Linear Algebra

Math 314 Fall 2025

January 13, 2026

Chapter 1

Systems of Equations

1.1 What is Linear Algebra?

Linear algebra is the study of linear equations.

Definition 1.1. A **linear equation** in the variables $x_1, x_2, \dots, x_n$ is an equation that can be written in the form:

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$$

where $a_1, \dots, a_n, b$ are constants (real numbers). The constant a_i is the **coefficient** of x_i, and b is the **constant term**.

Example 1.2.

(a) The equation

$$2x_1 - 5x_2 + 2 = -x_1$$

is a linear equation, as it is equivalent to the equation

$$3x_1 - 5x_2 = -2.$$

(b) The equation

$$x_2 = 2(\sqrt{6} - x_1) + x_3$$

is also a linear equation: note that

$$x_2 = 2(\sqrt{6} - x_1) + x_3 \iff 2x_1 + x_2 - x_3 = 2\sqrt{6}.$$

(c) The equation

$$x_1 x_2 = 6$$

is **not** a linear equation.

(d) The equations

$$x_1 + \log x_2 - x_3 = 2 \quad \text{and} \quad x_1^2 = 7$$

are **not** linear equations.

In this class, we will study systems of linear equations:

Definition 1.3. A **system of linear equations** or **linear system** is a collection of one or more linear equations. A **solution** to a system of equations in the variables $x_1, \dots, x_n$ is a list $s = (s_1, \dots, s_n)$ of numbers that satisfy every equation in the system, meaning that if we replace x_1 by s_1, x_2 by s_2, and so on, then we obtain a true equality.

The **solution set** of a system is the set of all possible solutions.

Example 1.4.

(a) The system of linear equations

$$\begin{cases} x_1 = 4 \\ 2x_1 + x_2 = 0 \end{cases}$$

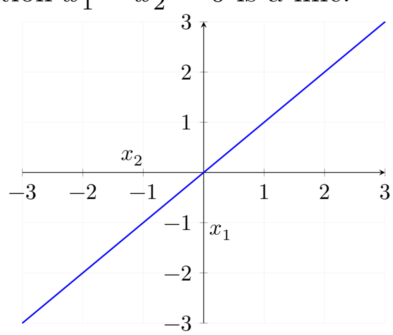
has one solution, the point $(4, -8)$. The solution set is $\{(4, -8)\}$, which is how we denote the set that has only one element $(4, -8)$.

(b) The system of linear equations

$$\begin{cases} x_1 = 4 \\ x_1 = 7 \end{cases}$$

is impossible, and it has no solutions. The solution set is the **empty set** $\emptyset$.

(c) The solution set of the equation $x_1 - x_2 = 0$ is a line:



We will later explain why the following holds:

Important

In general, a system of linear equations may have:

- No solutions,
- Exactly one solution, or
- Infinitely many solutions.

But it can never have a finite number of solutions greater than one.

Definition 1.5. Two systems of linear equations in the same variables $x_1, \dots, x_n$ are **equivalent** if they have the same solution set.

To study linear systems of equations, we keep replacing our system by an equivalent system, until the solution set becomes easy to find. To do this, we will use matrices.

Definition 1.6. An $m \times n$ (read *m by n*) matrix is a rectangular array of numbers with m rows and n columns. The (i, j) entry of A is the value on the ith row and jth column.

Example 1.7. The following is a 2×3 matrix:

$$\begin{bmatrix} 2 & 5 & 0 \\ 7 & -3 & 13 \end{bmatrix}.$$

Definition 1.8. A system of linear equations

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \qquad\qquad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases}$$

has **coefficient matrix** A and **constant vector b** below:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

The **augmented matrix** of the system is

$$[A|\mathbf{b}] = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix} \quad \text{also written} \quad \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right].$$

Sometimes we will write $Ax = \mathbf{b}$ to refer to the system in a more compact way.

Remark 1.9. In the coefficient matrix for a system of linear equations,

$$\begin{gathered} \text{number of rows} = \text{number of equations} \\ \text{number of columns} = \text{number of variables.} \end{gathered}$$

In contrast, the augmented matrix always has exactly one extra column.

Example 1.10. Given the system

$$\begin{cases} 3x_1 + x_2 = 5 \\ 2x_1 - x_3 = 6 \end{cases}$$

has coefficient matrix $\begin{bmatrix} 3 & 1 & 0 \\ 2 & 0 & -1 \end{bmatrix}$ and augmented matrix $\begin{bmatrix} 3 & 1 & 0 & 5 \\ 2 & 0 & -1 & 6 \end{bmatrix}$.

How do we solve systems of linear equations?

Theorem 1.11. *Any system of linear equations can be solved using the following* **elementary row operations** *on the augmented matrix:*

(a) Replace: Replace one row by the sum of itself and a multiple of another row.
(b) Swap: Swap two rows.
(c) Scale: Multiply all entries of a row by a nonzero constant.

How does this work in practice?

Example 1.12. Let us take the first step in resolving the following system of linear equations:

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_2 - 8x_3 = 8 \\ 5x_1 - 5x_3 = 10. \end{cases} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

Replace
$R_3 \to R_3 - 5R_1$

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_2 - 8x_3 = 8 \\ 10x_2 - 10x_3 = 10. \end{cases} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

$$\begin{array}{r|ccccc} \text{Old R}_3 & & 5 & 0 & -5 & 10 \\ -5\text{R}_1 & + & -5 & 10 & -5 & 0 \\ \hline \text{New R}_3 & & 0 & 10 & -10 & 10 \end{array}$$

Definition 1.13. We say two $n \times m$ matrices A and B are **row equivalent** if there exists a finite sequence of row operations that converts A into B. We will write $A \sim B$ to say that A and B are equivalent.

Example 1.14. The calculation we did in Example 1.12 shows that

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{bmatrix}.$$

Remark 1.15. Row operations are always reversible. If matrix A is row equivalent to B, then B is also row equivalent to A. So if we write $A \sim B$, it is also true that $B \sim A$.

Theorem 1.16. *If the augmented matrices of two linear systems are row equivalent, then the systems have the same solution set.*

In other words, if the augmented matrices are row equivalent, then the corresponding linear systems are equivalent. We use this idea to solve systems of linear equations: we keep performing row operations until we have a simpler system we can solve.

Example 1.17. Consider the linear system below, and its augmented matrix:

$$\begin{cases} 2x_2 - 8x_3 = 8 \\ x_1 - 2x_2 = 0 \\ 5x_1 - 5x_3 = 10 \end{cases} \qquad \left[\begin{array}{ccc|c} 0 & 2 & -8 & 8 \\ 1 & -2 & 0 & 0 \\ 5 & 0 & -5 & 10 \end{array}\right].$$

Let us row reduce step by step:

$$\left[\begin{array}{ccc|c} 0 & 2 & -8 & 8 \\ 1 & -2 & 0 & 0 \\ 5 & 0 & -5 & 10 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{ccc|c} 1 & -2 & 0 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{array}\right]$$

$$\xrightarrow[R_3 \to \frac{1}{5} R_3]{R_2 \to \frac{1}{2} R_2} \left[\begin{array}{ccc|c} 1 & -2 & 0 & 0 \\ 0 & 1 & -4 & 4 \\ 1 & 0 & -1 & 10 \end{array}\right]$$

$$\xrightarrow{R_3 \to R_3 - R_1} \left[\begin{array}{ccc|c} 1 & -2 & 0 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 2 & -1 & 10 \end{array}\right]$$

$$\xrightarrow{R_3 \to R_3 - 2R_2} \left[\begin{array}{ccc|c} 1 & -2 & 0 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 7 & 2 \end{array}\right].$$

This system is now in **triangular form**, which is sufficient to allow us to find the solutions by **back substitution**: we see that

$$7x_3 = 2,$$

so

$$x_3 = \frac{2}{7}.$$

We can now substitute this back in the second equation to obtain

$$x_2 = 4x_3 + 4 = \frac{8}{7} + 4 = \frac{36}{7}$$

and substituting into the first equation gives us

$$x_1 = 2x_2 = \frac{72}{7}.$$

The solution set is $\{(\frac{72}{7}, \frac{36}{7}, \frac{2}{7})\}$.

The big question is how to apply elementary row operations efficiently. This is where Gauss Elimination will come in.

1.2 Gaussian Elimination and Row Echelon form

Definition 1.18. Given a matrix, the **leading entry** of a particular row is the first nonzero entry in that row (from the left).

Definition 1.19. A rectangular matrix is in **row echelon form** if:

- Any rows consisting entirely of zeros are at the bottom.
- The leading entry of each nonzero row is to the right of the leading entry of the row above.
- All entries below a leading entry (in the same column) are zero.

Example 1.20. In each of the matrices below, we circled the leading entries.

(a) The matrix

$$\begin{bmatrix} 2 & -3 & 0 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in echelon form.

(b) The matrix

$$\begin{bmatrix} 2 & -3 \\ 1 & 4 \end{bmatrix}$$

is not in echelon form. (The leading entries are below each other!)

(c) The matrix

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

is not echelon form. (The rows should be switched!)

(d) The matrix

$$\begin{bmatrix} 1 & 3 & -3 \\ 0 & 0 & 0 \\ 0 & 3 & 4 \end{bmatrix}$$

is not in echelon form. (The second row should be at the bottom!)

Definition 1.21. A matrix is in **reduced row echelon form (RREF)** if it is in row echelon form and:

- The leading entry in each nonzero row is 1.
- Each leading 1 is the only nonzero entry in its column.

Remark 1.22. A typical matrix in RREF has the following format:

$$\begin{bmatrix} \cdots & 0 & 1 & \star & 0 & \star & 0 & \cdots & 0 & \star & \star \\ \cdots & 0 & 0 & 0 & 1 & \star & 0 & \cdots & 0 & \star & \star \\ \cdots & 0 & 0 & 0 & 0 & 0 & 1 & \cdots & 0 & \star & \star \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & \star & \star \\ \multicolumn{11}{c}{\text{if there are zero rows}} \\ \multicolumn{11}{c}{\text{they are at the bottom}} \end{bmatrix}$$

Example 1.23. In each of the matrices below, we circled the leading entries.

(a) The matrix

$$\begin{bmatrix} 2 & -3 & 0 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in row echelon form, but not in reduced row echelon form.

(b) The matrix

$$\begin{bmatrix} 1 & -3 & 0 & 1 \\ 0 & 1 & 4 & 8 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is in row echelon form, but not reduced.

(c) The matrix

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

is in echelon form, not reduced.

(d) The matrix

$$\begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 36 \\ 0 & 0 & 1 & 7 \end{bmatrix}$$

is in reduced row echelon form.

Definition 1.24. A **pivot position** in a matrix, often shortened to **pivot**, is a position that corresponds to a leading 1 in the reduced echelon form of the matrix. A **pivot column** is a column that contains a pivot position.

Remark 1.25. Important note: a pivot is a *position*, not a value.

Example 1.26. The matrix

$$A = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

has RREF

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

We can see the pivots easily from the RREF B:

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

Thus its pivot columns are the first three columns, and we can easily mark the pivots in A:

$$A = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}.$$

Theorem 1.27. *Each matrix is row equivalent to one and only one matrix in reduced echelon form.*

Since the reduced echelon form is unique, we can talk about *the* reduced echelon form of a matrix. In particular, the number of pivot positions in a matrix is well-defined.

Definition 1.28. The **rank** of a matrix A is the number of pivot positions in A.

However, while the reduced echelon form is unique, note that there are many different paths to the reduced echelon form.

Important

To solve a linear system of equations, we are going to:

(a) Write out the augmented matrix corresponding to the system.

(b) Get the augmented matrix in row reduced echelon form.
Remember: there is only <u>one</u> possible row reduced echelon form.

(c) Read the solution to the system from the row reduced echelon form.

To get the augmented matrix in row reduced echelon form, we will use an algorithm known as Gauss Elimination, or sometimes also called Gauss–Jordan Elimination.

Algorithm 1 (Gaussian Elimination)**.** To get a matrix into reduced row echelon form:

() Start with the leftmost nonzero column. This will be the first pivot column, with a pivot at the very top.

(a) Choose a nonzero entry in this pivot column; swap rows if needed to move it into the top position, and do nothing if the pivot is already in place. From this point on, we will <u>not</u> switch this row with another ever again.

(b) Use row operations to eliminate all other entries in this pivot column.

(c) Move (right) to the next pivot column and repeat.

(d) Scale pivot rows so that each pivot is 1. This can be done together with the previous steps, or all together at the end.

(e) Eliminate all entries *above* each pivot.

Example 1.29. Let us solve the following system of linear equations:

$$\begin{cases} x_1 - x_2 + x_3 = 2 \\ 2x_1 - 2x_2 + 3x_3 = 5 \\ -x_1 + x_2 - 2x_3 = -3. \end{cases}$$

First, we write the augmented matrix of the system:

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 2 & -2 & 3 & 5 \\ -1 & 1 & -2 & -3 \end{array}\right].$$

Since the first column is nonzero, that will be our first pivot column, with a pivot on the first row. Luckily, the top entry of the first column is already nonzero, so the first row is not going anywhere.

The next step is to eliminate the rest of the first column using the first row:

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 2 & -2 & 3 & 5 \\ -1 & 1 & -2 & -3 \end{array}\right] \xrightarrow{R_2 \to R_2 - 2R_1} \left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ -1 & 1 & -2 & -3 \end{array}\right] \xrightarrow{R_3 \to R_3 + R_1} \left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & -1 \end{array}\right].$$

Where is the next pivot column? To identify it, we need to now ignore the first row and find the next column with nonzero elements *in another row*. Since the second column has all zeroes outside of the first row, the next pivot column is actually the third column. So now we use the second row to zero out everything below the pivot in the third column.

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & -1 \end{array}\right] \xrightarrow{R_3 \to R_3 + R_2} \left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This matrix is now in row echelon form, and the pivot positions already have all 1s, but the matrix is not in RREF yet. To achieve that, we need to clear the entries above the pivots too.

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R_1 \to R_1 - R_2} \left[\begin{array}{ccc|c} 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Thus the RREF is

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

In fact, we have circled the pivots below:

$$\begin{bmatrix} 1 & -1 & 1 & 2 \\ 2 & -2 & 3 & 5 \\ -1 & 1 & -2 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now that we know how to apply Gauss Elimination, how will we read the solutions from the row reduced echelon form of the augmented matrix?

Example 1.30. The system

$$\begin{cases} x_1 = 4 \\ 2x_1 + x_2 = 0 \end{cases} \qquad \text{has augmented matrix} \quad \begin{bmatrix} 1 & 0 & 4 \\ 2 & 1 & 0 \end{bmatrix}.$$

Applying the elementary row operation $R_2 \mapsto R_2 - 2R_1$, we see that its reduced row echelon form is

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -8 \end{bmatrix}.$$

This means that

$$\begin{cases} x_1 = 4 \\ x_2 = -8 \end{cases}$$

and so $(4, -8)$ is only solution.

The key point that made the previous example easy is that every column corresponding to one of the variables $x_1, \ldots, x_n$ has a pivot. But this will not happen in general.

Discussion 1.31 (How to read the solutions from the RREF?)**.** Consider the columns of the augmented matrix corresponding to each of the variables $x_1, \ldots, x_n$, and ignore the last column (corresponding to the constant vector). The columns without pivots give us **free variables**, meaning that these are variables that can take any value. Each choice of values for the free variables will correspond to one solution to the system, because they impose conditions on the variables that are not free. We might call the variables that are not free **leading variables**. We then write an expression for the remaining variables (leading variables) depending on the free variables.

Important

Once we obtain the RREF of a system:

- Columns without pivots among $x_1, \ldots, x_n$ correspond to free variables.
- Free variables can take arbitrary values.
- Each choice of free variables gives one solution to the system.

Example 1.32. We saw in Example 1.29 that the augmented matrix of the system

$$\begin{cases} x_1 - x_2 + x_3 = 2 \\ 2x_1 - 2x_2 + 3x_3 = 5 \\ -x_1 + x_2 - 2x_3 = -3. \end{cases}$$

has RREF

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

The second column has no pivot, so x_2 is a free variable. The columns corresponding to the variables x_1 and x_3 have pivots, so they are not free. This means we can write x_1 and x_3 in terms of the free variable x_2: looking at our system, which has now been reduced to

$$\begin{cases} x_1 - x_2 = 1 \\ x_3 = 1 \\ 0 = 0 \end{cases}$$

we get

$$x_1 = 1 + x_2 \quad \text{and} \quad x_3 = 1.$$

The variable x_2 can take any value, say $x_2 = t$, where t is a parameter that varies. The solutions to the system are all the points of the form

$$(1 + t, t, 1)$$

where t can take any value. The solution set is

$$\{(1 + t, t, 1) \mid t \text{ any real value}\}.$$

Example 1.33. Suppose the RREF of the augmented matrix of a system is

$$\left[\begin{array}{cccc|c} 1 & 0 & 2 & -1 & 4 \\ 0 & 1 & -3 & 2 & -7 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Then x_3 and x_4 are free variables, while x_1 and x_2 are not. To write down all solutions, we need to let the free variables take any values possible. Setting $x_3 = s$ and $x_4 = t$, where s and t are now parameters that will vary over all real numbers, we get

$$\begin{cases} x_1 &= 4 - 2s + t, \\ x_2 &= -7 + 3s - 2t. \end{cases}$$

So the solution set is

$$\{(4-2s+t,\ -7+3s-2t,\ s,\ t) \mid s, t \in \mathbb{R}\}.$$

Given a system of linear equations, rather than finding the solution set we might just want to know the answers to the following questions:

- Does the system have at least one solution?
- If a solution exists, is it unique? Meaning, does the system have only one solution, or infinitely many?

Definition 1.34. A system of linear equations is:

- **Consistent** if it has at least one solution.
- **Inconsistent** if it has no solutions.

Remark 1.35. A consistent linear system might have one solution or infinitely many solutions.

Theorem 1.36 (Consistency Criterion)**.** *A linear system of equations is inconsistent if and only if the reduced echelon form of its augmented matrix has a pivot in the last column.*

Remark 1.37. A simpler way to say this: a system is inconsistent if the RREF has a row of the form

$$[0\ 0\ \cdots\ 0 \mid 1].$$

Example 1.38. The system with augmented matrix

$$\begin{bmatrix} 1 & 3 & 5 & 7 \\ 0 & 0 & 0 & 42 \end{bmatrix}$$

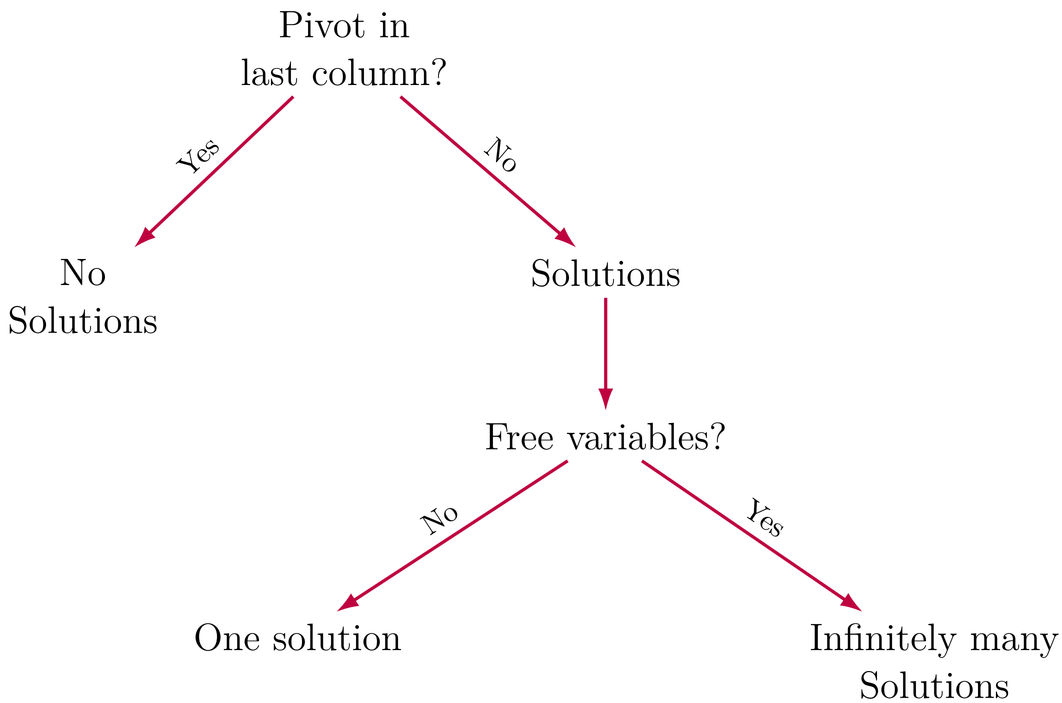
is inconsistent: we can see that the second row corresponds to the impossible equation $0 = 42$. We can also check the system is inconsistent by seeing that the reduced echelon form is

$$\begin{bmatrix} 1 & 3 & 5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which has a pivot on the last column.

In summary: to determine how many solutions a system has, look at the reduced row echelon form of the augmented matrix:

	Free variables: Yes	**Free variables: No**
Pivot in last column: Yes	0	0
Pivot in last column: No	∞	1



Example 1.39.

(a) The system whose augmented matrix has reduced row echelon form

$$\begin{bmatrix} 1 & 3 & 5 \\ 0 & 0 & 0 \end{bmatrix}$$

has a free variable (x_2) and no pivot in the last column, so it has infinitely many solutions. In fact, the solution set is

$$\{(5-3t, t) \mid t \in \mathbb{R}\}.$$

(b) The system whose augmented matrix has reduced row echelon form

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

has a free variable (x_2) and a pivot in the last column, so it has no solutions.

(c) The system whose augmented matrix has reduced row echelon form

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

has no free variables and no pivot in the last column, so it has exactly one solution. In fact, the unique solution is $(0, 3)$, so the solution set is $\{(0, 3)\}$.

1.3 The geometry of the solution set of a linear system of equations

Discussion 1.40 (One equation in two variables)**.** The solution set of one linear equation in two variables

$$a_1x_1 + a_2x_2 = b$$

is typically a line, except:

- If $a_1 = a_2 = 0$ and $b \neq 0$, the system is **inconsistent**, as it is equivalent to the equation
$$0 = b,$$
which is false. The solution set is the empty set $\emptyset$.
- If $a_1 = a_2 = b = 0$, the system is equivalent to
$$0 = 0$$
and the solution set is the entire plane $\mathbb{R}^2$.

Discussion 1.41 (Two equations in two variables)**.** What is the solution set of a system of two linear equations in two variables? Consider the system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases}$$

where a_{11} and a_{12} are not both zero, and a_{21} and a_{22} are not both zero. Each equation determines a line, so the solution set to this system of equations is the intersection of two lines. This can be:

- A point (one solution),
- A line (infinitely many solutions),
- The empty set (no solution, i.e. if the two lines are parallel).

Example 1.42.

(a) The system of equations
$$\begin{cases} x_1 = x_2 \\ x_1 + x_2 = 2 \end{cases}$$
has one solution: the solution set is $\{(1, 1)\}$. If we were to represent this geometrically, we only draw one point.

(b) The system of equations
$$\begin{cases} x_1 - x_2 = 2 \\ x_1 - x_2 = 0 \end{cases}$$
has no solutions: the solution set is the empty set $\emptyset$. (The two lines corresponding to each equation are parallel!)

(c) The system of equations

$$\begin{cases} x_1 = x_2 \\ x_1 + x_2 = 2 \end{cases}$$

has infinitely many solution: the solution set is a whole line. A fancy mathematical way to indicate that line is

$$\{(x_1, x_2) \mid x_1 - 2x_2 = -1\}.$$

Example 1.43 (Planes in three dimensions)**.** A linear equation in three variables, such as

$$x + y + z = 0$$

determines a plane in three-dimensional space. The solution to a system of linear equations in three variables such as

$$\begin{cases} 3x - y + z = 0 \\ 2x + y + 2z = 2 \\ x + 4y - 2z = 11 \end{cases}$$

is the intersection of the three planes corresponding to each equation.

Chapter 2

Vectors

2.1 Introduction to vectors

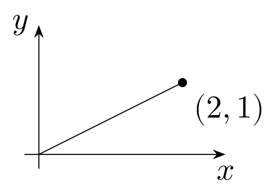
Definition 2.1. A **vector** is a matrix with only one column, that is, an $n \times 1$ matrix

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

The real number v_1 is the **first component** of v, and v_i is the ith component of v.
We write $\mathbb{R}^n$ for the set of all vectors with n components in the real numbers. The **zero vector** in $\mathbb{R}^n$ is the vector whose entries are all zero:

$$0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{in } \mathbb{R}^n.$$

Discussion 2.2. $\mathbb{R}^2$ is a two-dimensional plane, when we think of the point (a, b) in the plane as corresponding to the vector $\begin{bmatrix} a \\ b \end{bmatrix}$:



$$\rightsquigarrow \quad \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

When we represent our vector with its tail at the origin, we say the vector is in **standard position**. We might also represent a vector with its head at point $A = (a_1, \dots, a_n)$ and its tail at point $B = (b_1, \dots, b_n)$, in which case the vector is

$$v = \begin{bmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{bmatrix}.$$

Definition 2.3. We can sum vectors:

$$\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{bmatrix}.$$

We can also multiply vectors by **scalars** (real numbers):

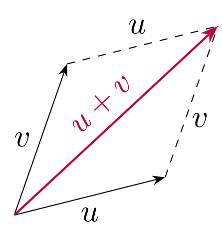
$$c \cdot \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} c\,u_1 \\ \vdots \\ c\,u_n \end{bmatrix}.$$

Example 2.4.

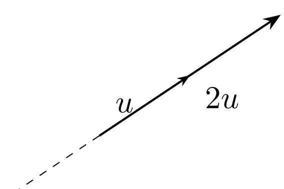
(a) $\begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 8 \\ 7 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \end{bmatrix}.$

(b) $2 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$

Remark 2.5. Here is a geometric visualization of sums (parallelogram rule):



Here is a geometric visualization of scalar multiples:



Theorem 2.6 (Properties of vector operations)**.** *For all vectors $u, v, w \in \mathbb{R}^n$ and scalars c, d:*

(a) $u + v = v + u$

(b) $(u + v) + w = u + (v + w)$

(c) $u + 0 = 0 + u = u$

(d) $u + (-u) = -u + u = 0$

(e) $c(u + v) = cu + cv$

(f) $(c + d)u = cu + du$

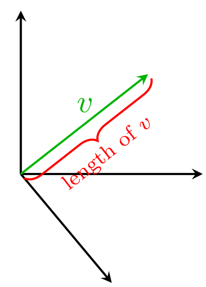
(g) $c(du) = (cd)u$

(h) $1u = u$

Definition 2.7. Let v be a vector in $\mathbb{R}^n$. The **length** or **norm** of v is the nonnegative real number

$$\|v\| = \sqrt{v_1^2 + \cdots + v_n^2}.$$

Remark 2.8. If we identify the vector $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ with the point $(v_1, \dots, v_n)$ in n-dimensional space, the norm of v is the length of the line segment between that point and the origin.



Theorem 2.9. *If v is a vector in $\mathbb{R}^n$ and c is any scalar, $\|cv\| = |c| \cdot \|v\|$.*

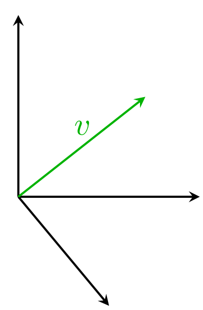
Definition 2.10. A vector in $\mathbb{R}^n$ whose length is 1 is called a **unit** vector.

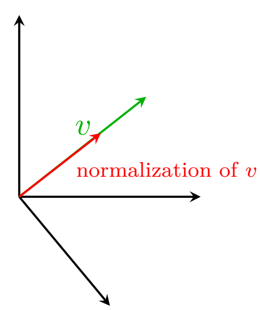
We can always find a unit vector with the same direction as a given vector v by *normalizing* v:

Definition 2.11. Let $v \neq \mathbf{0}$ be a vector in $\mathbb{R}^n$. The **normalization** of v is the unit vector

$$\frac{v}{\|v\|}$$

which has the same direction as v.





The most important unit vectors are the standard unit vectors:

Definition 2.12. The ith standard basis vector in $\mathbb{R}^n$ is the unit vector

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \\ \\ \\ \longrightarrow \text{position } i \\ \\ \\ \\ \end{array}$$

Notation 1. In $\mathbb{R}^3$, one sometimes writes $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ for the standard basis vectors:

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{e}_1 \qquad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \mathbf{e}_2 \qquad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{e}_3.$$

2.2 Linear combinations

Definition 2.13. Given vectors $v_1, \dots, v_p$ and scalars $c_1, \dots, c_p$, the vector

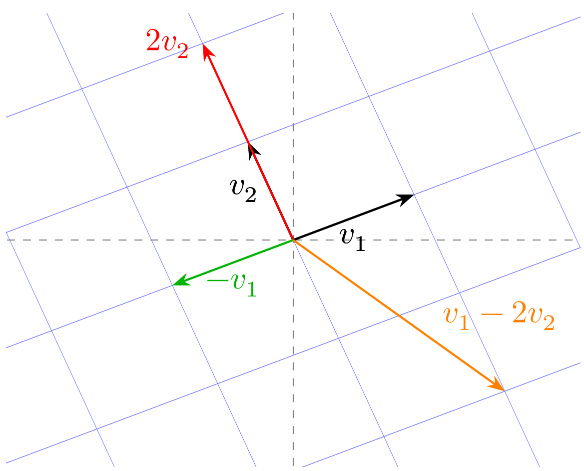
$$c_1 v_1 + \cdots + c_p v_p$$

is a **linear combination** of $v_1, \dots, v_p$ with coefficients $c_1, \dots, c_p$.

Definition 2.14 (Span)**.** Let $v_1, \dots, v_p$ be vectors in $\mathbb{R}^n$. The set of all linear combinations of $v_1, \dots, v_p$ is the **span** of $v_1, \dots, v_p$, written

$$\operatorname{span}\left(\{v_1, \dots, v_p\}\right) = \{c_1 v_1 + \cdots + c_p v_p \mid c_i \in \mathbb{R}\}.$$

Remark 2.15. What does this look like geometrically? Here is a depiction of some linear combinations of v_1 and v_2:



Any point on the plane determined by v_1 and v_2 is a linear combination of v_1 and v_2.

Remark 2.16. Note that any vector in $\mathbb{R}^n$ can be written as a linear combination of the standard vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$: the vector $v \in \mathbb{R}^n$ is

$$v = v_1 \mathbf{e}_1 + \cdots + v_n \mathbf{e}_n.$$

A typical question we would like to answer is the following: given vectors $v_1, \dots, v_p, b \in \mathbb{R}^n$, is b a linear combination of $v_1, \dots, v_p$?

Discussion 2.17. Given vectors $v_1, \dots, v_p, b \in \mathbb{R}^n$, b is a linear combination of $v_1, \dots, v_p$, or equivalently b is in the span of $v_1, \dots, v_p$, if and only if the vector equation

$$x_1 v_1 + \cdots + x_p v_p = b$$

has solutions. This vector equation has the same solutions as the linear system with augmented matrix

$$\left[\begin{array}{ccc|c} v_1 & \cdots & v_p & b \end{array}\right].$$

Example 2.18. Is $b = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$ a linear combination of $v_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$?
Equivalently, is b in $\operatorname{span}(\{v_1, v_2\})$?
Solution: We are asking if the system

$$x_1 \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

has a solution; equivalently, whether the linear system with augmented matrix

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix}$$

is consistent (= has solutions). We see that

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \xrightarrow[R_3 \to R_3 + 5R_1]{R_2 \to R_2 + 2R_1} \begin{bmatrix} 1 & 2 & 7 \\ 0 & 9 & 18 \\ 0 & 16 & 32 \end{bmatrix} \xrightarrow{R_2 \to \frac{1}{9} R_2} \begin{bmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 16 & 32 \end{bmatrix}$$

$$\xrightarrow{R_1 \to R_1 - 2R_2} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 16 & 32 \end{bmatrix} \xrightarrow{R_3 \to R_3 - 16R_2} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{(reduced row echelon form).}$$

No pivots in the last column $\Rightarrow$ the system is consistent.
Answer: yes.
In fact, if we wanted to find an explicit way of writing our vector as a linear combination of the other two, all we need is a solution to the system. From the RREF, we see that there is a unique solution (no free variables!), given by $(3, 2)$. Thus

$$3 \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}.$$

Example 2.19.

(a) $\operatorname{span}\left(\left\{\begin{bmatrix}0\\0\end{bmatrix}\right\}\right) = \left\{\begin{bmatrix}0\\0\end{bmatrix}\right\}$.

$\left\{\begin{bmatrix}a\\0\end{bmatrix} : a \text{ any value}\right\}$.

(b) $\operatorname{span}\left(\left\{\begin{bmatrix}1\\0\end{bmatrix}\right\}\right) =$

Remark 2.20. Note that for any vector $v \in \mathbb{R}^n$,

$$\operatorname{span}(\{v\}) = \{\lambda v \mid \lambda \in \mathbb{R}\}$$

is the set of all scalar multiples of v.

Example 2.21. Let u, v be vectors in $\mathbb{R}^3$, both nonzero. If u is a scalar multiple of v, then $\operatorname{span}(\{u, v\}) = \operatorname{span}(\{u\})$ is a line. Otherwise, $\operatorname{span}(\{u, v\})$ is a plane!

2.3 Matrix Equations

Definition 2.22 (Matrix-vector multiplication)**.** Let A be an $m \times n$ matrix and consider a vector $x \in \mathbb{R}^n$. The product Ax is the vector in $\mathbb{R}^m$ given by

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

We might shorten this by setting the columns of A to be $a_1, \dots, a_n$, so that we can write

$$Ax = x_1 a_1 + \cdots + x_n a_n.$$

This indicates a linear combination of the columns of A with coefficients $x_1, \dots, x_n$.

Remark 2.23. For the product Ax of a matrix A with a vector x to be defined, we need the number of columns of A to match the number of rows of the vector x.

Notation 2. Given a system of linear equations

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \qquad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases}$$

with coefficient matrix A and constant vector b, we can write our system in matrix notation

$$Ax = b$$

where

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

a vector of *variables.*

Remark 2.24. The matrix equation $Ax = b$ has the *exact same solution set* as the vector equation $x_1a_1 + \cdots + x_na_n = b$ and as the linear system with augmented matrix

$$\left[\, A \mid b \,\right] \;=\; \left[\, a_1 \;\cdots\; a_n \;\mid\; b \,\right].$$

Example 2.25. The linear system

$$\begin{cases} \;\; x_1 + 3x_2 = 4 \\ -x_1 + x_2 = 1 \end{cases}$$

has augmented matrix

$$\begin{bmatrix} 1 & 3 & 4 \\ -1 & 1 & 1 \end{bmatrix}$$

and can be written as a matrix equation as follows:

$$\begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \qquad \text{or equivalently} \qquad x_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}.$$

Remark 2.26. The system $Ax = b$ has a solution if and only if b is a linear combination of the columns of A.

Theorem 2.27. *Fix an $m \times n$ matrix A. The following are equivalent:*

(a) The system $Ax = b$ has a solution for every *vector $b \in \mathbb{R}^m$.*
(b) Every vector $b \in \mathbb{R}^m$ is a linear combination of the columns of A.
(c) The columns of A span $\mathbb{R}^m$.
(d) The coefficient matrix A has a pivot in every row.

Remark 2.28. The last statement is about A itself, the coefficient matrix of the system, and *not* an augmented matrix.

Example 2.29. The matrix

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

has a pivot in every row. Hence the equation $Ax = b$ has solutions for every $b \in \mathbb{R}^2$. Indeed,

$$\operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \right) = \mathbb{R}^2.$$

Theorem 2.30 (Properties of matrix-vector products)**.** *Let c be a scalar, let $u, v \in \mathbb{R}^n$, and let A be an $m \times n$ matrix. Then*

$$A(u + v) = Au + Av \qquad \textit{and} \qquad A(cu) = c(Au).$$

2.4 Homogeneous linear systems of equations

Definition 2.31. A linear system is **homogeneous** if we can write it as

$$Ax = 0.$$

Remark 2.32. A homogeneous system always has a solution, $x = 0$. This is called the **trivial solution**. A solution $x \neq 0$ is called **nontrivial**.

Remark 2.33. Given a homogeneous system $Ax = 0$, the system has a nontrivial solution if and only if the system has at least one free variable.

Example 2.34. Consider the homogeneous linear system

$$\begin{cases} 3x_1 + 5x_2 - 4x_3 = 0 \\ -3x_1 - 2x_2 + 4x_3 = 0 \\ 6x_1 + x_2 - 8x_3 = 0. \end{cases}$$

This can be written in matrix notation as

$$\underbrace{\begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{b} \qquad \text{(this is a homogeneous system).}$$

We can solve this system by finding the reduced row echelon form of the augmented matrix $[A\,|b]$. Since $b = 0$, row operations on A and on $[A\,|\,0]$ are the equivalent, and the augmented matrix does not add any new information. So it is sufficient to find the by finding the reduced row echelon form of A.

$$\left[\begin{array}{ccc|c} 3 & 5 & -4 & 0 \\ -3 & -2 & 4 & 0 \\ 6 & 1 & -8 & 0 \end{array}\right] \xrightarrow{R_2 \to R_2 + R_1} \left[\begin{array}{ccc|c} 3 & 5 & -4 & 0 \\ 0 & 3 & 0 & 0 \\ 6 & 1 & -8 & 0 \end{array}\right] \xrightarrow{R_3 \to R_3 - 2R_1} \left[\begin{array}{ccc|c} 3 & 5 & -4 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & -9 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{R_2 \to \frac{1}{3}R_2} \left[\begin{array}{ccc|c} 3 & 5 & -4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -9 & 0 & 0 \end{array}\right] \xrightarrow{R_1 \to R_1 - 5R_2} \left[\begin{array}{ccc|c} 3 & 0 & -4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -9 & 0 & 0 \end{array}\right] \xrightarrow{R_3 \to R_3 + 9R_2} \left[\begin{array}{ccc|c} 3 & 0 & -4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{R1 \to \frac{1}{3}R1} \left[\begin{array}{ccc|c} 1 & 0 & -\frac{4}{3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

From the reduced row echelon form we see that

$$x_3 \text{ is free}, \qquad x_1 = \tfrac{4}{3}x_3, \qquad \text{and } x_2 = 0.$$

Hence the solution set is

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} \qquad \text{(a one parameter family; nontrivial solutions occur when } x_3 \neq 0\text{).}$$

The general solution to our system is

$$x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} \qquad \text{or} \qquad t \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}$$

where t is a parameter that can take any real value. The trivial solution comes from choosing $t = 0$. Each choice of $t \neq 0$ gives a nontrivial particular solution: for example, taking $t = 1$ gives the solution

$$\begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}.$$

Remark 2.35. In summary, the general solution to a homogeneous system is a linear combination of vectors, with the free variables as coefficients. The general solution, when written in this format, is said to be in **parametric vector form**.

Definition 2.36. A **nonhomogeneous linear system** is a linear system of the form

$$Ax = b \qquad \text{for some } b \neq 0.$$

Theorem 2.37. *The general solution to the nonhomogeneous system $Ax = b$ is*

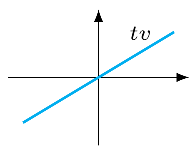
$$x = \textit{ one particular solution } + \textit{ general solution to the homogeneous system } Ax = 0.$$

Remark 2.38. Theorem 2.37 says that the solution set of the nonhomogeneous system $Ax = b$ is obtained by translating the solution set for $Ax = 0$ by a vector corresponding to one particular solution to $Ax = b$.
For example, suppose that the general solution to $Ax = 0$ is $x = tv$, where the parameter t can take the value of any real number, and $v \in \mathbb{R}^n$ is any nonzero vector; note that $x = tv$ is a line with direction v. Then the general solution to $Ax = b$ is

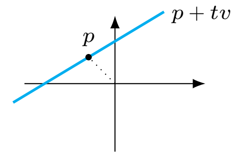
$$x = tv + p \qquad \text{for some vector } p.$$

Geometrically, this corresponds to a line parallel to v, but that goes through the point corresponding to p.

solutions to $Ax = 0$



solutions to $Ax = b$



Remark 2.39. Here are some useful geometric rules: given $u, v \in \mathbb{R}^n$,

- Parametric equation of the line through u parallel to v:

$$x = u + tv, \qquad t \in \mathbb{R}.$$

- Parametric equation of the line through u and v:

$$x = u + t(v - u), \qquad t \in \mathbb{R}.$$

Example 2.40. Consider the system

$$\begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & 8 \end{bmatrix} x = \begin{bmatrix} 7 \\ 1 \\ -4 \end{bmatrix}.$$

The reduced row echelon form of the augmented matrix is

$$\left[\begin{array}{ccc|c} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Thus the general solution to the system is

$$x = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \qquad \text{which is in parametric vector form.}$$

Important

To write the solution set of a consistent system:

(a) Row-reduce the augmented matrix into reduced echelon form.
(b) Write each non-free variable in terms of the free ones.
(c) Write the general solution x as a vector whose entries depend on the free variables (if there are free variables).
(d) Decompose this as a linear combination of vectors where each coefficient is a free variable (plus possibly one term with coefficient 1 for a particular solution).

Example 2.41. Let us find the general solution to the linear system with augmented matrix

$$\left[\begin{array}{ccccc|c} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{array}\right].$$

and write the solution in parametric vector form.

__Solution.__ First, one uses Gauss Elimination to see that the augmented matrix has reduced echelon form

$$\left[\begin{array}{ccccc|c} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{array}\right].$$

From the RREF, we see that the free variables are x_3 and x_4. So the general solution is

$$\begin{cases} x_3, x_4 \text{ are free variables} \\ \quad x_1 = 2x_3 - 3x_4 - 24, \\ \quad x_2 = 2x_3 - 2x_4 - 7, \\ \quad x_5 = 4. \end{cases}$$

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$. Setting $x_3 = x_4 = 0$ gives us the particular solution

$$\mathbf{x} = \begin{bmatrix} -24 \\ -7 \\ 0 \\ 0 \\ 4 \end{bmatrix}.$$

Now we can write the general solution in parametric vector form:

$$\mathbf{x} = \underbrace{x_3 \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ -2 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\text{general solution of the homogeneous system}} + \underbrace{\begin{bmatrix} -24 \\ -7 \\ 0 \\ 0 \\ 4 \end{bmatrix}}_{\text{particular solution}}.$$

We close this section with an important warning:

Important

Caution! Given a linear system $Ax = b$, there is a big difference between

the coefficient matrix A and the augmented matrix $[A \quad b]$.

- Is the system $Ax = b$ consistent? $\implies$ look at the augmented matrix.
- The system $Ax = 0$ is always consistent.

  We can solve the system by focusing only on A and then finding a particular solution, but if we do so we must remember A is *not* the augmented matrix of the system.

2.5 Linear Independence

Definition 2.42. A set of vectors $\{v_1, \dots, v_p\}$ in $\mathbb{R}^n$ is **linearly independent** if the vector equation

$$x_1 v_1 + \cdots + x_p v_p = 0$$

has only the trivial solution $x_1 = \cdots = x_p = 0$. We say that $v_1, \dots, v_p$ are linearly independent or that the set $\{v_1, \dots, v_p\}$ is linearly independent.
A set of vectors $\{v_1, \dots, v_p\}$ in $\mathbb{R}^n$ is **linearly dependent** if there exist scalars $c_1, \dots, c_p$, not all zero, such that

$$c_1 v_1 + \cdots + c_p v_p = 0.$$

Given such $c_1, \dots, c_p$, the equation

$$c_1 v_1 + \cdots + c_p v_p = 0.$$

is called a **relation of linear dependence** among $v_1, \dots, v_p$. We say that $v_1, \dots, v_p$ are linearly dependent or that the set $\{v_1, \dots, v_p\}$ is linearly dependent.

Remark 2.43. Equivalently, $\{v_1, \dots, v_p\}$ is linearly independent if and only if

$$c_1 v_1 + \cdots + c_p v_p = 0 \quad \implies \quad c_1 = \cdots = c_p = 0.$$

Remark 2.44. The singleton set $\{v\}$ is linearly independent $\iff v \neq 0$.

Example 2.45. $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ are linearly dependent: for example, we can take

$$0 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

More generally:

Theorem 2.46. *Any set of vectors in $\mathbb{R}^n$ that contains the zero vector is linearly dependent.*

Why? Because we can always take any nonzero coefficient for the zero vector and 0 for the coefficients of all the (nonzero) vectors.

Example 2.47. $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}2\\0\end{bmatrix}$ are linearly dependent, since one is a scalar multiple of the other. Indeed,

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

is a relation of linear dependence.

In fact, more generally, any two nonzero vectors that are scalar multiples of each other form a linearly dependent set.

Example 2.48. If v is any nonzero vector and $t \neq 1$, then v and tv are linearly dependent, since

$$t \cdot v + (-1) \cdot (tv) = 0,$$

and the coefficients t and -1 are not both zero. Thus any two nonzero vectors that are scalar multiples of each other form a linearly dependent set.

Remark 2.49. A set $\{v_1, \dots, v_p\}$ of two or more vectors is linearly dependent if and only if one of the vectors is a linear combination of the others. However, note that this does *not* say that *every* v_i is a linear combination of the rest.

Example 2.50. $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix}\right\}$ is linearly dependent, since

$$\begin{bmatrix}2\\2\end{bmatrix}=2\begin{bmatrix}1\\1\end{bmatrix}.$$

However, note that $\begin{bmatrix}1\\0\end{bmatrix}$ is not a linear combination of the other two vectors.

Remark 2.51. An equation of linear dependence among the vectors $v_1, \dots, v_n$ is a nontrivial solution to the homogeneous system

$$x_1v_1 + \cdots + x_nv_n = 0.$$

Thus to decide if the vectors $v_1, \dots, v_n$ are linearly independent, we consider the matrix

$$A = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$$

whose columns are the vectors $v_1, \dots, v_n$, and ask whether the system $Ax = 0$ has a nontrivial solution. The vertical lines above are just for visual effect, as a reminder that each v_i is a vector; the correct way to write this is

$$A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}.$$

Theorem 2.52. *The columns of a matrix A are linearly independent if and only if the homogeneous system $Ax = 0$ has only the trivial solution.*

The homogeneous system $Ax = 0$ has only the trivial solution if and only if the augmented matrix $[A \mid 0]$ of the homogeneous system $Ax = 0$ has no free variables. Thus we can check whether a set of vectors is linearly independent by looking at the RREF of a matrix with those vectors as columns:

Theorem 2.53. *Consider the matrix*

$$A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}.$$

The column vectors $v_1, \ldots, v_n$ are linearly independent if and only if A has a pivot in every column.

Example 2.54. Consider the vectors

$$v_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \qquad v_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

Question. Are v_1, v_2, and v_3 linearly independent? linearly independent?
Solution. Consider the homogeneous system whose coefficient matrix has columns v_1, v_2, and v_3:

$$\left[\begin{array}{ccc|c} 1 & 4 & 2 & 0 \\ 2 & 5 & 1 & 0 \\ 3 & 6 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

There is one free variable, so the system has nontrivial solutions. Hence the vectors are linearly dependent.

Question. How do we find a linear dependence relation among v_1, v_2, and v_3?
Solution. From the reduced row echelon form, we see that

general solution: $\begin{cases} x_3 \text{ free} \\ x_1 = 2x_3 \\ x_2 = -x_3 \end{cases}$ In parametric vector form: $x = t \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$.

To get a particular solution, we can take for example $t = 1$, giving us

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

Remember: these are the coefficients in our relation of linear dependence. This gives us the following relation of linear dependence:

$$2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - 1 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Theorem 2.55. *Any set of more than n vectors in $\mathbb{R}^n$ is linearly dependent.*

Note that if we have more than n vectors in $\mathbb{R}^n$, it is not possible for the matrix with those vectors as columns to have a pivot in every column.

Example 2.56. Using Theorem 2.55, we can see immediately that

$$v_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \qquad v_3 = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}, \qquad \text{and} \qquad v_3 = \begin{bmatrix} 10 \\ 11 \\ 12 \end{bmatrix}$$

are linearly dependent, since we have 4 vectors in $\mathbb{R}^3$.

2.6 Matrix Transformations

Definition 2.57 (Matrix transformation)**.** Any $m \times n$ matrix A determines a function $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ as follows: for each vector $x \in \mathbb{R}^n$,

$$\mathrm{T}(x) = Ax.$$

Such a function is called a **matrix transformation**.

Remark 2.58. Helpful visual aid: the matrix A gives a function $\mathrm{T}\colon \mathbb{R}^{\#\text{ columns}} \longrightarrow \mathbb{R}^{\#\text{ rows}}$.

Example 2.59. The matrix

$$A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$$

determines the function $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by $\mathrm{T}(x) = Ax$. For example,

$$\mathrm{T}\left(\begin{bmatrix} 3 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ -3 \end{bmatrix}.$$

and

$$\mathrm{T}\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2+3 \\ 6-5 \\ -2-7 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}.$$

Example 2.60. Consider the transformation $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $\mathrm{T}(x) = Ax$, where

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

Note that for all values of x_1 and x_2,

$$\mathrm{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

This is the *identity map*! And in fact, the matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is called the *identity matrix*.

Notation 3 (Identity matrix)**.** The $n \times n$ matrix

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

is the $n \times n$ **identity matrix**.

Example 2.61. The 3×3 identity matrix is

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

2.7 Linear transformations

Definition 2.62 (Linear transformation)**.** A function $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** if for all vectors $u, v \in \mathbb{R}^n$ and all scalars c,

$$\mathrm{T}(u+v) = \mathrm{T}(u) + \mathrm{T}(v) \quad \text{and} \quad \mathrm{T}(c\,u) = c\,\mathrm{T}(u).$$

Theorem 2.63 (Properties of linear transformations)**.** *If* T *is a linear transformation, then*

(a) $\mathrm{T}(0) = 0$.

(b) $\mathrm{T}(c_1 u + c_2 v) = c_1\,\mathrm{T}(u) + c_2\,\mathrm{T}(v)$ *for all scalars* c_1 *and* c_2 *and all vectors* u *and* v.

Let us first see some examples of functions that are *not* linear transformations.

Example 2.64.

(a) The function $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathrm{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1^2 \\ x_2 \end{bmatrix}$$

is not a linear transformation: it fails to preserve addition, since for example,

$$\mathrm{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}\right) = T\left(\begin{bmatrix} 3 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 9 \\ 0 \end{bmatrix}$$

while

$$\mathrm{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + \mathrm{T}\left(\begin{bmatrix} 2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}.$$

One can also see that this function does not preserve scaling:

$$\mathrm{T}\left(2\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \mathrm{T}\left(\begin{bmatrix} 2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 0 \end{bmatrix},$$

but

$$2\,\mathrm{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

(b) Let $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the function given by

$$\mathrm{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 + 1 \\ -x_2 \\ x_1 \end{bmatrix}.$$

This is not a linear function, since it fails to preserve addition and scaling. But an even easier way to see that it fails to preserve addition is to note that

$$\mathrm{T}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

It turns out that every linear transformation is actually a matrix transformation.

Theorem 2.65. *A function $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if it is a matrix transformation, meaning that there exists a matrix A such that*

$$\mathrm{T}(x) = Ax \quad \textit{for all } x \in \mathbb{R}^n.$$

To find this matrix A, we do the following:

Definition 2.66 (Standard matrix of a linear transformation)**.** Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the standard basis of $\mathbb{R}^n$. Given a linear transformation $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$, consider the matrix

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

The matrix A is called the **standard matrix** of T and it satisfies

$$\mathrm{T}(x) = Ax \qquad \text{for all } x \in \mathbb{R}^n.$$

Remark 2.67. Let us check that the standard matrix of a linear transformation does what we claim it does. Suppose that T is a linear transformation with standard matrix A. Given any vector $x \in \mathbb{R}^n$, we saw earlier that we can decompose x into its components and write it as a linear combination of the standard basis elements:

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

Then

$$\begin{aligned} \mathrm{T}(x) &= \mathrm{T}(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) \\ & x_1\,\mathrm{T}(\mathbf{e}_1) + \cdots + x_n\,\mathrm{T}(\mathbf{e}_n) && \text{since } T \text{ is a linear transformation} \\ &= Ax && \text{since the } \mathrm{T}(\mathbf{e}_i) \text{ are the columns of } A. \end{aligned}$$

This shows that T is in fact a matrix transformation, with associated matrix A.

Let us see some examples.

Example 2.68 (Dilation in $\mathbb{R}^2$)**.** Let us find the standard matrix for the dilation $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathrm{T}(x) = 2x.$$

It is not hard to check that this is indeed a linear transformation. Since

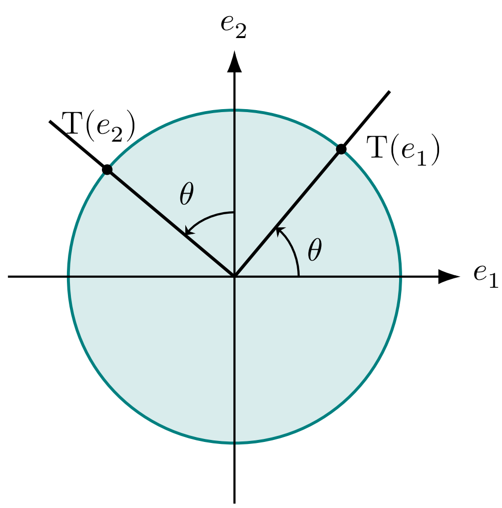
$$\mathrm{T}(\mathbf{e}_1) = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \mathrm{T}(\mathbf{e}_2) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

we conclude that the standard matrix for this linear transformation is

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

Example 2.69. Generalizing what we saw in Example 2.60, the standard matrix for the identity function $\mathbb{R}^n \to \mathbb{R}^n$ is the identity matrix.

Example 2.70 (Rotation in the plane)**.** Consider the function $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ that rotates each point counterclockwise by an angle θ (in radians).



Then using trigonometry, one can show that

$$\mathrm{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \qquad \text{and} \qquad \mathrm{T}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

and thus the standard matrix for this linear transformation is

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

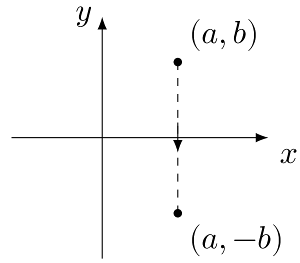
We conclude that

$$\mathrm{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Example 2.71. Consider the matrix transformation $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $\mathrm{T}(x) = Ax$, where

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

Note that $\mathrm{T}\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a \\ -b \end{bmatrix}$. Geometrically,

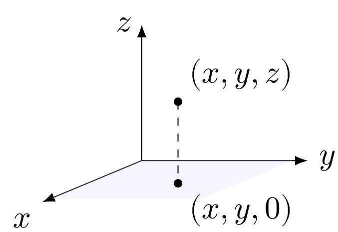


This is the reflection across the x-axis.

Example 2.72 (Geometric description in $\mathbb{R}^3$)**.** Let us give a geometric description of the matrix transformation with standard matrix

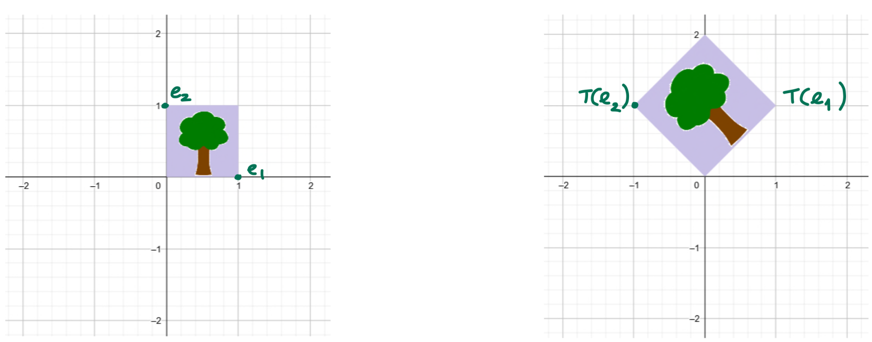
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Note that $\mathrm{T}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$. Geometrically,



This is the orthogonal projection of $\mathbb{R}^3$ onto the xy-plane.

Example 2.73. Consider the linear transformation $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ that does the following:



Note how we explicitly marked the images of $\mathbf{e}_1$ and $\mathbf{e}_2$. How can we tell? We know that $\mathrm{T}(0) = 0$, so looking for the other bottom corner of the tree we find the image of $\mathrm{T}(e_1)$.

Moreover, the $T(e_2)$ must be the opposite corner. While a linear transformation might stretch things, it will not change the fact that e_1 and e_2 are on opposite corners of the tree. This is sufficient for us to find the standard matrix, and thus to completely describe the linear transformation: the matrix is

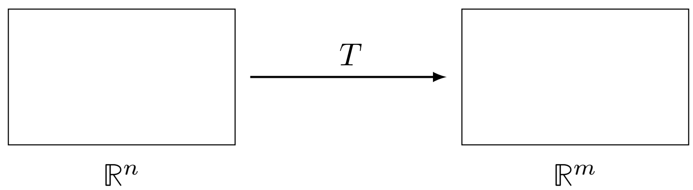
$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

2.8 Injective and surjective maps

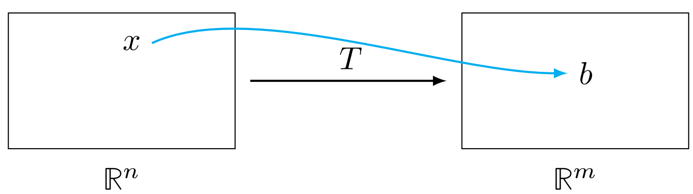
Definition 2.74. A function $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ has **domain** $\mathbb{R}^n$ and **codomain** $\mathbb{R}^m$.

Informally, the domain is the set of all inputs, and the codomain is the set of all *possible* outputs, whether or not they are *actual* outputs. Saying the codomain is $\mathbb{R}^m$ means that all the outputs are vectors in $\mathbb{R}^m$, but not that every vector in $\mathbb{R}^m$ can be obtained as a specific output.

Example 2.75 (In a picture)**.** We can visualize T as mapping the "domain box" to the "codomain box":



Here is a visual depiction of an input x going to an output b:

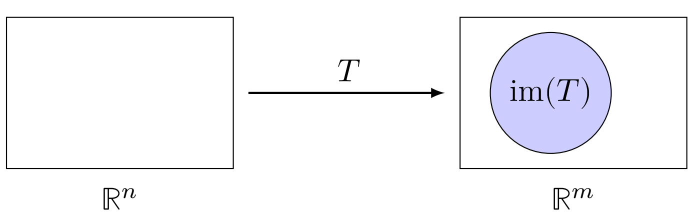


Definition 2.76 (Image or range)**.** The **image** or **range** of a function $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ is

$$\operatorname{im}(\mathrm{T}) = \{\, \mathrm{T}(x) \mid x \in \mathbb{R}^n \,\} \subseteq \mathbb{R}^m.$$

Definition 2.77 (Surjective function)**.** Let $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function. We say T is **surjective** if for every $b \in \mathbb{R}^m$ there exists *at least one* $x \in \mathbb{R}^n$ such that $\mathrm{T}(x) = b$. Equivalently, T is surjective if $\operatorname{im}(\mathrm{T}) = \mathbb{R}^m$, meaning the image is the entire codomain. Some authors also use the word **onto**.

Example 2.78. A function is <u>not</u> surjective if $\operatorname{im}(T)$ is a proper subset of $\mathbb{R}^m$.



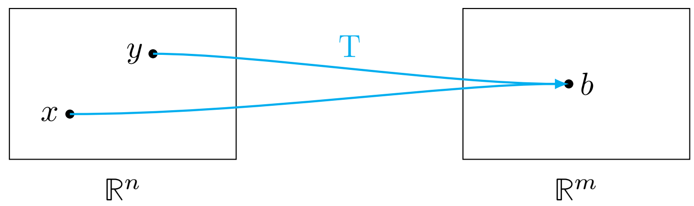
not surjective since $\operatorname{im}(T) \neq \mathbb{R}^m$

Definition 2.79 (Injective function)**.** Let $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function. We say T is **injective** if for each $b \in \mathbb{R}^m$ there exists *at most one* $x \in \mathbb{R}^n$ such that $\mathrm{T}(x) = b$. Equivalently,

$$\mathrm{T}(x_1) = \mathrm{T}(x_2) \implies x_1 = x_2.$$

Remark 2.80. Some authors use the word **one-to-one** to refer to injective functions, but that can lead to some ambiguity, so we will avoid those words.

Remark 2.81 (In a picture)**.** A function is not injective if two different inputs map to the same output.



not injective since
$T(x) = b = T(y)$ and $x \neq y$

Definition 2.82. A function $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ is **bijective** if it is both injective and surjective.

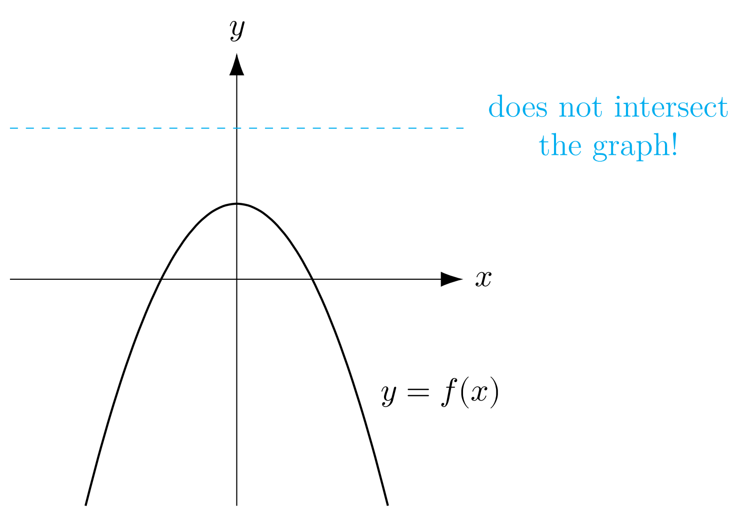
Example 2.83. Consider the (nonlinear) function $\mathrm{T}\colon \mathbb{R} \to \mathbb{R}$ given by $\mathrm{T}(x) = x^2$. This function is not injective: for example, $\mathrm{T}(1) = 1 = \mathrm{T}(-1)$. It is also not surjective: $\operatorname{im}\mathrm{T}$ is just the set of nonnegative real numbers:

$$\operatorname{im}\mathrm{T} = \{x \in \mathbb{R} \mid x \geqslant 0\}.$$

Discussion 2.84. Let us focus on the more familiar case of functions $f\colon \mathbb{R} \to \mathbb{R}$, and consider the graph of such a function f. We can describe the injective and surjective properties visually:

- Surjective: f is surjective if and only if every horizontal line crosses the graph of f at least once.
  Note that if the horizontal line $y = b$ crosses the graph at (a, b), then that means that

$$f(a) = b.$$

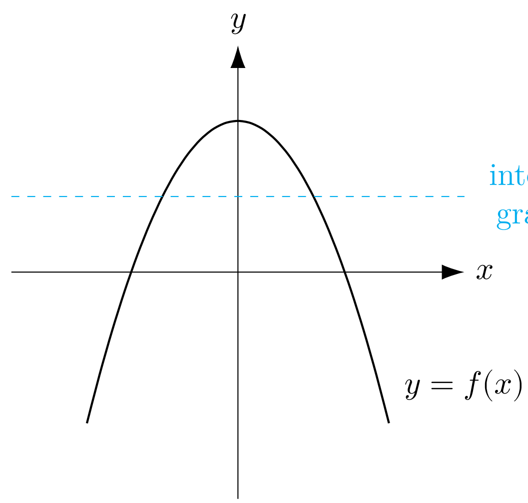


Example: f not surjective

Injective: f is injective if and only if every horizontal line crosses the graph of f at most once.

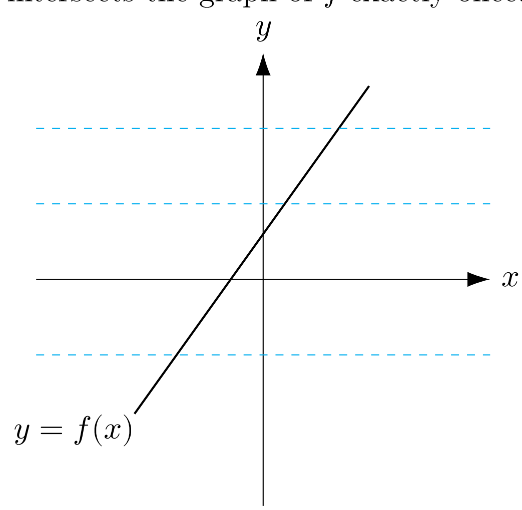
Note that if the horizontal line $y = b$ crosses the graph twice, say at (a, b) and (c, b), then that means that

$$f(a) = b = f(c).$$



Example: f not injective

Putting these ideas together: f is bijective if and only if every horizontal line intersects the graph of f exactly once.



Example: f bijective

Remark 2.85. Injectivity and surjectivity are <u>different</u> properties. A function $\mathrm{T} : \mathbb{R}^n \to \mathbb{R}^m$ can be

- injective but not surjective,
- surjective but not injective,
- bijective (both injective and surjective),
- or neither injective nor surjective.

We are of course interested specifically in the case where our function is a linear transformation.

Example 2.86 (Identity on $\mathbb{R}^2$ is bijective)**.** Let $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the identity map, so that

$$\mathrm{T}\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a \\ b \end{bmatrix}.$$

- T is surjective: for any $\begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$ we have $T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a \\ b \end{bmatrix}$.
- T is injective: each vector maps to itself, so equal outputs force equal inputs.

Example 2.87 (injective but not surjective)**.** Define $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^3$ by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ x+y \end{bmatrix}.$$

- This map T is not surjective because, for instance, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \notin \mathrm{im}(T)$.
- But T is injective: if $\mathrm{T}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = T\left(\begin{bmatrix} u \\ v \end{bmatrix}\right)$, comparing the first two coordinates gives $x = u$ and $y = v$, so
$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}.$$

Example 2.88 (surjective but not injective)**.** Define $\mathrm{T}\colon \mathbb{R}^3 \to \mathbb{R}^2$ by

$$\mathrm{T}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix}.$$

- This function T is surjective because for any $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ we have $T\left(\begin{bmatrix} x \\ y \\ 0 \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix}$.

- But T is not injective since, for example,

$$T\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = \begin{bmatrix}1\\0\end{bmatrix} = T\left(\begin{bmatrix}1\\0\\1\end{bmatrix}\right).$$

We can characterize injectivity via the kernel:

Definition 2.89. The **kernel** of a linear transformation $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ is the set

$$\ker(\mathrm{T}) = \{x \in \mathbb{R}^n \mid \mathrm{T}(x) = 0\}.$$

Remark 2.90. Note that the kernel of any linear transformation always contains the zero vector.

Theorem 2.91. *Let* $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation. Then T is injective if and only if the equation* $\mathrm{T}(x) = 0$ *has only the trivial solution* $x = 0$*. Equivalently,* T *is injective if and only if* $\ker(T) = \{0\}$.

We can also decide if a linear transformation is injective or surjective by looking at the RREF of the corresponding standard matrix.

Theorem 2.92. *Let* $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation with standard matrix A.*

(a) The linear transformation T is surjective if and only if the columns of A span $\mathbb{R}^m$.
Equivalently: T *is surjective if and only if A has a pivot in every row.*

(b) The linear transformation T is injective if and only if the columns of A are linearly independent.
Equivalently: T *is surjective if and only if A has no free variables, meaning it has a pivot in every column.*

Example 2.93 (Identity map is bijective)**.** The identity map $\mathrm{T}\colon \mathbb{R}^3 \to \mathbb{R}^3$ is both surjective and injective. And in fact, this is the linear transformation with standard matrix

$$I_3 = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}$$

which has a pivot in every column and every row.

Example 2.94 (surjective but not injective)**.** Let $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$\mathrm{T}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = x_1.$$

Its standard matrix is $A = \begin{bmatrix} 1 & 0 \end{bmatrix}$. Thus T is surjective, as A has a pivot in every row. In fact, we can see that for each $b \in \mathbb{R}$ we can take

$$\mathrm{T}\left(\begin{bmatrix} b \\ 0 \end{bmatrix}\right) = b.$$

On the other hand, T is not injective since there is no pivot on the second column. In fact, we can see that for example

$$T\!\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = T\!\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = 1.$$

Example 2.95 (injective but not surjective)**.** Let us again consider the map $\mathrm{T}\colon \mathbb{R}^2 \to \mathbb{R}^3$ from Example 2.87, given by

$$T\!\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ x+y \end{bmatrix}.$$

Its standard matrix is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Thus T is injective but not surjective.

Example 2.96. Consider the linear transformation with standard matrix

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix}.$$

Row-reducing gives

$$A \sim \begin{bmatrix} 1 & 0 & -\frac{4}{3} \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

There is a pivot in each of the first two rows but none in the third, so T is not surjective. There is a missing pivot in the third column (a free variable), so T is not injective.

Chapter 3

Matrix operations

3.1 Adding and multiplying matrices

Let A be a matrix. Recall that the (i, j)-th entry of A is the value on the ith row and jth column. In what follows, we will write $A = [a_{ij}]$ to indicate that the matrix A has a_{ij} in the (i, j)-th entry. More precisely,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Definition 3.1 (Sum of matrices)**.** Let A and B be two $m \times n$ matrices. The **sum** of A and B is the $m \times n$ matrix $A + B$ whose (i, j)-th entry is the sum of the (i, j)-th entries of A and B. More precisely, if $A = [a_{ij}]$ and $B = [b_{ij}]$, then

$$A + B = [a_{ij} + b_{ij}].$$

Note that the sum of two matrices is only defined if they have the same size.

Example 3.2. We have $\begin{bmatrix} 3 & -1 \\ 2 & -11 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 3+1 & -1+3 \\ 2+4 & -11+5 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 6 & -6 \end{bmatrix}$.

Definition 3.3 (Multiplication by Scalars)**.** If c is a scalar and A is an $m \times n$ matrix, then cA is the $m \times n$ matrix whose entries are obtained by multiplying all the entries of A by c. If $A = [a_{ij}]$, then $cA = [ca_{ij}]$.

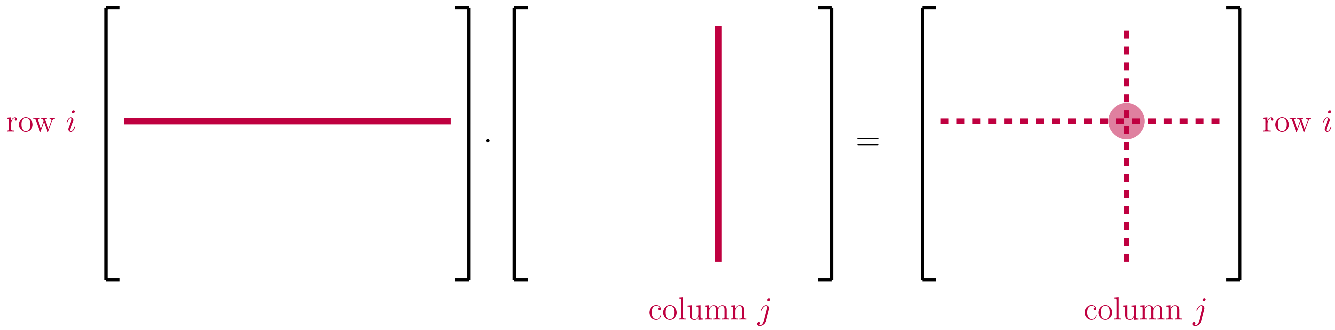
Example 3.4. We have $3 \cdot \begin{bmatrix} -1 & 1 \\ -1 & 5 \end{bmatrix} = \begin{bmatrix} -3 & 3 \\ -3 & 15 \end{bmatrix}$.

Definition 3.5 (Matrix Multiplication)**.** If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then AB is the $m \times p$ matrix obtained by multiplying rows of A with columns of B: if b_j is the j-th column of B, then

$$AB = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_p \end{bmatrix}$$

Note that Ab_i is a vector with m rows.

In a picture: to find the (i, j)th entry of AB, we focus on row i of A and column j of B



and the (i, j)th entry of AB is obtained by summing up the products of the successive elements in this row and this column, as follows:

$$(i, j)\text{th entry of } AB = a_{i1}b_{1j} + \cdots + a_{in}b_{nj}.$$

Remark 3.6. The multiplication AB is only defined if the number of columns of A matches the number of rows of B.

Example 3.7. The product

$$\begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}$$

is not defined.

Example 3.8. Here is an example of two matrices we can multiply:

$$\begin{aligned} \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} &= \begin{bmatrix} 2 \times 4 + 3 \times 1 & 2 \times 3 + 3 \times (-2) & 2 \times 6 + 3 \times 3 \\ 1 \times 4 - 5 \times 1 & 1 \times 3 + (-5) \times (-2) & 1 \times 6 + (-5) \times 3 \end{bmatrix} \\ &= \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix} \end{aligned}$$

Theorem 3.9 (Properties of Matrix Multiplication)**.** *Let A, B, and C be matrices, and assume that their sizes are such that the products AB and BC make sense.*

(a) Associativity: $(AB)C = A(BC)$.
(b) Left distributivity: $A(B + C) = AB + AC$.
(c) Right distributivity: $(B + C)A = BA + CA$.
(d) For any scalar α, $\alpha(AB) = (\alpha A)B = A(\alpha B)$.
(e) Let I_m denote the $m \times m$ identity matrix. Then $I_m A = A = AI_n$.

Important

Warning! The order of the matrices in a product matters!

In fact, it could even be that one product is defined and the other one is not. But even if the two matrices are square, we may have $AB \neq BA$. Here is an example:

Example 3.10. Consider

$$A = \begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 5 \cdot 2 + 1 \cdot 4 & 5 \cdot 0 + 1 \cdot 3 \\ (-1) \cdot 2 + 3 \cdot 4 & (-1) \cdot 0 + 3 \cdot 3 \end{bmatrix} = \begin{bmatrix} 14 & 3 \\ 10 & 9 \end{bmatrix}$$

while

$$BA = \begin{bmatrix} 2 \cdot 5 + 0 \cdot (-1) & 2 \cdot 1 + 0 \cdot 3 \\ 4 \cdot 5 + 3 \cdot (-1) & 4 \cdot 1 + 3 \cdot 3 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 17 & 13 \end{bmatrix}.$$

Hence $AB \neq BA$.

Definition 3.11 (Zero matrix)**.** The zero $m \times n$ matrix is the $m \times n$ matrix whose entries are all 0. We sometimes denote it simply by 0, if the size is clear from context.

Important

Warning! Cancellation fails. If $AB = AC$, it does *not* follow that $B = C$. Similarly, $BA = CA$ does *not* imply that $B = C$, and $AB = 0$ does *not* imply $A = 0$ or $B = 0$.

Here are some examples illustrating this:

Example 3.12. Let

$$A = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 2 \end{bmatrix} = C.$$

Then

$$AB = [1 \times 2 + 1 \times 0] = [2] = [1 \times 1 + 1 \times 1] = AC,$$

but as we noted above $B \neq C$.

Example 3.13. Let

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

Then $AB = 0$, but $A \neq 0$ and $B \neq 0$.

Let us now focus on square matrices.

Definition 3.14. A **square matrix** is an $n \times n$ matrix, meaning that it is a matrix with the same number of rows and columns.

Definition 3.15. Given a square matrix A, its consists of the (i, i) entries of A for all i.

Definition 3.16 (Powers of a matrix)**.** If A is an $n \times n$ (square) matrix, we define the powers of A by

$$A^2 = AA, \quad A^3 = AAA, \ \ldots, \quad A^k = \underbrace{A \cdot A \cdots A}_{k \text{ times}}.$$

We also set

$$A^1 = A \quad \text{and} \quad A^0 = I_n.$$

Definition 3.17. Given a square $n \times n$ matrix A, its **trace** $\operatorname{tr}(A)$ is the sum of the entries on the main diagonal:

$$\operatorname{tr}(A) = a_{11} + \cdots + a_{nn}.$$

Example 3.18. We have

$$\operatorname{tr}\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right) = 1 + 4 = 5.$$

We close this section with a quick note on linear transformations.

Remark 3.19. If $T\colon \mathbb{R}^a \to \mathbb{R}^b$ and $S\colon \mathbb{R}^b \to \mathbb{R}^c$ are linear transformations and T has standard matrix A and S has standard matrix B, then $S \circ T$ has standard matrix BA. If $U\colon \mathbb{R}^a \to \mathbb{R}^b$ is also a linear transformation, with standard matrix C, then $U + T$ has standard matrix $A + C$.

3.2 The transpose of a matrix

Definition 3.20. If A is an $m \times n$ matrix, then its **transpose** A^{T} is the $n \times m$ matrix whose rows are the columns of A.

Example 3.21. The transpose of

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 3 & 4 & 6 \end{bmatrix} \quad \text{is} \quad A^{\mathsf{T}} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 5 & 6 \end{bmatrix}.$$

Theorem 3.22 (Facts about Transposes)**.** *Let A and B be matrices whose sizes make the following make sense.*

(a) $(A^{\mathsf{T}})^{\mathsf{T}} = A$.
(b) $(A + B)^{\mathsf{T}} = A^{\mathsf{T}} + B^{\mathsf{T}}$.
(c) $(\alpha A)^{\mathsf{T}} = \alpha A^{\mathsf{T}}$ *for any scalar* α.
(d) $(AB)^{\mathsf{T}} = B^{\mathsf{T}} A^{\mathsf{T}}$.

Definition 3.23. A square matrix A is **symmetric** if $A = A^{\mathsf{T}}$.

Discussion 3.24. If $A = [a_{ij}]$ is a symmetric matrix, note that the (i, j)th entry of A^{T} is a_{ji}, and thus

$$A = A^{\mathsf{T}} \implies a_{ij} = a_{ji} \text{ for all } i, j.$$

Note that this does not impose any conditions on the main diagonal, as it simply says that $a_{ii} = a_{ii}$, which is not very surprising. We conclude that a square matrix is symmetric if and only if $a_{ij} = a_{ji}$ for all $i \neq j$. This explains the name *symmetric*.

Example 3.25. A 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is symmetric if and only if $b = c$. Thus a 2×2 symmetric matrix is one of the form

$$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}.$$

3.3 Invertible matrices

Example 3.26. Let

$$A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} -14+15 & -10+10 \\ 21-21 & 15-14 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -14+15 & -35+35 \\ 6-6 & 15-14 \end{bmatrix} = BA.$$

Thus A and B are inverses!

Definition 3.27 (Inverse Matrix)**.** Let A be an $n \times n$ matrix and let $I = I_n$ be the $n \times n$ identity matrix. The **inverse** of A, if it exists, is an $n \times n$ matrix B such that

$$AB = I \quad \text{and} \quad BA = I.$$

If A has an inverse, we say A is an **invertible matrix**.

Not every square matrix is invertible!

Example 3.28. The matrix

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

does not have an inverse. To check this, we could write

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and try to solve the equations obtained from $AB = I = BA$; we would find no solution, and thus A has no inverse.

But if a matrix is invertible, then its inverse is unique.

Remark 3.29. Suppose that A is an invertible matrix, and let B and C be two inverses of A, meaning that

$$AB = I = BA \quad \text{and} AC = I = CA.$$

Then multiplying $AC = I$ by B on the left, we conclude that

$$\begin{aligned} B &= BI && \text{since } BI = B \\ &= B(AC) && \text{since } AC = I \\ &= (BA)C && \text{by associativity} \\ &= IC && \text{since } BA = I \\ &= C. \end{aligned}$$

Thus $B = C$, and A has a unique inverse.

Notation 4. If A is an invertible matrix, we write A^{-1} for its unique inverse.

Theorem 3.30. *The* 2×2 *matrix*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is invertible if and only if $ad - bc \neq 0$*. If* $ad - bc \neq 0$*, then the inverse of* A *is given by*

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

The number $ad - bc$ is called the **determinant** of A, and written

$$x.$$

We will discuss the determinant in more detail later on in the class.

Example 3.31. Let

$$A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

Then

$$\det(A) = 3 \cdot 6 - 5 \cdot 4 = 18 - 20 = -2 \neq 0.$$

So A is invertible, with

$$A^{-1} = \tfrac{1}{-2} \begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} = \begin{bmatrix} -3 & 2 \\ \frac{5}{2} & -\frac{3}{2} \end{bmatrix}.$$

Theorem 3.32 (Properties of Invertible Matrices)**.** *Let* A *and* B *be* $n \times n$ *matrices.*

(a) If A *is invertible, then* A^{-1} *is also invertible, with inverse* A*, that is,*

$$(A^{-1})^{-1} = A.$$

(b) If A *and* B *are invertible, then* AB *is invertible, with*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

(c) If A is invertible, then A^T is invertible, and

$$(A^\mathsf{T})^{-1} = (A^{-1})^\mathsf{T}.$$

Theorem 3.33. *If A is an invertible $n \times n$ matrix, then for each $b \in \mathbb{R}^n$ the equation $Ax = b$ has a unique solution, which is given by $x = A^{-1}b$.*

Remark 3.34. Recall that

$$\begin{aligned} Ax = b \text{ has solutions for all } b &\iff A \text{ has a pivot in every row.} \\ Ax = b \text{ has only one solution} &\iff A \text{ has a pivot in every column.} \end{aligned}$$

Thus A is invertible if and only if A has a pivot in every row and every column. Note that since A is a square matrix to begin with, having a pivot in every row is equivalent to having a pivot in every column.

Given this, we can now say precisely what the reduced row echelon form of an invertible matrix is, since we know there must be a pivot in every row and every column.

Theorem 3.35. *A matrix A is invertible if and only if the reduced row echelon form of A is the $n \times n$ identity matrix.*

Example 3.36. Suppose we want to solve the system

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} x = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

The coefficient matrix of the system is

$$A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix},$$

and we calculated the inverse of this 2×2 matrix in Example 3.31, so we can use Theorem 3.33 to solve our system:

$$x = \begin{bmatrix} -3 & 2 \\ \frac{5}{2} & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} -9 + 14 \\ \frac{15}{2} - \frac{21}{2} \end{bmatrix} = \begin{bmatrix} 5 \\ -\frac{6}{2} \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}.$$

Algorithm 2 (How to compute inverses)**.** To find the inverse of an $n \times n$ matrix A, consider the extended matrix $[A|I]$. We then perform row reduction, until we get a matrix of the form $[I|B]$:

$$[A \,|\, I] \xrightarrow{\text{row reduction}} [I \,|\, B]\,.$$

Finally, $A^{-1} = B$.

Example 3.37. Consider the matrix

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

Is it invertible? To decide that, the only method we have at this point is to row-reduce and determine whether A has pivots in every row, or equivalently, if A has pivots in every column. That means we need to row reduce and decide whether

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix} \simeq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

One can check that this does indeed hold, and thus A is invertible. To find the inverse of A, we follow Algorithm 2: we need to row-reduce the extended matrix

$$[A|I] = \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right].$$

$$\begin{aligned}
\left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right] &\xrightarrow{R_2 \leftrightarrow R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right] \\
&\xrightarrow{R_3 \to R_3 - 4R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{array}\right] \\
&\xrightarrow{R_3 \to R_3 + 3R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right] \\
&\xrightarrow{R_2 \to R_2 - R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right] \\
&\xrightarrow{R_3 \to \frac{1}{2} R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{array}\right] \\
&\xrightarrow{R_1 \to R_1 - 3R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{9}{2} & -5 & -\frac{3}{2} \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{array}\right]
\end{aligned}$$

Therefore,

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}^{-1} = \begin{bmatrix} -\frac{9}{2} & -5 & -\frac{3}{2} \\ -2 & 4 & -1 \\ \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix}.$$

We will discuss other ways to compute the inverse later.

Once more, we can think about what this means for linear transformations.

Remark 3.38. Consider the linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ with standard matrix A. If A is invertible, then T is invertible, and the standard matrix for T^{-1} is A^{-1}. In fact,

$$T \text{ is bijective} \iff T \text{ is invertible} \iff \text{there exists } S\colon \mathbb{R}^n \to \mathbb{R}^n \text{ such that } S \circ T = I = T \circ S.$$

We finish this chapter by collecting many equivalent conditions to being invertible. We all add a few more equivalent conditions later on in the class, when we discuss determinants in more detail.

Theorem 3.39 (Inverse Matrix Theorem). *Let A be any $n \times n$ matrix, and write $I = I_n$. The following are equivalent:*

(a) A is invertible.

(b) There exists B such that $BA = I$.

(c) There exists B such that $AB = I$.

(d) We have $A \sim I$.

(e) The matrix A has rank n.

(f) The equation $Ax = 0$ has only the trivial solution.

(g) The columns of A form a linearly independent set.

(h) The linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = Ax$ is injective.

(i) The linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = Ax$ is surjective.

(j) The equation $Ax = b$ has at least one solution for each b.

(k) The transpose A^T is invertible.

Recall that the rank of a matrix is the number of pivots.

Definition 3.40. A square matrix whose reduced row echelon form is the identity matrix is called **nonsingular**. A square matrix that is not nonsingular is called **singular**.

Given Theorem 5.23, a square matrix is nonsingular if and only if it is invertible.

Chapter 4

Vector spaces

Linear Algebra is the study of vector spaces. These will be our main objects of study for the rest of the semester.

4.1 Examples of vector spaces

Definition 4.1. A **vector space** is a nonempty set V, whose elements we call **vectors**, together with two operations, called **addition** (of vectors in V) and **multiplication by scalars** (of a real scalar by a vector in V), satisfying the following properties for all vectors u, v, and w in V and all (real) scalars c and d:

(a) The addition $u + v$ of any vectors u and v in V is also a vector in V.

(b) The multiplication cv of a vector v by a scalar c is a vector in V.

(c) Commutativity: $u + v = v + u$ for all $u, v \in V$.

(d) Associativity: $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$.

(e) There is a **zero vector** in V, denoted 0, such that $0 + v = v + 0 = v$.

(f) For every vector v there is a vector $-v$ such that $v + (-v) = 0$.

(g) Distributivity: $c(u + v) = cu + cv$ and $(c + d)v = cv + dv$ for $u, v \in V$ and all scalars c and d.

(h) Associativity of multiplication by scalars: $c(dv) = (cd)v$.

(i) $1v = v$.

Remark 4.2. This is actually the definition of a *real* vector space, where the scalars are real numbers. In this class, all vectors spaces will be real vector spaces, but in other contexts you might learn about vector spaces with other types of scalars, such as complex numbers.

Remark 4.3. As a consequence of properties a) through j) in the definition of a vector space, any vector v in a vector space V must satisfy the following additional properties:

- $0v = \mathbf{0}$.
- $c\mathbf{0} = \mathbf{0}$.
- $-v = (-1)v$.

Example 4.4. The spaces $\mathbb{R}^n$ we have been talking about all semester are vector spaces with the addition of vectors and multiplication by scalars we defined in Definition 2.3.

Example 4.5. The set $M_{n\times m}$ of all $m \times n$ matrices is also a vector space! The addition of matrices we defined in Definition 3.1 and multiplication by scalars we defined in Definition 3.3 make this a vector space.

Example 4.6. The set $\mathbb{S}$ of doubly infinite sequences of real numbers

$$\{y_n\}_n = (\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots)$$

is also a vector space, with addition defined by

$$\{y_n\} + \{z_n\} = \{y_n + z_n\} = (\dots, y_{-2} + z_{-2}, y_{-1} + z_{-1}, y_0 + z_0, y_1 + z_1, y_2 + z_2, \dots)$$

and multiplication by scalars defined by

$$c\{y_n\} = \{cy_n\} = (\dots, cy_{-2}, cy_{-1}, cy_0, cy_1, cy_2, \dots).$$

The elements of this vector space appear in engineering applications in situations where a signal is measured in discrete time.

Example 4.7. For each fixed $n \geqslant 0$, the set $\mathbb{P}_n$ of polynomials of degree at most n is the set of polynomials of the form

$$p(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n,$$

where t is a variable and the coefficients $a_0, \dots, a_n$ are real numbers. The **degree** of $p(t)$ is the largest power of t whose coefficient is not zero. For example, $p(t) = 2$ has degree 0, and $q(t) = 2 + t^3 - 3t^4$ has degree 4. We can add polynomials of degree at most n, and the result is a polynomial of degree at most n:

$$(a_0 + a_1 t + \dots + a_n t^n) + (b_0 + b_1 t + \dots + b_n t^n) = (a_0 + b_0) + (a_1 + b_1)t + \dots + (a_n + b_n)t^n.$$

We can also multiply polynomials of degree at most n by a scalar, and the result is a polynomial of degree at most n:

$$c(a_0 + a_1 t + \dots + a_n t^n) = ca_0 + ca_1 t + \dots + ca_n t^n.$$

For example, if $p(t) = 7t + t^2 - 3t^3$, $q(t) = 2 + t^3 - 3t^4$, and $c = 5$, then

$$p(t) + q(t) = 2 + 7t + t^2 - 2t^3 - 3t^4 \text{ and } cp(t) = 35t + 5t^2 - 15t^3.$$

The set $\mathbb{P}_n$ with this addition and multiplication rules is a vector space.

Example 4.8. Let V be the set of all functions $f\colon [a, b] \to \mathbb{R}$, where $[a, b]$ is an interval in $\mathbb{R}$. We can add two functions in V, by saying that $f + g$ is the function with values

$$(f + g)(x) = f(x) + g(x),$$

and multiply a function f by a scalar c, by saying cf is the function with values

$$(cf)(x) = cf(x).$$

With these definitions, V is a vector space. The same idea also works if we replace $[a, b]$ with any set D of real numbers, and consider all the functions $D \to \mathbb{R}$. Note that in this example, each function is one vector.

4.2 Subspaces

Definition 4.9. A subset W of V is a **subspace** of V if

- The zero vector $\mathbf{0}$ of V is in W.
- W is closed under addition: if u and v are in W, then $u+v$ is in W.
- W is closed under multiplication by scalars: for all v in W and scalars c, we have cv in V.

Important

To check that a particular subset W of a vector space V is a subspace of V, we need to check that all three of properties are satisfied:

- The zero vector $\mathbf{0}$ of V is in W.
- W is closed under addition: for all vectors u and v in W, the vector $u+v$ is also in W.
- W is closed under multiplication by scalars: for all vectors v in W, and all scalars c, the vector cv is in V.

Remark 4.10. With these properties, a subspace W of a vector space V is also a vector space in its own right, with an addition and multiplication rules that are compatible with the addition and multiplication rules of V.

Example 4.11. If V is any vector space, then V is a subspace of V. Also, the set $\{\mathbf{0}\}$ with just the zero vector of V is also a subspace of V. These are called the **trivial** subspaces of V, since any vector space always has these two subspaces (which might be the same, if V only has the zero vector!)

Example 4.12. In $\mathbb{R}^3$, any plane through the origin is a subspace of $\mathbb{R}^3$, and any plane that does not go through the origin is not a subspace of $\mathbb{R}^3$.

Example 4.13. The subset of $\mathbb{R}^2$ given by

$$W = \{(x, y) \mid x, y \geqslant 0\}$$

is *not* a subspace of $\mathbb{R}^2$, since it is not closed for scalar multiplication. For example, $(1, 0)$ is in W, but $(-1, 0) = -1 \cdot (1, 0)$ is not.

Example 4.14. Suppose that V is any vector space, and consider any fixed vector v in V. The set of all scalar multiples of v is a subspace of V. Note that here it is important that

$$cv + dv = (c + d)v,$$

so that the sum of any two multiples of v is also a multiple of v.
For a more concrete example, consider the vector space $\mathbb{R}^2$. The set of all multiples of the vector

$$v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

is a subspace of $\mathbb{R}^2$, which coincides with the entire horizontal axis.

We can also write this more formally as the set

$$W = \left\{ \begin{bmatrix} c \\ 0 \end{bmatrix} : c \text{ any scalar} \right\}.$$

Let's check that W is actually a subspace of $\mathbb{R}^2$. We need to check three things:

- When we choose $c = 0$, we see that the vector $\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix}$ is in W.
- Given any two vectors in V, say $u = \begin{bmatrix} c \\ 0 \end{bmatrix}$ and $v = \begin{bmatrix} d \\ 0 \end{bmatrix}$, we have

$$u + v = \begin{bmatrix} c \\ 0 \end{bmatrix} + \begin{bmatrix} d \\ 0 \end{bmatrix} = \begin{bmatrix} c+d \\ 0 \end{bmatrix} \text{ is in W.}$$

- Given any vector in W, say $v = \begin{bmatrix} d \\ 0 \end{bmatrix}$, and any scalar c, we have

$$cv = c \begin{bmatrix} d \\ 0 \end{bmatrix} = \begin{bmatrix} cd \\ 0 \end{bmatrix}.$$

Example 4.15. The set $\mathbb{P}_2$ of all polynomials of degree at most 2 is a subspace of the set $\mathbb{P}_3$ of all polynomials of degree at most 3. Let us check it carefully:

- The zero vector is the constant polynomial 0, which is a constant polynomial, and thus has degree at most 3. We conclude that 0 is in $\mathbb{P}_3$.
- Consider two polynomials in $\mathbb{P}_3$, say

$$v = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \qquad \text{and} \qquad w = b_0 + b_1 t + b_2 t^2 + b_3 t^3.$$

  Then

$$v+w = (a_0+a_1t+a_2t^2+a_3t^3)+(b_0+b_1t+b_2t^2+b_3t^3) = (a_0+b_0)+(a_1+b_1)t+(a_2+b_2)t^2+(a_3+b_3)t^3$$

  is also a polynomial of degree up to 3, and thus $v + w$ is also in $\mathbb{P}_3$.
- Consider any polynomial in $\mathbb{P}_3$, say

$$v = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

  and any scalar c. Then

$$cv = c(a_0 + a_1 t + a_2 t^2 + a_3 t^3) = (ca_0) + (ca_1)t + (ca_2)t^2 + (ca_3)t^3$$

  is also a polynomial of degree up to 3, and thus cv is in $\mathbb{P}_3$.

Example 4.16. The set $\mathbb{P}$ of all real polynomials in one variable (of any degree) is also a vector space. For any n, the set $\mathbb{P}_n$ of polynomials of degree at most n is a subspace of $\mathbb{P}$.

Example 4.17. The vector space $\mathbb{R}^2$ is *not* a subspace of $\mathbb{R}^3$ because $\mathbb{R}^2$ is not a subset of $\mathbb{R}^3$. However, we can consider the subset

$$W = \left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} : a, b \text{ any scalars} \right\}$$

of $\mathbb{R}^3$, which is indeed a subspace of $\mathbb{R}^3$, and which looks a lot like $\mathbb{R}^2$.

4.3 Linear combinations and span

Here are some definitions we have seen before in $\mathbb{R}^n$, but now generalized to any vector space.

Definition 4.18. Let V be a vector space and consider vectors $v_1, \dots, v_n$ in V. A **linear combination** of $v_1, \dots, v_n$ is any vector of the form $c_1 v_1 + \cdots + c_n v_n$ for some scalars $c_1, \dots, c_n$.

Definition 4.19. Let V be a vector space and $v_1, \dots, v_n$ be vectors in V. The **span** of $v_1, \dots, v_n$ is the set of all linear combinations of $v_1, \dots, v_n$, which we denote by

$$\operatorname{span}(\{v_1, \dots, v_n\}) = \{c_1 v_1 + \cdots + c_n v_n \mid c_i \text{ any scalars}\}.$$

Theorem 4.20. *Given any vector space V, and vectors $v_1, \dots, v_n$ in V, the set $\operatorname{span}(\{v_1, \dots, v_n\})$ is a subspace of V.*

How would we prove this theorem? We would check that the span of any set of vectors satisfies the three properties it needs to in order to be a subspace.

Example 4.21. Consider the subset of $\mathbb{R}^4$ given by

$$W = \left\{ \begin{bmatrix} a-3b \\ b-a \\ a \\ b \end{bmatrix} : a, b \text{ any scalars} \right\}.$$

Is W a vector subspace of $\mathbb{R}^4$? Notice that we can rewrite all the vectors in W as follows:

$$\begin{bmatrix} a-3b \\ b-a \\ a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

So we can see that the vectors in W are actually just all the linear combinations of the vectors

$$\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix},$$

so

$$W = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\} \right),$$

which means that W is in fact a vector subspace of $\mathbb{R}^4$. Notice also that geometrically, this vector space is a plane.

Example 4.22. In $M_{2\times 2}$, consider the set of all diagonal matrices

$$D = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} : a, b \text{ any scalars} \right\}.$$

We can prove carefully that this is a subspace of $M_{2\times 2}$, by checking that it contains the zero matrix, it is closed for sums, and it is closed for multiplication by scalars.
Alternatively, we can note that this is also the span of two matrices:

$$D = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \right).$$

Thus Theorem 4.20 says that D is a subspace of $M_{2\times 2}$. This is known as the subspace of diagonal matrices, since the matrices in D are precisely the matrices that only have nonzero entries in the main diagonal.

Definition 4.23. Let V be a vector space and consider vectors $v_1, \dots, v_n$ in V. If

$$V = \operatorname{span}\left(\{v_1, \dots, v_n\}\right),$$

we say that $\{v_1, \dots, v_n\}$ is a **spanning set** for V.

Example 4.24. We saw back in Remark 2.16 that every vector in $\mathbb{R}^n$ can be written as a linear combination of the standard basis vectors $e_1, \dots, e_n$. This means that $\{e_1, \dots, e_n\}$ is a spanning set for $\mathbb{R}^n$.
More generally, given vectors $v_1, \dots, v_k$ in $\mathbb{R}^n$, they form a spanning set for $\mathbb{R}^n$ if and only if we can obtain any vector $b \in \mathbb{R}^n$ as a linear combination of $v_1, \dots, v_k$. Equivalently, $v_1, \dots, v_k$ form a spanning set for $\mathbb{R}^n$ if and only if the matrix

$$A = \begin{bmatrix} v_1 & \cdots v_k \end{bmatrix}$$

with columns $v_1, \dots, v_k$ has a pivot in every row.

This trick of taking the vectors and putting them in the columns of a matrix is special for vectors in $\mathbb{R}^n$. So let us see some other examples.

Example 4.25. We saw in Example 4.22 that the diagonal matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

form a spanning set for the vector space of all diagonal matrices.
There are however many spanning sets for the same vector space. For example, the vectors

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

also form a spanning set for the space of diagonal matrices.

Definition 4.26. Let V be a vector space. A set of vectors $\{v_1, \dots, v_n\}$ in V is a **linearly independent set**, or the vectors $v_1, \dots, v_n$ are **linearly independent**, if the only scalars $c_1, \dots, c_n$ such that

$$c_1 v_1 + \cdots + c_n v_n = 0$$

are the scalars $c_1 = \cdots = c_n = 0$. Equivalently, given any scalars $c_1, \dots, c_n$ that are not all zero,

$$c_1 v_1 + \cdots + c_n v_n \neq 0.$$

A set of vectors $\{v_1, \dots, v_n\}$ in V is a **linearly dependent set** if it is not linearly independent, that is, if there exist scalars $c_1, \dots, c_n$, not all zero, such that

$$c_1 v_1 + \cdots + c_n v_n = 0.$$

Given specific such scalars $c_1, \dots, c_n$, the equation

$$c_1 v_1 + \cdots + c_n v_n = 0$$

is called a **linear dependence relation** for $v_1, \dots, v_n$.

This is of course a generalization of Definition 2.42, and we have already seen many examples of linearly independent vectors in $\mathbb{R}^n$. So let us see some examples in other vector spaces.

Example 4.27. In $M_{2\times 2}$, the vectors

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

are linearly independent, and the vectors

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

are linearly dependent, since

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

The next result is the general version of Remark 2.49.

Theorem 4.28. *The vectors $v_1, \dots, v_n$ are linearly dependent if and only if one of the vectors is a linear combination of the remaining ones.*

Example 4.29. In $\mathbb{R}^2$, the vectors

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

form a spanning set for $\mathbb{R}^2$, but they are linearly dependent. In contrast, the set with just the vector v_1 is linearly independent, but it is not a spanning set for $\mathbb{R}^2$. Finally, the set with just the vectors

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } v_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

is both linearly independent and a spanning set for $\mathbb{R}^2$.

4.4 Null space and column space

We now focus on two subspaces of $\mathbb{R}^n$ that we can naturally assign to any matrix.

Definition 4.30. The **column space** of an $m \times n$ matrix A is the set of all linear combinations of the columns of A. So if $A = [a_1 \cdots a_n]$, the column space of A is the subspace of $\mathbb{R}^m$ given by

$$\mathrm{col}(A) = \mathrm{span}\left(\{a_1, \dots, a_n\}\right).$$

Similarly, the row space of A, written $\mathrm{row}(A)$, is the span of the rows of A.

Remark 4.31. We can also describe the column space of A as the image of the linear transformation $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(x) = Ax$, meaning the set

$$\mathrm{col}(A) = \mathrm{im}(T) = \{b \text{ in } \mathbb{R}^m : Ax = b \text{ has at least one solution}\} = \{T(x) \mid x \in \mathbb{R}^n\}.$$

Theorem 4.32. *The column space of any $m \times n$ matrix is a subspace of $\mathbb{R}^m$, while the row space is a subspace of $\mathbb{R}^n$.*

Example 4.33. Consider

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

The column space of A is

$$\mathrm{col}(A) = \mathrm{span}\left(\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}3\\-2\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}\right\}\right) = \mathrm{span}\left(\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}\right\}\right),$$

and

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

forms a spanning set for $\mathrm{col}(A)$.

Definition 4.34. Let A be an $m \times n$ matrix. The **null space** of A, denoted $\mathrm{Nul}(A)$, is the set of all solutions to the homogeneous equation $Ax = 0$, meaning,

$$\mathrm{Nul}(A) = \{x \text{ in } \mathbb{R}^n : Ax = 0\}.$$

Remark 4.35. The null space of A is the set of all vectors that are sent to zero by the linear transformation T defined by $\mathrm{T}(x) = Ax$. This is the same as the kernel of T.

Theorem 4.36. *The null space of any $m \times n$ matrix is a subspace of $\mathbb{R}^n$. Equivalently, the set of solutions of any homogeneous linear system is a vector space.*

Example 4.37. Consider

$$A = \begin{bmatrix} 1 & 4 & -5 & 2 \\ 0 & 2 & -4 & 0 \\ -1 & 1 & -5 & 2 \\ 3 & -1 & 11 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The general solution to the homogeneous system $Ax = 0$ is

$$x = t \begin{bmatrix} -3 \\ 2 \\ 1 \\ 0 \end{bmatrix} \quad \text{where } t \text{ is any scalar.}$$

Notice that this vector space can be rewritten as

$$\mathrm{Nul}(A) = \mathrm{span}\left(\left\{\begin{bmatrix} -3 \\ 2 \\ 1 \\ 0 \end{bmatrix}\right\}\right).$$

Example 4.38. Let

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The general solution to the homogeneous system $Ax = 0$ is

$$x = a \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}, a, b, c \text{ any scalars.}$$

So we can write the null space of A as

$$\mathrm{Nul}(A) = \mathrm{span}\left(\left\{\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}\right\}\right).$$

Remark 4.39. In the last example above, we found that

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

is a spanning set for the null space of

$$\begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}.$$

Notice that when we write the general solution to a homogeneous system $Ax = 0$ in parametric vector form, writing the solutions as linear combinations with one vector for each free variable, the vectors we get are always going to be a spanning set for $\mathrm{Nul}(A)$. Moreover, note that the vectors we obtain will also automatically be linearly independent, because each one has a 1 in the row corresponding to one of the free variables, and the remaining ones have a 0 in that same row.

Do not confuse the column space and the null space of a matrix!

Example 4.40. Consider the matrix

$$A = \begin{bmatrix} 1 & 4 & -5 & 2 \\ 0 & 2 & -4 & 0 \\ -1 & 1 & -5 & 2 \\ 3 & -1 & 11 & 1 \end{bmatrix}.$$

If we say

$$w = \begin{bmatrix} -3 \\ 2 \\ 1 \\ 0 \end{bmatrix} \text{ is in } \mathrm{Nul}(A),$$

that means $Aw = 0$, or that w is a solution to the linear system

$$\begin{cases} x_1 + 4x_2 - 5x_3 + 2x_4 = 0 \\ 2x_2 - 4x_3 = 0 \\ -x_1 + x_2 - 5x_3 + 2x_4 = 0 \\ 3x - 1 - x_2 + 11x_3 + x_4 = 0. \end{cases}$$

In contrast, if we say that

$$u = \begin{bmatrix} 5 \\ 2 \\ 0 \\ 2 \end{bmatrix} \text{ is in } \mathrm{col}(A),$$

that means that the u is a linear combination of the columns of A, or equivalently that it is in the span of the columns of A. This is also equivalent to saying that the system

$$\begin{cases} x_1 + 4x_2 - 5x_3 + 2x_4 = 5 \\ 2x_2 - 4x_3 = 2 \\ -x_1 + x_2 - 5x_3 + 2x_4 = 0 \\ 3x - 1 - x_2 + 11x_3 + x_4 = 2 \end{cases}$$

is consistent.

4.5 Bases

Definition 4.41. Let V be a vector space V. A **basis** for V is a set of linearly independent vectors that span V.

Remark 4.42. In practice, we will think of finite-dimensional vector spaces, which have a basis with finitely many elements. In that case, a basis for a vector space V is a set of vectors $\{v_1, \dots, v_n\}$ in V such that:

- $v_1, \dots, v_n$ are linearly independent, and
- $\operatorname{span}(\{v_1, \dots, v_n\}) = V$.

Remark 4.43. The plural of **basis** is **bases**; **basis** is singular and **bases** is plural.

Example 4.44. The vectors

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

form a basis for $\mathbb{R}^2$.

Example 4.45. The vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

are a basis for $\mathbb{R}^n$, called the **standard basis** for $\mathbb{R}^n$.

We already saw in Remark 2.16 that the standard basis is a spanning set for $\mathbb{R}^n$, we just did not have this language back then. We also know that the standard basis vectors are linearly independent: for example, because the matrix $A = [e_1 \cdots e_n]$ is none other than the identity matrix.

Important

Given a basis $\mathcal{B} = \{b_1, \dots, b_n\}$ for a vector space V, any vector $v \in V$ can be written as a linear combination

$$v = v_1 b_1 + \cdots + v_n b_n$$

in a unique way, meaning there is only one choice of coefficients $c_1, \dots, c_n$ that work.

Example 4.46. In $\mathbb{P}_n$, the vectors

$$\{1, t, t^2, \dots, t^n\}$$

form a basis for $\mathbb{P}_n$, called the **standard basis** for $\mathbb{P}_n$. There are other basis for $\mathbb{P}_n$: for example,

$$\{1, t-1, t^2\}$$

is also a basis for $\mathbb{P}_2$, different from the standard basis

$$\{1, t, t^2\}.$$

Important

Warning: Just because a set of vectors spans a particular subspace, it does not have to necessarily be a basis for that subspace; we need to check that the set of vectors is also linearly independent in order to be a basis.

Example 4.47. The vectors

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

span the subspace of $M_{2\times 2}$ of all diagonal matrices. However, these vectors are linearly dependent. For example, because

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

On the other hand, note that we can drop one of these and get a basis for the subspace of diagonal matrices: for example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

are a basis for the set of diagonal matrices.

Theorem 4.48 (Spanning set theorem)**.** *Let H be a subspace of a vector space V, and consider a set of vectors $S = \{v_1, \dots, v_n\}$ such that $H = \operatorname{span}(\{v_1, \dots, v_n\})$.*

(a) Suppose $v_1, \dots, v_n$ are linearly dependent. If the vector v_k is a linear combination of the remaining vectors, then if we drop v_k from S, the remaining set of vectors still spans H.

(b) If $H \neq \{\mathbf{0}\}$, then some subset of S is a basis for H.

Example 4.49. The vectors in $\mathbb{R}^3$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

are linearly independent, but they are not a basis for $\mathbb{R}^3$. They are, however, a basis for the vector space they span.
We can however extend this set of vectors to a basis for $\mathbb{R}^3$, by adding one more vector:

$$\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Theorem 4.50. *Let H be a subspace of a vector space V. Any linearly independent set of vectors in H can be expanded to a basis for H.*

Theorem 4.51. *Every basis for a vector space V has the same number of elements.*

This is actually a very deep theorem, and one we will not prove in this class.

Definition 4.52. The number of vectors in a basis of V is called the **dimension** of V, which we write $\dim(V)$.

Remark 4.53. The dimension of the zero vector space $\{\mathbf{0}\}$ is 0, and the empty set is a basis for $\{\mathbf{0}\}$.

Example 4.54. The dimension of $\mathbb{R}^n$ is n.

Example 4.55. The dimension of the space $M_{2\times 2}$ of all 2×2 matrices has dimension 4. For example, the following is a basis:

$$\left\{ \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} \right\}.$$

In general, the space $M_{m\times n}$ of all $m \times n$ matrices has dimension mn.

Example 4.56. The space $\mathbb{P}_n$ of polynomials of degree up to n has dimension $n+1$: we saw in Example 4.46 that $\{1, t, t^2, \dots, t^n\}$ is a basis for $¶_n$.

Example 4.57. The dimension of a line through the origin (zero) in $\mathbb{R}^n$ is 1: it is the span of one nonzero vector. The dimension of a plane through the origin in $\mathbb{R}^n$ is 2: it is the span of two linearly independent vectors.

Definition 4.58. A vector space is **finite-dimensional** if it has a basis with finitely many elements. A vector space is **infinite dimensional** if it has a basis with infinitely many vectors.

Example 4.59. The vector space $\mathbb{P}$ of all polynomials of any degree is an infinite dimensional vector space. The set

$$\{1, t, t^2, t^3, \dots\}$$

is a basis for $\mathbb{P}$.

Theorem 4.60. *If a vector space V has (finite) dimension n, then any set of more than n vectors in V is linearly dependent.*

We have seen an application of this theorem before! If A is a matrix with more columns than rows, then the columns of A are linearly dependent. How is this an application of the theorem? Let's say that A has m rows and n columns, and $m < n$. The columns of A are vectors in $\mathbb{R}^m$, and since there are more than m vectors, we now know they must be linearly dependent.

Theorem 4.61. *If H is a subspace of a finite-dimensional vector space V, then H also has finite dimension, and* $\dim(H) \leqslant \dim(V)$.

Theorem 4.62 (Basis theorem)**.** *Let V be a vector space of dimension n.*

- *Any set of n linearly independent vectors is a basis for V.*
- *Any set of n vectors that spans V is a basis for V.*

Here are some immediate consequences of the Basis Theorem:

- If W is a subspace of $\mathbb{R}^n$ of dimension n, then $W = \mathbb{R}^n$.
- If $W \neq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$, then $\dim(W) < n$.

Example 4.63. Let's find the dimension of the vector space

$$W = \left\{ \begin{bmatrix} a-3b+6c \\ 5a+4d \\ b-2c-d \\ 5d \end{bmatrix} : a, b, c, d \text{ any real numbers} \right\}.$$

This is a subspace of $\mathbb{R}^4$, which we can rewrite as

$$W = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 0 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix} \right\} \right).$$

Since we have a spanning set with 4 vectors, the dimension is at most 4. Notice, however, that just because these vectors span W, that doesn't necessarily make them linearly independent. In fact, we notice that

$$-2 \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ -2 \\ 0 \end{bmatrix},$$

so we can eliminate one of these two vectors and still have a spanning set for W. We claim that the remaining 3 vectors form a basis for W. Indeed, if

$$a \begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

then

$$\begin{bmatrix} a-3b \\ 5a+4c \\ b-c \\ 5c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

so $c = 0$, and thus $5a = 0 \Rightarrow a = 0$ and $b = 0$ as well.
Alternatively, we could have found a basis for W as follows: W is spanned by the columns of

$$A = \begin{bmatrix} 1 & -3 & 6 & 0 \\ 5 & 0 & 0 & 4 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We will see in the next section that the pivot columns of A form a basis for W, so the first, second, and fourth vectors form a basis for W. Therefore, $\dim(W) = 3$.

This leads us to how to find bases for the null space and column spaces of a matrix.

4.6 Bases for the null and column space of a matrix

Theorem 4.64. *The pivot columns of a matrix A form a basis for* $\operatorname{col}(A)$*. Therefore, the dimension of the column space of A is the number of pivot columns of A.*

Recall that we call the number of pivot columns of a matrix A the **rank** of A. Thus the dimension of the column space of A is the rank of A, written $\operatorname{rank}(A)$.

This gives us an algorithm to find a basis for the column space of a matrix:

Important

To find a basis for the column space of A:

- Step 1: Row reduce the matrix A to reduced row echelon form.
- Step 2: Collect the pivot columns of A.

Warning: Make sure to use the pivot columns of A, not of its reduced echelon form!

Now we can rewrite an old theorem in a new way:

Theorem 4.65. *The column space of an $m \times n$ matrix is $\mathbb{R}^m$ if and only if A has a pivot in every row.*

Example 4.66. Consider

$$A = \begin{bmatrix} 1 & 4 & -5 & 2 \\ 0 & 2 & -4 & 0 \\ -1 & 1 & -5 & 2 \\ 3 & -1 & 11 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The pivot columns of A are the first, second, and fourth columns, so

$$\operatorname{col}(A) = \operatorname{span}\left(\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \\ 1 \end{bmatrix}\right\}\right),$$

and in fact

$$\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \\ 1 \end{bmatrix}\right\}$$

form a basis for $\operatorname{col}(A)$. In particular,

$$\dim(\operatorname{col}(A)) = 3.$$

Our discussion earlier about finding spanning sets for the null space of a matrix actually produced a basis for that null space:

Important

To find a basis for the null space of A:

- Step 1: Find the general solution for $Ax = 0$.
- Step 2: Write that general solution in parametric vector form, writing the general solution as a linear combination of vectors using one vector for each free variable.
- Step 2: The vectors we used form a basis for $\text{Nul}(A)$.

In particular, this says the following about the dimension of $\text{Nul}(A)$:

Theorem 4.67. *The dimension of the null space of A is the number of free variables of A.*

Let us see this in an example.

Example 4.68. Consider

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The general solution to the homogeneous system $Ax = 0$ is

$$x = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \text{ with } x_2, x_4, x_5 \text{ any scalars,}$$

so

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

is a basis for $\text{Nul}(A)$, and $\dim(\text{Nul}(A)) = 3$.

4.7 Rank

Recall that two matrices are **row equivalent** if there is a sequence of elementary row operations that transforms one into the other.

Important

Warning: If A and B are row equivalent $m \times n$ matrices, their their column spaces have the same dimension, but they are *different* subspaces of $\mathbb{R}^m$.

Example 4.69. Since

$$A = \begin{bmatrix} 1 & -3 & 0 \\ 5 & 0 & 4 \\ 0 & 1 & -1 \\ 0 & 0 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = B,$$

then both $\operatorname{rank}(A) = \operatorname{rank}(B) = 3$, meaning that

$$\dim \operatorname{col}(A) = \dim \operatorname{col}(B).$$

However, $\operatorname{col}(A)$ and $\operatorname{col}(B)$ are <u>not</u> the same subspace of $\mathbb{R}^4$. For example, the vector

$$\begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix}$$

is in $\operatorname{col}(A)$ but not in $\operatorname{col}(B)$, since all the vectors of $\operatorname{col}(B)$ have a zero on the last coordinate. We have

$$\operatorname{col}(A) = \operatorname{span}\left(\left\{\begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix}\right\}\right) \text{ and } \operatorname{col}(B) = \operatorname{span}\left(\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\right\}\right).$$

In contrast, to describe the null space of a matrix of A, we only need to know the reduced echelon form of A.

Remark 4.70. If two matrices A and B are row equivalent, then they have the same null space.

We can also talk about the row space of a matrix.

Definition 4.71. The **row space** of a matrix A is the set of all linear combinations of the rows of A.

Theorem 4.72. *If A and B are row equivalent matrices, then their row spaces are the same. If B is in echelon form, the nonzero rows of B form a basis for* $\mathrm{row}(A) = \mathrm{row}(B)$. *In particular, the dimension of* $\mathrm{row}(A)$ *is the number of pivots of A.*

Theorem 4.73. *Let A $m \times n$ matrix. The row space and column space of A have the same dimension.*

Note, however, that they are not the same space! The column space of an $m \times n$ matrix is a subspace of $\mathbb{R}^m$, and the row space is a subspace of $\mathbb{R}^n$.

We now redefine the rank in a fancier way, which we already knew to be the number of pivots in A:

Definition 4.74. The **rank** of a matrix A is

$$\mathrm{rank}(A) = \dim(\mathrm{col}(A)) = \dim(\mathrm{row}(A)).$$

The following is a very powerful and very famous theorem:

Theorem 4.75 (Rank–Nullity theorem)**.** *For any $m \times n$ matrix A,*

$$\mathrm{rank}(A) + \dim(\mathrm{Nul}(A)) = n.$$

The proof is easy with what we have said so far: the point is that the rank of A is the number of pivot columns, while the dimension of the null space of A is the number of columns that are *not* pivot columns. Adding the two together, we get the total number of columns, which is n.

The Rank–Nullity Theorem gets its name from the fact that the dimension of the null space is often called the nullity of the matrix.

Definition 4.76. The **nullity** of a matrix A is $\dim(\mathrm{Nul}(A))$.

4.8 Coordinates

Definition 4.77. Let $B = \{v_1, \dots, v_n\}$ be a basis for a vector space V. The **coordinates** of a vector v in V **relative to the basis** B are the unique scalars $c_1, \dots, c_n$ such that

$$v = c_1 v_1 + \dots + c_n v_n.$$

The **coordinate vector** of v **with respect to** B is the vector in $\mathbb{R}^n$ given by

$$[v]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Example 4.78. The coordinate vector of $v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ with respect to $B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ is

$$[v]_B = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

The coordinate vector of $v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ with respect to $C = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ is

$$[v]_C = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Example 4.79. Consider the following basis for $M_{2\times 2}$:

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

The coordinate vector of

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$$

with respect to B is

$$[A]_B = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}.$$

4.9 Linear Transformations

Definition 4.80. Let V and W be vector spaces. A **linear transformation** from V to W is a function $T\colon V \to W$ such that

(a) $T(u+v) = T(u) + T(v)$ for all u and v in V.

(b) $T(cu) = cT(u)$ for all u in V and all scalars c.

Example 4.81. Let V be any vector space. The **identity function** $\mathrm{id}_V\colon V \to V$

$$\mathrm{id}_V(v) = v$$

is a linear transformation.

Example 4.82. Let V and W be vector spaces. The **zero function** $Z\colon V \to W$

$$Z(v) = 0 \quad \text{for all } v \in V$$

that maps every vector in V to the zero vector in W is a linear transformation.

Example 4.83. The matrix transformations $\mathbb{R}^m \to \mathbb{R}^n$ we discussed before are linear transformations.

Linear transformations can be between very different vector spaces.

Example 4.84. Consider the derivative function $D\colon \mathbb{P}_2 \to \mathbb{P}_2$, which is given by

$$D(p(t)) = \frac{d}{dt}p(t).$$

Since taking derivatives preserves sums and multiplication by scalars, D is a linear transformation. More precisely, given polynomials $p(t), q(t)$ in $\mathbb{P}_2$ and a constant c, we see that

$$D(p(t) + q(t)) = \frac{d}{dt}\left(p(t) + q(t)\right) = \frac{d}{dt}p(t) + \frac{d}{dt}q(t) = D(p(t)) + D(q(t))$$

and

$$D(c\,p(t)) = \frac{d}{dt}(c\,p(t)) = c\frac{d}{dt}p(t) = cD(p(t)).$$

Example 4.85. We claim that the function

$$\begin{aligned} T : \mathbb{P}_2 &\longrightarrow \mathbb{R}^3 \\ a_0 + a_1 t + a_2 t^2 &\longmapsto \begin{bmatrix} a_1 \\ a_2 \\ a_1 + a_2 \end{bmatrix} \end{aligned}$$

is a linear transformation. To check that this T is indeed a linear transformation, we need to check it satisfies two things:

(a) We claim that $T(p + q) = T(p) + T(q)$ for all p and q in V. Indeed, for any two polynomials $p = a_0 + a_1 t + a_2 t^2$ and $q = b_0 + b_1 t + b_2 t^2$, we have

$$\begin{aligned} T(p + q) &= T((a_0 + a_1 t + a_2 t^2) + (b_0 + b_1 t + b_2 t^2)) \\ &= T((a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2) \\ &= \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ (a_1 + b_1) + (a_2 + b_2) \end{bmatrix} \\ &= \begin{bmatrix} a_1 \\ a_2 \\ a_1 + a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ a_2 + b_2 \end{bmatrix} \\ &= T(a_0 + a_1 t + a_2 t^2) + T(b_0 + b_1 t + b_2 t^2) \\ &= T(p) + T(q). \end{aligned}$$

(b) $T(cp) = cT(p)$ for all p in V and scalars c. Indeed, for all polynomials $p = a_0 + a_1 t + a_2 t^2$ and any scalar c, we have

$$\begin{aligned}
T(cp) &= T((ca_0) + (ca_1)t + (ca_2)t^2) \\
&= \begin{bmatrix} ca_1 \\ ca_2 \\ ca_1 + ca_2 \end{bmatrix} \\
&= c \begin{bmatrix} a_1 \\ a_2 \\ a_1 + a_2 \end{bmatrix} \\
&= cT(a_0 + a_1 t + a_2 t^2) \\
&= cT(p(t)).
\end{aligned}$$

Definition 4.86. The **kernel** of a linear transformation $T\colon V \to W$ is the set

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

The **range** or **image** of T is the set

$$\operatorname{im}(T) = T(V) = \{w \text{ in } W : w = T(v) \text{ for some } v \text{ in } V\}.$$

Theorem 4.87. *Let $T\colon V \to W$ be a linear transformation. The kernel of T is a subspace of V, and the image of T is a subspace of W.*

Example 4.88. Consider the linear transformation

$$\begin{aligned}
T : \mathbb{P}_2 &\longrightarrow \mathbb{R}^3 \\
a_0 + a_1 t + a_2 t^2 &\longmapsto \begin{bmatrix} a_1 \\ a_2 \\ a_1 + a_2 \end{bmatrix}
\end{aligned}$$

Theorem 4.87 says that the kernel of T is a subspace of $\mathbb{P}_2$:

$$\ker(T) = \{\text{constant polynomials in } \mathbb{P}_2\}.$$

Moreover, the image of T is a subspace of $\mathbb{R}^2$:

$$\operatorname{im}(T) = T(\mathbb{P}_2) = \left\{ \begin{bmatrix} a_1 \\ a_2 \\ a_1 + a_2 \end{bmatrix} : a_1, a_2 \text{ any real numbers} \right\} = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\} \right).$$

Discussion 4.89. Let V and W be vector spaces and consider a linear transformation $T\colon V \to W$. Once we know the images of the elements of a basis $\mathcal{B}$ for V, that completely determines the linear transformation T. Indeed, given any vector $v \in V$, there is a unique way to write

$$v = v_1 b_1 + \cdots + v_n b_n$$

in terms of basis vectors $b_1, \ldots, b_n$, and since T is linear, we see that

$$T(v) = T(v_1 b_1 + \cdots + v_n b_n) = v_1 T(b_1) + \cdots v_n T(v_n).$$

Definition 4.90. Let V and W be finite-dimensional vector spaces and let $\mathcal{B} = \{b_1, \ldots, b_n\}$ and $\mathcal{C} = \{c_1, \ldots, c_m\}$ be basis for V and W, respectively. Consider a linear transformation $T\colon V \to W$. For each vector w in W, write $[w]_{\mathcal{C}}$ for the vector representing w in the basis $\mathcal{C}$. The **matrix representing** T in the bases $\mathcal{B}$ and $\mathcal{C}$ is the matrix

$$A_{\mathcal{C}\leftarrow\mathcal{B}} = \left[[T(b_1)]_{\mathcal{C}} \quad \cdots \quad [T(b_n)]_{\mathcal{C}} \right].$$

Remark 4.91. Earlier in the semester we talked about the standard matrix for a linear transformation $T\colon \mathbb{R}^m \to \mathbb{R}^n$. In this new language, we were really writing the matrix for T with respect to the standard basis for $\mathbb{R}^m$ and $\mathbb{R}^n$.

Example 4.92. Consider the vector space $\mathbb{P}_2$ and its standard basis $\mathcal{B} = \{1, t, t^2\}$. Let us write the matrix representing the derivative transformation $D\colon \mathbb{P}_2 \to \mathbb{P}_2$ we discussed in Example 4.84 in the basis $\mathcal{B}$. Note here that the domain and codomain of our linear transformation are the same, so we can use the same basis on both sides. So we need to find the images of 1, t, and t^2, and then write them with respect to the basis $\mathcal{B}$.

$$\left.\begin{aligned} D(1) = 0 &\implies [D(1)]_{\mathcal{B}} = [0]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ D(t) = 1 &\implies [D(t)]_{\mathcal{B}} = [1]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ D(t^2) = 2t &\implies [D(t^2)]_{\mathcal{B}} = [2t]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}. \end{aligned}\right. \qquad \implies A_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

We can now use this to compute the image of any vector via our linear transformation. For example, suppose we want to calculate $D(3 + 7t - t^2)$. We could of course use our knowledge that D simply takes the derivative with respect to t; or we can use the matrix directly. To do that, first we need to write $p = 3 + 7t - t^2$ in our chosen basis $\mathcal{B}$:

$$[p]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 7 \\ -1 \end{bmatrix}.$$

Now we mutiply the matrix by this vector:

$$[D(p)]_{\mathcal{B}} = A_{\mathcal{B}\leftarrow\mathcal{B}}[p]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ -1 \end{bmatrix} = \begin{bmatrix} 7 \\ -2 \\ 0 \end{bmatrix}.$$

This tells us the coordinates of $D(p)$ with respect to $\mathcal{B}$, meaning

$$D(p) = 7 - 2t.$$

Example 4.93. Consider the linear transformation $T\colon M_{2\times 2} \to \mathbb{P}_2$ given by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = d + (b-c)t + at^2.$$

Fix the basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

for $M_{2\times 2}$ and the basis

$$\mathcal{C} = \{1, t, t^2\}$$

for $\mathbb{P}_2$. Let us compute the matrix for T relative to the bases $\mathcal{B}$ and $\mathcal{C}$:

$$T\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = t^2 \quad T\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = t \quad T\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) = -t \quad T\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = 1$$

so

$$[T(b_1)]_{\mathcal{C}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad [T(b_2)]_{\mathcal{C}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad [T(b_3)]_{\mathcal{C}} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \quad [T(b_4)]_{\mathcal{C}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

And therefore the matrix for T relative to $\mathcal{B}$ and $\mathcal{C}$ is

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

Now suppose we want to compute $T(v)$, where $v = \begin{bmatrix} 2 & 3 \\ 1 & -2 \end{bmatrix}$. Since

$$[v]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \\ 1 \\ -2 \end{bmatrix} \implies [T(v)]_{\mathcal{C}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ 2 \end{bmatrix},$$

so $T(v) = -2 + 2t^2 + 2t^3$.

Chapter 5

Determinants

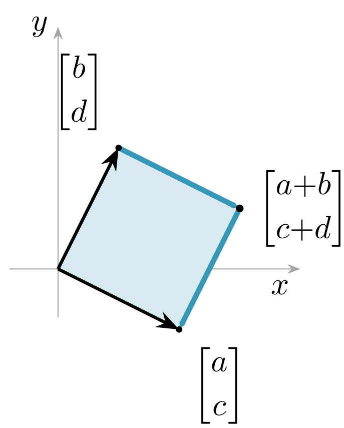
5.1 Geometric Meaning of the determinant

We will soon be defining the determinant of any square matrix. Before we give the definition, let us talk about the geometric meaning of the determinant.

Discussion 5.1. Consider any 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and draw the parallelogram determined by the columns of A.



The determinant

$$\det(A) = ad - bc$$

is (up to sign) the area of this parallelogram, meaning

$$\det(A) = \pm \text{area of the parallelogram determined by the columns of } A.$$

The sign is determined as follows:

- $+$ if the angle from the first column to the second goes counterclockwise;
- $-$ if the angle from the first column to the second goes clockwise.

In particular, for a 2×2 matrix A:

$$\begin{aligned}\det A = 0 &\iff \text{the area determined by the columns of } A \text{ is not 2-dimensional}\\ &\iff \text{the two columns of } A \text{ are linearly dependent.}\end{aligned}$$

Note that the area determined by two vectors is not 2-dimensional precisely when the two vectors are multiples of each other, which as we know is equivalent to the two vectors being linearly dependent.
For a 3×3 matrix A, $\det(A)$ is (up to sign) the volume of the parallelepiped spanned by the three column vectors; the sign is given by the right-hand rule.

In general, for an $n \times n$ matrix,

$$\det(A) = \pm n\text{-dimensional volume of the parallelotope spanned by the columns of } A.$$

If the columns fail to span an n-dimensional region (i.e., they are linearly dependent), then $\det(A) = 0$.

Notation 5. Some authors write $|A|$ for $\det(A)$. More precisely,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{ denotes } \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right).$$

5.2 Computing determinants

Definition 5.2. Let $A = (a_{ij})$ be an $n \times n$ matrix. For $1 \leqslant i, j \leqslant n$, let A_{ij} denote the matrix obtained from A by deleting row i and column j. The (i, j)**-minor** M_{ij} is the determinant of the $(n-1) \times (n-1)$ matrix A_{ij} obtained by deleting row i and column j from A. The (i, j)**-cofactor** is

$$C_{ij} = (-1)^{i+j} M_{ij} = (-1)^{i+j} \det(A_{ij}).$$

Discussion 5.3 (Cofactor expansion along a Row)**.** The determinant of A can be obtained by expanding along the first row, as follows:

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}) = a_{11} \det(A_{11}) \pm a_{12} \det(A_{12}) \pm \cdots \pm a_{1n} \det(A_{1n}).$$

Moreover, we can also expand along a different row: for any fixed row i,

$$\det A = \sum_{j=1}^{m} (-1)^{i+j} a_{ij} \det(A_{ij}) = \pm a_{i1} \det(A_{i1}) \pm a_{i2} \det(A_{i2}) \pm \cdots \pm a_{in} \det(A_{in}).$$

The sign for $a_{i1} \det(A_{i1})$ is $(-1)^{i+1}$, and after that the signs alternate.

Discussion 5.4 (Cofactor Expansion Along a Column)**.** The determinant of A can be obtained by expanding along the first column, as follows:

$$\det A = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(A_{i1}) = a_{11} \det(A_{11}) - a_{21} \det(A_{21}) \pm \cdots + (-1)^{n+1} a_{n1} \det(A_{n1}).$$

For any fixed column j,

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

Note that the sign for $a_{1j} \det(A_{1j})$ is $(-1)^{1+j}$, and after that the signs alternate.
The sign pattern $(-1)^{i+j}$ follows the standard checkerboard of plus/minus signs:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Remark 5.5 (Expanding Along Different Rows/Columns Gives the Same Determinant)**.** When we compute $\det A$ by expanding along different rows/columns, we obtains the same value each time. You might try experimenting with a few examples to see this.

Example 5.6. Let us calculate the determinant of

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}.$$

To do this, we will expand along row 1:

$$\begin{aligned} \det A &= (+)\, 1 \cdot \det\left(\begin{bmatrix} 4 & -1 \\ -2 & 0 \end{bmatrix} \right) \;-\; 5 \cdot \det\left(\begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} \right) \;+\; 0 \cdot \det\left(\begin{bmatrix} 2 & 4 \\ 0 & -2 \end{bmatrix} \right) \\ &= 1 \cdot \big(4 \cdot 0 - (-1)(-2)\big) \;-\; 5 \cdot \big(2 \cdot 0 - (-1) \cdot 0\big) \;+\; 0 \cdot \big(2 \cdot (-2) - 4 \cdot 0\big) \\ &= 1 \cdot (0 - 2) \;-\; 5 \cdot (0 - 0) \;+\; 0 \cdot (-4 - 0) \\ &= -2. \end{aligned}$$

If instead we use row 2, we get

$$\begin{aligned} \det A &= (-)\, 2 \cdot \det\left(\begin{bmatrix} 5 & 0 \\ -2 & 0 \end{bmatrix} \right) \;+\; 4 \cdot \det\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) \;-\; (-1) \cdot \det\left(\begin{bmatrix} 1 & 5 \\ 0 & -2 \end{bmatrix} \right) \\ &= -2 \cdot \big(5 \cdot 0 - 0 \cdot (-2)\big) \;+\; 4 \cdot \big(1 \cdot 0 - 0 \cdot 0\big) \;+\; 1 \cdot \big(1 \cdot (-2) - 5 \cdot 0\big) \\ &= -2 \cdot (0 - 0) \;+\; 4 \cdot (0 - 0) \;+\; (-2 - 0) \\ &= -2. \end{aligned}$$

Note that this is the same result as above (as it should be!).
We can also, for example, use column 3:

$$\det A = (+)\,0 \cdot \det\begin{bmatrix} 2 & 4 \\ 0 & -2 \end{bmatrix} \;-\; (-1)\cdot\det\begin{bmatrix} 1 & 5 \\ 0 & -2 \end{bmatrix} \;+\; 0\cdot\det\begin{bmatrix} 1 & 5 \\ 2 & 4 \end{bmatrix}$$

$$\begin{aligned} \det A = \qquad & = 0\cdot\Big(2\cdot(-2) - 4\cdot 0\Big) \;+\; 1\cdot\Big(1\cdot(-2) - 5\cdot 0\Big) \;+\; 0\cdot\Big(1\cdot 4 - 5\cdot 2\Big) \\ & = 0\cdot(-4-0) \;+\; (-2-0) \;+\; 0\cdot(4-10) \\ & = -2. \end{aligned}$$

And once more, we get the same result (yay!).

Example 5.7. It might seem a little silly, but this works even for a 2×2 matrix: to compute the determinant of

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

by expanding along the first row, we get

$$\det(A) = +a\det([d]) - b\det([c]) = ad - bc.$$

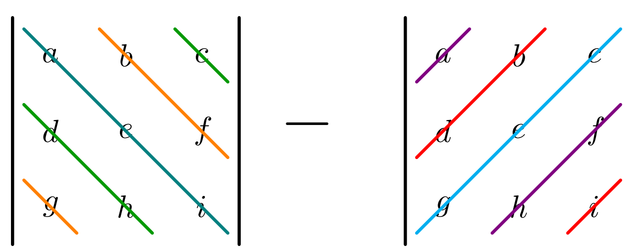
Example 5.8 (Determinant Formula for a 3×3 Matrix)**.** For a 3×3 matrix

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

a common explicit formula for its determinant is

$$\det(A) = aei + bfg + cdh - ceg - bdi - afh.$$

Here is is in a picture:



Here you should interpret the same color as meaning those entries get multiplies together, and the products of each color get added together. Ultimately, we compute

$$\det(A) = (\ \mathrm{aei} + \mathrm{bfg} + \mathrm{cdh}\ ) - (\ \mathrm{ceg} + \mathrm{bdi} + \mathrm{afh}\ ).$$

Example 5.9. Let us calculate the determinant of

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

again, this time using the formula in Example 5.8. Just like before, we get

$$\begin{aligned} \det(A) &= (1 \cdot 5 \cdot 0 + 5 \cdot (-1) \cdot 0 + 2 \cdot (-2) \cdot 0) - (0 \cdot 4 \cdot 0 + 5 \cdot 2 \cdot 0 + (-1) \cdot (-2) \cdot 1) \\ &= 2. \end{aligned}$$

5.3 Properties of the determinant

The most important fact about determinants is that the determinant detects invertibility.

Theorem 5.10 (Invertibility and the Determinant)**.** *A square matrix A is invertible if and only if* $\det(A) \neq 0$.

The idea is that A is invertible if and only if the columns of A form a basis for $\mathbb{R}^n$, which is equivalent to asking that the columns of A do indeed determine an n-dimensional object as opposed to an object of smaller dimension, in which case the determinant would be zero.

We can easily compute the determinant of a triangular matrix.

Definition 5.11. A square matrix A is **upper triangular** if all entries below the main diagonal are zero, and **lower triangular** if all entries above the main diagonal are zero. A **triangular matrix** is any matrix that is either upper triangular or lower triangular.

Example 5.12. Consider the matrices

$$U = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \quad \text{and} \quad L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 3 & 4 \end{bmatrix}.$$

Note that U is upper triangular and L is lower triangular.

Theorem 5.13 (Determinant of a Triangular Matrix)**.** *If A is a triangular matrix, then the determinant of A is just the product of the entries in the main diagonal. More precisely, if A is an $n \times n$ triangular matrix,*

$$\det(A) = \prod_{i=1}^{n} a_{ii}.$$

Corollary 5.14. *The determinant of the identity matrix is* 1*.*

Example 5.15. The matrix

$$A = \begin{bmatrix} -1 & -1 & 7 & 5 \\ 0 & -5 & 42 & 2 \\ 0 & 0 & 2 & 13 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

is upper triangular, so

$$\det(A) = (-1) \cdot (-5) \cdot 2 \cdot 5 = 50.$$

Let us now collect some nice properties of determinants:

Theorem 5.16. *Let A and B be $n \times n$ matrices and $c \in \mathbb{R}$ any scalar. Then*

(a) $\det(A^{\mathsf{T}}) = \det(A)$.

(b) $\det(AB) = \det(A)\det(B)$.

(c) $\det(cA) = c^n \det(A)$.

Note that there is no formula for $\det(A+B)$, and that in fact anything can happen. (Try it out in some examples!)

Remark 5.17. As an easy consequence of the properties in Theorem 5.16, we see that if A is invertible, then

$$\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I) = 1 \implies \det(A^{-1}) = \frac{1}{\det(A)}.$$

Remark 5.18. We can also see from here that the product of two invertible matrices is invertible. Indeed, if A and B are both invertible, then

$$\det(A) \neq 0, \det(B) \neq 0 \implies \det(AB) = \det(A)\det(B) \neq 0.$$

Therefore, AB is also invertible.

Theorem 5.19 (Effect of Elementary Row Operations on det)**.** *Let A and B be two square matrices of the same size. Then:*

(a) Swapping rows multiplies the determinant by -1: if B is obtained from A by switching two rows, then $\det(B) = -\det(A)$.

(b) Adding a multiple of one row to another row leaves the determinant unchanged: if B is obtained from A by adding a multiple of a row to another, then $\det(B) = \det(A)$.

(c) Multiplying a row by a scalar k multiplies det *by k: if B is obtained from A by multiplying a row by k, then* $\det(B) = k\det(A)$.

Example 5.20. Consider the matrix

$$A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}.$$

Let us try to find its determinant by row reducing to a simpler matrix. First, we note that

$$A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix} \xrightarrow{R_1 \to \frac{1}{2}R_1} \begin{bmatrix} 1 & -4 & 3 & 4 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix} = B.$$

The rules from Theorem 5.19 tell us that

$$\det(B) = \frac{1}{2}\det(A) \implies \det(A) = 2\det(B).$$

Now let us find the determinant of B. We have

$$B = \begin{bmatrix} 1 & -4 & 3 & 4 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix} \xrightarrow[R_3 \to R_3 + 3R_1]{R_2 \to R_2 - 3R_1 \quad R_4 \to R_4 - R_1} \begin{bmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & -12 & 10 & 10 \\ 0 & 0 & -3 & 2 \end{bmatrix}$$

$$\xrightarrow{R_3 \to R_3 + 4R_2} \begin{bmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & -3 & 2 \end{bmatrix} \xrightarrow{R_4 \to R_4 - \frac{1}{2}R_3} \begin{bmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} = C.$$

Note that every step we have taken consists of adding a multiple of one row to another, so by Theorem 5.19 the determinant does not change. Thus

$$\det(C) = \det(B).$$

But now C is upper triangular, so its determinant is just the product of the diagonal entries. Thus

$$\det(C) = 1 \cdot 3 \cdot (-6) \cdot 1 = -18.$$

We conclude that

$$\det(A) = 2\det(B) = 2\det(C) = 2 \cdot (-18) = -36.$$

More generally, we can always calculate the determinant by looking at any echelon for of our matrix.

Theorem 5.21 (Echelon Form and Determinant)**.** *Suppose A can be brought to a (not necessarily reduced) echelon form E using only no rescaling, meaning that we only do r many row switches and operations where we add one row to another. Then*

$$\det A = \begin{cases} (-1)^r \cdot \text{product of the entries in the pivot positions of } E & \text{if } A \text{ is invertible} \\ 0 & \text{otherwise.} \end{cases}$$

The fact that $\det(A) = \det(A^{\mathsf{T}})$ also has some nice consequences: anything we can say about the columns of A that relates to the determinant has a counterpart for the rows.

Theorem 5.22. *Let A be any square matrix.*

(a) If A has a column of zeroes, then $\det(A) = 0$.

(b) If A has a row of zeroes, then $\det(A) = 0$.

(c) If one of the columns of A is a scalar multiple of another, then $\det(A) = 0$.

(d) If one of the rows of A is a scalar multiple of another, then $\det(A) = 0$.

We are now ready to update the Inverse Matrix Theorem:

Theorem 5.23 (Inverse Matrix Theorem)**.** *Let A be any $n \times n$ matrix, and write $I = I_n$. The following are equivalent:*

(a) A is invertible.

(b) There exists B such that $BA = I$.

(c) There exists B such that $AB = I$.

(d) We have $A \sim I$.

(e) The matrix A has rank n.

(f) The equation $Ax = 0$ has only the trivial solution.

(g) The columns of A form a linearly independent set.

(h) The rows of A form a linearly independent set.

(i) The linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = Ax$ is injective.

(j) The linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = Ax$ is surjective.

(k) The equation $Ax = b$ has at least one solution for each b.

(l) The transpose A^{T} is invertible.

(m) The determinant of A is nonzero: $\det(A) \neq 0$.

As a consequence, given any n vectors $v_1, \dots, v_n$ in $\mathbb{R}^n$, we can determine if they form a basis for $\mathbb{R}^n$ by determining wheter

$$\det\left(\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}\right).$$

Equivalently, this will tell us if $v_1, \dots, v_n$ are linearly independent, or equivalently if they span $\mathbb{R}^n$.

5.4 Inverse matrices using determinants

Definition 5.24 (Cofactor Matrix and Adjoint)**.** Let A be a square $n \times n$ matrix. The **cofactor matrix** of A is the $n \times n$ matrix $\operatorname{cof}(A) = (C_{ij})$ where

$$C_{ij} = (-1)^{i+j} \det(A_{ij}).$$

The **adjoint** of A is the transpose of the cofactor matrix:

$$\operatorname{adj}(A) = C^{\mathsf{T}}.$$

Example 5.25. Consider any 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Its cofactor and adjoint matrices are

$$\operatorname{cof}(A) = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix} \quad \text{and} \quad \operatorname{adj}(A) = \begin{bmatrix} d & -b \\ -d & a \end{bmatrix}.$$

Sound familiar? Indeed, it should!

Theorem 5.26 (Inverse via Adjoints)**.** *If A is invertible, then*

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A).$$

Thus

$$A \cdot \operatorname{adj}(A) = \operatorname{adj}(A) \cdot A = \det(A) \cdot I.$$

Example 5.27. Consider

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

and lets find its cofactor matrix, adjoint, and inverse. First, we calculate the 9 cofactors:

$$\begin{aligned} C_{11} &= \begin{vmatrix} -1 & 1 \\ 4 & -2 \end{vmatrix} & C_{12} &= -\begin{vmatrix} 1 & 1 \\ 1 & -2 \end{vmatrix} & C_{13} &= \begin{vmatrix} 1 & -1 \\ 1 & 4 \end{vmatrix} \\ &= (-1)(-2) - (1)(4) & &= -\big((1)(-2) - 1 \cdot 1\big) & &= (1)(4) - (-1)(1) \\ &= -2 & &= -(-3) = 3 & &= 5 \end{aligned}$$

$$\begin{aligned} C_{21} &= -\begin{vmatrix} 1 & 3 \\ 4 & -2 \end{vmatrix} & C_{22} &= \begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix} & C_{23} &= -\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} \\ &= -\big((1)(-2) - (3)(4)\big) & &= (2)(-2) - (3)(1) & &= -\big((2)(4) - (1)(1)\big) \\ &= -(-14) = 14 & &= -4 - 3 = -7 & &= -(8 - 1) = -7 \end{aligned}$$

$$\begin{aligned} C_{31} &= \begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} & C_{32} &= -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} & C_{33} &= \begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} \\ &= (1)(1) - (3)(-1) & &= -\big((2)(1) - (3)(1)\big) & &= (2)(-1) - (1)(1) \\ &= 1 + 3 = 4 & &= -(2 - 3) = 1 & &= -2 - 1 = -3. \end{aligned}$$

Thus we get

$$\operatorname{cof}(A) = \begin{bmatrix} -2 & 3 & 5 \\ 14 & -7 & -7 \\ 4 & 1 & -3 \end{bmatrix} \qquad \text{and} \qquad \operatorname{adj}(A) = (\operatorname{cof} A)^{\mathsf{T}} = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix}.$$

Now note that

$$A \cdot \operatorname{adj} A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix} \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix} = \begin{bmatrix} 14 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 14 \end{bmatrix}.$$

The numbers on the diagonal are simply the determinant of A. We conclude that

$$\det A = 14.$$

Finally, we get

$$A^{-1} = \frac{1}{14} \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{7} & 1 & \frac{2}{7} \\ \frac{3}{14} & -\frac{1}{2} & \frac{1}{14} \\ \frac{5}{14} & -\frac{1}{2} & -\frac{3}{14} \end{bmatrix}.$$

Chapter 6

Eigenvalues and eigenvectors

6.1 Eigenvalues

Definition 6.1. Let A be a square matrix. A real number λ is called an **eigenvalue** of A if there exists a nontrivial solution x to the equation $Ax = \lambda x$. A vector x is a **eigenvector** of A if $Ax = \lambda x$ for some real number λ. In this case, we say x is an eigenvector associated to the eigenvalue λ.

Example 6.2. Let

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}.$$

Since

$$\begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4 \begin{bmatrix} 6 \\ -5 \end{bmatrix},$$

the vector $\begin{bmatrix} 6 \\ -5 \end{bmatrix}$ is an eigenvector associated to the eigenvalue -4. On the other hand,

$$\begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda \begin{bmatrix} 3 \\ -2 \end{bmatrix},$$

so $\begin{bmatrix} 3 \\ -2 \end{bmatrix}$ is not an eigenvector of A.

How do we find the eigenvalues for a given square matrix? An eigenvector x of A satisfies the equation $Ax = \lambda x$ for some real number λ, or equivalently,

$$Ax - \lambda x = 0 \Leftrightarrow (A - \lambda I)x = 0.$$

So the eigenvalues λ of A are the real numbers λ such that

$$(A - \lambda I)x = 0$$

has nontrivial solutions. When does that happen? When $A - \lambda I$ is not invertible.

Theorem 6.3. *Let A be a square matrix. A real number λ is an eigenvalue of A if and only if*
$$\det(A - \lambda I) = 0.$$

Definition 6.4. The equation
$$\det(A - \lambda I) = 0$$
is called the **characteristic equation** of A, and the polynomial $\det(A - \lambda I)$ is called the **characteristic polynomial** of A.

Example 6.5. Let
$$A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}.$$
The eigenvalues of A are the solutions to the characteristic equation:
$$\begin{aligned} \det\left(\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) &= 0 \\ \iff \begin{vmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{vmatrix} &= 0 \\ \iff (2-\lambda)(-6-\lambda) - 9 &= 0 \\ \iff -12 - 2\lambda + 6\lambda + \lambda^2 - 9 &= 0 \\ \iff \lambda^2 + 4\lambda - 21 &= 0, \end{aligned}$$
so
$$\lambda = \frac{-4 \pm \sqrt{4^2 - 4 \cdot 1 \cdot (-21)}}{2} = \frac{-4 \pm \sqrt{16 + 84}}{2} = \frac{-4 \pm 10}{2},$$
so the eigenvalues of A are
$$\lambda = \frac{-4 + 10}{2} = \frac{6}{2} = 3 \quad \text{and} \quad \lambda = \frac{-4 - 10}{2} = -\frac{14}{2} = -7.$$

Theorem 6.6. *The eigenvalues of a triangular matrix are the entries on its main diagonal.*

Example 6.7. The eigenvalues of
$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 5 & 7 \\ 0 & 0 & 6 & 8 \\ 0 & 0 & 0 & 12 \end{bmatrix}$$
are 1, 3, 6, and 12.

Remark 6.8. Let A be any $n \times n$ matrix. Notice that 0 is an eigenvalue of A if and only if $Ax = 0$ has nontrivial solutions.

Theorem 6.9. *Let A be any $n \times n$ matrix. The following are equivalent:*

(a) *The number* 0 *is an eigenvalue of A.*
(b) *The null space of A has dimension at least* 1.
(c) *A has free variables.*
(d) $\det(A) = 0$.
(e) *A is not invertible.*

6.2 Eigenspaces

We have discussed how to find the eigenvalues of a given matrix. How do we find the corresponding eigenvectors?

Example 6.10. Let

$$A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}.$$

The eigenvalues of A are the solutions to the equation

$$\det\left(\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = 0,$$

and we have computed these before: the eigenvalues of A are $\lambda = 3$ and $\lambda = -7$. How do we find eigenvectors corresponding to these eigenvalues?
The eigenvectors associated to 3 are solutions to the equation $(A - 3I)x = 0$, so the set of *all* the eigenvectors of A associated to 3 is precisely $\mathrm{Nul}(A - 3I)$.

$$\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - 3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix}.$$

So the eigenvectors of A associated to 3 are of the form

$$\begin{cases} x_1 = 3x_2 \\ x_2 \text{ is a free variable} \end{cases} \qquad \Longrightarrow \qquad x = x_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

This tells us two things: that $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ is an eigenvector for A associated to the eigenvalue 3, and in the eigenvectors of A associated to 3 are simply all the multiples of $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

Definition 6.11. Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. The **eigenspace** of A associated to λ is the set of all eigenvectors of A associated to λ: all the solutions to the equation $(A - \lambda I)x = 0$.

Since the eigenspace of A associated to λ is simply the null space of the matrix $A - \lambda I$, the next theorem is an immediate corollary of the fact that the null space of an $n \times n$ matrix is a subspace of $\mathbb{R}^n$.

Theorem 6.12. *Let A be an $n \times n$ matrix, and let λ be an eigenvalue of A. The eigenspace of A associated to λ is a subspace of $\mathbb{R}^n$.*

To find the eigenspace of A associated to λ is the same as finding the null space of $A - \lambda I$.

Example 6.13. Let's find the eigenspace of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$ associated to -7. Since

$$\begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & 0 \end{bmatrix},$$

the vectors in the eigenspace are of the form

$$\begin{cases} x_1 = -\frac{1}{3}x_2 \\ x_2 \text{ is a free variable} \end{cases} \qquad \implies \qquad x = x_2 \begin{bmatrix} -\frac{1}{3} \\ 1 \end{bmatrix}.$$

So the eigenspace of A associated to the eigenvalue -7 is the set of multiples of $\begin{bmatrix} -\frac{1}{3} \\ 1 \end{bmatrix}$.

Example 6.14. Let's find the eigenvalues of

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

We need to find the solutions to the characteristic equation of this matrix, which is

$$\begin{aligned} \begin{vmatrix} 3-\lambda & -2 \\ 2 & -1-\lambda \end{vmatrix} &= 0 \\ \Leftrightarrow (3-\lambda)(-1-\lambda) + 4 &= 0 \\ \Leftrightarrow -3 - 3\lambda + \lambda + \lambda^2 + 4 &= 0 \\ \Leftrightarrow \lambda^2 - 2\lambda + 1 &= 0 \\ \Leftrightarrow (\lambda - 1)^2 &= 0. \end{aligned}$$

so there is only one eigenvalue, 1, but it has **multiplicity** 2. Since

$$A - I = \begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

the eigenspace of A associated to 1 is the set of vectors of the form

$$\begin{cases} x_1 = x_2 \\ x_2 \text{ is a free variable} \end{cases} \qquad \implies \qquad x = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

or

$$\operatorname{span}\left(\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}\right).$$

Definition 6.15. The **(algebraic) multiplicity** of the eigenvalue λ is the multiplicity of λ as a root of the equation $\det(A - \lambda I) = 0$.

Note, however, that not all eigenspaces are lines; here are some extreme examples:

Example 6.16. Fix $n > 1$, and let Z be the zero $n \times n$ matrix. Notice that

$$Z - \lambda I = -\lambda I \implies \det(Z - \lambda I) = (-\lambda)^n,$$

so $\lambda = 0$ is the only eigenvalue of Z, with multiplicity n. The eigenspace of Z corresponding to 0 is $\mathbb{R}^n$.

Example 6.17. Fix $n > 1$, and let I be the $n \times n$ identity matrix. Notice that

$$I - \lambda I = (1-\lambda)I \implies \det(Z - \lambda I) = (1-\lambda)^n,$$

so $\lambda = 1$ is the only eigenvalue of I. The eigenspace of Z corresponding to 1 is $\mathbb{R}^n$. Notice here that the characteristic polynomial of I is

$$(1-\lambda)^n.$$

This means that the unique eigenvalue 1 has multiplicity n.

Example 6.18. Consider a 6×6 matrix A with characteristic polynomial $\lambda^6 - 4\lambda^5 - 12\lambda^4$. This can be factored as

$$\lambda^6 - 4\lambda^5 - 12\lambda^4 = \lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2).$$

So 0 is an eigenvalue of A with multiplicity 4, 6 has multiplicity 1 and -2 has multiplicity 1.

Theorem 6.19. *Let A be a square matrix and λ be an eigenvalue of A. The dimension of the eigenspace of A associated to λ is at most the multiplicity of λ as a root of the characteristic polynomial of A.*

In particular, the dimensions of the eigenspaces associated to A add up to at most the number of columns (equivalently, rows) of A. We will explore this more in the next section. This is why sometimes one refers to the dimension of the eigenspace associated to λ as the **geometric multiplicity** of λ, while its multiplicity as a root of the characteristic polynomial is the **algebraic multiplicity** of λ.

Theorem 6.20. *Let A be a square matrix. Any two nonzero eigenvectors associated to different eigenvalues of A must be linearly independent.*

Remark 6.21. Suppose that x_1 and x_2 are two eigenvectors for A corresponding to two *different* eigenvalues λ_1 and λ_2. Then the set $\{x_1, x_2\}$ is definitely linearly independent: if it was a linearly dependent set, then x_2 would be a multiple of x_1. But all multiples of x_1 are eigenvectors for A associated to the eigenvalue λ_1, so they cannot also be eigenvectors for A associated to the (different) eigenvalue λ_2. Therefore, x_2 is not a multiple of x_1, and $\{x_1, x_2\}$ is a linearly independent set.

Definition 6.22. Let A and B be two $n \times n$ matrices. We say that A and B are **similar** if there exists an invertible matrix P such that $A = PBP^{-1}$.

Notice that if A is similar to B, then B is also similar to A, since we can take $Q = P^{-1}$ to obtain $B = QAQ^{-1}$.

Theorem 6.23. *Let A and B be two $n \times n$ matrices. If A is similar to B, then A and B have the same eigenvalues with the same multiplicity.*

Warning! Just because two matrices A and B have the same eigenvalues, that does not mean that A and B are similar. For example,

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

both have the eigenvalue 2 with multiplicity 2, but these matrices are not similar.

Warning! Similarity is not the same as row equivalence. Row operations on a matrix usually change its eigenvalues.

6.3 Diagonalization

What is all this for?

Definition 6.24. An $n \times n$ matrix A is said to be **diagonalizable** if it is similar to a diagonal matrix.

Theorem 6.25. *An $n \times n$ matrix A is diagonalizable if and only if there exist n linearly independent eigenvectors for A. Moreover, if A is diagonalizable, then we can find a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$ by doing the following:*

- *The columns of the matrix P are linearly independent eigenvectors for A.*
- *The matrix D is a diagonal matrix whose main diagonal entries are the eigenvalues of A, each appearing as many times as their multiplicity, and appearing in the same order as the corresponding eigenvectors in P.*

So A is diagonalizable if and only if we can find a basis for $\mathbb{R}^n$ whose elements are all eigenvectors of A.

Theorem 6.26. *An $n \times n$ matrix A is diagonalizable if and only if the sum of the dimensions of its eigenspaces is n.*

In this case, to find a basis for $\mathbb{R}^n$ of eigenvectors of A we just need to find basis for each eigenspace; collecting all those vectors together will give us a basis for $\mathbb{R}^n$. In general, the eigenspace associated to λ has dimension at most the multiplicity of λ. So if at least one of the eigenspaces of A has dimension strictly smaller than the multiplicity of the corresponding eigenvalue, then A is not diagonalizable.

Example 6.27. Let us try to diagonalize the matrix

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}.$$

Step 1: Find the eigenvalues of A.

$$\begin{vmatrix} 1-\lambda & 3 & 3 \\ -3 & -5-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{vmatrix} = (1-\lambda)(-5-\lambda)(1-\lambda) - 27 - 27 - (9(-5-\lambda) - 9(1-\lambda) - 9(1-\lambda))$$
$$= -\lambda^3 - 3\lambda^2 + 4$$
$$= -(\lambda-1)(\lambda+2)^2.$$

So the eigenvalues are 1 with multiplicity 1 and -2 with multiplicity 2.

Step 2: Find three linearly independent eigenvalues for A.

This is the most important step. If this step fails, then A is not diagonalizable. If this step works out, then A is definitely diagonalizable.
First, let us focus on the eigenspace associated to 1.

$$A - 1I = \begin{bmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

so

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\} \text{ is a basis for the eigenspace associated to 1.}$$

Moreover,

$$A + 2I = \begin{bmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

so

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\} \text{ is a basis for the eigenspace associated to } -2.$$

So the following is a basis for $\mathbb{R}^3$ of eigenvectors of A:

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

In particular, A is diagonalizable.

Step 3: Write the matrix P, whose columns are the basis of eigenvectors we found.

$$P = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

Step 4: Write the diagonal matrix D, whose diagonal entries are the eigenvalues of A, in the the same order as we put our eigenvectors in P.

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

We have diagonalized A: its diagonalization is $A = PDP^{-1}$, where

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \text{ and } P = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

Not every matrix is diagonalizable.

Example 6.28. Consider

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

Its characteristic equation is $(3-\lambda)^2$, so 3 is the only eigenvalue of A, and it has multiplicity 2. Since

$$A - 3I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

has only one free variable, the dimension of the eigenspace of A associated to 3 is 1, not 2. Therefore, A is not diagonalizable.

6.4 Applications of diagonalization

If we have an $n \times n$ matrix A and we want to compute A^k for some $k > 1$, in general that can be a computationally expensive process. But things get a lot simpler if A is diagonalizable. First, because taking powers of diagonal matrices is very simple:

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ & & \ddots & \\ & & & d_n \end{bmatrix}^k = \begin{bmatrix} d_1^k & 0 & \cdots & 0 \\ 0 & d_2^k & \cdots & 0 \\ & & \ddots & \\ & & & d_n^k \end{bmatrix}.$$

Moreover, if P is an invertible matrix and D is a diagonal matrix such that $A = PDP^{-1}$, then

$$A^k = \underbrace{A \cdots A}_{k \text{ times}} = PD\underbrace{(P^{-1}P)}_{I}D\underbrace{(P^{-1}P)}_{I} \cdots \underbrace{(P^{-1}P)}_{I}DP^{-1} = P\underbrace{D \cdots D}_{k \text{ times}}P^{-1} = PD^kP^{-1}.$$

This means we only need to take the powers of the n entries in the diagonal of D and multiply 3 $n \times n$ matrices, instead of multiplying k many $n \times n$ matrices.

Example 6.29. Consider the diagonalizable matrix

$$A = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}.$$

Now

$$A^2 = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 25 & 0 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 41 & 16 \\ -32 & -7 \end{bmatrix}.$$

We can even use this to give a general formula for A^k:

$$A^k = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ 2 \cdot 3^k - 2 \cdot 5^k & 2 \cdot 3^k - 5^k \end{bmatrix}.$$

Let us now talk about diagonalization and linear transformations. Where does diagonalization come in? The easiest possible linear transformations $T\colon \mathbb{R}^m \to \mathbb{R}^n$ are those that are given by multiplication by a diagonal matrix; if the standard matrix for T is diagonalizable, that gives us a basis $\mathcal{B}$ such that the $\mathcal{B}$-matrix for T is diagonal.

Theorem 6.30. *Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix A. If $A = PDP^{-1}$ with D a diagonal matrix, and $\mathcal{B}$ is the basis for $\mathbb{R}^n$ whose elements are the columns of P, then D is the matrix representing T in the basis $\mathcal{B}$.*

Example 6.31. Consider the linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x) = Ax$, and assume the following is a diagonalization of A:

$$A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}.$$

Then in the basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\},$$

the $\mathcal{B}$-matrix for T is diagonal, and given by

$$A_{\mathcal{B} \leftarrow \mathcal{B}} = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}.$$

6.5 Complex eigenvalues

Some $n \times n$ matrices have less than n real eigenvalues, or even none.

Example 6.32. The characteristic polynomial of

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

is $\lambda^2 + 1$, which has no real roots. However, if we allow ourselves to also consider complex numbers, this polynomial does have 2 roots: i and $-i$. So A does have two *complex eigenvalues*. To find eigenvectors associated to a complex eigenvalue, we need to allow for complex valued vectors also.

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} i \\ 1 \end{bmatrix} = i \begin{bmatrix} 1 \\ -i \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} -i \\ 1 \end{bmatrix} = -i \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

So $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ is a complex eigenvector associated to the complex eigenvalue i, and $\begin{bmatrix} 1 \\ i \end{bmatrix}$ is a complex eigenvector associated to the complex eigenvalue $-i$.

Chapter 7

Orthogonality

7.1 Inner product

Definition 7.1. Given two vectors u and v in $\mathbb{R}^n$, think of these vectors as $n \times 1$ matrices. The **inner product**, sometimes referred to as the **dot product**, of u and v is the real number

$$u \bullet v = u^T v.$$

If

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \text{ and } v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \text{ then } u \bullet v = u_1 v_1 + \cdots + u_n v_n.$$

Example 7.2. We have

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \bullet \begin{bmatrix} 5 \\ -1 \\ 0 \end{bmatrix} = 1 \cdot 5 + 2 \cdot (-1) + 3 \cdot 0 = 5 - 2 = 3.$$

Theorem 7.3. *Let u, v, w be vectors in $\mathbb{R}^n$ and $c \in \mathbb{R}$ be any scalar. Then:*

(a) $u \bullet v = v \bullet u$.
(b) $(u + v) \bullet w = u \bullet w + v \bullet w$.
(c) $(cu) \bullet v = v(u \bullet v) = u \bullet (cv)$.
(d) $u \bullet u \geqslant 0$.
(e) $u \bullet u = 0$ *if and only if* $u = 0$.

As a consequence of these properties, we also obtain that for all scalars $c_1, \dots, c_p$ and vectors $u_1, \dots, u_p$,

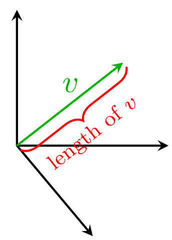
$$(c_1 u_1 + \cdots c_p u_p) \bullet w = c_1 (u_1 \bullet w) + \cdots c_n (u_n \bullet w).$$

Back in Definition 2.7, we defined the norm of a vector to be its length. We can now rewrite that in terms of the inner product.

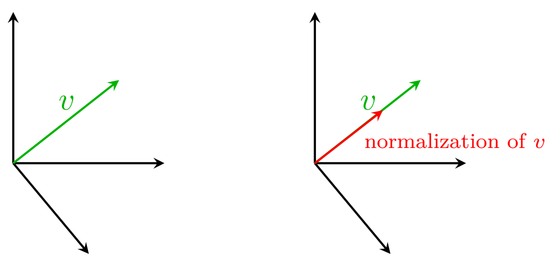
Definition 7.4. Let v be a vector in $\mathbb{R}^n$. The **length** or **norm** of v is the nonnegative real number

$$\|v\| = \sqrt{v \bullet v} = \sqrt{v_1^2 + \cdots + v_n^2}.$$

Remark 7.5. As we noted before, the norm has a geometric meaning: it is the length of the vector, in the colloquial sense; more precisely, it is the length of the line segment between that point and the origin.



Recall that a vector $v \in \mathbb{R}^n$ is a **unit** vector if $\|v\| = 1$. As we saw in Definition 2.11, we can always find a unit vector with the same direction as a given vector by *normalizing* it: the **normalization** of v is the unit vector $\frac{v}{\|v\|}$.



Example 7.6. Let L be the line in $\mathbb{R}^2$ spanned by $\begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix}$. We can replace this vector by a unit vector which spans the same line. First, let us replace it with a nicer vector, to avoid dealing with fractions:

$$v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

also spans the line L. Now we normalize v, and that will give us a unit vector spanning W.

$$\|v\|^2 = 2^2 + 3^2 = 13 \implies \|v\| = \sqrt{13},$$

so

$$u = \frac{v}{\|v\|} = \begin{bmatrix} \frac{2}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \end{bmatrix}$$

is a unit vector that gives us a basis for L. Notice that we could have also chosen the unit vector

$$-\frac{v}{\|v\|} = \begin{bmatrix} -\frac{2}{\sqrt{13}} \\ -\frac{3}{\sqrt{13}} \end{bmatrix}$$

Definition 7.7. Two vectors u and v in $\mathbb{R}^n$ are **orthogonal** if $u \bullet v = 0$.

In $\mathbb{R}^2$, two vectors are orthogonal if and only if they are perpendicular to each other.

Theorem 7.8 (Pythagorean Theorem)**.** *Two vectors u and v in $\mathbb{R}^n$ are orthogonal if and only if*

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 .$$

Definition 7.9. A set of vectors $\{v_1, \dots, v_p\}$ in $\mathbb{R}^n$ is an **orthogonal set** if any two distinct vectors from this set is orthogonal, meaning $v_i \bullet v_j = 0$ for all $i \neq j$.

Two nonzero vectors v and w in $\mathbb{R}^n$ are orthogonal if and only if in the plane they span, the two vectors are perpendicular.

Theorem 7.10. *If $S = \{u_1, \dots, u_p\}$ is a set of orthogonal vectors in $\mathbb{R}^n$, then S is a linearly independent set, and thus a basis for span(S).*

Example 7.11. Consider the set $S = \{u_1, u_2, u_3\}$, where

$$u_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, u_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, u_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}.$$

Note that:

$$\begin{aligned} u_1 \bullet u_2 &= 3 \cdot (-1) + 1 \cdot 2 + 1 \cdot 1 = 0 \\ u_1 \bullet u_3 &= 3 \cdot (-\frac{1}{2}) + 1 \cdot (-2) + 1 \cdot \frac{7}{2} = 0 \\ u_2 \bullet u_3 &= -1 \cdot (-\frac{1}{2}) + 2 \cdot (-2) + 1 \cdot \frac{7}{2} = 0 \end{aligned}$$

So S is an orthogonal set, and thus this set is linearly independent. Since there are 3 vectors, we conclude this is a basis for $\mathbb{R}^3$.

Theorem 7.12. *Suppose that $S = \{u_1, \dots, u_p\}$ is an orthogonal basis for the subspace W of $\mathbb{R}^n$. For each y in W, the unique weights $c_1, \dots, c_p$ such that*

$$y = c_1 u_1 + \dots + c_p u_p$$

are given by

$$c_j = \frac{y \bullet u_j}{u_j \bullet u_j} \text{ for each } j = 1, \dots, p.$$

Example 7.13. We saw before that

$$u_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, u_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, u_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}$$

form an orthogonal basis for $\mathbb{R}^3$. So we can now compute the coordinates of any vector in $\mathbb{R}^3$ with respect to this basis. For example, consider

$$y = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}.$$

Since

$$\begin{aligned} \frac{y \bullet u_1}{u_1 \cdot u_1} &= \frac{18+1-8}{11} = 1 \\ \frac{y \bullet u_2}{u_2 \cdot u_2} &= \frac{-6+2-8}{6} = -2 \\ \frac{y \bullet u_3}{u_3 \cdot u_3} &= \frac{-3-2-28}{\frac{33}{2}} = -2 \end{aligned}$$

we conclude that $y = u_1 - 2u_2 - 2u_3$, or more explicitly,

$$\begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}.$$

Thus if we call our basis above $\mathcal{B}$, the coordinate vector of y with respect to $\mathcal{B}$ is

$$[y]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}.$$

One thing that was really annoying in this calculation was that $u_i \bullet u_i$ might be complicated numbers. The best case scenario is when we pick an orthogonal basis where all the vectors are unit vectors.

Definition 7.14. An **orthonormal set** of vectors in $\mathbb{R}^n$ is an orthogonal set of unit vectors. An **orthonormal basis** for the subspace W of $\mathbb{R}^n$ is an orthonormal set that forms a basis for W.

Example 7.15. The standard basis $\{e_1, \dots, e_n\}$ for $\mathbb{R}^n$ is an orthonormal basis for $\mathbb{R}^n$.

Given an orthogonal basis for a subspace W, we can always turn it into an orthonormal basis by normalizing the vectors.

Example 7.16. We saw before that

$$u_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, u_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, u_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}$$

form an orthogonal basis for $\mathbb{R}^3$. We also computed

$$\begin{aligned} u_1 \cdot u_1 = 11 &\implies \|u_1\| = \sqrt{11} \\ u_2 \cdot u_2 = 6 &\implies \|u_2\| = \sqrt{6} \\ u_3 \cdot u_3 = \frac{33}{2} &\implies \|u_3\| = \sqrt{\frac{33}{2}}, \end{aligned}$$

so

$$v_1 = \begin{bmatrix} \frac{3}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \end{bmatrix}, v_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, v_3 = \begin{bmatrix} -\frac{1}{\sqrt{66}} \\ -\frac{4}{\sqrt{66}} \\ \frac{7}{\sqrt{66}} \end{bmatrix}$$

is an orthonormal basis for $\mathbb{R}^n$.

7.2 Orthogonal projections

Given a vector y and a subspace W of $\mathbb{R}^n$, there exists a unique vector $\hat{y}$ in W with the following properties:

- $y - \hat{y}$ is orthogonal to W, and
- $\hat{y}$ is the unique vector in W that is closest to y.

Theorem 7.17. *Let W be a subspace of $\mathbb{R}^n$. Each y in $\mathbb{R}^n$ can be written uniquely in the form*

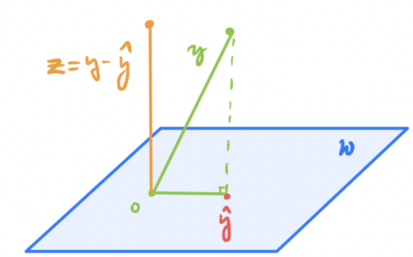
$$y = \hat{y} + z,$$

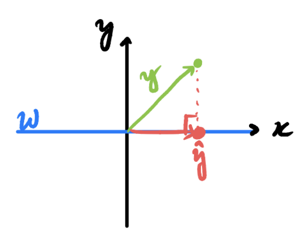
where $\hat{y}$ is in W and z is orthogonal to W. If $\{u_1, \dots, u_p\}$ is an orthogonal basis for W, then this vector $\hat{y}$ is given by

$$\hat{y} = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \cdots + \frac{y \bullet u_p}{u_p \bullet u_p} u_p$$

and $z = y - \hat{y}$.

This vector $\hat{y}$ is called the **orthogonal projection** of y in W. The notation $\hat{y}$ can be confusing, since it doesn't indicate which subspace W we are projecting onto. We often write $\text{proj}_W(y)$ to indicate the **orthogonal projection** of y onto W.





Example 7.18. Consider the vectors

$$u_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Notice that $\{u_1, u_2\}$ is an orthogonal basis for $W = \text{span}(\{u_1, u_2\})$, since

$$u_1 \bullet u_2 = 2 \cdot (-2) + 5 \cdot 1 + (-1) \cdot 1 = -4 + 5 - 1 = 0.$$

The orthogonal projection of y onto W is

$$\text{proj}_W(y) = \hat{y} = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \frac{y \bullet u_2}{u_2 \bullet u_2} u_2 = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix}.$$

Thus the closest point to y in W is

$$\operatorname{proj}_W(y) = \hat{y} = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \frac{y \bullet u_2}{u_2 \bullet u_2} u_2 = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix}.$$

Moreover, the following vector is orthogonal to W:

$$y - \hat{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix} = \begin{bmatrix} \frac{7}{5} \\ 0 \\ \frac{14}{5} \end{bmatrix}.$$

So now y can be decomposed as a sum of a vector in W and a vector that is orthogonal to W:

$$y = \underbrace{\begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix}}_{\in W} + \underbrace{\begin{bmatrix} \frac{7}{5} \\ 0 \\ \frac{14}{5} \end{bmatrix}}_{\in W^{\perp}}$$

Theorem 7.19 (The best approximation theorem)**.** *Let W be a subspace of $\mathbb{R}^n$, y any vector in $\mathbb{R}^n$, and let $\hat{y}$ be the orthogonal projection of y onto W. Then $\hat{y}$ is the closest point to y in W, meaning that*

$$\|y - \hat{y}\| < \|y - w\|$$

for all vectors $w \neq \hat{y}$ in W.

Remark 7.20. Notice that if y is a vector in W, then $\operatorname{proj}_W(y) = y$.

Definition 7.21. The **distance** of y to W is $\left\|y - \operatorname{proj}_W(y)\right\|$.

Example 7.22. Consider the vectors

$$u_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

and the subspace $W = \operatorname{span}(\{u_1, u_2\})$. The closest point to y in W is

$$\operatorname{proj}_W(y) = \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix}.$$

The distance of y to W is

$$\left\|y - \operatorname{proj}_W(y)\right\| = \sqrt{\left(\frac{7}{5}\right)^2 + \left(\frac{14}{5}\right)^2} = \frac{\sqrt{245}}{5}.$$

Theorem 7.23. *If $\{u_1, \dots, u_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then*

$$\operatorname{proj}_W(y) = (y \bullet u_1)u_1 + \cdots + (y \bullet u_p)u_p.$$

If $U = \begin{bmatrix} u_1 & \cdots u_p \end{bmatrix}$, then

$$\operatorname{proj}_W(y) = UU^T y \textit{ for all } y \textit{ in } \mathbb{R}^n.$$

7.3 Graham–Schmidt

Orthonormal bases are the best kind of bases we can find for a subspace W. To find an orthonormal basis, we can first find an orthogonal basis, and then normalize all the vectors.

Example 7.24. The set

$$\left\{ \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} \right\}$$

is an orthogonal basis for $\mathbb{R}^3$. Normalizing, we get an orthogonal basis for $\mathbb{R}^3$:

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \right\}$$

To find an orthogonal basis for a subspace of $\mathbb{R}^n$, we can follow the Graham–Schmidt Process:

Theorem 7.25 (The Gram–Schmidt Process)**.** *Given a basis $\{x_1, \dots, x_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, set*

$$\begin{aligned} v_1 =& x_1 \\ v_2 =& x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1 \\ v_3 =& x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2 \\ &\vdots \\ v_p =& x_p - \frac{x_p \bullet v_1}{v_1 \bullet v_1} v_1 - \cdots - \frac{x_p \bullet v_{p-1}}{v_{p-1} \bullet v_{p-1}} v_{p-1} \end{aligned}$$

Then $\{v_1, \dots, v_p\}$ is an orthogonal basis for W. Moreover,
$\operatorname{span}(\{x_1, \dots, x_i\}) = \operatorname{span}(\{v_1, \dots, v_i\})$ *for each $i = 1, \dots, p$.*

Example 7.26. Consider the subspace $W = \operatorname{span}(\{x_1, x_2, x_3\})$ of $\mathbb{R}^4$ with

$$x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \; x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \text{ and } x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Note that x_1, x_2, x_3 are linearly independent, so they form a basis for W. Let us find an orthonormal basis for W by using the Gram-Schmidt process.

Step 1: Take $v_1 = x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

Step 2: Take

$$v_2 = x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}.$$

Optional step: since we are only searching for an orthogonal basis for now, we can rescale our vectors to make the numbers easier to deal with. For example, we can replace v_2 with

$$v_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

and the set $\{v_1, v_2\}$ is still orthogonal.

Step 2: Take

$$v_3 = x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{12} \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{2}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}.$$

We can rescale this and use instead

$$v_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}.$$

We have found the following orthogonal basis for W:

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

We can now rescale these to unit vectors to obtain an orthonormal basis for W:

$$\left\{ \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{-3}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \end{bmatrix}, \begin{bmatrix} 0 \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \right\}.$$

Note that in the Gram-Schmidt process, we started with a basis for W and turned it into an orthogonal basis by replacing each vector u_{i+1} with a vector that is orthogonal to $\operatorname{span}(\{u_1, \dots, u_i\})$ by subtracting from u_{i+1} its orthogonal projection onto W.

7.4 Orthogonal complements

Definition 7.27. Let W be a subspace of $\mathbb{R}^n$. The **orthogonal complement** of W is the set of vectors that are orthogonal to every vector in W:

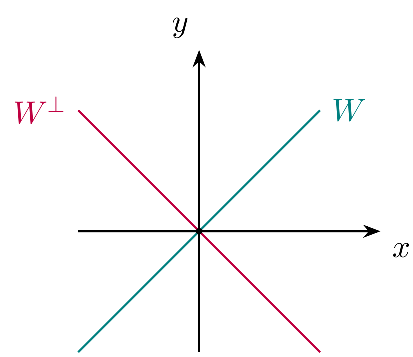
$$W^\perp = \{v \text{ in } \mathbb{R}^n \mid v \bullet w = 0 \text{ for every } w \text{ in } W\}.$$

Example 7.28. In $\mathbb{R}^2$, let W be the line

$$W = \left\{ \begin{bmatrix} a \\ a \end{bmatrix} : a \text{ any real numbers} \right\} = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \right).$$

We will show that its orthogonal complement is the line

$$W^\perp = \left\{ \begin{bmatrix} b \\ -b \end{bmatrix} : b \text{ any real number} \right\} = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\} \right).$$



We will see how soon.

Example 7.29. Consider the xy-plane

$$W = \left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} : a, b \text{ any real numbers} \right\} = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} \right)$$

in $\mathbb{R}^3$. We will soon learn how to show that its orthogonal complement is the line

$$W^\perp = \left\{ \begin{bmatrix} 0 \\ 0 \\ c \end{bmatrix} : c \text{ any real number} \right\} = \operatorname{span}\left(\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \right).$$

Theorem 7.30. *Let W be a subspace of $\mathbb{R}^n$. Then $W^\perp$ is a subspace of $\mathbb{R}^n$.*

To check if a vector is in $W^\perp$, it is sufficient to look at a basis for W.

Theorem 7.31. *Let W be a subspace of $\mathbb{R}^n$. A vector v is in $W^\perp$ if and only v is orthogonal to every vector in a spanning set for W. In particular, if $\mathcal{B}$ is a basis for W, then v is in $W^\perp$ if and only $v \bullet w = 0$ for every w in $\mathcal{B}$.*

Example 7.32. Consider the subspace of $\mathbb{R}^3$ given by

$$W = \left\{ \begin{bmatrix} x+y \\ y \\ x-y \end{bmatrix} \mid x, y \in \mathbb{R} \right\}.$$

Is the vector

$$v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

in $W^\perp$? To answer this question, we need a basis for W; consider the basis for W given by

$$\left\{ u_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, u_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\}.$$

We have

$$u_1 \bullet v = 1 + 0 + 3 = 4 \neq 0 \quad \text{and} \quad u_2 \bullet v = 1 + 2 - 3 = 0.$$

Since $u_1 \bullet v \neq 0$, then v is not in $W^\perp$.
In contrast, the vector

$$y = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}$$

is in $W^\perp$, since

$$u_1 \bullet y = 1 - 1 = 0 \quad \text{and} \quad u_2 \bullet y = 1 - 2 + 1 = 0.$$

To find the orthogonal complement of a subspace W, we can use the following theorem:

Theorem 7.33. *Let A be an $m \times n$ matrix. Then*

$$(\mathrm{col}(A))^\perp = \mathrm{Nul}(A^T).$$

Remark 7.34. We can easily see why. Suppose that

$$W = \mathrm{span}\left(\{u_1, \dots, u_m\}\right)$$

or equivalently that

$$W = \mathrm{col}(A), \text{ where } A = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}.$$

Note that v is in $W^\perp$ exactly when

$$\begin{aligned} u_1 \bullet v &= 0 \\ &\vdots \\ u_m \bullet v &= 0 \end{aligned}$$

but this is equivalent to asking that

$$\begin{bmatrix} {u_1}^\mathsf{T} \\ \vdots \\ {u_m}^\mathsf{T} \end{bmatrix} v = 0,$$

or equivalently, v is in $\mathrm{Nul}(A^\mathsf{T})$.

Important

Given a subspace W, to find a basis for $W^\perp$, we can:

- Step 1: Take a spanning set $\{v_1, \dots, v_n\}$ for W, and consider the matrix with these vectors as columns:

$$A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}.$$

Note that $W = \mathrm{col}(A)$.

- Step 2: Find a basis for $\mathrm{Nul}(A^\mathsf{T})$. This is a basis for $W^\perp$.

Alternatively, we can simply do the following:

Important

Given a subspace W, to find a basis for $W^\perp$, we can:

- Step 1: Take a spanning set $\{v_1, \dots, v_n\}$ for W, and consider the matrix with these vectors as rows:

$$B = \begin{bmatrix} {v_1}^\mathsf{T} \\ \vdots \\ {v_n}^\mathsf{T} \end{bmatrix}.$$

- Step 2: Find a basis for $\mathrm{Nul}(B)$. This is a basis for $W^\perp$.

Example 7.35. In $\mathbb{R}^2$, let W be the line

$$W = \left\{ \begin{bmatrix} a \\ a \end{bmatrix} : a \text{ any real numbers} \right\} = \mathrm{span}\left(\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \right).$$

To find the orthogonal complement $W^\perp$, we need to find the null space of

$$B = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

This matrix is already in RREF, and the solutions to $Bx = 0$ are of the form

$$\begin{cases} x_1 = -x_2 \\ x_2 \text{ free variable} \end{cases} \implies x = \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Thus we get a basis for $W^\perp$; in fact, we can even rescale the vector to get the basis

$$\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$

is a basis for the orthogonal complement of W, which is the line

$$W^\perp = \left\{ \begin{bmatrix} b \\ -b \end{bmatrix} : b \text{ any real number} \right\} = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\} \right).$$

Example 7.36. Consider the subspace of $\mathbb{R}^3$ given by

$$W = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} \right).$$

To find a basis for its orthogonal complement, consider the matrix

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \text{so} \qquad A^\mathsf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Since A^T is already in RREF, a basis for its null space is given by

$$\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Notice these are exactly the two orthogonal spaces W and $W^\perp$ from Example 7.29.

Example 7.37. Consider

$$W = \operatorname{span}\left(\left\{ \begin{bmatrix} -3 \\ 6 \\ -1 \\ 1 \\ -7 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -4 \\ 5 \\ 8 \\ -4 \end{bmatrix} \right\} \right).$$

Its orthogonal complement is the null space of

$$B = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The general solution to the homogeneous system $Bx = 0$ is

$$x = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \text{ with } x_2, x_4, x_5 \text{ any scalars,}$$

so

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

is a basis for $\mathrm{Nul}(B) = W^{\perp}$, and $\dim(W^{\perp}) = 3$.
Note that W is the column space of

$$\begin{bmatrix} -3 & 1 & 2 \\ 6 & -2 & -4 \\ -1 & 2 & 5 \\ 1 & 3 & 8 \\ -7 & -1 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & \frac{1}{5} \\ 0 & 1 & \frac{13}{5} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus $\dim(W) = 2$. Note moreover that W is a subspace of $\mathbb{R}^5$, and that

$$\dim(W) + \dim(W^{\perp}) = 2 + 3 = 5.$$

This was not a coincidence! In fact, we have the following theorem:

Theorem 7.38. *Let W be any subspace of $\mathbb{R}^n$. Then*

$$\dim(W) + \dim(W^{\perp}) = n.$$

Remark 7.39. We can justify this using the Rank–Nullity Theorem. Indeed, taking any spanning set for W, say $v_1, \dots, v_m$. Then W is the span of the rows of the matrix

$$B = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}$$

whose rows are the vectors $v_1, \dots, v_m$. Note that this matrix has m rows and n columns. By the Rank–Nullity Theorem,

$$\dim(\mathrm{Nul}(B)) = n - \mathrm{rank}(B).$$

The rank of B is the number of pivots of B, which coincides with the number of pivots of

$$B^{\mathsf{T}} = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix},$$

and that is the dimension of W. Thus

$$\dim(W^{\perp}) = n - \dim(W).$$

We can now check that this indeed holds for the previous examples as well. For example,

Example 7.40. The subspace

$$W = \operatorname{span}\left(\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}\right)$$

of $\mathbb{R}^3$ has dimension 2, and its orthogonal complement

$$W^{\perp} = \operatorname{span}\left(\left\{\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}\right)$$

has dimension 1, so indeed

$$\dim(W) + \dim(W^{\perp}) = 2 + 1 = 3.$$

7.5 Symmetric and orthogonal matrices

Recall that a matrix A is **symmetric** if $A^T = A$. Notice that for a matrix to be symmetric, it must be a square matrix; its main diagonal entries can be anything, but the entry on the ith row and jth column must match the entry in the ith column and jth row.

Theorem 7.41. *If A is a symmetric matrix, then any two eigenvectors associated to different eigenvalues are orthogonal.*

Thus if a symmetric $n \times n$ matrix A is diagonalizable, we can choose an orthogonal basis of eigenvectors of A for $\mathbb{R}^n$. Therefore, we can write A as $A = PDP^{-1}$ where D is a diagonal matrix and the columns of P form an orthonormal basis for $\mathbb{R}^n$.

Definition 7.42. A square matrix U is **orthogonal** (sometimes also called **orthonormal**) if the columns of U form an orthonormal basis for $\mathbb{R}^n$. Equivalently, U is an orthogonal matrix if its inverse is its transpose, meaning that

$$U^{\mathsf{T}}U = I = UU^{\mathsf{T}}.$$

Definition 7.43. An $n \times n$ matrix A is **orthogonally diagonalizable** if there exists an orthogonal matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

Theorem 7.44. *An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.*

In particular, every symmetric matrix is diagonalizable.

Theorem 7.45 (Spectral Theorem for Symmetric matrices)**.** *Every $n \times n$ symmetric matrix has the following properties:*

(a) A has n real eigenvalues, counting multiplicities.

(b) The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation of A.
(c) Eigenvectors corresponding to different eigenvalues of A are orthogonal.
(d) A is orthogonally diagonalizable.

Given a symmetric matrix, we now know that we can always orthogonally diagonalize A, meaning we can find a diagonalization for A of the form

$$A = PDP^{-1} = PDP^{\intercal}$$

where

D is diagonal and P is orthogonal.

Here is how we do this in an example:

Example 7.46. Let us orthogonally diagonalize the symmetric matrix

$$A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}.$$

Its characteristic equation is

$$-\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda - 8)(\lambda - 6)(\lambda - 3) = 0.$$

Let us find a basis for each eigenspace:

$$A - 8I = \begin{bmatrix} -2 & -2 & -1 \\ -2 & -2 & -1 \\ -1 & -1 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{Basis for the eigenspace of } A \text{ associated to } 8: \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

$$A - 6I = \begin{bmatrix} 0 & -2 & -1 \\ -2 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \quad \text{Basis for the eigenspace of } A \text{ associated to } 6: \left\{ \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix} \right\}.$$

$$A - 3I = \begin{bmatrix} 3 & -2 & -1 \\ -2 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{Basis for the eigenspace of } A \text{ associated to } 3: \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

The 3 vectors we got form an orthogonal basis for $\mathbb{R}^3$. We can normalize to obtain an orthonormal basis for $\mathbb{R}^3$, and use it to write an orthonormal matrix P:

$$P = \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

We can also take the diagonal matrix

$$D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

and now we have a diagonalization for $A = PDP^{-1} = PDP^{\mathsf{T}}$.

If one of the eigenspaces has dimension 2 or more, we need to carefully choose an orthogonal basis for each such eigenspace. In the example above, each eigenspace has dimension one, so the vectors we got for each eigenspace are all automatically orthogonal. Below is an example where one of the eigenspaces has dimension 2:

Example 7.47. Let us orthogonally diagonalize

$$A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}.$$

Eigenvalues of A: First, we need to solve the characteristic equation of A:

$$\begin{aligned} \det(\lambda I - A) &= 0 \\ \det \begin{bmatrix} \lambda - 3 & 2 & -4 \\ 2 & \lambda - 6 & -2 \\ -4 & -2 & \lambda - 3 \end{bmatrix} &= 0 \\ \lambda^3 - 12\lambda^2 + 21\lambda + 98 &= 0 \\ (\lambda - 7)^2(\lambda + 2) &= 0 \end{aligned}$$

Thus the eigenvalues are

$$\lambda = -2 \quad \text{with multiplicity 1} \qquad \text{and} \qquad \lambda = 7 \quad \text{with multiplicity 2.}$$

Eigenvectors associated to 7: To find a basis for the eigenspace associated to -2, we need to find a basis for the null space of $A - 7I$. To do that, note that

$$A - 7I = \begin{bmatrix} -4 & -2 & 4 \\ -2 & -1 & 2 \\ 4 & 2 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{1}{2} & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus the solutions to $(A - 7I)x = 0$ are of the form

$$\begin{cases} x_1 = -\frac{1}{2}x_2 + x_3 \\ x_2 \text{ and } x_3 \text{ free variables.} \end{cases} \implies x = \begin{bmatrix} -\frac{1}{2}x_2 + x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

This gives us the following basis for the eigenspace associated to 7:

$$\left\{ \begin{bmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Note that we can always rescale our vectors without changing the basis, so let us take instead the two basis vectors

$$y_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \text{and} \qquad y_2 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}.$$

However, a basis for the eigenspace is not good enough to orthogonally diagonalize A: we need an orthogonal basis. Our two vectors are not orthogonal, since

$$y_1 \cdot y_2 = 1 \cdot 1 + 0 + 0 = 1.$$

Gram–Schmidt to find an orthogonal basis for the eigenspace associated to 7: Following the Graham–Schmidt process, we keep $v_1 = y_1$ and replace y_2 with

$$y_2 - \frac{y_2 \cdot y_1}{y_1 \cdot y_1} y_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -2 \\ -\frac{1}{2} \end{bmatrix}.$$

We can rescale to clear fractions, getting

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \text{and} \qquad v_2 = \begin{bmatrix} 1 \\ -4 \\ -1 \end{bmatrix}.$$

Now $v_1 \cdot v_2 = 0$, so $\{v_1, v_2\}$ is an orthogonal basis for the eigenspace associated to 7.

Eigenvectors associated to -2: To find a basis for the eigenspace associated to -2, we need to find a basis for the null space of $A + 2I$. To do that, note that

$$A + 2I = \begin{bmatrix} 5 & -2 & 4 \\ -2 & 8 & 2 \\ 4 & 2 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus the solutions to $(A + 2I)x = 0$ are of the form

$$\begin{cases} x_1 = -x_3 \\ x_2 = -\frac{1}{2} x_3 \\ x_3 \text{ is a free variable.} \end{cases} \qquad \Longrightarrow x = \begin{bmatrix} -x_3 \\ -\frac{1}{2} x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ -\frac{1}{2} \\ 1 \end{bmatrix}.$$

This gives us the basis

$$\left\{ \begin{bmatrix} -1 \\ -\frac{1}{2} \\ 1 \end{bmatrix} \right\}.$$

Again we rescale to get nicer numbers, and obtain

$$v_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}.$$

Orthogonal basis of eigenvectors: We now have a orthogonal basis for $\mathbb{R}^3$ consisting of eigenvectors of A: $\{v_1, v_2, v_3\}$. However, we need an orthonormal basis.

Normalize the eigenvectors: since $\|v_1\| = \sqrt{2}$, $\|v_2\| = \sqrt{18}$, and $\|v_3\| = 3$, then we can take the orthonormal basis

$$\left\{ \frac{\sqrt{2}}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{18}} \begin{bmatrix} 1 \\ -4 \\ -1 \end{bmatrix}, \frac{1}{3} \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{18}} \\ -\frac{4}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \end{bmatrix}, \begin{bmatrix} -\frac{2}{3} \\ -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix} \right\}.$$

Orthogonal diagonalization: Finally, we have

$$A = PDP^{-1} = PDP^{\mathsf{T}}$$

where

$$P = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & -\frac{4}{\sqrt{18}} & -\frac{1}{3} \\ \frac{\sqrt{2}}{2} & -\frac{1}{\sqrt{18}} & \frac{2}{3} \end{bmatrix} \qquad \text{and} \qquad D = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

Chapter 8

Applications of Linear Algebra

8.1 Least squares approximations

Given an inconsistent system of equations $Ax = b$, the best we can do is find x such that Ax is as close as possible to b, a *best approximation* to b. This idea is very helpful when dealing with real data. Given a set of data points, we can try to find a line that passes as close as possible to the points, to try to give an approximate model for the data. However, this is not always possible – not all data fits a linear model.

Definition 8.1. Let A be an $m \times n$ matrix and b be a vector in $\mathbb{R}^m$. A **least squares solution** to $Ax = b$ is a vector y in $\mathbb{R}^n$ such that

$$\|Ay - b\| \leqslant \|Ax - b\|$$

for every vector x in $\mathbb{R}^n$.

Remark 8.2. If $Ax = b$ is a consistent system, there is some vector x such that $Ax - b = 0$, so $\|Ax - b\| = 0$. If the system is inconsistent, then $\|Ax - b\|$ is never zero. A best possible approximation y has $\|Ay - b\|$ as small as possible.

Remark 8.3. If $Ax = b$ is consistent, any solution x *is* a least squares solution to $Ax = b$.

The least squares solution to $Ax = b$ is the best approximation of b to the set of vectors of the form Ax – meaning, to $\operatorname{col}(A)$. We saw before that the best approximation of x to W is $\operatorname{proj}_W(x)$, so a best approximation of b to $\operatorname{col}(A)$ should be $\widehat{b} = \operatorname{proj}_{\operatorname{col}(A)}(b)$. So for $\widehat{x}$ to be a least squares solution to $Ax = b$, we must have $A\widehat{x} = \widehat{b}$. Notice also that $b - \operatorname{proj}_{\operatorname{col}(A)}(b)$ is orthogonal to $\operatorname{col}(A)$, so if $\widehat{x}$ is a least squares solution to $Ax = b$, then $b - \widehat{b} = b - A\widehat{x}$ is orthogonal to A. Therefore,

$$A^T(b - A\widehat{x}) = 0 \iff A^\intercal A\widehat{x} = A^T b.$$

Theorem 8.4. *The least square solutions to $Ax = b$ coincide with the solutions to $A^\intercal Ax = A^T b$.*

We call the system $A^\mathsf{T} Ax = A^T b$ the **normal equations** of $Ax = b$.

Example 8.5. Consider the system $Ax = b$ with

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \text{ and } b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

Since

$$[A|b] = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 1 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

the system is inconsistent. Since

$$A^\mathsf{T} A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \quad \text{and} \quad A^T b = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix},$$

the normal equations of our system are

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

We can solve this system by row reduction, but we can actually solve this faster by using the fact that

$$(A^T A)^{-1} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}^{-1} = \frac{1}{85-1} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

so the least squares solution to $Ax = b$ is

$$y = (A^T A)^{-1} b = \frac{1}{85-1} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Remark 8.6. Note that for any matrix A, AA^T is a symmetric matrix. Indeed,

$$(AA^\mathsf{T})^\mathsf{T} = (A^\mathsf{T})^\mathsf{T} A^\mathsf{T} = AA^\mathsf{T}.$$

Theorem 8.7. *If the columns of A are linearly independent, then $A^\mathsf{T} A$ is invertible.*

Theorem 8.8. *Let A be an $m \times n$ matrix, and let b be a column vector in $\mathbb{R}^m$. Consider the equation*

$$Ax = b.$$

(a) Any solution z to the normal equations

$$(A^\mathsf{T} A)z = A^\mathsf{T} b$$

is a best approximation to a solution to $Ax = b$, in the sense that $\|b - Az\|$ is smallest possible.

(b) If the columns of A are linearly independent, then $A^\mathsf{T}A$ is invertible, and there is a unique least squares approximation z to the solutions of $Ax = b$, given by

$$z = (A^\mathsf{T}A)^{-1}A^\mathsf{T}b.$$

But of course $A^\mathsf{T}A$ is not always invertible, in which case there are multiple least squares solutions to $Ax = b$.

Example 8.9. Consider the system $Ax = b$ with

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \text{ and } b = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}.$$

Once more, this system is inconsistent:

$$[A|b] = \begin{bmatrix} 1 & 1 & 0 & 0 & -3 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 2 \\ 1 & 0 & 0 & 1 & 5 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

But this time, $A^\mathsf{T}A$ is not invertible:

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix},$$

the augmented matrix for the normal system of equations $A^\mathsf{T}Ax = A^Tb$ is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

so the general least squares solution to $Ax = b$ is

$$y = x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix}.$$

8.2 Singular value decomposition

While not every square matrix can be diagonalized, we will now see that every matrix A (even if A is rectangular) can be decomposed as a product in a nice way, called its singular value decomposition.

Let A be any $m \times n$ matrix. First, note that the matrix $A^\mathsf{T} A$ is symmetric:

$$(A^\mathsf{T} A)^\mathsf{T} = A^\mathsf{T} (A^\mathsf{T})^\mathsf{T} = A^\mathsf{T} A.$$

Since $A^\mathsf{T} A$ is symmetric, it is also orthogonally diagonalizable, so there is a basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^\mathsf{T} A$. Let $\{v_1, \dots, v_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^\mathsf{T} A$, and let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ be the corresponding eigenvalues. Notice we are allowing for repetition of eigenvalues here. Then for each i,

$$\|Av_i\|^2 = (Av_i)^T (Av_i) = v_i^T ((A^\mathsf{T} A) v_i) = v_i^T (\lambda_i v_i) = \lambda_i (v_i^T v_i) = \lambda_i (v_i \bullet v_i) = \lambda_i.$$

This shows that all the eigenvalues of $A^\mathsf{T} A$ are nonnegative. Moreover,

$$\sqrt{\lambda_i} = \|Av_i\|.$$

Definition 8.10. Let A be any $m \times n$ matrix, and let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ be the eigenvalues of $A^\mathsf{T} A$. The **singular values** of A are the $\sigma_1 = \sqrt{\lambda_1}, \dots, \sigma_n = \sqrt{\lambda_n}$.

Given an orthonormal basis $\{v_1, \dots, v_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors of $A^\mathsf{T} A$, the calculation above shows that the singular values of A are the lengths of the vectors $Av_1, \dots, Av_n$.

Example 8.11. Consider the matrix

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}.$$

Then

$$A^\mathsf{T} A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}.$$

One can check that the characteristic polynomial of $A^\mathsf{T} A$ is

$$\det(A^\mathsf{T} A - \lambda I) = -\lambda^3 + 450\lambda^2 - 32400\lambda = -\lambda(\lambda - 90)(\lambda - 360).$$

Thus the eigenvalues of $A^\mathsf{T} A$ are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$, and the singular values of A are $\sigma_1 = \sqrt{360} = 6\sqrt{10}$, $\sigma_2 = \sqrt{90} = 3\sqrt{10}$, and $\sigma_3 = 0$.

One can also show that that the largest possible value of $\|Ax\|$ with x a unit vector is precisely the largest singular value of A, and to obtain this value of $\|Ax\|$ we can use any unit eigenvector of $A^\mathsf{T} A$ associated to the largest eigenvalue.

Theorem 8.12. *Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^\mathsf{T}A$, corresponding (in order) to the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. Suppose $A^\mathsf{T}A$ has r many nonzero eigenvalues, meaning that*

$$\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0 \qquad \textit{and} \qquad \lambda_{r+1} = \cdots = \lambda_n = 0.$$

Then $\{Av_1, \ldots, Av_r\}$ is an orthogonal basis for $\mathrm{col}(A)$, *and* $\mathrm{rank}(A) = r$.

Example 8.13. Consider the matrix

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}.$$

We saw in Example 8.11 that the eigenvalues of $A^\mathsf{T}A$ are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$. Moreover, one can find basis for each of the eigenspaces of $A^\mathsf{T}A$. Since each eigenspace has dimension one, we automatically get an orthogonal basis for $\mathbb{R}^3$ by taking these three vectors. By normalize, we get the following unit eigenvectors:

$$v_1 = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}, v_2 = \begin{bmatrix} -\frac{2}{3} \\ -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix}, \text{ and } v_3 = \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}.$$

Here v_1 is associated to $\lambda_1 = 360$, v_2 is associated to $\lambda_2 = 90$, and v_3 is associated to $\lambda_3 = 0$. Therefore, A has rank 2, and $\{Av_1, Av_2\}$ form an orthogonal basis for $\mathrm{col}(A)$.

Theorem 8.14 (Singular Value Decomposition)**.** *Let A be an $m \times n$ matrix of rank r. Then we can decompose A as a product $A = U\Sigma V^T$, where:*

- *Σ is an $m \times n$ matrix that is diagonal in blocks, of the form*

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

with D a diagonal $r \times r$ matrix whose main diagonal entries are the singular values of A.

- *U is an orthogonal $m \times m$ matrix, and V is an orthogonal $n \times n$ matrix.*

Definition 8.15. Any factorization $A = U\Sigma V^T$ as above is called a **singular value decomposition** of A. The columns of U are called **left singular vectors** of A, and the columns of V are called **right singular vectors** of A.

A singular value decomposition is not unique, but the matrix Σ is completely determined by A, and it is unique if we insist that the singular values of A appear in decreasing order.

Here is a summary of what to do, followed by a detailed recipe:

- Σ is the $m \times n$ matrix

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},$$

where D is the diagonal matrix with $\sigma_1 \geqslant \cdots \geqslant \sigma_r$ in the main diagonal:

$$D = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}$$

- We find an orthonormal basis $\{v_1, \dots, v_n\}$ for $\mathbb{R}^n$ of eigenvectors of $A^\mathsf{T} A$ we found:
$$V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}.$$
- We extend $\{u_1 = \frac{Av_1}{\sigma_1}, \dots, u_r = \frac{Av_r}{\sigma_r}\}$ to an orthonormal basis $\{u_1, \dots, u_m\}$ for $\mathbb{R}^m$, and take
$$U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}.$$

Steps to construct a singular value decomposition:

(a) Find the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ of $A^\mathsf{T} A$. Set r to be the number of eigenvalues that is nonzero, so $\lambda_1, \ldots, \lambda_r > 0$ and $\lambda_{r+1} = \cdots = \lambda_n = 0$.

(b) Find the singular values $\sigma_1 = \sqrt{\lambda_1} \geqslant \cdots \geqslant \sigma_n = \sqrt{\lambda_n}$ of A.

(c) Σ is the $m \times n$ matrix of the form

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \text{ where } D = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}$$

is the diagonal matrix with $\sigma_1 \geqslant \cdots \geqslant \sigma_r$ in the main diagonal. Notice we complete Σ with zeroes until it's exactly the same shape as A.

(d) Find an orthonormal basis $\{v_1, \ldots, v_n\}$ for $\mathbb{R}^n$ where v_i is an eigenvector of $A^\mathsf{T} A$ corresponding to λ_i. There are a few things to pay attention to. After we find a basis of eigenvectors of $A^\mathsf{T} A$, we might need to use Gram–Schmidt on some of the eigenspaces: eigenvectors coming from different eigenvalues are automatically orthogonal to each other, but vectors coming from the same eigenvalue do not have to be orthogonal to each other. Whenever one of the eigenspaces has dimension 2 or more, we need to first check if the basis we obtained is orthogonal, unless we miraculously got an orthogonal basis by accident, we need to apply Gram-Schmidt to our chosen basis for that eigenspace.
The columns of V are the orthonormal basis for $\mathbb{R}^n$ of eigenvectors of $A^\mathsf{T} A$ we found:

$$V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}.$$

(e) We know that $\{Av_1, \ldots, Av_r\}$ is an orthogonal basis for $\operatorname{col}(A)$; normalize each Av_i to obtain an orthonormal basis $u_1, \ldots, u_r$ for $\operatorname{col}(A)$. More precisely, take

$$u_i = \frac{Av_i}{\|Av_i\|} = \frac{Av_i}{\sigma_i}.$$

(f) Extend $\{u_1, \ldots, u_r\}$ to an orthonormal basis $\{u_1, \ldots, u_m\}$ for $\mathbb{R}^m$. To do that, we need to find an orthonormal basis for $(\operatorname{span}(\{u_1, \ldots, u_r\}))^\perp$, which we can do as follows:

- Find a basis for the null space of

$$\begin{bmatrix} {u_1}^\mathsf{T} \\ \vdots \\ {u_r}^\mathsf{T} \end{bmatrix}.$$

- That basis you found is probably not orthonormal yet. If not orthogonal yet, use Gram-Schmidt. Then normalize. Those new vectors are $u_{r+1}, \ldots, u_m$.
Take

$$U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}.$$

Example 8.16. Consider the matrix

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}.$$

We saw before that the eigenvalues of $A^{\mathsf{T}}A$ are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$, corresponding to the following unit eigenvectors:

$$v_1 = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}, v_2 = \begin{bmatrix} -\frac{2}{3} \\ -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix}, \text{ and } v_3 = \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}.$$

Thus the singular values of A are $\sigma_1 = 6\sqrt{10}$, $\sigma_2 = 3\sqrt{10}$, and $\sigma_3 = 0$, and

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} D & 0 \end{bmatrix} = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}.$$

Now the columns of V are an orthonormal basis of eigenvectors of $A^{\mathsf{T}}A$, in the same order as the corresponding singular values show up in Σ:

$$V = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

Finally, we consider the orthogonal basis for $\text{col}(A)$ given by

$$Av_1 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix}, \quad Av_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} -\frac{2}{3} \\ -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \end{bmatrix}.$$

Normalize, to obtain

$$u_1 = \begin{bmatrix} \frac{18}{6\sqrt{10}} \\ \frac{6}{6\sqrt{10}} \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix} \qquad u_2 = \begin{bmatrix} \frac{3}{3\sqrt{10}} \\ -\frac{9}{3\sqrt{10}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} \end{bmatrix}.$$

Now in this case we do not need to extend this to an orthogonal basis for $\mathbb{R}^2$, since we already have 2 vectors (notice that $r = m$), so we are done. Our singular value decomposition is $A = U\Sigma V^T$, where

$$\Sigma = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}, \text{ and } U = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{bmatrix}.$$

It is sufficient to simply give the three matrices above as our final answer, since they give us all the information we need to determine the singular value decomposition of A, but for emphasis, we have just shown that

$$\begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

Example 8.17. Let us find a singular value decomposition for

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

Then

$$A^{\mathsf{T}}A = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 2 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$$

has eigenvalues $\lambda_1 = 18$ and $\lambda_2 = 0$. So the singular values of A are $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ and $\sigma_2 = 0$ and $r = 1$. So

$$D = \begin{bmatrix} 3\sqrt{2} \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

$$A^{\mathsf{T}}A - 18I = \begin{bmatrix} -9 & -9 \\ -9 & -9 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \qquad \text{Basis for the eigenspace associated to 18 : } \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}.$$

$$A^{\mathsf{T}}A - 0I = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \qquad \text{Basis for the eigenspace associated to 18 : } \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Notice that $\{v_1, v_2\}$ must be an orthogonal basis for $\mathbb{R}^2$ because we got two vectors associated to distinct eigenvalues (of our symmetric matrix $A^{\mathsf{T}}A$). We now normalize these vectors to obtain

$$v_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \text{and} \qquad v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Therefore,

$$V = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Now we compute

$$Av_1 = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} \\ -\frac{4}{\sqrt{2}} \end{bmatrix}.$$

Normalizing, we get

$$u_1 = \frac{1}{\sqrt{18}} \begin{bmatrix} -\frac{2}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} \\ -\frac{4}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ \frac{2}{3} \\ -\frac{2}{3} \end{bmatrix}.$$

Now we need to complete this to an orthonormal basis for $\mathbb{R}^3$. We can do this by finding a basis for $(\operatorname{span}(\{u_1\}))^{\perp}$, and then applying Gram–Schmidt to turn it into an orthogonal basis for $(\operatorname{span}(\{u_1\}))^{\perp}$. So we need a basis for the null space of

$$\begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 2 \end{bmatrix},$$

which is given by

$$w_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \text{ and } w_3 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}.$$

Applying Gram-Schmidt, we get:

$$\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \qquad \text{Normalizing} : u_2 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} - \frac{4}{5}\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} \\ -\frac{4}{5} \\ -1 \end{bmatrix} \qquad \text{Normalizing} : u_3 = \begin{bmatrix} \frac{2}{\sqrt{45}} \\ -\frac{4}{\sqrt{45}} \\ -\frac{5}{\sqrt{45}} \end{bmatrix}.$$

So now we take

$$U = \begin{bmatrix} -\frac{1}{3} & \frac{2}{5} & \frac{2}{\sqrt{45}} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & -\frac{4}{\sqrt{45}} \\ -\frac{2}{3} & 0 & -\frac{5}{\sqrt{45}} \end{bmatrix}.$$

So our singular value decomposition for A is $A = U\Sigma V^T$, with

$$U = \begin{bmatrix} -\frac{1}{3} & \frac{2}{5} & \frac{2}{\sqrt{45}} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & -\frac{4}{\sqrt{45}} \\ -\frac{2}{3} & 0 & -\frac{5}{\sqrt{45}} \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \text{and} \qquad V = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Example 8.18. Here is one more example: let us find a singular value decomposition for

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

First, we need to compute $A^{\mathsf{T}}A$, its eigenvalues, and find a basis of eigenvectors of A for $\mathbb{R}^2$.

$$A^{\mathsf{T}}A = \begin{bmatrix} 3 & 0 & 1 \\ -3 & 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & -3 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 10 & -8 \\ -8 & 10 \end{bmatrix}.$$

The characteristic equation for $A^{\mathsf{T}}A$ is

$$0 = (10-\lambda)^2 - 64 = \lambda^2 - 20\lambda + 100 - 64 = \lambda^2 - 20\lambda + 36.$$

so its roots are

$$\lambda = \frac{20 \pm \sqrt{20^2 - 4\cdot 36}}{2} = \frac{20 \pm \sqrt{400-144}}{2} = \frac{20 \pm \sqrt{256}}{2} = \frac{20 \pm 16}{2}$$

and the eigenvalues of $A^\mathsf{T}A$ are

$$\lambda = \frac{20+16}{2} = \frac{36}{2} = 18 \text{ or } \lambda = \frac{20 \pm 16}{2} = \frac{4}{2} = 2.$$

So the singular values of A are $\sigma_1 = \sqrt{18}$ and $\sigma_2 = \sqrt{2}$. Thus

$$\Sigma = \begin{bmatrix} \sqrt{18} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}.$$

Next step: eigenvectors.

$$A^\mathsf{T}A - 18I = \begin{bmatrix} -8 & -8 \\ -8 & -8 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{Basis for this eigenspace}: \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}.$$

$$A^\mathsf{T}A - 2I = \begin{bmatrix} 8 & -8 \\ -8 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \quad \text{Basis for this eigenspace}: \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

These two vectors are already orthogonal to each other, because they come from different eigenvalues. So we can just normalize them and get

$$v_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Therefore,

$$V = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Now

$$Av_1 = \begin{bmatrix} 3 & -3 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{6}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} \qquad Av_2 = \begin{bmatrix} 3 & -3 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{2}{\sqrt{2}} \end{bmatrix}.$$

Normalizing these, we get

$$u_1 = \frac{Av_1}{\|Av_1\|} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad u_2 = \frac{Av_2}{\|Av_2\|} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.$$

Now we need to find a basis for the null space of

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The vector

$$u_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

forms a basis for the null space, and it is a unit vector. Finally,

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

In real world applications, the linear systems $Ax = b$ one considers are often very large (large numbers of variables and equations), and the entries of A and B are often complicated numbers. Any computations involving the system $Ax = b$ will lead to lots of rounding errors. In practice, substituting A by its singular value decomposition will lead to better roundoff errors.

Remark 8.19. What is A itself is symmetric already? In that case,

$$A^{\mathsf{T}} A = A^2$$

and its eigenvalues are precisely the squares of the eigenvalues of A, associated to the same eigenvectors. Indeed, if λ is an eigenvalue of A associated to the eigenvector v, then

$$A^2 v = A(Av) = A(\lambda v) = A(\lambda v) = \lambda(Av) = \lambda(\lambda v) = \lambda^2 v.$$

Thus V is an orthonormal matrix whose columns form a basis for $\mathbb{R}^n$ of eigenvectors of A itself (which are also eigenvectors for $A^{\mathsf{T}} A = A^2$). Moreover, if $r = \operatorname{rank}(A)$, then for each $1 \leqslant i \leqslant r$ we have

$$Av_i = \lambda_i v_i,$$

and its normalization is v_i. Thus $u_1 = v_1, \dots, u_r = v_r$, and to find U, we need only to complete $\{v_1, \dots, v_r\}$ to an orthonormal basis for $\mathbb{R}^n$. (Note that $m = n$ since A must be a square matrix) Therefore, $U = V$, and the singular value decomposition we get is

$$A = U \Sigma V^{\mathsf{T}} = U \Sigma U^{\mathsf{T}} = U \Sigma U^{-1}$$

simply an orthogonal diagonalization for A.

8.3 The return of the Invertible Matrix Theorem

We can now add even more to our growing list of equivalent conditions to being invertible.

Theorem 8.20 (Invertible Matrix Theorem)**.** *The following are equivalent for an $n \times n$ matrix A:*

(a) A is invertible.
(b) A has a pivot in every row.
(c) A has a pivot in every column.
(d) $\operatorname{col}(A) = \mathbb{R}^n$.
(e) $\dim(\operatorname{col}(A)) = n$.

(f) $\dim(\text{row}(A)) = n$.
(g) $\text{rank}(A) = n$.
(h) $\text{Nul}(A) = \{\mathbf{0}\}$.
(i) $\dim(\text{Nul}(A)) = 0$.
(j) $(\text{col}(A))^{\perp} = \{0\}$.
(k) $(\text{Nul}(A))^{\perp} = \mathbb{R}^n$.
(l) *A has n nonzero singular values.*

We are done, we swear. This is the very last time we write the Invertible Matrix Theorem.

8.4 Low Rank approximations

Given a singular value decomposition for A, we can approximate A by taking a low-rank approximation, as follows:

Definition 8.21 (Rank r approximation)**.** Let A be an $m \times n$ matrix, where $m \geqslant n$, and consider a singular value decomposition $A = U\Sigma V^{\mathsf{T}}$. Recall that U is $m \times m$, Σ is $m \times n$, and V is $n \times n$. For each value $1 \leqslant r \leqslant n$, decompose U, Σ, and V as

$$U = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \qquad V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$$

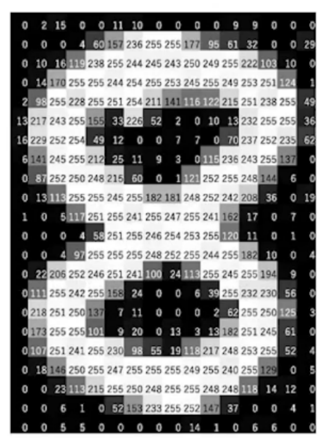
with U_1 an $m \times r$ matrix, Σ_1 an $r \times r$ matrix, and V_1^T an $r \times n$ matrix. Then

$$U_1 \Sigma_1 {V_1}^{\mathsf{T}}$$

is a **rank r approximation** of A.

If $r \leqslant \operatorname{rank}(A)$, a rank r approximation of A is an $m \times n$ matrix of rank r.

This has lots of applications, including to image processing. A digital image is composed of a collection of pixels (picture elements). A digital image contains a fixed number of rows and columns of pixels, each of which is corresponds to a certain color. In a grayscale image, each pixel has an integer value from 0 to 255 on a scale, where 0 represents black and 255 represents white. Here's an example:[1]



This is a 24×16 digital image (so there are 24×16 many pixels). To a computer, this picture is a 24×16 matrix:

[1]image from Himanshi Singh's blog

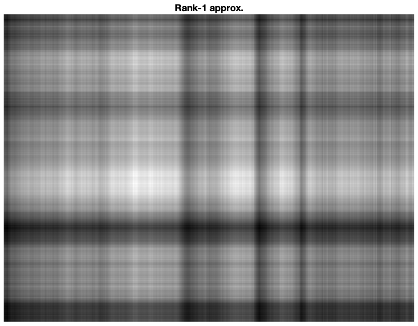
0	2	15	0	0	11	10	0	0	0	0	9	9	0	0	0
0	0	0	4	60	157	236	255	255	177	95	61	32	0	0	29
0	10	16	119	238	255	244	245	243	250	249	255	222	103	10	0
0	14	170	255	255	244	254	255	253	245	255	249	253	251	124	1
2	98	255	228	255	251	254	211	141	116	122	215	251	238	255	49
13	217	243	255	155	33	226	52	2	0	10	13	232	255	255	36
16	229	252	254	49	12	0	0	7	7	0	70	237	252	235	62
6	141	245	255	212	25	11	9	3	0	115	236	243	255	137	0
0	87	252	250	248	215	60	0	1	121	252	255	248	144	6	0
0	13	113	255	255	245	255	182	181	248	252	242	208	36	0	19
1	0	5	117	251	255	241	255	247	255	241	162	17	0	7	0
0	0	0	4	58	251	255	246	254	253	255	120	11	0	1	0
0	0	4	97	255	255	255	248	252	255	244	255	182	10	0	4
0	22	206	252	246	251	241	100	24	113	255	245	255	194	9	0
0	111	255	242	255	158	24	0	0	6	39	255	232	230	56	0
0	218	251	250	137	7	11	0	0	0	2	62	255	250	125	3
0	173	255	255	101	9	20	0	13	3	13	182	251	245	61	0
0	107	251	241	255	230	98	55	19	118	217	248	253	255	52	4
0	18	146	250	255	247	255	255	255	249	255	240	255	129	0	5
0	0	23	113	215	255	250	248	255	255	248	248	118	14	12	0
0	0	6	1	0	52	153	233	255	252	147	37	0	0	4	1
0	0	5	5	0	0	0	0	0	14	1	0	6	6	0	0

But most digital images we look at contain a lot more than 24×16 pixels... For example, here is a 1024×768 photo of my dog, taken with an iPhone (and then shrunken a bit):

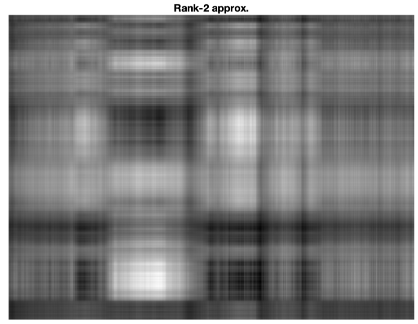
Actually, this photo is in color (RGB), which is stored in a more elaborate way: instead of one 1024×768 matrix, the image corresponds to 3 matrices (RGB), all 1024×768, the entries of which are all integers from 0 to 255. We can take those three matrices and create one grayscale image by taking the average of the RGB values in each pixel.

Even the black and white grayscale picture takes up a lot of data: a 1024×768 digital image contains 786432 pixels. We can instead consider a low rank approximation of the corresponding matrix.

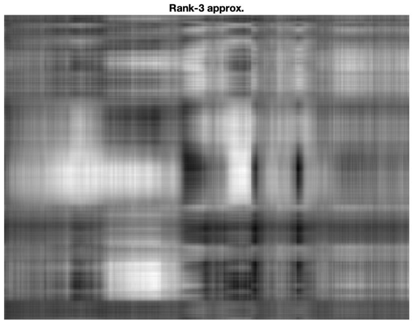
Rank-1 approx.



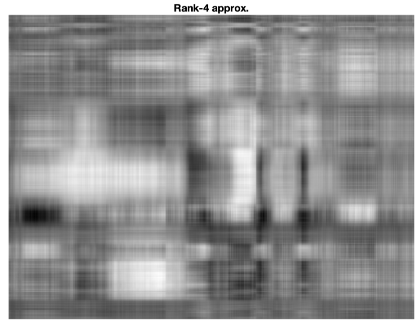
Rank-2 approx.



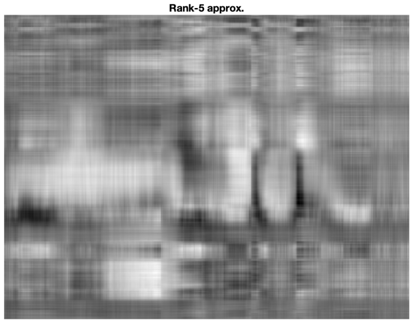
Rank-3 approx.



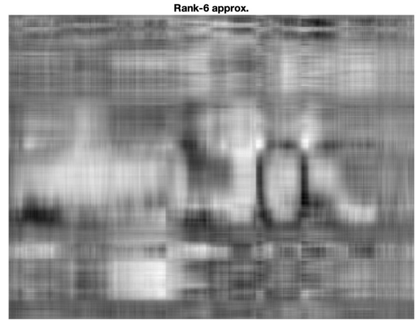
Rank-4 approx.



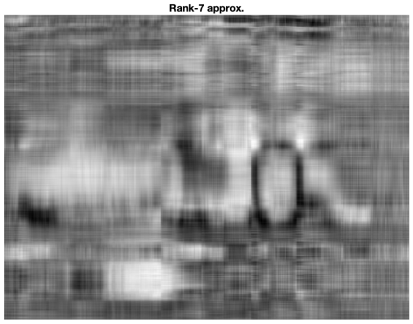
Rank-5 approx.



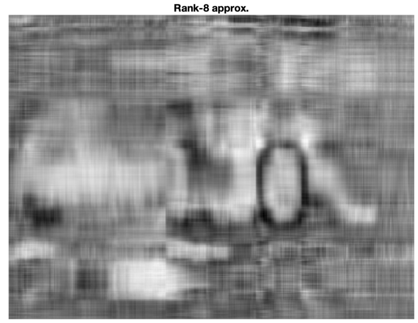
Rank-6 approx.



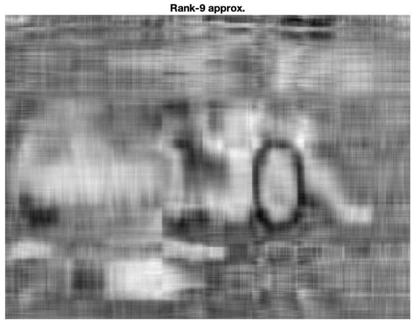
Rank-7 approx.



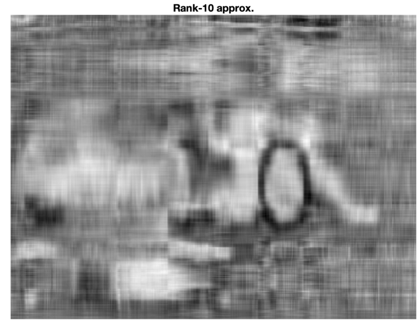
Rank-8 approx.



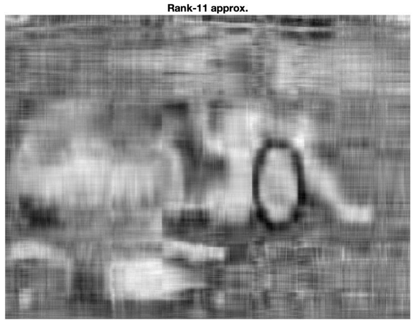
Rank-9 approx.



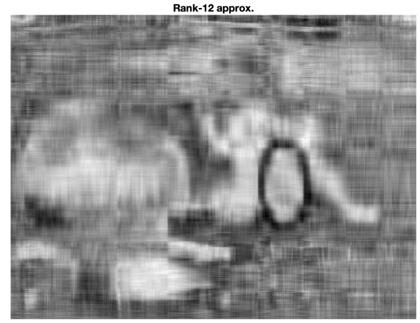
Rank-10 approx.



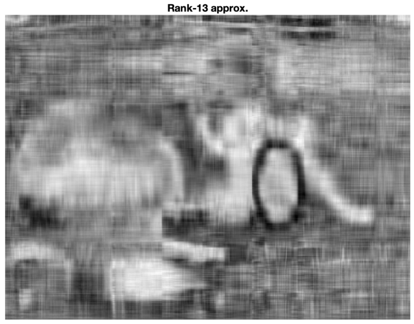
Rank-11 approx.



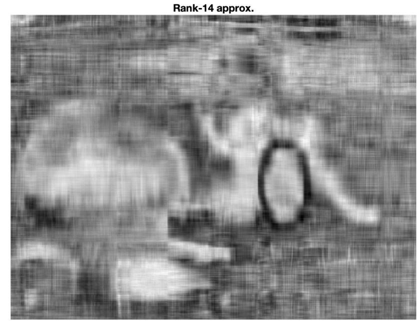
Rank-12 approx.



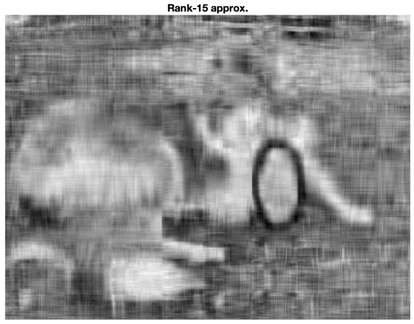
Rank-13 approx.



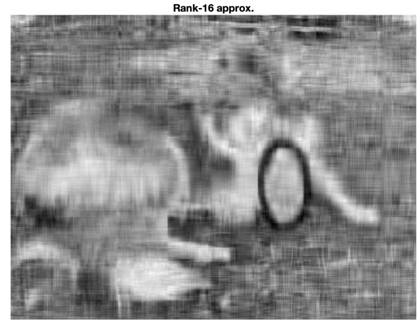
Rank-14 approx.



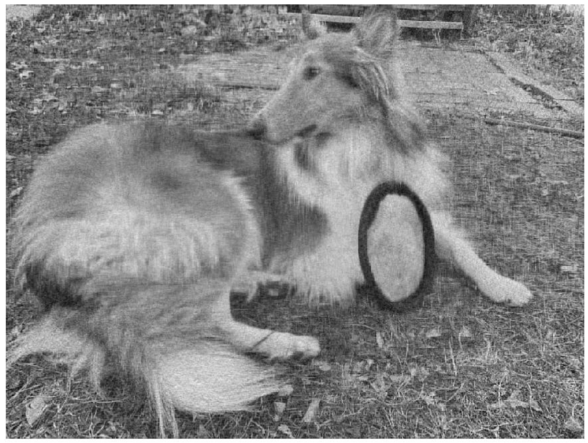
Rank-15 approx.



Rank-16 approx.



Finally, here is a rank 100 approximation of Ollie chilling:



The original matrix corresponding to the black and white photo has rank 768. But this rank 100 photo just looks like a slightly grainier version of the original photo! Here is the original photo again:

In conclusion: the largest eigenvalues have the strongest impact on the photo. Collecting only some of the data can still produce spectacularly accurate results.

Index