1 Last time: multiplying vectors matrices

Given a matrix $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$ and a vector $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$ we define

$$\boxed{Av = v_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + v_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + v_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}}.$$

We refer to Av as the product of A and v, or the vector given by multiplying v by A.

Example. We have $\begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1+0+3 \\ -5+0+7 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

Multiplying $v \in \mathbb{R}^n$ by an $m \times n$ matrix A transforms v to a new vector $Av \in \mathbb{R}^m$.

This transformation is *linear* in the sense that:

1. $A(u+v) = Au + Av$ if $u, v \in \mathbb{R}^n$.
2. $A(cv) = c(Av)$ if $v \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

If A is an $m \times n$ matrix and $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $b \in \mathbb{R}^m$, then we call $\boxed{Ax = b}$ a *matrix equation.*

A matrix equation $Ax = b$ has the same solutions as the linear system with augmented matrix $\begin{bmatrix} A & b \end{bmatrix}$.

Theorem. Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a solution for any $b \in \mathbb{R}^m$.
2. The span of the columns of A is all of $\mathbb{R}^m$.
3. A has a pivot position in every row.

Example. The matrix equation

$$\begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

may fail to have a solution since

$$\texttt{RREF}\left(\begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & * \\ 0 & 1 & * \\ 0 & 0 & 0 \end{bmatrix}$$

has pivot positions only in rows 2 and 3.

A *homogeneous* linear system is one that can be written $Ax = 0$.

Such a system has one *trivial solution* given by $x = 0$.

A homogeneous linear system has a nontrivial solution if and only if it has at least one free variable.

A homogeneous linear system has a free variable if not every column is a pivot column in its *coefficient matrix* (remember that this is the augmented matrix without the last column).

2 Linear independence

We briefly introduced the notion of linear independence last time.

Vectors $v_1, v_2, \ldots, v_p \in \mathbb{R}^n$ are *linearly independent* if the homogeneous matrix equation

$$\begin{bmatrix} v_1 & v_2 & \ldots & v_p \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = 0$$

has no nontrivial solution.

If $c_1v_1 + c_2v_2 + \cdots + c_pv_p = 0$ where $c_1, c_2, \ldots, c_p \in \mathbb{R}$ and some $c_i \neq 0$, then we refer to "$c_1v_1 + c_2v_2 + \cdots + c_pv_p = 0$" as a *linear dependence* among the vectors $v_1, v_2, \ldots, v_p$.

Vectors are *linearly independent* if there is no linear dependence among them.

Vectors which are not linearly independent are *linearly dependent*.

Example. The vectors $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$, and $\begin{bmatrix} 5 \\ 9 \\ 16 \end{bmatrix}$ are linear dependent since

$$-\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + 3\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} - \begin{bmatrix} 5 \\ 9 \\ 16 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0.$$

But $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$, and $\begin{bmatrix} 5 \\ 9 \\ 15 \end{bmatrix}$ are linearly independent, since

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 0 & 3 & 9 \\ -1 & 5 & 15 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 5 \\ 0 & 3 & 9 \\ 0 & 7 & 20 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \texttt{RREF(A)}$$

where $\sim$ denotes row equivalence. Every column of A contains a pivot position, so the linear system with coefficient matrix A has no free variables, so $Ax = 0$ have no nontrivial solutions, meaning the columns of A are linearly independent.

Some useful facts about linear independence.

1. A single v is linearly independent if and only if $v \neq 0$.

   A list of vectors is linearly dependent if it includes the 0 vector.

2. Vectors $v_1, v_2, \ldots, v_p \in \mathbb{R}^n$ are linearly dependent if and only if some vector v_i is a linear combination of the other vectors $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_p$.

   We saw this in the previous example: $\begin{bmatrix} 5 \\ 9 \\ 16 \end{bmatrix} = 3\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$.

The last thing we'll note about linear independence (for now) is this useful, non-obvious fact:

Theorem. Assume $p > n$ and $v_1, v_2, \ldots, v_p \in \mathbb{R}^n$. Then these vectors are linearly dependent.

Proof. Let $A = \begin{bmatrix} v_1 & v_2 & \dots & v_p \end{bmatrix}$.

This matrix has more columns than rows.

Each row contains at most one pivot position, so there are fewer pivot positions than columns.

Therefore some column is not a pivot column.

This means the linear system $Ax = 0$ has a free variable, so has a nontrivial solution.

This implies that $v_1, v_2, \dots, v_p$, the columns of A, are linearly dependent. □

Example. Suppose $v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $v_3 = \begin{bmatrix} 5 \\ 60 \end{bmatrix}$. Then

$$A = \begin{bmatrix} 1 & 1 & 5 & 0 \\ 2 & 3 & 60 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 5 & 0 \\ 0 & 1 & 50 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -45 & 0 \\ 0 & 1 & 50 & 0 \end{bmatrix} = \texttt{RREF(A)}.$$

Column 3 of A contains no pivot position, so x_3 is a free variable in the vector equation $x_1v_1+x_2v_2+x_3v_3$.

Therefore v_1, v_2, v_3 are linearly dependent.

In fact we have $x_1v_1 + x_2v_2 + x_3v_3 = 0$ if and only if $x_1 - 45x_3 = x_2 + 50x_3 = 0$.

Take $x_3 = 1$. Then $x_1 = 45$ and $x_2 = -50$, so $45v_1 - 50v_2 + v_3 = 0$.

3 Linear transformations

A function f (like the ones we see in calculus) takes an input x from some set X (for example, $\mathbb{R}$) and produces an output $f(x)$ in another set Y

We write $f : X \to Y$ to mean that f is a function that takes inputs from X and gives outputs in Y.

X is called the *domain* of the function f.

Y is sometimes called the *codomain* of f.

Remark. For every x in the domain X of f, we get an output $f(x)$.

It possible that some values y in the codomain Y may never occur as outputs of f, however.

The *image* of an input x in X under f is the ouput $f(x)$.

Define the *image* or *range* of the function f to be the subset $\{f(x) : x \in X\}$ of the codomain Y. This is the set of all possible outputs of f. We denote the range of f by range(f).

Example. An $m \times n$ matrix A is a function $A : \mathbb{R}^n \to \mathbb{R}^m$. Given an input vector $v \in \mathbb{R}^n$, the corresponding output is $Av \in \mathbb{R}^m$.

Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$ whose domain and codomain are the sets of all vectors in dimensions m and n. The function f is a *linear transformation* or *linear function* if both of these properties hold:

(i) $f(u + v) = f(u) + f(v)$ for all vectors $u, v \in \mathbb{R}^n$.

(ii) $f(cv) = cf(v)$ for all vectors $v \in \mathbb{R}^n$ and scalars $c \in \mathbb{R}$.

Linear transformations have the following additional properties:

Proposition. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation then

(iii) $f(0) = 0$.

(iv) $f(u - v) = f(u) - f(v)$ for $u, v \in \mathbb{R}^n$.

(v) $f(au+bv) = af(u) + bf(v)$ for all $a, b \in \mathbb{R}$ and $u, v \in \mathbb{R}^n$.

Proof. (iii) We have $f(0) = f(0+0) = 2f(0)$ so $f(0) = 0$.

(iv) We have $f(u-v) = f(u) + f(-v) = f(u) + (-1)f(v) = f(u) - f(v)$.

(v) We have $f(au+bv) = f(au) + f(bv) = af(u) + bf(v)$. □

Example. We have already seen that an $m \times n$ matrix A defines a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$.

Suppose $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$ and $T : \mathbb{R}^2 \to \mathbb{R}^3$ is the function defined by $T(v) = Av$.

(a) The image of a vector $v \in \mathbb{R}^2$ under T is by definition $T(v) = Av$.

The image of $v = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ under T is $T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$.

(b) Is the range of T all of $\mathbb{R}^3$? If it was, then (from results last time) A would have to have a pivot position in every row. This is impossible since each column can only contain one pivot position, but A has three rows and only two columns. Therefore range$(T) \neq \mathbb{R}^3$.

Example. Fix $\theta \in [0, 2\pi)$. The notation $[a, b)$ means "the set of numbers $x \in \mathbb{R}$ with $a \leq x < b$." Define

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

and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation $T(v) = Av$.

How does $T(v)$ compare to v, in terms of geometric representations of vectors in $\mathbb{R}^2$?

1. If $v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is a vector parallel to the x-axis, then $T(v) = Av = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$. Draw a picture of this:

   In words: $T(v)$ is the arrow from the origin to the point on the unit circle which is angle θ counterclockwise from $(x, y) = (1, 0)$.

2. If $v = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is a vector parallel to the y-axis, then $T(v) = Av = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} = \begin{bmatrix} \cos(\theta + \frac{\pi}{2}) \\ \sin(\theta + \frac{\pi}{2}) \end{bmatrix}$. Draw a picture of this:

   In words: $T(v)$ is the arrow from 0 to the point on the unit circle which is angle $\theta + \frac{\pi}{2}$ counterclockwise from $(x, y) = (1, 0)$, which is the point angle θ counterclockwise from $(x, y) = (0, 1)$.

It appears from these examples that $T(v)$ is the vector given by rotating v counterclockwise by angle θ.

How do we know this is true for any $v \in \mathbb{R}^2$? Use linearity and the fact that the sum of two vectors u and v in $\mathbb{R}^2$ corresponds to the arrow from the origin to the opposite vertex of the parallelogram with sides u and v.

Any vector $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ can be written $v = \begin{bmatrix} v_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \end{bmatrix}$, so is the arrow to the opposite vertex in the parallelogram with sides $\begin{bmatrix} v_1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ v_2 \end{bmatrix}$. Since

$$T(v) = T\left(\begin{bmatrix} v_1 \\ 0 \end{bmatrix}\right) + T\left(\begin{bmatrix} 0 \\ v_2 \end{bmatrix}\right)$$

and since T rotates by angle θ the two vectors on the right, it follows that $T(v)$ is the arrow from 0 to the opposite vertex in our previous parallelogram, now rotated counterclockwise by angle θ. Draw a picture to convince yourself:

Theorem. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then there is a unique $m \times n$ matrix A such that $T(v) = Av$ for all $v \in \mathbb{R}^n$.

Moral: Matrices uniquely represent all linear transformations $\mathbb{R}^n \to \mathbb{R}^m$.

Proof. Define $e_1, e_2, \dots, e_n \in \mathbb{R}^n$ as the vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad e_{n-1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

so that e_i has a 1 in the ith row and 0 in all other rows.

Define $a_i = T(e_i) \in \mathbb{R}^m$ and $A = \begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_n \end{bmatrix}$.

If w is any vector $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{R}^n$ then

$$\begin{aligned} T(w) &= T(w_1 e_1 + w_2 e_2 + \dots + w_n e_n) \\ &= w_1 T(e_1) + w_2 T(e_2) + \dots + w_n T(e_n) = w_1 a_1 + w_2 a_2 + \dots + w_n a_n = Aw. \end{aligned}$$

Thus A is one matrix such that $T(v) = Av$ for all vectors $v \in \mathbb{R}^n$.

To show that A is the only such matrix, suppose B is a $m \times n$ matrix with $T(v) = Bv$ for all $v \in \mathbb{R}^n$.

Then $T(e_i) = Ae_i = Be_i$ for all $i = 1, 2, \dots, n$.

But Ae_i and Be_i are the ith columns of A and B. For example,

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} e_3 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

Therefore A and B have the same columns, so they are the same matrix: $A = B$. □

We call the matrix A in this theorem the *standard matrix* of the linear transformation T.

Example. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is the function $T(v) = 3v$.

This is a linear transformation. (Why?) What is the standard matrix A of T?

As we saw in the proof of the theorem, the standard matrix of $T : \mathbb{R}^n \to \mathbb{R}^n$ is

$$A = \begin{bmatrix} T(e_1) & T(e_2) & \dots & T(e_n) \end{bmatrix} = \begin{bmatrix} 3e_1 & 3e_2 & \dots & 3e_n \end{bmatrix} = \begin{bmatrix} 3 & 0 & \dots & 0 \\ 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 3 \end{bmatrix}.$$

In words, A is the matrix with 3 in each position $(1,1),(2,2),\dots,(n,n)$ and 0 in all other positions.

One calls such a matrix *diagonal*.

Example. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is the function

$$T\left(\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}\right) = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1^2 + v_2^2 + \cdots + v_n^2.$$

This function is *not* linear: we have $T(2v) = 4T(v) \neq 2T(v)$ for any nonzero vector $v \in \mathbb{R}^n$.

Example. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is the function

$$T\left(\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}\right) = \begin{bmatrix} v_n \\ \vdots \\ v_2 \\ v_1 \end{bmatrix}.$$

This function is a linear transformation. (Why?) Its standard matrix is

$$A = \begin{bmatrix} T(e_1) & T(e_2) & \dots & T(e_{n-1}) & T(e_n) \end{bmatrix} = \begin{bmatrix} e_n & e_{n-1} & \dots & e_2 & e_1 \end{bmatrix} = \begin{bmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{bmatrix}.$$

In the matrix on the right, we adopt the convention of only writing the nonzero entries: all positions in the matrix which are blank contain zero entries.

Definition. A function $f : X \to Y$ is *one-to-one* or *injective* if $f(a) = f(b)$ implies $a = b$. In words: f does not send two different inputs to the same output.

Theorem. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear then T is one-to-one if and only if the only solution to $T(x) = 0$ is $x = 0 \in \mathbb{R}^n$, i.e., the columns of the standard matrix A of T are linearly independent.

Proof. If T is not one-to-one, then there are vectors $u, v \in \mathbb{R}^n$ with $u \neq v$ and $T(u) = T(v)$.

In this case $u - v \neq 0$ and $T(u - v) = T(u) - T(v) = 0$ so $T(x) = 0$ has a nontrivial solution.

If T is one-to-one, then $T(x) = T(0) = 0$ implies $x = 0$, so $T(x) = 0$ has only trivial solutions. □

Definition. A function $f : X \to Y$ is *onto* or *surjective* if $\text{range}(f) = \{f(x) : x \in X\} = Y$. In words: the range of f is equal to its codomain.

Theorem. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear then T is onto if and only if the columns of the standard matrix A of T span $\mathbb{R}^m$.

Proof. The vectors in the range of T are precisely the linear combinations of the columns of A.

The range is $\mathbb{R}^m$ precisely when the span of the columns of A is $\mathbb{R}^m$. □