Determinants in detail

Kyle Miller

September 27, 2016

The book's introduction to the determinant $\det A$ of an $n \times n$ square matrix A is to say there is a quantity which determines exactly when A is invertible, followed by a "definition" of cofactor expansion along the first row. The book follows this up with the claim that cofactor expansion along any row or column is equal, since it would avoid a "lengthy digression." These definitions are opaque, not giving any insight, and it is possible the authors decided that these might as well be the definition since cofactor expansion is a reasonable way to calculate determinants by hand for small matrices. Let us try to do better!

While it is true that the digression is somewhat lengthy, many things in chapters 3.1 and 3.2 are a part of it anyway, so the claim is somewhat indistinguishable from the authors giving up explaining the theory simply.

I cannot promise that I will succeed, but I hope to show as clearly as I can that the determinant, and by extension the cofactor expansion, is the consequence of three reasonable properties.

1 Multilinear transformations

We have seen linear transformations already, which are functions $T : \mathbb{R}^n \to \mathbb{R}^m$ with two properties: $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$ and $T(c\vec{u}) = cT(\vec{u})$ for all $\vec{u}, \vec{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

A *multilinear transformation* is a function $(\mathbb{R}^n)^k \to \mathbb{R}^m$ which takes k vectors of $\mathbb{R}^n$ to produce a vector of $\mathbb{R}^m$, and which is linear in each input. That is:

1. $T(\dots, \vec{u} + \vec{v}, \dots) = T(\dots, \vec{u}, \dots) + T(\dots, \vec{v}, \dots)$
2. $T(\dots, c\vec{u}, \dots) = cT(\dots, \vec{u}, \dots)$

where everything in the $\dots$ remains the same in each equality.

For example, a multilinear transformation which takes two vectors has the following rule:

$$\begin{aligned} T(\vec{u}_1 + \vec{v}_1, \vec{u}_2 + \vec{v}_2) &= T(\vec{u}_1, \vec{u}_2 + \vec{v}_2) + T(\vec{u}_2, \vec{u}_2 + \vec{v}_2) \\ &= T(\vec{u}_1, \vec{u}_2) + T(\vec{u}_1, \vec{v}_2) + T(\vec{u}_2, \vec{u}_2) + T(\vec{u}_2, \vec{v}_2) \end{aligned}$$

This is sort of like saying that multilinear transformations are a multiplication rule for vectors (it expands like $(u_1 + u_2)(v_1 + v_2)$ would for numbers). Beware that $T(\vec{u}_1 + \vec{v}_1, \vec{u}_2 + \vec{v}_2)$ isn't necessarily equal to $T(\vec{u}_1, \vec{u}_2) + T(\vec{v}_1, \vec{v}_2)$!

A concrete example of a multilinear transformation is the dot product $\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$, which is $(\mathbb{R}^n)^2 \to \mathbb{R}$ (and will be introduced later in the course). Another concrete example is the cross product $\vec{u} \times \vec{v}$ which is $(\mathbb{R}^3)^2 \to \mathbb{R}$ (and will not be).

It is possible to show that the function which takes a pair of vectors in $\mathbb{R}^2$ and gives the area of the parallelogram between them is a multilinear transformation, but we won't do this here. For this to work out, swapping the vectors must give *negative* the area. This is because $T(\vec{x}, \vec{x}) = 0$ for any $\vec{x}$ (this is a degenerate parallelogram), and

$$\begin{aligned} 0 &= T(\vec{u} + \vec{v}, \vec{u} + \vec{v}) \\ &= T(\vec{u}, \vec{u}) + T(\vec{u}, \vec{v}) + T(\vec{v}, \vec{u}) + T(\vec{v}, \vec{v}) \\ &= T(\vec{u}, \vec{v}) + T(\vec{v}, \vec{u}) \end{aligned}$$

so $T(\vec{u}, \vec{v}) = -T(\vec{v}, \vec{u})$.

One thing you can do with a multilinear transformation is move scalars wherever you want For instance:

$$T(c\vec{u}, \vec{v}) = cT(\vec{u}, \vec{v}) = T(\vec{u}, c\vec{v})$$

Beware that $cT(\vec{u}, \vec{v})$ is not necessarily equal to $T(c\vec{u}, c\vec{v})$.

Another thing you can do with a multilinear transformations is come up with something like a matrix of the transformation, but matrices won't do anymore: it takes a tensor. (A matrix is a kind of tensor.) It ends up being fairly similar. Put in all possible sets of standard basis vectors to get a "column." As an example, for $T : (\mathbb{R}^n)^2 \to \mathbb{R}^m$, "column" $\vec{a}_{ij}$ is a vector of $\mathbb{R}^m$, where $\vec{a}_{ij} = T(\vec{e}_i, \vec{e}_j)$. Then, $T(\vec{u}, \vec{v}) = \sum_{i,j} u_i v_j \vec{a}_{ij}$ by multilinearity. The tensor is sort of like a matrix, but a rectangular prism of numbers rather than a rectangle. For sake of justifying this equation:

$$\begin{aligned} T(\vec{u}, \vec{v}) &= T(\sum_i u_i \vec{e}_i, \sum_j v_j \vec{e}_j) \\ &= \sum_i u_i T(\vec{e}_i, \sum_j v_j \vec{e}_j) \\ &= \sum_i \sum_j u_i v_j T(\vec{e}_i, \vec{e}_j) \\ &= \sum_{i,j} u_i v_j \vec{a}_{ij} \end{aligned}$$

2 Characterization of determinants

A determinant for $\mathbb{R}^n$ is a function $\det : (\mathbb{R}^n)^n \to \mathbb{R}$ which takes n vectors of $\mathbb{R}^n$ and produces a real number with the following properties:

1. det is multilinear.
2. $\det(\vec{e}_1, \dots, \vec{e}_n) = 1$. (That is, $\det I_n = 1$.)
3. $\det(\dots, \vec{u}, \dots, \vec{v}, \dots) = -\det(\dots, \vec{v}, \dots, \vec{u}, \dots)$. (That is, det is *antisymmetric*.)

What we will show is that these three properties characterize the determinant: anything which has these three properties is *the* determinant. We will start this by showing that these three properties are sufficient to compute the determinant of a matrix.

Using the standard basis of $\mathbb{R}^n$, we can do the following expansion (whose notation is hard to make nice, sorry!):

$$\begin{aligned} \det A &= \det \begin{pmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{pmatrix} \\ &= \det \begin{pmatrix} \sum_{i_1} a_{i_1 1} \vec{e}_{i_1} & \sum_{i_2} a_{i_2 2} \vec{e}_{i_2} & \dots & \sum_{i_n} a_{i_n n} \vec{e}_{i_n} \end{pmatrix} \\ &= \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \det \begin{pmatrix} \vec{e}_{i_1} & \vec{e}_{i_2} & \dots & \vec{e}_{i_n} \end{pmatrix} \end{aligned}$$

So, the determinant so far is some expression involving a product of entries of A from different columns, multiplied by some determinant of a matrix whose columns are some standard basis vectors. Antisymmetry implies that whenever two columns are the same, the determinant is zero. This is because swapping those two columns leaves the matrix unchanged, and yet the determinant becomes negative of itself. The only solution to $x = -x$ is $x = 0$.

Thus, every time any pair of indices i_j and i_k are equal, then that determinant is zero. That means $(i_1, i_2, \dots, i_n)$ is some permutation of $(1, 2, \dots, n)$. The set S_n is the set of all permutations of $(1, 2, \dots, n)$, so we can rewrite the determinant expression as

$$\det A = \sum_{i \in S_n} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \det \begin{pmatrix} \vec{e}_{i_1} & \vec{e}_{i_2} & \dots & \vec{e}_{i_n} \end{pmatrix}$$

This means each matrix appearing in the sum has the property that each column is a distinct standard basis vector: that is, the matrix is the result of permuting the columns of the identity matrix. The property of antisymmetry says that these determinants are either 1 or -1 since we assume $\det I_n = 1$. A common notation is to write $(-1)^i$ for this determinant, which is called the *sign* of the permutation. Therefore, we have the equation

$$\det A = \sum_{i \in S_n} (-1)^i a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \tag{1}$$

One thing we should check is that $(-1)^i$ is actually meaningful. Given a permutation $(i_1, \dots, i_n)$ in S_n, we may obtain $(1, 2, \dots, n)$ by the following process: look for 1 and swap it with location 1, look for 2 and swap it with location 2, and so on. If the number of swaps was even, then $(-1)^i = 1$, and if the number of swaps was odd, then $(-1)^i = -1$. We don't count swaps which don't move anything.

Let us test out this formula to come up with 2×2 and 3×3 determinant formulas. For 2×2, there are two permutations of $(1, 2)$, namely $(1, 2)$ and $(2, 1)$. $(-1)^{(1,2)} = (-1)^0 = 1$, and $(-1)^{(2,1)} = (-1)^1 = -1$ since it takes one swap to sort $(2, 1)$. Then,

$$\begin{aligned} \det \begin{pmatrix} a_{11} & 12 \\ a_{21} & a_{22} \end{pmatrix} &= (-1)^{(1,2)} a_{11} a_{22} + (-1)^{(2,1)} a_{21} a_{12} \\ &= a_{11} a_{22} - a_{21} a_{12}, \end{aligned}$$

just as the book claims. For 3×3, there are six permutations of $(1, 2, 3)$. Let us calculate the signs of each of these permutations ahead of time.

$$\begin{aligned} (-1)^{(1,2,3)} &= 1 \\ (-1)^{(1,3,2)} &= -(-1)^{(1,2,3)} = -1 \\ (-1)^{(2,1,3)} &= -(-1)^{(1,2,3)} = -1 \\ (-1)^{(2,3,1)} &= -(-1)^{(1,3,2)} = 1 \\ (-1)^{(3,1,2)} &= -(-1)^{(1,3,2)} = 1 \\ (-1)^{(3,2,1)} &= -(-1)^{(1,2,3)} = -1 \end{aligned}$$

Then,

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

(which is the formula mentioned on page 156 of the textbook). Try doing cofactor expansion of this matrix to see it really does match the book.

We have almost shown that determinants exist. All that remains is to show that $(-1)^i$ is actually actually meaningful. Why is it that if a permutation can be gotten to with an even number of swaps it can't be gotten to with an odd number of swaps, and vice versa? One reason I know is that there is a strange function defined in the following way from S_n: let $f(i) = \prod_{j<k} (x_{i_j} - x_{i_k})$. For instance, $f(2, 1, 3) = (x_2 - x_1)(x_2 - x_3)(x_1 - x_3) = -(x_1 - x_2)(x_1 - x_3)(x_2 - x_3) = -f(1, 2, 3)$. With this, $f(1, 2, \dots, n) = \prod_{j<k} (x_j - x_k)$, and $f(i) = \pm f(1, \dots, n)$ for any $i \in S_n$, and we define $(-1)^i$ to be the number where $f(i) = (-1)^i f(1, \dots, n)$. This matches our previous definition, because whenever we swap two entries of i to get i', then in $f(i')$ every term involving those two terms becomes negative of what it was, and every term involving only one term swaps with a term involving the other, except for the single term involving both, which becomes negative. Thus, there are an odd number of -1's which can be factored out, and $f(i') = -f(i)$. (I apologize for this reason; I'd need more time to make it better, but I doubt anyone will get this far anyway. If you're one who did read this, I'm happy to explain it to you.)

So: $(-1)^i$ is a well-defined expression, and since we derived the determinant from the properties, and there was only one choice for everything, there is a determinant, whose formula is given in equation 1.

3 Determinants and row reduction

The multilinearity properties were all in terms of columns, but we may also speak of multilinearity of rows for the reason that $\det A^T = \det A$. This equation is because

$$\sum_{i\in S_n}(-1)^i a_{i_1 1}a_{i_2 2}\cdots a_{i_n n} = \sum_{j\in S_n}(-1)^j a_{1 j_1}a_{2 j_2}\cdots a_{n j_n}$$

which isn't immediately clear, but which is from seeing that a swap of columns corresponds to a swap of rows, so we can translate $(-1)^i$ for a column permutation i into $(-1)^j$ for a row permutation j. That this has anything to do with the transpose is from the rule $(A^T)_{ij} = A_{ji}$.

Every elementary row operation in row reduction then has a clear effect on the determinant. For ease of writing these things down, they will be given as column operations, but keep the $\det A^T = \det A$ formula in mind. These are simply applications of multilinearity and antisymmetry:

- Scaling. $\det\left(\vec{a}_1 \ \dots \ c\vec{a}_i \ \dots \ \vec{a}_n\right) = c\det\left(\vec{a}_1 \ \dots \ \vec{a}_i \ \dots \ \vec{a}_n\right)$.
- Swapping. $\det\left(\vec{a}_1 \ \dots \ \vec{a}_i \ \dots \ \vec{a}_j \ \dots \ \vec{a}_n\right) = -\det\left(\vec{a}_1 \ \dots \ \vec{a}_j \ \dots \ \vec{a}_i \ \dots \ \vec{a}_n\right)$.
- Replacement. $\det\left(\vec{a}_1 \ \dots \ \vec{a}_i{+}c\vec{a}_j \ \dots \ \vec{a}_n\right) = \det\left(\vec{a}_1 \ \dots \ \vec{a}_i \ \dots \ \vec{a}_n\right) + c\det\left(\vec{a}_1 \ \dots \ \vec{a}_j \ \dots \ \vec{a}_n\right) = \det\left(\vec{a}_1 \ \dots \ \vec{a}_i \ \dots \ \vec{a}_n\right)$ (since the second determinant in the sum had a matrix with column $\vec{a}_j$ appearing twice).

Since an elementary matrix is the transformation for an elementary row operation, that is, the result of applying the row operation to the identity matrix, we then have

- Scaling. $\det(\text{scaling a row by } c) = c$.
- Swapping. $\det(\text{swapping two rows}) = -1$.
- Replacement. $\det(\text{adding } c \text{ of a row to another row}) = 1$.

With these in mind, we can show the all-important $\det(AB) = \det(A)\det(B)$, which lets us go back and forth between the complexity of matrix multiplication and the simplicity of real number multiplication.

If a matrix A is not invertible, then one column is a linear combination of the others, so by multilinearity (and reasoning similar to that for replacement) $\det A = 0$. Then, for any matrix B, since we know AB is not invertible, we see $\det(AB) = \det(A)\det(B)$ when A is not invertible, since both sides are 0.

Otherwise, if A is invertible, we have some more work to do. Previously in the course we showed that every invertible matrix is the product of some number of elementary matrices: $A = E_1E_2\cdots E_k$. Then, given $\det B$, we can get $\det(E_kB)$ by realizing E_kB is the result of applying the row operation E_k to B, so then $\det(E_kB) = \det(E_k)\det(B)$ since all of the numbers happen to match up no matter which of the three row operations E_k may happen to be. By repeatedly applying this reasoning, we get $\det(E_1\cdots E_kB) = \det(E_1)\det(E_2\cdots E_kB)$ and so on, until $\det(AB) = \det(E_1)\det(E_2)\cdots\det(E_k)\det(B)$. By then reversing this process, we can write $\det(E_1\cdots E_k)$ for $\det(E_1)\cdots\det(E_k)$ since each matrix represents a row operation. Thus, $\det(AB) = \det(A)\det(B)$ when A is invertible.

A nice side effect of this is that it suggests another way to compute determinants: row reduce a matrix until all the pivots are 1. If there are not enough pivots, the determinant is 0 since the matrix is not invertible. Otherwise, go back through which elementary row operations were performed and multiply the reciprocals of their determinants together. Here is an example of one way of keeping track of this:

$$\det\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} = \det\begin{pmatrix}1 & 1\\ 0 & -2\end{pmatrix} = -2\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} = -2\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = -2.$$

One calculation-accelerating fact is that an upper- (or lower-) triangular matrix has a determinant which is a product of the diagonal entries. This is because the only permutation in S_n which does not involve entries below (or above) the diagonal is $(1, 2, \dots, n)$. This is the only one which matters, since every other permutation involves a multiplication by one of those zero elements below (or above) the diagonal. Another reason is that an upper triangular matrix can be row reduced to a diagonal matrix using only replacement.

4 Cofactor expansion

Now we will show that cofactor expansion is a way to compute the determinant. This requires some temporary notation to maintain sanity. Let A_{ij} for now denote a *minor* of matrix A, which is the $(n-1)\times(n-1)$ matrix obtained by omitting row i and column j. We will show that expansion along the first row works:

Theorem 1 (Cofactor expansion along the first row)**.** *If A is 1×1, then $\det A = a_{11}$. Otherwise,*

$$\det A = \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \det A_{1j}$$

Proof. When A is 1×1, then (1) is the only permutation in S_1, so $\det A = a_{11}$.

Now suppose $n > 1$. The permutations in S_n can be split into n different classes, where class j is all permutations i with $i_j = 1$. For example, in S_3, the second class consists of $(2,1,3)$ and $(3,1,2)$. We may split the determinant sum according to class:

$$\det A = \sum_{j=1}^{n} \sum_{\substack{i\in S_n\\ i_j=1}} (-1)^i a_{i_1 1} a_{i_2 2}\cdots a_{i_n n} \tag{2}$$

$$= \sum_{j=1}^{n} a_{1j} \sum_{\substack{i\in S_n\\ i_j=1}} (-1)^i a_{i_1 1} a_{i_2 2}\cdots \widehat{a_{i_j j}} \cdots a_{i_n n} \tag{3}$$

where the $\widehat{a_{i_j j}}$ represents omitting that term from the product. The terms in the inner sum do not contain any entries of the first row of A, since they have been omitted by the condition $i_j = 1$ and by factoring a_{1j} out of the expression. The permutation $i' = (1, i_1, \ldots, \widehat{i_j}, \ldots, i_n)$ is related to i itself by swapping i_j with the term before it repeatedly until it is at the beginning, so $(-1)^{i'} = (-1)^{j-1}(-1)^i$. We may use this to see that

$$\sum_{\substack{i\in S_n\\ i_j=1}} (-1)^i a_{i_1 1} a_{i_2 2}\cdots \widehat{a_{i_j j}} \cdots a_{i_n n} = (-1)^{j-1} \sum_{\substack{i\in S_n\\ i_1=1}} (-1)^i a_{i_2 1} a_{i_3 2}\cdots a_{i_j (j-1)} \widehat{a_{i_{j+1} j}} a_{i_{j+1}(j+1)} \cdots a_{i_n n}$$

$$= (-1)^{j-1} \det A_{1j}$$

The second expression is supposed to represent using entries 2 through n of a permutation after the swapping the 1 to the front, which involves skipping over column j. This is hard to describe, but is easier if you (personally) think about an example. The third expression comes from realizing that a permutation of $(2,3,\ldots,n)$ can be thought of as a permutation of S_{n-1} by renumbering, and that the expression is the determinant of the minor A_{1j}, directly from equation 1. It will probably take many moments to believe this.

By substituting this back into the determinant expression in equation 3, $\det A = \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \det A_{1j}$. ☐

Theorem 2 (Cofactor expansion along a row)**.** *If A is $n\times n$ with $n > 1$, and if $1 \le k \le n$, then cofactor expansion along row k is*

$$\det A = \sum_{j=1}^{n} (-1)^{j+k} a_{kj} \det A_{kj}.$$

Proof. By swapping row k repeatedly with the row before it, we may obtain a matrix A' with row 1 being k, and with $\det A = (-1)^{k-1}\det A'$ since there are $k-1$ swaps between them. The relationship between minors is that $A'_{1j} = A_{kj}$, and among entries $a'_{1j} = a_{kj}$, so by cofactor expansion along the first row of A',

$$\det A = (-1)^{k-1} \sum_{j=1}^{n} (-1)^{j-1} a_{kj} \det A_{kj} = \sum_{j=1}^{n} (-1)^{k+j-2} a_{kj} \det A_{kj} = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det A_{kj}.$$

☐

Cofactor expansion along columns follows from $\det A^T = \det A$.

5 A word about efficiency

We have two methods of calculating a determinant, cofactor expansion and row reduction. Both are terrible to do by hand, but which is less terrible? One proxy for terribleness is the number of multiplications each takes, which we will find upper bounds for given a matrix of size $n \times n$.

For a 1×1 matrix, cofactor expansion takes 0 multiplications. For $n \times n$, there are n terms in the sum, each which has one multiplication by each of the n determinants of size $(n-1) \times (n-1)$. If $T_c(n)$ is the number of multiplications, we thus have the recurrence relations $T_c(1) = 0$ and $T_c(n) = n + nT_c(n-1)$.

n	$T_c(n)$
1	0
2	2
3	9
4	40
5	205
6	1236

Table 1: Worst-case number of multiplications for cofactor expansion

For row reduction, we have to think about the algorithm carefully. We don't count swaps, and we do not need to scale any rows throughout the algorithm since we may multiply down the diagonal at the very end ($n-1$ multiplications there), and for every pivot row, we perform a replacement operation for every row below it. Row i has $n-i+1$ nonzero entries and $n-i$ rows below it, each with $n-i+1$ entries themselves. Row i then takes $(n-i)(n-i+1)$ multiplications, which means $T_r(n) = n - 1 + \sum_{i=1}^{n-1}(n-i)(n-i+1)$ counts the worst-case number of multiplications to row reduce and then multiply down the diagonal.

n	$T_r(n)$
1	0
2	3
3	10
4	23
5	44
6	75

Table 2: Worst-case number of multiplications for row reduction

This suggests that it might be worth doing cofactor expansion for up to 3×3 matrices, and then considering row reduction for anything larger. Even if we scale rows while reducing, 4×4 takes 32 multiplications.

If a matrix has a lot of zeros, it might be worth cofactor expanding even if it is a large matrix, and it also might be worth performing some row or column operations just before beginning cofactor expansion to produce a lot of zeros.

If the point is to tell whether a matrix is invertible, it might be worth row reducing anyway because you might want the inverse.

6 Determinant of a product, again

While we showed that $\det(AB) = \det(A)\det(B)$ using row reduction, can we show it directly from the formula we derived for determinants? If so, this is a kind of calculation which is non-illuminating, but it can be nice to know it is possible. Let $C = AB$, so then $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$.

$$\begin{aligned}
\det(AB) &= \sum_{\sigma \in S_n} (-1)^\sigma c_{\sigma_1 1} \cdots c_{\sigma_n n} \\
&= \sum_{\sigma \in S_n} (-1)^\sigma \left(\sum_{k_1} a_{\sigma_1 k_1} b_{k_1 1} \right) \cdots \left(\sum_{k_n} a_{\sigma_n k_n} b_{k_n n} \right) \\
&= \sum_{k_1} \cdots \sum_{k_n} \sum_{\sigma \in S_n} (-1)^\sigma a_{\sigma_1 k_1} b_{k_1 1} \cdots a_{\sigma_n k_n} b_{k_n n} \\
&= \sum_{k_1} \cdots \sum_{k_n} \sum_{\sigma \in S_n} (-1)^\sigma a_{\sigma_1 k_1} \cdots a_{\sigma_n k_n} b_{k_1 1} \cdots b_{k_n n}
\end{aligned}$$

If two indices k_i and k_j are equal, then for a permutation σ, the $a_{\sigma_{k_i} k_i}$ and $a_{\sigma_{k_j} k_j}$ components of a term are swapped by a permutation σ' which is σ except whose components k_i and k_j are swapped. Such σ and σ' come in pairs with $(-1)^\sigma = -(-1)^{\sigma'}$, so the inner sum is zero in this case. This means $(k_1, \ldots, k_n)$ must be a permutation for the inner sum to be nonzero, so

$$\det(AB) = \sum_{k \in S_n} \sum_{\sigma \in S_n} (-1)^\sigma a_{\sigma_1 k_1} \cdots a_{\sigma_n k_n} b_{k_1 1} \cdots b_{k_n n}$$

For each permutation k and permutation σ, there is a permutation τ where $\tau_i = \sigma_{k_i^{-1}}$. Here, k^{-1} represents the inverse permutation: whenever $k_i = j$, then $k_j^{-1} = i$. This has the property that when $k_j = i$ then $\sigma_j = \tau_i$. Thus, having τk to mean the permutation with $(\tau k)_i = \tau_{k_i}$, we may rewrite the sum as

$$\det(AB) = \sum_{\tau \in S_n} \sum_{k \in S_n} (-1)^{\tau k} a_{\tau_1 1} \cdots a_{\tau_n n} b_{k_1 1} \cdots b_{k_n n}$$

One fact about permutations is that $(-1)^{\tau k} = (-1)^\tau (-1)^k$ since τk can be sorted to be τ by using the swaps to bring k to the $(1, 2, \ldots, n)$ permutation. With this in mind,

$$\begin{aligned}
\det(AB) &= \sum_{\tau, k \in S_n} (-1)^\tau (-1)^k a_{\tau_1 1} \cdots a_{\tau_n n} b_{k_1 1} \cdots b_{k_n n} \\
&= \left(\sum_{\tau \in S_n} (-1)^\tau a_{\tau_1 1} \cdots a_{\tau_n n} \right) \left(\sum_{k \in S_n} (-1)^k b_{k_1 1} \cdots b_{k_n n} \right) \\
&= \det(A) \det(B)
\end{aligned}$$

To give you the words behind this: S_n is called the *symmetric group*, and $k \mapsto (-1)^k$ is the *sign homomorphism*. We used group properties of composition of permutations and inverse permutations.