Determinants

Patrick Yin

May 30, 2020

1 Introduction

At UC Berkeley, most of my linear algebra foundation lied in EECS16A/B for the majority of my freshman year. While I really enjoyed EECS16A/B's application-based approach to linear algebra, I felt that they didn't do justice to some linear algebra topics. Particularly, I felt that determinants were not taught to its fullest. So in this note, I wanted to write down a rigorous derivation for many ideas related to determinants. Specifically, I develop the definition of determinations from multilinearity and use this definition to derive its permutation definition, cofactor expansion, and various other determinant properties.

2 Multilinear Map

Multilinearity serves as the core of the determinant definition, so we will first define multilinearity. Note that this definition is slightly simplified for our purposes.

Definition 1 (Multilinear Map Definition). *Let V_1, V_2, ..., V_k, and W be finite-dimensional vector spaces. The mapping $T : V_1 \times V_2 \times ... \times V_k \to W$ is multilinear if it is linear in each of its variables. In other words,*

$$\begin{aligned} &T(v_1,..., v_{i-1}, av_i + bv_i', v_{i+1}, ..., v_k) \\ &\qquad = aT(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_k) + bT(v_1, ..., v_{i-1}, v_i', v_{i+1}, ..., v_k) \end{aligned} \tag{1}$$

for all a, $b \in \mathbb{R}$, $v_j \in V_j$ for $j = 1$, ..., k and $v_i' \in V_i$ for $i = 1$, ..., k.

3 Determinant Definition

Intuitively, determinants for a 2x2 matrix represent the signed area of the parallelogram formed by its two column vectors. Determinants for a 3x3 matrix represent the signed volume of the parallelopiped formed by its three column vectors. For the general nxn matrix, although its impossible to visualize, the determinant should represent the signed "volume" of the hyperparallelopiped formed by its column vectors.

More rigorously defined, a determinant is a function det: $(\mathbb{R}^n)^n \rightarrow \mathbb{R}$ with the following properties:

1. det is multilinear (Intuitively, calculating volume is multilinear. Scaling one vector by some factor will proportionally scale the volume of the resulting parallelpiped by the same factor).

2. $\det(\vec{e}_1,...,\ \vec{e}_n) = \det(I_n) = 1$ (the "volume" of the unit hypercube is 1).

3. $\det(...,\vec{u},...,\vec{v}...) = -\det(...,\vec{v},...\vec{u},...)$ (This is the antisymmetry property. Intuitively, changing order of two vectors flips space over, so sign inverts. 3Blue1Brown's video on determinants gives a really good visual intuition for this).

Anything with these three properties must be the determinant. With this definition, we can now derive many ideas relating to the determinant.

Corollary 1. *If two vectors in the determinant are the same, the determinant is zero. In other words,*

$$det(..., \vec{u}, ..., \vec{u}, ...) = 0 \tag{2}$$

Proof. Using property 3 of determinants, we know that if we swap the two repeated columns, the determinant becomes the negative of itself:

$$\det(..., \vec{u}, ..., \vec{u}, ...) = -\det(..., \vec{u}, ..., \vec{u}, ...) \tag{3}$$

Thus, the determinant must equal 0. ☐

4 Determinants and Permutations

Let $A \in \mathbb{R}^{n\times n}$. We want to find det(A). To do this, let us first split A into its column vectors $\vec{a}_1, ..., \vec{a}_n$. Then,

$$\begin{aligned}
\det(A) &= \det(\vec{a}_1, \vec{a}_2, ..., \vec{a}_n) \\
&= \det(\sum_{i_1=1}^{n} a_{i_1 1}\vec{e}_{i_1}, \sum_{i_2=1}^{n} a_{i_2 2}\vec{e}_{i_2}, ..., \sum_{i_n=1}^{n} a_{i_n n}\vec{e}_{i_n}) \\
&= \sum_{i_1=1}^{n} a_{i_1 1}\det(\vec{e}_{i_1}, \sum_{i_2=1}^{n} a_{i_2 2}\vec{e}_{i_2}, ..., \sum_{i_n=1}^{n} a_{i_n n}\vec{e}_{i_n}) \\
&= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} ... \sum_{i_n=1}^{n} a_{i_1 1}a_{i_2 2}...a_{i_n n}\det(\vec{e}_{i_1}, \vec{e}_{i_2}, ..., \vec{e}_{i_n})
\end{aligned} \tag{4}$$

In lines 3 and 4, we can pull out the coefficients of the determinant due to its multilinearity property.

With corollary 1, we now know that every time $i_j = i_k$ for $j \neq k$, $\det(\vec{e}_{i_1}, \vec{e}_{i_2}, ..., \vec{e}_{i_n}) = 0$. This means we don't have to consider these cases when we calculate the determinant since they contribute zero to the value of the determinant. Thus, we only have to consider $(i_1, i_2, ..., i_n)$ where they are some permutation of $(1, 2, ..., n)$. Define S_n to be the set of all permutation of $(1, 2, ..., n)$. Then,

$$\det(A) = \sum_{i \in S_n} a_{i_1 1} a_{i_2 2} ... a_{i_n n} \det(\vec{e}_{i_1}, \vec{e}_{i_2}, ..., \vec{e}_{i_n}) \tag{5}$$

where $i := (i_1, i_2, ..., i_n)$.
We can now find the value of $\det(\vec{e}_{i_1}, \vec{e}_{i_2}, ..., \vec{e}_{i_n})$. Since we know that $\det(\vec{e}_1, \vec{e}_2, ..., \vec{e}_n) = 1$ by property 2 and each column swap flips the sign of the determinant by property 3, we just need to count the number of column swaps needed to get from $\det(\vec{e}_{i_1}, \vec{e}_{i_2}, ..., \vec{e}_{i_n})$ to $\det(\vec{e}_1, \vec{e}_2, ..., \vec{e}_n)$. If the number of swaps required is even, $\det(\vec{e}_{i_1}, \vec{e}_{i_2}, ..., \vec{e}_{i_n}) = 1$. If it is odd, $\det(\vec{e}_{i_1}, \vec{e}_{i_2}, ..., \vec{e}_{i_n}) = -1$. Therefore our equation is,

$$\det(A) = \sum_{i \in S_n} (-1)^i a_{i_1 1} a_{i_2 2} ... a_{i_n n} \tag{6}$$

where $(-1)^i$ is the sign of the determinant.

Remark. *How do we know that $(-1)^i$ is meaningful? In other words, how do we know that the number of steps needed to order an unordered list is deterministically even or odd? Can we find a counterexample where an unordered list can be ordered in an even number of steps with one sorting method and ordered in an odd number of steps with a different sorting method? Turns out, the answer is no. The number of steps being even or odd is deterministic, so $(-1)^i$ is meaningful. But why? I've spent a couple of hours trying to prove this myself, but to no avail. I have not been able to come up with a rigorous proof yet. However, I do have a intuitive explanation to why this is the case:*

Consider an optimal sorting method that sorts an unordered list in the least number of steps possible. It seems that every other sorting method is just some less efficient variation of that optimal sorting method. In other words, every other sorting method is just the optimal sorting method with unnecessary steps. It also seems that the number of unnecessary steps will always be even since every step that isn't a necessary step needs to be backtracked in some manner at some point (This is the iffy part that is hard to prove). Since an even number plus an even number is still even and an odd number plus an even number is still odd, the number of unnecessary steps doesn't change whether the total number of steps is even or odd.

5 Cofactor Expansion

Cofactor expansion is a method that computes the determinant of a matrix. It equates the determinant of a $n \times n$ matrix to the summation of its $(n-1) \times (n-1)$

minors. We denote A_{ij} to be a minor of matrix A obtained by omitting row i and column j.

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

$$det(A) = \sum_{j=1}^{n} (-1)^{j-1} a_{1j} det(A_{1j}) \tag{7}$$

Proof. When A is 1×1, $\det(A) = \det(\vec{a}_1) = \det(a_{11}\vec{e}_1) = a_{11} det(\vec{e}_1) = a_{11}$. Now lets consider $n > 1$. Since we want to prove cofactor expansion along the first row, we want to isolate terms where $i_j = 1$ in our determinant formula.

$$\begin{aligned} \det(A) &= \sum_{i \in S_n} (-1)^i a_{i_1 1} a_{i_2 2} ... a_{i_n n} \\ &= \sum_{j=1}^{n} \sum_{\substack{i \in S_n \\ i_j = 1}} (-1)^i a_{i_1 1} a_{i_2 2} ... a_{i_n n} \\ &= \sum_{j=1}^{n} a_{1j} \sum_{\substack{i \in S_n \\ i_j = 1}} (-1)^i a_{i_1 1} a_{i_2 2} ... \hat{a}_{i_j j} ... a_{i_n n} \end{aligned} \tag{8}$$

where $\hat{a}_{i_j j}$ represents omitting that term. Notice that currently $i = (i_1, ..., i_n)$. But now with the knowledge that $i_j = 1$, we can modify i to be $(1, i_1, ..., \hat{i}_j, ..., i_n)$. In order to go from $(i_1, ..., i_n)$ to $(1, i_1, ..., \hat{i}_j, ..., i_n)$, we need to shift i_j left $j - 1$ times to get i_j to the first position. In other words, any permutation $i \in S_n | (i_j = 1)$ can be modified to be $i \in S_n | (i_1 = 1)$ through $j - 1$ pairwise swaps. Thus,

$$\begin{aligned} \det(A) &= \sum_{j=1}^{n} a_{1j} \sum_{\substack{i \in S_n \\ i_j = 1}} (-1)^i a_{i_1 1} a_{i_2 2} ... \hat{a}_{i_j j} ... a_{i_n n} \\ &= \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \sum_{\substack{i \in S_n \\ i_1 = 1}} (-1)^i a_{i_2 1} a_{i_3 2} ... a_{i_j j-1} \hat{a}_{i_{j+1} j} a_{i_{j+1} j+1} ... a_{i_n n} \\ &= \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \det(A_{1j}) \end{aligned} \tag{9}$$

This part may take a while to understand. But essentially, we have factored out the j^{th} column where $i_j = 1$ (i.e. first row) from our equation as $(-1)^{j-1} \sum_{j=1}^{n} a_{1j}$. What is left turns out to be the determinant formula for A_{1j} since $i \in S_n | (i_i = 1)$ is just a permutation of S_{n-1}. $\square$

Remark. *This proof may take a while to understand because there are a lot of things at play here. Not only are we dealing with $j - 1$ column swaps to* *shift the original matrix to a new one, we are also doing this n times for each $j = 1, ..., n$. To really understand this process, I would try this proof by hand and write down simple examples to get an idea of what is going on.*

Theorem 2 (Cofactor Expansion along a row)**.** *If $A \in \mathbb{R}^{n \times n}$ and $n > 1$, then the cofactor expansion along row k is*

$$det(A) = \sum_{j=1}^{n} (-1)^{j+k} a_{kj} det(A_{kj}) \tag{10}$$

Proof. By swapping row k repeatedly with the row above it, we obtain matrix A' with row 1 of A' being row k of A. This takes $k-1$ swaps. This means that $\det(A) = (-1)^{k-1} \det(A')$. Also, we know that $A_{kj} = A'_{1j}$ and $a_{kj} = a'_{1j}$ for $j = 1, ..., n$. Then,

$$\begin{aligned}
\det A &= (-1)^{k-1} \det(A') \\
&= (-1)^{k-1} \sum_{j=1}^{n} (-1)^{j-1} a'_{1j} \det(A'_{1j}) \\
&= (-1)^{k-1} \sum_{j=1}^{n} (-1)^{j-1} a_{kj} \det(A_{kj}) \\
&= \sum_{j=1}^{n} (-1)^{k+j-2} a_{kj} \det(A_{kj}) \\
&= \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det(A_{kj})
\end{aligned} \tag{11}$$

□

Since $\det(A) = \det(A^T)$ (this is proved below), cofactor expansion along the columns of A is the same as cofactor expansion along the rows of A except that $a_{kj} \to a_{jk}$ and $A_{kj} \to A_{jk}$.

6 Properties of Determinants

Corollary 2. $det(A) = det(A^T)$

Proof. We know that

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

and

$$\det(A^T) = \textstyle\sum_{j \in S_n} (-1)^j a_{1 j_1} a_{2 j_2} ... a_{n j_n}$$

These two terms are actually identical. This is because for every product chain of the n a_{ij}s, all the is have to be distinct from one another and all the js have to be distinct from one another between each a_{ij}. This applies to both det(A) and det(A^T). Thus there is a one-to-one mapping between $a_{i_1 1}a_{i_2 2}...a_{i_n n}$ and $a_{1j_1}a_{2j_2}...a_{nj_n}$ where $a_{i_1 1}a_{i_2 2}...a_{i_n n} = a_{1j_1}a_{2j_2}...a_{nj_n}$. For each bijective pair, $(-1)^i = (-1)^j$ since $a_{i_1 1}a_{i_2 2}...a_{i_n n} = a_{1j_1}a_{2j_2}...a_{nj_n}$. □

Corollary 3. *Every elementary row operation in Gaussian row reduction has a clear effect on the determinant.*

1. *Scaling: det(matrix with scales a row by c) = c (e.g.* $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ *scales the second row by 2)*
2. *Swapping: det(matrix with swaps two rows) = -1 (e.g.* $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ *swaps the two rows)*
3. *Replacement: det(adding c of one row to another row) = 1 (e.g.* $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ *adds two times the second row to the first row)*

Proof. Using the determinant's multilinearity and antisymmetry properties, we can prove for all 3 types of elementary matrices.

1. Scaling: $\det(\vec{v_1}, ..., c\vec{v_i}, ..., \vec{v_n}) = c \det(\vec{v_1}, ..., \vec{v_i}, ..., \vec{v_n})$
2. Swapping: $\det(...,\vec{u},...,\vec{v}...) = -\det(...,\vec{v},...\vec{u},...)$
3. Replacement: $\det(\vec{v_1}, ..., \vec{v_i} + c\vec{v_j}, ..., \vec{v_n})$
   $= \det(\vec{v_1}, ..., \vec{v_i}, ..., \vec{v_n}) + c \det(\vec{v_1}, ..., \vec{v_j}, ..., \vec{v_n}) = \det(\vec{v_1}, ..., \vec{v_i}, ..., \vec{v_n})$

□

Corollary 4. *det(AB)= det(A)det(B)*

Proof. We will break our proof into two cases.
If A is not invertible, AB is also not invertible so det(AB)= 0 = det(A)det(B).
If A is invertible, $A = E_1E_2...E_k$ where E_i are elementary matrices (this is just reversed Gaussian elimination). Corollary 3 implies that $\det(E_kB) = \det(E_k)$det(B) since the numbers match up no matter which elementary row operation are used. We can repeat this reasoning to show that $\det(E_1E_2...E_kB) = \det(E_1)\det(E_2)...\det(E_k)$det(B). So $\det(AB) = \det(E_1)\det(E_2)...\det(E_k)$det(B). Again using Corollary 3, we find that $\det(AB) = \det(A)\det(B)$.

Remark. *There is another proof, which I won't go into, that uses cofactor expansion to prove this property. However, this is the most elegant proof in my opinion.*

□

Corollary 5. *det(αA)= α^ndet(A) where $A \in \mathbb{R}^{n\times n}$*

Proof. αA scales each row of A by α. By the multilinear property of determinants, we can factor out an α for each row of A when calculating the determinant of αA. ☐

Corollary 6. *Consider the block matrices A_1 and A_2:*

$$A_1 = \begin{bmatrix} I_{k\times k} & 0 \\ 0 & A \end{bmatrix}, A_2 = \begin{bmatrix} A & 0 \\ 0 & I_{k\times k} \end{bmatrix} \tag{12}$$

such that A is square. Then,

$$det(A_1) = det(A_2) = det(A) \tag{13}$$

Proof. Applying cofactor expansion on the first row of A_1, we get $\det(A_1) = \det(\begin{bmatrix} I_{(k-1)\times(k-1)} & 0 \\ 0 & A \end{bmatrix})$. We repeatedly apply this process until we get to det(A). We apply a similiar process to prove for A_2 where we repeatedly apply cofactor expansion on the last row of A_2. ☐

Corollary 7. *Consider the block matrices X_1, X_2, and X_3:*

$$X_1 = \begin{bmatrix} X_{11} & 0 \\ 0 & X_{22} \end{bmatrix}, X_2 = \begin{bmatrix} X_{11} & X_{12} \\ 0 & X_{22} \end{bmatrix}, X_3 = \begin{bmatrix} X_{11} & 0 \\ X_{12} & X_{22} \end{bmatrix} \tag{14}$$

such that X_{11} and X_{22} are both square matrices. Then,

$$det(X_1) = det(X_2) = det(X_3) = det(X_{11})det(X_{22}) \tag{15}$$

Proof.

$$\begin{aligned} X_1 &= \begin{bmatrix} X_{22} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & X_{22} \end{bmatrix} \\ \det(X_1) &= \det(\begin{bmatrix} X_{22} & 0 \\ 0 & I \end{bmatrix})\det(\begin{bmatrix} I & 0 \\ 0 & X_{22} \end{bmatrix}) \\ \det(X_1) &= \det(X_{11})\det(X_{22}) \end{aligned} \tag{16}$$

The last line of equation 16 uses corollary 6.

We will now proof for X_3. If X_{11} is not invertible, X_3 is not invertible so $\det(X_3) = 0 = \det(X_{11})\det(X_{22})$. Otherwise, if X_{11} is invertible,

$$\begin{aligned} X_1 &= \begin{bmatrix} I & 0 \\ -X_{12}X_{11}^{-1} & I \end{bmatrix} X_3 \\ \det(X_1) &= \det(\begin{bmatrix} I & 0 \\ -X_{12}X_{11}^{-1} & I \end{bmatrix})\det(X_3) \\ \det(X_1) &= \det(X_3) \end{aligned} \tag{17}$$

$\det(\begin{bmatrix} I & 0 \\ -X_{12}X_{11}^{-1} & I \end{bmatrix}) = I$ can be derived by repeatedly applying cofactor expansion on the first row until we end up with just I.

Since det(A)=det(A^T), det(X_2)=det(X_3). ☐

Remark. *Intuitively, we can see that X_{12} cannot form any nonzero permutations since it will always match up with the adjacent zero matrix, so the corollary must hold.*

Corollary 8. *$det(A^{-1}) = \frac{1}{det(A)}$*

Proof.

$$\begin{aligned} AA^{-1} &= I \\ \det(AA^{-1}) &= \det(I) \\ \det(A)\det(A^{-1}) &= 1 \\ \det(A^{-1}) &= \frac{1}{\det(A)} \end{aligned} \tag{18}$$

☐

Corollary 9. *The determinant of a triangular matrix is the product of its diagonal entries.*

Proof. Assume we are working with a lower-triangular matrix first. Apply cofactor expansion on the first row repeatedly. Automatically, we see that the determinant of a lower-triangular matrix is the product of its diagonal entries. Since the determinant of a matrix and the matrix's tranpose is the same, we know this is true to upper-triangular matrices as well. ☐

Lemma 3 (Characteristic polynomial)**.**

$$det(A - \lambda I) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)...(\lambda - \lambda_n) \tag{19}$$

where $\lambda_1, \lambda_2, ..., \lambda_n$ are the eigenvalues of A.

Proof. We can see that

$$\det(A - \lambda I) = (-1)^n \det(\lambda I - A) \tag{20}$$

This equation is true because $(\lambda I - A)$ is just the negative of $(A - \lambda I)$. In other words, each row of $(A - \lambda I)$ is scaled by -1 to get to $(\lambda I - A)$. Using the scaling property of corollary 3, we pull out all the -1s (one for each row of the matrix) to get $(-1)^n$.
Next, we know that

$$\lambda I - A = \begin{bmatrix} \lambda - a_{11} & . & a_{1n} \\ . & . & . \\ a_{n1} & . & \lambda - a_{nn} \end{bmatrix} \tag{21}$$

In order to calculate $\det(\lambda I - A)$, we do cofactor expansion along the first row repeatedly. We can see that the highest power of the characteristic polynomial must be λ^n with coefficient one (comes from $(\lambda - a_{11})(\lambda - a_{22})...(\lambda - a_{nn})$). Thus the polynomial must be in the form of $\lambda^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + ... + c_1\lambda + c_0\lambda$ where c_i are the scalar coefficients for the polynomial. This polynomial must then equal to $(\lambda - \lambda_1)(\lambda - \lambda_2)...(\lambda - \lambda_n)$ since the polynomial's roots have to be at the eigenvalues of A (i.e. $\lambda = \lambda_i$ is where $\det(\lambda I - A) = 0$). Plugging this polynomial back into $\det(\lambda I - A)$, we have proved our lemma. □

Corollary 10. *det(A) is the product of the eigenvalues of A.*

Proof.

$$\begin{aligned} \det(A - \lambda I) &= (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)...(\lambda - \lambda_n) \\ &= (\lambda_1 - \lambda)(\lambda_2 - \lambda)...(\lambda_n - \lambda) \end{aligned} \tag{22}$$

The first equation comes directly from lemma 3.
By setting λ to 0, we get

$$\det(A) = \lambda_1\lambda_2...\lambda_n \tag{23}$$

□

7 2x2 Matrices

Corollary 11 (Determinant of 2x2 matrix)**.**

$$det(\begin{bmatrix} a & b \\ c & d \end{bmatrix}) = ad - bc \tag{24}$$

Proof. We can prove this geometrically. Draw a vector $\begin{bmatrix} a \\ c \end{bmatrix}$ and another vector $\begin{bmatrix} b \\ d \end{bmatrix}$. Create a parallelogram with the two vectors as two of the sides. With some geometry, we can see the the area of the parallelogram must be $ad - bc$. □