1 Last time: theorems about bases and rank

A *subspace* of $\mathbb{R}^n$ is a nonempty subset H with the property that $u + v \in H$ and $cv \in H$ whenever $u, v \in H$ and $c \in \mathbb{R}$. (Requiring H that be nonempty is equivalent to requiring that $0 \in H$.)

A *basis* of a subspace is a linearly independent set of vectors whose span is the whole subspace.

(The plural of "basis" is "bases.")

Two crucial facts:

- Every subspace has a basis.
- Every basis of a given subspace has the same number of elements.

The *dimension* of a subspace is the common size of all of its bases. If H is a subspace with $\dim H = p$ then any set of p vectors in H which are linearly independent, or which span H, form a basis for H.

The dimension of $\mathbb{R}^n$ is n, while the dimension of $\{0\}$ is 0.

Be sure to know how to (1) construct a basis of $\mathrm{Nul}\, A$ and (2) construct a basis of $\mathrm{Col}\, A$.

Theorem (Rank theorem)**.** Let A be an $m \times n$ matrix.

1. The dimension of the nullspace $\mathrm{Nul}\, A = \{v \in \mathbb{R}^n : Av = 0\}$ is the number of free variables in the linear system $Ax = 0$.
2. The dimension of the column space $\mathrm{Col}\, A$ (given by the span of the columns of A) is the number of pivot columns in A.
3. It holds that $\mathrm{rank}\, A + \dim \mathrm{Nul}\, A = n$, where we define $\mathrm{rank}\, A = \dim \mathrm{Col}\, A$.

(Exercise: why does the third statement follow from the first two?)

Corollary. For an $n \times n$ matrix A, the following are equivalent:

1. A is invertible.
2. $\mathrm{rank}\, A = n$.
3. $\dim \mathrm{Nul}\, A = 0$.

Remember that if U and V are two sets then $U \subset V$ means that every element of U is also an element of V. Thus, the only way that we can have both $U \subset V$ and $V \subset U$ is if $U = V$.

Last time we also proved this proposition:

Proposition. Suppose U, V are two subspaces of $\mathbb{R}^n$ with $U \subset V$. Then $\dim U \leq \dim V$, and if the two subspaces have the same dimension $\dim U = \dim V$, then $U = V$.

Another way of defining a basis of a subspace H of $\mathbb{R}^n$ is as a set of vectors $b_1, b_2, \dots, b_k$ with the property that if m is any positive integer and $v_1, v_2, \dots, v_k$ are any vectors in $\mathbb{R}^m$, there there is a unique linear transformation $T : H \to \mathbb{R}^m$ with $T(b_i) = v_i$ for $i = 1, 2, \dots, k$.

For our applications, it is not essential to know how to prove this. But if you wanted to try: first show that the existence of such a linear transformation T follows from the linear independence of $b_1, b_2, \dots, b_k$. Then check that T is unique exactly when $b_1, b_2, \dots, b_k$ span H.

If $\mathcal{B} = (b_1, b_2, \dots, b_k)$ is an ordered basis of a k-dimensional subspace H, then we define

$$[\cdot]_{\mathcal{B}} : H \to \mathbb{R}^k$$

as the unique linear function with $[b_i]_{\mathcal{B}} = e_i \in \mathbb{R}^k$ for $i = 1, 2, \dots, k$.

Recall that $e_i \in \mathbb{R}^k$ is the vector with 1 in row i and 0 in all other rows.

We call $[v]_\mathcal{B}$ the *coordinate vector* of $v \in H$ in the basis $\mathcal{B}$.

2 Determinants

The subject of the next few lectures is the *determinant* of a square matrix. We will approach this first as an abstract function with a few special properties. The determinant ends up being ubiquitous and important in various parts of math and physics, for example, in computing integrals in multivariable calculus and defining eigenvalues later in this course.

Our first "definition" of the determinant is via the following theorem, which essentially says that a set of three special properties uniquely identifies the determinant among all functions on $n \times n$ matrices.

Theorem. Let n be any positive integer. There exists a unique function

$$\det : \{n \times n \text{ matrices}\} \to \mathbb{R},$$

called the *determinant*, with the following properties:

(i) $\det I_n = 1$. In words: the determinant of the identity matrix is 1.

(ii) If $a_1, a_2, \dots, a_n \in \mathbb{R}^n$ and $1 \leq i < j \leq n$ then

$$\det \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_j & \cdots & a_n \end{bmatrix} = -\det \begin{bmatrix} a_1 & \cdots & a_j & \cdots & a_i & \cdots & a_n \end{bmatrix}$$

In words: interchanging two columns in an $n \times n$ matrix reverses the sign of the determinant.

(iii) If $a_1, a_2, \dots, a_n, u, v \in \mathbb{R}^n$ then $\det \begin{bmatrix} a_1 & \cdots & a_{i-1} & u+v & a_{i+1} & \cdots & a_n \end{bmatrix}$ is equal to

$$\det \begin{bmatrix} a_1 & \cdots & a_{i-1} & u & a_{i+1} & \cdots & a_n \end{bmatrix} + \det \begin{bmatrix} a_1 & \cdots & a_{i-1} & v & a_{i+1} & \cdots & a_n \end{bmatrix},$$

and if $c \in \mathbb{R}$ then $\det \begin{bmatrix} a_1 & \cdots & a_{i-1} & cv & a_{i+1} & \cdots & a_n \end{bmatrix}$ is equal to

$$c \cdot \det \begin{bmatrix} a_1 & \cdots & a_{i-1} & v & a_{i+1} & \cdots & a_n \end{bmatrix}.$$

In words: if all but one column of an $n \times n$ matrix are fixed, and the determinant is viewed as a function of the remaining column, then we get a function $\mathbb{R}^n \to \mathbb{R}$ which is linear.

This is a very abstract way of defining a function. At this point there is a lot to digest, and it is not at all clear, even if we knew the theorem were true, how we could compute $\det A$ for any particular square matrix. However, the advantage in abstraction is that we can quickly derive several different concrete formulas for the determinant, each of which would be hard to derive from the others.

We spend the rest of this lecture proving the theorem. To do this, we start by *assuming there exists a function* det *with the given properties*. We will use these properties to narrow the possibilities for det down to one function given by a certain formula, and then check that this formula does satisfy (i)-(iii).

Let A be an $n \times n$ matrix with columns $a_1, a_2, \dots, a_n \in \mathbb{R}^n$.

Lemma. If A has two equal columns then $\det A = 0$.

Proof. Suppose $A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$ where $a_i = a_j$ for $i < j$.

Then $\det A = -\det \begin{bmatrix} a_1 & \cdots & a_j & \cdots & a_i & \cdots & a_n \end{bmatrix} = -\det A$ so $2 \det A = 0$ and $\det A = 0$. ☐

Corollary. If A has a column which is a linear combination of its other columns, i.e., if the columns of A are not linearly independent, then $\det A = 0$.

Proof. If $A = \begin{bmatrix} a_1 & a_2 & \dots a_n \end{bmatrix}$ and $a_1 = c_2a_2 + c_3a_3 + \dots + c_na_n$ for some numbers $c_2, c_3, \dots, c_n \in \mathbb{R}$, then $\det A = c_2 \det \begin{bmatrix} a_2 & a_2 & \dots a_n \end{bmatrix} + c_3 \det \begin{bmatrix} a_3 & a_2 & a_3 & \dots a_n \end{bmatrix} + \dots + c_n \det \begin{bmatrix} a_n & a_2 & \dots a_n \end{bmatrix}$. Each determinant in the sum is zero by the previous lemma so $\det A = 0$.

If a different column of A is a linear combination of the other columns, then define B by swapping that column and the first column of A. Then $\det A = -\det B$ and the argument in the previous paragraph shows that $\det B = 0$, so again $\det A = 0$. $\square$

This leads to the following nontrivial property of the determinant:

Corollary. If A is not invertible then $\det A = 0$.

Proof. If A is not invertible then its columns are not linearly independent. $\square$

With these facts, we can already derive an explicit formula for $\det A$ when $n = 1$ or $n = 2$.

Example. For 1×1 matrices we have $\det \begin{bmatrix} a \end{bmatrix} = a \det \begin{bmatrix} 1 \end{bmatrix} = a$.

Example. For 2×2 matrices we have

$$\begin{aligned}
\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \det \begin{bmatrix} \begin{bmatrix} a \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ c \end{bmatrix} & \begin{bmatrix} b \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ d \end{bmatrix} \end{bmatrix} \\
&= \det \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} + \det \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} + \det \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix} + \det \begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix} \\
&= ab \underbrace{\det \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}}_{=0} + ad \underbrace{\det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{=\det I_2 = 1} + bc \underbrace{\det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{=-\det I_2 = -1} + cd \underbrace{\det \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}}_{=0} = ad - bc.
\end{aligned}$$

The first equality just rewrites the two columns of our first matrix as sums of simpler vectors. The second and third equalities follow by extensive use of property (iii) in the theorem defining det.

A formula to remember:

$$\boxed{\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc}$$

3 Permutation matrices

We digress to discuss a particular class of square matrices whose determinants are easy to compute.

A *permutation matrix* is an $n \times n$ matrix whose entries are all 0 or 1, and which has exactly one nonzero entry in each row and in each column.

Let S_n be the set of $n \times n$ permutation matrices.

Example. The elements of S_2 are $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Example. The elements of S_3 are

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

Let R_n be the set of $n \times n$ matrices whose entries are all 0 or 1, and which have exactly one nonzero entry in each column (but possibly multiple nonzero entries in a given row).

Example. The elements of R_2 are $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and $\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$.

Note that $S_n \subset R_n$. The size of S_n is $n!$ while the size of R_n is n^n.

Lemma. If $X \in R_n$ but $X \notin S_n$ then $\det X = 0$.

Proof. In this case X must have two equal columns. □

Given $X \in S_n$, define $\mathrm{inv}(X)$ as the number of 2×2 submatrices of X equal to $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

To form a 2×2 submatrix of X, choose any two rows and any two columns, not necessarily adjacent, and then take the 4 entries in those rows and columns.

Equivalently, $\mathrm{inv}(X)$ is the number of pairs of 1s in X with one 1 below and to the left of the other.

For example,

$$\mathrm{inv}\left(\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\right) = 2, \qquad \mathrm{inv}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = 0, \qquad \mathrm{inv}\left(\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\right) = 3.$$

Lemma. If $X \in S_n$ then $\det X = (-1)^{\mathrm{inv}(X)}$.

Proof. If $X \in S_n$ and $\mathrm{inv}(X) > 0$, then X must have two adjacent columns where the 1 in the left column is below the 1 in the right column. Form Y by interchanging these two columns. One can check (try drawing a picture of the matrices X and Y) that $\mathrm{inv}(Y) = \mathrm{inv}(X) - 1$. We know that $\det Y = -\det X$.

If $\mathrm{inv}(Y) > 0$, thes construct a permutation matrix Z from Y in the same way. Continuing this process will eventually give a permutation matrix $A \in S_n$ with $\det X = (-1)^{\mathrm{inv}(X)} \det A$ and $\mathrm{inv}(A) = 0$. But the only permutation matrix $A \in S_n$ with $\mathrm{inv}(A) = 0$ is $A = I_n$, so $\det(A) = 1$ and $\det(X) = (-1)^{\mathrm{inv}(X)}$. □

4 A formula for $\det A$

Given a matrix $X \in R_n$ and an arbitrary $n \times n$ matrix A, define

$$\Pi(X, A) = \text{ the product of the entries of } A \text{ in the nonzero positions of } X.$$

For example,

$$\Pi\left(\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = cdh.$$

We can now give the first *concrete* (though still complicated) description of the determinant.

Theorem. Suppose A is an $n \times n$ matrix. Then

$$\boxed{\det A = \sum_{X \in S_n} \Pi(X, A)(-1)^{\mathrm{inv}(X)}}$$

where the notation $\sum_{X \in S_n}$ means "compute $\Pi(X, A)(-1)^{\mathrm{inv}(X)}$ for each $n \times n$ permutation matrix X and then take the sum of all of the resulting numbers."

The function given by this formula has the defining properties of the determinant. This confirms our first theorem: the only function with the properties we originally ascribed to the determinant is this formula.

This theorem subsumes our first theorem. Before proving it, let's do an example.

Example. We can use the general formula for $\det A$ to compute the determinant of a 3×3 matrix. Suppose $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. Then our formula becomes

$$\begin{aligned} \det A = \quad & \Pi\left(\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, A\right)(-1)^0 + \Pi\left(\begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix}, A\right)(-1)^1 + \\ & \Pi\left(\begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix}, A\right)(-1)^1 + \Pi\left(\begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix}, A\right)(-1)^2 + \\ & \Pi\left(\begin{bmatrix} & & 1 \\ 1 & & \\ & 1 & \end{bmatrix}, A\right)(-1)^2 + \Pi\left(\begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix}, A\right)(-1)^3 = aei - afh - bdi + bfg + cdh - ceg. \end{aligned}$$

The 0s are omitted in the permutation matrices to improve readability. We can rewrite this as

$$\boxed{\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a(ei - fh) - b(di - fg) + c(dh - eg)}$$

Note that each term in parentheses is the determinant of a 2×2 submatrix of A.

Next time we will see that this type of formula can be generalised to higher dimensions.

Proof of theorem. The most difficult part of the proof is our notation, which gets fairly complicated.

Suppose $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$. We can then also write

$$A = \begin{bmatrix} \sum_{i=1}^n a_{i1} e_i & \sum_{j=1}^n a_{j2} e_j & \dots & \sum_{k=1}^n a_{kn} e_k \end{bmatrix}.$$

In words: express each column of A as the linear combination of the basis vectors $e_1, e_2, \dots, e_n$ of $\mathbb{R}^n$.

Using the fact that the determinant is linear as a function of each column of A, it follows that

$$\begin{aligned} \det A &= \det \begin{bmatrix} \sum_{i=1}^n a_{i1} e_i & \sum_{j=1}^n a_{j2} e_j & \dots & \sum_{k=1}^n a_{kn} e_k \end{bmatrix} \\ &= \sum_{i=1}^n a_{i1} \cdot \det \begin{bmatrix} e_i & \sum_{j=1}^n a_{j2} e_j & \dots & \sum_{k=1}^n a_{kn} e_k \end{bmatrix} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{i1} a_{j2} \cdot \det \begin{bmatrix} e_i & e_j & \dots & \sum_{k=1}^n a_{kn} e_k \end{bmatrix} \\ &\vdots \\ &= \underbrace{\sum_{i=1}^n \sum_{j=1}^n \cdots \sum_{k=1}^n}_{n \text{ summations}} \underbrace{a_{i1} a_{j2} \cdots a_{kn}}_{=\Pi(X,A)} \det \underbrace{\begin{bmatrix} e_i & e_j & \dots & e_k \end{bmatrix}}_{\text{this is a matrix } X \in R_n}. \end{aligned}$$

If this sequence of equalities is confusing, try to see if the corresponding step in our calculation of $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ makes more sense. We are really just generalising that calculation from 2 to n dimensions.

Key observation: the matrix $\begin{bmatrix} e_i & e_j & \dots & e_k \end{bmatrix}$ varies over all elements of R_n as the indices $i, j, \dots, k$ vary in the summations $\sum_{i=1}^n \sum_{j=1}^n \cdots \sum_{k=1}^n$. This means we can rewrite the last formula as

$$\det A = \sum_{X \in R_n} \Pi(X, A) \det X.$$

If $X \in R_n$ then $\det X = (-1)^{\mathrm{inv}(X)}$ if $X \in S_n$ and otherwise $\det X = 0$. Therefore, we actually have

$$\det A = \sum_{X \in S_n} \Pi(X, A)(-1)^{\mathrm{inv}(X)}. \tag{*}$$

This formula was computed *under the assumption that a function* det *exists with the properties in our first theorem.* This means that if our first theorem is true, then the determinant must be given by the formula we just derived. The last thing we need to do is to check *that the function (*) actually has the properties we require for the determinant.*

This is not too hard, and mostly involves some exercises in algebra manipulating the expression (*):

1. We have $\det I_n = \sum_{X \in S_n} \Pi(X, I_n)(-1)^{\mathrm{inv}(X)} = 1$.

   Proof. This holds since $\Pi(X, I_n) = 0$ unless $X = I_n$ if $X \in S_n$. □

2. If we interchange two columns in A then $\det A$ changes by a factor of -1.

   Proof. Let $\tilde{X}$ be the matrix given by interchanging columns i and j in X. If $X \in S_n$ then $\tilde{X} \in S_n$ and $\mathrm{inv}(\tilde{X}) - \mathrm{inv}(X)$ is an odd number. (This is not obvious but can be shown by an elementary argument: try drawing a picture of X compared to $\tilde{X}$.) Hence $(-1)^{\mathrm{inv}(X)} = -(-1)^{\mathrm{inv}(\tilde{X})}$.

   If $X \in S_n$ then $\Pi(X, A) = \Pi(\tilde{X}, \tilde{A})$ for all matrices A. (Why?)

   Thus $\det A = \sum_{X \in S_n} \Pi(X, A)(-1)^{\mathrm{inv}(X)} = -\sum_{X \in S_n} \Pi(\tilde{X}, \tilde{A})(-1)^{\mathrm{inv}(\tilde{X})} = -\det \tilde{A}$. □

3. If we fix all but one column of A, then the formula (*) is linear as a function $\mathbb{R}^n \to \mathbb{R}$ of the remaining column.

   Proof. If column i of A is the vector $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix}$ and all other columns of A are fixed numbers, then the formula (*) reduces to a function

   $$\det A = c_1 x_1 + c_2 x_2 + \dots c_n x_n = \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{**}$$

   where $c_1, c_2, \dots, c_n$ are some numbers *depending only on the other columns of* A. To see why this is true, note that for each $X \in S_n$ the value of $\Pi(X, A)(-1)^{\mathrm{inv}(X)}$ is ± 1 times the product of n entries in A, only one of which occurs in column i.

   The formula (**) shows that as a function x, the determinant $\det A$ is linear. □

This confirms that (*) does have the properties we stated in our first theorem. □

The formula $\det A = \sum_{X \in S_n} \Pi(X, A)(-1)^{\mathrm{inv}(A)}$ is not an efficient way of computing the determinant of most matrices since the sum involves a huge number of terms if n is large. There are 2 terms for $n = 2$, 6 for $n = 3$, 24 terms for $n = 4$, and 120 terms for $n = 5$.

Next time: more properties of determinants and how to compute them efficiently.

5 Vocabulary

Keywords from today's lecture:

1. **Permutation matrix**.

   A square matrix P whose entries are each 0 or 1, that has exactly one nonzero entry equal to 1 in each row and each column.

   If P is an $n \times n$ permutation matrix and A is a matrix with n rows then PA is a matrix formed by rearranging ("permuting") the rows of A. If A is a matrix with n columns then AP is a matrix formed by rearranging the columns of A.

   Example: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \text{ or } \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$

2. **Determinant**.

   The unique function
   $$\det : \{n \times n \text{ matrices}\} \to \mathbb{R}$$
   with $\det I_n = 1$, with the property that interchanging two columns in an $n \times n$ matrix A reverses the sign of $\det A$, and with the property that if all but one column in an $n \times n$ matrix A are fixed, then $\det A$ is a linear function of the remaining column.

   Example: $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$