Systems of Linear Equations

The Punch Line: We can solve systems of linear equations by manipulating a matrix that represents the system.

Warm-Up: Which of these equations are linear?

(a) $y = mx + b$, with x and y as variables

(b) $(y - y_0) + 4(x - x_0) = 0$, with x and y as variables

(c) $4x + 2y - 9z = 12$, with x, y, and z as variables

(d) $x_1^2 + x_2^2 = 1$, with x_1 and x_2 as variables

(e) $a^2x + 3b^3y = 6$, with x and y as variables

(f) $a^2x + 3b^3y = 6$, with a and b as variables

(a) This system is linear, because it can be written $-mx + y = b$.

(b) This system is linear, because it can be written $4x + y = (y_0 + 4x_0)$.

(c) This system is linear, because it is already in the standard form.

(d) This system is not linear, because it involves squared variables, so it can't be put into the standard form.

(e) This system is linear with these variables, because it is already in standard form (it is okay for coefficients to be squared).

(f) It is not linear with these variables, because it can't be written in the standard form (the variables have operations done to them that aren't simply scaling or shifting).

The Setup: When we have a linear system of equations, we can make an *augmented matrix* representing the system by arranging the coefficients on the left side of the matrix (keeping them in the same order for each equation, and writing a 0 whenever a variable is missing from one of the equations), and the constants on the right side. This is often easiest to do when each equation is written in the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, so it can help to rewrite equations like this if they are given to you differently.

1. Write down an augmented matrix representing these linear systems.

(a) The system for x_1 and x_2 given by

$$\begin{aligned} x_1 + x_2 &= 4 \\ x_1 - 2x_2 &= 1 \end{aligned}$$

(b) The system for x, y, and z given by

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

(c) The system for x and y given by

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

(a) $\left[\begin{array}{cc|c} 1 & 1 & 4 \\ 1 & -2 & 1 \end{array}\right]$

(b) $\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 1 & 1 & 0 & 2 \\ 0 & 1 & -1 & 1 \end{array}\right]$

(c) $\left[\begin{array}{cc|c} 1 & 1 & 4 \\ 1 & -2 & 1 \end{array}\right]$

2. Write down a linear system of equations represented by these augmented matrices.

(a) $\left[\begin{array}{cc|c} 1 & 2 & 0 \\ -3 & 1 & 0 \end{array}\right]$

(b) $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & -2 & 1 & 0 \\ 1 & 0 & -1 & 0 \end{array}\right]$

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

(a) The system in x_1 and x_2 given by $\begin{aligned} x_1 + 2x_2 &= 0 \\ x_2 - 3x_1 &= 0 \end{aligned}$

(b) The system in x_1, x_2, and x_3 given by $\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - 2x_2 + x_3 &= 0 \\ x_1 - x_3 &= 0 \end{aligned}$

(c) The system in x_1, x_2, and x_3 given by $\begin{aligned} x_1 + 2x_2 - 4x_3 &= -1 \\ 0 &= 0 \end{aligned}$ (the second equation may be omitted)

The Execution: Once we have a matrix representing a linear system of equations, we can use *elementary row operations* on the matrix to find equivalent system of equations. These operations are

1) Replacement: Replace a row with itself plus a multiple of a different row,

2) Interchange: Switch the order of two rows,

3) Scaling: Multiply everything in the row by the same constant (other than 0).

The goal is to use these three operations to find an equivalent system of equations that is easier to solve.

3. Solve each of these linear systems of equations.

(a) The two variable system given by

$$\begin{aligned} x_1 + x_2 &= 4 \\ x_1 - 2x_2 &= 1 \end{aligned}$$

(b) The three variable system given by

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

(c) The three variable system given by

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - 2x_2 + x_3 &= 0 \\ x_1 - x_3 &= 0 \end{aligned}$$

(a) We know from part **1.** that the augmented matrix of this system is $\left[\begin{array}{cc|c} 1 & 1 & 4 \\ 1 & -2 & 1 \end{array}\right]$. If we replace Row 2 with itself minus Row 1 (that is, itself plus -1 times Row 1), we get the matrix $\left[\begin{array}{cc|c} 1 & 1 & 4 \\ 0 & -3 & -3 \end{array}\right]$. Then, we can scale Row 2 by $\frac{-1}{3}$ to get the matrix $\left[\begin{array}{cc|c} 1 & 1 & 4 \\ 0 & 1 & 1 \end{array}\right]$. Finally, we can replace Row 1 with itself minus Row 2 to get $\left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 1 \end{array}\right]$. This represents the linear system of equations $\begin{aligned} x_1 &= 3 \\ x_2 &= 1 \end{aligned}$, which we can clearly see has the (unique) solution $s_1 = 3$, $s_2 = 1$.

(b) The augmented matrix of this system is $\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 1 & 1 & 0 & 2 \\ 0 & 1 & -1 & 1 \end{array}\right]$. If we replace Row 2 with itself minus Row 1, we get $\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 3 & -1 & 2 \\ 0 & 1 & -1 & 1 \end{array}\right]$. We can then interchange Row 2 and Row 3 to get $\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 3 & -1 & 2 \end{array}\right]$. Then we can replace Row 3 with itself minus three times Row 2, to get $\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 2 & -1 \end{array}\right]$. We scale Row 3 by $\frac{1}{2}$ to get $\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & \frac{-1}{2} \end{array}\right]$. The next step is to replace Row 2 with itself plus Row 3, which gives $\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{-1}{2} \end{array}\right]$. We replace Row 1 with itself plus twice Row 2 to get $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{-1}{2} \end{array}\right]$. Finally, we replace Row 1 with itself minus Row 3 to get $\left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{3}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{-1}{2} \end{array}\right]$. From this, we can see that $\left(\frac{3}{2}, \frac{1}{2}, \frac{-1}{2}\right)$ is the solution to this system.

(c) The augmented matrix of this system is $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & -2 & 1 & 0 \\ 1 & 0 & -1 & 0 \end{array}\right]$. We replace Row 2 with itself minus Row 1 to get $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & -3 & 0 & -3 \\ 1 & 0 & -1 & 0 \end{array}\right]$, then scale Row 2 by $\frac{-1}{3}$ to get $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{array}\right]$. We replace Row 1 with itself minus

Row 2 to get $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{array}\right]$. We replace Row 3 with itself minus Row 1 to get $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -2 & -2 \end{array}\right]$. Then, we scale Row 3 by $\frac{-1}{2}$ to get $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right]$. Finally, we replace Row 1 with itself minus Row 3 to get $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right]$. From this, we see that $s_1 = s_2 = s_3 = 1$ is the solution to this system.

Under the Hood: Why do elementary row operations result in equivalent systems of equations? Each equation is just some true statement about the solutions, and the system of equations is a collection of true statements that together give us enough information to figure out exactly what the solutions are. Elementary row operations are tools we use to make new true statements that contain the same amount of information about the solutions as the old ones. We know the new statements contain the same amount of information because they're reversible—if we started with them, we could do a different series of operations to get the original system. Come see me if you want to talk about why we are sure they make true statements—or try to prove it on your own!

The Row Reduction Algorithm

The Punch Line: Given any linear system of equations, there is a procedure which finds a particularly simple equivalent system.

Warm-Up: For each of these matrices, determine if it is in Echelon Form, Reduced Echelon Form, or neither.

(a)
$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

(b)
$$\begin{bmatrix} 1 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

(c)
$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 4 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

(d)
$$\begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(e)
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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

(a) Neither

(b) Reduced Echelon Form

(c) Neither

(d) Echelon Form

(e) Reduced Echelon Form

(f) Reduced Echelon Form

Using the Algorithm: Five steps transform any matrix into a row-equivalent Reduced Echelon Form matrix:

1) Identify the pivot column. This will be the leftmost column with a nonzero entry.
2) Select a nonzero entry in that column to be the pivot for that column. If necessary, interchange rows to put it at the top of the matrix.
3) Eliminate all of the nonzero entries in the pivot column by using row replacement operations.
4) Repeat steps 1)-3) on all rows you haven't yet used.
5) Eliminate all nonzero entries above each pivot, and scale each nonzero row so its pivot is 1.

1. Transform each of these matrices into Reduced Echelon Form with the above procedure:

(a)
$$\begin{bmatrix} 1 & 1 & 1 \\ -2 & 4 & -2 \\ -1 & 0 & 1 \end{bmatrix}$$

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

(c)
$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}$$

(a) We already have a 1 in leading position in the first row, first column, so we identify the first column as the pivot column and select the 1 in the upper left to be the pivot. Then, we add twice Row 1 to Row 2, and add Row 1 to Row 3, to yield the matrix $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 6 & 0 \\ 0 & 1 & 2 \end{bmatrix}$.

Having eliminated all of the other nonzero entries in the pivot column, we repeat for the second pivot position. There are entries in the second column, so that will be our new pivot column. To save future work, I will select the 6 in the second row as our pivot—it would be equally valid to select the 1 in the third row and interchange Rows 2 and 3 to put it as the next pivot entry. With 6 as the pivot, though, we can add $\frac{-1}{6}$ times Row 2 to Row 3 to get the matrix $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, which is in Echelon Form.

In order to get Reduced Echelon Form, we apply step 5. First, we can multiply Row 2 by $\frac{1}{6}$ and Row 3 by $\frac{1}{2}$ to get the matrix $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Then, we eliminate the entries above the pivots by subtracting both Row 2 and Row 3 from Row 1 to get $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. This is in Reduced Echelon Form, so we're done!

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

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

Interpreting the Results: The Reduced Echelon Form of the augmented matrix of a linear system can be used to find all solutions (the *solution set*) of the system at once. To do this, we write out the system corresponding to the Reduced Echelon Form matrix, then solve for all of the variables in pivot positions (we can do this easily because each one only appears in a single equation). Any remaining variables are called *free variables*, and can take on any value in a solution.

2. Find the solution set of each of these linear systems:

(a)

$$\begin{aligned} 2x_1 + 4x_3 &= 0 \\ x_2 + x_3 &= 3 \\ x_1 + x_2 + 3x_3 &= 3 \end{aligned}$$

(b)

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

(c)

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 + 2x_2 &= 1 \\ 2x_1 + 3x_2 &= 3 \end{aligned}$$

(a) The augmented matrix of this system is $\begin{bmatrix} 2 & 0 & 4 & 0 \\ 0 & 1 & 1 & 3 \\ 1 & 1 & 3 & 3 \end{bmatrix}$. From Part 1, we know that this has Reduced Echelon Form $\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. This corresponds to the system of equations

$$\begin{aligned} x_1 + 2x_3 &= 0 \\ x_2 + x_3 &= 3 \\ 0 &= 0. \end{aligned}$$

Since x_1 and x_2 correspond to the pivot positions in the augmented matrix, we solve for them in terms of the free variable x_3: $x_1 = -2x_3$ and $x_2 = 3 - x_3$. Since x_3 is not in a pivot position, it is a free variable, which means that we can't pick a definite value for it. Instead, there is an infinite number of solutions to the system of equations, one for *every possible* value of x_3. In these solutions, x_1 and x_2 are defined by the equations we just calculated.

(b) The augmented matrix of this system is $\begin{bmatrix} 1 & 1 & 1 \\ -2 & 4 & -2 \\ -1 & 0 & 1 \end{bmatrix}$ (pay special attention to Row 2 there; it's important to keep the variables in the same order in all rows of the augmented matrix, so because we had x before y in Row 1, we need to do the same in Row 2, even though the equation is given to us the other way around). This has Reduced Echelon form $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. The last row corresponds to the equation $0 = 1$, which is never true. This means that the system can't have any solutions, so it is *inconsistent*.

(c) The augmented matrix of this system is $\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 3 & 3 \end{bmatrix}$, which has Reduced Echelon Form $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$. This corresponds to the system

$$\begin{aligned} x_1 &= 3 \\ x_2 &= -1 \\ 0 &= 0. \end{aligned}$$

The equation $0 = 0$ is always true, so we can read off the solution to the system $x_1 = 3$ and $x_2 = -1$.

Under the Hood: The Reduced Echelon Forms of any two equivalent systems are the same. Since equivalent systems have the same solution set (by definition!), it is in some sense the simplest system with that solution set. Thus, the Row Reduction Algorithm is a way to find the simplest description of the solution set of a linear system—that works every time!

Vector Equations

The Punch Line: Vector equations allow us to think about systems of linear equations as geometric objects, and are an efficient notation to work with.

Warm-Up: Sketch the following vectors in $\mathbb{R}^2$:

(a) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

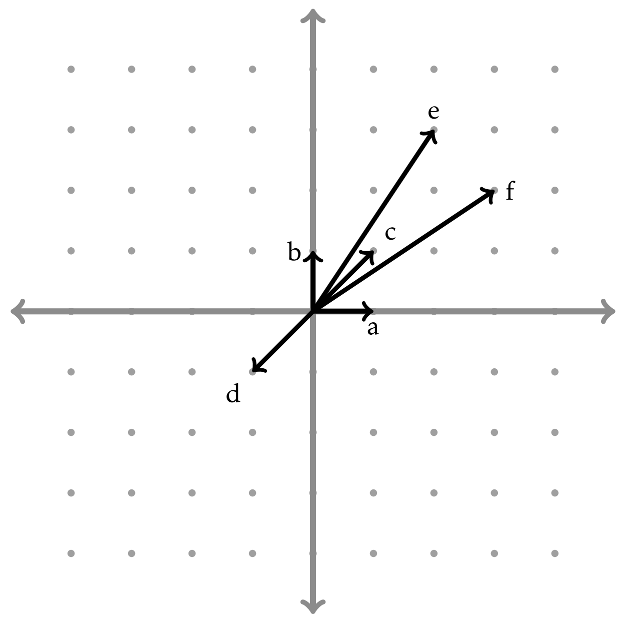
(b) $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(d) $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$

(e) $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$

(f) $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$



Linear Combinations: A *linear combination* of the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ with *weights* $w_1, w_2, \ldots, w_n$ is the vector $\vec{y}$ defined by

$$\vec{y} = w_1\vec{v}_1 + w_2\vec{v}_2 + \cdots + w_n\vec{v}_n.$$

That is, it's a sum of multiples of the vectors. Geometrically, it corresponds to stretching each vector $\vec{v}_i$ (where i is one of $1, 2, \ldots, n$) by the weight w_i, then laying them end to end and drawing $\vec{y}$ to the endpoint of the last vector.

1 Compute the following linear combinations:

(a) $\begin{bmatrix}1\\0\end{bmatrix}+\begin{bmatrix}0\\1\end{bmatrix}$

(b) $(-1)\begin{bmatrix}1\\1\end{bmatrix}$

(c) $\begin{bmatrix}2\\3\end{bmatrix}-\begin{bmatrix}3\\2\end{bmatrix}$

(d) $\begin{bmatrix}1\\2\\3\end{bmatrix}-2\begin{bmatrix}1\\1\\1\end{bmatrix}+\begin{bmatrix}1\\0\\-1\end{bmatrix}$

(e) $\begin{bmatrix}1\\2\end{bmatrix}+\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}$

(f) $4\begin{bmatrix}1\\\frac{1}{2}\end{bmatrix}-2\begin{bmatrix}\frac{1}{3}\\1\end{bmatrix}+3\begin{bmatrix}\frac{2}{9}\\2\end{bmatrix}$

Think about what each of these linear combinations mean geometrically (try sketching them).

(a) Addition of vectors is componentwise, so this linear combination yields $\begin{bmatrix}1\\1\end{bmatrix}$.

(b) Multiplication of a number and a vector (called *scalar multiplication* because the number is acting to scale the vector) is also componentwise, so this is $\begin{bmatrix}-1\\1\end{bmatrix}$.

(c) Applying the rules in sequence, we get $\begin{bmatrix}2\\3\end{bmatrix}-\begin{bmatrix}3\\2\end{bmatrix}=\begin{bmatrix}2\\3\end{bmatrix}+\begin{bmatrix}-3\\-2\end{bmatrix}=\begin{bmatrix}-1\\-1\end{bmatrix}$.

(d) The answer here is $\begin{bmatrix}0\\0\\0\end{bmatrix}$.

(e) This one is $\begin{bmatrix}1\\1\end{bmatrix}$.

(f) Finally, $\begin{bmatrix}4\\6\end{bmatrix}$.

Span: The *span* of the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is the set of all linear combinations of them. If $\vec{x}$ is in Span$\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, then we will be able to find some weights $w_1, w_2, \ldots, w_n$ to make the linear combination using those weights result in $\vec{x}$:

$$w_1\vec{v}_1 + w_2\vec{v}_2 + \cdots + w_n\vec{v}_n = \vec{x}.$$

Often, we are interested in determining if a given vector is in the span of some set of other vectors. In particular, a system of linear equations has a solution precisely when the rightmost column of the augmented matrix is in the span of the columns to the left of it. This means a system of linear equations is equivalent to a single vector equation.

2 Determine if $\vec{x}$ is in the span of the given vectors:

(a) $\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$; $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$

(b) $\vec{x} = \begin{bmatrix} 12 \\ 14 \end{bmatrix}$; $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(c) $\vec{x} = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$; $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\vec{v}_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

If it is, describe the linear combination that yields it.

(a) To check this, we write down the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 = \vec{x},$$

which says "the linear combination with weights a_1 and a_2 of vectors $\vec{v}_1$ and $\vec{v}_2$ is $\vec{x}$". If $\vec{x}$ is in Span$\{\vec{v}_1, \vec{v}_2\}$, then this equation will have a solution. We can write it out in components to see that this is equivalent to the system of linear equations

$$\begin{aligned} a_1 - 2a_2 &= 1 \\ a_1 &= 1 \\ a_1 + 2a_2 &= 1. \end{aligned}$$

By computing the Reduced Echelon Form of the augmented matrix of this system, we can identify any solutions, if they exist. However, the REF is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Since the last column has a pivot entry, we can see that this system is inconsistent. This means that the system of linear equations, and therefore the vector equation, does not have a solution. This means no linear combination of $\vec{v}_1$ and $\vec{v}_2$ yields $\vec{x}$, so it is not in their span.

(b) By following the above procedure, we can find that $\vec{x}$ is in the span of $\vec{v}_1$ and $\vec{v}_2$, with weights $a_1 = 13$ and $a_2 = -1$.

(c) Similarly, we find here that $\vec{x}$ is in the span of $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$. Our REF is $\begin{bmatrix} 1 & 0 & -1 & 5 \\ 0 & 1 & 2 & -4 \end{bmatrix}$, so we see that we have a free variable x_3, so there are infinitely many linear combinations that give $\vec{x}$. In particular, if $a_1 = 5 + a_3$ and $a_2 = -4 - 2a_3$ (and a_3 is anything) we have $a_1\vec{v}_1 + a_2\vec{v}_2 + a_3\vec{v}_3 = \vec{x}$.

Under the Hood: The span of a collection of vectors is essentially the set of all vectors that can be constructed using the members of the collection as components. This means that if a vector is *not* in the span of the collection, it has some additional component that's different from everything in the collection.

Matrix-Vector Products

The Punch Line: We can use even more compact notation than vector equations by introducing matrices. This will allow us to study systems of linear equations by studying matrices.

Warm-Up: Write the following systems of linear equations as vector equations:

(a) The system with variables z_1 and z_2

$$\begin{aligned} z_1 + 2z_2 &= 6 \\ 2z_1 - 5z_2 &= 3. \end{aligned}$$

(b) The system with variables x, y, and z

$$\begin{aligned} x &= x_0 \\ y &= y_0 \\ z &= z_0. \end{aligned}$$

(c) The system with variables x_1, x_2, and x_3

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - 2x_2 + x_3 &= 0 \\ x_1 - x_3 &= 0. \end{aligned}$$

(a) $z_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + z_2 \begin{bmatrix} 2 \\ -5 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$

(b) $x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$

(c) $x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$

The Technique: The linear combination $x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n$ is represented by the matrix-vector product

$$A\vec{x} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

This means that to compute a matrix-vector product, we can just write it back out as a linear combination of the columns of the matrix. This means that matrix-vector products only work when there are precisely as many columns in the matrix as there are entries in the vector.

1 Compute the following matrix-vector products:

(a) $\begin{bmatrix} 1 & 2 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$

(a) We write this as

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} &= 4 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ -5 \end{bmatrix} \\ &= \begin{bmatrix} 4(1) + 1(2) \\ 4(2) + 1(-5) \end{bmatrix} \\ &= \begin{bmatrix} 6 \\ 3 \end{bmatrix}. \end{aligned}$$

(b) This is $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

(c) This is $\begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$.

(d) This is $\begin{bmatrix} 5 \\ 17 \end{bmatrix}$.

Applications: The matrix equation $A\vec{x} = \vec{b}$ can be rephrased as the assertion that $\vec{b}$ is in the span of the columns of A. This gives us a geometric interpretation of systems of linear equations when we write them in matrix form—an equation being true means a particular vector, $\vec{b}$, is in the span of the collection of vectors $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\}$ that make up the matrix A. In this case, the vector $\vec{x}$ is the collection of weights in a linear combination that proves $\vec{b}$ is in the span of the columns of A.

2 If possible, find at least one solution to each of these matrix equations (if not, explain why it is impossible):

(a) $\begin{bmatrix} 1 & 2 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

(a) We have seen that $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ is a solution. To verify (and find any others), we write the augmented matrix $\begin{bmatrix} 1 & 2 & 6 \\ 2 & -5 & 3 \end{bmatrix}$. This has Reduced Echelon Form $\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 1 \end{bmatrix}$. From this, we can see that $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ is the unique solution (and, if we hadn't already done the multiplication from the previous problem, we have derived it from just the equations).

(b) We start with the augmented matrix $\begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & -2 & 1 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix}$. This has REF $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$, so we see the unique solution is $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

(c) The augmented matrix here is $\begin{bmatrix} 1 & 1 & b_1 \\ 1 & -1 & b_2 \\ 2 & 0 & b_3 \end{bmatrix}$. We work just with the left columns of the augmented matrix, and find that in REF, it looks like $\begin{bmatrix} 1 & 0 & * \\ 0 & 1 & * \\ 0 & 0 & * \end{bmatrix}$. This only works for some values that we could put into the $*$s, but not in general. This means that this matrix equation is inconsistent for (most) $\vec{b}$ (and, therefore, that the columns of the matrix do not span $\mathbb{R}^3$).

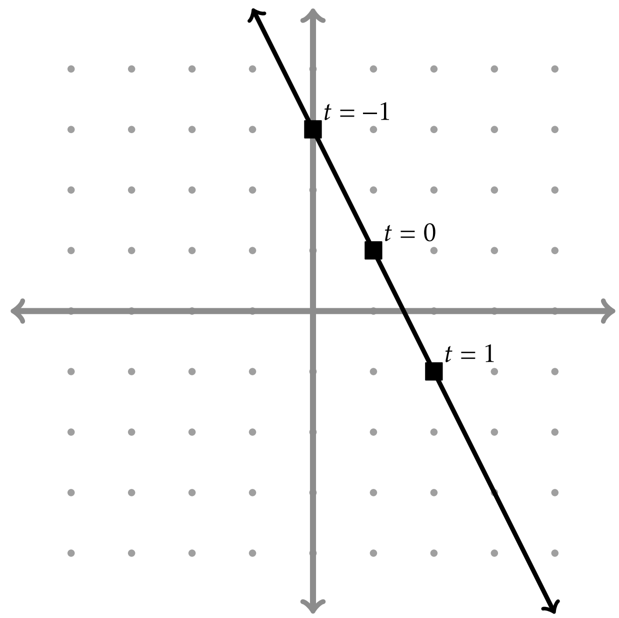
(d) Here, the REF of the augmented matrix is $\begin{bmatrix} 1 & 0 & -3 & 0 \\ 0 & 1 & 3 & 0 \end{bmatrix}$. We have a free variable in this, so there are infinitely many solutions. We can choose a value for x_3 to get a particular solution—choosing $x_3 = 0$ gives the solution $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, while choosing $x_3 = 1$ yields $\begin{bmatrix} 3 \\ -3 \\ 1 \end{bmatrix}$. In fact, the set of all solutions can be represented as $\vec{x} = t \begin{bmatrix} 3 \\ -3 \\ 1 \end{bmatrix}$, which forms a line (more on this in the next section of the book).

Under the Hood: Given any vector $\vec{b}$, the equation $A\vec{x} = \vec{b}$ means that $\vec{b}$ is in the span of the columns of A. This means that the span of the columns of A is related to the set of all possible matrix equations that could be solved with $A\vec{x}$ as the left hand side—there's one for each $\vec{b}$ in the span!

Solution Sets of Linear Systems

The Punch Line: There is a geometric interpretation to the solution sets of systems 0f linear equations, which allows us to explicitly describe them with *parametric equations*.

Warm-Up: Draw the line in $\mathbb{R}^2$ defined by $y = 3 - 2x$.



Verify that $x(t) = 1 + t$ and $y(t) = 1 - 2t$ satisfy the equation $y(t) = 3 - 2x(t)$ for all t, and plot $\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$ for $t = -1$, 0, and 1.

We compute $2x(t) + y(t) = 2(1 + t) + (1 - 2t) = 2 + t + 1 - 2t = 3$. Thus, $y(t) = 3 - 2x(t)$.

Homogeneous Equations: A matrix equation of the form $A\vec{x} = \vec{0}$ is called *homogeneous*. It always has the solution $\vec{x} = \vec{0}$, which is called the *trivial solution*. Any other solution is called a *nontrivial solution*; nontrivial solutions arise precisely when there is at least one free variable in the equation.

If there are m free variables in the homogeneous equation, the solution set can be expressed as the span of m vectors:

$$\vec{x} = s_1\vec{v}_1 + s_2\vec{v}_2 + \cdots + s_m\vec{v}_m.$$

This is called a *parametric equation* or a *parametric vector form* of the solution. A common parametric vector form uses the free variables as the parameters s_1 through s_m.

1 Find a parametric vector form for the solution set of the equation $A\vec{x} = \vec{0}$ for the following matrices A:

(a) $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & -3 \\ 2 & -1 & -1 \\ 4 & -4 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & -2 & 0 \\ -2 & 0 & 4 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

(a) We compute the REF of $\begin{bmatrix} A & \vec{0} \end{bmatrix}$, finding it to be $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. This means x_2 is a free variable, so solve for the relation $x_1 = -2x_2$. Since we know $x_2 = x_2$ (and can't say anything more, because x_2 is free), we can express our solution in the parametric form

$$\vec{x} = x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \text{Span}\left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}.$$

(b) Here our REF is $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, so the only solution is $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. We could write this as the parameterized version $\vec{x} = s \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \text{Span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$, but it's simplest to just leave it as $\vec{0}$ (which is still an explicit solution to the equation).

(c) The REF is $\begin{bmatrix} 1 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. This means x_2, x_3, and x_4 are all free variables. We express $x_1 = 2x_3$. With this, we can write our parametric solution as

$$\vec{x} = x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

(d) The REF here is again $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, so the only solution is $\vec{0}$.

Nonhomogeneous Equations: A matrix equation of the form $A\vec{x} = \vec{b}$ where $\vec{b} \neq \vec{0}$ is called *nonhomogeneous*. As we've seen, a nonhomogeneous system may be inconsistent and fail to have solutions. If it does have a solution, though, we can find a parametric form for them as well as in the homogeneous case. Here, we express the solutions as $\vec{x} = \vec{p} + \vec{v}_h$, where $\vec{p}$ is some particular solution to the nonhomogeneous system (which we can get by picking simple values for the parameters, such as taking all free variables to be zero), and $\vec{v}_h$ is a parametric form for the solution to the *homogeneous* equation $A\vec{v}_h = \vec{0}$.

2 If possible, find a parametric vector form for the solution set of the nonhomogeneous equation $A\vec{x} = \vec{b}$ for the following matrices A and vectors $\vec{b}$ (otherwise explain why it is impossible):

(a) $\begin{bmatrix} 1 & 2 \end{bmatrix}; \begin{bmatrix} 3 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}; \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 0 & 1 \\ 0 & -1 & 0 \\ 4 & 2 & 1 \end{bmatrix}; \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 2 \\ 1 & -1 \\ 2 & 2 \end{bmatrix}; \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(a) Since the augmented matrix of this system, $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, is already in REF, we can immediately solve for the pivot variable x_1 in terms of the free variable x_2 and constants. We get $x_1 = 3 - x_2$. To get a parametric vector form for the solution, we first choose the value $x_2 = 0$ to see that the vector $\vec{p} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ is *a* solution of the nonhomogeneous equation. To get the rest, we look at the homogeneous system with the same matrix A (this amounts to looking at the REF with the last column set to all zeros, or seeing how changing the free variables change the pivot variables). In this case, any vector of the form $s\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is a solution to the homogeneous equation (increasing x_2 decreases x_1 by the same amount), so our final parametric vector form of the answer is

$$\vec{x} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} + s\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

(b) The REF of this system is $\begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{-1}{2} \end{bmatrix}$. There are no pivot variables, so the only solution we get is

$$\vec{x} = \begin{bmatrix} \frac{1}{2} \\ \frac{-1}{2} \end{bmatrix}.$$

This is in parametric form, although there are no parameters.

(c) The REF here is $\begin{bmatrix} 1 & 0 & 0 & \frac{5}{2} \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -4 \end{bmatrix}$. This again has a unique solution, $\vec{x} = \begin{bmatrix} \frac{5}{2} \\ -2 \\ -4 \end{bmatrix}$.

(d) The REF here is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Since the last column is a pivot, the system is inconsistent, so the solution set is empty—there are no vectors $\vec{x}$ which solve this equation.

Under the Hood: Why do the solution sets to nonhomogeneous solutions have a "homogeneous part"? Imagine we are given two vectors, $\vec{x}_1$ and $\vec{x}_2$, and we're assured that $A\vec{x}_1 = \vec{b}$ and $A\vec{x}_2 = \vec{b}$. That is, we have two solutions to the nonhomogeneous equation. We can take the difference between these two equations to see that $A\vec{x}_1 - A\vec{x}_2 = \vec{b} - \vec{b}$. A property of matrix-vector multiplication lets us write the left-hand side as $A(\vec{x}_1 - \vec{x}_2)$, while the right-hand side is clearly $\vec{0}$, so we're left with the equation $A(\vec{x}_1 - \vec{x}_2) = \vec{0}$. That is, we've just shown the *difference* between two solutions to the nonhomogeneous equation is always a solution to the homogeneous equation with the same matrix!

Applications of Linear Systems

The Punch Line: Linear systems of equations can describe many interesting situations.

Set-Up: In a situation you can model with linear equations, there will be a number of *constraints*: things which must be equal because of the laws governing what's going on (e.g., laws of physics, economic principles, or definitions of quantities and the values you observe for them). These will give you the equations that you can solve to get information about the variables you care about

1 In the past three men's soccer games, the Gauchos averaged $\frac{5}{3}$ goals per game. They scored the same number of goals in the most recent two games, but three games ago they scored an additional two goals. How many points did they score in each game?

(This problem was actually true as of the 11^{th}, but even if you know the scores, it's probably helpful to set up the system and see how they come out of the equations.)

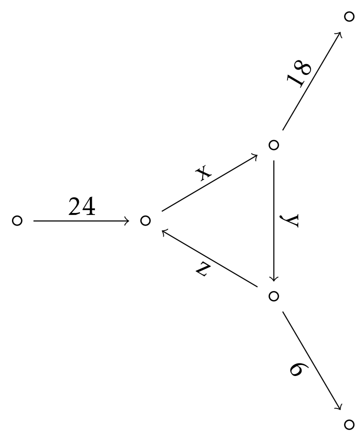
We'll call the number of goals in the past three games g_1, g_2, and g_3, with g_1 being the most recent. That the average is $\frac{5}{3}$ means that $\frac{1}{3}(g_1 + g_2 + g_3) = \frac{5}{3}$, which we can simplify to $g_1 + g_2 + g_3 = 5$. That the past two games had the same number of goals means $g_1 = g_2$, which we can put into the more standard form (with variables on one side and constants on the other) of $g_1 - g_2 = 0$. That three games ago there were two more goals can be represented with the equation $g_3 = g_2 + 2$ (we could equally well have used g_1, or even $\frac{1}{2}(g_1 + g_2)$—any quantity which is equal to their shared value). This can be written as $g_3 - g_2 = 2$. Thus, we have the system of linear equations

$$\begin{cases} g_1 + g_2 + g_3 & = 5 \\ g_1 - g_2 & = 0 \\ g_3 - g_2 & = 2. \end{cases}$$

The REF of the augmented matrix of the system is $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix}$, so the unique solution is

$$\vec{g} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}.$$

2 Suppose you're watching a bike loop on campus and writing down the net number of bicycles travelling through each part of the loop (the number of bikes going one direction minus the number going the other direction). You're able to observe how many net bikes per minute enter and leave through each of the three spokes, but aren't able to count well inside the loop. Luckily, you can use linear algebra to learn about how many net bikes per minute travel through each part of the loop (which is to say, find all solutions for x, y, and z that are consistent with the rest of the information about the problem)!



Here we're going to appeal to a principle I'd call "conservation of bikers": assuming that the (net) number of bicycles entering a junction of paths equals the (net) number of bicycles exiting that junction (this is an example of a "conservation law", which is a broad class of physical laws that often generate constraints that can be used in modelling systems). Applying this rule to each of the three junctions in the above diagram gives the system of equations

$$\begin{cases} 24+z &= x \\ x &= 18+y \\ y &= 6+z. \end{cases}$$

Rearranging these into standard form (with variables on one side and constants on the other) gives

$$\begin{cases} x-z &= 24 \\ x-y &= 18 \\ y-z &= 6. \end{cases}$$

Finding the REF of the augmented matrix of this system yields $\begin{bmatrix} 1 & 0 & -1 & 24 \\ 0 & 1 & -1 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. We extract the parametric vector form of the solution from this, obtaining

$$\vec{x} = \begin{bmatrix} 24 \\ 6 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

This means the set of solutions has an infinite number of possibilities, one for each possible (net) number of bikes travelling along path z.

3 (Example 1 in Section 1.6) In an *exchange model* of economics, an economy is divided into different sectors which depend on each others' products to produce their output. Suppose we know for each sector its total output for one year and exactly how this output is divided or "exchanged" among the other sectors of the economy. The total dollar (or other monetary unit) value of each sector's output is called the *price* of that output. There is an *equilibrium price* for this kind of model, where each sectors income exactly balances its expenses. We wish to find this equilibrium.

Suppose we have an economy described by the following table:

Distribution of output from: Coal	Electric	Steel	Purchased by:
0.0	0.4	0.6	Coal
0.6	0.1	0.2	Electric
0.4	0.5	0.2	Steel

If we denote the price of the total annual outputs of the Coal, Electric, and Steel sectors by p_C, p_E, and p_S respectively, what is the equilibrium price (or describe them if there is more than one).

We're trying to balance the input into each sector with the output, which is represented by the price variable for that sector. The input is the proportion of each sector's output devoted to the purchasing sector (as given in the table). This means we get the system of equations

$$\begin{cases} p_C &= 0.4p_E + 0.6p_S \\ p_E &= 0.6p_C + 0.1p_E + 0.2p_S \\ p_S &= 0.4p_C + 0.5p_E + 0.2p_S. \end{cases}$$

We write them in standard form (moving all variables to one side), and for computational convenience multiply each resulting equation by 10 to deal with integers rather than decimals, noting that this won't change our answer. This gives the homogeneous system of equations

$$\begin{cases} 10p_C - 4p_E - 6p_S &= 0 \\ -6p_C + 9p_E - 2p_S &= 0 \\ -4p_C - 5p_E + 8p_S &= 0. \end{cases}$$

The REF of the augmented matrix of this system is $\begin{bmatrix} 1 & 0 & -\frac{31}{33} & 0 \\ 0 & 1 & -\frac{28}{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. This means that our solutions set has the parametric vector form $\vec{p} = t\begin{bmatrix} 31 \\ 28 \\ 33 \end{bmatrix}$ (where the parameter $t = \frac{1}{33}p_S$ was chosen for integer entries in the vector). This means that the equilibrium is reached whenever the ratios between the prices are precisely 31 : 28 : 33 (regardless of the actual magnitude; if all prices doubled, say, there would still be an equilibrium).

Under the Hood: When can we use linear equations to model something? The basic setup of a linear system involves a collection of quantities that we know are equal to known values (or each other), and a collection of variables. We can use a linear system when the way the quantities depend on changes to the variables is independent of the actual values of the variables (adding the same amount to a variable changes each quantity in the same way, no matter what value any of the variables have).

Linear Independence

The Punch Line: *Linear independence* is a property describing a collection of vectors whose span is "as big as it can be."

Warm-Up: Are each of these situations possible?

(a) The vectors $\{\vec{u}, \vec{v}\}$ in $\mathbb{R}^2$ span all of $\mathbb{R}^2$.

(b) The vectors $\{\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4\}$ span all of $\mathbb{R}^3$.

(c) The vectors $\{\vec{x}, -\vec{x}\}$ span all of $\mathbb{R}^2$.

(d) The vectors $\{\vec{u}, \vec{v}\}$ span all of $\mathbb{R}^3$.

(e) The span of $\{\vec{u}_1, \vec{u}_2\}$ and the span of $\{\vec{v}_1, \vec{v}_2\}$ in $\mathbb{R}^3$ intersect only at $\vec{0}$.

(f) The span of $\{\vec{u}_1, \vec{u}_2\}$ and the span of $\{\vec{v}_1, \vec{v}_2\}$ in $\mathbb{R}^3$ are both planes and intersect only at $\vec{0}$.

(a) Yes, two vectors can span a plane.

(b) Yes, four vectors can span 3-space.

(c) No, they are both on the same line, so their span is that line.

(d) No, two vectors can span at most a plane.

(e) Yes, but only if at least one of the spans is just a line and not a whole plane.

(f) No, any two planes in $\mathbb{R}^3$ that include $\vec{0}$ (or any specific point in common) must intersect on at least a line.

Tests for Linear Independence: That the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is linearly independent means that if $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0}$, the only possibility is that $c_1 = c_2 = \cdots = c_n = 0$. That is, the only solution to the homogeneous matrix equation

$$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \vec{0}$$

is the trivial solution: the zero vector. If there's only one vector in the set, it is linearly independent unless it happens to be $\vec{0}$. If there's only two vectors, they are linearly independent unless one is a multiple of the other (including $\vec{0}$, which is 0 times any vector). If a subset of the vectors is linearly dependent, the whole set is. Finally, a set of vectors is linearly dependent if (and only if) at least one vector is in the span of the others.

1 Are these sets linearly dependent or independent? Why?

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

(b) $\left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ \frac{3}{4} \end{bmatrix} \right\}$

(c) $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$

(d) $\left\{ \begin{bmatrix} 2 \\ 6 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 12 \\ 1 \end{bmatrix} \right\}$

(a) No, the second vector is -1 times the first.

(b) No, the third vector is $\frac{1}{4}$ times the first, plus there are three vectors in $\mathbb{R}^2$.

(c) Yes, and we can check by observing there are no free variables in the REF of the coefficient matrix of the homogeneous equation described above.

(d) No, the third vector is the sum of the first and -2 times the second.

2 Are each of these situations possible?

(a) You have a set of vectors that spans $\mathbb{R}^3$. You remove two of them, and the set of vectors left behind is linearly dependent.

(b) You have two sets of vectors in $\mathbb{R}^6$. One has four vectors, and one has two vectors, and both sets are linearly independent. When you put both sets together, the resulting set of six vectors is linearly dependent.

(c) You have a set of three vectors which span $\mathbb{R}^3$, but it is linearly dependent.

(d) You have a linearly dependent set of three vectors in $\mathbb{R}^2$. If you remove any one of them, the other pair do not span $\mathbb{R}^2$.

(a) Yes, for example, if we started with $\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}\right\}$ and removed the last two.

(b) Yes, for example, if each vector in the set of two was a multiple of one from the set of four.

(c) No, because if the set is linearly dependent, one vector would be in the span of the other two, and the span of two vectors is a plane, while $\mathbb{R}^3$ is a 3-space.

(d) Yes, for example if the three vectors are all multiples of each other, so their span is a line.

Linear Transformations

The Punch Line: Matrix multiplication defines a special kind of function, known as a *linear transformation*.

Warm-Up: What do each of these situations mean (geometrically, algebraically, in an application, and/or otherwise)?

(a) The product of the matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

(b) The vector $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is in the span of $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$.

(c) The equation $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ has a solution.

(d) The set of vectors $\vec{x}$ such that the matrix equation $A\vec{x} = \vec{b}$ is satisfied forms a plane in $\mathbb{R}^3$.

(e) The set of vectors $\vec{b}$ such that the matrix equation $A\vec{x} = \vec{b}$ is satisfied forms a line in $\mathbb{R}^2$.

(f) For two particular vectors $\vec{x}$ and $\vec{b}$, and a matrix A, the matrix equation $A\vec{x} = \vec{b}$ is satisfied.

You might have more answers (and I would love to talk about them in office hours!), but here are some helpful ones:

(a) The matrix rotates the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ by 90° (or $\frac{\pi}{2}$ radians) counterclockwise to the vector $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$. In fact, the matrix rotates *any* vector by that angle, as you can check.

(b) There is a way to take a linear combination of the three vectors that yields the all-ones vector.

(c) There is a vector $\vec{x}$ that the matrix sends to (or transforms into) the all-ones vector.

(d) There are multiple linear combinations of the columns of A that yield $\vec{b}$, and A sends (infinitely) many vectors in $\mathbb{R}^3$ to $\vec{b}$.

(e) The span of the columns of A is a line, and A transforms any vector it multiplies into a multiple of some particular vector.

(f) The matrix A transforms the vector $\vec{x}$ into $\vec{b}$.

What They Are: A *linear transformation* is a mapping T that obeys two rules:

(a) $T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v})$ for all $\vec{u}$ and $\vec{v}$ in its domain,

(b) $T(c\vec{u}) = cT(\vec{u})$ for all scalars c and $\vec{u}$ in its domain.

These rules lead to the rule $T(c\vec{u}+d\vec{v}) = cT(\vec{u}) + dT(\vec{v})$ for c,d scalars and $\vec{u},\vec{v}$ in the domain of T, and in fact $T(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n)$. That is, the transformation of a linear combination of vectors is a linear combination of the transformations of the vectors (with the same coefficients).

1 Are each of these operations linear transformations? Why or why not?

(a) $T(\vec{x}) = 4\vec{x}$

(b) $T(\vec{x}) = A\vec{x}$ for some matrix A (with the right number of columns)

(c) $T(\vec{x}) = \vec{0}$

(d) $T(\vec{x}) = \vec{b}$ for some nonzero $\vec{b}$

(e) $T(\vec{x}) = \vec{x} + \vec{b}$ for some nonzero $\vec{b}$

(f) $T(\vec{x})$ takes a vector in $\mathbb{R}^2$ and rotates it by 45° ($\frac{\pi}{4}$ radians) counter-clockwise in the plane

(a) Yes, because $4(\vec{u}+\vec{v}) = 4\vec{u}+4\vec{v}$ by the distributive property, and $4(c\vec{u}) = 4c\vec{u} = c(4\vec{u})$ by the associative and commutative properties of scalar multiplication.

(b) Yes, the two properties of linear transformations are properties of matrix multiplication.

(c) Yes, because $T(\vec{u}+\vec{v}) = \vec{0} = \vec{0}+\vec{0} = T(\vec{u}) + T(\vec{v})$ and $T(c\vec{u}) = \vec{0} = c\vec{0} = cT(\vec{u})$. Note that we can find a matrix O (all of whose entries are zero) such that $O\vec{x} = \vec{0}$.

(d) No, because $T(c\vec{u}) = \vec{b}$, and $cT(\vec{u}) = c\vec{b}$, but $\vec{b} \neq c\vec{b}$ if $c \neq 1$ and $\vec{b} \neq \vec{0}$.

(e) No, because $T(\vec{u}+\vec{v}) = (\vec{u}+\vec{v}) + \vec{b} = \vec{u}+\vec{v}+\vec{b}$, but $T(\vec{u}) + T(\vec{v}) = (\vec{u}+\vec{b}) + (\vec{v}+\vec{b}) = \vec{u}+\vec{v}+2\vec{b}$, which is different for $\vec{b} \neq \vec{0}$.

(f) Yes. It's pretty clear the $T(c\vec{x}) = cT(\vec{x})$, because rotating a vector doesn't change its length, so if the input was a multiple of $\vec{x}$, the output will be that same multiple of $T(\vec{x})$. It's probably easiest to convince yourself that the vector addition property works with a sketch, but the gist is that rotating both vectors by the same amount doesn't change the *relative* angle between them, so laying them tail-to-head after the rotation looks essentially the same except for the initial angle. As it turns out, there's a matrix that accomplishes this linear transformation as well:

$$T(\vec{x}) = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \vec{x}.$$

It's not necessary to find this, but it does prove it's linear (by part (b)), and it's suggestive of things that will happen further along in the course...

What They Do: Linear transformations convert between two different spaces, such as $\mathbb{R}^n$ and $\mathbb{R}^m$. If $n = m$, then we can also think of them moving around the vectors inside $\mathbb{R}^n$ (e.g., by rotation or stretching).

2 What do the linear transformations corresponding to multiplication by these matrices do, geometrically? (Try applying the matrix to a vector composed of variables, then examining the result, or multiplying by a few simple vectors and sketching what happens.)

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(e) $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$

(a) This matrix does not change vectors it multiplies against.

(b) This matrix switches the coordinates of vectors it multiplies against, which reflects them about the line $y = x$.

(c) This rotates the x and y components by 90° (or $\frac{\pi}{2}$ radians), while leaving z alone.

(d) This "projects" a vector onto the z axis (it gives the vector that matches the input in height, but doesn't have any x or y components).

(e) This doubles the length of the vector.

(f) This quadruples the y coordinate while leaving x unchanged (this is sometimes called a shear transformation).

The Matrix of a Linear Transformation

The Punch Line: Linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ are *all* equivalent to matrix transformations, even when they are described in other ways.

Warm-Up: What does the linear transformation corresponding to multiplication by each of these matrices do geometrically (don't worry too much about the exact values for things like rotation or scaling)?

(a) $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

(a) This is a reflection about the x-axis—the y coordinate of each point is negated, so vectors above the axis are moved an equal distance below it, without changing in x.

(b) This is both a rotation by 45° ($\frac{\pi}{4}$ radians), and a scaling by $\sqrt{2}$.

(c) This is a projection to the xy-plane—the z coordinate collapses down to zero while the other coordinates remain unchanged.

(d) This maps a vector in $\mathbb{R}^3$ to the vector in $\mathbb{R}^2$ that looks like its projection in the xy plane. While this is a very similar transformation to the previous one, it's important to note that this time, the result is in a different space (honest-to-goodness $\mathbb{R}^2$, rather than a plane in $\mathbb{R}^3$).

Getting the Matrix: We can write down a matrix that accomplishes any linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ by writing down what the transformation does to the vectors corresponding to each component (these have a single 1 and the rest of their entries as zeros, and make up the columns of the $n \times n$ *identity matrix*, which has ones down the diagonal and zeros elsewhere).

1 Write down a matrix for each of these linear transformations.

(a) In $\mathbb{R}^2$, rotation by $180°$ (π radians) counterclockwise.

(b) In $\mathbb{R}^3$, rotation by $180°$ (π radians) counterclockwise in the xz plane.

(c) In $\mathbb{R}^2$, stretching the x direction by a factor of 2 then reflecting about the line $y = x$.

(d) In $\mathbb{R}^3$, the transformation that looks like a "vertical" (that is, the z direction is the one which moves) shear in both the xz and yz planes, each with a "shear factor" (the amount the corner of the unit square moves) of 2.

[Note: Don't worry too much if this one's harder than the rest, shear transformations are hard to describe. If you get stuck, it might be a good idea to work on Problem 2 rather than sink in too much time here.]

(a) We see that $\begin{bmatrix}1\\0\end{bmatrix}$ is sent to $\begin{bmatrix}-1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ is sent to $\begin{bmatrix}0\\-1\end{bmatrix}$. Putting these together, we get the matrix $\begin{bmatrix}-1 & 0\\0 & -1\end{bmatrix}$.

(b) Here we get $\begin{bmatrix}-1 & 0 & 0\\0 & 1 & 0\\0 & 0 & -1\end{bmatrix}$.

(c) Here $\begin{bmatrix}0 & 2\\1 & 0\end{bmatrix}$.

(d) Here $\begin{bmatrix}1 & 0 & 2\\0 & 1 & 2\\0 & 0 & 1\end{bmatrix}$ (ask me if you want to talk about why).

One to One and Onto: When describing a linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^m$, we say T is *one to one* if each vector in $\mathbb{R}^m$ is the image of at most one vector in $\mathbb{R}^n$ (it can fail to be the image of any vector, it just can't be the image of two different ones). We say T is *onto* if each vector in $\mathbb{R}^m$ is the image of at least one vector in $\mathbb{R}^n$ (it can be the image of more than one).

We can test these conditions with ideas we already know: T is one-to-one if and only if the columns of its matrix are linearly independent, and onto if and only if they span $\mathbb{R}^m$. An equivalent test for T being one-to-one is that the equation $A\vec{x} = \vec{0}$ (where A is the matrix of T) has only the trivial solution if and only if T is one-to-one. An equivalent test for onto is that $A\vec{x} = \vec{b}$ is consistent for all $\vec{b}$ in $\mathbb{R}^m$.

2 Determine if the linear transformations with the following matrices are one-to-one, onto, both, or neither.

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

(f) $\begin{bmatrix} 2 & 1 & 0 \\ 6 & -3 & 12 \\ 5 & 2 & 1 \end{bmatrix}$

(a) This is both, as $\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$ both spans $\mathbb{R}^2$ and is linearly independent.

(b) This is also both.

(c) This is one-to-one but not onto, as $\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right\}$ is linearly independent, but does not span $\mathbb{R}^3$ (in particular, $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ is not in their span).

(d) This is onto but not one-to-one, as $\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$ spans $\mathbb{R}^2$, but is not linearly independent (in particular, $\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$).

(e) This is neither, as $A\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \vec{0}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ (for example) is not in the span of the columns.

(f) This is also neither, as the columns are linearly dependent and do not span $\mathbb{R}^3$.

Why does the $A\vec{x} = \vec{0}$ test work? If $A\vec{x} = A\vec{y}$, then $A(\vec{x} - \vec{y}) = \vec{0}$. If x and y weren't the same to begin with, then their difference is mapped to $\vec{0}$ by A as a consequence of them having the same value for the product. Similarly, if $A\vec{z} = \vec{0}$ for a nonzero $\vec{z}$, then $A(\vec{x} + \vec{z}) = A\vec{x} + A\vec{z} = A\vec{x}$, even though $\vec{x} \neq \vec{x} + \vec{z}$.

Matrix Operations

The Punch Line: Various operations combining linear transformations can be realized with some standard matrix operations.

Addition and Scalar Multiplication: Just like with vector operations, the sum of matrices and the multiplication by a *scalar* (just a number, as opposed to a vector or matrix) are done component-by-component.

1 Try the following matrix operations:

(a) $3\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 \\ -2 & 4 \end{bmatrix} - \begin{bmatrix} 2 & -1 \\ 3 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + 2\begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

(a) This gives $\begin{bmatrix} 3 & 0 \\ 0 & -3 \end{bmatrix}$.

(b) This gives $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$.

(c) This gives $\begin{bmatrix} -1 & 3 \\ -5 & 4 \end{bmatrix}$.

(d) This one is $\begin{bmatrix} 3 & 2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$.

Matrix Multiplication: To multiply two matrices, we create a new matrix, each of whose columns is the result of the matrix-vector product of the left matrix with the corresponding column of the right matrix (the product will have the same number of rows as the left matrix, and the same number of columns as the right matrix). To get the ij entry (ith row and jth column) we could multiply the ith row of the left matrix with the jth column of the right matrix.

2 Multiply these matrices (if possible, otherwise say why it isn't):

(a) $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 \\ 0 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 4 & 2 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -2 \end{bmatrix}$

(a) We can compute this as $\left[\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right] = \begin{bmatrix} 2 & -1 \\ 3 & -1 \end{bmatrix}$

(b) We can compute the upper left $(1,1)$ entry as $\begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 0$, the upper right $(1,2)$ as $\begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -1$, the lower left $(2,1)$ as $\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1$, and the lower right $(2,2)$ as $\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1$ to get the resulting matrix $\begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}$.

(c) Using either method (I think the "linear combination of columns" is slightly easier, but your results may vary), we get this as $\begin{bmatrix} 2 & 1 & 3 \\ 2 & -2 & 0 \\ 1 & 2 & 3 \end{bmatrix}$.

(d) This is impossible—the left matrix has 4 columns, the right has 2 rows, and these numbers must match up for the matrix multiplication procedure to be well-defined.

Transpose: The last matrix operation for today is the *transpose*, where you switch the roles of rows and columns. That is, if you get an $n \times m$ matrix, its transpose will be $m \times n$.

3 Compute the following operations for the matrices given:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & -1 \end{bmatrix}.$$

(a) A^T

(b) B^T

(c) C^T

(d) $(BA)^T$

(e) $A^T B^T$

(f) $(BAC)^T$

(g) AA^T

(h) $A^T A$

(i) $(AA^T - B)^T$

(a) This is $\begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & -1 \end{bmatrix}$.

(b) This is $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$.

(c) This is $\begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 2 & -1 \end{bmatrix}$.

(d) The product $BA = \begin{bmatrix} 0 & 2 & 4 \\ 0 & -2 & -4 \end{bmatrix}$, so its transpose is $\begin{bmatrix} 0 & 0 \\ 2 & -2 \\ 4 & -4 \end{bmatrix}$.

(e) The product of transposes is also $\begin{bmatrix} 0 & 0 \\ 2 & -2 \\ 4 & -4 \end{bmatrix}$.

(f) This is $C^T A^T B^T = \begin{bmatrix} 0 & 0 \\ -2 & 2 \\ 0 & 0 \end{bmatrix}$.

(g) This is $\begin{bmatrix} 14 & -2 \\ -2 & 2 \end{bmatrix}$.

(h) This is $\begin{bmatrix} 2 & 2 & 2 \\ 2 & 4 & 6 \\ 2 & 6 & 10 \end{bmatrix}$. Note that it has different dimensions than AA^T.

(i) This is $AA^T - B^T = \begin{bmatrix} 13 & -1 \\ -1 & 1 \end{bmatrix}$.

What do these operations mean? Matrix addition and scalar multiplication correspond to adding and scaling the results of applying the linear transformation of the matrix, respectively. Matrix multiplication corresponds to composing the two linear transformations (applying one to the result of another). Transposition is a little weirder, and corresponds to switching the roles of variables and coefficients in a linear equation.

The Inverse of a Matrix

The Punch Line: Undoing a linear transformation given by a matrix corresponds to a particular matrix operation known as *inverse*.

Warm-Up: Are the following vector operations reversible/invertible?

(a) $T(\vec{x}) = 4\vec{x}$

(b) $T(\vec{x})$ is counterclockwise rotation in the plane by 45° ($\frac{\pi}{4}$ radians)

(c) $T(\vec{x}) = \vec{0}$

(d) $T(\vec{x}) = \vec{x} + \vec{b}$

(e) $T(\vec{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\vec{x}$

(f) $T(\vec{x}) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\vec{x}$

(a) Yes, with inverse $T^{-1}(\vec{x}) = \frac{1}{4}\vec{x}$.

(b) Yes, with an inverse given by clockwise 45° rotation.

(c) No, because we can't tell what the input was if everything goes to the same place.

(d) Yes, with inverse $T^{-1}(\vec{x}) = \vec{x} - \vec{b}$; note that this is not a linear transformation unless $\vec{b} = \vec{0}$.

(e) Yes, by multiplying by the same matrix again (switching components twice puts them back where they began).

(f) No, because $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -2 \end{bmatrix}$ are both sent to $\vec{0}$, so you can't tell which you started with by looking at the result.

The Inverse: The *inverse* of an $n \times n$ matrix A is another matrix B that satisfies the two matrix equations $AB = I_n$ and $BA = I_n$, where the *identity matrix* I_n has ones on the diagonal and zeroes everywhere else. We use the notation A^{-1} to refer to such a B (which, if it exists, is unique).

We can find the inverse of a matrix by applying row operations to the augmented matrix $\begin{bmatrix} A & I_n \end{bmatrix}$ (which is augmented with the n columns of the identity matrix, rather than a single vector). If the left part of the augmented matrix can be transformed by row operations to I_n, then the right part will be transformed by those row operations to A^{-1}. If the system is inconsistent, the matrix A is not invertible (and we may call it *singular*).

1 Find the inverse of these matrices (you may want to check your results by multiplying the result with the original matrix):

(a) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$

(c) $\begin{bmatrix} 3 & -1 \\ 7 & -2 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

(a) We find the REF of the augmented matrix $\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$, which is $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$, so the inverse of the matrix is $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Indeed, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Compare to the argument in part e) of the Warm-Up.

(b) This is $\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$.

(c) This is $\begin{bmatrix} -2 & 1 \\ -7 & 3 \end{bmatrix}$.

(d) This is $\begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$.

Relevance to Matrix Equations: The inverse of a matrix allows you to "reverse engineer" a matrix equation, in the sense that if $A\vec{x} = \vec{b}$ and A is invertible, then $\vec{x} = A^{-1}\vec{b}$ is a solution to the original equation. In fact, it is the unique solution to the equation!

---

2 Use the inverses computed previously to solve these matrix equations:

(a) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

(c) $\begin{bmatrix} 3 & -1 \\ 7 & -2 \end{bmatrix}\vec{x} = \begin{bmatrix} a \\ a+1 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\vec{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$

(a) We use the inverse calculated previously to get $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

(b) Here we get $\begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}$.

(c) We use the matrix as follows: $\begin{bmatrix} -2 & 1 \\ -7 & 3 \end{bmatrix}\begin{bmatrix} a \\ a+1 \end{bmatrix} = \begin{bmatrix} 1-a \\ 3-4a \end{bmatrix}$.

(d) Similarly, $\begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{-a+b+c}{2} \\ \frac{a-b+c}{2} \\ \frac{a+b-c}{2} \end{bmatrix}$.

Computing the inverse of a matrix reveals the structure of how to invert the linear transformation it represents. As the book notes, it can be faster to simply perform row operations to find a solution to any particular matrix equation. However, looking at the inverse matrix can give a more geometric idea of what undoing some particular operation is—to undo a rotation and shear requiring a different shear and rotation in the opposite direction, for example.

Characterizing Invertible Matrices

The Punch Line: There are many equivalent conditions to determine if a matrix is invertible, and describe properties of ones that we know are invertible.

Warm-Up: How big is the solution set of the homogeneous equation with these matrices (is it finite or infinite? what is its dimension?)? How about the span of their columns?

(a) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$

(a) The solution set to the homogeneous is finite, and consists of just the origin. The columns span $\mathbb{R}^3$.

(b) The solution set has one free variable, so is infinite and one-dimensional. There are three pivot rows, so the columns span a three-dimensional space.

(c) The solution set has two free variables, so is infinite and two-dimensional. There are two pivots, so the span of the columns is two-dimensional as well.

Matrix Conditions: Our first definition for invertible matrices states that A is invertible if some other matrix C makes the equations $AC = I_n$ and $CA = I_n$ simultaneously true. We can show that if A is invertible, so is A^T, that the inverse of A^T is C^T, and that if either one of the two equations in the definition is true, the other one must be as well, by playing around with transposing the equations.

1 Can each of these things happen? Do they have to be true?

(a) A is invertible and the matrix C with $AC = I_n$ is also invertible

(b) C is an inverse to both A and A^T (that is, $CA = I_n$ and $CA^T = I_n$)

(c) $CA = I_n$ and $AD = I_n$, but $C \neq D$

(d) $ABCD = I_n$, but $AB \neq I_n$ and $CD \neq I_n$

(a) This is always true—if A is invertible, its inverse (C in the above equation, often written A^{-1}) is also invertible, with inverse A. That is, *invertible matrices come in pairs*.

(b) This is possible, but requires that $C = C^T$, and implies $A = A^T$ (because we can just transpose the equations, and inverses are unique).

(c) This is not possible: if we take the transpose of the right equation, we get $D^T A^T = I_n^T = I_n$, but we know that C^T is the inverse for A^T, and inverses are unique.

(d) This is possible—it is saying that AB and CD are a pair of inverse matrices. Consider the case where all of them are $\frac{\sqrt{2}}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. The two products become $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which aren't identity matrices, but multiply to one.

Equation Conditions: We also have conditions based on the homogeneous and inhomogeneous equations involving the matrix. We know A is invertible if it is square ($n \times n$) and its columns span $\mathbb{R}^n$ or are linearly independent (for square matrices these are equivalent, though not in general). That is, it has n pivots (so its EF's have pivots in every column, and its REF is I_n) or no free variables. That is, the equation $A\vec{x} = \vec{b}$ has a solution for all $\vec{b}$ (which will turn out to be unique) or $A\vec{x} = \vec{0}$ has *only* the trivial solution.

2 Are these matrices invertible?

(a) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 & 3 \\ 0 & 2 & 2 \\ -1 & 1 & -1 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 3 & 1 & -4 \\ 0 & 1 & -1 & 2 \\ 1 & 0 & -2 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

(a) Yes—the columns are linearly independent, because the only way to get the right third component is to use the third column, then the only way to get the right second component is to set the second column to compensate, and so on.

(b) Yes—reducing to Echelon Form such as $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ shows there are 4 pivots.

(c) No—the third column is the sum of the second plus twice the first, so they are not linearly independent.

(d) No—there is no fourth component to any of the vectors, so $A\vec{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ has no solutions.

Why do the columns have to be linearly independent and span all of $\mathbb{R}^n$? If they were not linearly independent, there would be multiple solutions to $A\vec{x} = \vec{b}$, so we couldn't define $A^{-1}\vec{b} = \vec{x}$—we wouldn't know which to choose! And if they did not span $\mathbb{R}^n$, then there would be some $\vec{b}$ outside their span where we couldn't find any $\vec{x}$ so that $A\vec{x} = \vec{b}$—we'd again have a problem defining the inverse, but this time instead of having to many possible answers $\vec{x}$, we wouldn't have any!

Determinants!

The Punch Line: We can compute a value from the entries of a matrix to get yet *another* way of characterizing invertible matrices. ***SPOILER ALERT***: The determinant will also give us a variety of other useful pieces of information in understanding a matrix and its associated linear transformation!

Warm-Up: Are these matrices invertible? Are there conditions that make them so or not so depending on certain values? Try to answer without reducing them to REF (and in general, with as few computations as possible).

(a) $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 \\ 0 & 6 \end{bmatrix}$

(c) $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$

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

(e) $\begin{bmatrix} 1 & 1 & 2 \\ 1 & -2 & -1 \\ 1 & 1 & 2 \end{bmatrix}$

(f) $\begin{bmatrix} a & b & c \\ 0 & e & f \\ 0 & 0 & i \end{bmatrix}$

(a) This is not invertible—the columns are linearly dependent, violating one of the 10 conditions from last time.

(b) This is invertible—the columns are linearly independent and we can see 2 pivots, so two of the conditions are clearly met.

(c) This is invertible so long as neither a nor b are zero, for the same reason (we need them to be pivots).

(d) This is invertible so long as the columns aren't linearly dependent. Since there's only two of them, this would only happen if one was a multiple of the other. That is, $b = sa$, $d = sc$ for some s not zero. We could write that as $\frac{b}{a} = s = \frac{d}{c}$ (so long as a and c aren't zero), which we can see implies that $ad = bc$ is the condition for the columns being linearly dependent (this works even if a or c is zero—we could have derived it another way in that case). Rearranging, we could have said the condition is $ad - bc = 0$. Thus, the matrix *is* invertible precisely when $ad - bc \neq 0$.

(e) Here, the columns are again linearly dependent, although it's a bit harder to see—the third is the sum of the first two. This might suggest that the condition for a 3×3 is a little more complicated than the 2×2 condition from the previous part, but we can get a condition.

(f) This is invertible so long as a, e, and i are all nonzero—we want them to be pivots.

The Definition: We define the *determinant* of a matrix in general to be

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}).$$

In this definition, we're moving along the first row, taking (-1) to be one power higher than the column we're in (this means take a positive value for odd columns and a negative one for even columns), multiplying by the entry we find, then taking the determinant of the smaller matrix obtained by ignoring the top row and the column we're in. This involves the determinant of smaller matrices, so if we drill down enough layers, we'll get back to 2×2 matrices, where we can just use the formula $ad - bc$ from earlier.

1 Find the determinants for each of these matrices:

(a) $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 \\ 0 & 6 \end{bmatrix}$

(c) $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 4 \\ 1 & -1 & 2 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 1 \\ -1 & -2 & 1 \end{bmatrix}$

(f) $\begin{bmatrix} a & b & c \\ 0 & e & f \\ 0 & 0 & i \end{bmatrix}$

(a) This is a 2×2, so we can just use the formula, obtaining $6 - 3(2) = 0$.

(b) Here $6 - 0(2) = 6$.

(c) More generally, $ad - 0b = ad$.

(d) We get $(-1)^{1+1} 1 \begin{vmatrix} 3 & 4 \\ -1 & 2 \end{vmatrix} = 6 - (-4) = 10$.

(e) This is $1 \begin{vmatrix} 2 & 1 \\ -2 & 1 \end{vmatrix} - (-1) \begin{vmatrix} 0 & 1 \\ -1 & 1 \end{vmatrix} = 4 - (-1)(1) = 5$.

(f) More generally, $a \begin{vmatrix} e & f \\ 0 & i \end{vmatrix} - b \begin{vmatrix} 0 & f \\ 0 & i \end{vmatrix} + c \begin{vmatrix} 0 & e \\ 0 & 0 \end{vmatrix} = aei$. The same argument shows that the determinant of *any* triangular matrix can be found by multiplying the diagonal entries.

Cofactor Expansion: We can actually expand along *any* row or column. In that case, the (-1) has exponent $(-1)^{i+j}$ (where i marks the row and j the column we're in), the matrix entry is a_{ij}, and the subdeterminant is $\det(A_{ij})$. The goal here is to find the simplest row or column to move along to minimize the amount of computation. Mostly, this means finding the row or column with the most zeros, and the "nicest" (e.g., smallest) nonzero entries.

2 Try to compute the determinants of the following matrices by computing as few subdeterminants as possible.

(a) $\begin{bmatrix} 0 & 1 & 1 \\ 2 & 0 & -1 \\ -2 & 0 & 3 \end{bmatrix}$ (min is 1 subdeterminant)

(b) $\begin{bmatrix} 1 & 2 & 4 \\ 2 & 1 & 4 \\ 1 & 0 & -1 \end{bmatrix}$ (min is 2 subdeterminants)

(c) $\begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 3 & 9 \\ 0 & 0 & 1 & 4 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ (min is 3 subdeterminants if you use a result from problem 1, and 4 otherwise)

(d) $\begin{bmatrix} 2 & -1 & 0 & 3 \\ 7 & -2 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ -3 & -2 & 0 & 1 \end{bmatrix}$ (min is 2 subdeterminants)

For many of these, there are several ways to go which are equivalently easy, these are just some of them. When doing this kind of thing, just do what you find easiest.

(a) Going down the second column, we get $(-1)^{1+2}(1)\begin{vmatrix} 2 & -1 \\ -2 & 3 \end{vmatrix} = -4$

(b) Going across the third row, we get $\begin{vmatrix} 2 & 4 \\ 1 & 4 \end{vmatrix} + (-1)\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = 4 - (-3) = 7.$

(c) Going down the first column, we get $\begin{vmatrix} 1 & 3 & 9 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{vmatrix} - \begin{vmatrix} 2 & 4 & 8 \\ 1 & 3 & 9 \\ 0 & 1 & 4 \end{vmatrix} = 1 - \left(2\begin{vmatrix} 3 & 9 \\ 1 & 4 \end{vmatrix} - \begin{vmatrix} 4 & 8 \\ 1 & 4 \end{vmatrix}\right) = 1 - (6 - 8) = 3.$ We went down the first column in the 3×3.

(d) Going down the third column, we get $(-1)^{2+3}\begin{vmatrix} 2 & -1 & 3 \\ 1 & 0 & 0 \\ -3 & -2 & 1 \end{vmatrix} = (-1)\left((-1)^{2+1}\begin{vmatrix} -1 & 3 \\ -2 & 1 \end{vmatrix}\right) = 7.$ We went across the second row in the 3×3.

Determinants II: Return of Row Operations

The Punch Line: We can use row operations to calculate determinants if we're careful.

The Process: Our three row operations—interchange, scaling, and replacement with a sum—have predictable effects on the determinant. By tracking the operations we use to get a matrix that is easy to compute a determinant for—generally a matrix in echelon form (which is also in triangular form)—we can avoid most of the work involved. In particular, interchanging two rows multiplies the determinant by -1, scaling a row by k scales the determinant by k, and replacing a row with its sum with a multiple of a *different* row does not change the determinant. In practice, we mostly want to interchange rows and use scaled sums to get to Echelon Form (not necessarily reduced!), then multiply the diagonal entries.

1 Use row operations to help compute these determinants:

(a) $\begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix}$

(b) $\begin{vmatrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 0 & 2 & 4 & 6 \\ 0 & 0 & 1 & -1 \end{vmatrix}$

(c) $\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix}$

(d) $\begin{vmatrix} 1 & 2 & 3 & 0 & 4 \\ 12 & \sqrt{\pi} & e^{e^e} & 1 & \sqrt{1+\sqrt{1+\sqrt{1+\sqrt{2}}}} \\ 0 & 0 & 2 & 0 & 6 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{2}{3} \\ 0 & \frac{1}{2} & 0 & 0 & 1 \end{vmatrix}$

(a) A pair of interchanges (top and bottom, followed by the middle pair) yields the identity (with determinant 1), so the determinant here is $(-1)^2 \det(I) = 1$.

(b) Putting this in Echelon Form (*not* reduced) gives $\begin{bmatrix} 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & -3 \end{bmatrix}$, with no interchanges necessary, so the determinant is $(1)(1)(2)(-3) = -6$.

(c) Here an Echelon Form with no interchanges gives $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 0 & -2 \\ 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 4 \end{bmatrix}$, so the determinant is $(1)(-2)(-2)(4) = 16$.

(d) For this one, we probably want to start with a cofactor expansion down the fourth column, getting

$$\begin{vmatrix} 1 & 2 & 3 & 0 & 4 \\ 12 & \sqrt{\pi} & e^{e^e} & 1 & \sqrt{1+\sqrt{1+\sqrt{1+\sqrt{2}}}} \\ 0 & 0 & 2 & 0 & 6 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{2}{3} \\ 0 & \frac{1}{2} & 0 & 0 & 1 \end{vmatrix} = (-1)^{4+2} \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 2 & 6 \\ 0 & \frac{1}{2} & 0 & \frac{2}{3} \\ 0 & \frac{1}{2} & 0 & 1 \end{vmatrix}.$$

We can then interchange the second and third rows (at the cost of a negative sign) and subtract the new second row from the fourth (without changing the determinant to see that

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 2 & 6 \\ 0 & \frac{1}{2} & 0 & \frac{2}{3} \\ 0 & \frac{1}{2} & 0 & 1 \end{vmatrix} = (-1) \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & \frac{1}{2} & 0 & \frac{2}{3} \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & \frac{1}{3} \end{vmatrix} = -\frac{1}{3}.$$

Column Operations and Other Properties: Since $\det(A) = \det(A^T)$ (which takes a bit of argument to show), we can also do column operations analogous to the row operations, with the same effect on the determinant. Interspersing them can be helpful. Another useful property is that $\det(AB) = \det(A)\det(B)$ (although $\det(A+B)$ is often not $\det(A)+\det(B)$).

2 Find expressions for the following determinants (and justify them):

(a) $\det\left(A^2\right)$

(b) $\det(A^n)$

(c) $\det(BA)$

(d) $\det\left(A^{-1}\right)$

(e) $\det(kA)$ (where k is some real number)

(f) $\begin{vmatrix} A & O \\ O & B \end{vmatrix}$

In the last problem, A and B are standing for the entries of matrices A and B filling out those portions of the matrix, and O stands for zeros in those entries (so if A is $n\times n$ and B is $m\times m$, this matrix is $(n+m)\times(n+m)$. This is something of a challenge problem—I expect it's more abstract than most problems you'll be given.

(a) Since $\det(AB) = \det(A)\det(B)$, we have $\det\left(A^2\right) = \det(AA) = \det(A)\det(A) = \det(A)^2$.

(b) Repeating the above argument gives that $\det(A^n) = \det(A)^n$.

(c) We know that $\det(BA) = \det(B)\det(A)$. Since the determinant of a real matrix is just a real number, though, this is $\det(A)\det(B)$ (the determinants commute even if the matrices do not!).

(d) Since $\det(I) = 1$, and $I = AA^{-1}$, we get $\det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right) = 1$, which we can solve as $\det\left(A^{-1}\right) = \det(A)^{-1}$, as $\det(A) \neq 0$ so long as it's invertible.

(e) Since kA has *each* row multiplied by k, we see that $\det(kA) = k^n \det(A)$ (assuming we're in $\mathbb{R}^n$).

(f) If we use interchanges and add rows to each other to put this in echelon form, we see we get the pivot values of A along the upper part of the diagonal and of B along the lower part, so this determinant is $\det(A)\det(B)$.

Vector Spaces

The Punch Line: The same ideas we've been using translate to work on more abstract vector spaces, which describe many things which occur in "nature" (at least, in the mathematics we use to describe nature).

The Rules: A *vector space* is a set of objects V that satisfy these 10 axioms:

1. $\vec{x} + \vec{y} \in V$ (we say V is closed under addition)
2. $\vec{x} + \vec{y} = \vec{y} + \vec{x}$ (addition is commutative)
3. $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$ (addition is associative)
4. There is a $\vec{0}$ with the property that $\vec{x} + \vec{0} = \vec{x} = \vec{0} + \vec{x}$
5. There is a $\overrightarrow{-x}$ for every $\vec{x}$ so $\vec{x} + \overrightarrow{-x} = \vec{0}$
6. $c\vec{x} \in V$ (V is closed under scalar multiplication)
7. $c(\vec{x} + \vec{y}) = c\vec{x} + c\vec{y}$ (left distributivity)
8. $(c + d)\vec{x} = c\vec{x} + d\vec{x}$ (right distributivity)
9. $c(d\vec{x}) = (cd)\vec{x}$ (scaling is associative)
10. $1\vec{x} = \vec{x}$ (multiplicative identity)

1 Are these things vector spaces?

(a) The subset $\{\vec{0}\}$ in any $\mathbb{R}^n$

(b) $\mathbb{R}^2$ but scalar multiplication $c\vec{x}$ is defined as $c\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x/c \\ y/c \end{bmatrix}$.

(c) $\mathbb{R}^2$ but addition is defined as $\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} y + w \\ x + z \end{bmatrix}$

(d) All functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = 0$

(e) The set of all functions of the form $f(\theta) = a\sin(\theta + \phi)$

(f) The set of vectors $\vec{b}$ such that $\vec{b} = A\vec{x}$ (where A is fixed)

(a) Yes—adding only the zero vector and scaling the zero vector don't do anything to it, and the operations work like in $\mathbb{R}^n$.

(b) No—the only axiom which fails is right distributivity, because $\frac{1}{c+d} \neq \frac{1}{c} + \frac{1}{d}$ in general.

(c) No—this addition is not associative, but more interestingly there is no zero element (adding the zero vector switches the components).

(d) Yes—adding two functions which vanish at zero won't make a nonzero entry there, nor will scaling it, and the operations being well-behaved can be checked.

(e) Yes—the sum of sines is another sine with a different amplitude and phase (you may need to refresh yourself on some trig identities for this).

(f) Yes—if $\vec{b} = A\vec{x}$ and $\vec{b}' = A\vec{x}'$, then $\vec{b} + \vec{b}' = A\vec{x} + A\vec{x}' = A(\vec{x} + \vec{x}')$, and $c\vec{b} = cA\vec{x} = A(c\vec{x})$, and the other axioms are properties of $\mathbb{R}^n$.

Subspaces: A subset U of a vector space V is a *subspace* if it contains $\vec{0}$ and is closed under addition and scaling. A subspace is a vector space in its own right.

2 Are these subsets subspaces?

(a) The vectors in $\mathbb{R}^3$ whose entries sum to zero

(b) The vectors in $\mathbb{R}^2$ which lie on one of the axes

(c) The vectors in $\mathbb{R}^3$ that are mapped to zero by matrices A and B

(d) The functions of the form $f(\theta) = A\sin(\theta + \phi)$ with ϕ rational

(e) The functions of the form $f(\theta) = A\sin(\theta + \phi)$ with ϕ irrational

(f) The functions of the form $f(\theta) = A\sin(\theta + \phi)$ with A rational

(a) Yes—$\vec{0}$ is obviously in, and its quick to check that the sum of the entries of the sum of the two vectors is the sum of the sums of their entries, which is zero, and scaling zero leaves it at zero.

(b) No—the sum of two of them may be off the axis (consider $\vec{e}_1 + \vec{e}_2$).

(c) Yes—in class, it was shown the vectors mapped to zero by one matrix form a subspace, and if a vector is in both *nullspaces*, its multiples and sums with other vectors with that property will be in each one separately, and so be in both of them.

(d) Yes—by trig identities, the phase change depends on the input phases, and the sum of rational numbers is rational.

(e) No—consider the phases π and $\pi + 1$.

(f) No—consider scaling by e.

Why do we want sets to be vector spaces? In some sense, vector spaces all work in the same way, so if we can show that some set we're interested in is a vector space, we get to import all kinds of results "for free." We're taking something we know how to work with—$\mathbb{R}^n$—and leveraging it to get answers to things that are harder to deal with—like differential equations (see future math courses).

Column and Null Spaces

The Punch Line: The sets of vectors we've been most interested in so far in the course—solution sets (to homogeneous systems) and spans—are in fact subspaces!

Warm-Up: Can these situations happen?

(a) A vector $\vec{x}$ is in both the null space and column space of a 3×5 matrix

(b) A vector $\vec{x}$ is in both the null space and column space of a 2×2 matrix

(c) A vector $\vec{x}$ is in neither the null space nor column space of a 2×2 matrix

(d) A vector $\vec{x}$ is in neither the null space nor column space of an invertible 4×4 matrix

(a) No, the null space is a subspace of $\mathbb{R}^5$ and the column space is a subspace of $\mathbb{R}^3$—they live in different "universes."

(b) Yes—consider the matrix $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

(c) Yes—consider the matrix $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and the vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

(d) No—for an invertible matrix, the columns span $\mathbb{R}^n$, so the column space includes every vector.

Null Spaces: The *null space* (also called the *kernel*) of a linear transformation T in the vector space V is the set of all vectors $\vec{x}$ that are mapped to $\vec{0} \in V$ by T: $T(\vec{x}) = \vec{0}$. For $\mathbb{R}^n$ and $T(\vec{x}) = A\vec{x}$ for a matrix A, we can explicitly describe the vectors in the null space by finding a parametric form for the solution set of the homogeneous equation $A\vec{x} = \vec{0}$. The vectors attached to each parameter span the null space.

1 Describe the null spaces of the following linear transformations:

(a) $T(\vec{x}) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\vec{x}$

(b) $T(\vec{x}) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 1 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 \end{bmatrix}\vec{x}$

(c) $T(\vec{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\vec{x}$

(d) $T(\vec{x}) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\vec{x}$

(e) $T(\vec{x}) = \begin{bmatrix} 8 & 6 & 7 & 5 \\ 3 & 0 & 9 & 9 \end{bmatrix}\vec{x}$

(f) $T(\vec{x}) = \begin{bmatrix} 1 & -2 \\ 2 & -4 \\ -1 & 2 \\ 0 & 0 \end{bmatrix}\vec{x}$

(g) $T(f(x)) = f(x) - f(0)$ acting on the space of all $\mathbb{R} \to \mathbb{R}$ functions*

*This is something of a challenge problem; it should help you understand null spaces, but it probably won't be on an exam.

(a) We find the REF of the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, from which we read out the solution set as $\vec{x} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = s \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ (I've chosen $s = -x_2$ to make the solution set look a little nicer). So, the null space is Span$\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$.

(b) Here the REF is $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, so the solution set is $\vec{x} = x_4 \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \end{bmatrix}$, so the null space is Span$\left\{ \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \end{bmatrix} \right\}$.

(c) This matrix is invertible, so the solution set of the homogeneous equation is just the zero matrix, so the null space is the subspace consisting only of the zero vector.

(d) The matrix is already in REF, so we read off the solution set as $\vec{x} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Thus, the null space is Span$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$.

(e) The REF here is $\begin{bmatrix} 1 & 0 & 3 & 3 \\ 0 & 1 & -\frac{17}{6} & -\frac{19}{6} \end{bmatrix}$. Then the solution set is $\vec{x} = x_3 \begin{bmatrix} -3 \\ \frac{17}{6} \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ \frac{19}{6} \\ 0 \\ 1 \end{bmatrix} = s \begin{bmatrix} -18 \\ 17 \\ 6 \\ 0 \end{bmatrix} + t \begin{bmatrix} -18 \\ 19 \\ 0 \\ 6 \end{bmatrix}$. Thus, the null space is the two-dimensional space in $\mathbb{R}^4$ described as Span$\left\{ \begin{bmatrix} -18 \\ 17 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} -18 \\ 19 \\ 0 \\ 6 \end{bmatrix} \right\}$.

(f) The REF here is $\begin{bmatrix} 1 & -2 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$, so the solution set is $\vec{x} = x_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, so the null space is Span$\left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$.

(g) Here we can't appeal to a REF, because we don't have a matrix. But, we *can* consider that if $T(f(x)) = 0$, then $f(x) - f(0) = 0$, or $f(x) = f(0)$. This means that the null space of T here is the set of all functions which, for every x, give the same answer as at $x = 0$—that is, constant functions. Thus, the null space is the span of the constant function $f_1(x) = 1$.

Column Spaces and Range: The *column space* of a matrix is the span of its columns. For more general linear transformations, the analogous concept is *range*—the set of vectors in the vector space V that can be reached by applying the linear transformation. In $\mathbb{R}^n$, we can get the column space as just the span of the columns (although we can describe it more succinctly if we eliminate linearly dependent columns).

> 2 Describe the range of these linear transformations. What is their dimension? Try to find a spanning set with only that many vectors. See if you can relate these situations to the null spaces you found on the last page.
>
> (a) $T(\vec{x}) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \vec{x}$
>
> (b) $T(\vec{x}) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 1 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 \end{bmatrix} \vec{x}$
>
> (c) $T(\vec{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \vec{x}$
>
> (d) $T(\vec{x}) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \vec{x}$
>
> (e) $T(\vec{x}) = \begin{bmatrix} 8 & 6 & 7 & 5 \\ 3 & 0 & 9 & 9 \end{bmatrix} \vec{x}$
>
> (f) $T(\vec{x}) = \begin{bmatrix} 1 & -2 \\ 2 & -4 \\ -1 & 2 \\ 0 & 0 \end{bmatrix} \vec{x}$
>
> (g) $T(f(x)) = f(x) - f(0)$ acting on the space of all $\mathbb{R} \to \mathbb{R}$ functions*
>
> *This is again a challenge problem. What could the dimension be here?

(a) The range here is $\text{Span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$—we don't need to write it twice. This is a distinct subspace of $\mathbb{R}^2$ from the null space, and we can see that any vector in $\mathbb{R}^2$ has a part in the column space and a part in the null space (that is, $\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$ spans all of $\mathbb{R}^2$, so every vector is a linear combination of vectors from both subspaces). This is a nice property, and not one that always holds.

(b) Since columns one, two, and three are pivot columns, they are linearly independent (you can check this), so the column space is $\text{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 5 \\ 2 \end{bmatrix}\right\}$. It is three-dimensional (there are three pivot rows), and a little checking shows that the null space and the column space again don't intersect.

(c) The two columns are linearly independent, so their span is all of $\mathbb{R}^2$.

(d) The span of the columns is $\text{Span}\left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\} = \text{Span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$, which is one-dimensional. Note that this is precisely the null space of this transformation! The subspace $\text{Span}\left\{\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$ is *neither* in the null space nor column space of the matrix. This is a rather weird situation, but it shows that while the *sizes* of the null and column spaces match up to describe the size of the whole domain, they don't necessarily describe it themselves.

(e) The REF has two pivots, so the column space is all of $\mathbb{R}^2$. Just because there's a nontrivial null space doesn't mean there's anything missing from the range, if the domain is in a "bigger" (higher dimensional) vector space.

(f) Since the columns are linearly dependent, we only need one to describe the column space as $\text{Span}\left\{\begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}\right\}$, which is one-dimensional. This means that even though the null space is "small" compared to the target space, the range can be as well if the domain is lower-dimensional compared to the target space.

(g) We can see that any function with $f(0) = 0$ is unchanged by T, so it is in the range. If $f(0) \neq 0$, then $T(f(x))$ is zero at $x = 0$. This means that all and only functions which are zero at $x = 0$ are in the range of T.

What's going on with the linear transformation in part (d)? When (part of) the column space is in the null space, the matrix is sending vectors somewhere it will send to zero. If we applied the transformation twice (or, in general, enough times), it would send all vectors to zero. It's kind of a drawn-out process: send vectors matching some description (in some span) to zero, then change other vectors to take their places. It's important to remember that the null space is describing where vectors are *before* the transformation, while the column space is describing *after*.

Bases

The Punch Line: We have an efficient way to define subspaces using collections of vectors in them.

Warm-Up: Are these sets linearly independent? What do they span?

(a) $\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}1\\2\\2\end{bmatrix}\right\} \subset \mathbb{R}^3$

(b) All vectors in $\mathbb{R}^{42}$ with a zero in at least one component

(c) $\left\{\begin{bmatrix}2\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\-1\\1\end{bmatrix}, \begin{bmatrix}-2\\2\\-3\\2\end{bmatrix}\right\} \subset \mathbb{R}^4$

(d) $\left\{1, t-1, (t-1)^2 + 2(t-1)\right\} \subset \mathscr{P}_2$

(a) These are linearly independent. We can check this by examining the homogeneous linear equation $\begin{bmatrix}1 & 1 & 1\\1 & 1 & 2\\1 & 2 & 2\end{bmatrix}\vec{x} = \vec{0}$, or by observing that no linear combination of the first two can have different first and second component, no linear combination of the first and third can have different second and third components, and no linear combination of the last two can have all components the same. They span $\mathbb{R}^3$, as we can check by showing the inhomogeneous equation $\begin{bmatrix}1 & 1 & 1\\1 & 1 & 2\\1 & 2 & 2\end{bmatrix}\vec{x} = \vec{b}$ is consistent for all $\vec{b}$.

(b) This set is super dependent—it contains multiples of every vector in it. It spans $\mathbb{R}^42$, though, as it contains the vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_{42}$ which have a 1 in the position in their subscript and zeros elsewhere, and this subset spans $\mathbb{R}^42$.

(c) This set is linearly dependent—the last column is twice the second minus the first. It spans a plan in $\mathbb{R}^2$, in particular $s\begin{bmatrix}2\\0\\1\\0\end{bmatrix} + t\begin{bmatrix}0\\1\\-1\\1\end{bmatrix}$. This is because eliminating vectors linearly dependent on the rest doesn't change the span (because by assumption you can get at them with a linear combination of the rest), and the first two vectors are linearly independent.

(d) This set is linearly independent—the third entry is the only one with a t^2, so it is not dependent on either of the previous two, and they can't depend on it (there would be no way to eliminate that term). Clearly, the first two are linearly independent.

Bases: A *basis* for a vector space is a linearly independent spanning set. Every finite spanning set contains a basis by removing linearly dependent vectors, and many finite linearly independent sets may be extended to be a basis by adding vectors (if eventually this process terminates in a spanning set).

1 Are these sets bases for the indicated vector spaces? If not, can vectors be removed (which?) or added (how many?) to make it a basis?

(a) $\left\{\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}1\\1\\2\end{bmatrix},\begin{bmatrix}1\\2\\2\end{bmatrix}\right\}\subset\mathbb{R}^3$

(b) $\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}2\\1\end{bmatrix},\begin{bmatrix}3\\2\end{bmatrix}\right\}\subset\mathbb{R}^2$

(c) $\left\{\begin{bmatrix}2\\0\\1\\0\end{bmatrix},\begin{bmatrix}0\\1\\-1\\1\end{bmatrix},\begin{bmatrix}-2\\2\\-3\\2\end{bmatrix}\right\}\subset\mathbb{R}^4$

(d) $\left\{(t-1),(t-1)^2,(t-1)^3\right\}\subset\mathscr{P}_3$

(a) We saw previously that this set spans $\mathbb{R}^3$ and is linearly independent, so it is a basis.

(b) This set spans $\mathbb{R}^2$ (we could choose a coefficient on the second vector to match the second component of a given vector, then choose a coefficient on the first vector to match the first). It isn't linearly independent, though, as the second two vectors are linear combinations of the first. We could remove them to get a linearly independent set, hence a basis. In fact, no pair of vectors is linearly dependent, so any pair from this collection is a basis (we have to check that they span $\mathbb{R}^2$ on their own, but this is true in this case).

(c) We saw that this set is neither linearly independent nor spans $\mathbb{R}^4$. We could find a basis for $\mathbb{R}^4$ by removing the last vector (which is in the span of the first two) then adding in the vectors $\begin{bmatrix}1\\0\\0\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\\0\\0\end{bmatrix}$. The resulting collection $\left\{\begin{bmatrix}2\\0\\1\\0\end{bmatrix},\begin{bmatrix}0\\1\\-1\\1\end{bmatrix},\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\right\}$ is a basis, as you can check.

(d) This collection doesn't span $\mathscr{P}_3$ (you can verify that 1 is not in their span), but is linearly independent. We can simply add in 1, and verify that the result is indeed a basis by showing that if $p(x) = a + bx + cx^2 + dx^3$, then a linear combination of polynomials in this set yields p. Since this covers every polynomial in $\mathscr{P}_3$, we're good.

Finding Bases in $\mathbb{R}^n$: We're often interested in subspaces of the form Nul A and Col A for some matrix A. Fortunately, we can extract both by examining the Reduced Echelon Form of A.

A basis for Col A consists of all columns in A itself which correspond to pivot columns in the REF of A. A basis for Nul A consists of the vector parts corresponding to each free variable in a parametric vector representation of the solution set of the homogeneous equation $A\vec{x} = \vec{0}$, which we can find from the REF of A. Caution: In general, although free variables correspond to non-pivot columns in the REF, the basis for Nul A will *not* consist of those columns—in fact, they will often be of the wrong size!

2 Find bases for Nul A and Col A for each matrix below:

(a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & -1 \end{bmatrix}$

(d) $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ where $a \neq 0$

(a) This is in REF, so we identify the pivot columns, and see they are a basis for Col A, $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$. The parametric vector form for the solution set is $s \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, so Nul A has basis $\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$.

(b) This has REF $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$, so a basis for Col A is $\left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right\}$, and a basis for Nul A is $\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$. Note the different sizes of the vectors.

(c) The REF here is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$. The column space basis is $\left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \right\}$, while the null space is $\{\vec{0}\}$, so there is no linearly independent spanning set.

(d) If $ad - bc \neq 0$, then the REF is I_2, so a basis for the column space is $\left\{ \begin{bmatrix} a \\ c \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix} \right\}$, and the null space is $\{\vec{0}\}$. Otherwise, since $a \neq 0$ the REF is $\begin{bmatrix} 1 & b/a \\ 0 & 0 \end{bmatrix}$, so a basis for the column space is $\left\{ \begin{bmatrix} a \\ c \end{bmatrix} \right\}$ and for the null space $\left\{ \begin{bmatrix} -b \\ a \end{bmatrix} \right\}$.

Coordinates

The Punch Line: If we have a basis of n vectors for any vector space, we can describe (and work with) any vector from the space or equation in it as if it were in $\mathbb{R}^n$ all along!

Coordinate Vectors: If we have an *ordered* basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for vector space V, then any vector $v \in V$ has a unique representation

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n,$$

where each c_i is a real number. Then we can write the *coordinate vector* $[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$.

1 Find the representation of the given vector $\vec{v}$ with respect to the ordered basis $\mathcal{B}$.

(a) $\mathcal{B} = \left\{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right\}$, $\vec{v} = \begin{bmatrix} 8 \\ 0 \\ 5 \end{bmatrix}$

(b) $\mathcal{B} = \left\{1, t, t^2, t^3\right\}$, $\vec{v} = t^3 - 2t^2 + t$

(c) $\mathcal{B} = \left\{1, (t-1), (t-1)^2, (t-1)^3\right\}$, $\vec{v} = t^3 - 2t^2 + t$

(d) $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right\}$, $\vec{v} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$

(e) $\mathcal{B} = \left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right\}$,
$\vec{v} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$

(a) Here we have the second original component first, followed by the third original component, followed by the first original. Thus, $\begin{bmatrix} 8 \\ 0 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 8 \end{bmatrix}_{\mathcal{B}}$.

(b) Here, we get the coordinate vector $\begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{B}}$.

(c) We can rearrange our polynomial as $t^3 - 2t^2 + t = (t-1)^2 + (t-1)^3$, so its coordinates in this basis are $\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}_{\mathcal{B}}$.

(d) We can see that to match the middle component, we need $c_1 = -\frac{1}{2}$. This leaves $\begin{bmatrix} 3/2 \\ 0 \\ -3/2 \end{bmatrix}$, so $c_2 = \frac{3}{2}$ and $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 3/2 \end{bmatrix}_{\mathcal{B}}$. This raises the important point that *the number of entries in a coordinate vector depends on the length of the basis it relates to*, not *the original vector space*!

Change of Coordinates in $\mathbb{R}^n$: If we have a basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for $\mathbb{R}^n$, we can recover the standard representation by using the matrix P whose columns are the (ordered) basis elements represented in the standard basis:

$$P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}.$$

The matrix P^{-1} takes vectors in the standard encoding and represents them with respect to $\mathcal{B}$. Thus, if $\mathcal{C}$ is another basis for the same space and Q is the matrix bringing representations with respect to $\mathcal{C}$ to the standard basis, then $Q^{-1}P$ is a matrix which takes a vector encoded with respect to $\mathcal{B}$ and returns its encoding with respect to $\mathcal{C}$. That is,

$$[\vec{v}]_{\mathcal{C}} = Q^{-1}P\,[\vec{v}]_{\mathcal{B}}.$$

2 Compute the change of basis matrices for the following bases (into and from the standard basis).

(a) $\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ (b) $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ (c) $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ (d) $\left\{ \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\}$

(a) We have $P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, and $P^{-1} = P$ (which we can see as P just transposes the first and third components).

(b) We have $P = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, and $P^{-1} = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ (check this!).

(c) We have $P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $P^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

(d) We have $P = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$ and $P^{-1} = \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix}$.

3 Compute the change of basis matrices between the two bases:

(a) $\mathcal{B} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}, \mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$

(b) $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}, \mathcal{C} = \left\{ \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\}$

(a) The transition from encoding in $\mathcal{B}$ to $\mathcal{C}$ is given by $\begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$. Its inverse is $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$.

(b) The transition from $\mathcal{B}$ to $\mathcal{C}$ is given by $\begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ -3 & -7 \end{bmatrix}$. Its inverse is $\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 7 & 4 \\ -3 & -2 \end{bmatrix}$.

Dimension and Rank

The Punch Line: We can compare the "size" of different vector spaces and subspaces by looking at the size of their bases.

Warm-Up: Are these bases for the given vector space?

(a) $\left\{\begin{bmatrix}2\\5\end{bmatrix}, \begin{bmatrix}5\\-2\end{bmatrix}\right\}$ in $\mathbb{R}^2$

(b) $\{1, 1+t, t\}$ in $\mathscr{P}_1$

(c) $\left\{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}\right\}$ in the vector space (you can check it is one) of all 2×2 matrices

(d) $\left\{1, t, t^2, \ldots, t^n\right\}$ for some fixed n in the space of all polynomials

(a) Yes, they are linearly independent, and span $\mathbb{R}^2$.

(b) No, they are not linearly independent—$1+t$ is the sum of two other basis elements.

(c) No, they do not span the space of all matrices—any matrix with a nonzero bottom right entry isn't in their span.

(d) No, because t^{n+1} is not in their span.

Dimension: If one basis for a vector space V has n vectors, then all others do. We can see this by writing the other basis' coordinates with respect to the first basis, then looking at the Reduced Echelon Form of this matrix—there can't be any free variables, and there must be n pivots, so there must be n vectors in the new basis.

1 Find the dimension for each of the following subspaces.

(a) $\text{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$

(b) $\text{Span}\left\{1-t+t^2, 1+t-2t^2, t^2-t, t^3-t, t^3-t^2\right\}$

(c) $\text{Nul}\begin{bmatrix}1 & 1 & 1\\1 & -1 & 1\\-1 & 1 & -1\end{bmatrix}$

(d) $\text{Col}\begin{bmatrix}1 & 2 & 3\\0 & 1 & 2\\-1 & 0 & 1\end{bmatrix}$

(a) These are linearly independent, and clearly span their span, so the span has a basis of size 2, so is of dimension 2.

(b) We look at the coordinates of this with respect to the basis $\left\{1, t, t^2, t^3\right\}$. This gives the vectors $\left\{\begin{bmatrix}1\\-1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\-2\\0\end{bmatrix}, \begin{bmatrix}0\\-1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\-1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\0\\-1\\1\end{bmatrix}\right\}$.

We examine the matrix $\begin{bmatrix}1 & 1 & 0 & 0 & 0\\-1 & 1 & -1 & -1 & 0\\1 & -2 & 1 & 0 & -1\\0 & 0 & 0 & 1 & 1\end{bmatrix}$, and find that it has REF $\begin{bmatrix}1 & 0 & 0 & 0 & 0\\0 & 1 & 0 & 0 & 0\\0 & 0 & 1 & 0 & -1\\0 & 0 & 0 & 1 & 1\end{bmatrix}$. Thus, the first four vectors are linearly independent and span the whole span, so they are a basis. Thus, this span has dimension 4 (and is therefore all of $\mathscr{P}_3$).

(c) The REF here is $\begin{bmatrix}1 & 0 & 1\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$, so there is one free variable, so the null space has dimension 1.

(d) The REF here is $\begin{bmatrix}1 & 0 & -1\\0 & 1 & 2\\0 & 0 & 0\end{bmatrix}$, which has two pivots, so the column space has dimension 2.

Rank of a Matrix: For any $n \times m$ matrix A, the dimension of the null space is the number of free variables and the dimension of the column space is the number of pivots. These add up to m, the number of columns (a column is either a pivot or corresponds to a free variable). We call the dimension of the column space the *rank* of a matrix.

2 Find the ranks of the following matrices:

(a) $\begin{bmatrix} 1 & 3 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -2 \\ 1 & 0 & 0 & 6 \end{bmatrix}$

(d) An invertible $n \times n$ matrix

(a) This matrix clearly has two pivots, so the column space will have dimension 2, so this is the rank of the matrix.

(b) We can see that these columns are all linearly independent, so form a basis for their span, so the rank is 3.

(c) We can see through interchanging rows that this matrix has 3 pivots, hence has rank 3.

(d) One equivalent property to being invertible is having no free variables, which means that $\dim \text{Nul} A = 0$, so the dimension of the column space must be the number of columns n, so this must be the rank.

The statement that rank A + dim Nul A is the number of columns of A is an important theorem known as the Rank Nullity Theorem (some people call dim Nul A the *nullity* of A). It is basically saying that the input space to A has only two important parts: the null space, and the vectors which contain the information for knowing what the column space looks like. There's a bit more to it than that, but the gist is there isn't some third kind of vector lurking around that isn't related to either the null or column spaces.

Eigenvalues

The Punch Line: If we have a linear transformation from an n-dimensional vector space to itself, we can choose a basis that makes the matrix of the linear transformation especially simple—characterized by just n constants.

Warm-Up: Are the following matrices invertible? If not, what is the dimension of their null space?

(a) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 4 \\ 0 & 0 & 4 \end{bmatrix}$

(e) $\begin{bmatrix} -1 & 2 & 4 \\ 0 & 0 & 4 \\ 0 & 0 & 2 \end{bmatrix}$

(f) $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$

(a) Invertible—its columns are clearly linearly independent.

(b) Invertible—again, the columns are linearly independent.

(c) Not invertible—the columns are the same. The null space has dimension 1—it is Span$\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$.

(d) Invertible—the determinant is $8 \neq 0$.

(e) Not invertible—the determinant is 0. The null space has dimension 1: we can see that there will be two pivots in the REF of the matrix, so it will have rank 2, hence by the Rank-Nullity Theorem ($\text{rank}A + \dim \text{Nul}A = n$ for an $m \times n$ matrix), $\dim \text{Nul}A = 1$.

(f) Not invertible—the determinant is 0. The null space will have dimension 1 here, as it consists of all vectors with 0 as their second component (check this). It might be a good idea to keep this matrix in mind as we start exploring how to find the eigenvalues of a matrix—this one doesn't have "enough" eigenvectors (we'll see what this means later).

Eigenvalues and Eigenvectors: If V is an n-dimensional vector space and T is a linear transformation from V back into itself, and we find a (nonzero) vector $\vec{v} \in V$ and $\lambda \in \mathbb{R}$ that make the equation $T(\vec{v}) = \lambda\vec{v}$, we call λ and *eigenvalue* and $\vec{v}$ an *eigenvector* for T. The eigenvectors of the linear transformation are vectors whose direction does not change when you apply the transformation (except possibly reversing if $\lambda < 0$). The eigenvalues of the linear transformation are the different "scaling factors" that the transformation uses (as well as containing information about whether it reverses direction of certain vectors).

1 Are these vectors eigenvectors of the given linear transformation? If so, what are their eigenvalues?

(a) $\vec{v} = t^2 \in \mathscr{P}_2$ with $T(p(t)) = t\frac{d}{dt}[p(t)]$

(b) $\vec{v} = t \in \mathscr{P}_2$ with $T(p(t)) = t\frac{d}{dt}[p(t)]$

(c) $\vec{v} = 1 \in \mathscr{P}_2$ with $T(p(t)) = t\frac{d}{dt}[p(t)]$

(d) $\vec{v} = \begin{bmatrix}1\\1\end{bmatrix}$ with $T(\vec{x}) = \begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\vec{x}$

(e) $\vec{v} = \begin{bmatrix}1\\0\end{bmatrix}$ with $T(\vec{x}) = \begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\vec{x}$

(f) $\vec{v} = \begin{bmatrix}2\\3\end{bmatrix}$ with $T(\vec{x}) = \begin{bmatrix}1 & -5/3\\0 & -3/2\end{bmatrix}\vec{x}$

(g) $\vec{v} = \begin{bmatrix}0\\0\end{bmatrix}$ with $T(\vec{x}) = \vec{0}$

(a) Yes, with eigenvalue 2: $T(t^2) = t(2t) = 2t^2$.

(b) Yes, with eigenvalue 1: $T(t) = t(1) = t$.

(c) Yes, with eigenvalue 0: $T(1) = t(0) = 0$.

(d) No, because $\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$.

(e) No, because $\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$.

(f) Yes, because $\begin{bmatrix}1 & -5/3\\0 & -3/2\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}2-\frac{5}{3}(3)\\-\frac{3}{2}(3)\end{bmatrix} = \begin{bmatrix}-3\\-\frac{9}{2}\end{bmatrix} = -\frac{3}{2}\begin{bmatrix}2\\3\end{bmatrix}$. (Sorry for the numbers on this one, but I wanted you to see that eigenvectors can be a little messy sometimes.)

(g) No, because in our definition we insist that eigenvectors be nonzero. In fact, $\vec{0}$ is the *only* vector that isn't an eigenvector of this T: all the other vectors are eigenvectors with eigenvalue 0. There's a good reason for excluding $\vec{0}$ as an eigenvector—we want the set of eigenvalues that a matrix has to be a description of what it does, and $\vec{0}$ would be an eigenvector for *every* eigenvalue.

Eigenspaces: The set of eigenvectors for eigenvalue λ of a given linear transformation is almost a subspace—all it's missing is the zero vector, which we may as well add in (after all, it also satisfies the equation $T(\vec{v}) = \lambda\vec{v}$). This means we can find a subspace corresponding to each eigenvalue of the linear transformation—we call it the *eigenspace for eigenvalue* λ, and denote it E_λ.

This is the null space of a linear transformation which is a slight modification of the original: if our transformation had matrix A, then E_λ is the null space of $A-\lambda I_n$. This is because if $(A-\lambda I_n)\vec{v} = 0$, then $A\vec{v}-\lambda I_n\vec{v} = A\vec{v}-\lambda\vec{v} = \vec{0}$, or $A\vec{v} = \lambda\vec{v}$. This means that if we know λ is an eigenvalue of the transformation, we can find its eigenspace by using techniques we already know for describing null spaces! We can also prove some number $\mu \in \mathbb{R}$ is *not* an eigenvalue by showing that its "eigenspace" (the null space of $A - \mu I_n$) is just $\left\{\vec{0}\right\}$, which isn't an eigenvector.

2 Determine if λ is an eigenvalue for the (transformation given by the) matrix A by computing E_λ:

(a) $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \lambda = 1$

(b) $A = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}, \lambda = 2$

(c) $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \lambda = 1$

(d) A is an invertible matrix, $\lambda = 0$

(a) We look at $A - 1I_2 = \begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix}$. The REF of this matrix is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, so there is only the trivial solution to the homogeneous equation for it, so the space $E_1 = \left\{\vec{0}\right\}$, so 1 is not an eigenvalue.

(b) We look at $A - 2I_2 = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix}$. This has REF $\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$, hence the homogeneous equation has parametric solution

$$\vec{x} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

This means the null space of $A-2I_2$ is Span $\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$, so any vector with both components equal and nonzero is an eigenvector of A with eigenvalue 2.

(c) We look at $A - I_3 = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}$. The REF of this matrix is $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, so the null space is Span $\left\{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$. Thus, $E_1 = \text{Span}\left\{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$.

(d) The eigenspace E_0 is the null space of $A - 0I_n = A$. If A is invertible, though, this means that its null space consists only of the zero vector. This shows that 0 is not an eigenvalue of any invertible matrix (or any invertible linear transformation in general)!

Under the Hood: Eigenspaces are a very interesting and important class of *invariant subspaces*—subspaces that are preserved by the transformation, in that any vector in the subspace will be mapped to another vector in the same subspace. The action of the transformation is very simple in the eigenspaces, so this is a huge win—if we can break any vector up into pieces that are all in eigenspaces, we can describe what happens to it just by seeing how each of those pieces gets scaled by the appropriate eigenvalue.

As it turns out, eigenspaces tell most of the story of invariant subspaces. There are only two other kinds in $\mathbb{R}^n$: spaces where the transformation looks like a rotation rather than a scaling, and spaces where the transformation "eventually" works like a scaling (after you apply it enough times). In this course, though, we'll just be focused on eigenspaces as they're presented here (ask me if you want to know more, or take Math 108).

Eigenvalues and Where to Find Them

The Punch Line: Finding the eigenvalues of a matrix boils down to finding the roots of a polynomial.

Warm-Up: What are the eigenvalues of these matrices? What is the dimension of each eigenspace?

[Note: you shouldn't have to do many computations here—just look at Echelon Forms and try to see how many free variables there will be.]

(a) $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 2 & 4 & 8 \\ 0 & 0 & 2 & 8 \\ 0 & 0 & 0 & 4 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 2 & 0 & 8 \\ 0 & 0 & 2 & 8 \\ 0 & 0 & 0 & 4 \end{bmatrix}$

(d) $\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$

(e) $\begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$

(f) $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(a) We can see that the only value of λ for which $A - \lambda I_3$ has a non-trivial null space is $\lambda = 2$. In this case, $A - 2I_3 = O$ (the matrix of all zeroes), so the null space is all of $\mathbb{R}^3$, which is three-dimensional. This means that E_2, the 2-eigenspace, has dimension three, and thus that A acts everywhere as if it were the scalar 2 (in some sense, this kind of matrix—with the same value along the diagonal and zeroes everywhere else—is the simplest possible).

(b) The eigenvalues here are $\lambda = 1$, 2, and 4. For 1 and 4, we can see $A - \lambda I_4$ will have three pivots remaining, so the eigenspaces will have dimension 1 (three pivots means one free variable here). Looking at $A - 2I_4$, we see it is

$$A - 2I_4 = \begin{bmatrix} -1 & 2 & 4 & 8 \\ 0 & 0 & 4 & 8 \\ 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

The first, third, and fourth columns are clearly pivot columns here, so E_2 also has dimension 1, even though 2 appears twice on the diagonal.

(c) The eigenvalues are again $\lambda = 1$, 2, and 4. Again, $\dim E_0 = \dim E_4 = 1$, but this time

$$A - 2I_4 = \begin{bmatrix} -1 & 2 & 4 & 8 \\ 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 2 \end{bmatrix},$$

which only has two pivots. Thus, $\dim E_2 = 2$ for this matrix.

(d) The eigenvalues are $\lambda = a$, b, and c. Each will have a one-dimensional eigenspace, unless two or all three are the same, in which case that eigenvalue will have a two- or three-dimensional eigenspace.

(e) The only eigenvalue is $\lambda = a$. If $b = 0$, then E_a has dimension 2, otherwise it has dimension 1.

(f) The only eigenvalue here is $\lambda = 0$. There are three pivots, so E_0 has $5 - 3 = 2$ dimensions.

The Characteristic Equation: If λ is an eigenvalue of the matrix A, that means there is some nonzero $\vec{v} \in \mathbb{R}^n$ that satisfies the equation $A\vec{v} = \lambda\vec{v}$. Then $(A - \lambda I_n)\vec{v} = \vec{0}$ (from putting all terms with $\vec{v}$ on the same side), so $(A - \lambda I_n)$ is a non-invertible matrix (it has nontrivial null space, because $\vec{v} \neq \vec{0}$). Since we know that a matrix being not invertible is equivalent to its determinant being zero, we can check when the equation $\det(A - \lambda I_n) = 0$ is true. This gives a polynomial equation in λ of degree n (why?), so if we can find the roots of the polynomial, we know all of the eigenvalues. This equation is known as the *characteristic equation*.

1 What is the characteristic equation for each of these matrices?
[Note: You need not solve the characteristic equation yet.]

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} -3/2 & 1/2 \\ 1/2 & -3/2 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \end{bmatrix}$

(f) $\begin{bmatrix} -1 & 2 \\ -1 & 1 \end{bmatrix}$

(a) We look at $A - \lambda I_2 = \begin{bmatrix} 1-\lambda & 0 \\ 0 & 2-\lambda \end{bmatrix}$. This has determinant $(1-\lambda)(2-\lambda)$, so our characteristic equation is $(1-\lambda)(2-\lambda) = 0$. This is already factored, so we can confirm it gives eigenvalues 1 and 2.

(b) Here we have $A - \lambda I_2 = \begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix}$, so the characteristic equation is $(-\lambda)(-\lambda) - (1)(1) = \lambda^2 - 1 = 0$.

(c) Here we similarly have $A - \lambda I_3 = \begin{bmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{bmatrix}$. Taking the determinant and setting it to zero gives the characteristic equation

$$-\lambda \begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} - \begin{vmatrix} 1 & 0 \\ 1 & -\lambda \end{vmatrix} = -\lambda(\lambda^2 - 1) + \lambda = -\lambda^3 + 2\lambda = 0.$$

(d) Here we get $\left(-\frac{3}{2} - \lambda\right)^2 - \left(\frac{1}{2}\right)^2 = 0$. Note that if we wanted to solve it, we could multiply this out and then factor, but since we have a difference of squares we could just write $\left(-\frac{3}{2} - \lambda + \frac{1}{2}\right)\left(-\frac{3}{2} - \lambda - \frac{1}{2}\right) = (1-\lambda)(2-\lambda) = 0$ from here. Be careful about "simplifying" an equation—always keep in mind the most useful form of it (for us, factored)! Also, note that this has the same characteristic equation as the matrix in (a), despite being a very different matrix.

(e) Successive cofactor expansions along the middle three rows of

$$\det(A - \lambda I_5) = \begin{vmatrix} -\lambda & 0 & 0 & 0 & 2 \\ 0 & -\lambda & 0 & 0 & 0 \\ 0 & 0 & -\lambda & 0 & 0 \\ 0 & 0 & 0 & -\lambda & 0 \\ 2 & 0 & 0 & 0 & -\lambda \end{vmatrix}$$

gives the determinant as $(-\lambda)^3 \begin{vmatrix} -\lambda & 2 \\ 2 & -\lambda \end{vmatrix} = (-\lambda)^3 \left((-\lambda)^2 - (2)^2\right) = (-\lambda)^3 \left(\lambda^2 - 4\right)$. Thus, our characteristic equation is $-\lambda^3 \left(\lambda^2 - 4\right) = 0$.

(f) Here we get $\begin{vmatrix} -1-\lambda & 2 \\ -1 & 1-\lambda \end{vmatrix} = (-1-\lambda)(1-\lambda) - (2)(-1) = \lambda^2 + 1$.

2 What are the eigenvalues of these matrices? What are the dimensions of each eigenspace?

[Note: Again, try to minimize computation—we're not after the eigenspace itself, just its dimension, so you only need to manipulate the matrix into an Echelon Form matrix, not fully solve for its null space.]

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} -3/2 & 1/2 \\ 1/2 & -3/2 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \end{bmatrix}$

(f) $\begin{bmatrix} -1 & 2 \\ -1 & 1 \end{bmatrix}$

(a) The characteristic equation we computed was $(1-\lambda)(2-\lambda)$, so the eigenvalues are 1 and 2. We can see that each will have a one-dimensional eigenspace (there will be one pivot after subtracting the appropriate multiple of the identity).

(b) The characteristic equation we got above was $\lambda^2 - 1 = (\lambda-1)(\lambda+1) = 0$, so the eigenvalues are $\lambda = \pm 1$. We can see that $A \mp I_2$ will have rank 1 (that is, one pivot), so each eigenspace is again of dimension 1.

(c) The characteristic equation we got was $-\lambda^3 + 2\lambda = -\lambda(\lambda^2 - 2) = 0$, so the eigenvalues are $-\sqrt{2}$, 0, and $\sqrt{2}$. Again, we'll find each has a one-dimensional eigenspace.

(d) This has the same characteristic equation as part (a), so also has 1 and 2 as eigenvalues, and we can check that it also has only a one-dimensional eigenspace for each.

(e) Our characteristic polynomial here is $-\lambda^3\left(\lambda^2 - 4\right) = -\lambda^3(\lambda-2)(\lambda+2) = 0$. We thus have eigenvalues 0 and ± 2. We can check that we'll get 4 pivots for both positive and negative two (the three middle rows will be linearly independent, and the top and bottom will be multiples of each other), so they both have one-dimensional eigenspaces. The matrix for eigenvalue 0 clearly has two pivots (the first and fifth columns are the only nonzero ones, and are clearly linearly independent), so has a three-dimensional eigenspace (there are three free variables).

(f) Our characteristic polynomial is $\lambda^2 + 1 = 0$, which has no real solutions, so there are no real eigenvectors.

Under the Hood: You may have noticed we often got one-dimensional eigenspaces. If we know something is an eigenvalue, we know that its eigenspace is *at least* one dimensional, and the eigenspaces for different eigenvalues are distinct except for the zero vector (otherwise A would act on a vector in both by scaling by the different eigenvalues, which would give two different answers!). Thus, if we have n distinct eigenvalues for a matrix in $\mathbb{R}^n$, we know we have found n distinct subspaces, each of which is at least one-dimensional. This means they *have* to be one-dimensional, otherwise $\mathbb{R}^n$ would have more than n dimensions! As it turns out, the characteristic equation gives us information on the maximum size of each eigenspace, through the *multiplicity* of each root (how many times it appears).

Diagonalization

The Punch Line: Eigenvalues and -vectors can be used to factor a matrix in a way that makes computation easier.

Warm-Up What are the eigenvalues of these matrices? What are their eigenspaces?

(a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ (f) $\begin{bmatrix} -1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & -1 \end{bmatrix}$

(a) This has eigenvalues $\lambda = 1, 2, 3$, with eigenspaces $\text{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right\}, \text{Span}\left\{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}, \text{Span}\left\{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}$, respectively.

(b) This has eigenvalue $\lambda = 1, 2, 3$, with eigenspaces $\text{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right\}, \text{Span}\left\{\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}\right\}, \text{Span}\left\{\begin{bmatrix} 9/2 \\ 3 \\ 1 \end{bmatrix}\right\}$, respectively.

(c) This has eigenvalues $\lambda = -1, 1$, with eigenspaces $\text{Span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}, \text{Span}\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$, respectively.

(d) The characteristic equation here is

$$\begin{vmatrix} 1-\lambda & 2 \\ 3 & 6-\lambda \end{vmatrix} = (1-\lambda)(6-\lambda) - 6 = \lambda^2 - 7\lambda = 0.$$

Thus, the eigenvalues are $\lambda = 0, 7$, with eigenspaces $\text{Span}\left\{\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right\}, \text{Span}\left\{\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right\}$, respectively.

(e) This has only the eigenvalue $\lambda = 3$, with eigenspace $\text{Span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$.

(f) We find the characteristic equation

$$\begin{vmatrix} -1-\lambda & -2 & 1 \\ -2 & 2-\lambda & -2 \\ 1 & -2 & -1-\lambda \end{vmatrix} = (-1-\lambda)\begin{vmatrix} 2-\lambda & -2 \\ -2 & -1-\lambda \end{vmatrix} - (-2)\begin{vmatrix} -2 & 1 \\ -2 & -1-\lambda \end{vmatrix} + \begin{vmatrix} -2 & 1 \\ 2-\lambda & -2 \end{vmatrix} = 16 + 12\lambda - \lambda^3 = 0.$$

We can factor this polynomial as $(4-\lambda)(2+\lambda)^2 = 0$, so our eigenvalues are $\lambda = -2, 4$. Checking their eigenspaces yields $\text{Span}\left\{\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}, \text{Span}\left\{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\right\}$, respectively.

Diagonalizing: If the matrix A has eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and eigenvectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ corresponding to them, then we write $P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}$ for the matrix whose columns are the eigenvectors and D for the matrix with the eigenvalues down the diagonal and zeroes elsewhere. Then

$$AP = A\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \cdots & \lambda_n\vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = PD.$$

If the eigenvectors are linearly independent, then P is invertible, and $A = PDP^{-1}$.

1 Are these matrices diagonalizable? If so, what are P and D?

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

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

(c) $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$

(e) $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$

(f) $\begin{bmatrix} -1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & -1 \end{bmatrix}$

(a) Yes, it is already diagonal! Here $P = I_3$ and D is the original matrix.

(b) Yes—we've found the eigenvalues and vectors in the warm-up, so $P = \begin{bmatrix} 1 & 2 & 9/2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

(c) Yes—$P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $D = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. Of course, we could also have chosen $P = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$—the order doesn't matter so long as the same order is used for the eigenvectors and -values.

(d) Yes again. Here $P = \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}$ and $D = \begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix}$. Note that P is invertible, while D and our original matrix are not. Diagonalizability is a different condition from invertibility—non-invertible matrices like this one can be diagonalizable, and invertible matrices can be non-diagonalizable!

(e) This matrix is not diagonalizable—there is only one eigenvector, so we can't make an invertible matrix P out of eigenvectors! Note that this matrix is invertible—the inverse is $\frac{1}{9}\begin{bmatrix} 3 & -1 \\ 0 & 3 \end{bmatrix}$—which shows off a sort of essential form of non-diagonalizable matrices (ask me if you're curious about this, there's some interesting stuff going on, but it's outside the scope of this course).

(f) This is diagonalizable! We have $P = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & -2 \\ 0 & 1 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$. Again, we could have chosen P and D differently—in this case, it's worth noting that we could have chosen $P = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 4 & -2 \\ 1 & 5 & 1 \end{bmatrix}$ with the same D, because we can take *any* two linearly independent eigenvectors from the two-dimensional eigenspace E_{-2}.

Using the Diagonalization: If we have written $A = PDP^{-1}$ with D a diagonal matrix, then we can easily compute the kth power of A as $A^k = PD^kP^{-1}$ (adjacent P and P^{-1} matrices will cancel, putting all of the D matrices together and just leaving the ones on the end).

2 The Fibonacci numbers are a *very* famous sequence of numbers. The first one is $F_1 = 0$, the second is $F_2 = 1$, and from then on out, each number is the sum of the previous two $F_n = F_{n-1} + F_{n-2}$ (this is sometimes used as a simple model for population growth—although it assumes immortality). Since it's annoying to compute F_n if n is very large (we'd have to do a lot of backtracking to get to known values), it would be nice to have a closed form for F_n. We can derive one with the linear algebra we already know!

(a) Since the equation defining F_n in terms of F_{n-1} and F_{n-2} is linear, we can use a matrix equation to represent the situation. In particular, we want a matrix A such that

$$A\begin{bmatrix} F_{n-1} \\ F_{n-2} \end{bmatrix} = \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix},$$

so that we can keep applying A to get further along in the sequence. What is this A?

(b) Since $\begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = A\begin{bmatrix} F_{n-1} \\ F_{n-2} \end{bmatrix} = A^2\begin{bmatrix} F_{n-2} \\ F_{n-3} \end{bmatrix} = \cdots$, we can find F_n by computing $A^{n-2}\begin{bmatrix} F_2 \\ F_1 \end{bmatrix} = A^{n-2}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (we only need to advance by $n-2$ steps, because the top entry starts at 2). It's easier to raise matrices to powers after we diagonalize them, so find an invertible P and diagonal D so that $A = PDP^{-1}$ (the numbers are a little gross, so don't be alarmed).

(a) We want $A\begin{bmatrix} F_{n-1} \\ F_{n-2} \end{bmatrix} = \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \begin{bmatrix} F_{n-1} + F_{n-2} \\ F_{n-1} \end{bmatrix}$, so we can find the matrix of the linear transformation

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

(b) The characteristic equation of A is

$$\begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = (1-\lambda)(-\lambda) - 1 = \lambda^2 - \lambda - 1 = 0.$$

Using the quadratic formula on this, we see that the eigenvalues are $\lambda = \frac{1\pm\sqrt{5}}{2}$ (the positive root is the Golden Ratio φ, while the negative root is sometimes denoted $\overline{\varphi}$). Looking at $A - \varphi I_2 = \begin{bmatrix} -\frac{\sqrt{5}}{2} & 1 \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix}$, we see it has eigenvector $\begin{bmatrix} \sqrt{5} \\ 2 \end{bmatrix}$. Similarly, an eigenvector for $\overline{\varphi}$ is $\begin{bmatrix} -\sqrt{5} \\ 2 \end{bmatrix}$. Thus, we can write

$$A = \begin{bmatrix} \sqrt{5} & -\sqrt{5} \\ 2 & 2 \end{bmatrix}\begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}\frac{1}{20}\begin{bmatrix} 2 & \sqrt{5} \\ -2 & \sqrt{5} \end{bmatrix}.$$

2 cont.

(c) Since $A^k = PD^kP^{-1}$, we can write out F_n as the first component of $PD^{n-2}P^{-1}\begin{bmatrix}F_2\\F_1\end{bmatrix} = PD^{n-2}P^{-1}\begin{bmatrix}1\\0\end{bmatrix}$ (if we wanted to be clever, we could write this as

$$F_n = \begin{bmatrix}1 & 0\end{bmatrix} PD^{n-2}P^{-1}\begin{bmatrix}1\\0\end{bmatrix},$$

as the row vector picks out the first component). Use this to write down a formula for F_n (don't worry about multiplying out powers of any terms involving square roots, just leave them as whatever they are)! Nifty!!!

(c) Since $P^{-1}\begin{bmatrix}1\\0\end{bmatrix} = \frac{1}{20}\begin{bmatrix}2\\-2\end{bmatrix} = \begin{bmatrix}1/10\\-1/10\end{bmatrix}$ and $D^{n-2} = \begin{bmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} & 0\\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\end{bmatrix}$, we get

$$F_n = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\sqrt{5} & -\sqrt{5}\\ 2 & 2\end{bmatrix}\begin{bmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} & 0\\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\end{bmatrix}\begin{bmatrix}1/10\\-1/10\end{bmatrix} = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\frac{\sqrt{5}}{10}\left(\varphi^{n-2}+\overline{\varphi}^{n-2}\right)\\ \frac{2}{10}\left(\varphi^{n-2}-\overline{\varphi}^{n-2}\right)\end{bmatrix} = \frac{\varphi^{n-2}+\overline{\varphi}^{n-2}}{2\sqrt{5}}.$$

Under the Hood: The right way to think about the matrices P and P^{-1} is as change-of-coordinates matrices to an *eigenbasis*—then the requirement for diagonalizability is that the eigenvectors of A form a basis for the space they're in. Essentially, what we're doing is choosing a clever basis so that A looks like a diagonal matrix in that basis.

Inner Products, Length, and Orthogonality

The Punch Line: We can compute a real number relating two vectors—or a vector to itself—that gives information on both length and angle.

Warm-Up What are the lengths of these vectors, as found geometrically (using things like the Pythagorean Theorem)?

(a) $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$

(b) $\begin{bmatrix} 0 \\ -2 \end{bmatrix}$

(c) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(d) $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$

(e) $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$

(f) $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$

(a) This vector is along one of the axes, so we can see that it's length is 3.

(b) Similarly for this one; although it is in the negative y direction, its length is positive 2.

(c) We can think of the vector as the sum of its x and y components, and because these are perpendicular, we can use the Pythagorean Theorem to see that the square of the length of the vector is the sum of the squares of the lengths of the legs. For this vector, it means the square of the length is the sum of $1^2 + 1^2 = 2$, so the length itself is $\sqrt{2}$. This should be familiar as the length of the diagonal of a square of side length 1.

(d) Here we get a right triangle formed by the vector and its x and y components. It's a Pythagorean Triple, so we see that the length squared is $3^2 + 4^2 = 5^2$, or the length is 5.

(e) Here we note that one leg is of length 1 and one of length 2, so the total length is $\sqrt{1^2 + 2^2} = \sqrt{5}$. Note again that we are only using positive lengths.

(f) Here we can work in two stages: the "footprint" in the xy plane is $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, with length $\sqrt{2}$ as before. This then forms one leg of the vector, with the other leg being the z component. Then the length is $\sqrt{\left(\sqrt{2}\right)^2 + 3^2} = \sqrt{11}$. This is a rather cumbersome way to get the length, so we'd like to find a way to do it all in one step.

The Inner Product: If we think about a vector $\vec{v} \in \mathbb{R}^n$ as a $n \times 1$ matrix (a single column), then $\vec{v}^T$ is a $1 \times n$ matrix (a single row, sometimes called a row vector). Then we can multiply $\vec{v}^T$ against a vector (on the left) to get a 1×1 matrix, which we can consider a scalar. This is the idea behind the *inner product* in $\mathbb{R}^n$, also called the *dot product*: we take two vectors, $\vec{u}$ and $\vec{v}$, and define their inner product as $\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v}$. This corresponds to multiplying together corresponding entries in the vectors, then adding all of the results to get a single number.

1 Find the inner product of the two given vectors:

(a) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$

(b) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 \\ -1 \\ 1 \\ -2 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 2 \\ -1 \\ 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

(e) $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$

(f) $\begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} -y \\ x \end{bmatrix}$

(a) Here we get $\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = (1)(2) + (1)(3) = 5.$

(b) Here we get $\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 + 0 = 1.$

(c) Here $(1)(3) + (-1)(2) + (1)(-1) + (-2)(0) = 3 - 2 - 1 + 0 = 0.$

(d) Here $(1)(0) + (0)(1) = 0.$

(e) Here $(0)(x) + (0)(y) = 0.$

(f) Here $(x)(y) + (-y)(x) = xy - yx = 0.$

Length and Orthogonality: We observe that in $\mathbb{R}^2$, the quantity $\sqrt{\vec{v}\cdot\vec{v}}$ is the length of $\vec{v}$ as given by the Pythagorean Theorem. This motivates us to define the length of a vector in *any* $\mathbb{R}^n$ as $\left\|\vec{v}\right\| = \sqrt{\vec{v}\cdot\vec{v}}$ (encouraged that it also agrees with our idea of length in $\mathbb{R}^1$ and $\mathbb{R}^3$). Then the *distance* between $\vec{u}$ and $\vec{v}$ is $\left\|\vec{u}-\vec{v}\right\|$, the length of the vector between them.

We also observe that in $\mathbb{R}^2$, if $\vec{u}$ and $\vec{v}$ are perpendicular then $\vec{u}\cdot\vec{v} = 0$, and vice versa. To generalize this, we say $\vec{u}$ and $\vec{v}$ are *orthogonal* if $\vec{u}\cdot\vec{v} = 0$ (and indeed, this matches with perpendicularity in three dimensions as well).

2 What are the lengths of these vectors (computed with inner products)?

(a) $\begin{bmatrix}3\\4\end{bmatrix}$

(b) $\begin{bmatrix}2\\-3\\1\\-1\end{bmatrix}$

(c) $\begin{bmatrix}1\\1\\1\end{bmatrix}$

(d) The vector of all 1s in $\mathbb{R}^n$ (this is something of a challenge problem)

(a) Here we get $\begin{bmatrix}3 & 4\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix} = (3)(3)+(4)(4) = 25$ as the inner product, so the length is $\sqrt{25} = 5$.

(b) Here the inner product is $(2)^2+(-3)^2+(1)^2+(-1)^2 = 4+9+1+1 = 15$, so the length is $\sqrt{15} \approx 3.87$ (the decimal expansion isn't necessary to compute).

(c) Here the inner product is $(1)^2+(1)^2+(1)^2 = 3(1)^2 = 3$, so the length is $\sqrt{3} \approx 1.73$.

(d) Here the inner product will be n copies of $(1)^2$ summed, which will come out to be n. Thus, the vector will have length $\sqrt{n}$.

3 What is the distance between these two vectors? Are they orthogonal?

(a) $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\-1\end{bmatrix}$

(b) $\begin{bmatrix}1\\2\\3\end{bmatrix}$ and $\begin{bmatrix}1\\2\\-3\end{bmatrix}$

(c) $\begin{bmatrix}2\\5\end{bmatrix}$ and $\begin{bmatrix}-2\\-5\end{bmatrix}$

(d) Two (different) standard basis vectors in $\mathbb{R}^n$

(a) The distance between these is the length of $\begin{bmatrix}1\\1\end{bmatrix}-\begin{bmatrix}1\\-1\end{bmatrix}=\begin{bmatrix}0\\2\end{bmatrix}$. This is $\sqrt{(0)^2+(2)^2}=2$. Their dot product is $\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix}=0$, so they are orthogonal.

(b) The distance here is the length of their difference $\begin{bmatrix}0\\0\\6\end{bmatrix}$, which is 6. The dot product is $\begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}1\\2\\-3\end{bmatrix}=$ $1+4-9=-5\neq 0$, so they are not orthogonal.

(c) The distance here is the length of $\begin{bmatrix}4\\10\end{bmatrix}$, which is $\sqrt{16+100}=\sqrt{116}=2\sqrt{29}$. Note that the length of each vector is $\sqrt{29}$. Their inner product is $\begin{bmatrix}2 & 5\end{bmatrix}\begin{bmatrix}-2 & -5\end{bmatrix}=-4-25=-29\neq 0$, so they are not orthogonal.

(d) The distance here is the length of a vector which has two nonzero entries: one 1 and one −1. Thus, it is $\sqrt{1^2+(-1)^2}=\sqrt{2}$. Since the 1 in one standard basis vector will multiply against a 0 in the other, the inner product will be zero, and thus the standard basis vectors are orthogonal.

Under the Hood: This idea of orthogonality can be used to find the collection of *all* vectors which are orthogonal to some given $\vec{u}$. These are the solutions to the equation $\vec{u}\cdot\vec{v}=\vec{u}^T\vec{v}=0$. This is just finding the nullspace of the matrix $\vec{u}^T$, but now it has a nice geometric interpretation. The solution set is a subspace, known as the *orthogonal complement* of $\vec{u}$.

Orthogonal Sets

The Punch Line: With an inner product, we can find especially nice bases called orthonormal sets.

Warm-Up What are the inner products and lengths of the following pairs of vectors?

(a) $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(e) $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

(f) $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(a) Here we have an inner product $\begin{bmatrix} 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = (1)(1) + (-2)(0) + (1)(-1) = 0$. The vector lengths are $\sqrt{(1)^2 + (-2)^2 + (1)^2} = \sqrt{6}$ and $\sqrt{(1)^2 + (0)^2 + (-1)^2} = \sqrt{2}$.

(b) Here the inner product is 0 and the length of the all-ones vector is $\sqrt{3}$.

(c) The inner product here is again zero.

(d) The inner product is 0 and both vectors have length $\sqrt{2}$.

(e) The inner product here is 1, the first vector has length 1 and the second has length $\sqrt{2}$.

(f) The inner product here is 2, and the second vector has length $\sqrt{3}$.

Orthogonal and Orthonormal Sets: If the inner product of every pair of vectors in a set $\{\vec{u}_1, \ldots, \vec{u}_m\}$ is zero, we call the set *orthogonal*. In this case, it's a linearly independent set, and so a basis for its span. If there are n vectors in the set, it is a basis for $\mathbb{R}^n$.

If in addition to begin orthogonal, every vector in the set is a *unit vector* (has length 1), we call the set *orthonormal*. Since an orthogonal set is a basis, there is a unique representation of any vector $\vec{v} = c_1\vec{u}_1 + \cdots + c_n\vec{v}_n$; as it turns out the coefficients $c_i = \frac{\vec{u}_i \cdot \vec{v}}{\vec{u}_i \cdot \vec{u}_i}$. If the set is orthonormal, this means the coefficients are just the inner products with the basis vectors.

1 Are these sets orthogonal? If so, find an orthonormal set by rescaling them.

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

(b) $\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\}$

(c) $\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$

(d) $\left\{\begin{bmatrix}3\\-1\\0\\3\\1\end{bmatrix}, \begin{bmatrix}0\\0\\0\\4\\-12\end{bmatrix}, \begin{bmatrix}3\\-11\\-7\\-6\\-2\end{bmatrix}, \begin{bmatrix}1\\1\\2\\-2\\\frac{1}{13}\end{bmatrix}, \begin{bmatrix}\sqrt{5}\\1\\0\\0\\-\sqrt{7}\end{bmatrix}, \begin{bmatrix}83\\18\\27\\-1\\0\end{bmatrix}\right\}$

(a) We found that the vectors are orthogonal in the warm-up. They aren't orthonormal, but we can get an orthonormal set by dividing by their lengths. This yields $\left\{\frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}, \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$.

(b) Again, these vectors are orthogonal but not orthonormal. The orthonormal set is $\left\{\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}\right\}$.

(c) We found previously that these vectors are not orthogonal, because they have nonzero inner products with each other.

(d) Rather than compute the 15 inner products we'd need to check if this set is orthogonal, we can use the knowledge that an orthogonal set is linearly independent. This set is in $\mathbb{R}^5$ and has 6 vectors, so it can't be linearly independent, so it isn't orthogonal. We can also see that the inner product of the second and fifth vectors is $12\sqrt{7} \neq 0$, which also shows it is not orthogonal.

Orthogonal Matrices: In an unfortunate twist of terminology, we call a matrix an *orthogonal matrix* if its columns are an orthonormal set (not just orthogonal like the name might make you think). These matrices are precisely those matrices U where $U^T U = I_n$.

2 Are these matrices orthogonal?

(a) $\begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ -2/\sqrt{6} & 0 & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

(c) $\frac{1}{3}\begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}$

(d) $\frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \end{bmatrix}$

(e) $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

(f) The change-of-coordinates matrices to and from an orthonormal set [Challenge problem]

(a) Yes, we've previously found these to be an orthonormal set.

(b) No, although the columns are orthogonal, they are not orthonormal.

(c) Yes, we can check that $\frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \end{bmatrix}\frac{1}{3}\begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix} = I_2$.

(d) No, since $\frac{1}{3}\begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}\frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 & 4 & -2 \\ 4 & 5 & 2 \\ -2 & 2 & 8 \end{bmatrix} \neq I_3$. This shows that even though $U^T U$ is an identity matrix, it's not necessarily true that UU^T is if U is not square (if it's square, the condition means $U^T = U^{-1}$, so it does commute with U).

(e) No, since although the columns are clearly orthogonal, they are not unit vectors.

(f) Yes, since one will have the orthonormal set as its columns, and the other will be the inverse of that matrix, and because for *square* matrices $U^T U = I_n$ implies $UU^T = I_n$ (by the way inverse matrices work), the inverse must also be orthonormal.

Under the Hood: Orthogonal transformations from $\mathbb{R}^n$ to itself are precisely those which do not change inner products (where $(U\vec{u})\cdot(U\vec{v}) = \vec{u}\cdot\vec{v}$ for all pairs of vectors). This means they do not change the geometry involved (lengths, relative angles, or distances), so they are particularly interesting transformations.

This is an example of an incredibly common pattern in mathematics: when there is some kind of structure (like a vector space structure, or geometric relationships), mathematicians are interested in finding the collection of functions which preserve that structure (linear transformations and transformations by orthogonal matrices, in those two cases). There are also other classes of linear transformations that preserve things like areas (determinant has absolute value 1), or orientation (determinant is precisely 1), or just angles and not lengths (columns are orthogonal but not necessarily orthonormal), and many more.

Orthogonal Projection

The Punch Line: Inner products make it quite easy to compute the component of vectors that lie in interesting subspaces—in particular, components in the direction of any other vector.

Warm-Up What is the closest vector on the x-axis to the following vectors?

(a) $\begin{bmatrix} 4 \\ 0 \end{bmatrix}$ (b) $\begin{bmatrix} 4 \\ 9 \end{bmatrix}$ (c) $\begin{bmatrix} x \\ y \end{bmatrix}$

What is the closest point on the y-axis to these vectors? On the xy-plane?

(d) $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ (e) $\begin{bmatrix} 9 \\ 1 \\ 2 \end{bmatrix}$ (f) $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$

(a) This vector is on the x-axis, so it is itself the closest vector on that axis.

(b) Here we have a y-component, but $\begin{bmatrix} 4 \\ 0 \end{bmatrix}$ is still the closest vector on the x-axis (any other vector would have a longer distance, as you can see by drawing out a triangle with the x-component, y-component, and distance between the given vector and another vector on the x-axis as legs).

(c) In general, the closest vector on the x-axis will be the vector $\begin{bmatrix} x \\ 0 \end{bmatrix}$.

(d) The closest on the y-axis is $\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$. The vector itself lives in the xy-plane.

(e) The closest vector on the y-axis is $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ (which is not particularly close, but changing y won't get you closer in the x or z directions). The closest vector in the xy-plane is $\begin{bmatrix} 9 \\ 1 \\ 0 \end{bmatrix}$, which is much closer (it's "right below" the real vector).

(f) In general, the closest vector in the y-axis is $\begin{bmatrix} 0 \\ y \\ 0 \end{bmatrix}$, and on the xy-plane is $\begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$. This is the right kind of picture to have in mind for orthogonal projection—just keeping the components in the directions we specify. If they aren't already our axes, we have to do some computation (often with inner products), but geometrically we're doing the same thing.

Orthogonal Projection: If we have some vector $\vec{u}$ that we're interested in, we can compute the *orthogonal projection* of any other vector $\vec{v}$ onto $\vec{u}$ as $\frac{\vec{v}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\vec{u}$. That is, the ratio of the inner product of $\vec{v}$ and $\vec{u}$ to the inner product of $\vec{u}$ with itself is the coefficient on $\vec{u}$ giving the closest vector in Span$\{\vec{u}\}$ to $\vec{v}$. This coefficient can be thought of as "the amount of $\vec{v}$ in the direction of $\vec{u}$", and the projection (which is a vector) as "the component of $\vec{v}$ in the direction of $\vec{u}$."

1 Compute the projection of $\vec{v}$ onto $\vec{u}$.

(a) $\vec{v} = \begin{bmatrix}1\\2\\3\end{bmatrix}$ onto $\vec{u} = \begin{bmatrix}1\\1\\1\end{bmatrix}$

(b) $\vec{v} = \begin{bmatrix}2\\-1\end{bmatrix}$ onto $\vec{u} = \begin{bmatrix}1\\3\end{bmatrix}$

(c) $\vec{v} = \begin{bmatrix}2\\-1\end{bmatrix}$ onto $\vec{u} = \begin{bmatrix}1\\2\end{bmatrix}$

(d) $\vec{v} = \begin{bmatrix}1\\0\end{bmatrix}$ onto $\vec{u} = \begin{bmatrix}x\\y\end{bmatrix}$

(a) Here we get $\vec{v}\cdot\vec{u} = (1)(1)+(2)(1)+(3)(1) = 6$ and $\vec{u}\cdot\vec{u} = (1)^2+(1)^2+(1)^2 = 3$, so the projection is $\frac{6}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}2\\2\\2\end{bmatrix}$.

(b) Here we get $\vec{v}\cdot\vec{u} = -1$ and $\vec{u}\cdot\vec{u} = 10$, so the projection is $-\frac{1}{10}\begin{bmatrix}1\\3\end{bmatrix}$. This vector is fairly short (its length is $\frac{1}{\sqrt{10}} \approx 0.32$) compared to $\vec{v}$ (which has length $\sqrt{5} \approx 2.24$, about seven times larger), which means that most of $\vec{v}$ is in a direction orthogonal to $\vec{u}$. Also, the negative coefficient means that the part of $\vec{v}$ that lies "in the direction of $\vec{u}$" is in fact going away from $\vec{u}$—by in the direction, we mean along that same line, which includes going backwards.

(c) Here we get $\vec{v}\cdot\vec{u} = 0$. This means that $\vec{v}$ is orthogonal to $\vec{u}$. Geometrically, there is no component of $\vec{v}$ that is parallel to $\vec{u}$—they in a sense have "nothing to do with each other."

(d) Here we get $\vec{v}\cdot\vec{u} = x$ and $\vec{u}\cdot\vec{u} = x^2+y^2$. Thus, the projection is $\frac{x}{x^2+y^2}\begin{bmatrix}x\\y\end{bmatrix}$. It's worth noting that the length of this vector is less than (or equal to, if $y = 0$) 1—the length of the projection of a vector onto some other vector is at most the length of the original vector, and less unless they are parallel.

Projection Onto Subspaces: If W is a subspace of $\mathbb{R}^n$, we can compute the projection of a vector onto W. This is found by taking all and only the component of a vector which lie in W, which is most easily done if we have an orthogonal (or orthonormal) basis for W. Then we can simply compute the relevant inner products to project onto each basis vector, then add up all the results. (Note that this won't work if the basis isn't orthogonal.)

2 Project the vector $\vec{v}$ onto the subspace spanned by the given vectors.

(a) $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $W = \text{Span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$

(b) $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $W = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$

(a) Here we can check that the basis given for W is orthogonal, so we compute the inner product of $\vec{v}$ with the given vectors—0 and -2, respectively—and their inner products with themselves—6 and 2, respectively, to see that the projection of $\vec{v}$ onto W is $\frac{0}{6}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + \frac{-2}{2}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

(b) Here our basis is not orthogonal. We first remove the component of the second basis vector in the direction of the first to get an orthogonal vector—the projection of $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ onto $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ is $\frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, so our new basis vector (the second minus its component in the direction of the first) is $\begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$. You can check that this is indeed orthogonal to the first and in W.

Then we can project $\vec{v}$ onto $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$ to get the projection as $\frac{3}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{5/2}{3/2}\begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 7/3 \\ 2/3 \\ 5/3 \end{bmatrix}$.

Under the Hood: Why are orthogonal bases so much easier to project onto (we don't even have a good way to project onto the span of non-orthogonal vectors other than finding an orthogonal basis for that same subspace)? Heuristically, each vector in an orthogonal set is giving "independent information" about a vector in their span. Travelling in the direction of one of them doesn't move at all in the direction of the others, while for non-orthogonal vectors, increasing in one direction also moves in some of the others, and it's hard to separate the effects.

So, a basis gives enough information to describe any vector (it spans the space) and doesn't have redundant information (it's linearly independent), while an *orthogonal* basis also has the property that pieces of that description don't interfere with each other. An orthonormal basis is even nicer, in that the information requires less processing to get information about lengths—the coefficient on each component is the length in that direction (in other bases, the length of the basis vector changes this).

The Gram-Schmidt Process

The Punch Line: We can turn any basis into an orthonormal basis using a (relatively) simple procedure.

Warm-Up For what choices of the variables are these bases orthogonal? Can they be made orthonormal by choosing variables correctly?

(a) $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}x\\y\end{bmatrix}\right\}$

(b) $\left\{\begin{bmatrix}4\\y\end{bmatrix}, \begin{bmatrix}x\\1\end{bmatrix}\right\}$

(c) $\left\{\begin{bmatrix}1/2\\1/2\\1/2\end{bmatrix}, \begin{bmatrix}x\\y\\0\end{bmatrix}, \begin{bmatrix}-x\\0\\z\end{bmatrix}\right\}$

(a) The inner product of $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}x\\y\end{bmatrix}$ is x, so if $x = 0$ (and we'll want $y \neq 0$ to prevent the zero vector), then they are orthogonal. If in particular $y = 1$, we have an orthonormal set.

(b) The inner product of $\begin{bmatrix}4\\y\end{bmatrix}$ and $\begin{bmatrix}x\\1\end{bmatrix}$ is $4x + y$, so as long as $y = -4x$, we'll have an orthogonal set. The length of the first vector is $\sqrt{4^2 + y^2} \geq \sqrt{4^2} = 4$, though, so we can't make this set orthonormal.

(c) The inner product of $\begin{bmatrix}x\\y\\0\end{bmatrix}$ and $\begin{bmatrix}-x\\0\\z\end{bmatrix}$ is $-x^2$, so if we were to have an orthogonal set, we would need $x = 0$. But then the inner product of the first two vectors would be $\frac{1}{2}y$, so we'd need $y = 0$ to make the inner product zero. However, this would mean the second vector must be zero, which means we won't be able to find an orthogonal basis by choosing variables. If only there were another way...

The Gram-Schmidt Process: Suppose we know $\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$ is a basis for some subspace W we are interested in. We can make an orthogonal basis $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ for the same subspace by repeatedly stripping away the parts of vectors that are not orthogonal to the previous ones.

In particular, we set $\vec{v}_1 = \vec{w}_1$ (there aren't previous vectors that it could be nonorthogonal to). Then we set $\vec{v}_2 = \vec{w}_2 - \frac{\vec{w}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1$ (we take off any part of $\vec{w}_2$ that's in the direction $\vec{v}_1$ with a projection). Similarly, we set $\vec{v}_3 = \vec{w}_3 - \frac{\vec{w}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \frac{\vec{w}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2$ (we have to remove parts in the first *two* directions now). In general, we set

$$\vec{v}_k = \vec{w}_k - \frac{\vec{w}_k \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \dots - \frac{\vec{w}_k \cdot v_{k-1}}{\vec{v}_{k-1} \cdot v_{k-1}}\vec{v}_{k-1}$$

(subtracting off the projection onto all previous vectors in the basis we are constructing).

1 Apply the Gram-Schmidt Process to the following (ordered) bases:

(a) $\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$ (b) $\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}\right\}$ (c) $\left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}2\\0\\-3\end{bmatrix}, \begin{bmatrix}17\\-3\\1\end{bmatrix}\right\}$

(a) Here, we choose $\vec{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$. Then the projection of the second vector onto $\vec{v}_1$ is $\begin{bmatrix}1\\0\\0\end{bmatrix}$, so we choose $\vec{v}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix} - \begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}0\\1\\0\end{bmatrix}$. The projection of the third onto $\vec{v}_1$ is also $\begin{bmatrix}1\\0\\0\end{bmatrix}$ and onto $\vec{v}_2$ is $\begin{bmatrix}0\\1\\0\end{bmatrix}$, so we choose $\vec{v}_3 = \begin{bmatrix}1\\1\\1\end{bmatrix} - \begin{bmatrix}1\\0\\0\end{bmatrix} - \begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\end{bmatrix}$. Thus, when we apply the Gram-Schmidt Process to this basis, we get the standard basis out.

(b) Here, we choose $\vec{v}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}$. Then $\vec{v}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix} - \begin{bmatrix}2/3\\2/3\\2/3\end{bmatrix} = \begin{bmatrix}1/3\\1/3\\-2/3\end{bmatrix}$ and $\vec{v}_3 = \begin{bmatrix}1\\0\\0\end{bmatrix} - \begin{bmatrix}1/3\\1/3\\1/3\end{bmatrix} - \begin{bmatrix}1/6\\1/6\\-1/3\end{bmatrix} = \begin{bmatrix}1/2\\-1/2\\0\end{bmatrix}$. This set is dramatically different than the standard basis—the order of the basis you start with has a large impact on the Gram-Schmidt Process. To put them all together, our orthogonal basis is then $\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1/3\\1/3\\-2/3\end{bmatrix}, \begin{bmatrix}1/2\\-1/2\\0\end{bmatrix}\right\}$.

(c) Here we get $\vec{v}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}$. Then we get $\vec{v}_2 = \begin{bmatrix}2\\0\\-3\end{bmatrix} - \frac{10}{14}\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}5/2\\1\\-3/2\end{bmatrix}$, and $\vec{v}_3 = \begin{bmatrix}17\\-3\\1\end{bmatrix} - \frac{14}{14}\begin{bmatrix}1\\2\\3\end{bmatrix} - \frac{76/2}{19/2}\begin{bmatrix}5/2\\1\\-3/2\end{bmatrix} = \begin{bmatrix}6\\-9\\4\end{bmatrix}$. To put them all together, our orthogonal basis is $\left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}5/2\\1\\-3/2\end{bmatrix}, \begin{bmatrix}6\\-9\\4\end{bmatrix}\right\}$.

Orthonormal Bases: After applying the Gram-Schmidt Process, it's easy to get an orthonormal basis—just rescale the results. It's important to note that the rescaling can be done right after subtracting off the projections onto the previous vectors, but shouldn't be done before doing so, as subtracting vectors changes lengths (it won't harm the process, but you won't get unit vectors out of it).

2 Find the orthonormal bases from the results of Problem 1.

(a) The standard basis is already orthonormal, so here we're done.

(b) Here, we find the norm of the first vector is $\sqrt{3}$, of the second $\sqrt{\frac{2}{3}}$, and of the third $\sqrt{\frac{1}{2}}$. Thus, our orthonormal basis is $\left\{ \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \sqrt{\frac{3}{2}} \begin{bmatrix} 1/3 \\ 1/3 \\ -2/3 \end{bmatrix}, \sqrt{2} \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} \right\}$.

(c) Here, the resulting orthonormal basis is $\left\{ \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \frac{1}{\sqrt{38}} \begin{bmatrix} 5 \\ 2 \\ -3 \end{bmatrix}, \frac{1}{\sqrt{133}} \begin{bmatrix} 6 \\ -9 \\ 4 \end{bmatrix} \right\}$.