1 Last time: inverses

The *transpose* of an $m \times n$ matrix A is the $n \times m$ matrix A^T whose rows are the columns of A.

For example, $\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}^T = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$.

If A is $m \times n$ and B is $n \times k$, then AB is defined and $(AB)^T = B^T A^T$.

The following all mean the same thing for a function $f : X \to Y$:

1. f is *invertible*.
2. f is one-to-one and onto.
3. For each $b \in Y$ there is exactly one $a \in X$ with $f(a) = b$.
4. There is a unique function $f^{-1} : Y \to X$, called the *inverse* of f, such that

$$f^{-1}(f(a)) = a \qquad \text{and} \qquad f(f^{-1}(b)) = b \qquad \text{for all } a \in X \text{ and } b \in Y.$$

Proposition. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear and invertible then $m = n$ and T^{-1} is invertible.

The following all mean the same thing for an $n \times n$ matrix A:

1. A is *invertible*.
2. A is the standard matrix of an invertible linear function $T : \mathbb{R}^n \to \mathbb{R}^n$.
3. There is a unique $n \times n$ matrix A^{-1}, called the *inverse* of A, such that

$$A^{-1}A = AA^{-1} = I_n = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

4. For each $b \in \mathbb{R}^n$ the equation $Ax = b$ has a unique solution.
5. $\texttt{RREF}(A) = I_n$
6. The columns of A span $\mathbb{R}^n$ and are linearly independent.

Proposition. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be an arbitrary 2×2 matrix.

(1) A is invertible if and only if $ad - bc \neq 0$.

(2) If $ad - bc \neq 0$ then $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

For example, $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} = \frac{1}{-2} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}$.

Proposition. Let A and B be $n \times n$ matrices.

1. If A is invertible then $(A^{-1})^{-1} = A$.
2. If A and B are both invertible then AB is invertible with $(AB)^{-1} = B^{-1}A^{-1}$.

3. If A is invertible then A^T is invertible with $(A^T)^{-1} = (A^{-1})^T$.

Process to compute A^{-1}

Let A be an $n \times n$ matrix. Consider the $n \times 2n$ matrix $\begin{bmatrix} A & I_n \end{bmatrix}$.

If A is invertible then $\texttt{RREF}\left(\begin{bmatrix} A & I_n \end{bmatrix}\right) = \begin{bmatrix} I_n & A^{-1} \end{bmatrix}$.

So to compute A^{-1}, row reduce $\begin{bmatrix} A & I_n \end{bmatrix}$ to reduced echelon form, and then take the last n columns.

Example. $\begin{bmatrix} 1 & 3 & 1 & 0 \\ 5 & 8 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 1 & 0 \\ 0 & -7 & -5 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 1 & 0 \\ 0 & 1 & 5/7 & -1/7 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -8/7 & 3/7 \\ 0 & 1 & 5/7 & -1/7 \end{bmatrix}$.

Therefore $\begin{bmatrix} 1 & 3 \\ 5 & 8 \end{bmatrix}^{-1} = \frac{1}{-7}\begin{bmatrix} 8 & -3 \\ -5 & 1 \end{bmatrix}$, which agrees with our formula in the 2×2 case.

2 Stronger characterization of invertible matrices

Remember that a matrix can only be invertible if it has the same number of rows and columns.

Theorem. When A is a square matrix, the following are equivalent:

(a) A is invertible.

(b) The columns of A are linearly independent.

(c) The columns of A span $\mathbb{R}^n$

We said earlier that a matrix is invertible if and only if its columns both are linearly independent and span $\mathbb{R}^n$. This is still true, but it turns out that if we know ahead of time that A is a square matrix, then either condition (b) or (c) implies the other.

Proof. We already know that (a) implies both (b) and (c).

Assume just (b) holds. Then A has a pivot position in every column, so $\texttt{RREF}(A) = I_n$ since A has the same number of rows and columns. But this implies that A is invertible.

Similarly, if (c) holds then A has a pivot position in every row, so $\texttt{RREF}(A) = I_n$ and A is invertible. □

Corollary. Suppose A and B are $n \times n$ matrices. If $AB = I_n$ then $BA = I_n$.

This means that if we want to show that $B = A^{-1}$ then it is enough to just check that $AB = I_n$.

Proof. Assume $AB = I_n$. Then the columns of A span $\mathbb{R}^n$ since if $v \in \mathbb{R}^n$ then $Au = v$ for $u = Bv \in \mathbb{R}^n$, so A is invertible. Therefore $B = A^{-1}AB = A^{-1}I_n = A^{-1}$ so $BA = A^{-1}A = I_n$. □

Aside (optional reading). The corollary is equivalent to saying that if $T, U : \mathbb{R}^n \to \mathbb{R}^n$ are linear and $T \circ U = \mathrm{id}_{\mathbb{R}^n}$ is the identity function, then $U \circ T = \mathrm{id}_{\mathbb{R}^n}$. Curiously, this fails in "infinite dimensions."

Define $\mathbb{R}^\infty$ as the set of "infinite" column vectors $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \end{bmatrix}$ where $v_1, v_2, \dots$ are real numbers but only finitely many are nonzero. Every element $v \in \mathbb{R}^\infty$ is formed by taking an ordinary vector in $\mathbb{R}^n$ for some n and then adding on infinitely many extra rows of zeros. If an infinite column vector seems strange, another way to view elements of $\mathbb{R}^n$ is as functions $v : \{1, 2, 3, \dots\} \to \mathbb{R}$ with the property that the number of positive integers i with $v(i) \neq 0$ is finite.

Sums and scalar multiples of vectors in $\mathbb{R}^\infty$ are defined as coordinate-wise operations exactly as for vectors in $\mathbb{R}^n$, and so we can define linear transformations $\mathbb{R}^\infty \to \mathbb{R}^\infty$ by the same pair of conditions as we use to define linear transformations $\mathbb{R}^n \to \mathbb{R}^m$. (Can you write down the details?)

Now consider the functions $T, U : \mathbb{R}^\infty \to \mathbb{R}^\infty$ given by the *shift operators*

$$U\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{bmatrix}\right) = \begin{bmatrix} 0 \\ v_1 \\ v_2 \\ \vdots \end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{bmatrix}\right) = \begin{bmatrix} v_2 \\ v_3 \\ v_4 \\ \vdots \end{bmatrix}.$$

In words, U shifts a vector down by adding a zero row at the top, while T shifts a vector up by forgetting the first row. Both of these functions are linear transformations. (Check this!)

We have $T \circ U = \mathrm{id}_{\mathbb{R}^\infty}$ since

$$T\left(U\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{bmatrix}\right)\right) = T\left(\begin{bmatrix} 0 \\ v_1 \\ v_2 \\ \vdots \end{bmatrix}\right) = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{bmatrix}.$$

However, $U \circ T \neq \mathrm{id}_{\mathbb{R}^\infty}$ since

$$U\left(T\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{bmatrix}\right)\right) = U\left(\begin{bmatrix} v_2 \\ v_3 \\ v_4 \\ \vdots \end{bmatrix}\right) = \begin{bmatrix} 0 \\ v_2 \\ v_3 \\ \vdots \end{bmatrix}.$$

Linear algebra is the study of the linear transformations $\mathbb{R}^n \to \mathbb{R}^m$ where n and m are finite numbers. The study of linear transformations $\mathbb{R}^\infty \to \mathbb{R}^\infty$ is *functional analysis* (MATH 4063).

3 Subspaces of $\mathbb{R}^n$

Returning to our usual convention, let n be a positive integer (not ∞).

Recall that $0 \in \mathbb{R}^n$ denotes the zero vector $0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

Subsets of $\mathbb{R}^n$ that are closed under scalar multiplication and addition are called *subspaces*. To be precise:

Definition. Let H be a subset of $\mathbb{R}^n$. The subset H is a *subspace* if these three conditions hold:

1. $0 \in H$.
2. $u + v \in H$ if $u, v \in H$.
3. $cv \in H$ if $c \in \mathbb{R}$ and $v \in H$.

<u>Common examples</u>

$\mathbb{R}^n$ is a subspace of itself.

The set $\{0\}$ consisting of just the zero vector is a subspace of $\mathbb{R}^n$.

The empty set $\varnothing$ is *not* a subspace since it does not contain 0.

A subset $H \subset \mathbb{R}^2$ is a subspace if and only if $H = \{0\}$ or $H = \mathbb{R}^2$ or H is a line through 0.

The span of any set of vectors in $\mathbb{R}^n$ is a subspace.

(Later, we will see that every subspace is the span of some set of vectors.)

Example. The set X of vectors $v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3$ with $v_1 + v_2 + v_3 = 1$ is *not* a subspace since $0 \notin X$.

The set H of vectors $v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3$ with $v_1 + v_2 + v_3 = 0$ is a subspace since if $u, v \in H$ and $c \in \mathbb{R}$ then

$$(u_1 + v_1) + (u_2 + v_2) + (u_3 + v_3) = (u_1 + u_2 + u_3) + (v_1 + v_2 + v_3) = 0 + 0 = 0$$

and

$$cv_1 + cv_2 + cv_3 = c(v_1 + v_2 + v_3) = 0$$

so $u + v \in H$ and $cv \in H$.

Any matrix A gives rise to two subspaces, called the *column space* and *null space*.

Definition. The *column space* of an $m \times n$ matrix A is the subspace

$$\operatorname{Col} A \subset \mathbb{R}^m$$

given by the span of the columns of A.

Remark. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is the linear function $T(x) = Ax$ then $\operatorname{Col} A = \operatorname{range}(T)$.

Note that $\operatorname{Col} A = \mathbb{R}^m$ if and only if $Ax = b$ has a solution for each $b \in \mathbb{R}^m$.

A vector $b \in \mathbb{R}^m$ belongs to $\operatorname{Col} A$ if and and only if $Ax = b$ has a solution.

Definition. The *null space* of an $m \times n$ matrix A is the subspace

$$\operatorname{Nul} A \subset \mathbb{R}^n$$

given by the set of vectors $v \in \mathbb{R}^n$ with $Av = 0$.

Proof that $\operatorname{Nul} A$ *is a subspace.* If $u, v \in \operatorname{Nul} A$ and $c \in \mathbb{R}$ then $A(u + v) = Au + Av = 0 + 0 = 0$ and $A(cv) = c(Av) = 0$, so $u + v \in \operatorname{Nul} A$ and $cv \in \operatorname{Nul} A$. Thus $\operatorname{Nul} A$ is a subspace of $\mathbb{R}^n$. □

Remark. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is the linear function $T(x) = Ax$ then $\operatorname{Nul} A = \{x \in \mathbb{R}^n : T(x) = 0\}$. This is usually called the *kernel* of T.

Note: the column space is a subspace of $\mathbb{R}^m$ where m is the number of rows of A, while the null space is a subspace of $\mathbb{R}^n$ where n is the number of columns of A.

At first, subspaces seem like big, complicated objects. But it turns out that each subspace is completely determined by a finite amount of data. This data will be called a *basis*. Let H be a subspace of $\mathbb{R}^n$.

Definition. A *basis* for H is a set of vectors $\{v_1, v_2, \dots, v_k\} \subset H$ which are linearly independent and have span equal to H.

The empty set $\varnothing$ is considered to be a basis for the zero vector space $\{0\}$.

Example. The vectors $\{e_1, e_2, \dots, e_n\} \subset \mathbb{R}^n$ where $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, etc., is a basis for $\mathbb{R}^n$.

We call this the *standard basis* of $\mathbb{R}^n$.

Theorem. Every subspace H of $\mathbb{R}^n$ has a basis of size at most n.

Proof. If $H = \{0\}$ then $\varnothing$ is a basis.

Assume $H \neq \{0\}$. Let $\mathcal{B}$ be a set of linearly independent vectors in H that is as large as possible. The size of $\mathcal{B}$ must be at most n since any $n+1$ vectors in $\mathbb{R}^n$ are linearly dependent by a result proved in an earlier lecture.

Let $w_1, w_2, \dots, w_k$ be the elements of $\mathcal{B}$. Since $\mathcal{B}$ is as large as possible, if $v \in H$ is any vector then $w_1, w_2, \dots, w_k, v$ are linearly dependent so we can write $c_1 v_1 + c_2 v_2 + \dots + c_k v_k + cv = 0$ for some numbers $c_1, c_2, \dots, c_k, c \in \mathbb{R}$ which are not all zero. Since the vectors in $\mathcal{B}$ are linearly independent, we must have $c \neq 0$ (why?) so it follows that

$$v = \tfrac{c_1}{c} w_1 + \tfrac{c_2}{c} w_2 + \dots + \tfrac{c_k}{c} w_k.$$

Thus, not only is $\mathcal{B}$ a set of linearly independent vectors, but these vectors also span H, so $\mathcal{B}$ is a basis. $\square$

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

How can we find a basis for $\operatorname{Nul} A$? Well, finding a basis for $\operatorname{Nul} A$ is more or less the same task as finding all solutions to the homogeneous equation $Ax = 0$. So let's first try to solve that equation.

If we row reduce the 3×6 matrix $\begin{bmatrix} A & 0 \end{bmatrix}$, we get

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

This tells us that $Ax = 0$ if and only if $\begin{cases} x_1 - 2x_2 - x_4 + 3x_5 = 0 \\ x_3 + 2x_4 - 2x_5 = 0. \end{cases}$

Therefore $x \in \operatorname{Nul} A$ if and only if

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}.$$

The vectors

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

are a basis for $\mathrm{Nul}\, A$: we just computed that these vectors span the null space, and they are linearly independent since each has a nonzero entry in a row (namely, either row 2, 4, or 5) whether the others have zeros. (Why does this imply linear independence?)

This example is important: the procedure just described works to construct a basis of $\mathrm{Nul}\, A$ for any matrix A. The size of this basis will always be equal to the number of free variables in the linear system $Ax = 0$. How to find a basis for $\mathrm{Nul}\, A$ is something you should remember at the end of this course.

Example. Let $B = \begin{bmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

This matrix is in reduced echelon form.

Finding a basis for $\mathrm{Col}\, B$ is in some ways easier than finding a basis for $\mathrm{Nul}\, B$.

The columns of B automatically span $\mathrm{Col}\, B$, but they might not be linearly independent.

The largest linearly independent subset of the columns of B will be a basis for $\mathrm{Col}\, B$, however.

In our example, the pivot columns 1, 2 and 5 are linearly independent since each has a row with a 1 where the others have 0s. These columns span columns 3 and 4, so it follows that

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

is a basis for $\mathrm{Col}\, B$.

This example was special since the matrix B was already in reduced echelon form. To find a basis of the column space of an arbitrary matrix, we rely on the following observation:

Proposition. Let A be any matrix. The pivot columns of A form a basis for $\mathrm{Col}\, A$.

Proof. This proof sketches the main ideas but doesn't spell out all the details.

Suppose A is $m \times n$. The reduced echelon form of A is obtained by multiplying A by an invertible matrix E on the left, so we can write $\texttt{RREF}(A) = EA$.

If $a_1, a_2, \dots, a_k$ are the pivot columns of A, then $E \begin{bmatrix} a_1 & a_2 & \dots & a_k \end{bmatrix}$ is the $m \times k$ matrix $\begin{bmatrix} I_k \\ 0 \end{bmatrix}$ where the 0 means an $(m-k) \times n$ submatrix of zeros. These columns are linearly independent since if

$$\begin{bmatrix} a_1 & a_2 & \dots & a_k \end{bmatrix} v = 0$$

for $v \in \mathbb{R}^k$ then

$$0 = E \begin{bmatrix} a_1 & a_2 & \dots & a_k \end{bmatrix} v = \begin{bmatrix} I_k \\ 0 \end{bmatrix} v = \begin{bmatrix} v \\ 0 \end{bmatrix}$$

which implies that $v = 0$.

Suppose w is a non-pivot column of A. The definition of reduced echelon form implies that the corresponding column Ew of $\texttt{RREF}(A) = EA$ is in the span of $Ea_1, Ea_2, \dots, Ea_k$. (Why?) If we have

$Ew = c_1Ea_1 + \cdots + c_kEa_k$ then multiplying both sides by E^{-1} gives $w = c_1a_1 + \cdots + c_ka_k$ so w is in the span $a_1, a_2, \ldots, a_k$. Therefore the pivot columns of A span the other columns, and hence span $\operatorname{Col} A$.

Since the pivot columns are linearly independent and have span equal to $\operatorname{Col} A$, they form a basis. $\square$

Example. The matrix

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix}$$

is row equivalent to the matrix B in the previous example. The pivot columns of A are therefore also columns 1, 2, and 5, so

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -9 \\ 2 \\ 1 \\ -8 \end{bmatrix} \right\}$$

is a basis for $\operatorname{Col} A$.

Next time: we will show that if H is a subspace of $\mathbb{R}^n$ then all of its bases have the same size. The common size of these basis is the *dimension* of H.