear function

$$\begin{aligned} f\colon \mathbb{R}[x]_{\leq 2} &\to \mathbb{R}[x] \leq 3 \\ p &\mapsto xp + 2p'. \end{aligned}$$

Choose ordered bases $\mathcal{B} = \langle 1, x, x^2 \rangle$ and $\mathcal{D} = \langle 1, x, x^2, x^3 \rangle$ for the domain and codomain, respectively. Find the matrix representing f with respect to these bases, and use the matrix to computer $f(3 + 2x + x^2)$.

Solution. Compute the images of the basis vectors in $\mathcal{B}$:

$$\begin{aligned} f(1) &= x \cdot 1 + 2(1)' = x \\ f(x) &= x \cdot x + 2(x)' = x^2 + 2 \\ f(x^2) &= x \cdot x^2 + 2(x^2)' = x^3 + 4x. \end{aligned}$$

Next, find the coordinates of each of these with respect to $\mathcal{D}$:

$$\begin{aligned} [x]_{\mathcal{D}} &= (0, 1, 0, 0) \\ [x^2 + 2]_{\mathcal{D}} &= (2, 0, 1, 0) \\ [x^3 + 4]_{\mathcal{D}} &= (0, 4, 0, 1). \end{aligned}$$

Therefore,

$$[f]_{\mathcal{B}}^{\mathcal{D}} = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Here is a helpful way to think about this matrix:

$$\begin{array}{c c} & \begin{array}{ccc} f(1) & f(x) & f(x^2) \end{array} \\ \begin{array}{c} 1 \\ x \\ x^2 \\ x^3 \end{array} & \begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}.$$

The columns are labeled by the images of the basis vectors of the domain, and the rows are labeled by basis vectors of codomain.

To find $f(3 + 2x + x^2)$, we first do the calculation using coordinates:

$$[f(3 + 2x + x^2)]_{\mathcal{D}} = [f]_{\mathcal{B}}^{\mathcal{D}} [3 + 2x + x^2]_{\mathcal{B}}$$

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

$$= \begin{pmatrix} 4 \\ 7 \\ 2 \\ 1 \end{pmatrix}.$$

It follows that

$$f(3 + 2x + x^2) = 4 + 7x + 2x^2 + x^3.$$

Check using the definition of f:

$$\begin{aligned} f(3 + 2x + x^2) &= x(3 + 2x + x^2) + 2(3 + 2x + x^2)' \\ &= (3x + 2x^2 + x^3) + 2(2 + 2x) \\ &= 4 + 7x + 2x^2 + x^3. \end{aligned}$$

We would next like to prove that the rank of a linear function is equal to the rank of any matrix representative of that function. Recall that the rank of a linear function is the dimension of its image, and the rank of a matrix is the dimension of its column space (which we saw is equal to the dimension of its row space—it is the number of pivot columns in the reduced row echelon form of the matrix).

Proposition. Let V and W be finite-dimensional vector spaces with ordered bases $\mathcal{B}$ and $\mathcal{D}$, respectively. Let $f\colon V \to W$ be a linear transformation. Then

$$\operatorname{rank}(f) = \operatorname{rank}([f]_{\mathcal{B}}^{\mathcal{D}}).$$

Proof. We first consider the special case of a linear mapping $L_A\colon F^n \to F^m$ where $A \in M_{m\times n}$. Thus, $L_A(x) = Ax$. We saw last time that the image of L_A is the span of the column of A, i.e., the column space of A. Thus, the result holds in this case:

$$\operatorname{rank}(L_A) := \dim(\operatorname{im}(L_A)) = \dim(\operatorname{colspace}(A)) = \operatorname{rank}(A).$$

Now consider the general case. We have the commutative diagram

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\phi_{\mathcal{B}}"', "\wr"] & W \arrow[d, "\phi_{\mathcal{D}}", "\wr"'] \\ F^n \arrow[r, "L_A"] & F^m \end{tikzcd}$$

where, in this case, $A = [f]_{\mathcal{B}}^{\mathcal{D}}$. Since $\phi_{\mathcal{B}}$ and $\phi_{\mathcal{D}}$ are isomorphism and the diagram commutes,

$$\operatorname{rank}(L_A) := \dim(\operatorname{im}(L_A)) = \dim(\operatorname{im}(L_A \circ \phi_{\mathcal{B}})) = \dim(\operatorname{im}(\phi_{\mathcal{D}} \circ f)) = \dim(\operatorname{im}(f)) =: \operatorname{rank}(f).$$

We have seen that $\operatorname{rank}(L_A) = \operatorname{rank}(A)$. So the result follows, in general. □

Corollary. With notation as above, let $A = [f]_{\mathcal{B}}^{\mathcal{D}} \in M_{m \times n}(F)$.

(a) f is surjective if and only if $\operatorname{rank}(A) = m = \dim(W)$.
(b) f is injective if and only if $\operatorname{rank}(A) = n = \dim(V)$.
(c) f is an isomorphism if and only if $\operatorname{rank}(A) = m = n$.

Proof.

(a) The function f being surjective means that $\operatorname{im}(f) = W$, which is equivalent to saying that $\dim(\operatorname{im}(f)) = \dim(W)$, or that $\operatorname{rank}(f) = m$, and we have just seen that $\operatorname{rank}(f) = \operatorname{rank}(A)$.
(b) We know that f is injective if and only if $\dim(\ker(f)) = 0$. By the rank-nullity theorem,

$$n = \dim V = \dim(\operatorname{im}(f)) + \dim(\ker(f)) = \operatorname{rank}(f) + \dim(\ker(f)).$$

From the Proposition, we have $\operatorname{rank}(f) = \operatorname{rank}(A)$. Therefore, $\dim(\ker f) = 0$ if and only if $\operatorname{rank}(A) = n$.
(c) This part follows from the previous two.

□

Composition of linear functions. Consider the linear functions

$$\begin{aligned} f\colon \mathbb{R}^4 &\to \mathbb{R}^2 \\ (x, y, z, w) &\mapsto (2x - z + 3w, x - y + 4z) \end{aligned}$$

and

$$\begin{aligned} g\colon \mathbb{R}^2 &\to \mathbb{R}^3 \\ (s, t) &\mapsto (5s - t, 2t, -3s). \end{aligned}$$

Let's compute the composition $g \circ f \colon \mathbb{R}^4 \to \mathbb{R}^3$:

$$\begin{aligned}(g \circ f)(x, y, z, w) &= g(\underbrace{2x - z + 3w}_{s}, \underbrace{x - y + 4z}_{t}) \\ &= (5(2x - z + 3w) - (x - y + 4z), 2(x - y + 4z), -3(2x - z + 3w)) \\ &= (9x + y - 9z + 15w, 2x - 2y + 8z, -6x + 3z - 9w).\end{aligned}$$

The matrices associated with f and g (with respect to the standard bases) are, respectively,

$$\begin{pmatrix} 2 & 0 & -1 & 3 \\ 1 & -1 & 4 & 0 \end{pmatrix}, \quad \begin{pmatrix} 5 & -1 \\ 0 & 2 \\ -3 & 0 \end{pmatrix}, \quad \begin{pmatrix} 9 & 1 & -9 & 15 \\ 2 & -2 & 8 & 0 \\ -6 & 0 & 3 & -9 \end{pmatrix}.$$

What is the relation among these matrices? We will take up this question next time.

Week 6, Friday: Linear transformations and matrices III

The goal today is to formally define the algebraic structure for matrices (linear structure and multiplication). Multiplication of matrices corresponds with composition of their corresponding linear transformations.

Composition of linear functions.

Proposition. Let $f\colon V \to W$ and $g\colon W \to U$ be linear functions. Their the composition $g \circ f\colon V \to U$ is a linear function.

Proof. Let $u, v \in V$ and $\lambda \in F$. Then, since f and g are linear,

$$\begin{aligned}(g \circ f)(u + \lambda v) &:= g(f(u + \lambda v)) \\ &= g\left(f(u) + \lambda f(v)\right) \\ &= g(f(u)) + \lambda g(f(v)) \\ &= (g \circ f)(u) + \lambda (g \circ f)(v).\end{aligned}$$

$\square$

Let $f\colon V \to W$ and $g\colon W \to U$ be a linear functions. We are interested in a matrices representing the composition

$$g \circ f\colon V \xrightarrow{f} W \xrightarrow{g} U.$$

Fix ordered bases $\mathcal{B} = \langle v_1, \ldots, v_n \rangle$ for V, $\mathcal{C} = \langle w_1, \ldots, w_\ell \rangle$ for W, and $\mathcal{D} = \langle u_1, \ldots, u_m \rangle$ for U. Let

$$P := [g]_{\mathcal{C}}^{\mathcal{D}} \quad \text{and} \quad Q = [f]_{\mathcal{B}}^{\mathcal{C}}.$$

Thus, $P \in M_{m \times \ell}(F)$ and $Q \in M_{\ell \times n}$. The relevant commutative diagram is

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\phi_{\mathcal{B}}"', "\wr"] & W \arrow[r, "g"] \arrow[d, "\phi_{\mathcal{C}}", "\wr"'] & U \arrow[d, "\phi_{\mathcal{D}}", "\wr"'] \\ F^n \arrow[r, "Q"] & F^\ell \arrow[r, "P"] & F^m \end{tikzcd}$$

Let's compute $[g\circ f]_{\mathcal{B}}^{\mathcal{D}}$. To find its j-th column, we find the coordinates of $(g\circ f)(v_j)$ with respect to the ordered basis $\mathcal{D}$:

$$\begin{aligned}
(g\circ f)(v_j) &= g(f(v_j)) \\
&= g\left(\sum_{k=1}^{\ell} Q_{kj} w_k\right) && (\text{ j-th column of } Q) \\
&= \sum_{k=1}^{\ell} Q_{kj} g(w_k) \\
&= \sum_{k=1}^{\ell} Q_{kj}\left(\sum_{i=1}^{m} P_{ik} u_i\right) && (\text{ k-th column of } P) \\
&= \sum_{i=1}^{m}\left(\sum_{k=1}^{\ell} P_{ik} Q_{kj}\right) u_i.
\end{aligned}$$

So the j column of $[g\circ f]_{\mathcal{B}}^{\mathcal{D}}$ is given by the coefficients of the u_i in the above some. That means that the (i,j)-th entry of the matrix $[g\circ f]_{\mathcal{B}}^{\mathcal{D}}$, i.e., the entry in its i-row and j-th column is

$$\left([g\circ f]_{\mathcal{B}}^{\mathcal{D}}\right)_{ij} = \sum_{k=1}^{m} P_{ik} Q_{kj}.$$

Definition. (Multiplication of matrices) Let $P\in M_{m\times \ell}(F)$ and $Q\in M_{\ell\times n}(F)$, then the product $PQ\in M_{m\times n}(F)$ is defined by

$$(PQ)_{ij} = \sum_{k=1}^{\ell} P_{ik} Q_{kj}.$$

Note: The formula says that the (i,j)-th entry of the product PQ is the dot product of the i-th row of P with the j-th column of Q. That's what one thinks about when performing the calculation of PQ in practice.

Example. Here is an example of the product of two matrices. For instance, to find the $(2,3)$-entry of the product, we take the dot product of the second row of the first matrix with the third column of the second:

$$\begin{pmatrix} 5 & -1 \\ 0 & 2 \\ -3 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 & -1 & 3 \\ 1 & -1 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 9 & 1 & -9 & 15 \\ 2 & -2 & 8 & 0 \\ -6 & 0 & 3 & -9 \end{pmatrix}.$$

Recall the relevance of this computation: the first two matrices encode linear functions g and f, and their product is a matrix encoding the composition $g \circ f$.

Proposition. Let $f\colon V \to W$ and $g\colon W \to U$ be a linear functions, and fix ordered bases $\mathcal{B} = \langle v_1, \ldots, v_n \rangle$ for V, $\mathcal{C} = \langle w_1, \ldots, w_\ell \rangle$ for W, and $\mathcal{D} = \langle u_1, \ldots, u_m \rangle$ for U. Then we have

$$[g \circ f]_{\mathcal{B}}^{\mathcal{D}} = [g]_{\mathcal{C}}^{\mathcal{D}} [f]_{\mathcal{B}}^{\mathcal{C}}.$$

Proof. The proof is exactly the motivation we just gave for the definition of the matrix product. □

We summarize some basic properties of matrix algebra.

Proposition. Let A be an $m \times n$ matrix, B an $n \times r$ matrix, both over a field F, and $\lambda \in F$.

(a) $\lambda(AB) = (\lambda A)B = A(\lambda B)$.

(b) $A(BC) = (AB)C$ for all $r \times s$ matrices C.

(c) $A(B + C) = AB + AC$ for all $n \times r$ matrices C.

(d) $(C + D)A = CA + DA$ for all $r \times m$ matrices C and D.

Proof. We will just prove part (b), associativity of multiplication. So let C be an $r \times s$ matrix. We have

$$\begin{aligned}
(A(BC))_{ij} &= \sum_{k=1}^{n} A_{ik}(BC)_{kj} \\
&= \sum_{k=1}^{n} \left(A_{ik} \left(\sum_{\ell=1}^{r} B_{k\ell} C_{\ell j} \right) \right) \\
&= \sum_{k=1}^{n} \sum_{\ell=1}^{r} A_{ik}(B_{k\ell} C_{\ell j}) \\
&= \sum_{\ell=1}^{r} \sum_{k=1}^{n} A_{ik}(B_{k\ell} C_{\ell j}) \\
&= \sum_{\ell=1}^{r} \sum_{k=1}^{n} (A_{ik} B_{k\ell}) C_{\ell j}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\ell=1}^{r} \left(\sum_{k=1}^{n} A_{ik} B_{k\ell} \right) C_{\ell j} \\
&= \sum_{\ell=1}^{r} (AB)_{i\ell} C_{\ell j} \\
&= ((AB)C)_{ij}.
\end{aligned}
$$

□

Warning. Matrix multiplication is not commutative, in general. First of all, if the dimensions aren't right, multiplication for both AB and BA might not make sense. For instance, if

$$
A = \begin{pmatrix} 1 & 0 \\ 3 & -1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 1 & 4 \end{pmatrix},
$$

then AB is defined, but not BA.

However, even if AB and BA are both defined, it is usually not the case that $AB = BA$. Try just about any example with 2×2 matrices to see this.

Definition. Let V and W be vector spaces over a field F. The set of linear transformations (homomorphisms) from V to W is denoted $\mathcal{L}(V, W)$ or $\mathrm{Hom}(V, W)$. It forms a vector space with operations defined as follows: for $f, g \in \mathrm{Hom}(V, W)$ and $\lambda \in F$,

$$
(f + g)(v) = f(v) + g(v) \quad \text{and} \quad (\lambda f)(v) = \lambda f(v)
$$

for all $v \in V$.

Proposition. Let V and W be vectors spaces over F of dimension n and m, respectively. Then there is an isomorphism of vector spaces

$$
\mathrm{Hom}(V, W) \to M_{m \times n}(F).
$$

Sketch of proof. Choose ordered bases $\mathcal{B}$ and $\mathcal{D}$ for V and W, respectively. Then an isomorphism is given by

$$
\begin{aligned}
\mathrm{Hom}(V, W) &\to M_{m \times n}(F) \\
f &\mapsto [f]_{\mathcal{B}}^{\mathcal{D}}.
\end{aligned}
$$

This isomorphism will change to a different isomorphism if different bases are chosen.

□

Week 7, Monday: Matrix inversion

Last time, we defined matrix multiplication: if A is an $m \times p$ matrix and B is a $p \times n$ matrix, then AB is the $m \times n$ matrix with i, j-entry

$$(AB)_{ij} := \sum_{k=1}^{p} A_{ik} B_{kj}.$$

If $m = n$, then BA would also be defined, but it is usually that case that $AB \neq BA$. Another peculiar thing is that for matrices, there are “zero divisors”, i.e., matrices A, B such that $AB = 0$, but neither A nor B is a zero matrix. For example,

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Diagonal matrices. The matrix A is a *diagonal matrix* if its only nonzero entries appear along the diagonal: $A_{ij} = 0$ if $i \neq j$. This terminology makes sense regardless of the dimensions of A, but is usually used in the case of square matrices, i.e., for the case where A is an $n \times n$ matrix. In that case, we write

$$A = \operatorname{diag}(a_1, \dots, a_n)$$

where $A_{ii} = a_i$ for $i = 1 \dots, n$ (and $A_{ij} = 0$, otherwise.). For instance,

$$\operatorname{diag}(1, 4, 0, 6) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix}.$$

Identity matrices. The $n \times n$ *identity matrix* is the $n \times n$ matrix

$$I_n = \operatorname{diag}(1, \dots, 1).$$

It has the following property: $AI_n = A$ and $I_n B = B$ whenever these products make sense. For instance,

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

and

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

Inverses. Let A be an $m \times n$ matrix, and let B be an $n \times m$ matrix. If $AB = I_n$, we say A is a *left-inverse* for B and B is a *right-inverse* for A. For example,

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Hence, A is a left-inverse for B and B is a right-inverse for A. On the other hand,

$$BA = \begin{pmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

So B is not a left-inverse of A and A is not a right-inverse of B. (In fact, B does not have a left-inverse and A does not have a right-inverse. This has to do with their ranks not being high enough. The connection with solving systems of equations we describe below explains that.)

We will mainly be interested in inverses for square matrices. Suppose that A is an $n \times n$ matrix. Suppose B is a right-inverse. So B is an $n \times n$ matrix such that $AB = I_n$. Since matrix multiplication is not commutative, the value of BA is not immediately clear. However, in fact, we have the following important result:

Theorem. Let A and B be $n \times n$ matrices. The following are equivalent:

(a) $AB = I_n$.
(b) $BA = I_n$.

If $AB = I_n$, we say A and B are *invertible* and write $A^{-1} = B$ and $B^{-1} = A$. The following are equivalent:

(i) A is invertible.
(ii) $\operatorname{rank}(A) = n$.
(iii) The reduced echelon form of A is I_n.

The proof of this theorem will follow from an elegant algorithm for computing the inverse of a matrix which we present below. The equivalence of the last to items on the list is something we already know.

Calculating the inverse. Our problem now is to determine whether an inverse for a matrix exists, and if so, to calculate that inverse. The methods we present here would also be applicable to calculating right- and left-inverses of non-square matrices—it boils down to solving systems of linear equations, after all—but we will concentrate on the case of square matrices.

Example. Let

$$A = \begin{pmatrix} 0 & 3 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

A right-inverse for A would satisfy the following:

$$\begin{pmatrix} 0 & 3 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

So we need to find the entries $a, b, \ldots, i$. We can break this into three problems:

$$\begin{pmatrix} 0 & 3 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} a \\ d \\ g \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 3 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} b \\ e \\ h \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 3 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} c \\ f \\ i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Equivalently, we need to solve three systems of linear equations:

$$0x + 3y - z = 1$$

$$\begin{aligned} x + 0y + z &= 0 \\ x - y + 0z &= 0 \end{aligned}$$

$$\begin{aligned} 0x + 3y - z &= 0 \\ x + 0y + z &= 1 \\ x - y + 0z &= 0 \end{aligned}$$

$$\begin{aligned} 0x + 3y - z &= 0 \\ x + 0y + z &= 0 \\ x - y + 0z &= 1 \end{aligned}$$

Their augmented matrices would like:

$$\left(\begin{array}{ccc|c} 0 & 3 & -1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \end{array}\right), \quad \left(\begin{array}{ccc|c} 0 & 3 & -1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \end{array}\right), \quad \left(\begin{array}{ccc|c} 0 & 3 & -1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{array}\right).$$

The row operations needed to determine the solvability of this system are the same in all three cases. So we can combine all three of these systems at once in one "super"-augmented matrix calculation:

$$\left(\begin{array}{ccc|ccc} 0 & 3 & -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{r_1 \leftrightarrow r_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 3 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{r_3 \to r_3 - r_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 3 & -1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & -1 & 1 \end{array}\right)$$

$$\xrightarrow[r_3 \to -r_3]{r_2 \leftrightarrow r_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & -1 \\ 0 & 3 & -1 & 1 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{r_3 \to r_3 - 3r_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & -1 \\ 0 & 0 & -4 & 1 & -3 & 3 \end{array}\right)$$

$$\xrightarrow{r_3 \to -r_3/4} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1/4 & 3/4 & -3/4 \end{array}\right)$$

$$\xrightarrow[r_2 \to r_2 - r_3]{r_1 \to r_1 - r_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/4 & 1/4 & 3/4 \\ 0 & 1 & 0 & 1/4 & 1/4 & -1/4 \\ 0 & 0 & 1 & -1/4 & 3/4 & -3/4 \end{array}\right).$$

Going back to the original systems of equations, we see that we need

$$\begin{pmatrix} a \\ d \\ g \end{pmatrix} = \begin{pmatrix} 1/4 \\ 1/4 \\ -1/4 \end{pmatrix}, \quad \begin{pmatrix} b \\ e \\ h \end{pmatrix} = \begin{pmatrix} 1/4 \\ 1/4 \\ 3/4 \end{pmatrix}, \quad \begin{pmatrix} c \\ d \\ i \end{pmatrix} = \begin{pmatrix} 3/4 \\ -1/4 \\ -3/4 \end{pmatrix}.$$

In other words, the following matrix is a right-inverse for A:

$$\begin{pmatrix} 1/4 & 1/4 & 3/4 \\ 1/4 & 1/4 & -1/4 \\ -1/4 & 3/4 & -3/4 \end{pmatrix}.$$

The argument we've just given for a particular matrix easily generalizes to give the following algorithm.

Algorithm for computing the inverse of a matrix. Let A be an $n \times n$ matrix. Perform row operations on the "super"-augmented matrix $[A \mid I_n]$ to compute the reduced echelon form of A:

$$(A \mid I_n) \longrightarrow \left(\tilde{A} \mid B\right).$$

There are two possibilities: either $\tilde{A} = I_n$ or not. If $\tilde{A} = I_n$, then $B = A^{-1}$. Next, we consider what happens when $\tilde{A} \neq I_n$. Since B is derived by performing row operations on I_n, we have $\operatorname{rank}(B) = \operatorname{rank}(I_n) = n$. Thus, B cannot have a row of zeros. If $\tilde{A} \neq I_n$, it must have a row of zeros. It follows that the system of equations is inconsistent, and A has no inverse.

Now suppose that $\operatorname{rank}(A)$ so that

$$(A \mid I_n) \longrightarrow (I_n \mid B) \tag{18.1}$$

and $AB = I_n$. Consider trying to find C so that $BC = I_n$. In this case, reverse the row operations in (18.1) to get

$$(B \mid I_n) \longrightarrow (I_n \mid A),$$

and thus, $C = A$, i.e., $BA = I_n$.

In summary:

- If $\tilde{A} = I_n$ (equivalently, $\text{rank}(A) = n$) then $AB = BA = I_n$. (So $B = A^{-1}$ and $A = B^{-1}$.)
- If $\tilde{A} \neq I_n$ (equivalently, $\text{rank}(A) < n$), then $\tilde{A}$ has a row of zeros and A has no inverse.

In particular: $A \in M_{n\times n}$ is invertible if and only if $\text{rank}(A) = n$.

Week 7, Wednesday: Change of basis

Let V be a vector space with ordered basis $\alpha = \langle v_1, \ldots, v_n \rangle$. Recall the coordinate mapping

$$\begin{aligned} \phi_\alpha \colon V &\xrightarrow{\sim} F^n \\ v = a_1 v_1 + \cdots + a_n v_n &\mapsto (a_1, \ldots, a_n). \end{aligned}$$

In particular, we have $\phi_\alpha(v_j) = e_j$. Consider the special case where $V = F^n$ so that

$$\phi_\alpha \colon F^n \to F^n,$$

and each v_j is an element of F^n. Since ϕ_α is now a mapping between tuples, is represented by the $n \times n$ matrix M with the property that $\phi_\alpha(v) = Mv$ for all $v \in F^n$. The j-th column of M is $\phi_\alpha(e_j)$ for $j = 1, \ldots, n$.

Proposition. With notation as above, let P be the $n \times n$ matrix whose j-th column is v_j for $j = 1, \ldots, n$. Then $M = P^{-1}$.

Proof. If X is any matrix, then Xe_j is the j-th column of X. So in our case, $Pe_j = v_j$. Since the v_j form a basis, the columns of P are linearly independent. So P has rank n and is, thus, invertible. We have

$$Pe_j = v_j \quad \Rightarrow \quad P^{-1}Pe_j = P^{-1}v_j \quad \Rightarrow \quad e_j = P^{-1}v_j.$$

Therefore, $P^{-1}v_j = e_j = \phi_\alpha(v_j)$ for all j. Since the v_j form a basis for F^n, it follows that $P^{-1}v = \phi_\alpha(v)$ for all $v \in F^n$. So P^{-1} is the matrix representing ϕ_α, which means that $P^{-1} = M$. $\square$

Next, consider a linear function

$$L_A \colon F^n \to F^m$$

given by the $m \times n$ matrix A, i.e., $L_A(v) = Av$ for all $v \in F^n$. Let $\alpha = \langle v_1, \ldots, v_n \rangle$ and $\beta = \langle w_1, \ldots, w_m \rangle$ be ordered bases for F^n and F^m respectively. What is the matrix representing L_A with respect to these new bases? We have the diagram

$$\begin{array}{ccc} F^n & \xrightarrow{L_A} & F^m \\ \phi_\alpha \big\downarrow \wr & & \wr \big\downarrow \phi_\beta \\ F^n & \xrightarrow{[L_A]_\alpha^\beta} & F^m. \end{array}$$

For ease of notation, let $B := [L_A]_\alpha^\beta$. Our main goal today is to give a formula for calculating B. We already know how to find the matrices representing the vertical coordinate mappings: let P and Q be the matrices whose columns are the elements of α and β, respectively, in order. Our diagram becomes

$$\begin{array}{ccc} F^n & \xrightarrow{A} & F^m \\ P^{-1} \big\downarrow \wr & & \wr \big\downarrow Q^{-1} \\ F^n & \xrightarrow{B} & F^m. \end{array}$$

Therefore, $B = Q^{-1}AP$. We summarize our result:

Proposition. Let $A \in M_{m\times n}(F)$, and consider the linear mapping $L_A \colon F^n \to F^m$ determined by A, i.e., $L(v) = Av$ for each $v \in F^n$. Let $\alpha = \langle v_1, \ldots, v_n\rangle$ and $\beta = \langle w_1, \ldots, w_m\rangle$ be ordered bases for F^n and F^m, respectively. Let P be the $n \times n$ matrix with j-th column v_j for $j = 1, \ldots, n$, and let Q be the $m \times m$ matrix with j-th column w_j for $j = 1, \ldots, m$. Then the matrix B representing L_A with respect to the bases α and β is

$$B = Q^{-1}AP,$$

and we have the commutative diagram

$$\begin{array}{ccc} F^n & \xrightarrow{A} & F^m \\ P^{-1} \big\downarrow \wr & & \wr \big\downarrow Q^{-1} \\ F^n & \xrightarrow{B} & F^m. \end{array}$$

Example. Let $\mathbb{Q}$ be the field of rational numbers, and consider the linear function

$$\begin{aligned} f \colon \mathbb{Q}^3 &\to \mathbb{Q}^2 \\ (x, y, z) &\mapsto (x + 3y + 2z, 2y + z), \end{aligned}$$

with corresponding matrix

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

Choose the following bases for the domain and codomain:

$$\mathbb{Q}^3 : \quad \alpha = \langle (1,0,0), (1,1,0), (1,1,1)\rangle$$

$$\mathbb{Q}^2: \quad \beta = \langle (0,1), (1,1) \rangle.$$

To find the matrix representing f with respect to these new bases, create matrices whose columns are the basis vectors:

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Calculate the inverse of Q:

$$\left(\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right) \xrightarrow{r_1 \leftrightarrow r_2} \left(\begin{array}{cc|cc} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right) \xrightarrow{r_1 \to r_1 - r_2} \left(\begin{array}{cc|cc} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right).$$

The matrix representing f with respect to the bases α and β is then:

$$B = Q^{-1}AP = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 3 & 2 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -2 & -3 \\ 1 & 4 & 6 \end{pmatrix}.$$

This agrees with the fact that

$$\begin{aligned} f(1,0,0) &= (1,0) = -1(0,1) + 1(1,1) \\ f(1,1,0) &= (4,2) = -2(0,1) + 4(1,1) \\ f(1,1,1) &= (6,3) = -3(0,1) + 6(1,1). \end{aligned}$$

An important special case. The special case of the Proposition that arises most frequently in practice is where $m = n$ and $\alpha = \beta$. In other, words, we start with a mapping $L_A \colon F^n \to F^n$ represented by the matrix A, and we choose the same new basis $\alpha = \langle v_1, \ldots, v_n \rangle$ for F^n for both the domain and codomain. We are then interested in the matrix representing L_A with respect to this new basis α. In that case, let P be the matrix whose columns are $v_1, \ldots, v_n$, and we get the commutative diagram

$$\begin{tikzcd} F^n \arrow[r, "A"] \arrow[d, "P^{-1}"', "\wr"] & F^n \arrow[d, "\wr"', "P^{-1}"] \\ F^n \arrow[r, "B"] & F^m \end{tikzcd}$$

and the matrix we are looking for is

$$B = P^{-1}AP.$$

We say B is formed by *conjugating* A.

Exercise. Say $A, B \in M_{n\times n}(F)$ are *similar* and write $A \sim B$ if there exists an invertible matrix $P \in M_{n\times n}(F)$ such that $P^{-1}AP = B$. Prove that $\sim$ is an equivalence relation. What does an equivalence class represent?

Example. Consider the real matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

What matrix represents the linear function $L_A \colon F^3 \to F^3$ with respect to the ordered basis $\alpha = \langle (1,1,1), (1,0,-1), (0,1,-1) \rangle$?

Solution. Use the vectors of α as columns to define the matrix

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

Compute the inverse of P using our algorithm (omitted):

$$P^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix}.$$

Then the matrix representing L_A with respect to the ordered basis α is

$$B = P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Notice that the matrix representing L_A in the example becomes the much simpler diagonal matrix after a change of basis. We can then apply an important trick to compute A^k for all integers k. First note that

$$A^2 = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 2 & 3 & 3 \\ 3 & 2 & 3 \\ 3 & 3 & 2 \end{pmatrix}, \quad A^4 = \begin{pmatrix} 6 & 5 & 5 \\ 5 & 6 & 5 \\ 5 & 5 & 6 \end{pmatrix}, \quad A^5 = \begin{pmatrix} 10 & 11 & 11 \\ 11 & 10 & 11 \\ 11 & 11 & 10 \end{pmatrix}.$$

What happens in general? Here is the trick that can be applied here:

$$B^k = (P^{-1}AP)^k$$

$$
\begin{aligned}
&= \underbrace{(P^{-1}AP)(P^{-1}AP)(P^{-1}AP)\cdots(P^{-1}AP)(P^{-1}AP)}_{k \text{ times}} \\
&= P^{-1}A(PP^{-1})A(PP^{-1})A(PP^{-1})\cdots(PP^{-1})A(PP^{-1})AP \\
&= P^{-1}A^kP.
\end{aligned}
$$

Since $B^k = P^{-1}A^kP$, we can solve for A^k by multiply both sides of the equality on the left by P and on the right by P^{-1} to get

$$
\begin{aligned}
A^k = PB^kP^{-1} &= \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}^k \left(\frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix} \right) \\
&= \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 2^k & 0 & 0 \\ 0 & (-1)^k & 0 \\ 0 & 0 & (-1)^k \end{pmatrix} \left(\frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix} \right) \\
&= \frac{1}{3} \begin{pmatrix} 2^k + 2\,(-1)^k & 2^k - (-1)^k & 2^k - (-1)^k \\ 2^k - (-1)^k & 2^k + 2\,(-1)^k & 2^k - (-1)^k \\ 2^k - (-1)^k & 2^k - (-1)^k & 2^k + 2\,(-1)^k \end{pmatrix} \\
&= \frac{1}{3} \begin{pmatrix} a & b & b \\ b & a & b \\ b & b & a \end{pmatrix}
\end{aligned}
$$

where for $k = 1, 2, 3, \ldots$,

$$
a = \begin{cases} 2^k + 2 & \text{if } k \text{ is even} \\ 2^k - 2 & \text{if } k \text{ is odd} \end{cases} \quad \text{and} \quad b = \begin{cases} 2^k - 1 & \text{if } k \text{ is even} \\ 2^k + 1 & \text{if } k \text{ is odd.} \end{cases}
$$

Exercise. Show that $2^k \pm 2$ and $2^k \pm 1$ are divisible by 3 for $k = 1, 2, \ldots$

Week 7, Friday: Determinants

Definition. The *determinant* is a multilinear, alternating function of the rows of square matrix, normalized so that its value on the identity matrix is 1.

To explain this terminology, start with the fact that the determinant is a function of the form

$$\det\colon M_{n\times n}(F) \to F.$$

Given a square matrix $A \in M_{n\times n}(F)$ with rows $r_1, \ldots, r_n \in F^n$, we write $\det(A) = \det(r_1, \ldots, r_n)$, i.e., we consider the determinant as a function of the rows of A. The determinant function has the following properties:

(a) *Multilinear.* The determinant is a linear function with respect to each row. Thus, if $r_1, \ldots, r_n$ are the row vectors of A (elements of F^n), r_i' is another row vector, and $\lambda \in F$, then

$$\begin{aligned}\det(r_1, \ldots, r_{i-1}, \lambda\, r_i + r_i', r_{i+1}, \ldots, r_n) &= \lambda \det(r_1, \ldots, r_{i-1}, r_i, r_{i+1}, \ldots, r_n) \\ &\quad + \det(r_1, \ldots, r_{i-1}, r_i', r_{i+1}, \ldots, r_n).\end{aligned}$$

The above expresses the fact that, in particular, the determinant is linear with respect to the i-th row.

(b) *Alternating.* The determinant is zero if two of its arguments are equal:

$$\det(r_1, \ldots, r_n) = 0$$

if $r_i = r_j$ for some $i \neq j$.

(c) *Normalized.* $\det(I_n) = \det(e_1, \ldots, e_n) = 1$.

We will prove the following theorem later:

Theorem. For each $n \geq 0$, there exists a unique determinant function.

For now we will accept this theorem on faith and explore some of the consequences. The following proposition shows that we can compute the determinant through row reduction.

Proposition 1. (Behavior of the determinant with respect to row operations.) Let $A, B \in M_{n\times n}(F)$.

(a) If B is obtained from A by swapping two rows, then $\det(B) = -\det(A)$.

(b) If B is obtained from A by scaling a row by a scalar λ, then $\det(B) = \lambda \det(A)$ (even if $\lambda = 0$).

(c) If B is obtained from A by adding a scalar multiple of one row to another row, then $\det(B) = \det(A)$.

Proof. For part (a), let $r_1, \ldots, r_n \in F^n$ be the rows of A. For ease of notation, we will assume that B is obtained from A by swapping the first two rows. The argument we present clearly generalizes to the case of swapping arbitrary rows. Replace the first two rows of A with $r_1 + r_2$ to obtain a matrix whose determinant is 0 by the alternating property:

$$0 = \det(r_1 + r_2, r_1 + r_2, r_3, \ldots, r_n).$$

Expand my multilinearity to get:

$$\begin{aligned}
0 &= \det(r_1 + r_2, r_1 + r_2, r_3, \ldots, r_n) \\
&= \det(r_1, r_1 + r_2, r_3, \ldots, r_n) + \det(r_2, r_1 + r_2, r_3, \ldots, r_n) \\
&= \det(r_1, r_1, r_3, \ldots, r_n) + \det(r_1, r_2, r_3, \ldots, r_n) \\
&\qquad\qquad + \det(r_2, r_1, r_3, \ldots, r_n) + \det(r_2, r_2, r_3, \ldots, r_n) \\
&= 0 + \det(A) + \det(B) + 0.
\end{aligned}$$

It follows that $\det(B) = -\det(A)$.

Part (a) follows immediately from the fact that the determinant is linear with respect to each row:

$$\det(r_1, \ldots, r_{i-1}, \lambda r_i, r_{i+1}, \ldots, r_n) = \lambda \det(r_1, \ldots, r_{i-1}, r_i, r_{i+1}, \ldots, r_n).$$

For Part (c), we use multilinearity and the alternating property. For ease of notation, we'll consider the case where B is obtained from A by adding a multiple of row 1 to row 2:

$$\det(B) = \det(r_1, \lambda r_1 + r_2, r_3, \ldots, r_n)$$

$$\begin{aligned}
&= \lambda \det(r_1, r_1, r_3, \ldots, r_n) + \det(r_1, r_2, r_3, \ldots, r_n) \\
&= 0 + \det(r_1, r_2, r_3, \ldots, r_n) \\
&= \det(A).
\end{aligned}$$

□

Corollary. Let $A \in M_{n\times n}(F)$, and let E be the reduced row echelon form of A. Then there exists a non-zero $k \in F$ such that $\det(A) = k \det(E)$.

Proof. The proof is immediate from Proposition 1. □

Example 1. Here we compute the determinant of a 2×2 matrix using the fact that the determinant is a multilinear alternating mapping with value 1 on the identity matrix.

$$\begin{aligned}
\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \det((a, b), (c, d)) \\
&= \det(a\, e_1 + b\, e_2, c\, e_1 + d\, e_2) \\
&= a \det(e_1, c\, e_1 + d\, e_2) + b \det(e_2, c\, e_1 + d\, e_2) \\
&= ac \det(e_1, e_1) + ad \det(e_1, e_2) + bc \det(e_2, e_1) + bd \det(e_2, e_2) \\
&= 0 + ad \det(e_1, e_2) + bc \det(e_2, e_1) + 0 \\
&= ad \det(e_1, e_2) - bc \det(e_1, e_2) \\
&= ad \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - bc \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= ad \cdot 1 - bc \cdot 1 = ad - bc.
\end{aligned}$$

Example 2. Here is an example of using row reduction to compute the determinant of a matrix. Let

$$A = \begin{pmatrix} 1 & 2 & -2 \\ 9 & 4 & 0 \\ 2 & 2 & 4 \end{pmatrix}$$

Using Proposition 1, we see that

$$\det(A) = \det \begin{pmatrix} 1 & 2 & -2 \\ 9 & 4 & 0 \\ 2 & 2 & 4 \end{pmatrix}$$

$$= \det \begin{pmatrix} 1 & 2 & -2 \\ 0 & -14 & 18 \\ 0 & -2 & 8 \end{pmatrix}$$

$$= -\det \begin{pmatrix} 1 & 2 & -2 \\ 0 & -2 & 8 \\ 0 & -14 & 18 \end{pmatrix}$$

$$= 2\det \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & -4 \\ 0 & -14 & 18 \end{pmatrix}$$

$$= 2\det \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & -4 \\ 0 & 0 & -38 \end{pmatrix}$$

$$= 2(-38)\det \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= 2(-38)\det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= 2(-38) = -76.$$

Example 3.

$$\det \begin{pmatrix} 4 & 2 & -3 & 8 \\ 0 & 5 & 1 & 3 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 3 \end{pmatrix} = (4 \cdot 5 \cdot 2 \cdot 3)\det \begin{pmatrix} 1 & 1/2 & -3/2 & 4 \\ 0 & 1 & 1/5 & 3/5 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= (4 \cdot 5 \cdot 2 \cdot 3)\det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= (4 \cdot 5 \cdot 2 \cdot 3) \cdot 1 = 120.$$

A matrix like that in the previous example, which has only zero entries below the diagonal, is called *upper-triangular*. So $A \in M_{n\times n}(F)$ is upper-triangular if $A_{ij} = 0$ whenever $i > j$.

Proposition 2. The determinant of an upper-triangular matrix is the product of its diagonal elements.

Proof. Let A be upper-triangular, and let E be its reduced echelon form. From Proposition 1, we know that $\det(A) = k\det(E)$ for some non-zero constant k. Imagine row-reducing an upper-triangular matrix, and you will see that E has a row of zeros if and only if A has some diagonal entry equal to zero. If E has a row of zeros, then $\det(E) = 0$. To see this, suppose the rows of E are $r_1, \ldots, r_n$ with $r_n = \vec{0}$. By multilinearity, we have:

$$\begin{aligned}
\det(E) &= \det(r_1, \ldots, r_{n-1}, \vec{0}) \\
&= \det(r_1, \ldots, r_{n-1}, 0 \cdot \vec{0}) \\
&= 0 \cdot \det(r_1, \ldots, r_{n-1}, \vec{0}) \\
&= 0.
\end{aligned}$$

So if A has a diagonal entry equal to 0, then $\det(E) = 0$, which implies $\det(A) = k\det(E) = 0$. So the result holds in this case.

Next, suppose that A has no diagonal entries equal to 0. Compute $\det(A)$ using multilinearity:

$$\det(A) = \det\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \ldots & a_{1n} \\ 0 & a_{22} & a_{23} & a_{24} & \ldots & a_{2n} \\ 0 & 0 & a_{33} & a_{34} & \ldots & a_{3n} \\ 0 & 0 & 0 & a_{44} & \ldots & a_{4n} \\ & & & & \ddots & \vdots \\ & & & & & a_{nn} \end{pmatrix}$$

$$= a_{11}\cdots a_{nn}\det\begin{pmatrix} 1 & a_{12}/a_{11} & a_{13}/a_{11} & a_{14}/a_{11} & \ldots & a_{1n}/a_{11} \\ 0 & 1 & a_{23}/a_{22} & a_{24}/a_{22} & \ldots & a_{2n}/a_{22} \\ 0 & 0 & 1 & a_{34}/a_{33} & \ldots & a_{3n}/a_{33} \\ 0 & 0 & 0 & 1 & \ldots & a_{4n}/a_{44} \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix}$$

$$= a_{11}\cdots a_{nn}\det(I_n)$$

$$= a_{11} \cdots a_{nn}.$$

Above, it is clear that once we get to the case of all 1s on the diagonal, we can row-reduce the matrix to the identity by adding multiples of rows to other rows—operations that do not change the determinant. □

Proposition 3. Let $A \in M_{n\times n}(F)$. The following are equivalent:

(a) $\det(A) \neq 0$,
(b) $\operatorname{rank}(A) = n$,
(c) A is invertible, i.e., A has an inverse.

Proof. Given our algorithm for computing the inverse of a matrix, the equivalence of parts 2 and 3 is evident. To show that parts 1 and 2 are equivalent, recall that by Proposition 1, we have $\det(A) = k\det(E)$ where E is the reduced echelon form of A and k is a non-zero scalar. Thus, $\det(A) = 0$ if and only if $\det(E) = 0$. The rank of A is n if and only if $E = I_n$, in which case $\det(A) = k \neq 0$. The rank of A is strictly less than n if and only if E has a row of zeros. Since E is upper-triangular, Proposition 2 implies that E has a row of zeros if and only if $\det(E) = 0$. □

To come:

(a) Define the *transpose*, A^t of A by $A^t_{ij} := A_{ji}$. Then $\det A^t = \det A$, and thus, the determinant is also the unique multilinear, alternating, normalized function on the *columns* of a matrix.

(b) The determinant is multiplicative: $\det(AB) = \det(A)\det(B)$.

(c) The determinant may be calculated by "expanding" along any row or column.

(d) We have the following formula for the determinant

$$\det A = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) A_{1\sigma(1)} \cdots A_{n\sigma(n)}$$

where $\mathfrak{S}_n$ is the collection of all permutations of $(1, \dots, n)$ and $\operatorname{sgn}(\sigma)$ is the *sign* of the permutation σ (i.e., 1 if the permutation is formed by an even number of flips and -1 if it is formed by an odd number of flips).

(e) Over the real numbers, the determinant gives the signed volume of the parallelepiped spanned by the rows (or by the columns) of the matrix.

Week 8, Monday: Determinant of the transpose

Goal: Our goal for the day is to show that the determinant of a matrix and the determinant of the transpose of that matrix are equal. It then immediately follows that the determinant is not only a multilinear, alternating, normalized function of the *rows* of a matrix (by definition), but it is also a multilinear, alternating, normalized function of its *columns*. So one may use both row and column operations to compute the determinant.

Elementary matrices. An $n \times n$ matrix is called an *elementary matrix* if it is obtained from the identity matrix, I_n, through a single elementary row operation (scaling a row by a nonzero scalar, swapping rows, or adding one row to another).

Here is why elementary matrices are interesting: Let E be an $n \times n$ elementary matrix corresponding to some row operation and let A be any $n \times k$ matrix. Then EA is the matrix obtained from A by performing that row operation. Thus, you can perform row operations through multiplication by elementary matrices.

Example. Let

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 0 & -1 & 2 \\ 1 & 5 & 6 & 7 \end{pmatrix}.$$

To find the elementary matrix that will subtract 3 times the first row of A from the second row, we do that same operation to the identity matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{r_2 \to r_2 - 3r_1} \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} =: E_1.$$

Multiplying by E_1 on the left performs the same elementary row operation on A:

$$E_1 A = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 0 & -1 & 2 \\ 1 & 5 & 6 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -6 & -10 & -10 \\ 1 & 5 & 6 & 7 \end{pmatrix}.$$

Next, let E_2 be the elementary matrix corresponding to subtracting the first row from the third:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{r_3 \to r_3 - r_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} =: E_2.$$

Multiplying $E_1 A$ on the left by E_2 performs the corresponding row operation:

$$E_2 E_1 A = E_2(E_1 A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -6 & -10 & -10 \\ 1 & 5 & 6 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -6 & -10 & -10 \\ 0 & 3 & 3 & 3 \end{pmatrix}.$$

Next, swap the second and third rows:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{r_2 \leftrightarrow r_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} =: E_3,$$

and, thus,

$$E_3 E_2 E_1 A = E_3(E_2 E_1 A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -6 & -10 & -10 \\ 0 & 3 & 3 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 3 & 3 \\ 0 & -6 & -10 & -10 \end{pmatrix}.$$

To continue row reduction, we would now scale the second row by $\frac{1}{3}$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{r_2 \leftrightarrow r_2/3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} =: E_4,$$

and, thus,

$$E_4 E_3 E_2 E_1 A = E_4(E_3 E_2 E_1 A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 3 & 3 \\ 0 & -6 & -10 & -10 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & -6 & -10 & -10 \end{pmatrix}.$$

As illustrated in the example, above, performing a sequence of row operations to a matrix is equivalent to multiplying on the left by a sequence of elementary matrices. In particular, if $\widetilde{A}$ is the reduced row echelon form of A, then there are elementary matrices $E_1, \ldots, E_\ell$ such that

$$\widetilde{A} = E_\ell \cdots E_2 E_1 A.$$

Determinant of the transpose. If A is an $m \times n$ matrix, recall that its *transpose* is the matrix A^t defined by

$$(A^t)_{ij} := A_{ji}.$$

Thus, the rows of A^t are the columns of A.

Example If

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 5 & 2 \\ 1 & 3 \end{pmatrix},$$

then

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

Our goal now is to prove the non-obvious fact that for an $n \times n$ matrix A,

$$\det(A) = \det(A^t).$$

Example. We have seen that

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Note that we also have

$$\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}^t \right) = \det \begin{pmatrix} a & c \\ b & d \end{pmatrix} = ad - bc.$$

Recall that we can compute the determinant of A by performing row operations and keeping track of swaps and scalings of rows. Once we have shown that $\det(A) = \det(A^t)$, it follows that, in order compute the determinant of A, we may also use column operations (again keeping track of swaps and scalings). That's because row operations applied to A^t are the same as column operations applied to A.

To prove this fact about determinants of transposes, we need the following theorem, proposition, and lemma:

Theorem. Let A and B be $n \times n$ matrices. Then

$$\det(AB) = \det(A)\det(B).$$

Proof. Upcoming homework. $\square$

Proposition. Let A and B be $n \times n$ matrices. Then

(a) $(AB)^t = B^t A^t$.

(b) If A is invertible, then $(A^t)^{-1} = (A^{-1})^t$.

Proof. Upcoming homework. $\square$

Lemma. Let E be an elementary matrix. Then $\det(E) = \det(E^t) \neq 0$.

Proof. There are three cases to consider[1]:

(a) Suppose E is formed from I_n by swapping rows i and j. In this case, E^t is also formed from I_n by swapping rows i and j. Thus, $E = E^t$, and $\det(E^t) = \det(E) = -1$.

(b) Suppose E is formed from I_n by scaling row i by $\lambda \neq 0$. In this case, E^t is also formed from I_n by scaling row i by λ. So in this case, $\det(E^t) = \det(E) = \lambda$.

(c) Suppose E is formed from I_n by adding λr_i to r_j for rows $r_i \neq r_j$. Then E^t is formed from I_n by adding λr_j to r_i. So in this case, $\det(E^t) = \det(E) = \det(I_n) = 1$.

$\square$

We can now prove our main result:

Theorem. Let A be an $n \times n$ matrix. Then $\det(A) = \det(A^t)$.

Proof. Let $\widetilde{A}$ be the reduced echelon form for A. Then $\widetilde{A} = I_n$ if and only if $\operatorname{rank} A = n$. So if $\widetilde{A} \neq I_n$, we have $\operatorname{rank}(A) < 0$, which means that $\det(A) = 0$. Since row rank and column rank are equal, we would then have $\operatorname{rank}(A^t) = \operatorname{rank}(A) < n$, which means $\det(A^t) = 0$, too. So the theorem holds in that case.

Now consider that case where $\widetilde{A} = I_n$. Thus, by applying row operations to A, we arrive at the identity matrix. Choose elementary matrices E_i such that

$$E_\ell \cdots E_2 E_1 A = I_n. \tag{21.1}$$

Taking determinants and using the fact that determinants preserve products yields:

$$\det(E_\ell) \cdots \det(E_2) \det(E_1) \det(A) = 1. \tag{21.2}$$

[1] The reader is strongly encouraged to create examples of each of these three cases.

Taking transposes in equation (21.1) gives

$$A^t E_1^t \cdots E_\ell^t = I_n^t = I_n.$$

Take determinants and use the fact from the Lemma that $\det(E) = \det(E^t)$ if E is an elementary matrix to get

$$\det(A^t)\det(E_1)\cdots\det(E_\ell) = 1. \tag{21.3}$$

The result follows by equating (21.2) and (21.3) and using the fact that the determinant of an elementary matrix is nonzero (so that we may cancel). □

Corollary. The determinant is a multilinear, alternating, normalized function of the *columns* of a square matrix.

For instance, the above corollary says the in order to compute the determinant, one may use column operations to simplify the calculation, just as we used row operations. Further, one may mix row and column operations, when convenient, as in the following example.

Example. Let

$$M = \begin{pmatrix} 1 & -1 & -1 & -1 \\ -2 & 3 & 0 & 1 \\ -2 & 1 & 4 & -1 \\ -5 & 1 & 1 & 5 \end{pmatrix}.$$

To compute the determinant of M, add the first row to each other row to get

$$M' = \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & 0 & 3 & -2 \\ -4 & 0 & 0 & 4 \end{pmatrix},$$

then, in M', add columns 2, 3, and 4 to column 1 to get

$$M'' = \begin{pmatrix} -2 & -1 & -1 & -1 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 3 & -2 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

The types of row and column operations we used to get from M to M'' have no effect on the determinant, and thus,

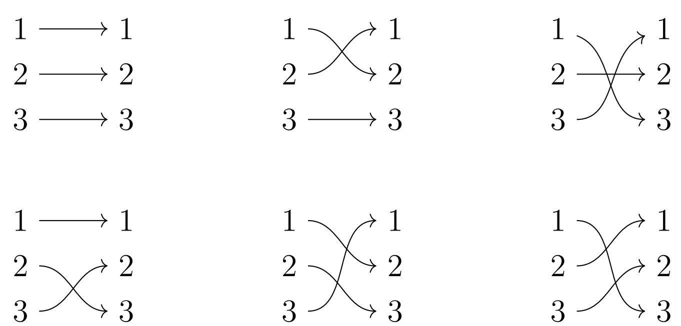
$$\det(M) = \det(M') = \det(M'') = -2 \cdot 2 \cdot 3 \cdot 4 = -48.$$

Remark. Start with I_n and perform a sequence of row operations to get a matrix $B \in M_{n\times n}(F)$. Then multiplying $A \in M_{n\times k}(F)$ on the *left* by B gives the matrix $BA \in M_{n\times k}(F)$ formed from A by performing the same sequence of row operations. Similarly, one can start with I_n, perform column operations to produce a matrix $C \in M_{n\times n}(F)$, then multiplying $D \in M_{k\times n}(F)$ on the *right* by C give the matrix $DC \in M_{k\times n}(F)$ obtained by performing the column operations on D.

Week 8, Wednesday: Permutation expansion of the determinant

Definition. A *permutation* of a set X is a bijective mapping of X to itself. If σ and τ are permutations of X, then so is their composition $\sigma \circ \tau$. The collection of all permutations of X along with the binary operation $\circ$ given by composition of functions is called the *symmetric group on* X. For each nonnegative integer n, let $[n] := \{1, \dots, n\}$. The symmetric group on $[n]$ is called the *symmetric group of degree* n and denoted by $\mathfrak{S}_n$.

Example. Here are six elements of $\mathfrak{S}_3$:



Note. Define the *factorial* of a natural number as follows: $0! = 1$, and for each integer $n > 0$, recursively define $n! = n(n-1)!$. Thus, $1! = 1$, $2! = 2 \cdot 1$, $3! = 3 \cdot 2 \cdot 1 = 6$, $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$, etc. Then the number of elements of $\mathfrak{S}_n = n!$ since we can uniquely determine every permutation σ by first choosing one of n values for $\sigma(1)$, then any of the remaining values $n-1$ for $\sigma(2)$, then one of the remaining $n-2$ values of $\sigma(3)$, etc.

Definition. Let $\sigma \in \mathfrak{S}_n$. The *permutation matrix corresponding to* σ is the $n \times n$ matrix P_σ whose i-th row is $e_{\sigma(i)}$. Another way of saying this is that, P_σ is obtained by permuting the columns of the identity matrix, I_n, according to σ: put e_j in column $\sigma(j)$.

Example. Let $\sigma \in \mathfrak{S}_3$ be defined by $\sigma(1) = 2$, $\sigma(2) = 3$, and $\sigma(3) = 1$. Then

$$P_\sigma = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Exercise. Let $\sigma, \tau \in \mathfrak{S}_n$ and let A be an $n \times n$ matrix.

(a) If the rows of A are $(r_1, \ldots, r_n)$, then the i-th row of $P_\sigma A$ is $r_{\sigma(i)}$. In other words, the multiplying on the left by P_σ permutes the rows of A in the same way that the rows of I_n are permuted to form P_σ. We leave it as an exercise to the reader to investigate the effect of multiplying A on the right by P.

(b) $P_\sigma e_{\sigma(i)} = e_i$, and $P_\sigma P_\tau = P_{\tau\circ\sigma}$. (Note that the order of σ and τ have switched.)

Rook placements. Permutation matrices are exactly those that have a single 1 in each row and in each column. Thus, if the 1s in a permutation matrix were replaced by rooks in the game of chess, then no rook would be attacking another. We sometimes call a permutation matrix a *rook placement*.

Definition. The *sign* of $\sigma \in \mathfrak{S}_n$ is

$$\mathrm{sign}(\sigma) = \det(P_\sigma) = \pm 1.$$

A permutation is *even* if its sign is 1 and *odd* if its sign is -1.

Every permutation matrix P_σ may be obtained from I_n through a sequence of transpositions of columns, i.e., a sequence in which each step consists of swapping two columns. In the example above, P_σ is formed by permuting columns 1, 2, and 3 of I_3 as follows:

$$\begin{tikzcd} 1 \arrow[dr] & 1 \\ 2 \arrow[dr] & 2 \\ 3 \arrow[uur] & 3 \end{tikzcd}.$$

This permutation could have been obtained from two transpositions:

$$\begin{tikzcd} 1 \arrow[r] & 1 \arrow[dr] & 1 \\ 2 \arrow[dr] & 2 \arrow[ur] & 2 \\ 3 \arrow[ur] & 3 \arrow[r] & 3 \end{tikzcd}.$$

Thus, every permutation matrix can be obtained as the product of permutation matrices corresponding to transpositions. Swapping two columns in a matrix changes the sign of the determinant. Therefore, even though a permutation σ may be realized in different ways as sequences of transpositions, the parity (evenness or oddness) of the number of transpositions required is well-defined: the number is even if $\det(P_\sigma) = 1$ and odd if $\det(P_\sigma) = -1$.

Theorem. Let A be an $n \times n$ matrix. Then

$$\det(A) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sign}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}.$$

Example. Consider the case $n = 3$. Then

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Each term $A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}$ in the formula in the theorem should be thought of as the product of the entries corresponding to a rook placement. The permutations, rook placements, and corresponding summands appear in Figure 22.1.[1]

Proof of permutation formula for the determinant. We want to compute

$$\det(A_{11}e_1 + A_{12}e_2 + \cdots + A_{1n}e_n, \quad \ldots \quad, A_{n1}e_1 + A_{n2}e_2 + \cdots + A_{nn}e_n).$$

Each of the n components in the above expression consists of n summands where each of the summands has the form $a_{ij}e_j$. Using the multilinear properties of the determinant, when we expand the above express, we get $n!$ terms, each of the form

$$A_{1j_1} A_{2j_2} \cdots A_{nj_n} \det(e_{1j_1}, e_{2j_2}, \ldots, e_{nj_n}).$$

If any pair of these e_{kj_k} is the same, this term will evaluate to 0. Thus, for the nonzero terms, $e_{1j_1}, e_{2j_2}, \ldots, e_{nj_n}$ must be some permutation of $e_1, \ldots, e_n$. We then have

$$\det(e_{1j_1}, e_{2j_2}, \ldots, e_{nj_n}) = \pm 1,$$

depending on the sign of the permutation σ defined by

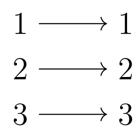
$$\sigma(1) = j_1, \ \sigma(2) = j_2, \ \ldots \ , \ \sigma(n) = j_n,$$

and we can write

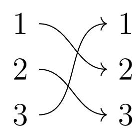
$$A_{1j_1} A_{2j_2} \cdots A_{n_{j_n}} \det(e_{1j_1}, e_{2j_2}, \ldots, e_{nj_n}) = \operatorname{sign}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}.$$

□

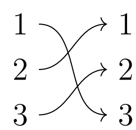
[1] Note that the sign of the permutation is also equal to $(-1)^c$ where c is the number of times two arrows cross in the diagram for the permutation in the left-most column.



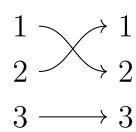
$$\begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & \boxed{a_{22}} & a_{23} \\ a_{31} & a_{32} & \boxed{a_{33}} \end{pmatrix} \qquad a_{11}a_{22}a_{33}$$



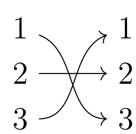
$$\begin{pmatrix} a_{11} & \boxed{a_{12}} & a_{13} \\ a_{21} & a_{22} & \boxed{a_{23}} \\ \boxed{a_{31}} & a_{32} & a_{33} \end{pmatrix} \qquad a_{12}a_{23}a_{31}$$



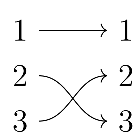
$$\begin{pmatrix} a_{11} & a_{12} & \boxed{a_{13}} \\ \boxed{a_{21}} & a_{22} & a_{23} \\ a_{31} & \boxed{a_{32}} & a_{33} \end{pmatrix} \qquad a_{13}a_{21}a_{32}$$



$$\begin{pmatrix} a_{11} & \boxed{a_{12}} & a_{13} \\ \boxed{a_{21}} & a_{22} & a_{23} \\ a_{31} & a_{32} & \boxed{a_{33}} \end{pmatrix} \qquad -a_{12}a_{21}a_{33}$$



$$\begin{pmatrix} a_{11} & a_{12} & \boxed{a_{13}} \\ a_{21} & \boxed{a_{22}} & a_{23} \\ \boxed{a_{31}} & a_{32} & a_{33} \end{pmatrix} \qquad -a_{13}a_{22}a_{31}$$



$$\begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & a_{22} & \boxed{a_{23}} \\ a_{31} & \boxed{a_{32}} & a_{33} \end{pmatrix} \qquad -a_{11}a_{23}a_{32}.$$

Figure 22.1: Computing the determinant of the 3×3 matrix $A = (a_{ij})$ via rook placements. The determinant is the sum of the terms in the right-most column.

See the next pages for all of the details of the above proof in the case $n = 3$.

Let's look at the proof again in the case $n = 3$. The i-th row vector of A is

$$r_i = a_{i1}e_1 + a_{i2}e_2 + a_{i3}e_3.$$

To compute the determinant of A we start by expanding using multilinearity:

$$\begin{aligned}
\det(A) &= \det(r_1, r_2, r_3) \\[1em]
&= \det(a_{11}e_1 + a_{12}e_2 + a_{13}e_3, a_{21}e_1 + a_{22}e_2 + a_{23}e_3, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\[1em]
&= \det(a_{11}e_1, a_{21}e_1 + a_{22}e_2 + a_{23}e_3, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\
&\qquad + \det(a_{12}e_2, a_{21}e_1 + a_{22}e_2 + a_{23}e_3, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\
&\qquad\qquad + \det(a_{13}e_3, a_{21}e_1 + a_{22}e_2 + a_{23}e_3, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\[1em]
&= \det(a_{11}e_1, a_{21}e_1, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\
&\qquad + \det(a_{11}e_1, a_{22}e_2, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\
&\qquad\qquad + \det(a_{11}e_1, a_{23}e_3, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\[1em]
&\qquad + \det(a_{12}e_2, a_{21}e_1, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\
&\qquad\qquad + \det(a_{12}e_2, a_{22}e_2, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\
&\qquad\qquad\qquad + \det(a_{12}e_2, a_{23}e_3, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\[1em]
&\qquad\qquad + \det(a_{13}e_3, a_{21}e_1, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\
&\qquad\qquad\qquad + \det(a_{13}e_3, a_{22}e_2, a_{31}e_1 + a_{32}e_2 + a_{33}e_3) \\
&\qquad\qquad\qquad\qquad + \det(a_{13}e_3, a_{23}e_3, a_{31}e_1 + a_{32}e_2 + a_{33}e_3)
\end{aligned}$$

There is one more step to go in the complete expansion, at which point, we'll have 27 terms. For completeness, I'll list these all on the next page.

$$
\begin{aligned}
&= \det(a_{11}e_1, a_{21}e_1, a_{31}e_1) \\
&\qquad + \det(a_{11}e_1, a_{21}e_1, a_{32}e_2) \\
&\qquad\qquad + \det(a_{11}e_1, a_{21}e_1, a_{33}e_3) \\[1ex]
&\qquad + \det(a_{11}e_1, a_{22}e_2, a_{31}e_1) \\
&\qquad\qquad + \det(a_{11}e_1, a_{22}e_2, a_{32}e_2) \\
&\qquad\qquad\qquad + \det(a_{11}e_1, a_{22}e_2, a_{33}e_3) \\[1ex]
&\qquad\qquad + \det(a_{11}e_1, a_{23}e_3, a_{31}e_1) \\
&\qquad\qquad\qquad + \det(a_{11}e_1, a_{23}e_3, a_{32}e_2) \\
&\qquad\qquad\qquad\qquad + \det(a_{11}e_1, a_{23}e_3, a_{33}e_3) \\[2ex]
&+ \det(a_{12}e_2, a_{21}e_1, a_{31}e_1) \\
&\qquad + \det(a_{12}e_2, a_{21}e_1, a_{32}e_2) \\
&\qquad\qquad + \det(a_{12}e_2, a_{21}e_1, a_{33}e_3) \\[1ex]
&\qquad + \det(a_{12}e_2, a_{22}e_2, a_{31}e_1) \\
&\qquad\qquad + \det(a_{12}e_2, a_{22}e_2, a_{32}e_2) \\
&\qquad\qquad\qquad + \det(a_{12}e_2, a_{22}e_2, a_{33}e_3) \\[1ex]
&\qquad\qquad + \det(a_{12}e_2, a_{23}e_3, a_{31}e_1) \\
&\qquad\qquad\qquad + \det(a_{12}e_2, a_{23}e_3, a_{32}e_2) \\
&\qquad\qquad\qquad\qquad + \det(a_{12}e_2, a_{23}e_3, a_{33}e_3) \\[2ex]
&+ \det(a_{13}e_3, a_{21}e_1, a_{31}e_1) \\
&\qquad + \det(a_{13}e_3, a_{21}e_1, a_{32}e_2) \\
&\qquad\qquad + \det(a_{13}e_3, a_{21}e_1, a_{33}e_3) \\[1ex]
&\qquad + \det(a_{13}e_3, a_{22}e_2, a_{31}e_1) \\
&\qquad\qquad + \det(a_{13}e_3, a_{22}e_2, a_{32}e_2) \\
&\qquad\qquad\qquad + \det(a_{13}e_3, a_{22}e_2, a_{33}e_3) \\[1ex]
&\qquad\qquad + \det(a_{13}e_3, a_{23}e_3, a_{31}e_1)
\end{aligned}
$$

$$+ \det(a_{13}e_3, a_{23}e_3, a_{32}e_2)$$
$$+ \det(a_{13}e_3, a_{23}e_3, a_{33}e_3)$$

Use linearity to pull out the constants:

$$
\begin{aligned}
&= a_{11}a_{21}a_{31}\ \det(e_1, e_1, e_1)\\
&\qquad + a_{11}a_{21}a_{32}\ \det(e_1, e_1, e_2)\\
&\qquad\qquad + a_{11}a_{21}a_{33}\ \det(e_1, e_1, e_3)\\[1ex]
&\qquad + a_{11}a_{22}a_{31}\ \det(e_1, e_2, e_1)\\
&\qquad\qquad + a_{11}a_{22}a_{32}\ \det(e_1, e_2, e_2)\\
&\qquad\qquad\qquad + a_{11}a_{22}a_{33}\ \det(e_1, e_2, e_3)\\[1ex]
&\qquad\qquad + a_{11}a_{23}a_{31}\ \det(e_1, e_3, e_1)\\
&\qquad\qquad\qquad + a_{11}a_{23}a_{32}\ \det(e_1, e_3, e_2)\\
&\qquad\qquad\qquad\qquad + a_{11}a_{23}a_{33}\ \det(e_1, e_3, e_3)\\[3ex]
&+ a_{12}a_{21}a_{31}\ \det(e_2, e_1, e_1)\\
&\qquad + a_{12}a_{21}a_{32}\ \det(e_2, e_1, e_2)\\
&\qquad\qquad + a_{12}a_{21}a_{33}\ \det(e_2, e_1, e_3)\\[1ex]
&\qquad + a_{12}a_{22}a_{31}\ \det(e_2, e_2, e_1)\\
&\qquad\qquad + a_{12}a_{22}a_{32}\ \det(e_2, e_2, e_2)\\
&\qquad\qquad\qquad + a_{12}a_{22}a_{33}\ \det(e_2, e_2, e_3)\\[1ex]
&\qquad\qquad + a_{12}a_{23}a_{31}\ \det(e_2, e_3, e_1)\\
&\qquad\qquad\qquad + a_{12}a_{23}a_{32}\ \det(e_2, e_3, e_2)\\
&\qquad\qquad\qquad\qquad + a_{12}a_{23}a_{33}\ \det(e_2, e_3, e_3)\\[3ex]
&+ a_{13}a_{21}a_{31}\ \det(e_3, e_1, e_1)\\
&\qquad + a_{13}a_{21}a_{32}\ \det(e_3, e_1, e_2)\\
&\qquad\qquad + a_{13}a_{33}a_{21} \det(e_3, e_1, e_3)\\[1ex]
&\qquad + a_{13}a_{22}a_{31}\ \det(e_3, e_2, e_1)\\
&\qquad\qquad + a_{13}a_{22}a_{32}\ \det(e_3, e_2, e_2)\\
&\qquad\qquad\qquad + a_{13}a_{22}a_{33}\ \det(e_3, e_2, e_3)\\[1ex]
&\qquad\qquad + a_{13}a_{23}a_{31}\ \det(e_3, e_3, e_1)
\end{aligned}
$$

$$
\begin{aligned}
&+ a_{13}a_{23}a_{32}\ \det(e_3, e_3, e_2)\\
&\qquad + a_{13}a_{23}a_{33}\ \det(e_3, e_3, e_3)
\end{aligned}
$$

Now we use the alternating property of the determinant. If any row is repeated, the determinant is 0. Getting rid of those terms leaves:

$$\begin{aligned}\det(A) = a_{11}a_{22}a_{33} \det(e_1, e_2, e_3) \\ + a_{11}a_{23}a_{32} \det(e_1, e_3, e_2) \\ + a_{12}a_{21}a_{33} \det(e_2, e_1, e_3) \\ + a_{12}a_{23}a_{31} \det(e_2, e_3, e_1) \\ + a_{13}a_{21}a_{32} \det(e_3, e_1, e_2) \\ + a_{13}a_{22}a_{31} \det(e_3, e_2, e_1).\end{aligned}$$

Next notice that each determinant appearing above is the determinant of a permutation matrix. For instance, the term

$$a_{12}a_{23}a_{31} \det(e_2, e_3, e_1)$$

contains $\det(e_2, e_3, e_1)$, which is the determinant of the permutation matrix for the permutation $\sigma(1) = 2$, $\sigma(2) = 3$, and $\sigma(3) = 1$. We have

$$\begin{aligned}a_{12}a_{23}a_{31} \det(e_2, e_3, e_1) &= a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)} \det(P_\sigma) \\ &= a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)}\mathrm{sign}(\sigma).\end{aligned}$$

In this way, the six terms in the sum can be expressed as follows:

$$\begin{aligned}\det(A) &= \sum_{\sigma \in \mathfrak{S}_3} \det(P_\sigma) a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)} \\ &= \sum_{\sigma \in \mathfrak{S}_3} \mathrm{sign}(P_\sigma) a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)}.\end{aligned}$$

Week 8, Friday: Existence and uniqueness of the determinant

Laplace expansion of the determinant. Let A be an $n \times n$ matrix. For each $i, j \in \{1, 2, \ldots, n\}$, define A^{ij} to be the matrix formed by removing the i-th row and j-th column from A. Fix $k \in \{1, 2, \ldots, n\}$. Then

$$\det(A) = \sum_{j=1}^{n} (-1)^{k+j} A_{kj} \det(A^{kj}).$$

This expresses $\det(A)$ in terms of an alternating sum of determinants of $(n-1) \times (n-1)$ matrices. We call this *expanding* $\det(A)$ *along the k-th row.* Applying the formula recursively leads to a complete evaluation of $\det(A)$. Since, $\det(A) = \det(A^t)$, you can also calculate the determinant by recursively expanding along *columns.*

Example. Let

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

Let's calculate the determinant by expanding along the first row:

$$\det(A) = 1 \cdot \det \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} - 2 \cdot \det \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} + 3 \cdot \det \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

$$= (-1) - 2(1) + 3(2) = 3.$$

To check, let's expand along the second row, instead, noting the signs:

$$\det(A) = -2 \cdot \det \begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} 0 \cdot \det \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$$

$$= -2(-1) + 0(-2) - 1(-1) = 3.$$

Finally, let's expand along the third *column*:

$$\det(A) = 3 \cdot \det \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} + 1 \cdot \det \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix}$$

$$= 3(2) - 1(-1) + 1(-4) = 3.$$

Note: if your matrix has a particular row or column with a lot of 0s in it, you might want to expand along that row or column since a lot of the terms will be 0. For example, to compute

$$\det \begin{pmatrix} 1 & 3 & 0 \\ 3 & 2 & 3 \\ 1 & 4 & 0 \end{pmatrix},$$

expand along the third column:

$$0\,(\text{blah}) - 3\det \begin{pmatrix} 1 & 3 \\ 1 & 4 \end{pmatrix} + 0\,(\text{blah}) = -3(1) = -3.$$

The "blah"s are there instead of explicit determinants since they are being multiplied by 0. Their exact values don't matter, so we don't need to waste time calculating them. We will not prove the formula for the Laplace expansion. It is very similar to that for the permutation expansion.

Existence and uniqueness of the determinant. Recall the definition that started our discussion of the determinant:

Definition. The *determinant* is a multilinear, alternating function $\det\colon M_{n\times n}(F) \to F$ of the rows of square matrix, normalized so that its value on the identity matrix is 1.

The definition says "the determinant", but for all we knew, there could be several different functions $M_{n\times n}(F) \to F$ all satisfying the criteria of being multilinear, alternating, and normalized. Or, it is possible there are no functions that satisfy the criteria? So the definition requires us to prove that, in fact, there exists exactly one determinant function (for each n).

Just after defining the determinant, we showed that if $d\colon M_{n\times n}(F) \to F$ is any multilinear, alternating, normalized function, then a choice of a row reduction for $A \in M_{n\times n}(F)$ determines the value of $d(A)$. The subtlety here is that, there are many different sequence of row operations that would produced the row echelon form for A. Do each of these produce the same value for $d(A)$ (in other words, is d *well-defined*)? We never proved that they would.

So let's begin again and consider the particular function $d\colon M_{n\times n}(F) \to F$ defined recursively as the Laplace expansion of a matrix along its first row:

$$d(A) := \sum_{j=1}^{n} (-1)^{1+j} A_{1j}\, d(A^{1j}) \tag{23.1}$$

if $n \geq 1$, and by $d(A) = a$ if $A = [a]$ is a 1×1 matrix. This function d is well-defined—there are no choices to be made is in calculation.

Exercise. Prove that d is multilinear, alternating, and normalized (i.e., its value at I_n is 1).

Thus, we see there exists at least one determinant function.

Having defined d by (23.1), now note that in addition to calculating d using the given recursive formula, since d is multilinear, alternating, and normalize, its value can be determined via row reductions, just as before. What's new now is that we see that no matter which choices are made in the row reduction, we must get the value determined by (23.1).

In sum, we have shown that a multilinear, alternating, normalized function exists and is unique. Its value is completely determined by choosing any sequence of row operations reducing a matrix to its row echelon form, and the choice of the sequence of row operations does not matter. So far, we have three different methods for calculating the determinant: using row operations, summing over permutations, and via Laplace expansion along any row or column.

BONUS CONTENT

Generalized Laplace expansion. Let $A \in M_{n\times n}(F)$, and fix a subset of k rows $r_{i_1}, \dots, r_{i_k}$ of A where $1 \leq k \leq n$. Let $I = \{i_1, \dots, i_k\}$ be the indices of these rows. For any subset $J \subseteq \{j_1, \dots, j_k\}$, define $|J| := j_1 + \cdots + j_k$, and define

$$\begin{aligned} A^{IJ} &= \text{the } k \times k \text{ submatrix of } A \text{ formed by the intersection of} \\ &\quad\ \text{rows indexed by } I \text{ and the columns indexed by } J \\ \bar{A}^{IJ} &= \text{the } (n-k) \times (n-k) \text{ submatrix of } A \text{ formed by the} \\ &\quad\ \text{intersection of rows indexed by } \{1, \dots, n\} \setminus I \text{ and the} \\ &\quad\ \text{columns indexed by } \{1, \dots, n\} \setminus J. \end{aligned}$$

Then

$$\det(A) = \sum_J (-1)^{|I|+|J|} \det(A^{IJ}) \det(\bar{A}^{IJ})$$

where the sum is over all k-element subsets J of $\{1, \dots, n\}$.[1]

Example. The case where $k = 1$ is the ordinary Laplace expansion formula.

[1] Since $\det(A) = \det(A^t)$, there is a similar formula for expansion along a fixed set of k columns.

Example. Let

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

We will compute $\det(A)$ using the generalize Laplace expansion along the first two rows of A. So, using the notation from above, $I = \{1,2\} \subset \{1,2,3,4\}$. There are six choices for a pair of columns:

$$J \in (\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}).$$

The term in the expansion corresponding to $J = \{1,3\}$ would be

$$\begin{aligned}(-1)^{|I|+|J|}\det(A^{IJ})\det(\bar{A}^{IJ}) &= (-1)^{(1+2)+(1+3)}\det\begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}\det\begin{pmatrix} 1 & 6 \\ 6 & 3 \end{pmatrix} \\ &= (-1)(1\cdot 2 - 0\cdot 2)(1\cdot 3 - 6\cdot 6) = 66.\end{aligned}$$

The entire expansion is

$$\begin{aligned}\det(A) &= \sum_J (-1)^{|I|+|J|}\det(A^{IJ})\det(\bar{A}^{IJ}) \\ &= \sum_J (-1)^{3+|J|}\det(A^{\{1,2\}J})\det(\bar{A}^{\{1,2\}J}) \\ &= (-1)^{3+(1+2)}\det(A^{\{1,2\}\{1,2\}})\det(\bar{A}^{\{3,4\}\{3,4\}}) + (-1)^{3+(1+3)}\det(A^{\{1,2\}\{1,3\}})\det(\bar{A}^{\{3,4\}\{2,4\}}) \\ &\quad + (-1)^{3+(1+4)}\det(A^{\{1,2\}\{1,4\}})\det(\bar{A}^{\{3,4\}\{2,3\}}) + (-1)^{3+(2+3)}\det(A^{\{1,2\}\{2,3\}})\det(\bar{A}^{\{3,4\}\{1,4\}}) \\ &\quad + (-1)^{3+(2+4)}\det(A^{\{1,2\}\{2,4\}})\det(\bar{A}^{\{3,4\}\{1,3\}}) + (-1)^{3+(3+4)}\det(A^{\{1,2\}\{3,4\}})\det(\bar{A}^{\{3,4\}\{1,2\}}) \\ &= \det(A^{\{1,2\}\{1,2\}})\det(\bar{A}^{\{3,4\}\{3,4\}}) - \det(A^{\{1,2\}\{1,3\}})\det(\bar{A}^{\{3,4\}\{2,4\}}) \\ &\quad + \det(A^{\{1,2\}\{1,4\}})\det(\bar{A}^{\{3,4\}\{2,3\}}) + \det(A^{\{1,2\}\{2,3\}})\det(\bar{A}^{\{3,4\}\{1,4\}}) \\ &\quad - \det(A^{\{1,2\}\{2,4\}})\det(\bar{A}^{\{3,4\}\{1,3\}}) + \det(A^{\{1,2\}\{3,4\}})\det(\bar{A}^{\{3,4\}\{1,2\}})\end{aligned}$$

Continuing the calculation:

$$A = \begin{pmatrix} 1 & 7 & 0 & 5 \\ 2 & 2 & 2 & 2 \\ 5 & 1 & 4 & 6 \\ 0 & 6 & 7 & 3 \end{pmatrix}$$

$$\begin{aligned}
\det(A) &= \det(A^{\{1,2\}\{1,2\}})\det(\bar{A}^{\{3,4\}\{3,4\}}) - \det(A^{\{1,2\}\{1,3\}})\det(\bar{A}^{\{3,4\}\{2,4\}}) \\
&\quad + \det(A^{\{1,2\}\{1,4\}})\det(\bar{A}^{\{3,4\}\{2,3\}}) + \det(A^{\{1,2\}\{2,3\}})\det(\bar{A}^{\{3,4\}\{1,4\}}) \\
&\quad - \det(A^{\{1,2\}\{2,4\}})\det(\bar{A}^{\{3,4\}\{1,3\}}) + \det(A^{\{1,2\}\{3,4\}})\det(\bar{A}^{\{3,4\}\{1,2\}}) \\
&= \det\begin{pmatrix} 1 & 7 \\ 2 & 2 \end{pmatrix}\det\begin{pmatrix} 4 & 6 \\ 7 & 3 \end{pmatrix} - \det\begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}\det\begin{pmatrix} 1 & 6 \\ 6 & 3 \end{pmatrix} \\
&\quad + \det\begin{pmatrix} 1 & 5 \\ 2 & 2 \end{pmatrix}\det\begin{pmatrix} 1 & 4 \\ 6 & 7 \end{pmatrix} + \det\begin{pmatrix} 7 & 0 \\ 2 & 2 \end{pmatrix}\det\begin{pmatrix} 5 & 6 \\ 0 & 3 \end{pmatrix} \\
&\quad - \det\begin{pmatrix} 7 & 5 \\ 2 & 2 \end{pmatrix}\det\begin{pmatrix} 5 & 4 \\ 0 & 7 \end{pmatrix} + \det\begin{pmatrix} 0 & 5 \\ 2 & 2 \end{pmatrix}\det\begin{pmatrix} 5 & 1 \\ 0 & 6 \end{pmatrix} \\
&= (-12)(-30) - (2)(-33) + (-8)(-17) + (14)(15) - (4)(35) + (-10)(30) \\
&= 332.
\end{aligned}$$

Example. Let

$$A = \begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 0 \\ 0 & 1 & 7 & 0 & 0 \\ 1 & 2 & 1 & 4 & 7 \\ 3 & 4 & 2 & 9 & 3 \end{pmatrix}.$$

The generalized Laplace expansion along the first three rows has only two nonzero terms, yielding

$$\det(A) = \det\begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 4 \\ 0 & 1 & 7 \end{pmatrix}\det\begin{pmatrix} 4 & 7 \\ 9 & 3 \end{pmatrix}.$$

Week 9, Monday Parametrizing linear subspaces

Parametrizing linear subspaces

Let V be a finite-dimension vector space of dimension n, and let $1 \leq k \leq n$. Our goal today is to construct a geometric object whose points are in one-to-one correspondence with the k-dimensional subspaces of V.

Projective space. We start with the case $k = 1$. Let $W \subseteq V$ be a one-dimensional subspace of V, and let $\{w\}$ be a basis for W. Then $w \neq 0$ and

$$W = \operatorname{Span}\{w\} = \{\lambda w : \lambda \in F\}.$$

One might be tempted to say that W *is* w in the sense that W is completely determined by w. From that point of view, $V \setminus \{0\}$ would be exactly the geometric object we are looking for: it points correspond to one-dimensional subspace (by take the span of a point). We have a mapping of sets

$$\begin{aligned} V \setminus \{0\} &\to \text{one-dimensional subspaces of } V \\ u &\mapsto \operatorname{Span}\{u\}. \end{aligned}$$

There only one problem: the above mapping is not a bijection—so it is not the one-to-one correspondence we are seeking. It is surjective since every one-dimensional subspace is spanned by some vector, but it is not injective since a one-dimensional space may have many different bases. How far away is the mapping from being injective? When do two different nonzero vectors have the same span? The answer: exactly when the two vectors are scalar multiples of each other.

The above discussion motivates the following equivalence relation:

Definition. If $w, w' \in V \setminus \{0\}$, write $w \sim w'$ if there exists a *nonzero* scalar $\lambda \in F$ such that $w' = \lambda w$.

Exercise. The relation $\sim$ is an equivalence relation on $V \setminus \{0\}$.

Definition. The set of equivalences classes for $\sim$ is *projective space on* V, denoted $\mathbb{P}(V)$.

Using the usual notation for a set modulo an equivalence relation, we write

$$\mathbb{P}(V) := \quad (V \setminus \{0\}) \,/\!\sim \quad = \{[w] : w \in V \setminus \{0\}\}$$

where $[w]$ is the equivalence class of w, i.e., the set of all nonzero scalar multiples of w. We then have a bijection of sets

$$\begin{aligned} \mathbb{P}(V) &\to \text{one-dimensional subspaces of } V \\ [u] &\mapsto \operatorname{Span}\{u\}. \end{aligned}$$

From now on, we use this bijection to identify each point $[u] \in \mathbb{P}(V)$ with its corresponding one-dimensional subspace $\operatorname{Span}(u) \subseteq V$.

We will next consider a special case, which should justify calling projective space a "geometric" object.

Definition. *Real projective n-space* $\mathbb{P}(\mathbb{R}^{n+1})$, denoted $\mathbb{P}^n$.

A point in $\mathbb{P}^n$ is is a one-dimensional subspace of $\mathbb{R}^{n+1}$, which is the same as a line through the origin. So we think of $\mathbb{P}^n$ as the set of lines through the origin in $\mathbb{R}^{n+1}$.

We explain now why $\mathbb{P}^n$ should be considered n-dimensional in some sense. For $i = 1, \ldots, n+1$, define

$$U_i = \{[(a_1, \ldots, a_{n+1})] \in \mathbb{P}^n : a_i \neq 0\}\,.$$

Note that U_i is well-defined: if $[(a_1, \ldots, a_{n+1})] = [(b_1, \ldots, b_{n+1})]$, then there exists a nonzero $\lambda \in \mathbb{R}$ such that $(a_1, \ldots, a_{n+1}) = \lambda(b_1, \ldots, b_{n+1})$. Thus, $a_i = \lambda b_i$, and it follows that $a_i \neq= 0$ if and only if $b_i \neq 0$. Let $[(a_1, \ldots, a_{n+1})] \in \mathbb{P}^n$. Then there exists at least one i such that $a_i \neq 0$. We can then scale by $\frac{1}{a_i}$ to get another name for the same point in projective space:

$$\begin{aligned} (a_1, \ldots, a_n) \sim \frac{1}{a_i}(a_1, \ldots, a_{n+1}) \quad \Rightarrow \quad [(a_1, \ldots, a_n)] &= [(a_1/a_i, \ldots, a_i/a_i, \ldots, a_{n+1}/a_i)]) \\ &= [(a_1/a_i, \ldots, 1, \ldots, a_{n+1}/a_i)]). \end{aligned}$$

In this way, every point in $U_i \subset \mathbb{P}^n$ is represented by whose i-th coordinate is 1. We have bijection $\mathbb{R}^n \leftrightarrow U_i$ that sends $(b_1, \ldots, b_n)$ to the equivalence class $[(b_1, \ldots, b_{i-1}, 1, b_i, \ldots, b_n)]$. (We have just squeezed a 1 between b_{i-1} and b_{i+1}.) The point, finally, is that it only takes n number to identify any point in U_i.

Grassmannians. We know generalize the above discussion to find a space parametrizing k-dimensional subspaces of V for any $1 \leq k \leq n$. Let $W \subseteq V$, and let $(w_1, \ldots, w_k)$ be a basis for W. We place these basis vectors in as the rows of a $k \times n$ matrix:

$$M = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{k1} & \cdots & w_{kn} \end{pmatrix}$$

where $w_i = (w_{i1}, \dots, w_{in})$. Conversely, if M is any $k \times n$ matrix of rank k, then the rowspace of M is a subspace of V of dimension k. So we get a surjection

$$\begin{aligned} k \times n \text{ matrices of rank } k &\to k\text{-dimensional subspace of } V \\ M &\mapsto \text{rowspace}(M). \end{aligned}$$

This mapping is not injective because a k-dimensional subspace of V may have many different bases. In fact, given M with rows $w_1, \dots, w_k$ spanning a k-dimensional subspace of V, we can describe all of the matrices M' with the same row space as M: they are exactly the matrices obtained by performing row operations to M. This, in turn, means that M' must be of the form PM where P is $k \times k$ invertible matrix. (The matrix P is the product of elementary matrices corresponding to the row operations that transform M to M'.) We proceed as we did earlier for projective space (i.e., for the case $k = 1$).

Definition. If M and M' are $k \times n$ matrices of rank k, write $M \sim M'$ if there exists an invertible $k \times k$ matrix P such that $M' = PM$.

Exercise. The relation $\sim$ is an equivalence relation on the set of $k \times n$ matrices of rank k.

Definition. The set of equivalences classes for $\sim$ is the *Grassmannian*, denoted $G(k, n)$.

Example. If $F = \mathbb{R}$, we have $\mathbb{P}^n = G(1, n + 1)$.

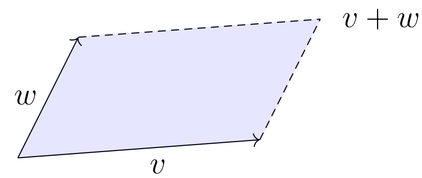
If J is a subset of $\{1, \dots, n\}$ of size k, let U_J be the subset of $G(k, n)$ consisting of equivalence classes of matrices M such that the submatrix of M consisting of the columns with indices in J has rank k. (Note: this notion is well-defined, just as the U_i were well-defined, above.) Then each element of U_J has a unique representative M that is a $k \times n$ matrix whose submatrix consisting of the columns in J is the identity matrix.

Week 9, Wednesday: Determinants and volume

The *parallelogram spanned by* $v, w \in \mathbb{R}^2$ is

$$P = \{\lambda v + \mu w \colon \lambda, \mu \in [0,1]\}$$

where $[0,1] = \{x \in \mathbb{R} \colon 0 \leq x \leq 1\}$:



Theorem. Let $A(v,w)$ be the area of the parallelogram spanned by $v, w \in \mathbb{R}^2$. Then

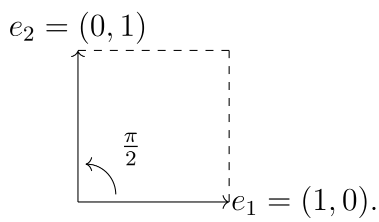
$$A(v,w) = |\det(v,w)|,$$

where $\det(v,w)$ is the determinant of the matrix with rows v and w.

Note: Since the determinant of a square matrix and its transpose are the same, $A(v,w)$ is also the absolute value of the matrix whose *columns* are v and w.

Proof of theorem. Define $SA(v,w)$ to be the signed area defined in the worksheet. We show that SA satisfies the properties required of a determinant function. Then, since the determinant is unique, it follows that $SA(v,w) = \det(v,w)$ and the result follows since $A(v,w) = |SA(v,w)|$.

- **Normalized.** We have $SA(e_1, e_2) = 1$:

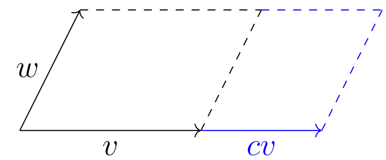


The sign is positive since the angle from e_1 to e_2 is less than π.

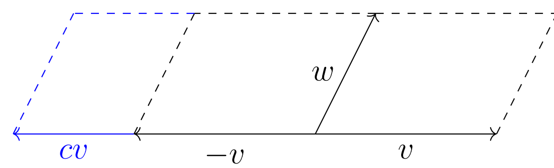
- **Alternating.** We have $SA(v, v) = 0$ since in this case, the corresponding parallelogram is degenerate.

- **Multilinear.**
  - $SA(cv, w) = cSA(v, w)$ and $SA(v, cw) = cSA(v, w)$:

    $c > 0$



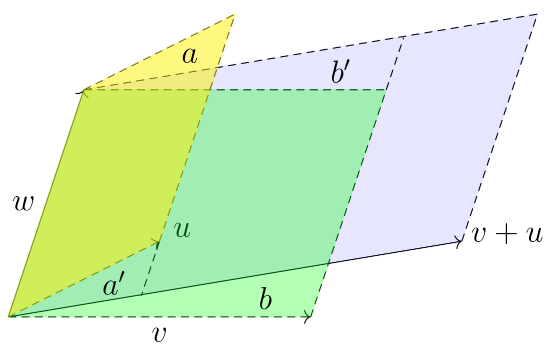
$c < 0$



The areas are scaled by $|c|$ in either case since the base is scaled by $|c|$ and the height does not change. The drawing assumes that the angle from v to w is less than π. There is a similar drawing for the case where the angle is greater than π. Either way, in the case where $c < 0$ note that although $SA(cv, w)$ and $SA(v, w)$ have opposite signs, $SA(cv, w)$ and $cSA(v, w)$ have the same sign.

Similar drawings show that $SA(v, cw) = cSA(v, w)$.

  - $SA(v + u, w) = SA(v, w) + SA(u, w)$:



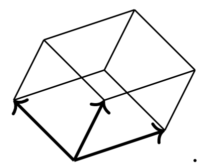
Note how to dissect the u-w and v-w parallelograms to get the $(v + u)$-w parallelogram: Cut section a in the u-w parallelogram and place it at section a', then cut section b in the v-w parallelogram and place it at section b'. The result is two parallelograms that can exactly cover the $(v + u)$-w parallelogram.

Of course, our drawing is just one case among the many possible angles between pairs of v, u, and w.

$\square$

Definition. The *parallelepided spanned by* $v_1, \dots, v_n \in \mathbb{R}^n$ is

$$P = \{\lambda_1 v_1 + \cdots + \lambda_n v_n : \lambda_i \in [0,1] \text{ for } i = 1, \dots, n\},$$



It turns out that the volume of P is given by the determinant of the matrix whose row (or columns) are $v_1, \dots, v_n$:

$$\operatorname{vol}(P) = |\det(v_1, \dots, v_n)|.$$

Note that one of the vertices of P is the origin (set $\lambda_1 = \cdots = \lambda_n = 0$). To get an arbitrary parallelepiped in $\mathbb{R}^n$ we can just translate by any vector $u \in \mathbb{R}^n$:

$$P + u := \{p + u : p \in P\} = \{\lambda_1 v_1 + \cdots + \lambda_n v_n + u : 0 \le \lambda_i \le 1 \text{ for } i = 1, \dots, n\}.$$

The volume does not change:

$$\operatorname{vol}(P + u) = |\det(v_1, \dots, v_n)|.$$

Theorem. Let P be the parallelepided spanned by $v_1, \dots, v_n \in \mathbb{R}^n$. Let $A \in M_{n\times n}(\mathbb{R})$, and let $L_A\colon \mathbb{R}^n \to \mathbb{R}^n$ be the corresponding linear function, $L_A(x) = Ax$. Then $L_A(P)$ is the parallelepiped spanned by the vectors $Av_1, \dots, Av_n$, and

$$\operatorname{vol}(L_A(P)) = |\det(A)|\operatorname{vol}(P).$$

Moreover, $L_A(P + u) = L_A(P) + L_A(u)$. Thus, *application of L_A scales the volumes of parallelepipeds in $\mathbb{R}^n$ be a factor of* $|\det(A)|$.

Proof. We have $x \in L_A(P)$ if and only if there exist $\lambda_1, \dots, \lambda_n \in [0,1]$ such that

$$x = A(\lambda_1 v_1 + \cdots + \lambda_n v_n) = \lambda_1 Av_1 + \cdots + \lambda_n Av_n,$$

i.e., if and only if x is in the parallelepiped determined by $Av_1, \dots, Av_n$.
Let B be the matrix with columns $v_1, \dots, v_n$. Then $\operatorname{vol}(P) = |\det(B)|$, Note that AB is the matrix whose columns are $Av_1, \dots, Av_n$. It follows that

$$\operatorname{vol}(L_A(P)) = |\det(AB)| = |\det(A)||\det(B)|.$$

We have $L_A(P + u) = L_A(P) + L_A(u)$ since L_A is linear. $\square$

Remark. To approximate the volume of an arbitrary shape in $\mathbb{R}^n$, one can try to dissect the shape into a union of parallelepipeds. One definition of the volume of an arbitrary shape is derived by taking limits of such approximations. One can then ask how the volume of a shape S changes under the application of a function $f : \mathbb{R}^n \to \mathbb{R}^n$. If the function is "nice" (differentiable), then at each point p in the shape, one creates a linear approximation $Df(p)$ of the function f (called the *derivative of f at p*), akin to L_A, above. Further assuming that f is injective, the volume of the image of the shape is then given by the *change of variables* formula in multivariable calculus:

$$\mathrm{vol}(f(S)) = \int_{p \in S} |\det(Df(p))|.$$

In light of the theorem we just proved, the determinant $|Df(p)|$ should be thought of as a scaling factor. It tells us how much f scales volumes (infinitesimally) at p.

Week 9, Friday: Eigenvectors and eigenvalues

Definition. Let $f\colon V \to V$ be a linear transformation of a vector space V over F. A *nonzero* vector $v \in V$ is an *eigenvector* for f with *eigenvalue* $\lambda \in F$ if

$$f(v) = \lambda v.$$

If $A \in M_{n\times n}(F)$, a *nonzero* vector $v \in F^n$ is an *eigenvector* for A with *eigenvalue* $\lambda \in F$ if

$$Av = \lambda v.$$

Thus, eigenvectors and eigenvalues for A are the same as eigenvectors and eigenvalues for the associated linear function $f_A\colon F^n \to F^n$ (defined by $f_A(v) = Av$).

Here is why we like eigenvectors: Suppose that $\alpha = \langle v_1, \dots, v_n\rangle$ is an ordered basis of eigenvectors for $f\colon V \to V$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, i.e., $f(v_i) = \lambda_i v_i$ for $i = 1, \dots, n$. Then the matrix $[f]_\alpha^\alpha$ representing f with respect to the basis α for the domain and codomain is the diagonal matrix $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$.

Example. Let

$$A = \begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix}.$$

with corresponding linear function

$$\begin{aligned} f_A\colon \mathbb{R}^2 &\to \mathbb{R}^2 \\ (x,y) &\mapsto (-x+2y, -6x+6y). \end{aligned}$$

It turns out that $(2,3)$ and $(1,2)$ are eigenvectors for f_A with eigenvalues 2 and 3, respectively:

$$\begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix} = 2\begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Find the matrix representing f_A with respect to the ordered basis

$$\alpha = \langle (2,3), (1,2) \rangle.$$

To do this we write the image of each vector in α as a linear combination of the vectors in α and pull off the coefficients to create columns:

$$\begin{aligned} f_A(2,3) &= 2(2,3) = 2 \cdot (2,3) + 0 \cdot (1,2) \\ f_A(1,2) &= 3(1,2) = 0 \cdot (2,3) + 3 \cdot (1,2). \end{aligned}$$

Hence,

$$[f_A]_\alpha^\alpha = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = \operatorname{diag}(2,3).$$

That is the point: a basis of eigenvectors gives a matrix representative that is diagonal, which is the simplest type of matrix to think about. Let's think abstractly about what just happened. The matrix A is the matrix representing f_A with respect to the standard basis, and the matrix

$$D = \operatorname{diag}(2,3) = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

represents f_A with respect to the basis α. Let ϕ_α be the mapping that takes coordinates with respect to α. We get the commutative diagram:

$$\begin{tikzcd}
(1,2) \arrow[rrrrr, bend left=10] \arrow[ddddd, bend right=10] & & & & & 3(1,2) \arrow[ddddd, bend left=10] \\
& (2,3) \arrow[rrr, bend left=10] \arrow[ddd, bend right=10] & & & 2(2,3) \arrow[ddd, bend left=10] & \\
& & \mathbb{R}^2 \arrow[r, "A"] \arrow[d, "\phi_\alpha"', "\wr"] & \mathbb{R}^2 \arrow[d, "\phi_\alpha", "\wr"'] & & \\
& & \mathbb{R}^2 \arrow[r, "D"] & \mathbb{R}^2. & & \\
& (1,0) \arrow[rrr, bend right=10] & & & 2(1,0) & \\
(0,1) \arrow[rrrrr, bend right=10] & & & & & 3(0,1)
\end{tikzcd}$$

Reviewing something we talked about earlier in the semester: The matrix for ϕ_α would be a bit of a chore to write down. It's j-column would be the image of e_j. So we would have to write each e_j as a linear combination of the basis vectors in α. However, the inverse of ϕ_α is *easy* to write down. Take a look at the commutative diagram. By construction of ϕ_α, we have

$$\phi_\alpha^{-1}(1,0) = (2,3) \quad \text{and} \quad \phi_\alpha^{-1}(0,1) = (1,2).$$

So the matrix for ϕ_α^{-1} is

$$P = \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}.$$

So the matrix for ϕ_α is P^{-1}. Therefore, another way to write the commutative diagram is

$$\begin{tikzcd} \mathbb{R}^2 \arrow[r, "A"] \arrow[d, "P^{-1}"', "\wr"] & \mathbb{R}^2 \arrow[d, "P^{-1}", "\wr"'] \\ \mathbb{R}^2 \arrow[r, "D"] & \mathbb{R}^2. \end{tikzcd}$$

From contemplating this diagram, we see that

$$D = P^{-1}AP.$$

Summary: having found eigenvectors $(2, 3)$ and $(1, 2)$, we place those eigenvectors as columns in a matrix P, and then $P^{-1}AP$ is a diagonal matrix with the corresponding eigenvalues on the diagonal.

To generalize:

Let $A \in M_{n\times n}(F)$ with corresponding linear function $f_A\colon F^n \to F^n$. Suppose $\alpha = \langle v_1, \ldots, v_n\rangle$ is an ordered basis of eigenvectors with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, i.e., $Av_i = \lambda_i v_i$ for $i = 1, \ldots, n$. Let P be the matrix whose columns are $v_1, \ldots, v_n$. Then

$$P^{-1}AP = D,$$

where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, and we have a commutative diagram

$$\begin{tikzcd} F^n \arrow[r, "A"] \arrow[d, "P^{-1}"', "\wr"] & F^n \arrow[d, "P^{-1}", "\wr"'] \\ F^n \arrow[r, "D"] & F^n. \end{tikzcd}$$

How does one find eigenvectors and eigenvalues? Let $A \in M_{n\times n}(F)$ with corresponding function $f_A\colon F^n \to F^n$ (so $f_A(v) := Av$). We are looking for a nonzero vector $v \in F^n$ and a scalar λ such that $Av = \lambda v$. To achieve that, the following argument is of **central importance**:

$$Av = \lambda v \quad \Leftrightarrow \quad (A - \lambda I_n)v = 0 \quad \Leftrightarrow \quad v \in \ker(A - \lambda v).$$

This says that:

$\lambda \in F$ **is an eigenvalue for** A **if and only if** $\ker(A - \lambda I_n) \neq \{0\}$.

So we would like to determine those λ for which the kernel of $A - \lambda I_n$ is nontrivial, for which the following is key:

$$\ker(A - \lambda I_n) \neq \{0\} \quad \Leftrightarrow \quad \operatorname{rank}(A - \lambda I_n) < n \quad \Leftrightarrow \quad \det(A - \lambda I_n) = 0.$$

Let's apply this to the matrix A in our example:

$$\begin{aligned}
\det\left(\begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) &= \det\left(\begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) \\
&= \det\begin{pmatrix} -1-\lambda & 2 \\ -6 & 6-\lambda \end{pmatrix} \\
&= (-1-\lambda)(6-\lambda) - 2(-6) \\
&= \lambda^2 - 5\lambda + 6 \\
&= (\lambda - 2)(\lambda - 3).
\end{aligned}$$

Thus, $\ker(A - \lambda I_n) \neq \{0\}$ if and only if $\lambda = 2, 3$. So the eigenvalues for A are 2 and 3.

Having found the eigenvalues, **how do we go about finding corresponding eigenvalues**? For each eigenvalue λ, there are nonzero elements $\ker(A - \lambda I_n)$. So we just apply our algorithm for finding the kernel of a matrix:

$\boxed{\lambda = 2}$

$$A - 2I_2 = \begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} -3 & 2 \\ -6 & 4 \end{pmatrix}.$$

So we need to find $(x, y) \in \mathbb{R}^2$ satisfying

$$\begin{pmatrix} -3 & 2 \\ -6 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Therefore, we perform Gaussian elimination:

$$\begin{pmatrix} -3 & 2 \\ -6 & 4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -\frac{2}{3} \\ 0 & 0 \end{pmatrix}.$$

Hence,

$$\ker(A - 2I_2) = \left\{\left(\frac{2}{3}y, y\right) : y \in \mathbb{R}\right\}.$$

For a basis we could take $\left(\frac{2}{3}, 1\right)$, or easier, $(2, 3)$.

Similarly for the other eigenvalue:

$\boxed{\lambda = 3}$

$$A - 3I_2 = \begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix} - \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

$$= \begin{pmatrix} -4 & 2 \\ -6 & 3 \end{pmatrix}$$

$$\rightsquigarrow \begin{pmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{pmatrix}.$$

Hence,

$$\ker(A - 3I_2) = \left\{ \left(\frac{1}{2}y, y \right) : y \in \mathbb{R} \right\}.$$

For a basis we could take $\left(\frac{1}{2}, 1\right)$, or easier, $(1, 2)$.

Let $A \in M_{n \times n}(F)$. The eigenvalues for A are exactly the solutions λ to

$$\det(A - \lambda I_n) = 0.$$

If $\lambda \in F$ is an eigenvalue, it corresponding eigenvectors are the nonzero vectors in

$$\ker(A - \lambda I_n).$$

Use our algorithm to find a basis for the matrix $A - \lambda I_n$.

Week 10, Monday: Diagonalization algorithm

Recall from last time: an *eigenvector* for a linear transformation $f\colon V \to V$ is a *nonzero* vector $v \in V$ such that

$$f(v) = \lambda v$$

for some $\lambda \in F$. In that case, λ is called an *eigenvalue* for f.

Definition. Let V be an n-dimensional vector space. A linear mapping $f\colon V \to V$ is *diagonalizable* if there exists an ordered basis α of V such that $[f]_\alpha^\alpha = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. A matrix $A \in M_{n\times n}(F)$ is *diagonalizable* if its corresponding linear mapping f_A is diagonalizable.

Proposition. A linear mapping $f\colon V \to V$ is diagonalizable if and only if V has a basis consisting solely of eigenvectors for f.

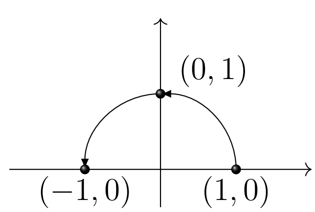
Proof. Let α be any ordered basis. Then $[f]_\alpha^\alpha$ is diagonal if and only if, for each $j = 1, \ldots, n$, the j-th column of $[f]_\alpha^\alpha$ has a single non-zero entry, in the j-th row. That j-th column is determined by

$$f(v_j) = 0 \cdot v_1 + \cdots + 0 \cdot v_{j-1} + \lambda_j \cdot v_j + 0 \cdot v_{j+1} + \cdots + 0 \cdot v_n,$$

for some scalar λ_j. However, the above condition is equivalent to $f(v_j) = \lambda_j v_j$ for $j = 1, \ldots, n$, i.e., to α being a basis of eigenvectors. $\square$

Example. Not all linear transformations of a vector space to itself are diagonalizable. For instance, consider the linear transformation $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ that is rotation of the plane by $90°$, having matrix

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



There is no point $0 \neq v \in \mathbb{R}^2$ such that $Av = \lambda v$ for some λ. (The matrix is diagonalizable over $\mathbb{C}$, though. Can you find two eigenvectors? Don't get your hopes up, though—there are matrices that are not diagonalizable over $\mathbb{C}$.)

Suppose $f\colon F^n \to F^n$ is a linear transformation, and let A be the matrix corresponding to f, i.e., the matrix whose j-th column is $f(e_j)$ for all j (i.e., the matrix for f with respect to the standard basis for F^n). Suppose we can find a basis $\alpha = \langle v_1, \ldots, v_n \rangle$ of eigenvectors for f with corresponding, not necessarily distinct, eigenvalues $\lambda_1, \ldots, \lambda_n$. Let P be the matrix with columns $v_1, \ldots, v_n$. Then, as we saw last time,

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Definition. Two $n \times n$ matrices A and B over F are *similar* or *conjugate* if there exists an invertible matrix P such that $A = P^{-1}BP$.

Exercise. The reader should verify that similarity is an equivalence relation.

Remark. Let $f\colon V \to V$ be a linear transformation of a finite-dimensional vector space, and let α and β be two ordered bases for V. Then we saw earlier in the semester that the matrices $A := [f]_\alpha^\alpha$ and $B := [f]_\beta^\beta$ are conjugate, i.e., the matrices for f with respect to any two bases for V are conjugate. The converse is also true: every matrix conjugate to A is the matrix representing f with respect to some basis.

Finding eigenvectors and eigenvalues. Let $A \in M_{n\times n}(F)$ with corresponding linear function

$$\begin{aligned} f_A\colon F^n &\to F^n \\ v &\mapsto Av. \end{aligned}$$

As mentioned last time, the following argument is of central importance in the story of eigenvectors and eigenvalues: We are looking for nonzero $v \in F^n$ and any $\lambda \in F$ such that $Av = \lambda v$. We have

$$Av = \lambda v \quad \Leftrightarrow \quad (A - \lambda I_n)v = 0 \quad \Leftrightarrow \quad v \in \ker(A - \lambda v).$$

This says that:

$\lambda \in F$ is an eigenvalue for A if and only if $\ker(A - \lambda I_n) \neq \{0\}$.

So we would like to determine those λ for which the kernel of $A - \lambda I_n$ is nontrivial. The following is key:

$$\ker(A - \lambda I_n) \neq \{0\} \quad \Leftrightarrow \quad \operatorname{rank}(A - \lambda I_n) < n \quad \Leftrightarrow \quad \det(A - \lambda I_n) = 0.$$

Definition. The *characteristic polynomial* of A is

$$p_A(x) := \det(A - xI_n).$$

We have just seen that

> $\lambda \in F$ is an eigenvalue for A if and only if it is a zero of the characteristic polynomial for A, i.e., if and only if $p_A(\lambda) = 0$.

Example. Let

$$A = \begin{pmatrix} 2 & -7 & 3 \\ 0 & -5 & 3 \\ 0 & 0 & 2 \end{pmatrix}.$$

The characteristic polynomial of A is

$$\begin{aligned}
p_A(t) &= \det\left(\begin{pmatrix} 2 & -7 & 3 \\ 0 & -5 & 3 \\ 0 & 0 & 2 \end{pmatrix} - x\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) \\
&= \det\left(\begin{pmatrix} 2 & -7 & 3 \\ 0 & -5 & 3 \\ 0 & 0 & 2 \end{pmatrix} - \begin{pmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{pmatrix}\right) \\
&= \det\begin{pmatrix} 2-x & -7 & 3 \\ 0 & -5-x & 3 \\ 0 & 0 & 2-x \end{pmatrix} \\
&= (2-x)(-5-x)(2-x) \\
&= -(x-2)^2(x+5).
\end{aligned}$$

Thus, $p_A(x) = 0$ if and only if $x \in \{2, -5\}$. So the eigenvalues of A are 2 (with *multiplicity* 2), and -5.

Recall that our goal is to diagonalize A by finding a basis of eigenvectors. That's not always possible, but we can try. The **first step** is to compute the zeros of the characteristic polynomial, $p_A(x)$. This tells us the eigenvalues for A. We then need to find the eigenvectors to go along with these eigenvalues. Recall that nonzero $v \in F^n$ is an eigenvector for A with eigenvalue λ if and only $v \in \ker(A - \lambda I_n)$.

Definition. Let λ be an eigenvalue of the $n \times n$ matrix A over F. Then the *eigenspace for* λ is

$$E_\lambda := E(A)_\lambda := \{v \in F^n : Av = \lambda v\} = \ker(A - \lambda I_n v).$$

The eigenspace, being the kernel of a matrix, is a linear subspace of F^n.

The **second step** in trying to diagonalize A is to compute a basis for each eigenspace E_λ.

Example. We have seen that the eigenvalues for

$$A = \begin{pmatrix} 2 & -7 & 3 \\ 0 & -5 & 3 \\ 0 & 0 & 2 \end{pmatrix}.$$

are 2 (with multiplicity 2) and -5. Let's compute the corresponding eigenspaces in the case $F = \mathbb{R}$.

$\boxed{E_2}$

$$\begin{aligned} A - 2I_3 &= \begin{pmatrix} 2 & -7 & 3 \\ 0 & -5 & 3 \\ 0 & 0 & 2 \end{pmatrix} - \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -7 & 3 \\ 0 & -7 & 3 \\ 0 & 0 & 0 \end{pmatrix} \\ &\longrightarrow \begin{pmatrix} 0 & 1 & -3/7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

The first and third variables are free. Hence,

$$\ker(A - 2I_3) = \left\{(x, \tfrac{3}{7}z, z) : x, z \in \mathbb{R}\right\}.$$

For a basis we could take $\left\{(1,0,0),(0,\frac{3}{7},1)\right\}$, or easier, $\{(1,0,0),(0,3,7)\}$.

$\boxed{E_{-5}}$

$$A-(-5)I_3=\begin{pmatrix}2&-7&3\\0&-5&3\\0&0&2\end{pmatrix}+\begin{pmatrix}5&0&0\\0&5&0\\0&0&5\end{pmatrix}$$

$$=\begin{pmatrix}7&-7&3\\0&0&3\\0&0&7\end{pmatrix}$$

$$\longrightarrow\begin{pmatrix}1&-1&0\\0&0&1\\0&0&0\end{pmatrix}.$$

Hence,

$$\ker(A+5I_3)=\{(y,y,0):y\in\mathbb{R}\}.$$

For a basis we could take $(1,1,0)$.

Thus, we have found three eigenvectors $(1,0,0)$, $(0,3,7)$, and $(1,1,0)$. It turns out that eigenvectors for distinct eigenvalues are always linearly independent (we'll see this later). Hence, we have found a basis of eigenvectors. Thus, A is diagonalizable, and if we use these eigenvectors as the columns for a matrix:

$$P=\begin{pmatrix}1&0&1\\0&3&1\\0&7&0\end{pmatrix},$$

then one may check that

$$P^{-1}AP=\operatorname{diag}(2,2,-5).$$

Example. Now consider a matrix that is just slightly different from A:

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

The characteristic polynomial for A and for B are the same:

$$\det(B-xI_3)=\det\begin{pmatrix}2-x&1&3\\0&-5-x&3\\0&0&2-x\end{pmatrix}=-(x-2)^2(x+5).$$

Thus, A and B have the same eigenvalues. Let's compute the eigenspaces for B over $\mathbb{R}$.

$\boxed{E_2}$

$$
\begin{aligned}
B - 2I_3 &= \begin{pmatrix} 2 & 1 & 3 \\ 0 & -5 & 3 \\ 0 & 0 & 2 \end{pmatrix} - \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 & 3 \\ 0 & -7 & 3 \\ 0 & 0 & 0 \end{pmatrix} \\
&\longrightarrow \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.
\end{aligned}
$$

Thus, $\ker(B-2I_3)$ has basis $\{(1,0,0)\}$. It is only one-dimensional. Recall that $\ker(A-2I_3)$ was two-dimensional. This is a crucial difference.

$\boxed{E_{-5}}$

$$
\begin{aligned}
A - (-5)I_3 &= \begin{pmatrix} 2 & 1 & 3 \\ 0 & -5 & 3 \\ 0 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix} \\
&= \begin{pmatrix} 7 & 1 & 3 \\ 0 & 0 & 3 \\ 0 & 0 & 7 \end{pmatrix} \\
&\longrightarrow \begin{pmatrix} 1 & 1/7 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.
\end{aligned}
$$

Hence,

$$\ker(A - 3I_3) = \{(y, -y/7, 0) : y \in \mathbb{R}\}.$$

For a basis we could take $(-7, 1, 0)$.

Our calculations prove that, at most, we can find two linearly independent vectors that are eigenvectors for B. Thus, there is no basis for $\mathbb{R}^3$ consisting of eigenvectors for B. Therefore, B is not diagonalizable.

Diagonalizing Algorithm Let $A \in M_{n\times n}(F)$.

(a) Find the eigenvalues of A as the zeros of its characteristic polynomial,

$$p_A(x) = \det(A - xI_n).$$

(b) For each eigenvalue λ, compute a basis for the eigenspace $E_\lambda = \ker A - \lambda I_n$.

(c) **The matrix A is diagonalizable if and only if of the total number of eigenvectors in the bases found in the previous step is n. i.e., if and only if the sum of the dimensions of the eigenspaces is n.** If so, the union of these vectors is a basis for F^n. Create a matrix P whose columns are these vectors. Then $P^{-1}AP = D$, where D is a diagonal matrix with the eigenvalues along the diagonal, and we get a corresponding commutative diagram:

$$\begin{tikzcd}
F^n \arrow[r, "A"] \arrow[d, "P^{-1}"', "\wr"] & F^n \arrow[d, "P^{-1}", "\wr"'] \\
F^n \arrow[r, "D"] & F^n.
\end{tikzcd}$$

The matrix P^{-1}, considered as a linear function, takes coordinates with respect to the basis of eigenvalues.

Remark. An $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. Step (c) of the diagonalization algorithm depends on a fact we will prove next time: eigenvectors with distinct eigenvalues are linearly independent. (We compute bases for each eigenspace, and of course the elements in a basis are linearly independent. But when we combine the bases for all of the eigenspaces, why is the resulting set independent?)

Week 10, Wednesday: Eigenspaces

Before getting started, we make an observation which should have probably already been mentioned:

Proposition. Let A, B be $n \times n$ matrices representing a linear function $f\colon V \to V$ with respect to different bases. Then their characteristic polynomials are the same: $p_A(x) = p_B(x)$.

Proof. We have $A = P^{-1}BP$ for some $n \times n$ matrix P. Then

$$\begin{aligned}
p_A(x) &= \det(A - xI_n) \\
&= \det(P^{-1}BP - xI_n) \\
&= \det(P^{-1}BP - xP^{-1}I_nP) \\
&= \det(P^{-1}BP - P^{-1}(xI_n)P) \qquad (x \text{ is a scalar}) \\
&= \det(P^{-1}(B - xI_n)P) \\
&= \det(P^{-1})\det(B - xI_n)\det(P) \\
&= \det(B - xI_n).
\end{aligned}$$

For the last step, recall that $\det(P^{-1}) = \det(P)^{-1}$, which follows from multiplicativity of the determinant:

$$1 = \det(I_n) = \det(P^{-1}P) = \det(P^{-1})\det(P).$$

$\square$

Thus, it makes sense to talk about the **characteristic polynomial of a linear transformation**: it the characteristic polynomial of any matrix representing the transformation.

Last time, we discussed the following algorithm that determines whether a matrix is diagonalizable and, if it is, shows how to diagonalize it.

Diagonalization Algorithm Let $A \in M_{n\times n}(F)$.

(a) Find the eigenvalues of A as the zeros of its characteristic polynomial, $p_A(x) = \det(A - xI_n)$.

(b) For each eigenvalue λ, compute a basis for the eigenspace $E_\lambda = \ker(A - \lambda I_n)$.

(c) **The matrix A is diagonalizable if and only if of the total number of eigenvectors in the bases found in the previous step is n.** In other words, A is diagonalizable if and only if $\sum_\lambda \dim E_\lambda = n$ where the sum is over all eigenvalues λ of A. If so, then the union of these vectors is a basis for F^n. Create a matrix P whose columns are these vectors. Then $P^{-1}AP = D$, where D is a diagonal matrix with the eigenvalues along the diagonal, and we get a corresponding commutative diagram:

$$\begin{tikzcd} F^n \arrow[r, "A"] \arrow[d, "P^{-1}"', "\wr"] & F^n \arrow[d, "P^{-1}", "\wr"'] \\ F^n \arrow[r, "D"] & F^n. \end{tikzcd}$$

The matrix P^{-1}, considered as a linear function, takes coordinates with respect to the basis of eigenvalues.

As mentioned last time, Step (c) of the diagonalization algorithm depends on the fact that eigenvectors with distinct eigenvalues are linearly independent. (Thus, when we combine the bases for all of the eigenspaces, we end up with a linearly independent set.) We now prove this.

Proposition. Let V be any vector space, and let $f \colon V \to V$ be a linear transformation. Let $v_1, \dots, v_k \in V$ be eigenvectors for f with corresponding eigenvalues λ_i:

$$f(v_i) = \lambda_i v_i$$

for $i = 1, \dots, k$. Suppose $\lambda_1, \dots, \lambda_k$ are *distinct*. Then $v_1, \dots, v_k$ are linearly independent.

Proof. We will prove this by induction on k. The case $k = 1$ is OK since, by definition, an eigenvector is a nonzero vector. Suppose $v_1, \dots, v_{k-1}$ are linearly independent for some $k > 1$ and that

$$a_1 v_1 + \cdots + a_k v_k = 0$$

for some $a_i \in F$. Let id_V be the identity transformation defined by $\mathrm{id}_V(v) = v$ for all $v \in V$. Apply the linear transformation $f - \lambda_k \mathrm{id}_V$ to the above dependence relation to get

$$
\begin{aligned}
& (f - \lambda_k \mathrm{id}_V)(a_1 v_1 + \cdots + a_k v_k) = (f - \lambda_k \mathrm{id}_V)(0) = 0 \\
\Rightarrow \quad & f(a_1 v_1 + \cdots + a_k v_k) - \lambda_k \mathrm{id}_V(a_1 v_1 + \cdots + a_k v_k) = 0 \\
\Rightarrow \quad & (a_1 \lambda_1 v_1 + \cdots + a_k \lambda_k v_k) - (a_1 \lambda_k v_1 + \cdots + a_k \lambda_k v_k) = 0 \\
\Rightarrow \quad & a_1(\lambda_1 - \lambda_k) v_1 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k) v_{k-1} + a_k(\lambda_k - \lambda_k) v_k = 0 \\
\Rightarrow \quad & a_1(\lambda_1 - \lambda_k) v_1 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k) v_{k-1} = 0
\end{aligned}
$$

Since $v_1, \ldots, v_{k-1}$ are linearly independent, all the coefficients are zero:

$$a_1(\lambda_1 - \lambda_k) = \cdots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

Since the λ_i are distinct, this implies $a_1 = \cdots = a_{k-1} = 0$. Therefore, the original equation, $a_1 v_1 + \cdots + a_k v_k = 0$ becomes $a_k v_k = 0$. Since v_k is an eigenvector, it is nonzero. Hence, $a_k = 0$, as well. □

Corollary. Suppose $\dim V = n$ and $f \colon V \to V$ is a linear transformation. Then if f has n distinct eigenvalues, it is diagonalizable.

Proof. Each eigenvalue has at least one corresponding eigenvector. From the above proposition, if f has n distinct eigenvalues, then it has n linearly independent eigenvectors. Since V has dimension n, these eigenvectors form a basis for V. Let α be an ordered basis consisting of those eigenvectors. Then $[f]_\alpha^\alpha$ is diagonal. □

Remark. The Proposition implies that the union of bases for the eigenspaces of A forms a linearly independent sets. For instance, for convenience, suppose that A has three (distinct) eigenvalues λ_1, λ_2, and λ_3, and suppose the corresponding eigenspaces have bases $\{u_1, \ldots, u_p\}$, $\{v_1, \ldots, v_q\}$, and $\{w_1, \ldots, w_r\}$, respectively. We would like to show that the union of these sets is linearly dependent. So suppose we have a relation

$$a_1 u_1 + \cdots + a_p u_p + b_1 v_1 + \cdots + b_q v_q + c_1 w_1 + \cdots + c_r w_r = 0.$$

Let $u = \sum_{i=1}^{p} a_i u_i$, $v = \sum_{i=1}^{q} b_i v_i$, and $w = \sum_{i=1}^{r} c_i w_i$. Then we have $u \in E_{\lambda_1}$, $v \in E_{\lambda_2}$, and $w \in E_{\lambda_3}$ and

$$u + v + w = 0.$$

By the Proposition, must have $u = v = w = 0$. Otherwise, this relation would be a nontrivial linear relation among eigenvectors with distinct eigenvalues. (Note that the only element of an eigenspace that is not an eigenvector is the zero vector.)

Warning. The converse to the corollary is not true. For instance, consider the identity function on F^n. Its matrix is I_n, which is already diagonal, and 1 is its only eigenvalue:

$$p_{I_n}(x) = \det(I_n - xI_n) = \det\left((1-x)I_n\right) = (1-x)^n \det(I_n) = (1-x)^n.$$

So I_n is diagonalizable (in fact, it's already diagonal) even though its eigenvalues are not distinct.

Cramer's rule

Definition. Let $A \in M_{n\times n}(F)$. For $i, j \in \{1, \dots, n\}$, let $A^{ij} \in M_{(n-1)\times(n-1)}(F)$ be the matrix formed by removing the i-th row and j-th column of A. The i, j*-th minor* of A is $\det(A^{ij})$, and the i, j*-th cofactor* of A is $(-1)^{i+j}\det(A^{ij})$. The *adjugate* of A is the matrix $\operatorname{adj}(A) \in M_{n\times n}(F)$ with i, j-th coordinate

$$\operatorname{adj}(A)_{ij} = (-1)^{i+j}\det(A^{ji}).$$

(Note we are using A^{ji}, not A^{ij}.)

Theorem (Cramer's rule). Let $A \in M_{n\times n}(F)$ be an invertible matrix, and let $b \in F^n$. Then the solution to the system of linear equations $Ax = b$ is given by

$$x_j = \frac{\det(M_j)}{\det(A)}$$

for $j = 1, \dots, n$ where $M_j \in M_{n\times n}(F)$ is the matrix formed by replacing the j-th column of A with b.

Corollary. If $A \in M_{n\times n}(F)$ is invertible, then

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A)$$

where $\operatorname{adj}(A)$ is the *adjugate* of A, defined by

Corollary. If $A \in M_{n\times n}(F)$ is invertible and $F = \mathbb{R}$ of $F = \mathbb{C}$, then

(a) the solution for the system of equations $Ax = b$ is a continuous function of the entries of A and b, and

(b) the entries of A^{-1} are continuous functions of the entries of A.

Proof. The entries in the determinant of a matrix B are polynomials in the entries of B. A quotient f/g of polynomials f and g is a continuous function wherever g is nonzero. $\square$

Example. Consider the matrix

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

The adjugate of A is

$$\operatorname{adj}(A) = \begin{pmatrix} 2 & 2 & -8 \\ 18 & -6 & -48 \\ -2 & -2 & -4 \end{pmatrix}.$$

For instance, to find the $1,2$-entry of $\operatorname{adj}(A)$ is

$$(-1)^{1+2}\det(A^{2,1}) = (-1)^3 \det\begin{pmatrix} -1 & 6 \\ 0 & 2 \end{pmatrix} = 2.$$

Using Cramer's rule to compute the inverse of A, we get

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A) = -\frac{1}{24}\begin{pmatrix} 2 & 2 & -8 \\ 18 & -6 & -48 \\ -2 & -2 & -4 \end{pmatrix} = \begin{pmatrix} -\frac{1}{12} & -\frac{1}{12} & \frac{1}{3} \\ -\frac{3}{4} & \frac{1}{4} & 2 \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{6} \end{pmatrix}.$$

Week 10, Friday: Algebraic and geometric multiplicity. Jordan form.

When does a transformation fail to be diagonalizable? We now introduce a sequence of ideas that will allow us to answer this question.

Example. Earlier, we considered the linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix

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

Geometrically, it rotates the plane counterclockwise by $90°$ and, hence, has no eigenvectors: an eigenvector would not rotate—it would just be scaled. The characteristic polynomial of A is

$$p_A(x) = \det \begin{pmatrix} -x & -1 \\ 1 & -x \end{pmatrix} = x^2 + 1.$$

The equation $x^2 + 1 = 0$ has no solutions over $\mathbb{R}$, and hence, the transformation has no eigenvalues.

Now consider the linear transformation $f \colon \mathbb{C}^2 \to \mathbb{C}^2$ given by the same matrix A. Over $\mathbb{C}$ we can solve $x^2 + 1 = 0$ to find two eigenvalues, $\pm i$. Each of these will have at least one eigenvector, and eigenvectors for distinct eigenvalues are linearly independent. Since $\mathbb{C}^2$ has dimension 2, that means we will get a basis of eigenvectors. Let's compute a basis for the eigenspace for i:

$$A - iI_2 = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \xrightarrow{r_1 \leftrightarrow r_2} \begin{pmatrix} 1 & -i \\ -i & -1 \end{pmatrix} \xrightarrow{r_2 \leftrightarrow r_2 + ir_1} \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix}.$$

So the kernel of $A - iI_2$ is $\{(iy, y) : y \in \mathbb{C}\}$, which has basis $\{(i, 1)\}$. Similarly, the eigenspace for $-i$ has basis $\{(-i, 1)\}$. Check:

$$A \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ i \end{pmatrix} = i \begin{pmatrix} i \\ 1 \end{pmatrix}$$

$$A \begin{pmatrix} -i \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -i \end{pmatrix} = -i \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Letting

$$P = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix},$$

we get

$$P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

This example illustrates one obstacle to diagonalization: the characteristic polynomial may not have enough roots in the field F.

Definition. A polynomial $p \in F[x]$ *splits over* F if there exist $c, \lambda_1, \ldots, \lambda_n \in F$ such that

$$p(x) = c(x_1 - \lambda_1) \cdots (x - \lambda_n).$$

Equivalently, $p(x)$ had n roots (zeros), $\lambda_1, \ldots, \lambda_n$, in F. These λ_i need not be distinct.

Remark. Let F be any field, and let $p(x)$ be a polynomial whose coefficients are in F, i.e., $p(x) \in F[x]$. It turns out that there exists a field $F \subseteq K$ such that $p(x)$ splits over K.

Example. The polynomial $p(x) = x^2 + 1$ splits over $\mathbb{C}$ but not over $\mathbb{R}$.

A useful fact from algebra:

Theorem. (Fundamental theorem of algebra) Every $p \in \mathbb{C}[x]$ splits over $\mathbb{C}$.

Proposition. Let V be a vector space over F with $\dim V = n$, and let $f : V \to V$ be a linear transformation. If f is diagonalizable, then its characteristic polynomial splits over F.

Proof. Let $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ be a diagonal matrix representing f. Then the characteristic polynomial for f (which, as we saw earlier, in the last lecture, does not depend on the choice of matrix representative) is

$$p_f(x) = p_D(x) = \det(D - xI_n) = (\lambda_1 - x) \cdots (\lambda_n - x) = (-1)^n (x_1 - \lambda_1) \cdots (x - \lambda_n).$$

$\square$

The converse of this proposition is not true:

Example. Let

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

The characteristic polynomial of A is

$$\begin{aligned} p_A(x) &= \det(A - xI_2) \\ &= \det\begin{pmatrix} 1-x & 1 \\ 0 & 1-x \end{pmatrix} \\ &= (x-1)^2. \end{aligned}$$

Thus, the characteristic polynomial splits over any field F. There is one eigenvalue, 1, which occurs with algebraic multiplicity 2 (the precise definition of *algebraic multiplicity* appears below). Let's proceed with the algorithm for diagonalization by computing a basis for the eigenspace for 1, i.e., for $\ker(A - I_2)$:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Therefore, $\ker(A - I_2) = \{(x, 0) : x \in F\}$. A basis is $\{(1, 0)\}$. Thus, there is no basis for F^2 consisting of eigenvectors: our theory says any eigenvector would have to have eigenvalue 1, and the space of eigenvectors with eigenvalue 1 is only one-dimensional!

Definition. Let $\dim V < \infty$. The *algebraic multiplicity* of an eigenvalue $\lambda \in F$ for a linear transformation $f\colon V \to V$ (or for any matrix representing f) is the largest number m such that $p_f(x) = (x - \lambda)^m q(x)$ for some polynomial $q(x) \in F[x]$.

The *geometric multiplicity* of λ is the dimension of the eigenspace $E_\lambda(f)$ for λ:

$$\dim E_\lambda(f) = \dim \ker(f - \lambda\,\mathrm{id}_V).$$

So if A is a matrix representing f, then the geometric multiplicity of $\lambda \in F$ is

$$\dim E_\lambda(A) = \dim \ker(A - \lambda\, I_n).$$

Remark. To rephrase something we already know: $A \in M_{n\times n}(F)$ **is diagonalizable if and only if the sum of its geometric multiplicities is** n**.** That's because this is the only case in which we have enough linearly independent eigenvectors to form a basis of eigenvectors.

Proposition. Let $\dim V < \infty$, and let λ be an eigenvalue of a linear transformation $f\colon V \to V$. Then the geometric multiplicity of λ is at most the algebraic multiplicity of λ.

Proof. Let $v_1, \ldots, v_k$ be a basis for $\ker(f - \lambda \operatorname{id}_V)$, and extend it to a basis $v_1, \ldots, v_n$ for all of V. We have $f(v_i) = \lambda v_i$ for $1 = 1, \ldots, k$. So with respect to our chosen basis, the matrix representing f has the form

$$A := \begin{pmatrix} \lambda I_k & B \\ 0 & C \end{pmatrix},$$

where B and C are $(n-k) \times (n-k)$ matrices. So the characteristic polynomial for f is

$$\begin{aligned} p_f(x) &= \det \begin{pmatrix} (\lambda - x) I_k & B \\ 0 & C - x I_{n-k} \end{pmatrix} \\ &= \det((\lambda - x) I_k) \det(C - x I_{n-k}) \\ &= (\lambda - x)^k \det(C - x I_{n-k}) \\ &= (\lambda - x)^k q(x), \end{aligned}$$

for some polynomial $q(x)$. (To see the second equality, above, expand the determinant in line 1 along the first *column*—there will only be one term, which will be $\lambda - x$ times a smaller determinant. Expand that determinant along its first column. Repeat k times, each time picking up a factor of $\lambda - x$.) This shows that the algebraic multiplicity of λ is at least k, the geometric multiplicity of λ. □

Corollary. Let $A \in M_{n \times n}(F)$. Then A is diagonalizable if and only if its characteristic polynomial splits over F and the geometric multiplicity and algebraic multiplicity of each eigenvalue are equal.

Proof. Suppose that A is diagonalizable. We saw earlier in this lecture that the characteristic polynomial for A then splits over F. So we can write

$$p_A(x) = (-1)^n \prod_{i-1}^{k} (x - \lambda_i)^{m_i}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of A. The degree of $p_A(x)$ is n (exercise!), from which it follows that $\sum_{i=1}^{k} m_i = n$, i.e, the sum of the algebraic multiplicities is n.

Let g_i be the geometric multiplicity of eigenvalue λ_i. Since A is diagonalizable, we know that the sum of its geometric multiplicities is also n.

Therefore, we have $n = \sum_{i=1}^k g_i = \sum_{i=1}^k m_i = n$, and by the Proposition, $g_i \leq m_i$ for all i. Since the m_i are nonnegative, it follows that $m_i = g_i$ for all i.

Conversely, suppose that $p_A(x)$ splits and that the algebraic and geometric multiplicities of each eigenvalue are equal. Factor $p_A(x)$ as above and use the same notation for algebraic and geometric multiplicities. As before, since the degree of $p_A(x) = n$, we have $n = \sum_{i=1}^k m_i$. By assumption, $m_i = g_i$ for all i. So it follows that the sum of the geometric multiplicities is n, and hence, A is diagonalizable. □

Jordan form. What can we say when a linear transformation is not diagonalizable? Can we still choose a basis to make the matrix for the transformation simple in some sense? We give one answer here. First, we need a couple definitions. A *Jordan block of size k for $\lambda \in F$* is the $k \times k$ matrix with λs on the diagonal and 1s on the "superdiagonal":

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & 1 & \cdots & 0 & 0 \\ & & \vdots & & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

For example,

$$J_4(3) = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Note the following example of the special case of a Jordan block of size 1:

$$J_1(5) = [5].$$

A matrix is in *Jordan form* if it is in block diagonal form with Jordan blocks for various λ along the diagonal:

$$\begin{pmatrix} J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{k_2}(\lambda_2) & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & J_{k_m}(\lambda_m) \end{pmatrix}$$

For example, here is a matrix in Jordan form:

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

It has two 2×2 Jordan blocks for 2, a 1×1 Jordan block for 5, and a 3×3 Jordan block for 4:

$$\begin{pmatrix} J_2(2) & 0 & 0 & 0 \\ 0 & J_2(2) & 0 & 0 \\ 0 & 0 & J_1(5) & 0 \\ 0 & 0 & 0 & J_3(4) \end{pmatrix}.$$

Theorem. Let $\dim V < \infty$. Suppose $f\colon V \to V$ is a linear transformation over F and that the characteristic polynomial for f splits, i.e., the field F contains all of the zeros of the characteristic polynomial. Then there exists an ordered basis for V such that the matrix representing f with respect to that basis is in Jordan form. The Jordan form is unique up to a permutation of the Jordan blocks.

So a matrix is diagonalizable if and only if its characteristic polynomial splits and all of its Jordan blocks have size 1. We also know that a matrix such as

$$\begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

which is already in Jordan form but not diagonal, is not diagonalizable.

Week 11, Monday: Walks on graphs

We have devoted a lot of energy to the problem of diagonalizing a matrix. One major motivation for diagonalization is that it makes taking powers of a matrix easier. Explicitly, suppose that $A \in M_{n\times n}(F)$ is diagonalizable. So there exists a matrix P such that

$$P^{-1}AP = D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

It is easy to take powers of a diagonal matrix: $D^\ell = \operatorname{diag}(\lambda_1^\ell, \ldots, \lambda_n^\ell)$. Here is the important trick:

$$\begin{aligned}
D^\ell &= (P^{-1}AP)^\ell \\
&= (P^{-1}AP)(P^{-1}AP)(P^{-1}AP)\cdots(P^{-1}AP)(P^{-1}AP) \\
&= P^{-1}A(PP^{-1})A(PP^{-1})A(PP^{-1})\cdots(PP^{-1})AP \\
&= P^{-1}A^\ell P.
\end{aligned}$$

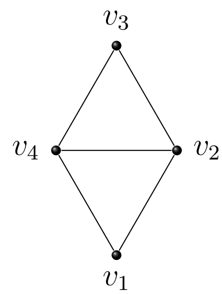
Therefore,

$$A^\ell = PD^\ell P^{-1} = P\operatorname{diag}(\lambda_1^\ell, \ldots, \lambda_n^\ell)P^{-1}.$$

In general, there will be many fewer arithmetic operations required on the right-hand side of this equation than on the left-hand side.

This lecture will consider one application of this idea.

Walks in graph. A graph (or network) consists of vertices connected by edges. Here is an example with 4 vertices connected by 5 edges:



The diamond graph.

A *walk of length* ℓ in a graph is a sequence of vertices $u_0 u_1 \dots u_\ell$ where u_{i-1} is connected to u_i for $i = 1, \dots, \ell$. So the length is the number of edges traversed. In our example, the following are walks from v_1 to v_4:

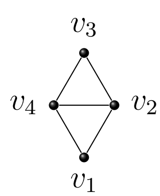
$$v_1 v_4 \quad \text{and} \quad v_1 v_2 v_3 v_4.$$

The first has length 1 and the second has length 3. We are interested in counting the number of closed walks between vertices.

Definition. Let G be a graph with vertices $v_1, \dots, v_n$. The *adjacency matrix* of G is the $n \times n$ matrix $A = A(G)$ defined by

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge connecting } v_i \text{ and } v_j \\ 0 & \text{otherwise.} \end{cases}$$

For example, the adjacency matrix of the diamond graph is



$$A = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \!\!\begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}.$$

Theorem. Let A be the adjacency matrix for a graph G with vertices $v_1, \dots, v_n$, and let $\ell \in \mathbb{Z} \geq 0$. Then then number of walks of length ℓ from v_i to v_j is $(A^\ell)_{ij}$.

Proof. Homework. $\square$

Example. Consider the diamond graph and its adjacency matrix A, displayed above. Then

$$A^0 = I_4, \quad A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 2 & 1 & 2 & 1 \\ 1 & 3 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ 1 & 2 & 1 & 3 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 2 & 5 & 2 & 5 \\ 5 & 4 & 5 & 5 \\ 2 & 5 & 2 & 5 \\ 5 & 5 & 5 & 4 \end{pmatrix}.$$

The highlighted entries in the matrix say there is 1 path of length 2 from v_2 to v_3 and there are 4 paths of length 3 from v_2 to itself. Can you find them? (The answer appears at the end of this lecture.)

So to count the number of walks, we need to compute powers of the adjacency matrix. Here is some good news:

Theorem. If A is an $n \times n$ symmetric matrix ($A = A^t$) over the real numbers, then it is diagonalizable over $\mathbb{R}$.

Proof. We may prove this later in the semester. (To look it up online, search for the "spectral theorem", which is usually stated for the more general class of Hermitian matrices. Over the real numbers, the Hermitian matrices are exactly the symmetric matrices.) □

This means that we can find a matrix P such that $P^{-1}AP = D$, where D is the diagonal matrix of the eigenvalues. Then $A^\ell = PD^\ell P^{-1}$. So we can find a nice closed form for the number of walks of length ℓ between any two vertices as a linear expression in the ℓ-th powers of the eigenvalues of A. If the eigenvalues are $\lambda_1, \dots, \lambda_n$, the equation $A^\ell = PD^\ell P^{-1}$ immediately implies that for each pair of vertices v_i and v_j there exist real numbers $c_1, \dots, c_n$, independent of ℓ, such that the number of closed walks of length ℓ from v_i to v_j is

$$c_1\lambda_1^\ell + \cdots + c_n\lambda_n^\ell.$$

The special case of closed walks is particularly nice.

Definition. A walk is *closed* if it begins and ends at the same vertex.

Definition. Let A be any $n \times n$ matrix. Then the *trace* of A is the sum of its diagonal entries:

$$\operatorname{tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

Proposition. Let A be the adjacency matrix of a graph G. Then the number of closed walks in G of length ℓ is $\operatorname{tr}(A^\ell)$.

Proof. For each $i = 1, \ldots, n$, the number of closed walks from v_i to v_i is $(A^\ell)_{ii}$. Summing over i gives the total number of closed walks. $\square$

Proposition. Let A be any $n \times n$ matrix. Then the trace of A is the sum of its eigenvalues, each counted according to its (algebraic) multiplicity.

Proof. Homework. $\square$

Corollary. Let A be the adjacency matrix of a graph G with n vertices, and let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be its list of (not necessarily distinct) eigenvalues. Then the number of closed walks in G of length ℓ is $\sum_{i=1}^n \lambda_i^\ell$.

Proof. The number of closed walks of length ℓ is $\operatorname{tr}(A^\ell)$, which is the sum of the eigenvalues of A^ℓ. By homework (an easy induction argument), if λ is an eigenvalue of A, then λ^ℓ is an eigenvalue of A^ℓ with unchanged eigenspace. It follows that the eigenvalues for A^ℓ are $\lambda_1^\ell, \ldots, \lambda_n^\ell$. $\square$

Example. Let A be the adjacency matrix of the diamond graph G. The characteristic polynomial of A is

$$\det(A - xI_4) = x^4 - 5x^2 - 4x = x(x+1)(x^2 - x - 4).$$

Using the quadratic equation, we find the eigenvalues for A:

$$0, -1, \frac{1+\sqrt{17}}{2}, \frac{1-\sqrt{17}}{2}.$$

Therefore, the number of closed walks in G of length ℓ is

$$w(\ell) = (0)^\ell + (-1)^\ell + \left(\frac{1+\sqrt{17}}{2}\right)^\ell + \left(\frac{1-\sqrt{17}}{2}\right)^\ell,$$

where

$$(0)^\ell = \begin{cases} 1 & \text{if } \ell = 0 \\ 0 & \text{if } \ell > 0. \end{cases}$$

The following table gives the number of closed walks for $\ell = 0, 1, \ldots, 6$:

ℓ	0	1	2	3	4	5	6
$w(\ell)$	4	0	10	12	50	100	298

Exercise. The complete graph, K_n, has vertices $1, \ldots, n$ and an edge between every pair of vertices. How many closed walks are there in K_n of lenght ℓ?

Questions.

(a) How would you generalize today's ideas to the case of a directed graph (in which the edges have directions)?

(b) How would you generalize today's ideas to the case in which the edges have weights? (A special case would be to let the weight of edge (u, v) be the probability that the edge is traversed given that the starting point is u. Another possibility is to think of the weight as a cost for traveling across the edge.)

Answer to example on page 2: $v_2v_4v_3$ has length 2 and the following have length 3: $v_2v_3v_4v_2$, $v_2v_4v_3v_2$, $v_2v_1v_4v_2$, and $v_2v_4v_1v_2$.

Week 11, Wednesday: Inner product spaces

We now add structure to a vector space allowing us to define length and angles.

Definition. Let V be a vector space over a field F where F is either $\mathbb{R}$ or $\mathbb{C}$. An *inner product* on V is a function

$$\begin{aligned} \langle\ ,\ \rangle \colon V \times V &\to F \\ (x, y) &\mapsto \langle x, y\rangle \end{aligned}$$

satisfying for all $x, y, z \in V$ and $c \in F$:

(a) linearity: $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ and $\langle cx, y\rangle = c\langle x, y\rangle$.

(b) conjugate symmetry: $\overline{\langle x, y\rangle} = \langle y, x\rangle$.

(c) positive-definiteness: $\langle x, x\rangle \in \mathbb{R}_{\geq 0}$, and $\langle x, x\rangle = 0$ iff $x = 0$.

Note. If $F = \mathbb{R}$, then an inner product is known as a *non-degenerate symmetric form.* If $F = \mathbb{C}$, an inner product is known as a *non-degenerate Hermitian form.*

Examples.

- The ordinary dot product on $\mathbb{R}^n$: Here, $V = \mathbb{R}^n$ and

  $$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = x \cdot y := \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots + x_n y_n.$$

  For example, in $\mathbb{R}^3$, we would have

  $\langle (1, 2, 3), (2, 3, 4)\rangle = 2+6+12 = 20$ and $\langle (1, 2, 3), (-2, 1, 0)\rangle = -2+2+0 = 0.$

- The ordinary inner product on $\mathbb{C}^n$: Here, $V = \mathbb{C}^n$ and

  $$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = x \cdot \bar{y} := \sum_{i=1}^{n} x_i \bar{y_i} = x_1 \bar{y_1} + \cdots + x_n \bar{y_n}.$$

For example, in $\mathbb{C}^2$, we would have

$$\begin{aligned}\langle (1+i, 1-i), (1+2i, 4)\rangle &= (1+i)\overline{(1+2i)} + (1-i)\,\overline{4}\\ &= (1+i)(1-2i) + (1-i)\,4\\ &= (3-i) + (4-4i) = 7-5i.\end{aligned}$$

- Let $V = \mathcal{C}_{\mathbb{R}}([0,1]) = \{f\colon [0,1]\to\mathbb{R} : f \text{ is continuous}\}$, the vector space of $\mathbb{R}$-valued continuous functions on the interval $[0,1]$, and

$$\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt.$$

To check positive-definiteness, note that if $f \neq 0$, then $f^2(t) > 0$ for t in some open interval in $[0,1]$. Hence,

$$\langle f, f\rangle = \int_0^1 f^2(t)\,dt > 0.$$

- $V = \mathbb{R}^2$, and

$$\langle (x_1, x_2), (y_1, y_2)\rangle = 3x_1y_1 + 2x_1y_2 + 2x_2y_1 + 4x_2y_2.$$

For positive-definiteness, we have

$$\langle (x_1, x_2), (x_1, x_2)\rangle = 3x_1^2 + 4x_1x_2 + 4x_2^2.$$

Complete the square:

$$\begin{aligned}3x_1^2 + 4x_1x_2 + 4x_2^2 &= 3\left(x_1^2 + \frac{4}{3}x_1x_2 + \frac{4}{3}x_2^2\right)\\ &= 3\left(\left(x_1 + \frac{2}{3}x_2\right)^2 - \frac{4}{9}x_2^2 + \frac{4}{3}x_2^2\right)\\ &= 3\left(\left(x_1 + \frac{2}{3}x_2\right)^2 + \frac{8}{9}x_2^2\right)\\ &\geq 0,\end{aligned}$$

with equality if and only if $x_1 = x_2 = 0$.

- Let $F = \mathbb{R}$ or $\mathbb{C}$, and let $V = M_{m\times n}(F)$. For $A \in M_{m\times n}(F)$, define the *conjugate transpose* of A by

$$A^* = \overline{A^t},$$

where the overline means taking the conjugate of each entry of A. If A has only real entries, the $A^* = A^t$. Next, define the inner product,

$$\langle A, B\rangle = \operatorname{tr}(B^*A) = \sum_{i=1}^{n}(B^*A)_{ii}.$$

(Note: The special case $m = 1$ gives the usual inner product on $\mathbb{R}^n$ or $\mathbb{C}^n$.) Proof of positive-definiteness is left as an exercise.

Proposition. Let $(V, \langle\, ,\, \rangle)$ be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$. Then for all $x, y, z \in V$ and $c \in F$,

(a) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.

(b) $\langle x, cy\rangle = \overline{c}\langle x, y\rangle$.

(c) $\langle x, 0\rangle = \langle 0, y\rangle = 0$.

(d) If $\langle x, y\rangle = \langle x, z\rangle$ for all $x \in V$, then $y = z$.

Proof. For part (a), notice that the definition of an inner product only guarantees sums on the left distribute. However, using properties of conjugation,

$$\begin{aligned}\langle x, y + z\rangle &= \overline{\langle y + z, x\rangle}\\ &= \overline{\langle y, x\rangle} + \overline{\langle z, x\rangle}\\ &= \langle x, y\rangle + \langle x, z\rangle.\end{aligned}$$

Parts (b) and (c) are left as exercises. For part (d), $\langle x, y\rangle = \langle x, z\rangle$ for all x implies

$$\begin{aligned}0 = \langle x, y\rangle - \langle x, z\rangle &= \langle x, y\rangle + (-1)\langle x, z\rangle\\ &= \langle x, y\rangle + \overline{(-1)}\langle x, z\rangle\\ &= \langle x, y\rangle + \langle x, (-1)z\rangle\\ &= \langle x, y\rangle + \langle x, -z\rangle\\ &= \langle x, y - z\rangle\end{aligned}$$

for all x. In particular, let $x = y - z$ to get $\langle y - z, y - z\rangle = 0$. By positive-definiteness, $y - z = 0$. $\square$

Week 11, Friday: Lengths, distances, components, angles

Definition. Let $(V, \langle\,,\rangle)$ be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$. The *norm* or *length* of $x \in V$ is

$$\|x\| = \sqrt{\langle x, x\rangle} \in \mathbb{R}_{\geq 0}.$$

Two vectors $x, y \in V$ are *orthogonal* or *perpendicular* if $\langle x, y\rangle = 0$. A *unit vector* is a vector of norm 1: so $x \in V$ is a unit vector if $\|x\| = 1$, which is equivalent to $\langle x, x\rangle = 1$.

Examples of norms.

- $V = \mathbb{R}^n$, $\langle x, y\rangle = x \cdot y$, the usual dot product. Then for $x \in \mathbb{R}^n$,

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

- $V = \mathbb{C}^n$, $\langle x, y\rangle = x \cdot \overline{y}$, the usual dot product on $\mathbb{C}^n$. Then for $z \in \mathbb{C}^n$,

$$\begin{aligned}\|z\| &= \sqrt{z_1\overline{z_1} + \cdots + z_n\overline{z_n}} \\ &= \sqrt{|z_1|^2 + \cdots + |z_n|^2}.\end{aligned}$$

  If $z_j \in \mathbb{C}$ is written as $z_j = x_j + iy_j$ with $x_j, y_j \in \mathbb{R}$, then $|z_j|^2 = x_j^2 + y_j^2$. So then

$$\|z\| = \sqrt{x_1^2 + y_1^2 + \cdots + x_n^2 + y_n^2}.$$

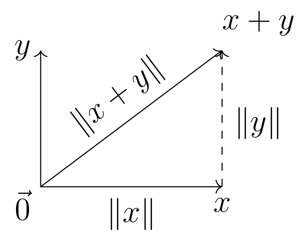
  Thus, if we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ via the isomorphism

$$(x_1 + iy_1, \ldots, x_n + iy_n) \to (x_1, y_1, \ldots, x_n, y_n),$$

  then the isomorphism preserves norms.

Proposition. (Pythagorean theorem) Let $(V, \langle\,,\rangle)$ be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$, and let $x, y \in V$ be perpendicular. Then

$$\|x\|^2 + \|y\|^2 = \|x + y\|^2.$$

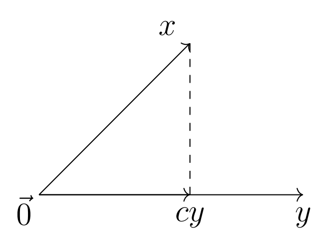


Proof. Since x and y are perpendicular, we have $\langle x, y\rangle = 0$. It follows that $\langle y, x\rangle = \overline{\langle x, y\rangle} = 0$, too. Therefore,

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle \\ &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\ &= \langle x, x\rangle + \langle y, y\rangle \\ &= \|x\|^2 + \|y\|^2.\end{aligned}$$

□

Suppose we are given two vectors x, y in an inner product space. A useful geometric operation is to break x into two parts, one of which lies along the vector y. Given any number c, the vector cy lies along y and we can evidently write x as the sum of two vectors: $x = (x - cy) + cy)$. In addition, though, we would like to require, by adjusting c, that the vector $x - cy$ is perpendicular to y. The picture in $\mathbb{R}^2$ would be:



We can calculate the required scalar c:

$$\langle x - cy, y\rangle = 0 \Longleftrightarrow \langle x, y\rangle - c\langle y, y\rangle = 0 \Longleftrightarrow c = \frac{\langle x, y\rangle}{\langle y, y\rangle} \Longleftrightarrow c = \frac{\langle x, y\rangle}{\|y\|^2},$$

which makes sense as long as $y \neq 0$.

Definition. Let $(V, \langle\ ,\ \rangle)$ be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$, and let $x, y \in V$ with $y \neq 0$. The *component* of x along y is the scalar

$$c = \frac{\langle x, y\rangle}{\langle y, y\rangle} = \frac{\langle x, y\rangle}{\|y\|^2}.$$

The *orthogonal projection* of x to y is the vector cy.

Example. Let $x \in V = \mathbb{R}^n$ or $\mathbb{C}^n$, and let e_j be the j-th standard basis vector. Then

$$\frac{\langle x, e_j \rangle}{\langle e_j, e_j \rangle} = \frac{x_j}{1} = x_j.$$

Thus, x_j is the component of x along e_j, and $x_j e_j$ is the projection of x to e_j.

Example. Let $x = (3,2)$ and $y = (5,0) = 5e_1$ in $\mathbb{R}^2$ with the usual inner product. Then the component of x along y is

$$\frac{\langle x, y \rangle}{\langle y, y \rangle} = \frac{(3,2) \cdot (5,0)}{(5,0),(5,0)} = \frac{15}{25} = \frac{3}{5}.$$

So the projection of x to y is

$$cy = \frac{3}{5}(5,0) = (3,0),$$

as expected.

Proposition. Let $(V, \langle\ ,\ \rangle)$ be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$. Let $x, y \in V$ and $c \in F$. Then

(a) $\|cx\| = |c|\|x\|$.

(b) $\|x\| = 0$ if and only if $x = 0$.

(c) Cauchy-Schwarz inequality: $|\langle x, y \rangle| \leq \|x\|\|y\|$.

(d) Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$.

Proof. Parts (a) and (b) are left as exercises. Part (c) is tricky. If $y = 0$, we're done. So assume $y \neq 0$, and let $c = \langle x, y \rangle / \langle y, y \rangle$ be the component of x along y. By construction, $x - cy$ is perpendicular to y and hence to cy. Therefore, by the Pythagorean theorem,

$$\|x - cy\|^2 + \|cy\|^2 = \|(x - cy) + cy\|^2 = \|x\|^2.$$

Since $\|x - cy\|^2 \geq 0$, if we drop that term in the above equation, we get

$$\|cy\|^2 \leq \|x\|^2.$$

Take square roots to get

$$\|x\| \geq \|cy\| = |c|\|y\| = \left| \frac{\langle x, y \rangle}{\|y\|^2} \right| \|y\| = \frac{|\langle x, y \rangle|}{\|y\|}.$$

Multiply through by $\|y\|$ to get Cauchy-Schwarz.

For the proof of the triangle inequality, we will need two basis results concerning complex numbers. Let $z = a + bi$ be any complex number. Then we have (i) $z + \overline{z} = (a + bi) + (a - bi) = 2a$. So

$$z + \overline{z} = 2\,\mathrm{Re}(z),$$

and (ii) $|z| = \sqrt{a^2 + b^2} \geq |a|$. So

$$\mathrm{Re}(z) \leq |z|.$$

The triangle inequality is then an easy consequence of Cauchy-Schwarz:

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y, x + y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2 \\
&= \|x\|^2 + \langle x, y\rangle + \overline{\langle x, y\rangle} + \|y\|^2 \\
&= \|x\|^2 + 2\,\mathrm{Re}(\langle x, y\rangle) + \|y\|^2 \\
&\leq \|x\|^2 + 2\,|\langle x, y\rangle| + \|y\|^2 \\
&\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\
&\leq (\|x\| + \|y\|)^2.
\end{aligned}$$

Take square roots to get the triangle inequality. $\square$

Distance. Let $(V, \langle\ ,\ \rangle)$ be an inner product space over $\mathbb{R}$ or $\mathbb{C}$. The *distance* between $x, y \in V$ is defined to be

$$d(x, y) := \|x - y\|.$$

The following properties then easily follow from what we have already done:

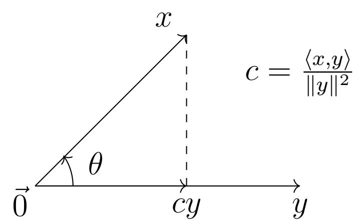
Proposition. For all $x, y, z \in V$,

(a) Symmetry: $d(x, y) = d(y, x)$.

(b) Positive-definiteness: $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$.

(c) Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

Angles. Now let $(V, \langle\ ,\ \rangle)$ be an inner product space over $F = \mathbb{R}$. (So we will not consider the case $F = \mathbb{C}$ in our discussion of angles.) We would like to define the notion of an *angle* between $x, y \in V$. Our picture for the component provides motivation:



The dashed vertical line and the vector y are perpendicular (by definition of c). The cosine of the displayed angle should be the length of the base, cy, divided by the length of the hypotenuse, x. That quotient is

$$\frac{\|cy\|}{\|x\|} = |c|\frac{\|y\|}{\|x\|} = \frac{|\langle x, y\rangle|}{\|y\|^2}\frac{\|y\|}{\|x\|} = \frac{|\langle x, y\rangle|}{\|x\|\|y\|}.$$

Omitting the absolute value on the real number $\langle x, y\rangle$ in the numerator provides the correct signs for the different quadrants (when θ is not between 0 and 90 degrees).

Definition. Let $(V, \langle\ ,\ \rangle)$ be an inner product space over $F = \mathbb{R}$, and let x, y be nonzero elements of V. The *angle* θ between x and y is

$$\theta = \cos^{-1}\left(\frac{\langle x, y\rangle}{\|x\|\|y\|}\right),$$

and thus,

$$\langle x, y\rangle = \|x\|\|y\|\cos(\theta).$$

Remarks.

- Cauchy-Schwarz says $|\langle x, y\rangle \leq \|x\|\|y\|$. Therefore,

$$-1 \leq \frac{\langle x, y\rangle}{\|x\|\|y\|} \leq 1.$$

  So the inverse cosine in the definition of the angle always makes sense.

- In the definition of the angle, it might make more sense conceptually to write

$$\cos(\theta) = \left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \right\rangle.$$

  In other words, the cosine of the angle between x and y is the inner product of their directions where the *direction* of a vector w is taken to be the scalar multiple of w with unit length, $w/\|w\|$.

Week 12, Monday: Gram-Schmidt

Let $(V, \langle\ ,\ \rangle)$ be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$.

Definition. Let $S \subseteq V$. Then S is an *orthogonal* subset of V if $\langle u, v\rangle = 0$ for all $u, v \in S$ with $u \neq v$. If S is an orthogonal subset of V and $\|u\| = 1$ for all $u \in S$, then S is an *orthonormal* subset of V.

Examples.

- The standard basis $e_1, \ldots, e_n$ for F^n is orthonormal with respect to the standard inner product on F^n.
- $\left\{\frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1)\right\}$ is orthonormal with respect to the standard inner product on $\mathbb{R}^2$.

Proposition. Let $S = \{v_1, \ldots, v_k\}$ be an orthogonal set of nonzero vectors in V, and let $y \in \operatorname{Span} S$. Then

$$y = \sum_{j=1}^{k} \frac{\langle y, v_j\rangle}{\langle v_j, v_j\rangle} v_j = \sum_{j=1}^{k} \frac{\langle y, v_j\rangle}{\|v_j\|^2} v_j.$$

Note that the coefficients are the components of y along each v_j.

Proof. Say $y = \sum_{i=1}^{k} a_i v_i$. Then for $j = 1, \ldots, k$,

$$\langle y, v_j\rangle = \langle \textstyle\sum_{i=1}^{k} a_i v_i, v_j\rangle = \sum_{i=1}^{k} a_i \langle v_i, v_j\rangle = a_j \langle v_j, v_j\rangle,$$

since $\langle v_i, v_j\rangle = 0$ for $i \neq j$. Hence,

$$a_j = \frac{\langle y, v_j\rangle}{\langle v_j, v_j\rangle} = \frac{\langle y, v_j\rangle}{\|v_j\|^2},$$

the component of y along v_j. $\square$

Corollary 1. If $S = \{v_1, \ldots, v_k\}$ is orthonormal and $y \in \operatorname{Span} S$, then

$$y = \sum_{i=1}^{k} \langle y, v_j \rangle v_i.$$

Corollary 2. Is $S = \{v_1, \ldots, v_k\}$ is an orthogonal set of nonzero vectors in V then S is linearly independent.

Proof. If $\sum_{i=1}^{k} a_i v_i = 0$, then for each $j = 1, \ldots, k$,

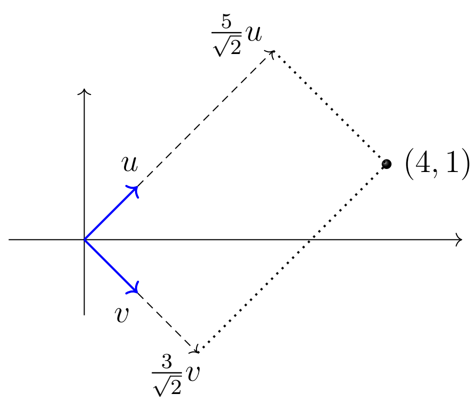
$$0 = \langle 0, v_j \rangle = \langle \textstyle\sum_{i=1}^{k} a_i v_i, v_j \rangle = a_j \langle v_j, v_j \rangle.$$

Since $v_j \neq 0$ and $\langle\ ,\ \rangle$ is positive-definite, we have $\langle v_j, v_j \rangle \neq 0$. Hence, $a_j = 0$ for $j = 1, \ldots, k$. $\square$

Example. Consider $\mathbb{R}^2$ with the standard inner product, and let

$$u = \frac{1}{\sqrt{2}}(1, 1) \quad \text{and} \quad v = \frac{1}{\sqrt{2}}(1, -1).$$

Then $\beta = \{u, v\}$ gives an orthonormal ordered basis for $\mathbb{R}^2$. What are the coordinates of $y = (4, 1)$ with respect to that basis?



Answer:

$$\begin{aligned} y &= \langle y, u \rangle u + \langle y, v \rangle v \\ &= (4, 1) \cdot \left(\frac{1}{\sqrt{2}}(1, 1) \right) u + (4, 1) \left(\frac{1}{\sqrt{2}}(1, -1) \right) v \end{aligned}$$

$$= \frac{5}{\sqrt{2}}\,u + \frac{3}{\sqrt{2}}\,v.$$

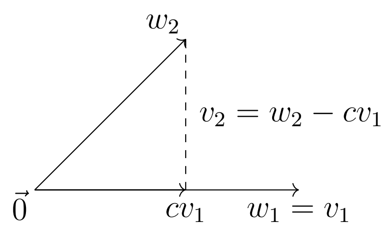
Check:

$$\frac{5}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(1,1)\right) + \frac{3}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(1,-1)\right) = \frac{5}{2}\,(1,1) + \frac{3}{2}(1,-1) = (4,1).$$

Gram-Schmidt. Given vectors $w_1, w_2 \in V$, we'd like to compute orthogonal vectors v_1, v_2 such that

$$\operatorname{Span}\{w_1, w_2\} = \operatorname{Span}\{v_1, v_2\}.$$

To do that, let $v_1 = w_1$, then "straighten out" w_2 to create v_2:



The number c is the component of w_2 along v_1. Recall, c is determined by requiring v_2 and v_1 to be orthogonal:

$$0 = \langle v_2, v_1\rangle = \langle w_2 - cv_1, v_1\rangle = \langle w_2, v_2\rangle - c\langle v_1, v_1\rangle.$$

Therefore,

$$c = \frac{\langle w_2, v_1\rangle}{\langle v_1, v_1\rangle} = \frac{\langle w_2, v_1\rangle}{\|v_1\|^2}.$$

(We've assumed $v_1 \neq 0$.)

The following algorithm generalizes this idea:

Algorithm. (Gram-Schmidt orthogonalization)

INPUT: $S = \{w_1, \ldots, w_n\}$, a linearly independent subset of V.

Let

$$v_1 := w_1.$$

For $k = 2, 3, \ldots, n$, define v_k by starting with w_k, then subtracting off the components of w_k along the previously found v_i :

$$v_k := w_k - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i\rangle}{\|v_i\|^2} v_i.$$

OUTPUT: $S' = \{v_1, \ldots, v_n\}$ an orthogonal set with $\operatorname{Span} S' = \operatorname{Span} S$.

or

OUTPUT: $S'' = \left\{ \dfrac{v_1}{\|v_1\|}, \ldots, \dfrac{v_n}{\|v_n\|} \right\}$ an orthonormal set with $\operatorname{Span} S' = \operatorname{Span} S$.

Proof of validity of the algorithm. We prove this by induction on n. The case $n = 1$ is clear. Suppose the algorithm works for some $n \geq 1$, and let $S = \{w_1, \ldots, w_{n+1}\}$ be a linearly independent set. By induction, running the algorithm on the first n vectors in S produces orthogonal $v_1, \ldots, v_n$ with

$$\operatorname{Span}\{v_1, \ldots, v_n\} = \operatorname{Span}\{w_1, \ldots, w_n\}.$$

Running the algorithm further produces

$$v_{n+1} = w_{n+1} - \sum_{i=1}^{n} \frac{\langle w_{n+1}, v_i \rangle}{\|v_i\|^2} v_i.$$

It cannot be that $v_{n+1} = 0$, since otherwise, the above equation we would say

$$w_{n+1} \in \operatorname{Span}\{v_1, \ldots, v_n\} = \operatorname{Span}\{w_1, \ldots, w_n\},$$

contradicting the assumption of the linear independence of the w_i. So $v_{n+1} \neq 0$. We now check that v_{n+1} is orthogonal to the previous v_i. For $j = 1, \ldots, n$, we have

$$\begin{aligned}
\langle v_{n+1}, v_j \rangle &= \left\langle w_{n+1} - \sum_{i=1}^{n} \frac{\langle w_{n+1}, v_i \rangle}{\|v_i\|^2} v_i, v_j \right\rangle \\
&= \langle w_{n+1}, v_j \rangle - \sum_{i=1}^{n} \frac{\langle w_{n+1}, v_i \rangle}{\|v_i\|^2} \langle v_i, v_j \rangle \\
&= \langle w_{n+1}, v_j \rangle - \frac{\langle w_{n+1}, v_j \rangle}{\|v_j\|^2} \langle v_j, v_j \rangle \\
&= \langle w_{n+1}, v_j \rangle - \langle w_{n+1}, v_j \rangle \\
&= 0.
\end{aligned}$$

We have shown $\{v_1, \ldots, v_{n+1}\}$ is an orthogonal set of vectors, and we would now like to show that its span is the span of $\{w_1, \ldots, w_{n+1}\}$. First, since $\{v_1, \ldots, v_{n+1}\}$ is orthogonal, it's linearly independent. Next, we have

$$\operatorname{Span}\{v_1, \ldots, v_{n+1}\} \subseteq \operatorname{Span}\{v_1, \ldots, v_n, w_{n+1}\} \subseteq \operatorname{Span}\{w_1, \ldots, w_n, w_{n+1}\}.$$

Since Span $\{v_1, \ldots, v_{n+1}\}$ is an $(n+1)$-dimensional subspace of the $(n+1)$-dimensional space Span $\{w_1, \ldots, w_n, w_{n+1}\}$, these spaces must be equal. $\square$

Corollary. Every nonzero finite-dimensional inner product space has an orthonormal basis.

Example. Let $V = \mathbb{R}_{\leq 1}[x]$, the space of polynomials of degree at most 1 with real coefficients and with inner product

$$\langle f, g \rangle = \int_0^1 f(t)g(t)\,dt.$$

Apply Gram-Schmidt to the basis $\{1, x\}$ to get an orthonormal basis. Note that 1 and x are not orthogonal:

$$\langle 1, x \rangle = \int_0^1 t\,dt = \frac{1}{2} \neq 0.$$

Gram-Schmidt: Start with $v_1 = 1$, then let

$$\begin{aligned} v_2 &= x - \frac{\langle x, v_1 \rangle}{\|v_1\|^2} v_1 \\ &= x - \frac{\langle x, 1 \rangle}{\|1\|^2} \cdot 1 \\ &= x - \frac{\int_0^1 t\,dt}{\int_0^1 dt} \cdot 1 \\ &= x - \frac{1}{2}. \end{aligned}$$

Check orthogonality:

$$\langle 1, x - 1/2 \rangle = \int_0^1 (t - 1/2)\,dt = 0.$$

Now scale $v_1 = 1$ and $v_2 = x - 1/2$ to create an orthonormal basis:

$$\begin{aligned} \|v_1\| &= \sqrt{\int_0^1 dt} = 1 \\ \|v_2\| &= \sqrt{\langle x - 1/2, x - 1/2 \rangle} \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{\int_0^1 (t-1/2)^2\,dt} \\
&= \sqrt{\int_0^1 (t^2 - t + 1/4)\,dt} \\
&= \sqrt{1/12}.
\end{aligned}
$$

So an orthonormal basis for V is

$$\left\{1, \sqrt{12}(x-1/2)\right\}.$$

Week 12, Wednesday: Orthogonal complements and projections

Definition. The *direct sum* of vector spaces U and W over a field F is the set

$$U \oplus W = \{(u, w) : u \in U \text{ and } w \in W\}$$

with scalar multiplication and vector addition defined by

$$\lambda(u, w) = (\lambda u, \lambda w) \quad \text{and} \quad (u, w) + (u', w') = (u + u', w + w'),$$

for all $u, u' \in U$, $w, w' \in W$, and $\lambda \in F$.

Proposition. Let U and W be subspaces of a vector space V over F such that: (i) the union of U and W spans V, and (ii) $U \cap W = \{0\}$. Then there is an isomorphism

$$\begin{aligned} U \oplus W &\to V \\ (u, w) &\mapsto u + w. \end{aligned}$$

Thus, every element of V has a unique expression of the form $u + w$ with $u \in U$ and $w \in W$.

Proof. Easy exercise. $\square$

Remark. In the case of the Proposition, we says that V is the *internal direct sum* of U and W and abuse notation by simply writing $V = U \oplus W$. The direct sum as we first defined it is sometimes called the *external direct sum* of U and W.

For the rest of this lecture, let $(V, \langle\ ,\ \rangle)$ be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$.

Definition. Let $S \subseteq V$ be nonempty. The *orthogonal complement* of S is

$$S^{\perp} = \{x \in V : \langle x, y \rangle = 0 \text{ for all } y \in S\}.$$

Exercise. Show that $S^\perp$ is a subspace of V.

Example. Consider $\mathbb{R}^3$ with the standard inner product, and let $S = \{(a, b, c)\}$. So S consists of the single vector $(a, b, c) \in \mathbb{R}^3$. Then

$$S^\perp = \{(x, y, z) \in \mathbb{R}^3 : (x, y, z) \cdot (a, b, c) = 0\} = \{(x, y, z) \in \mathbb{R}^3 : ax + by + cz = 0\},$$

a plane in $\mathbb{R}^3$ defined by the equation $ax + by + cz = 0$.

Proposition. Suppose $\dim V = n$ and $S = \{v_1, \ldots, v_k\}$ is an orthonormal subset of V.

(a) S can be extended to an orthonormal basis $\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for V.

(b) If $W = \operatorname{Span} S$, then $S' = \{v_{k+1}, \ldots, v_n\}$ is an orthonormal basis for $W^\perp$.

(c) If $W \subseteq V$ is any subspace, then

$$\dim W + \dim W^\perp = \dim V = n.$$

(d) If $W \subseteq V$ is any subspace, then $(W^\perp)^\perp = W$.

Proof. (a) To prove part (a), extend S to a basis $\{v_1, \ldots, v_k, w_{k+1}, \ldots, w_n\}$ for V, then apply Gram-Schmidt.

(b) The set $S' = \{v_{k+1}, \ldots, v_n\}$ is linearly independent since it's a subset of a basis. Since $\{v_1, \ldots, v_n\}$ is orthonormal, and $W = \operatorname{Span}\{v_1, \ldots, v_k\}$, we have $S' \subseteq W^\perp$. Therefore, $\operatorname{Span} S' \subseteq W^\perp$. For the opposite inclusion, take $x \in W^\perp$. Then since $\{v_1, \ldots, v_n\}$ is orthonormal, we have

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle\, v_i = \sum_{i=k+1}^{n} \langle x, v_i \rangle\, v_i \in \operatorname{Span} S'.$$

(c) If $W \subseteq V$ is any subspace, choose an orthonormal basis $\{v_1, \ldots, v_k\}$ for W. Then apply parts (a) and (b).

(d) It's clear that $W \subseteq (W^\perp)^\perp$ since

$$(W^\perp)^\perp = \left\{ x \in V : \langle x, y \rangle = 0 \text{ for all } y \in W^\perp \right\}.$$

Then, by part (c),

$$\dim(W^\perp)^\perp = n - \dim W^\perp = \dim W.$$

Hence, $W = (W^\perp)^\perp$. □

Proposition. Let W be a finite-dimensional subspace of V. Then

$$V = W \oplus W^{\perp}.$$

In other words, for each $y \in V$, there exist unique $u \in W$ and $z \in W^{\perp}$ such that

$$y = u + z.$$

We define u to be the *orthogonal projection* of y onto W.

If $u_1, \ldots, u_k$ is an orthonormal basis for W, then

$$u = \sum_{i=1}^{k} \langle y, u_i \rangle \, u_i.$$

Proof. By Gram-Schmidt, there exists an orthonormal basis $u_1, \ldots, u_k$ for W. Define $u = \sum_{i=1}^{k} \langle y, u_i \rangle \, u_i$ and $z = y - u$. Then $u \in W$ and $y = u + z$. Further, $z \in W^{\perp}$ since for each $j = 1, \ldots, k$, we have

$$\begin{aligned}
\langle z, u_j \rangle &= \langle y - u, u_j \rangle \\
&= \langle y, u_j \rangle - \langle \textstyle\sum_{i=1}^{k} \langle y, u_i \rangle \, u_i, u_j \rangle \\
&= \langle y, u_j \rangle - \textstyle\sum_{i=1}^{k} \langle y, u_i \rangle \, \langle u_i, u_j \rangle \\
&= \langle y, u_j \rangle - \langle y, u_j \rangle \, \langle u_j, u_j \rangle \\
&= \langle y, u_j \rangle - \langle y, u_j \rangle \\
&= 0.
\end{aligned}$$

For uniqueness, suppose there exist $u' \in W$ and $z' \in W^{\perp}$ such that

$$y = u + z = u' + z'.$$

Then $u - u' = z' - z \in W \cap W^{\perp} = \{0\}$. Thus, $u = u'$ and $z = z'$. (The reason $W \cap W^{\perp} = \{0\}$ is as follows: if $x \in W$, then we saw last time that $x = \sum_{i=1}^{k} \langle x, u_i \rangle \, u_i$. If it is also the case that $x \in W^{\perp}$, then $\langle x, u_i \rangle = 0$ for $i = 1, \ldots, k$ since each u_i is in W. Hence, $x = 0$.) $\square$

Corollary. The orthogonal projection u of y onto W is the closest vector in W to y:

$$\|y - u\| \leq \|y - w\|$$

for all $w \in W$ with equality if and only if $w = u$.

Proof. Write $y = u + z$ with $u \in W$ and $z \in W^\perp$, and let $w \in W$. Then $u - w \in W$ and $y - u \in W^\perp$. So $u - w$ and $z = y - u$ are perpendicular. By the Pythagorean theorem,

$$\begin{aligned}\|y - w\|^2 &= \|(u+z) - w\|^2 \\ &= \|(u-w) + z\|^2 \\ &= \|(u-w)\|^2 + \|z\|^2 \\ &\geq \|z\|^2 \\ &= \|y-u\|^2.\end{aligned}$$

Equality occurs above if and only if $\|u - w\|^2 = 0$, i.e., if and only if $u = w$. $\square$

Example. Let $V = \mathbb{R}^3$ with the standard inner product, and let's consider orthogonal projection onto the xy-plane. An orthonormal basis for the xy-plane is $\{e_1, e_2\}$. The projection of a point $u = (x, y, z) \in \mathbb{R}^3$ is given by

$$u = ((x,y,z)\cdot e_1)e_1 + ((x,y,z)\cdot e_2)e_2 = x\, e_1 + y\, e_2 = (x, y, 0).$$

The distance of (x, y, z) to the xy-plane is

$$\|(x,y,z) - u\| = \|(0,0,z)\| = |z|.$$

Application. Consider the vector space V of integrable functions $f\colon [0, 2\pi] \to \mathbb{R}$ with inner product

$$\langle f, g\rangle := \frac{1}{\pi}\int_0^{2\pi} f(t)g(t)\, dt.$$

Thus, the distance between $f, g \in V$ is

$$\|f - g\| = \sqrt{\frac{1}{\pi}\int_0^{2\pi} (f(t) - g(t))^2\, dt},$$

which will be small if $f(t) \approx g(t)$ for $t \in [0, 2\pi]$.

One may check that $S_n := \left\{\frac{1}{\sqrt{2}}, \cos(x), \sin(x), \cos(2x), \sin(2x), \dots, \cos(nx), \sin(nx)\right\}$ is an orthonormal subset. Given any integrable $f \in V$, the orthogonal projection of f to the subspace spanned by S_n gives the best approximation of the function using sines and cosines of frequencies $\frac{j}{2\pi}$ for $j = 0, \dots, n$. Write the projection of f to $\mathrm{Span}(S_n)$ as

$$\mathrm{proj}_{\mathrm{Span}(S_n)}(f)(x) = \alpha \cdot \frac{1}{\sqrt{2}} + \sum_{i=1}^n \beta_i \cos(ix) + \sum_{i=1}^n \gamma_i \sin(ix),$$

Since S_n is orthonormal, we may find the coefficients by taking inner products:

$$\alpha = \langle f, 1/\sqrt{2}\rangle = \frac{1}{\pi}\int_0^{2\pi} \frac{f(t)}{\sqrt{2}}\,dt$$

$$\beta_i = \langle f, \cos(ix)\rangle = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos(ix)\,dt$$

$$\gamma_i = \langle f, \sin(ix)\rangle = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin(ix)\,dt.$$

For instance, consider the function $f(x) = x$ for $x \in [0, 2\pi]$. We find

$$\alpha = \frac{1}{\pi}\int_0^{2\pi} \frac{t}{\sqrt{2}}\,dt = \sqrt{2}\pi.$$

Integrating by parts, we find

$$\beta_i = \frac{1}{\pi}\int_0^{2\pi} t\cos(it)\,dt = \frac{1}{\pi}\left(\frac{t\sin(it)}{i} + \frac{\cos(it)}{i^2}\right|_0^{2\pi} = 0$$

and

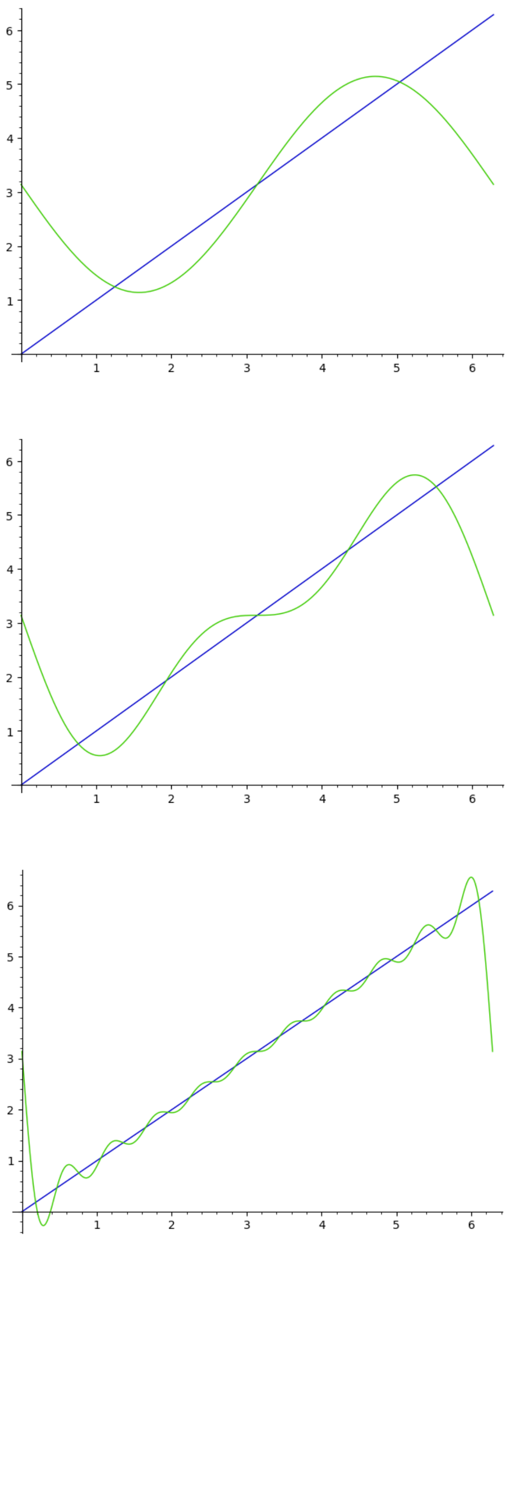
$$\gamma_i = \frac{1}{\pi}\int_0^{2\pi} t\sin(it)\,dt = \frac{1}{\pi}\left(-\frac{t\cos(it)}{i} + \frac{\sin(it)}{i^2}\right|_0^{2\pi} = -\frac{2}{i}.$$

Thus,

$$\operatorname{proj}_{\mathrm{Span}(S_n)}(f)(x) = \sqrt{2}\pi \cdot \frac{1}{\sqrt{2}} - \sum_{i=1}^{n} \frac{2}{i\pi}\sin(ix) = \pi - \frac{2}{\pi}\sum_{i=1}^{n}\frac{\sin(ix)}{i}.$$

See the next page to compare the graph of f with the graphs of these projections for various n.

The plot of $\mathrm{proj}_{\mathrm{Span}(S_n)}(f)$ versus the plot of $f(x) = x$ for $n = 1, 2,$ and 10.



Week 13, Monday: Systems of linear differential equations

Suppose that

$$x(t) = \text{amount of yeast at time } t$$

and that rate of growth of yeast (at least in the time frame in which we are interested) is proportional to the amount of yeast. So there exists a constant a such that

$$x'(t) = ax(t).$$

Integrating, we get

$$\int \frac{x'(t)}{x(t)}\,dt = \int a\,dt \quad \Rightarrow \quad \ln(x(t)) = at + b$$

for some constant b. Exponentiating then yields

$$x(t) = e^{at}c$$

where $c = e^b$. Evaluating at $t = 0$ shows that c is the *initial condition*: $x(0) = c$.

Now consider a two-dimensional system. Let

$$\begin{aligned} x_1(t) &= \text{ population of frogs in a pond} \\ x_2(t) &= \text{ population of flies in a pond,} \end{aligned}$$

and suppose the rate of change of these populations satisfies the following system of differential equations:

$$\begin{aligned} x_1'(t) &= ax_1(t) + bx_2(t) \\ x_2'(t) &= cx_1(t) + dx_2(t). \end{aligned}$$

So we are assuming that the rate of growth of these populations depends linearly on the sizes of the populations. Letting

$$x(t) := \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \quad \text{and} \quad x'(t) := \begin{pmatrix} x_1'(t) \\ x_2'(t) \end{pmatrix},$$

we can rewrite the system in matrix form:

$$x'(t) = Ax(t)$$

where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Our problem is to find $x(t)$. The key to solving higher-dimensional systems like this is the following:

Theorem. Let A be an $n \times n$ matrix over the real or complex numbers. Then the solution to $x' = Ax$ with initial condition $x(0) = p$ is

$$x = e^{At}p.$$

(Note that p is a column vector here.)

To make sense of this, we need to be able to exponentiate a matrix! To do that, recall that for a real or complex number a, we have

$$e^a = \sum_{k=0}^{\infty} \frac{1}{k!} a^k,$$

an infinite series that converges for all a. This formula generalizes: given any $n \times n$ matrix A over the real or complex numbers, define

$$e^{At} := \sum_{k=0}^{\infty} \frac{1}{k!}(At)^k = \sum_{k=0}^{\infty} \frac{1}{k!} A^k t^k = I_n + At + \frac{1}{2}A^2t^2 + \frac{1}{6}A^3t^2 + \frac{1}{24}A^4t^4 + \cdots.$$

Each entry of e^{At} is a power series in t, and that power series turns out to converge for all t. To compute e^{At} though, we need to somehow compute all of the powers of A. As you might expect, diagonalization comes to the rescue.

Computing e^{At}. If A is diagonalizable, then we can write

$$P^{-1}AP = D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

where the λ_i are the eigenvalues of A. As we have seen earlier, it follows that

$$A^k = (PDP^{-1})^k = PD^kP^{-1} = P\operatorname{diag}(\lambda_1^k, \dots, \lambda_n^k)P^{-1}.$$

Therefore, modulo some technicalities involving convergence, we have

$$e^{At} = \sum_{k=0}^{\infty} \frac{1}{k!} A^k t^k = \sum_{k=0}^{\infty} \frac{1}{k!}(PD^kP^{-1})t^k = P\left(\sum_{k=0}^{\infty} \frac{1}{k!} D^k t^k\right) P^{-1} = Pe^{Dt}P^{-1}.$$

Since D is diagonal, an easy calculation shows that

$$e^{Dt} = \operatorname{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t}).$$

So

$$e^{At} = P \operatorname{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t}) P^{-1}.$$

Example. Consider the following two-dimensional system:

$$\begin{aligned} x_1' &= x_2 \\ x_2' &= x_1. \end{aligned}$$

In matrix form,

$$x' = Ax$$

where

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

Applying our algorithm to diagonalize A, we find

$$P^{-1}AP = D = \operatorname{diag}(1, -1)$$

where

$$P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

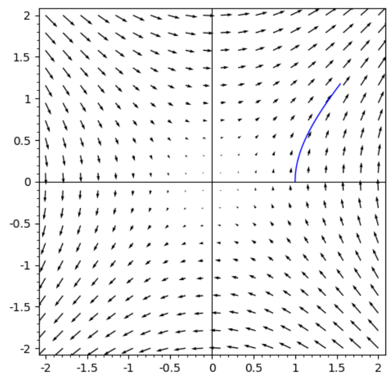
Therefore,

$$\begin{aligned} e^{At} &= Pe^{Dt}P^{-1} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} e^t + e^{-t} & e^t - e^{-t} \\ e^t - e^{-t} & e^t + e^{-t} \end{pmatrix}. \end{aligned}$$

So, for example, the solution with initial condition $x(0) = (1, 0)$ is

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} e^t + e^{-t} & e^t - e^{-t} \\ e^t - e^{-t} & e^t + e^{-t} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} e^t + e^{-t} \\ e^t - e^{-t} \end{pmatrix},$$

A plot of that solution $(x_1(t), x_2(t))$ in the plane appears in blue in the picture below. The arrows indicate the following: at each point $(x_1, x_2) \in \mathbb{R}^2$, we attach the velocity vector $(x_1', x_2') = (x_2, x_1)$.



The solution in blue has velocity vector $x'(0) = (x_2(0), x_1(0)) = (0, 1)$ at time $t = 0$. To repeat: geometrically, our solution is a parametrized curve in the plane:

$$\begin{aligned} x\colon \mathbb{R} &\to \mathbb{R}^2 \\ t &\mapsto x(t) = (x_1(t), x_2(t)). \end{aligned}$$

The differential equation specifies the tangent (velocity) vectors $x'(t)$ at each time t. It determines a "flow" as illustrated in the picture. Specifying an initial condition is like dropping a speck into the flow. We then get a unique solution, which is the trajectory of that speck over time (shown in blue, above).

Note: the arrows determine new "axes" pointed in the directions of the eigenvectors, $(1, 1)$ and $(1, -1)$.

Example. Next consider the following two-dimensional system:

$$\begin{aligned} x_1' &= x_2 \\ x_2' &= -x_1. \end{aligned}$$

It might approximate frog-fly populations since one would expect the frog population $x_1(t)$ to increase with the fly population and the fly population to decrease with the frog population. In matrix form,

$$x' = Ax$$

where

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

The characteristic polynomial is $p_A(x) = x^2 + 1$, so A is not diagonalizable over $\mathbb{R}$. However, it is diagonalizable over $\mathbb{C}$. So let's do that to see where that goes. Applying our algorithm to diagonalize A, we find

$$P^{-1}AP = D = \operatorname{diag}(i, -i)$$

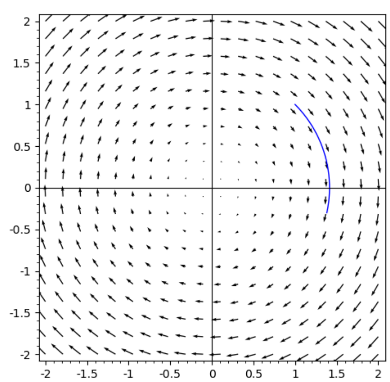
where

$$P = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}.$$

$$\begin{aligned}
e^{At} &= Pe^{Dt}P^{-1} \\
&= \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} \begin{pmatrix} e^{i\,t} & 0 \\ 0 & e^{-i\,t} \end{pmatrix} \begin{pmatrix} -\frac{1}{2}i & \frac{1}{2} \\ \frac{1}{2}i & \frac{1}{2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2}\,e^{i\,t} + \frac{1}{2}\,e^{-i\,t} & -\frac{1}{2}i\,e^{i\,t} + \frac{1}{2}i\,e^{-i\,t} \\ \frac{1}{2}i\,e^{i\,t} - \frac{1}{2}i\,e^{-i\,t} & \frac{1}{2}\,e^{i\,t} + \frac{1}{2}\,e^{-i\,t} \end{pmatrix} \\
&= \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix}.
\end{aligned}$$

So starting with equal populations of frogs and flies, $x(0) = (1, 1)$, we have

$$x(t) = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (\cos(t) + \sin(t), -\sin(t) + \cos(t)).$$



Note how this system of equations is not a great model for frogs and flies: starting at any initial population, the system evolves into one in which there are negative amounts of frogs or flies. One could hope that it applies locally, say near times at which the populations for frogs and flies is nearly equal. At any rate, it raises the question as to whether any *linear* system of equations would make a good model. Qualitatively, what are all of the possibilities for a two-dimensional linear system?

Week 13, Wednesday: Cross product

Let $v_1, \dots, v_{n-1}$ be a set of $n-1$ vectors in $\mathbb{R}^n$. Define the function

$$\begin{aligned} \chi \colon \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto \det(x, v_1, \dots, v_{n-1}). \end{aligned}$$

where we think of the determinant as a function of the rows $x, v_1, \dots, v_{n-1}$ of a matrix, as usual. The $1 \times n$ matrix representing χ has the form $(a_1 \cdots a_n)$. We define the *cross product* to be the row vector

$$v_1 \times \cdots \times v_{n-1} := (a_1, \dots, a_n).$$

The mapping χ is just dot product with the cross product:

$$\chi(x) = (a_1 \cdots a_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (a_1, \dots, a_n) \cdot x = (v_1 \times \cdots \times v_{n-1}) \cdot x.$$

matrix multiplication dot product

Theorem. (Properties of the cross product.)

(a) The cross product is a multilinear alternating function of $v_1, \dots, v_{n-1}$.

(b) Swapping v_i with v_j for $i \neq j$ changes the sign of the cross product.

(c) Adding a scalar multiple of v_i to v_j for some $i \neq j$ does not change the cross product.

(d) The cross product is orthogonal to the subspace spanned by $v_1, \dots, v_{n-1}$.

(e) The length of the cross product is the volume of the parallelepiped spanned by $v_1, \dots, v_{n-1}$.

(f) Given $w \in \mathbb{R}^n$, the volume of the parallelepiped spanned by w and $v_1, \ldots, v_{n-1}$ is $|w \cdot (v_1 \times \cdots \times v_{n-1})|$.

(g) Let A be the $(n-1) \times n$ matrix with rows $v_1, \ldots, v_{n-1}$, and let $A^{(j)}$ be the $(n-1) \times (n-1)$ matrix formed by removing the j-th column of A. Then

$$v_1 \times \cdots \times v_{n-1} = \left(\det(A^{(1)}), -\det(A^{(2)}), \det(A^{(3)}), \ldots, (-1)^{n-1}\det(A^{(n)})\right).$$

Proof. Properties (a)–(c) follow immediately from the properties of $\det(x, v_1, \ldots, v_{n-1})$. For property (d), note that

$$(v_1 \times \cdots \times v_{n-1}) \cdot v_i = \det(v_i, v_1, \ldots, v_{n-1}) = 0$$

since v_i is a repeated row.

For property (e), let P be the parallelepiped spanned by $v_1, \ldots, v_{n-1}$, and let Q be the parallelepiped spanned by $v_1 \times \cdots \times v_{n-1}$ and $v_1, \ldots, v_{n-1}$. Since $v_1 \times \cdots \times v_{n-1}$ is perpendicular to P, the volume of Q is given by the volume of the base, P, times the height $\|v_1 \times \cdots \times v_{n-1}\|$:

$$\operatorname{vol}(Q) = \|v_1 \times \cdots \times v_{n-1}\| \operatorname{vol}(P). \tag{36.1}$$

The volume of Q is the absolute value of the determinant of its spanning vectors. Therefore,

$$\begin{aligned}
\operatorname{vol}(Q) &= |\det(v_1 \times \cdots \times v_{n-1}, v_1, \ldots, v_{n-1})| \\
&= |\chi(v_1 \times \cdots \times v_{n-1}, v_1, \ldots, v_{n-1})| \\
&= (v_1 \times \cdots \times v_{n-1}) \cdot (v_1 \times \cdots \times v_{n-1}) \\
&= \|v_1 \times \cdots \times v_{n-1}\|^2.
\end{aligned}$$

Combining this with equation (36.1) yields the result:

$$\|v_1 \times \cdots \times v_{n-1}\| = \operatorname{vol}(P).$$

For property (f), note that

$$|w \cdot (v_1 \times \cdots \times v_{n-1})| = |\det(w, v_1, \ldots, v_{n-1})|,$$

which gives the volume of the parallelepiped in question.

Property (g) follows by expanding the determinant defining χ along its first row:

$$\chi(x) = \det(x, v_1, \ldots, v_{n-1})$$

$$= \det(A^{(1)}x_1 - \det(A^{(2)})x_2 + \cdots + (-1)^{n-1}\det(A^{(n)})x_n$$
$$= (\det(A^{(1)}, -\det(A^{(2)}), \ldots, (-1)^{n-1}\det(A^{(n)})) \cdot (x_1, \ldots, x_n).$$

$\square$

The cross product in $\mathbb{R}^3$. The cross product is most well-known in the case $n = 3$. Here, we have vectors $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$. The cross product is

$$x \times y = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1) \in \mathbb{R}^3.$$

The usual mnemonic is

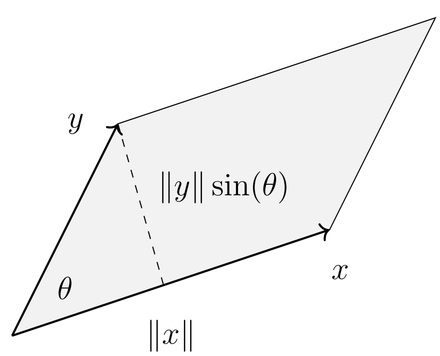
$$x \times y = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix} = (x_2y_3 - x_3y_2)\mathbf{i} - (x_1y_3 - x_3y_1)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k},$$

where $\mathbf{i} = e_1 = (1, 0, 0)$, $\mathbf{j} = e_2 = (0, 1, 0)$, and $\mathbf{k} = e_3 = (0, 0, 1)$. We get exactly the formula given by part (g) of the Theorem. The above is only a mnemonic since we have not defined a determinant in the case where the entries are vectors of various dimensions.

The cross product here is perpendicular to the parallelogram spanned by x and y, and its length is

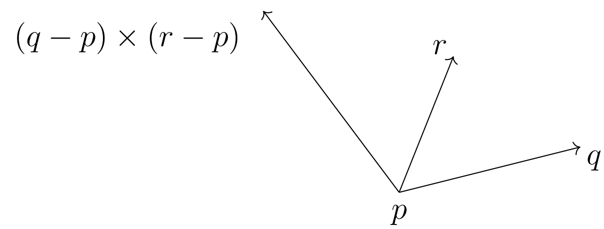
$$\|x \times y\| = \|x\|\|y\| \sin(\theta)$$

where θ is the angle between x and y. This last formula gives the area of the parallelogram spanned by x and y:



Example. Find an equation for the plane through the points $p = (1, 2, 3)$, $q = (1, 0, -2)$, and $r = (0, 7, 2)$.

SOLUTION: To find a vector perpendicular to the plane, we take the cross product of $q-p$ and $r-p$. Below is a picture that illustrates the geometry (with no attempt to get the actual coordinates correct!). The sides of the base parallelogram are spanned by the vectors $q - p$ and $r - p$.



Compute:

$$\begin{aligned}(q-p)\times(r-p) &= (0,-2,-5)\times(-1,5,-1)\\ &= \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k}\\ 0 & -2 & -5\\ -1 & 5 & -1\end{pmatrix}\\ &= 27\,\mathbf{i}+5\,\mathbf{j}-2\,\mathbf{k}\\ &= (27,5,-2).\end{aligned}$$

To double-check, note that the cross product is perpendicular to $q-p$ and $r-p$:

$$(0,-2,-5)\cdot(27,5,-2)=0 \quad \text{and} \quad (-1,5,-1)\cdot(27,5,-2)=0.$$

The set of all points (x,y,z) perpendicular to the cross product is the plane defined by

$$(27,5,-2)\cdot(x,y,z)=0,$$

i.e., the plane with equation

$$27x+5y-2z=0.$$

This plane passes through the origin, $(0,0,0)$. We want the translation of this plane that passes through p. (It will automatically then pass through q and r. So we could choose either q or r for this requirement, instead.) The equation of this translated plane will have the form

$$27x+5y-2z=c.$$

for some constant c. Plug in p (or q or r) to solve for c:

$$c=27(1)+5(2)-2(3)=31.$$

So the equation of the plane is

$$27x+5y-2z=31.$$

(Check that the equation is satisfied by p, q, and r!)

Parametric equation of the plane. As we saw earlier in the semester, we can parametrize this plane by

$$\begin{aligned} f(s,t) &= p + s(q-p) + t(r-p) \\ &= (1,2,3) + s(0,-2,-5) + t(-1,5,-1) \\ &= (1-t, 2-2s+5t, 3-5s-t). \end{aligned}$$

Thus, we get the function:

$$\begin{aligned} f\colon \mathbb{R}^2 &\to \mathbb{R}^3 \\ (s,t) &\mapsto (1-t, 2-2s+5t, 3-5s-t). \end{aligned}$$

The image of f is the plane passing through p, q, and r. One may check that if we let

$$x = 1-t, \quad y = 2-2s+5t, \quad z = 3-5s-t,$$

then $27x + 5y - 2z = 31$, i.e., the point satisfies the equation for the plane.

Week 13, Friday: The Spectral theorem

Theorem (Spectral theorem). Let A be an $n \times n$ symmetric matrix over $\mathbb{R}$. Then A is diagonalizable over $\mathbb{R}$, and there exists an orthonormal basis for $\mathbb{R}^n$ (with respect to the standard inner product) consisting of eigenvectors for A.

Example. Let

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

The characteristic polynomial of A is

$$p_A(t) = \det \begin{pmatrix} -1-t & -1 & -2 \\ -1 & -1-t & 2 \\ -2 & 2 & 2-t \end{pmatrix} = -t^3 + 12t + 16 = (4-t)(-2-t)^2.$$

So the eigenvalues are $4, -2, -2$. We next compute bases for the eigenspaces. For $\lambda = 4$,

$$A - 4I_4 = \begin{pmatrix} -5 & -1 & -2 \\ -1 & -5 & 2 \\ -2 & 2 & -2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

So the eigenspace for $\lambda = 4$ is $E_4\{(-\frac{1}{2}t, \frac{1}{2}t, t) : t \in \mathbb{R}\}$. One basis is $\{(-1, 1, 2)\}$. Normalizing gives the basis vector

$$v_1 = \frac{1}{\sqrt{6}}(-1, 1, 2).$$

For the eigenvalue $\lambda = -2$, we have

$$A + 2I_4 = \begin{pmatrix} 1 & -1 & -2 \\ -1 & 1 & 2 \\ -2 & 2 & 4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and the eigenspace is $E_{-2} = \{(s+2t, s, t) : s, t \in \mathbb{R}\}$. A basis is $\{(1,1,0),(2,0,1)\}$. Applying Gram-Schmidt to these two vectors yields an orthonormal basis for E_{-2} consisting of

$$v_2 = \frac{1}{\sqrt{2}}(1,1,0), \quad v_3 = \frac{1}{\sqrt{3}}(1,-1,1).$$

Now note that something surprising has happened: these vectors are orthogonal to v_1. We arrive at an orthonormal basis $\{v_1, v_2, v_3\}$ for $\mathbb{R}^3$ consisting of eigenvectors for A. Letting P be the 3×3 matrix whose columns are v_1, v_2, v_3, we have

$$P^{-1}AP = \operatorname{diag}(4,-2,2).$$

Since the v_i form an orthonormal set, it turns out that $P^{-1} = P^t$, the transpose of P:

$$P^tP = \begin{pmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Definition. A matrix $P \in M_{n\times n}(\mathbb{R})$ is *orthogonal* if its columns form an orthonormal set in $\mathbb{R}^n$.

Lemma. $P \in M_{n\times n}(\mathbb{R})$ is orthogonal if and only if $P^{-1} = P^t$.

Proof. Note that $(P^tP)_{ij} = v_i \cdot v_j$. So $P^tP = I_n$ if and only if the columns of P form and orthonormal set. □

Restatement of the spectral theorem. If A is a real $n \times n$ symmetric matrix, then there exists a real diagonal matrix D and an orthogonal matrix P such that

$$A = PDP^t.$$

Proof of the spectral theorem. We first prove that the characteristic polynomial of A splits over $\mathbb{R}$. By the Fundamental Theorem of Algebra, it splits over $\mathbb{C}$. So $p_A(t) = \prod_{k=1}^n (\lambda_k - t)$ for some $\lambda_k \in \mathbb{C}$. We must show that $\lambda_k \in \mathbb{R}$ for all k. So let $\lambda = \lambda_k$ for some k. Then there exists a nonzero $v \in \mathbb{C}^n$ such that $Av = \lambda v$. Recall the standard inner product on $\mathbb{C}^n$: for $y, z \in \mathbb{C}^n$, we have $\langle y, z\rangle = y \cdot \bar{z}$. Thinking of y and z as column vectors, we have $\langle y, z\rangle = z^*y$ where $(\ )^*$ denotes the conjugate transpose:

$$\langle y, z\rangle = y \cdot \bar{z} = \sum_{k=1}^n y_i\bar{z}_i = \begin{pmatrix} \bar{z}_1 & \cdots & \bar{z}_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = z^*y.$$

Therefore, for an arbitrary $n \times n$ complex matrix B, we have

$$\langle y, B^* z\rangle = (B^* z)^* y = z^* (B^*)^* y = z^* B y = \langle By, z\rangle.$$

Our matrix, A, is real and symmetric; so $A^* = \bar{A}^t = A^t = A$. Therefore,

$$\langle y, Az\rangle = \langle Ay, z\rangle.$$

Going back to $Av = \lambda v$, we have

$$\lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Av, v\rangle = \langle v, Av\rangle = \langle v, \lambda v\rangle = \bar{\lambda}\langle v, v\rangle.$$

Since $v \neq 0$ and inner products are positive-definite, it follows that $\lambda = \bar{\lambda}$, and hence $\lambda \in \mathbb{R}$.

We now prove the theorem by induction on n, the case $n = 1$ being trivial. Suppose $n > 1$ and let $\lambda_1 \in \mathbb{R}$ and $v_1 \in \mathbb{R}^n$ be an eigenvalue-eigenvector pair for A. Next, complete and apply Gram-Schmidt to construct and ordered orthonormal basis $\langle v_1, \cdots, v_n\rangle$ for $\mathbb{R}^n$. Let Q be the $n \times n$ matrix whose columns are the v_i. Then Q is orthogonal. Define

$$\tilde{A} := Q^{-1} A Q = Q^t A Q.$$

Then $\tilde{A}$ is symmetric:

$$\tilde{A}^t = \left(Q^t A Q\right)^t = Q^t A^t \left(Q^t\right)^t = Q^t A^t Q = Q^t A Q = \tilde{A}.$$

We would like to investigate the structure of $\tilde{A}$ further. To find its first column, we use the fact that $Av_1 = \lambda_1 v_1$. Let e_1 be the first standard basis vector of $\mathbb{R}^n$. Then the first column of $\tilde{A}$ is

$$\tilde{A} e_1 = Q^t A Q e_1 = Q^t A v_1 = Q^t \lambda_1 v_1 = \lambda_1 Q^t v_1.$$

The rows of Q^t are the orthonormal set $v_1, \ldots, v_n$. Therefore,

$$(Q^t v_1)_i = v_i \cdot v_1 = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{otherwise.} \end{cases}$$

So the first column of $\tilde{A}$ is the vector $(\lambda_1, 0, \cdots, 0)$. Since $\tilde{A}$ is symmetric, its first column and first row are the same vector. Therefore, $\tilde{A}$ has the form

$$\left(\begin{array}{c|ccc} \lambda_1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right)$$

where B is an $n \times n$ matrix. Since $\tilde{A}$ is symmetric, so is B. So we can apply induction to find an $(n-1) \times (n-1)$ orthogonal matrix T and a real diagonal matrix E such that $B = TET^t$. We then have

$$\tilde{A} = \underbrace{\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & T & \\ 0 & & & \end{array}\right)}_{S} \underbrace{\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & E & \\ 0 & & & \end{array}\right)}_{D} \underbrace{\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & T^t & \\ 0 & & & \end{array}\right)}_{S^t},$$

where the matrices S and T are defined as shown. Since T is orthogonal, so is S. Finally, define $P = QS$. Since Q and S are orthogonal, so is P (check: $(QS)^t(QS) = S^t(Q^tQ)S = S^tI_nS = I_n$). We have

$$A = Q\tilde{A}Q^t = Q(SDS^t)Q^t = (QS)D(QS)^t = PDP^t,$$

as desired. $\square$

We now discuss a more general version of the spectral theorem.

Definition. A matrix $A \in M_{n\times n}(\mathbb{C})$ is *Hermitian* if $A^* = A$ (so $A = \bar{A}^t$). A matrix $U \in M_{n\times n}(\mathbb{C})$ is *unitary* if its columns are orthonormal, or equivalently, if U is invertible with $U^{-1} = U^*$.

Theorem (Spectral theorem) Let A be an $n \times n$ Hermitian matrix. Then $A = UDU^*$ where U is unitary and D is a real diagonal matrix.

Homework assignments

Week 1, Friday

As for all Math 201 homework this semester, be sure to show your work for full credit, and please acknowledge your collaborators and tutors.

PROBLEM 1. *Calculations.* For each of the following systems of linear equations

- Find the associated augmented matrix M.
- Compute the reduced row echelon form E for M. **Show your work as in class, specifying your row operations.**
- From E determine whether there are solutions to the system. If there is a unique solution, state it. If there are infinitely many solutions, express the set of solutions in two ways: (i) parametrically, as in examples 2.4 and 2.5 in Chapter One, Section I.2, and (ii) in vector form as in Chapter One, Section I.3.

(a)

$$\begin{aligned} x - 2y + z &= 1 \\ -4x + 2y - z &= 0 \\ 3x + 3y - z &= 1. \end{aligned}$$

(b)

$$\begin{aligned} x + y + 3z &= 3 \\ -x + y + z &= -1 \\ 2x + 3y + 8z &= 4. \end{aligned}$$

(c)

$$\begin{aligned} x + y + 3z &= 3 \\ -x + y + z &= -1 \\ 2x + 3y + 8z &= 7. \end{aligned}$$

(d)

$$\begin{aligned} 2x - 2y - 3z &= -2 \\ 3x - 3y - 2z + 5w &= 7 \\ x - y - 2z - w &= -3. \end{aligned}$$

Problem 2. *Some questions about conics.*

(a) Let $y = px^2 + qx + r$ be the equation of a general parabola. By solving a system of equations, find the constants p, q, and r so that the resulting parabola passes through the points $(-2, 15)$, $(1, 3)$, and $(2, 11)$.

(b) A (real) plane conic is a set of points of the form

$$C = \{(x, y) \in \mathbb{R}^2 : ax^2 + bxy + cy^2 + dx + ey + f = 0\}$$

for some constants $a, b, c, d, e, f \in \mathbb{R}$, not all zero. For example, the unit circle centered at the origin is the conic specified by taking $a = c = 1$, $b = d = e = 0$, and $f = -1$ to get the defining equation, $x^2 + y^2 - 1 = 0$. Note that defining equation of a conic is only determined up to a scalar multiple: for instance, $2x^2 + 2y^2 - 2 = 0$, the conic with $a = c = 2$, $b = d = e = 0$, and $f = -2$, also determines the unit circle centered at the origin.

How many points in the plane do you think must be given to determine a specific conic, in general? Why? (Note: You probably don't have the tools yet to rigorously answer this question. What are your thoughts?)

Week 2, Tuesday

Problem 1. Let L be the line in $\mathbb{R}^3$ passing through the points $(1,1,1)$ and $(2,7,4)$.

(a) Find a system of two linear equations whose solution set is L. Show your work. (*Hint*: This will likely involve solving a system of linear equations.)

(b) Give a parametrization of L (you should only need one parameter).

Problem 2. Let H be the plane in $\mathbb{R}^3$ containing the points $(1,1,0)$, $(1,5,-3)$, and $(1,-2,4)$.

(a) Find a linear equation whose solution set is H. Show your work.

(b) Give a parametrization of H.

(c) What happens if we instead consider the three points $(1,1,0)$, $(1,5,-3)$ and $(1,-3,3)$? Is there such a plane? How does the process go differently for (a) and (b)?

Problem 3. Let H be the subset of vectors $(x_1,\dots,x_n)$ in $\mathbb{R}^n$ given by the set of solutions to the equation

$$a_1x_1+\cdots+a_nx_n=d,$$

for some constants $a_1,\dots,a_n,d$, with at least one a_i not equal to 0. Such set is called a *hyperplane*.

(a) Prove that this set of solutions can be parametrized using $n-1$ parameters.

(b) What do you expect the dimension of H to be? (We haven't defined dimension precisely, so use your intuition.)

(c) How many points do you expect need to be given to determine a hyperplane in $\mathbb{R}^n$?

Week 2, Friday

PROBLEM 1. Let $V = \mathbb{R}^2$. For the following pairs of operations, decide whether they make V into a vector space over $\mathbb{R}$. Justify your answer. In what follows, let $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$ and $r \in \mathbb{R}$.

(a)
$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 y_2) \quad \text{and} \quad r \cdot (x_1, x_2) = (rx_1, x_2).$$

(b)
$$(x_1, x_2) + (y_1, y_2) = (2x_1 + 2y_1, 3x_2 + 3y_2) \quad \text{and} \quad r \cdot (x_1, x_2) = (rx_1, rx_2).$$

PROBLEM 2. Here are two templates for showing a subset W of a vector space V over a field F is a subspace:

Proof 1. First note that $\vec{0} \in W$ since __________. Hence, $W \neq \emptyset$. Next, suppose that $u, v \in W$. Then __________. Hence, $u + v \in W$. Now suppose $r \in F$ and $w \in W$. Then __________. Therefore, $r \cdot w \in W$. $\square$

Proof 2. First note that $\vec{0} \in W$ since __________. Hence, $W \neq \emptyset$. Next, suppose that $r \in F$ and $u, v \in W$. Then __________. Hence, $u + r \cdot v \in W$. $\square$

Use one of these two templates for each of the following exercises.

(a) Show that $W = \{(x, y, z) \in \mathbb{R}^3 \mid 2x - y - 3z = 0\}$ is a subspace of $\mathbb{R}^3$.

(b) Show that the set $W = \{f \colon \mathbb{R} \to \mathbb{R} : f(t) = f(-t)\}$ is a subspace of the vector space of real-valued functions of one variable. (You will need to carefully use the definitions given in Example 1.12, p. 84, of the text.)

Week 3, Tuesday

PROBLEM 1. In each of the following:

- Determine whether the given vector v is in the span of the set S.
- If v is in the span of S, then explicitly write v as a linear combination of the vectors in S.

(a) $V = \mathcal{P}_3(\mathbb{Q})$, $v = x^3 + 8x^2 + 7x - 18$,
$S = \{x^3 + 3x - 2, x^3 + 4x^2 - x + 2, x^2 - 2x + 3\}$.

(b) $V = M_{2\times 2}(\mathbb{R})$, $v = \begin{pmatrix} -1 & 3 \\ 2 & 4 \end{pmatrix}$,
$$S = \left\{ \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}, \begin{pmatrix} -1 & 2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 4 \\ 3 & 5 \end{pmatrix} \right\}.$$

Definition. Let S be a subset of a vector space V. We say S *generates* V if $\mathrm{Span}(S) = V$.

PROBLEM 2. Let F be a field and consider $S = \{(1,1,0),(1,0,1),(0,1,1)\} \subseteq F^3$.

(a) Prove that if $F = \mathbb{Q}$, then S generates $\mathbb{Q}^3$.

(b) Prove that if $F = \mathbb{F}_2$, then S does not generate $\mathbb{F}_2^3$.

PROBLEM 3. Let V be a vector space over a field F, and let S_1 and S_2 be two subsets of V.

(a) Prove that $\mathrm{Span}(S_1 \cap S_2) \subseteq \mathrm{Span}(S_1) \cap \mathrm{Span}(S_2)$.

(b) Give an example in which $\mathrm{Span}(S_1 \cap S_2)$ and $\mathrm{Span}(S_1) \cap \mathrm{Span}(S_2)$ are equal, and one in which they are not equal.

Week 3, Friday

Problem 1. Determine whether the following sets are linearly dependent or linearly independent. If they are linearly dependent, find a subset that is linearly independent and has the same span.

(a) $\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}$ in $M_{3\times 2}(\mathbb{Q})$.

(b) $\{(1,3,2),(2,-1,2),(1,2,4)\}$ in $\mathbb{R}^3$.

(c) $\{(1,1,0),(1,0,1),(0,1,1)\}$ in $(\mathbb{F}_2)^3$ (recall that $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$, the field with two elements).

Problem 2. Let V be a vector space over $\mathbb{R}$. Let u and v be distinct vectors in V. Prove that $\{u, v\}$ is linearly independent if and only if $\{u+v, u-v\}$ is linearly independent.

Bonus: Would the same be true over $\mathbb{F}_2$?

Problem 3.

(a) For any field F, we have defined the vector space F^n of n-tuples with components in F. List all elements of (i) F^2 and (ii) F^3 in the case that $F = \mathbb{F}_2$.

(b) Let $S = \{u_1, ..., u_n\}$ be a set of linearly independent vectors in a vector space over $\mathbb{F}_2$. How many elements are in $\text{Span}(S)$? Justify your solution.

Week 4, Friday

Problem 1. Find the coordinates of each given vector v with respect to the ordered basis $B = \langle v_1, \ldots, v_n \rangle$ of V. Show your work.

(a) $v = (4, 1)$, $B = \langle (1, 2), (-2, 3) \rangle$, $V = \mathbb{R}^2$.

(b) $v = (4, 1)$, $B = \langle (1, 0), (0, 1) \rangle$, $V = \mathbb{R}^2$.

(c) $v = x^2 + 2x + 3$, $B = \langle 1, (x-1), (x-1)^2 \rangle$, $V = \mathcal{P}_2(\mathbb{R})$.

(d) $v = x^2 + 2x + 3$, $B = \langle 1, x, x^2, x^3 \rangle$, $V = \mathcal{P}_3(\mathbb{R})$.

Problem 2. Let $X = \{1, 2, 3\}$, and consider the vector space of functions

$$\mathbb{R}^X := \{f \colon X \to \mathbb{R}\}.$$

Recall that for $f, g \in \mathbb{R}^X$ and $r \in \mathbb{R}$, the vector space operations are defined as follows:

$$(f+g)(x) = f(x) + g(x) \quad \text{and} \quad (rf)(x) = r(f(x)).$$

Also recall that in order to prove that $f = g$, one would show that $f(i) = g(i)$, for $i = 1, 2, 3$—that's how one shows they are the same function.

The *zero function* is $z \in \mathbb{R}^X$, defined by $z(1) = z(2) = z(3) = 0$. It's the additive identity for the vector space. Also define the three *characteristic functions*: χ_1, χ_2, χ_3 by

$$\chi_i(j) := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

for $i = 1, 2, 3$. Thus, for instance, $\chi_2(1) = \chi_2(3) = 0$, and $\chi_2(2) = 1$. Define $B = \{\chi_1, \chi_2, \chi_3\}$. Show that B is a basis for $\mathbb{R}^X$ by completing the following steps.

(a) (Warm up) Let $f \in \mathbb{R}^X$ be defined by $f(1) = 5$, $f(2) = \pi$, and $f(3) = -7$. Write f as a linear combination of elements of B.

(b) Let g be an arbitrary element of $\mathbb{R}^X$. Show how to write g as a linear combination of elements of B. (Thus, B spans $\mathbb{R}^X$.)

(c) Show that B is a linearly independent set by proving that if

$$a\chi_1 + b\chi_2 + c\chi_3 = z$$

for some $a, b, c \in \mathbb{R}$, then $a = b = c = 0$.

Problem 3. Let V be a vector space over F and B a basis for V. Let $S \subseteq V$.

(a) Prove that if $B \subsetneq S$, then S is linearly dependent.

(b) Prove that if $S \subsetneq B$, then S does not span V.

Note 1: The first statement can be read as "a basis is a maximal linearly independent set in V". The second statement reads as "a basis is a minimal spanning set for V."

Bonus (ungraded): Prove the converse of both statements. Let B be a subset of V.

(c) Suppose that B is linearly independent and for every S with $B \subsetneq S$, S is linearly dependent. Prove that B is a basis.

(d) Suppose that B spans V and for every S with $S \subsetneq B$, S does not span V. Prove that B is a basis.

Week 5, Tuesday

PROBLEM 1. Let A be an $m \times n$ matrix with i,j-th entry A_{ij}. The *tranpose* of A, denoted A^T, is the $n \times m$ matrix with i,j-th entry A_{ji}: the i-th row of A^T is the i-th column of A. Thus, for example,

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}^T = \begin{pmatrix} a & c & e \\ b & d & f \end{pmatrix}.$$

A matrix A is *skew symmetric* if $A^T = -A$ (notice the minus sign!).

Let W be the set of 3×3 skew symmetric matrices over $\mathbb{R}$.

(a) Prove that W is a subspace of the vector space of all 3×3 matrices over $\mathbb{R}$.

(b) Give a basis for W.

(c) What is $\dim(W)$?

PROBLEM 2. Define the following matrix over the real numbers:

$$M = \begin{pmatrix} -14 & 56 & 40 & 92 \\ 6 & -24 & -17 & -39 \\ 8 & -32 & -23 & -53 \\ -1 & 4 & 3 & 7 \end{pmatrix}.$$

(a) What is the reduced echelon form for M? (You do not need to show your work for this.)

(b) Compute (i) a basis for the row space of M and (ii) a basis for the column space of M *using the algorithm presented in class* on Friday of Week 4. (There is a unique solution if you use the algorithm.)

PROBLEM 3.

(a) Prove that there exists a linear transformation $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ such that $f(2,1) = (0,-1,3)$ and $f(-1,2) = (1,0,-4)$. What is $f(5,10)$?

(b) Is there a linear transformation $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ such that $f(1,2,1) = (2,3)$, $f(3,1,4) = (6,2)$ and $f(7,-1,10) = (10,1)$? Explain your reasoning.

Week 6, Tuesday

PROBLEM 1. For the following functions f:

(i) prove that f is a linear transformation,

(ii) find bases for $\mathcal{N}(f)$ and $\mathcal{R}(f)$, and

(iii) compute the nullity and the rank of f.

(a) $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x, y, z) = (x - y + z, 2y + z)$.

(b) $f\colon \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ defined by $f(p(x)) = x \cdot p(x) + 2p'(x)$.

(Recall that $\mathcal{P}_n(F)$ denotes the vector space of polynomials with coefficients in F of degree less or equal to n. Here, $p'(x)$ denotes the standard derivative from calculus.)

PROBLEM 2. Let V, W and U be finite-dimensional vector spaces over F, and let $f\colon V \to W$, $g\colon W \to U$ be a linear transformations. One can prove[1] that $g \circ f$ is also a linear transformation.

(a) Show that

$$\operatorname{rank}(g \circ f) \leq \min\{\operatorname{rank}(f), \operatorname{rank}(g)\}.$$

(Hint: The rank-nullity theorem is useful for part of this problem.)

(b) Give an example in which the inequality is strict.

PROBLEM 3. Let V and W be vector spaces over F, and let $f\colon V \to W$ be linear an isomorphism (bijective linear transformation). Let $g\colon W \to V$ be the inverse function to f, that is, g satisfies that $g \circ f = \mathrm{id}_V$ and $f \circ g = \mathrm{id}_W$. Prove that g is a linear transformation.

[1] You don't have to turn that in, but it is a good exercise for you to try.

Week 6, Friday

PROBLEM 1. Let $\mathcal{P}_n(\mathbb{R})$ be the vector space of polynomials in x of degree at most n with coefficients in $\mathbb{R}$. Define

$$f\colon \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$$
$$p(x) \mapsto \int_0^x p(t)\,dt.$$

Note that this is a definite integral, so there is no constant of integration! One can show f is a linear transformation (but you don't need to do that for this problem).

(a) Find the matrix representing f with respect to the ordered bases $\langle 1, x, x^2\rangle$ for $\mathcal{P}_2(\mathbb{R})$ and $\langle 1, x, x^2, x^3\rangle$ for $\mathcal{P}_3(\mathbb{R})$.

(b) Find the matrix representing f with respect to the ordered bases $\langle 1+x+x^2, x+x^2, x^2\rangle$ for $\mathcal{P}_2(\mathbb{R})$ and $\langle 1+x+x^2+x^3, x+x^2+x^3, x^2+x^3, x^3\rangle$ for $\mathcal{P}_3(\mathbb{R})$.

PROBLEM 2. Let V be a finite-dimensional vector space over F, and let $f\colon V \to V$ be a linear transformation. Prove that f is one-to-one if and only if it is onto.

PROBLEM 3. Let $\mathcal{P}(\mathbb{R})$ be the vector space of polynomials in x with coefficients in $\mathbb{R}$. Define

$$f\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$$
$$p(x) \mapsto \int_0^x p(t)\,dt.$$

and

$$g\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$$
$$p(x) \mapsto p'(x).$$

Again, one can show that f and g are linear transformations, but you don't have to do that here.

(a) Prove that f is one-to-one, but not onto.

(b) Prove that g is onto, but not one-to-one.

(c) What can you say about $f \circ g$ and $g \circ f$?

NOTE: Contrast the situation in this problem with that in problem 2.

Week 7, Tuesday

Problem 1. Recall that if $P, Q \in M_{m\times n}(F)$, then their sum $P+Q \in M_{m\times n}(F)$ is defined to be the matrix with i,j-th entry

$$(P+Q)_{ij} := P_{ij} + Q_{ij},$$

On the other hand, if $P \in M_{m\times \ell}(F)$ and $Q \in M_{\ell\times n}(F)$ (note the change in dimensions), then their product $PQ \in M_{m\times n}(F)$ is defined to be the matrix whose i,j-th entry is

$$(PQ)_{ij} := \sum_{k=1}^{\ell} P_{ik}Q_{kj}.$$

Let $A \in M_{m\times \ell}(F)$, and let $B, C \in M_{\ell\times n}(F)$. Using only the definition of matrix addition and matrix multiplication, prove that

$$A(B+C) = AB + AC.$$

You do this by proving that the i,j-th entries on both sides are equal. **Please use summation notation, and be careful to specify the correct starting and ending points for the summation.** (The result you are proving is called the *left distributivity* property of matrix multiplication.)

Problem 2. For the following, recall that just as an example usually does not constitute a proof that something is true, a general discussion does not usually suffice to proof that something is not true. In the latter case, it is fine (but not necessary) to give a general discussion, but in the end, you should provide a concrete and simple counterexample.

(a) Prove that matrix multiplication of 2×2 matrices does not satisfy the commutative law, $AB = BA$.

(b) Prove that matrix multiplication of 2×2 matrices does not satisfy left cancellation.
Cancellation: If $AB = AC$ and A is not the zero matrix, then $B = C$.

Problem 3. Let $\mathcal{P}_n(\mathbb{R})$ be the vector space of polynomials in x of degree at most n with coefficients in $\mathbb{R}$. Let $f\colon \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ and $g\colon \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ be the linear transformations respectively defined as

$$f(p(x)) = (3+x)p'(x) + 2p(x) \quad \text{and} \quad g(a+bx+cx^2) = (a+b, c, a-b).$$

Let $\mathcal{B} = \langle 1, x, x^2\rangle$ and $\mathcal{D}$ be the standard ordered basis for $\mathbb{R}^3$.

(a) Compute the matrix representing f with respect to the basis $\mathcal{B}$ for both the domain and codomain.

(b) Is f one-to-one? Is it onto?

(c) Compute the matrix representing g with respect to the bases $\mathcal{B}$ and $\mathcal{D}$.

(d) Compute the matrix representing $g \circ f$ with respect to the bases $\mathcal{B}$ and $\mathcal{D}$. Then use Theorem 2.7 (Chapter Three, Section IV) to verify your result. (This theorem says the composition of linear maps is represented by the matrix product of the representatives of the linear maps.)

Week 7, Friday

Problem 1. The *trace* of an $n \times n$ matrix A is the sum of its diagonal elements:

$$\operatorname{tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

(a) If A and B are $n \times n$ matrices, prove that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. (Use the definition of matrix multiplication and summation notation in your proof.)

(b) If P is an invertible $n \times n$ matrix, prove that $\operatorname{tr}(PAP^{-1}) = \operatorname{tr}(A)$.

(c) Consider the following ordered basis for $\mathcal{M}_{2\times 2}(F)$:

$$\alpha = \left\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\rangle.$$

The trace defines a function $\operatorname{tr}\colon \mathcal{M}_{2\times 2}(F) \to F$, and it is not hard to check that it is a linear transformation of vector spaces over F (you don't have to prove that here, although it is a good exercise to do). Compute the matrix representing the trace function $\operatorname{tr}\colon \mathcal{M}_{2\times 2}(F) \to F$ with respect to α for the domain and with respect to the basis $\{1\}$ for the codomain.

Problem 2. Let

$$A = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & -3 \\ 4 & 1 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 & 4 \\ -1 & -2 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix}.$$

Compute, if possible, the following. If it is not possible, explain why.

(a) AB,

(b) $A(2B + 3C)$,

(c) $(AB)D$,

(d) $A(BD)$,

(e) AD.

Problem 3. Let V be a vector space over a field F. Recall that the identity function $\operatorname{id}_V\colon V \to V$ is given by $\operatorname{id}_V(v) = v$ for all $v \in V$. This function is linear (if you are not convinced, prove it, but you do not have to turn in that proof).

(a) Let V be a vector space of dimension n and let $\mathcal{B}$ be an ordered basis for V. Show that the matrix representing id_V with respect to the basis $\mathcal{B}$ for both the domain and the codomain is I_n (the $n \times n$ identity matrix).

(b) Let V and W be vector spaces of dimension n and let $f\colon V \to W$ be an isomorphism with inverse $f^{-1}\colon W \to V$. Let $\mathcal{B}$ and $\mathcal{D}$ be ordered bases for V and W, respectively. If A is the matrix representing f with respect to the bases $\mathcal{B}$ and $\mathcal{D}$, what is the matrix for f^{-1} with respect to the bases $\mathcal{D}$ and $\mathcal{B}$? Justify your answer.

(c) Consider the linear transformation $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x, y) = (3x + y, -x + 4y)$. Using part (b), find the inverse of f.

Week 8, Tuesday

PROBLEM 1. For each of the following matrices, use the algorithm from class to determine whether they have inverses, and if so, find the inverse. Show your work (i.e., the row reduction). (*Pointer:* as with many linear algebra problems, it's easy to make arithmetic mistakes, but it's also easy to check your answer!)

$$\text{(a)} \quad \begin{pmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{pmatrix} \qquad\qquad \text{(b)} \quad \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & -1 \\ 1 & 5 & 4 \end{pmatrix}.$$

PROBLEM 2. Let A and B be $n \times n$ matrices such that AB is invertible.

(a) Prove that A and B are invertible. *Hint:* Use rank.

(b) Give an example to show that A and B of arbitrary dimensions need not be invertible if AB is invertible.

PROBLEM 3. Given m in $\mathbb{R}$, consider the line L in $\mathbb{R}^2$ given by those points (x, y) that satisfy $y = mx$. (If you recall from high school, this is the line through the origin of slope m.) Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection of $\mathbb{R}^2$ about L. Geometrically, $f(x, y)$ is the point obtained by taking the mirror image of (x, y) across L, so that the line segment connecting (x, y) and $f(x, y)$ is bisected perpendicularly by L.

One can prove geometrically (but you don't have to do that here), that f is a linear transformation. The goal of this problem is to find a closed formula for f (without having to use any crazy trigonometry).

(a) What are $f(1, m)$ and $f(m, -1)$?

(Hint: note that $(m, -1)$ is perpendicular to L, you don't have to prove this for the homework, although you might want to figure out why.)

(b) Prove that $\{(1, m), (m, -1)\}$ is a basis for $\mathbb{R}^2$.

(c) Compute the matrix for f with respect to the ordered basis $\langle (1, m), (m, -1) \rangle$ for the domain and the codomain. Then use the change of basis result to compute the matrix for f with respect to the standard basis for the domain and the codomain. Conclude by giving a closed formula for f.

(d) Explain why your result makes sense in the cases $m = 0$ and $m = 1$.

Week 8, Friday

PROBLEM 1. Compute the determinant of the following matrices by using row operations.

(a) $\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 3 & -1 & 2 \\ 2 & 4 & 7 & -3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix}$

(c) $\begin{pmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix}$

(d) (BONUS) Generalize part (c) for the $n \times n$ matrix $\begin{pmatrix} n & -1 & \cdots & -1 \\ -1 & n & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n \end{pmatrix}$, with n in the main diagonal and -1 everywhere else.

PROBLEM 2. Let V be a vector space. For each integer $r > 0$, we now give a provisional definition of a new vector space called $\bigwedge^r V$. A spanning set for $\bigwedge^r V$ consists of expressions of the form $v_1 \wedge \cdots \wedge v_r$ where $v_1, \ldots, v_r \in V$. For example, if $u, v, w \in V$, the following would be typical elements of $\bigwedge^2 V$:

$$\begin{aligned} \omega &= 4\, u \wedge v - 5\, u \wedge w + 7\, v \wedge w \\ \mu &= 2\, u \wedge v + 9\, u \wedge w + 6\, v \wedge w. \end{aligned}$$

Addition is done by combining like terms, and scaling is done by scaling each term. For instance, continuing the example above, we get

$$\begin{aligned} \omega + \mu &= 6\, u \wedge v + 4\, u \wedge w + 13\, v \wedge w \\ 5\,\omega &= 20\, u \wedge v - 25\, u \wedge w + 35\, v \wedge w. \end{aligned}$$

We now add a couple of rules. First, these "wedge products" of vectors are linear in each component. For $v_i \in V$ and $a \in F$,

$$v_1 \wedge \cdots \wedge v_{i-1} \wedge (av_i + v_i') \wedge v_{i+1} \wedge \dots v_r =$$

$$a\, v_1 \wedge \cdots \wedge v_{i-1} \wedge v_i \wedge v_{i+1} \wedge \cdots \wedge v_r$$
$$+$$
$$v_1 \wedge \cdots \wedge v_{i-1} \wedge v_i' \wedge v_{i+1} \wedge \cdots \wedge v_r.$$

Second, we declare that $v_1 \wedge \cdots \wedge v_r = 0$ if $v_i = v_j$ for some $i \neq j$. To illustrate these rules in action suppose $u, v, w \in V$. Then in $\bigwedge^3 V$, we have the following:

$$\begin{aligned} u \wedge (2u + 3v + 5w) \wedge w &= u \wedge (2u) \wedge w + u \wedge (3v) \wedge w + u \wedge (5w) \wedge w \\ &= 2\, u \wedge u \wedge w + 3\, u \wedge v \wedge w + 5\, u \wedge w \wedge w \\ &= 0 + 3\, u \wedge v \wedge w + 0 \\ &= 3\, u \wedge v \wedge w. \end{aligned}$$

Another example, this time in $\bigwedge^2 V$:

$$\begin{aligned} (u + 2v) \wedge (u + 3v) &= u \wedge (u + 3v) + (2v) \wedge (u + 3v) \\ &= u \wedge u + u \wedge (3v) + (2v) \wedge u + (2v) \wedge (3v) \\ &= 0 + 3\, u \wedge v + 2\, v \wedge u + 6\, v \wedge v \\ &= 3\, u \wedge v + 2\, v \wedge u + 0 \\ &= 3\, u \wedge v + 2\, v \wedge u. \end{aligned}$$

It turns out there is a little more we can do to simplify this last example. By the second rule, we have $(u+v) \wedge (u+v) = 0$, since in this expression we have two copies of the same vector. But linearly expanding this expression, we get

$$\begin{aligned} 0 &= (u + v) \wedge (u + v) \\ &= u \wedge (u + v) + v \wedge (u + v) \\ &= u \wedge u + u \wedge v + v \wedge u + v \wedge v \\ &= 0 + u \wedge v + v \wedge u + 0 \\ &= u \wedge v + v \wedge u. \end{aligned}$$

Thus, $u \wedge v + v \wedge u = 0$. This means that

$$u \wedge v = -v \wedge u.$$

In fact, in a wedge product of vectors, swapping any two locations negates the expression. (The proof is similar to the one we just gave in the case of $r = 2$.) For instance,

$$u \wedge v \wedge w = -u \wedge w \wedge v = w \wedge u \wedge v = -w \wedge v \wedge u.$$

Continuing our example from above, we get

$$\begin{aligned}(u + 2v) \wedge (u + 3v) &= \dots \quad \text{(see earlier calculation)}\\ &= 3\,u \wedge v + 2\,v \wedge u\\ &= 3\,u \wedge v - 2\,u \wedge v\\ &= u \wedge v.\end{aligned}$$

Now for some problems:

(a) Let $V = \mathbb{R}^2$, and take two vectors $u = (a, b)$ and $v = (c, d)$ in $\mathbb{R}^2$. Let $e_1 = (1, 0)$ and $e_2 = (0, 1)$. Writing u and v as linear combinations of e_1 and e_2, find the number k in terms of a, b, c, d such that

$$u \wedge v = k\, e_1 \wedge e_2,$$

in $\bigwedge^2 V$. What is the relation between k and $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$?

(b) Now let $V = \mathbb{R}^3$, and take vectors $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$ in $\mathbb{R}^3$. Writing these vectors as linear combinations of the standard basis vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$, find numbers p, q, r in terms of the u_i and the v_i such that

$$u \wedge v = p\, e_2 \wedge e_3 - q\, e_1 \wedge e_3 + r\, e_1 \wedge e_2.$$

(Watch out for the minus sign in front of q.) Physics students may note a relation with the cross product of two vectors in $\mathbb{R}^3$.

Week 9, Friday

Problem 1. Let

$$A = \begin{pmatrix} 1 & -1 & 2 & -2 \\ -2 & 2 & 2 & 4 \end{pmatrix}.$$

Find elementary matrices $E_1, \ldots, E_\ell$ such that $E_\ell \cdots E_2 E_1 A$ is the reduced echelon form of A. (Check your work.)

Problem 2. This exercise will prove that the determinant is multiplicative, that is, for $n \times n$ matrices A and B,

$$\det(AB) = \det(A)\det(B).$$

Let B be a fixed $n \times n$ matrix over F such that $\det(B) \neq 0$. Consider the function

$$d\colon M_{n\times n}(F) \longrightarrow F$$

defined by $d(A) = \det(AB)/\det(B)$. You will prove that $d(A) = \det(A)$. For a matrix A, we write $(r_1, \ldots, r_n)$ for the rows of A, with each $r_i \in F^n$.

(a) Prove that d is multilinear on rows, that is, d satisfies that

$$d(r_1, \ldots, r_i + k \cdot r_i', \ldots, r_n) = d(r_1, \ldots, r_i, \ldots, r_n) + kd(r_1, \ldots, r_i', \ldots, r_n)$$

for all $r_1, \ldots, r_n, r_i' \in F^n$ and any $k \in F$.

(Some suggested notation to help in your proof: let $c_1, \ldots, c_n$ be the columns of B. Then $(AB)_{s,t} = r_s \cdot c_t$, i.e., the s,t-entry of AB is the dot product of the s-th row of A with the t-th column of B. Recall that the dot product is defined by $(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = x_1y_1 + \cdots + x_ny_n$. Letting A' be the matrix with rows $(r_1, \ldots, r_i', \ldots, r_n)$ and A'' the matrix with rows $(r_1, \ldots, r_i + kr_i', \ldots, r_n)$, you will need compare the rows of AB, $A'B$ and $A''B$.)

(b) Prove that d is alternating on rows, that is, d satisfies that $d(r_1, \ldots, r_n) = 0$ if $r_i = r_j$ for some $i \neq j$.

(c) Prove that $d(I_n) = 1$.

(d) Deduce that for all A, we have that $d(A) = \det(A)$, and that $\det(AB) = \det(A)\det(B)$.

(e) Prove that $\det(AB) = \det(A)\det(B)$ when $\det(B) = 0$. (*Hint:* You have basically proved this in an earlier homework.)

Week 10, Tuesday

PROBLEM 1. Compute the determinants of the following matrices by using the permutation expansion.

(a) $\begin{pmatrix} -1 & 2+i & 3 \\ 1-i & i & 1 \\ 3i & 2 & -1+i \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 4 & 0 \\ 5 & 0 & 0 & 0 & 0 \end{pmatrix}$

PROBLEM 2. Compute the determinants of the same matrices as in Problem 1 by using the Laplace expansion along any row or column. Be clear about which row or column you are using.

PROBLEM 3. Let $p(x_1, \ldots, x_n)$ be a polynomial on n variables with coefficients in a field F. An arbitrary term of this polynomial is of the form $ax_1^{d_1}x_2^{d_2}\ldots x_n^{d_n}$, where $a \in F$ and d_i is a nonnegative integer for all i. The total degree of this term is $d_1 + \cdots + d_n$.

For example, the polynomial

$$p(x_1, x_2, x_3) = 2 + x_1 + 3x_1x_2^2 - 4x_2^2x_3^3 + 9x_1x_2x_3$$

has five terms of total degree 0, 1, 3, 5, and 3, respectively.

Here is a result from polynomial algebra. If p satisfies the conditions:

(i) $p(x_1, \ldots, x_n) = 0$ whenever $x_i = x_j$ for $i \neq j$;

(ii) the total degree of every term is $n(n-1)/2$,

then

$$p(x_1, \ldots, x_n) = k(x_2 - x_1)(x_3 - x_1)\cdots(x_n - x_{n-1})$$

for some $k \in F$. Here the product contains all the terms of the form $x_j - x_i$ with $1 \leq i < j \leq n$. Note that the coefficient k is equal to the coefficient of $x_2x_3^2\cdots x_{n-1}^{n-2}x_n^{n-1}$ in p.

For example, when $n = 2$ and $F = \mathbb{R}$, $p(x_1, x_2) = x_1 - x_2$ satisfies both properties, $p(x_1, x_2) = x_1^2 - x_2^2$ satisfies (i) but not (ii), and $p(x_1, x_2) = x_1 + 2x_2$ satisfies (ii) but not (i).

Now consider the Vandermonde matrix

$$V(x_1, \ldots, x_n) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{pmatrix}$$

Let $p(x_1, \ldots, x_n) = \det(V(x_1, \ldots, x_n))$.

(a) Using properties of determinants, prove that p satisfies property (i).

(b) Using the permutation expansion of the determinant, prove that p satisfies property (ii).

(*Hint:* As always, it is useful to play around with small cases of n to understand what is really going on. Try $n = 2$ and $n = 3$, and then use what you learn to argue for the general case. When you write your solutions, you should write them for arbitrary n.)

(c) It follows from (a) and (b) and the discussion above that

$$p(x_1, \ldots, x_n) = k(x_2 - x_1)(x_3 - x_1) \cdots (x_n - x_{n-1})$$

for some $k \in F$. Find the value of the coefficient k.

(d) (BONUS.) A general polynomial of degree d in one variable over the real numbers has the form

$$q(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d,$$

where the a_i are real numbers. Pick n distinct real numbers $x_1, \ldots, x_n$, and pick arbitrary (not necessarily distinct) real numbers $b_1, \ldots, b_n$. Prove that there is a unique polynomial $q(x)$ of degree $n-1$ over the real numbers such that $q(x_i) = b_i$ for $i = 1, \ldots, n$.

(e) (BONUS.) Use the Vandermonde determinant to prove that the collection of functions $\{e^{\alpha x} : \alpha \in \mathbb{R}\}$ is linearly independent. (Recall that the set of functions from $\mathbb{R}$ to $\mathbb{R}$ is a vector space. The solution to this problem would show that this space is infinite dimensional.)

Week 10, Friday

PROBLEM 1. For each of the following matrices $A \in M_{n\times n}(F)$

(i) Determine all eigenvalues of A.

(ii) For each eigenvalue λ of A, find the set of eigenvectors corresponding to λ.

(iii) If possible, find a basis for F^n consisting of eigenvectors of A.

(iv) If successful in finding such a basis, determine an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

(a) $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$ for $F = \mathbb{R}$.

(b) $A = \begin{pmatrix} 0 & -2 & -3 \\ -1 & 1 & -1 \\ 2 & 2 & 5 \end{pmatrix}$ for $F = \mathbb{R}$.

(c) $A = \begin{pmatrix} 7 & -5 \\ 10 & -7 \end{pmatrix}$ for $F = \mathbb{R}$.

(d) $A = \begin{pmatrix} 7 & -5 \\ 10 & -7 \end{pmatrix}$ for $F = \mathbb{C}$.

(e) $A = \begin{pmatrix} 2 & 0 & -1 \\ 4 & 1 & -4 \\ 2 & 0 & -1 \end{pmatrix}$ for $F = \mathbb{R}$.

PROBLEM 2. Let $f\colon V \to V$ be a linear transformation. For a positive integer m, we define f^m inductively as $f \circ f^{m-1}$. Prove that if λ is an eigenvalue for f, then λ^m is an eigenvalue for f^m.

PROBLEM 3. Let $T\colon M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ defined as $T(A) = A^t$ (taking the transpose). One can prove that T is a linear transformation.

(a) Show that the only eigenvalues of T are 1 and -1. (*Hint:* Problem 2 might help.)

(b) For $n = 2$, describe the eigenvectors corresponding to each eigenvalue.

(c) Find an ordered basis $\mathcal{B}$ for $M_{2\times 2}(\mathbb{R})$ such that the matrix that represents T with respect to $\mathcal{B}$ is diagonal.

Week 11, Tuesday

REMINDER: You have a presentation proposal due next Tuesday, November 23. Please start thinking about the topic of your presentation: an application of linear algebra to other fields.

PROBLEM 1. Let $V = \mathcal{P}_3(\mathbb{R})$, the vector space over $\mathbb{R}$ consisting of all polynomials with real coefficients having degree at most 3. Define the following linear transformation on V (in which the prime denotes differentiation),

$$\begin{aligned} L\colon V &\to V \\ f &\mapsto xf' + f'. \end{aligned}$$

(a) Write the matrix of L with respect to the ordered basis $\langle 1, x, x^2, x^3 \rangle$ of V.

(b) What are the eigenvalues of L?

(c) Does V have a basis of eigenvectors of L? If so, give such a basis (**written as polynomials** not tuples of real numbers), and if not, explain why not.

PROBLEM 2. Consider the matrix

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

(a) What are the algebraic and geometric multiplicities of each of the eigenvalues of B?

(b) Explain whether B is diagonalizable in terms of the geometric multiplicities of its eigenvalues.

PROBLEM 3. Consider an $n \times n$ matrix A over $\mathbb{C}$. As mentioned in class, the characteristic polynomial of A is of the form

$$p_A(t) = \det(A - tI_n) = (-1)^n t^n + b_{n-1}t^{n-1} + \cdots + b_1 t + b_0.$$

(a) Prove that b_0 is equal to the determinant of A.

(b) Prove that $b_{n-1} = (-1)^{n-1}\operatorname{tr}(A)$, where tr denotes the trace.

(HINT: Using the permutation expansion identify where the possible terms of degree $n-1$ arise from.)

(c) Prove that the product of all eigenvalues of A (with multiplicity) is equal to $\det(A)$. (You may use the fact the characteristic polynomial will factor completely into linear factors, $p_A(t) = (-1)^n \prod_{i=1}^n (t - \lambda_i)$, since we are working over $\mathbb{C}$.)

(d) Prove that the sum of all eigenvalues of A (with multiplicity) is equal to $\operatorname{tr}(A)$.

NOTE: When $n = 2$, this result says that

$$p_A(t) = t^2 - \operatorname{tr}(A)t + \det(A).$$

Week 11, Friday

Problem 1. Consider a sequence of numbers p_n defined recursively by fixing constants a and b, next assigning initial values for p_0 and p_1, and then for $n \geq 1$ letting

$$p_{n+1} = ap_n + bp_{n-1}.$$

For instance, letting $a = 2$, $b = -1$, $p_0 = 0$, and $p_1 = 1$, we get

$$\begin{aligned} p_0 &= 0 \\ p_1 &= 1 \\ p_{n+1} &= 2p_n - p_{n-1} \quad \text{for } n \geq 1, \end{aligned}$$

which defines the sequence

$$0, 1, 2, 3, 4, 5, 6, \ldots$$

Given any sequence of this form, we can encode the recursive relation in the following matrix equation:

$$\begin{pmatrix} p_{n+1} \\ p_n \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_n \\ p_{n-1} \end{pmatrix}. \tag{38.1}$$

So we have

$$\begin{pmatrix} p_2 \\ p_1 \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_0 \end{pmatrix},$$

which implies

$$\begin{pmatrix} p_3 \\ p_2 \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_2 \\ p_1 \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix} \left[\begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_0 \end{pmatrix} \right] = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix}^2 \begin{pmatrix} p_1 \\ p_0 \end{pmatrix},$$

and so on. In general, we have

$$\begin{pmatrix} p_{n+1} \\ p_n \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} p_1 \\ p_0 \end{pmatrix}. \tag{38.2}$$

Let

$$A = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix}.$$

and suppose A is diagonalizable. Take P so that

$$P^{-1}AP = D = \operatorname{diag}(\lambda_1, \lambda_2).$$

We have seen that it follows that $A^n = PD^nP^{-1}$, so that equation (38.2) becomes

$$\begin{pmatrix} p_{n+1} \\ p_n \end{pmatrix} = PD^nP^{-1} \begin{pmatrix} p_1 \\ p_0 \end{pmatrix} = P \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} P^{-1} \begin{pmatrix} p_1 \\ p_0 \end{pmatrix}.$$

Thus, we get a closed form expression for p_n in terms of powers of the eigenvalues of A (just take the second component of the product on the right-hand side of the above equation).

Let $a = b = 1$, $p_0 = 0$, and $p_1 = 1$.

(a) Write out the first several values for the sequence (p_n).

(b) Write the corresponding matrix equation, as (38.1) above.

(c) Diagonalize the matrix A and compute the corresponding equation for p_n in terms of powers of the eigenvalues of A.

You may find the following notation useful:

$$\phi = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \overline{\phi} = \frac{1-\sqrt{5}}{2},$$

with useful relations $\phi\overline{\phi} = -1$, $\phi^2 = \phi + 1$, and $\phi + \overline{\phi} = 1$. (**Warning:** You will want to make sure you get the diagonalization perfect. This will take some time. Using the above notation as much as possible will help.)

(d) As in the first example above, let $a = 2$, $b = -1$, $p_0 = 0$ and $p_1 = 1$. What happens when you try to use the method above to find a closed formula for p_n?

Problem 2. One of the reasons we like diagonalization is because computing the powers of the matrix is easy if it is diagonalized (see previous exercise). In this exercise we explore the powers of Jordan blocks. Recall that a 2×2 Jordan block is a matrix of the form

$$J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix},$$

for some $\lambda \in F$.

(a) For J as above, compute J^2, J^3 and J^4. For a natural number n, make a conjecture for the value of J^n.

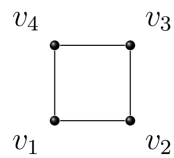
(b) Prove your conjecture. (*Hint:* You probably want to use induction.)

(c) (BONUS:) Repeat (a) and (b) above for a 3×3 Jordan block:

$$\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.$$

Week 12, Tuesday

PROBLEM 1. Consider the *cycle graph* C_4:



(a) Find the adjacency matrix $A = A(G)$.

(b) Compute A^4 and use it to determine the number of walks from v_1 to v_3 of length 4. List all of these walks (these will be ordered lists of 5 vertices).

(c) What is the total number of *closed* walks of length 4?

(d) Compute and factor the characteristic polynomial for A.

(e) What are the algebraic multiplicities of each of the eigenvalues?

(f) Diagonalize A using our algorithm: compute bases for the eigenspaces of each of the eigenvalues you just found, and use them to construct a matrix P such that $P^{-1}AP$ is a diagonal matrix with the eigenvalues along the diagonal.

(g) Use part (f) to find a closed expression for A^ℓ for each $\ell \geq 1$.

(h) Take the trace of A^ℓ to get a formula for the number of closed walks of length ℓ for each $\ell \geq 1$. (You can check your result against the formula given in class.)

PROBLEM 2. In this exercise we will prove the theorem from class:
"Let A be the adjacency matrix for a graph G with vertices $v_1, \ldots, v_n$, and let $\ell \in \mathbb{Z}_{\geq 0}$. Then then number of walks of length ℓ from v_i to v_j is $(A^\ell)_{ij}$."

(a) Let $p(i, j, \ell)$ denote the number of walks of length ℓ in G from v_i to v_j. Prove that for all $i, j = 1, \ldots, n$ and $\ell \geq 1$,

$$p(i, j, \ell) = \sum_{k=1}^{n} p(i, k, \ell - 1)p(k, j, 1).$$

(*Hint:* Part of the trick is to parse this formula appropriately.)

(b) Prove the theorem by induction on ℓ, using the result from part (a).

Week 13, Tuesday

Problem 1. Consider the inner product space $(\mathbb{R}^2, \langle\, , \,\rangle)$ presented as an example in class, where the inner product is defined as

$$\langle (x_1, x_2), (y_1, y_2) \rangle = 3x_1y_1 + 2x_1y_2 + 2x_2y_1 + 4x_2y_2.$$

(a) Compute the lengths of the vectors $(1,1)$ and $(1,-1)$.

(b) Compute the cosine of the angle between $(1,1)$ and $(1,-1)$. Are these vectors perpendicular?

(c) Find a non-zero vector perpendicular to $(1,1)$.

Problem 2. Recall the inner product defined on $M_{m\times n}(F)$, where $F = \mathbb{R}$ or $\mathbb{C}$: for $A, B \in M_{m\times n}(F)$, we define

$$\langle A, B \rangle = \operatorname{tr}(B^* A),$$

where $B^* = \overline{B^t}$ is the conjugate transpose. In this problem we will verify that this function does indeed satisfy the axioms of an inner product.

(a) Prove that this function is linear in the first coordinate: for all $A, B, C \in M_{m\times n}(F)$ and $r \in F$,

$$\langle A + rC, B \rangle = \langle A, B \rangle + r\langle C, B \rangle.$$

(b) Prove that this function is conjugate symmetric: for all $A, B \in M_{m\times n}(F)$,

$$\langle A, B \rangle = \overline{\langle B, A \rangle}.$$

(c) Prove that this function is positive-definite: for all $A \in M_{m\times n}(F)$ with $A \neq 0$,

$$\langle A, A \rangle > 0.$$

Problem 3. Let $(V, \langle\, , \,\rangle)$ be an inner product space over $\mathbb{R}$, and let $v, w \in V$ be nonzero vectors.

(a) Prove that if the vector

$$\frac{v}{\|v\|} + \frac{w}{\|w\|}$$

is nonzero, then it bisects the angle between v and w.

(b) Illustrate in $\mathbb{R}^2$ with the standard dot product.

Week 13, Friday

PROBLEM 1. Let $S = \{(1, 0, i), (1, 2, 1)\}$ in $\mathbb{C}^3$ (with the standard inner product). Compute $S^\perp$.

PROBLEM 2. Let $V = \mathcal{P}(\mathbb{R})$ be the vector space of all polynomials with coefficients in $\mathbb{R}$, with inner product $\langle f(x), g(x)\rangle = \int_0^1 f(t)g(t)\,dt$. Let $W = \mathcal{P}_1(\mathbb{R})$. (Warning: to get this problem right, you will need to be very careful with your calculations and double-check your solutions.)

(a) Find an orthonormal basis $\{u_1, u_2\}$ for W.

(b) Find the closest polynomial in W to $h(x) = 3 - 2x + x^2$. Express your solution in two forms: (i) as a linear combination of u_1 and u_2, and (ii) as a linear combination of 1 and x.

PROBLEM 3. Let V be an n-dimensional vector space over $F = \mathbb{R}$ or $\mathbb{C}$, and let $\langle\,,\,\rangle$ be an inner product on V. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be an ordered basis for V (not necessarily orthonormal). Let A be the $n \times n$ matrix given by

$$A_{ij} = \langle v_j, v_i\rangle.$$

Recall that for $x \in V$, $[x]_\mathcal{B}$ denotes the coordinate vector for x with respect to the basis $\mathcal{B}$, and as usual, we will think of this vector in F^n as an $n \times 1$ matrix.

(a) Prove that for all $x, y \in V$,

$$\langle x, y\rangle = ([y]_\mathcal{B})^* A\,([x]_\mathcal{B})\,.$$

(Recall that for a matrix C, we define $\overline{C}$ by $\overline{C}_{ij} = \overline{(C_{ij})}$, and then we define the conjugate transpose by $C^* = \overline{C^t}$. *Hint*: compute both sides using sum notation. On the right-hand side, you will be computing the $(1,1)$-entry of a 1×1 matrix.)

(b) Prove that the matrix A satisfies $A = A^*$.

(c) If the basis $\mathcal{B}$ is orthonormal, what is the matrix A?

(d) (Bonus) Let $\mathcal{D}$ be another ordered basis for V, and let C be the associated $n \times n$ matrix. How are A and C related?