STANDARD FORMULA FOR DETERMINANTS

This is a short writeup concerning determinants. Recall that we proved the following result.

Theorem 0.1. *Let A be an $n \times n$ matrix and $f(A)$ a function with the properties that $f(I) = 1$, $f(A)$ is linear in the first row and $f(A)$ changes sign under exchange of any two rows. If $g(A)$ is any function satisfying the same properties then $g(A) = f(A)$. This unique function, if it exists, is called the determinant and we denote it by* $\det A$.

Proof. The function $f(A)$ changes sign under arbitrary row swaps and this implies that it is linear in each row separately. Thus the function $f(A)$ does not change if any multiple of a row is subtracted from another row. It changes sign under row swaps. Hence we can bring the matrix A into diagonal form using by a sequence of subtraction of multiples of rows from other rows and a number of swaps, say K of them, i.e.,

$$A \to \begin{bmatrix} P_1 & 0 & . & . & 0 \\ 0 & P_2 & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & . & P_n \end{bmatrix}$$

and hence $f(A) = (-1)^K P_1 \cdots P_n$. Exactly the same moves show that $g(A) = (-1)^K P_1 \cdots P_n$ and hence $f(A) = g(A)$.

$\square$

This theorem makes an interesting point concerning the logic of the whole thing. What it says is that if the function *exists* then it is unique. The row reduction algorithm does not prove that the function exists because there are many ways how this can be carried out. The problem is not with the pivots $P_1, \dots, P_n$. They are always the same no matter how the row reduction is carried out. The problem is with the swaps. Alice might effect the row reduction using 4 swaps and Bob uses 5. Thus Alice would get $P_1 \cdots P_n$ but Bob would get $-P_1 \cdots P_n$ as the value of the function $f(A)$. If this were to happen then there is no function that satisfies the three requirements listed in the above theorem.

Below we give a formula for the determinant, (1). This formula is not suitable for numerical computations; it is a sum of $n!$ terms! The uses are mostly theoretical. Most of you know about eigenvalues and maybe remember that the product of the eigenvalues of a matrix equals the determinant or the sum of the eigenvalues of the matrix equals the trace of the matrix. It is these kind of insights you can get from formula (1). Another piece of insight are co-factors, a generalization of the cross product, which allow to compute $n-1$ dimensional volumes of parallelepipeds in n dimensions. First we have to explain a few facts about permutations.

1. Permutations

Imagine a set of n distinct objects. In order to distinguish them we give each one of them a label. A permutation is now a relabeling. Alternatively, consider the set $S = \{1, 2, \dots, n\}$. A permutation $\pi : S \to S$ is a function which one to one and hence onto (why?). We can express permutations in the following way

$$\begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ \pi(1) & \pi(2) & \dots & \pi(n-1) & \pi(n) \end{pmatrix}$$

E.g.

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$$

corresponds to the function $\pi(1) = 2, \pi(2) = 4, \pi(3) = 1$ and $\pi(4) = 3$. The set of permutations of n objects is called the symmetric group and is denoted by $\mathcal{S}_n$. Two permutations of n objects, π and σ can be composed

$$\pi \circ \sigma(i) = \pi(\sigma(i)) \ , i = 1, \dots, n \ .$$

E.g. composing the permutation π given above with the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$$

yields the permutation

$$\pi \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

which is the identity permutation. Thus, π is the inverse function, i.e., inverse permutation of σ. Thus, the composition of any two permutation is again a permutation and every permutation has an inverse permutation. This is why $\mathcal{S}_N$ is called a group.

Let $\pi \in \mathcal{S}_n$. Associated with this is an $n \times n$ matrix P_π which is the matrix that emerges from the identity by permuting the rows using the permutation π. For the permutation σ given above we have that

$$P_\sigma = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

also π in the example given above

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

Another way of describing P_π is to say that $P_\pi e_j = e_{\pi(j)}, j = 1, \dots, n$. One easily checks this also in the above example. The map $\pi \to P_\pi$ respects composition, i.e.,

$$P_{\pi \circ \sigma} = P_\pi P_\sigma \ ,$$

because

$$P_{\pi \circ \sigma} e_j = e_{\pi \circ \sigma(j)} = P_\pi e_{\sigma(j)} = P_\pi P_\sigma e_j \ .$$

Thus one easily checks that in the examples above $P_\pi P_\sigma = I_4$.

Here is an interesting theorem whose proof you can find at the end of this document. It is somewhat tricky.

Theorem 1.1. *Let $\pi \in \mathcal{S}_n$ and consider the corresponding permutation matrix P_π. Then the number of row swaps needed to bring P_π back to the identity matrix is either always even or always odd. If $P\pi$ can be brought back to I in an even number of swaps are called* **even** *permutations and the ones that need an odd number of swaps are called* **odd**. *Hence if π is even the* $\det P_\pi = 1$ *and if π is odd then* $\det P_\pi = -1$.

As we shall see, this innocuous theorem settles the question of existence. By the way in the literature 'swaps' are called transpositions. Now, that we know how to to compute the determinant of a permutation matrix the following theorem settles the question of existence.

Theorem 1.2. *Consider an $n \times n$ matrix A real or complex and denote the matrix elements by a_{ij} where both i and j vary between* 1 *and n. Then*

$$\det A = \sum_{\pi \in \mathcal{S}_n} (\det P_\pi) a_{1\pi(1)} \cdots a_{n\pi(n)} \tag{1}$$

Proof. In order to show this we call the right side of this equation $f(A)$. If $A = I$ we see that the only non-zero product appears when π is the identity permutation and hence $f(I) = 1$ since $\det I = 1$. Let A' be the matrix derived from A by exchanging any two rows. . Fix any i and j and consider the permutation σ given by $\sigma(i) = j$ and $\sigma(j) = i$ with all the other elements kept fixed. The matrix A' denotes the matrix with elements $a'_{k\ell} = a_{\sigma(k)\ell}$. It has the same rows as A except that the ith and jth are switched. We have to show that $f(A') = -f(A)$

To see this we observe that

$$a'_{1\pi(1)} a'_{2\pi(2)} \cdots a'_{n\pi(n)} = a_{\sigma(1)\pi(1)} a_{\sigma(2)\pi(2)} \cdots a_{\sigma(n)\pi(n)}$$

which we may rewrite as

$$a_{1\pi\circ\sigma^{-1}(1)} a_{2\pi\circ\sigma^{-1}(2)} \cdots a_{n\pi\circ\sigma^{-1}(n)}$$

Thus, we may write

$$a'_{1\pi(1)} a'_{2\pi(2)} \cdots a'_{n\pi(n)} = a_{1\pi\circ\sigma^{-1}(1)} a_{2\pi\circ\sigma^{-1}(2)} \cdots a_{n\pi\circ\sigma^{-1}(n)} \ .$$

Further

$$P_{\pi\circ\sigma^{-1}} = P_\pi P_{\sigma^{-1}}$$

and so

$$\det P_{\pi\circ\sigma^{-1}} = \det P_\pi \det P_{\sigma^{-1}} = -\det P_\pi \ ,$$

because any permutation matrix that switches two rows and keeps the others fixed has determinant -1. Thus

$$f(A') = -\sum_{\pi \in \mathcal{S}_n} \det P_{\pi\circ\sigma^{-1}} a_{1\pi\circ\sigma^{-1}(1)} \cdots a_{n\pi\circ\sigma^{-1}(n)}$$

and since $\pi \circ \sigma^{-1}$ runs through all the permutations precisely once we have that

$$f(A') = -\sum_{\pi \in \mathcal{S}_n} \det P_\pi a_{1\pi(1)} \cdots a_{n\pi(n)} = -f(A) \ .$$

Let us remark that the permuation σ is its own inverse and so we could have replaced σ^{-1} everywhere by σ. We also have to show that $f(A)$ is linear in the first row. Denote by A the matrix that has the first row αa_1^T and A' is the matrix with the first row βb^T and all the other rows of A and A' are the same. Consider now the matrix B that has the first row $\alpha a^T + \beta b^T$ but all the other rows are the same as the ones of A and hence A'. We find

$$f(B) = \sum_{\pi \in \mathcal{S}_n} b_{1\pi(1)} \cdots b_{n\pi(n)} \det P_\pi = \sum_{\pi \in \mathcal{S}_n} [\alpha a_{1\pi(1)} + \beta b_{1\pi(1)}] a_{2\pi(2)} \cdots a_{n\pi(n)} \det P_\pi$$

which equals

$$\alpha \sum_{\pi \in \mathcal{S}_n} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} \det P_\pi + \beta \sum_{\pi \in \mathcal{S}_n} \beta b_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} \det P_\pi = \alpha f(A) + \beta f(A') \ .$$

Hence by the uniqueness of any function that has the tree properties above, we find $f(A) = \det A$. $\square$

As an example, let us compute the determinant of a general 3×3 matrix. For this one needs a list of the permutations of three objects. There are 6 such permutations.

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} , \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} , \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} , \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} , \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} , \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} .$$

The determinant of the corresponding permutation matrices are, in the same order, $1, -1 - 1, -1, 1, 1$. Hence we have

$$\det A = a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \ .$$

2. Co-factors

We can rewrite the formula of the determinant of the 3×3 matrix in the interesting way

$$\det \mathrm{A} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \ .$$

The numbers

$$C_{11} = a_{22}a_{33} - a_{23}a_{32}, C_{12} = (a_{23}a_{31} - a_{21}a_{33}), C_{13} = (a_{21}a_{32} - a_{22}a_{31})$$

are called co-factors. The vector of these co-factors is the cross product of the second row with the third row, i.e.,

$$\begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} \times \begin{bmatrix} a_{31} \\ a_{32} \\ a_{33} \end{bmatrix} = \begin{bmatrix} a_{22}a_{33} - a_{23}a_{32} \\ a_{23}a_{31} - a_{33}a_{21} \\ a_{21}a_{32} - a_{31}a_{22} \end{bmatrix} .$$

Recall that this vector is perpendicular to the second and third row and its length is the area of the parallelogram spanned by these two vectors.

In general we can write the determinant of any square matrix A as

$$\det A = a_{11}C_{11} + \cdots a_{1n}C_{1n} \ . \tag{2}$$

This follows from the formula (1). These co-factors can be easily computed in terms of determinants of smaller matrices. The term in the determinant that is proportional to a_{11} is given by summing over all permutations π with $\pi(1) = 1$,

$$\sum_{\pi \in \mathcal{S}_n, \pi(1)=1} (\det P_\pi) a_{2\pi(2)} \cdots a_{n\pi(n)} = \sum_{\pi \in \mathcal{S}_{n-1}} (\det P_\pi) a_{2\pi(2)} \cdots a_{n\pi(n)}$$

which we recognize as the determinant of the $(n-1) \times (n-1)$ matrix M_{11} that is obtained by removing the first column and the first row. Thus

$$C_{11} = \det M_{11} \ .$$

Instead of concentrating on the first row we could use any other, say the ith row. The we write

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij}$$

and we call C_{ij} the ij co-factor. How do we compute these co-factors? Returning to (2) let us compute C_{12}. This is easy if we realize that by switching the first and second column we can reduce the problem to the previous case C_{11} albeit with a different matrix A' in which the first and second column are swapped. Thus

$$\det A' = a_{12} C'_{12} + a_{11} C'_{11} + a_{13} C_{13} + \cdots + a_{1n} C'_{1n}$$

and we have that

$$C'_{12} = \det M_{12}$$

where M_{12} is the $(n-1) \times (n-1)$ matrix that we get by removing the first row and second column from the matrix A. Since $\det A' = -\det A$ we have that

$$\det A = -(a_{12} C'_{12} + a_{11} C'_{11} + a_{13} C_{13} + \cdots + a_{1n} C'_{1n})$$

and thus

$$C_{12} = -\det M_{12} \ .$$

In general we have

Lemma 2.1.

$$C_{ij} = (-1)^{i+j} \det M_{ij}$$

where the $(n-1) \times (n-1)$ matrix M_{ij} is given by removing the ith rwo and the jth column from the matrix A.

Proof. With i row swaps we get the ith row to the top and j column swaps we get the jth column to be the first. Note, and this is important, these swaps *do not change the order of the other rows and columns.* This yields the formula. □

The point about all this is that *co-factors form the generalization of the cross product to higher dimensions.* We have

Theorem 2.2. *let A be an $n \times n$ matrix and $C_{1j}, j = 1, \dots n$ the $1j$ co-factors, i.e.,*

$$\det A = \sum_{j=1}^{n} a_{1j} C_{1j} \ .$$

Then

$$\sum_{j=1}^{n} a_{kj} C_{1j} = 0 \ , k = 2, 3, \dots, n$$

and

$$\sqrt{\sum_{j=1}^{n} C_{1j}^2}$$

is the $(n-1)$ dimensional volume of the parallelepiped spanned by rows $2, 3, \dots, n$.

Proof. The equation

$$\sum_{j=1}^{n} a_{kj} C_{1j} = 0 \ , k = 2, 3, \dots, n$$

follows from the fact that

$$\sum_{j=1}^{n} a_{kj} C_{1j}$$

is the determinant of the matrix where the kth row appears also in the first row and hence is zero. Consider the n dimensional parallelepiped spanned by the rows $1, \dots, n$. The volume is given by $|\det A|$ This parallelepiped can be viewed as one with a base given by the $n-1$ dimesnional parallelepiped spanned by the rows $2, \dots, n$. Since the vector of cofactors C_j is perpedicular to all the vectors that span the base we compute the height of the parallelepiped as the projection of the first row onto the unit vector in the direction of the vector of co-factors, i.e.,

$$h = \left| \frac{\sum_{j=1}^{n} a_{1j} C_{1j}}{|C|} \right| = \frac{|\det A|}{|C|} \text{ or } |C| h = |\det A|$$

where

$$|C| = \sqrt{\sum_{j=1}^{n} C_{1j}^2} \ .$$

Since the volume is given as the $n-1$ dimensional volume of the base times the height we obtain the stated result ∎

3. Cramer's rule

The rule is named after the French mathematician Gabriel Cramer (1704-1752). You know how to invert a 2×2 matrix. If

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

then we find the inverse as

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \ .$$

So far this is just a formula and to get some better insight let us rewrite this as the equations

$$\det A = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix} \begin{bmatrix} a_{22} \\ -a_{21} \end{bmatrix} \tag{3}$$

$$\det A = \begin{bmatrix} a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} -a_{12} \\ a_{11} \end{bmatrix} \tag{4}$$

$$0 = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix} \begin{bmatrix} -a_{12} \\ a_{11} \end{bmatrix} \tag{5}$$

$$0 = \begin{bmatrix} a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} a_{22} \\ -a_{2}1 \end{bmatrix} \tag{6}$$

$$\tag{7}$$

and we can identify right away the cofactors $C_{11} = a_{22}, C_{12} = -a_{21}, C_{21} = -a_{12}, C_{22} = a_{11}$. It is precisely these equations that we can write down in arbitrary dimensions using the co-factors.

$$\mathrm{detA} = \sum_{j=1}^{n} a_{kj} C_{kj} \ , \ \ k = 1, \ldots, n$$

and

$$0 = \sum_{j=1}^{n} a_{\ell j} C_{kj} \ , \ell \neq k \ .$$

Hence the inverse of A is given by

$$A^{-1} = \frac{1}{\mathrm{det} A} C^T$$

where the matrix C has the co-factors as matrix elements. While this is quite nice, it is not useful for computations for the same reasons as the determinant formula is not usfeul for this purpose.

4. Characteristic polynomial

From formula (1) we can glean some interesting facts about the characteristic polynomial concerning the trace and the determinant of a matrix. let A be an $n \times n$ matrix and consider its characteristic polynomial

$$\det(A - \lambda I) = \sum_{\pi \in \mathcal{S}_n} \mathrm{det} P_\pi (a_{1\pi(1)} - \lambda \delta_{1\pi(1)}) \cdots (a_{n\pi(n)} - \lambda \delta_{n\pi(n)}) \ .$$

Expanding this expression in powers of λ we see that it is first of all a polynomial of degree n and the coefficient of the highest power is $(-1)^n$. If we set $\lambda = 0$ we recover the term independent of λ which is $\mathrm{det} A$. More interesting is the term proportional to λ^{n-1} which is given by

$$-\lambda^{n-1} \sum_{\pi \in \mathcal{S}_n} \mathrm{det} P_\pi \sum_{i=1}^{n} \delta_{1\pi(1)} \cdots \delta_{(i-1)\pi(i-1)} a_{i\pi(i)} \delta_{(i+1)\pi(i+1)} \cdots \delta_{n\pi(n)} \ .$$

This term is non-zero only if π is the permutation that fixes $n-1$ indices and hence it must fix all of them, i.e., it is the identity permutation. Thus, this term equals

$$-\lambda^{n-1} \sum_{i=1}^{n} a_{ii} = -\lambda^{n-1} \mathrm{Tr} A \ .$$

Recall that the characteristic polynomial can be factored

$$\det(A - \lambda I) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

and by expanding we get that the constant term, i.e., the term independent of λ is given by $\lambda_1 \cdots \lambda_n$ and the term proportional to λ^{n-1} is given by

$$-\lambda^{n-1} \sum_{i=1}^{n} \lambda_i \ .$$

Thus we recover the identities

$$\lambda_1 \cdots \lambda_n = \mathrm{det} A$$

and

$$\sum_{i=1}^{n} \lambda_i = \mathrm{Tr} A \ .$$

This can be pushed further. Consider the term proportional to λ^{n-2}. For this we get

$$(-\lambda)^{n-2} \sum_{\pi \in \mathcal{S}_n} \det P_\pi \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{1\pi(1)} \cdots \delta_{i-1\pi(i-1)} a_{i\pi(i)} \delta_{i+1\pi(i+1)} \ldots \delta_{j-1\pi(j-1)} a_{j\pi(j)} \delta_{j+1\pi(j+1)} \ldots \delta_{n\pi n}$$

interchanging sums yields

$$(-\lambda)^{n-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\pi \in \mathcal{S}_n} \det P_\pi \delta_{1\pi(1)} \cdots \delta_{i-1\pi(i-1)} a_{i\pi(i)} \delta_{i+1\pi(i+1)} \ldots \delta_{j-1\pi(j-1)} a_{j\pi(j)} \delta_{j+1\pi(j+1)} \ldots \delta_{n\pi n} \ .$$

For i, j fixed the only permutations that yields nonzero terms must fix all elements except i and j. Hence the only ones are either the one with $\pi(i) = i, \pi(j) = j$, i.e. the identity permutation or the one with $\pi(i) = j, \pi(j) = i$. Hence we get that this term is

$$(-\lambda)^{n-2} \sum_{i=1}^{n} \sum_{j=1}^{n} [a_{ii} a_{jj} - a_{ij} a_{ji}] \ .$$

On the other hand the corresponding term in $(-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$ is

$$(-\lambda)^{n-2} \sum i \neq j \lambda_i \lambda_j$$

and hence we get that

$$\sum_{i \neq j} \lambda_i \lambda_j = \sum_{i=1}^{n} \sum_{j=1}^{n} [a_{ii} a_{jj} - a_{ij} a_{ji}] \ .$$

This last expression has a simple interpretation. Pick all the 2×2 diagonal sub-matrices of A, compute their determinants and sum them up. This yields $\sum_{i \neq j} \lambda_i \lambda_j$.

5. Proof of Theorem 1.1 (after van der Waerden)

Consider the polynomial in n variables $x_1, x_2, \ldots, x_n$ given by

$$\Delta(x_1, x_2, \ldots, x_n) = \Pi_{i<j} (x_i - x_j) \ .$$

If we act with a permutation π on the variables we get $\Delta(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$ which is again a polynomial in the same variables and with the same factors except for the signs. Hence there exists a function sign : $\mathcal{S}_n \to \{+1, -1\}$ such that $\Delta(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) = sign(\pi) \Delta(x_1, x_2, \ldots, x_n)$. Recall that $\mathcal{S}_n$ is the set of permutations. A permutation π is called even when $sign(\pi) = 1$ and odd if $sign(\pi) = -1$. An interesting exercise it to show that for two permutations π and σ we have that $sign(\pi \circ \sigma) = sign(\pi) sign(\sigma)$. If π is a transposition, i.e., a swap, then $sign(\pi) = -1$. To see this write

$$\Pi_{i<j} (x_i - x_j) = (x_1 - x_2) \Pi_{k=3}^{n} (x_1 - x_k) \Pi_{k=3}^{n} (x_2 - x_k) \Pi_{3 \leq i < j \leq n} (x_i - x_j)$$

from which we see that if we swap x_1 and x_2 but leave the others fixed then the expression changes the sign. Now if π is an even permutation which is realized by N swaps, then $1 = (-1)^N$ and N is even. Likewise, if N is an odd permutation one needs always an odd number of swaps.