The purpose of this document is to give a more economical axiomatic characterization of the determinant (see properties (i) and (ii) below) and proofs of basic results: Leibniz formula, cofactor expansion, explicit matrix inverse, and Cramer's rule.

Remark. In this framework, it's hard to prove that $\det(AB) = \det(A)\det(B)$ without setting up more machinery. The nicest proof of that formula is by interpreting the determinant as the factor by which the matrix scales volumes. A corollary of that formula is that $\det(A) \neq 0$ if $\det(A)$ is invertible (use the equation $AA^{-1} = \mathrm{Id}_{n\times n}$). The converse of this corollary is part of the 'formula for matrix inverse.'

1. Determinants

Let Mat_n be the set of real $n \times n$ matrices. For notational convenience, we also think of Mat_n as the set of n-tuples of vectors in $\mathbb{R}^n$:

$$\mathrm{Mat}_n \simeq \Big\{(v_1, \dots, v_n) \text{ where } v_i \in \mathbb{R}^n \text{ for each } i\Big\}.$$

The n-tuple $(v_1, \dots, v_n)$ corresponds to the matrix whose columns are $v_1, \dots, v_n$.

Definition. A function $f : \mathrm{Mat}_n \to \mathbb{R}$ is *multilinear* if it is linear in each 'argument':

$$\begin{aligned} f(v_1, \dots, v_{i-1}, \lambda_1 v_i + \lambda_2 v_i', v_{i+1}, \dots, v_n) &= \lambda_1 f(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_n) \\ &\quad + \lambda_2 f(v_1, \dots, v_{i-1}, v_i', v_{i+1}, \dots, v_n), \end{aligned}$$

where $v_1, \dots, v_n$ and v_i' are vectors in $\mathbb{R}^n$, and $\lambda_1, \lambda_2 \in \mathbb{R}$.

Remark. If f were *linear*, we would instead have

$$\begin{aligned} f(v_1, \dots, v_{i-1}, \lambda_1 v_i + \lambda_2 v_i', v_{i+1}, \dots, v_n) &= f(v_1, \dots, v_{i-1}, \lambda_1 v_i, v_{i+1}, \dots, v_n) \\ &\quad + f(0, \dots, 0, \lambda_2 v_i', 0, \dots, 0). \end{aligned}$$

This is because of the following linear relation:

$$(v_1, \dots, v_{i-1}, \lambda_1 v_i + \lambda_2 v_i', v_{i+1}, \dots, v_n) = (v_1, \dots, v_{i-1}, \lambda_1 v_i, v_{i+1}, \dots, v_n) + (0, \dots, 0, \lambda_2 v_i', 0, \dots, 0).$$

So, linear functions and multilinear functions behave very differently.

A multilinear function is uniquely determined by its values on tuples $(v_1, \dots, v_n)$ for which each v_i is an elementary basis vector.[1] For example, when $n = 2$, here is how it works:

$$\begin{aligned} f\begin{pmatrix} a & b \\ c & d \end{pmatrix} &= a\, f\begin{pmatrix} 1 & b \\ 0 & d \end{pmatrix} + c\, f\begin{pmatrix} 0 & b \\ 1 & d \end{pmatrix} \\ &= a\left(b\, f\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + d\, f\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) + c\left(b\, f\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + d\, f\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}\right) \\ &= ab\, f\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + ad\, f\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + bc\, f\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + cd\, f\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \end{aligned}$$

The first equation follows from linearity in the first argument ($i = 1$), and the second equation follows from linearity in the second argument ($i = 2$). The values of f on the matrices $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$ can be arbitrarily chosen, and the value of f on a general matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is determined by the formula above. To make sense of this formula, notice that, in $f\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, the 1's appear in the positions formerly occupied by a and b, which is why the product ab appears in front of this term, and similarly for the other terms.

[1] This means a vector with one entry 1 and all remaining entries 0. In the homework, this was called a 'one-hot vector.'

Similar reasoning applies for $n > 2$. For example, when $n = 3$, we have

$$f\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{12}a_{13}\, f\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + a_{11}a_{12}a_{23}\, f\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + a_{11}a_{12}a_{33}\, f\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (\text{other terms}).$$

There are 27 terms in total, corresponding to the 27 possible ways of putting a 1 in each of the three columns. The factors $a_{??}a_{??}a_{??}$ in front of each term correspond to the positions of the 1's.

Definition. The *determinant* is the unique multilinear function $\det : \mathrm{Mat}_n \to \mathbb{R}$ satisfying the following:

(i) $\det(\mathrm{Id}_{n\times n}) = 1$

(ii) For any i, we have $\det(v_1, \ldots, v_n) = 0$ if $v_i = v_{i+1}$.

Proposition. Multilinearity and axiom (ii) imply the following:

(iii) **(Duplicate column $\Rightarrow$ det = 0)** For any $i \neq j$, we have $\det(v_1, \ldots, v_n) = 0$ if $v_i = v_j$.

(iv) **(Column operations preserve determinant)** For any $i \neq j$, and any $\lambda \in \mathbb{R}$, we have

$$\det(\ldots, v_i, \ldots, v_j + \lambda v_i, \ldots) = \det(\ldots, v_i, \ldots, v_j, \ldots).$$

We have written the i-th and j-th arguments only, and all other arguments remain the same.

(v) **(Column swaps change the sign)** For any $i \neq j$, we have

$$\det(\ldots, v_j, \ldots, v_i, \ldots) = -\det(\ldots, v_i, \ldots, v_j, \ldots).$$

We have written the i-th and j-th arguments only, and all other arguments remain the same.

Proof. Taking $v_i = v_{i+1} = v + w$ for arbitrary vectors $v, w \in \mathbb{R}^n$, we find that

$$\det(\ldots, v + w, v + w, \ldots) = 0$$

by (ii). (We have written the i-th and $(i + 1)$-st arguments only.) On the other hand, multilinearity implies

$$\begin{aligned} \det(\ldots, v+w, v+w, \ldots) &= \det(\ldots, v, v+w, \ldots) + \det(\ldots, w, v+w, \ldots) \\ &= \det(\ldots, v, v, \ldots) + \det(\ldots, v, w, \ldots) \\ &\quad + \det(\ldots, w, v, \ldots) + \det(\ldots, w, w, \ldots) \\ &= \det(\ldots, v, w, \ldots) + \det(\ldots, w, v, \ldots). \end{aligned}$$

The first equation uses linearity in the i-th argument, the second equation uses linearity in the $(i + 1)$-st argument, and the third equation follows from (ii). Thus, we conclude that

$$\det(\ldots, v, w, \ldots) = -\det(\ldots, w, v, \ldots),$$

i.e. swapping two adjacent arguments changes the determinant by a sign.

This implies (v), because one can swap v_i and v_j by doing $2|j - i| - 1$ swaps of adjacent arguments. If $j > i$, then first exchange v_i and v_{i+1}, then exchange v_i with v_{i+2}, and keep going.

Next, (ii) and (v) imply (iii). Indeed, if $v_i = v_j$, then we can swap v_j with v_{i+1} (possibly changing the sign of the determinant), at which point we have two adjacent arguments which are equal. Then (ii) implies that the value of this determinant is zero.

Finally, (iv) follows from (iii):

$$\begin{aligned} \det(\ldots, v_i, \ldots, v_j + \lambda v_i, \ldots) &= \det(\ldots, v_i, \ldots, v_j, \ldots) + \lambda \det(\ldots, v_i, \ldots, v_i, \ldots) \\ &= \det(\ldots, v_i, \ldots, v_j, \ldots). \end{aligned}$$

The first equation uses linearity in the j-th argument, and the second equation uses (iii) to conclude that $\det(\ldots, v_i, \ldots, v_i, \ldots) = 0$, because the i-th and j-th arguments are equal. $\square$

With these properties, we can turn our discussion about general multilinear functions into a formula for the determinant. For example, if $n = 3$, then let's figure out the value of two of the 27 terms from before:

- Consider the term $a_{31}a_{22}a_{23}\, f\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$. When $f = \det$, this term equals zero because the matrix has two identical columns.
- Consider the term $a_{31}a_{22}a_{13}\, f\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$. When $f = \det$, we have
$$\begin{aligned} \det\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} &= -\det\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= -1. \end{aligned}$$
The first equation follows from (v) by swapping the first and third column, and the second equation follows from (i). Therefore, this term equals $-a_{31}a_{22}a_{13}$.

Applying this reasoning to all 27 terms, we conclude that

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{21}a_{12}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}.$$

Most of the terms are zero because they involve the determinant of a matrix with a duplicate column.

Note that each displayed term corresponds to a matrix P with exactly one 1 in each row and column. These are called *permutation matrices* because they specify permutations of the elements $\{1, \dots, n\}$, and $\det(P)$ is called the *sign* of the permutation. As we saw above, it equals $+1$ or -1 depending on the (parity of the) number of swaps required to go from P to $\mathrm{Id}_{n\times n}$. For example:

term	$a_{11}a_{22}a_{33}$	$-a_{11}a_{23}a_{32}$	$-a_{21}a_{12}a_{33}$	$+a_{21}a_{32}a_{13}$	$+a_{31}a_{12}a_{23}$	$-a_{31}a_{22}a_{13}$
permutation matrix	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$
sign of permutation	$+1$	-1	-1	$+1$	$+1$	-1
swaps to go to $\mathrm{Id}_{3\times 3}$	0	1	1	2	2	1

Corollary. For any matrix A, we have $\det(A) = \det(A^\top)$. Thus, $\det(-)$ also satisfies analogues of (iii), (iv), (v) for the rows instead of the columns.

Proof. In view of the formula for $\det(-)$ derived above, we only need to show that $\det(P) = \det(P^\top)$ for any permutation matrix P. As remarked above, $\det(P)$ is determined by the number of column swaps required to go from P to $\mathrm{Id}_{n\times n}$. For a permutation matrix, any column swap is equivalent to a row swap operation. Therefore, the number of column swaps equals the number of row swaps required to go from P to $\mathrm{Id}_{n\times n}$, and this computes $\det(P^\top)$. □

2. Cofactors

Let A be an $n \times n$ matrix.

Definition. For any $i, j \in \{1, \dots, n\}$, let $A^{(i,j)}$ be the matrix obtained by deleting the i-th row and j-th column of A. Let $C_{ij} := (-1)^{i+j}\det(A^{(i,j)})$. This is called the (i,j)-*cofactor of* A.

Proposition. **(Cofactor expansion for determinants)** For any j, we have

$$\det(A) = A_{1j}C_{1j} + A_{2j}C_{2j} + \cdots + A_{nj}C_{nj}.$$

This is called 'cofactor expansion along the j-th column.'

Proof. In the previous section, we saw that

$$\det(A) = \sum_P \det(P)\,(\text{product of the } n \text{ entries of } A \text{ in the locations where } P \text{ has entry } 1).$$

The sum is defined by letting P be any $n \times n$ permutation matrix (there are $n!$ of them) and adding the resulting terms together.

For each $i \in \{1, \ldots, n\}$, let's restrict attention to those P's for which $P_{ij} = 1$.

- Let $A^{(i,j)}$ and $P^{(i,j)}$ be obtained from A and P by deleting the i-th row and j-th column. Then the corresponding term equals

$$\underbrace{(-1)^{i+j}\det(P^{(i,j)})}_{\det(P)}\,\underbrace{(A_{ij})\,(\text{product of the } (n-1) \text{ entries of } A^{(i,j)} \text{ in the locations where } P^{(i,j)} \text{ has entry } 1)}_{(\text{product of the } n \text{ entries of } A \text{ in the locations where } P \text{ has entry } 1)}.$$

- As P ranges over all permutation matrices such that $P_{ij} = 1$, the matrix $P^{(i,j)}$ ranges over all $(n-1) \times (n-1)$ permutation matrices. Therefore, summing the term

$$(-1)^{i+j}\det(P^{(i,j)})(A_{ij})(\text{product of the } (n-1) \text{ entries of } A^{(i,j)} \text{ in the locations where } P^{(i,j)} \text{ has entry } 1)$$

  over all such P yields $(-1)^{i+j}A_{ij}\det(A^{(i,j)}) = A_{ij}C_{ij}$.

In the previous paragraph, we restricted to permutation matrices P for which $P_{ij} = 1$. In general, the 1 in the j-th column of P can appear in any row i, where $i \in \{1, \ldots, n\}$. Summing over all these possibilities for i yields the desired result:

$$\det(A) = A_{1j}C_{1j} + A_{2j}C_{2j} + \cdots + A_{nj}C_{nj}.$$

□

Corollary. **(Formula for matrix inverse)** Let C be the matrix whose (i,j)-th entry is C_{ij}. If $\det(A) \neq 0$, then A is invertible, with

$$A^{-1} = \frac{1}{\det(A)}\,C^{\top}.$$

The matrix $C^{\top}$ is called the *adjugate matrix* of A.

Proof. It suffices to show that $C^{\top}A = \det(A)\,\mathrm{Id}_{n\times n}$. For $j \in \{1, \ldots, n\}$, the (j,j)-th entry of $C^{\top}A$ is

$$A_{1j}C_{1j} + A_{2j}C_{2j} + \cdots + A_{nj}C_{nj},$$

which equals $\det(A)$ by the previous Proposition. Similarly, for $i \neq j$, the (i,j)-th entry of $C^{\top}A$ is

$$A_{1j}C_{1i} + A_{2j}C_{2i} + \cdots + A_{nj}C_{ni}.$$

If A' is obtained by replacing the i-th column of A by the entries of the j-th column of A, then the displayed sum is the formula for $\det(A')$ obtained by cofactor expansion along the i-th column. On the other hand, since A' has a duplicate column, $\det(A') = 0$ by (iii). Therefore, the (i,j)-th entry of $C^{\top}A$ is zero. □

Corollary. **(Cramer's rule)** If A is invertible, then the unique solution to

$$A\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = b$$

is given by $x_i = \frac{\det(A^{(i),b})}{\det(A)}$, where $A^{(i),b}$ is the matrix obtained by replacing the i-th column of A by the entries of b.

Proof. By the formula for matrix inverse, we have $x = \frac{1}{\det(A)} C^\top b$, so

$$x_i = \frac{1}{\det(A)} \underbrace{\left(b_1 C_{1i} + b_2 C_{2i} + \cdots + b_n C_{ni}\right)}_{\det(A^{(i),b})}.$$

To see that the number in the brace does in fact equal $\det(A^{(i),b})$, use cofactor expansion along the i-th column of the matrix $A^{(i),b}$. $\square$