Derivative of the Determinant

R. Scott McIntire

July 11, 2023

1 What is the Determinant of a Matrix?

The determinant of an $n \times n$ matrix A is the factor by which the volume of the unit cube is changed upon the action of A. Essentially, the cube is transformed into a parallelepiped. The primary edges of the unit cube are the vectors: $\{[1, 0, \ldots], [0, 1, 0, \ldots], \ldots [0, \ldots, 1]\}$. When A acts on these vectors through matrix multiplication produces the column vectors of A – by the definition of matrix multiplication.

So, to find this factor we need to compute the volume of the parallelepiped formed from the column vectors of A, since the volume of the unit cube is 1. Given n vectors in R^n, how do we find a formula for the parallelepiped volume? A good way is to use the properties of the function to help find a formula.

Let's call our volume function $V(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. Here's some properties we should agree on:

Multi-linear Property: V is linear in each slot. That is, it is a multi-linear function.

Degenerate Property: If any two edges are the same then the volume is 0.

Volume of Canonical Unit Cube: The volume of the unit cube using standard ordering of unit edges, $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, is 1.

Let's go through each property. The first property is true with respect to scaling. If we scale any slot (edge) by a factor k then the volume should of the resulting parallelepiped should scale by k. That addition should distribute in each slot takes more thought which we leave to the reader.

The degeneracy property is clear, if we have duplicate edges then a parallelogram in R^2 collapses to a line segment and consequently the area is zero. Similarly for higher dimensions.

The third property is clearly true, but why the need to care about the ordering of the standard edges? We show why by looking at a property implied by the multi-linearity and degeneracy properties.

Let's focus on R^3 as the general case is similar. Notice that

$$V(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_3) = 0 \tag{1}$$

because if the same vector is in two slots the volume is 0. But by the multi-linearity (1) becomes:

$$\begin{aligned} 0 &= V(\mathbf{x}_1, \mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_3) + V(\mathbf{x}_2, \mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_3) \\ &= V(\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_3) + V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) + V(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3) + V(\mathbf{x}_2, \mathbf{x}_2, \mathbf{x}_3) \\ &= V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) + V(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3) \end{aligned} \tag{2}$$

From this we see that

$$V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = -V(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3) \tag{3}$$

That is, V is *anti-symmetric*.

So, the reason the third property requires the unit cube to be ordered is that order matters because of this anti-symmetry property. *That is, a consequence of our assumptions is that the volume function will be signed!*

Let's now use these two properties to derive a formula for the volume function.

Each of the vectors: $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ can be written in terms of an orthonormal basis: $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ as

$$\begin{aligned} \mathbf{x}_1 &= a_{1,1}\mathbf{e}_1 + a_{2,1}\mathbf{e}_2 + a_{3,1}\mathbf{e}_3 \\ \mathbf{x}_2 &= a_{1,2}\mathbf{e}_1 + a_{2,2}\mathbf{e}_2 + a_{3,2}\mathbf{e}_3 \\ \mathbf{x}_3 &= a_{1,3}\mathbf{e}_1 + a_{2,3}\mathbf{e}_2 + a_{3,3}\mathbf{e}_3 \end{aligned}$$

If $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ are the mapping $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}$ by matrix A, then A is related to the $\{a_{i,j}\}_{i,j=1}^n$ by:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \tag{4}$$

So, coming up with a formula for the volume function in terms of the $\{a_{i,j}\}_{i,j=1}^n$ will translate directly to a formula for the determinant of A.

To illustrate the process of finding the volume I will do it for the case when $n = 2$. In this case we have:

$$\begin{aligned} V(\mathbf{x}_1, \mathbf{x}_2) &= V(a_{1,1}\mathbf{e}_1 + a_{2,1}\mathbf{e}_2, a_{1,2}\mathbf{e}_1 + a_{2,2}\mathbf{e}_2) \\ &= V(a_{1,1}\mathbf{e}_1, a_{1,2}\mathbf{e}_1 + a_{2,2}\mathbf{e}_2) + V(a_{2,1}\mathbf{e}_2, a_{1,2}\mathbf{e}_1 + a_{2,2}\mathbf{e}_2) \\ &= V(a_{1,1}\mathbf{e}_1, a_{1,2}\mathbf{e}_1) + V(a_{1,1}\mathbf{e}_1, a_{2,2}\mathbf{e}_2) \\ &+ V(a_{2,1}\mathbf{e}_2, a_{1,2}\mathbf{e}_1) + V(a_{2,1}\mathbf{e}_2, a_{2,2}\mathbf{e}_2) \\ &= a_{1,1}a_{1,2}V(\mathbf{e}_1, \mathbf{e}_1) + a_{1,1}a_{2,2}V(\mathbf{e}_1, \mathbf{e}_2) \\ &+ a_{2,1}a_{1,2}V(\mathbf{e}_2, \mathbf{e}_1) + a_{2,1}a_{2,2}V(\mathbf{e}_2, \mathbf{e}_2) \end{aligned}$$

Notice the we could pull out the $\{a_{i,j}\}_{i,j=1}^n$ because of multi-linearity – linearity refers not just to addition, but also scalar multiplication. Also, We have shown that the degeneracy condition implies anti-symmetry: $V(\mathbf{e}_1, \mathbf{e}_2) = -V(\mathbf{e}_2, \mathbf{e}_1)$. Finally, we know the volume (area in this case) of $V(\mathbf{e}_1, \mathbf{e}_2)$. It represents the volume of the unit cube (unit square in our case) and its volume is 1. Therefore,

$$V(\mathbf{x}_1, \mathbf{x}_2) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \tag{5}$$

This corresponds with the usual definition of the determinant of the matrix A.

The general case will have to do all of the permutations of the indices of a upon re-ordering the base vectors into standard form.

Let's see what this looks like in R^3.

$$\begin{aligned} V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) &= V(a_{1,1}\mathbf{e}_1 + a_{2,1}\mathbf{e}_2 + a_{3,1}\mathbf{e}_3, \\ &\quad a_{1,2}\mathbf{e}_1 + a_{2,2}\mathbf{e}_2 + a_{3,2}\mathbf{e}_3, \\ &\quad a_{1,3}\mathbf{e}_1 + a_{2,3}\mathbf{e}_2 + a_{3,3}\mathbf{e}_3) \end{aligned} \tag{6}$$

By multi-linearity this can be decomposed into the following:

$$V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} a_{i,1}a_{j,2}a_{k,3}V(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k) \tag{7}$$

There are 3^3 terms to consider. In the general case, there will be n^n terms to deal with – this is not looking promising. But notice, not all terms, $V(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k)$ survive. Any such term with a duplicate $\mathbf{e}_i$ will have value 0 because of the degeneracy condition. What terms do survive? The terms where the indices, i, j, k, differ. That is, i, j, k are *permutations* of $1, 2, 3$.[1] In the general case this reduces the number of terms in the sum from n^n to $n!$. In the general case, the sum becomes:[2]

$$V(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \sum_{\sigma \in S_n} a_{\sigma(1),i}a_{\sigma(2),2} \ldots a_{\sigma(n),n}V(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \ldots, \mathbf{e}_{\sigma(n)}) \tag{8}$$

Now, how do we compute each of the terms: $V(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \ldots, \mathbf{e}_{\sigma(n)})$? We know that we can switch the values of any two of the n slots and the value of V will change signs. It turns out that for any permutation one can apply some minimal number of these *transpositions* (switches) to get back to the canonical ordering $V(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$. So the value of $V(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \ldots, \mathbf{e}_{\sigma(n)})$ should be the value of $V(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ times $(-1)^k$, where k is the number of minimal transpositions needed to get to the canonical ordering of indices. This value, $(-1)^k$, can be represented by the *sgn* of the permutation.[3] The sgn of a permutation is set to 1 if k is odd and set to 0 if k is even. And since we know the volume of the canonical cube is 1, the formula

[1] More generally, permutations on n objects are one to one mappings from the set of integers $\{1, 2, \ldots, n\}$ to itself.
[2] S_n is the symmetric group of n objects.
[3] The mapping, sgn, is a group homomorphism: $S_n \to Z_2$.

for the parallelepiped – and hence the determinant of the matrix A – becomes:

$$\begin{aligned}
|A| &= V(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \\
&= V(a_{1,1}\mathbf{e}_1 + a_{2,1}\mathbf{e}_2 + \ldots + a_{n,1}\mathbf{e}_n, \\
&\qquad a_{1,2}\mathbf{e}_1 + a_{2,2}\mathbf{e}_2 + \ldots + a_{n,2}\mathbf{e}_n, \\
&\qquad \vdots \\
&\qquad a_{1,n}\mathbf{e}_1 + a_{2,n}\mathbf{e}_2 + \ldots + a_{n,n}\mathbf{e}_n) \quad \text{(expanding vectors into components)} \\
&= \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_n=1}^{n} a_{j_1,1} a_{j_2,2} \ldots a_{j_n,n} V(\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \ldots, \mathbf{e}_{j_n}) \quad \text{(by the Multi-linearity Property)} \\
&= \sum_{\sigma \in S_n} a_{\sigma(1),1} a_{\sigma(2),2} \ldots a_{\sigma(n),n} V(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \ldots, \mathbf{e}_{\sigma(n)}) \quad \text{(by the Degeneracy Property)} \\
&= \sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma)} a_{\sigma(1),1} a_{\sigma(2),2} \ldots a_{\sigma(n),n} V(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) \quad \text{(after placing edges in Canonical Order)} \\
&= \sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma)} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n} \quad \text{(Volume of Canonical Unit Cube is 1)} \qquad (9)
\end{aligned}$$

Note: There is a bug with this process. From our reasoning, it follows that the volume of $V(\mathbf{e}_2, \mathbf{e}_1)$ is -1. It seems like just as valid a way to describe the unit square, so why is the area negative? We can fixed this "bug" by claiming that it is, in fact, a feature. We can claim that we haven't just computed the area/volume, we have computed the *oriented volume*. If you want the volume you need only take the absolute value of the oriented volume.

We write the final formula down for reference but now with the caveat that the determinant of a matrix is not a positive stretching factor – the amount that the volume of the unit cube is stretched under A, but a *signed* stretching factor. Again, to find the magnitude of the stretching, take the absolute value of the determinant.

$$|A| = \sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma)} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n} \qquad (10)$$

1.1 Determinant of A^T

Notice that the formula (10) can also be written as:

$$|A| = \sum_{\sigma \in S_n} (-1)^{\text{sgn}(T(\sigma))} a_{T(\sigma)(1),1} a_{T(\sigma)(2),2} \cdots a_{T(\sigma)(n),n} \qquad (11)$$

Here, T is any one to one mapping from S_n to itself. Why? Because the sum in (10) is the sum over all elements of S_n. The mapping T just rearranges that sum. One such choice for T is the mapping that sends σ to its inverse, σ^{-1}. In this case, (11) becomes:

$$|A| = \sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma^{-1})} a_{\sigma^{-1}(1),1} a_{\sigma^{-1}(2),2} \cdots a_{\sigma^{-1}(n),n} \qquad (12)$$

But the multiplication can be rearranged in each term to give:

$$|A| \quad = \quad \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma^{-1})} a_{1,\sigma(1)} a_{2,\sigma(2)} \, \ldots \, a_{n,\sigma(n)} \tag{13}$$

For instance, after rearranging the multiplications in (12) there is an index j, so that $\sigma^{-1}(j) = 1$. So we have the first term in such an ordered product, $a_{1,j}$. But what must j be? It must be $\sigma(1)$ – since $\sigma^{-1}(\sigma(1)) = 1$. Similarly, the next term is $a_{2,\sigma(2)}$, etc. Then (13) can be written as:

$$|A| \quad = \quad \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma^{-1})} a'_{\sigma(1),1} a'_{\sigma(2),2} \, \ldots \, a'_{\sigma(n),n} \tag{14}$$

Here, we define $a'_{i,j}$ as the i^{th}, j^{th} element of the A^T.

Finally, for any transpositions: $\tau_1, \tau_2, \ldots, \tau_k$ that "unravel" the permutation σ back to the identity we have: $\tau_k \tau_{k-1} \ldots \tau_1 \sigma = \epsilon$.[4] Here, the multiplication is the group multiplication of applying one permutation on another. From this we know, $\sigma^{-1} = \tau_1^{-1} \tau_2^{-1} \ldots \tau_k^{-1}$. That is, for any collection of transpositions that returns σ to the identity, one can find a collection of transpositions *of the same number* that bring σ^{-1} back to the identity. Consequently, $\mathrm{sgn}(\sigma) = \mathrm{sgn}(\sigma^{-1})$.[5] With this, (14) may be written:

$$|A| \quad = \quad \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma)} a'_{\sigma(1),1} a'_{\sigma(2),2} \, \ldots \, a'_{\sigma(n),n} \tag{15}$$

But this is, by definition, nothing more than the determinant of A^T. From this we conclude that $|A| = |A^T|$.

2 Derivative of the Determinant

The determinant is a function from the space of matrices to R; that is, $|\cdot| : R^{n^2} \to R$. Therefore, its derivative at a particular "point" A, L_A, should satisfy:[6]

$$|A + H| \quad = \quad |A| + L_A(H) + o(H) \tag{16}$$

Here $\lim_{H \to 0} \frac{\|o(H)\|}{\|H\|} = 0$, where $\|H\|$ is any matrix norm and L_A is a linear mapping from R^{n^2} to R.

To start, let's try to find the derivative for an especially simple "point", the identity matrix, I.

$$|I + H| \quad = \quad \prod_{i=1}^{n} (1 + H_{i,i}) + o(H) \tag{17}$$

[4]We are using ϵ to denote the identity in S_n.

[5]As mentioned in a previous footnote, the mapping $\mathrm{sgn} : S_n \to Z_2$ is a group homomorphism. Therefore,

$$0 = \mathrm{sgn}(\epsilon) = \mathrm{sgn}(\sigma\sigma^{-1}) = \mathrm{sgn}(\sigma) + \mathrm{sgn}(\sigma^{-1}).$$

Consequently, $\mathrm{sgn}(\sigma^{-1}) = \mathrm{sgn}(\sigma)$, as they must both be 0 or both be 1 for their sum to be 0 in Z_2.

[6]Recall that for a function $f : R \to R$, the derivative at a point x is a linear function, $Df_x : R \to R$, such that $f(x+h) = f(x) + Df_x(h) + o(h)$.

In other words, the only $o(H)$ term comes from only one of the $n!$ permutations. The reason is that any other permutation will have 2 or more $\{H_{i,j}\}_{i,j=1}^{n}$ terms. To see this, consider any permutation, if it is not the identity then for some index, i, $a_{\sigma(i),i}$ is off the diagonal, and so its value is $h_{j,i}$ for some j. Now the value of $\sigma(j)$ can't be j, because i is already sent there (permutations are one to one mappings). Therefore, for this permutation, σ, the product term contains the product of two off diagonal terms times the other $n-2$ terms: $h_{i,j}h_{j,k} * ((n-2)\text{terms})$ which is $o(H)$. This product can be further decomposed into a constant term; a linear term in H; and $o(H)$ terms:

$$\begin{aligned} |I+H| &= 1+\sum_{i=1}^{n} H_{i,i} + o(H) \\ &= |I| + \mathrm{tr}(H) + o(H) \end{aligned} \tag{18}$$

So, the linear function acting on H to produce a number is the trace operator.

Now, let's do the same computations for a more general "point", an $n \times n$ matrix, A, with the stipulation that A is invertible.

$$\begin{aligned} |A+H| &= |A\left(I + A^{-1}H\right)| \\ &= |A|\,|I+A^{-1}H| \quad \text{(Multiplicative Property of Determinants)} \\ &= |A|\left(|I| + \mathrm{tr}(A^{-1}H) + o(H)\right) \quad \text{(From (18))} \\ &= |A| + |A|\mathrm{tr}(A^{-1}H) + o(H) \end{aligned} \tag{19}$$

Therefore, the derivative of the determinant at a non-singular "point", A, is the linear function $L_A : R^{n^2} \to R$ defined by:

$$L_A(B) = |A|\mathrm{tr}(A^{-1}B) \tag{20}$$

exercise: Check that L_A is a linear function.

exercise: What is the derivative at a "point" A where A is singular?

3 Applications: Liouville's Theorem

Suppose that a dynamical system is deforming n vectors in R^n over time by the following linear *vector differential equations*:

$$\frac{d\mathbf{x}_i(t)}{dt} = A\mathbf{x}_i(t) \quad i \in [1, n] \tag{21}$$

$$\mathbf{x}_i(0) = \mathbf{x0}_i \quad i \in [1, n] \tag{22}$$

Question: How is the volume of the parallelepiped that these vectors represent evolving in time?

To start, we could collect these vector equations into one linear *matrix differential equation*:

$$
\begin{aligned}
\frac{dX(t)}{dt} &= AX(t) && (23)\\
X(0) &= X0 && (24)
\end{aligned}
$$

Here, the n vectors form the columns of the matrix $X(0)$. The columns of $X(t)$ represent the vector's evolution in time.

To answer the question posed, we need to track $|X(t)|$. To do this, we find a differential equation that $|X(t)|$ satisfies and then solve.

$$
\begin{aligned}
\frac{d|X(t)|}{dt} &= |X(t)|\,\mathrm{tr}\left(X(t)^{-1}\frac{dX(t)}{dt}\right) \quad \text{(by (20) and the chain rule)}\\
\frac{d|X(t)|}{dt} &= |X(t)|\,\mathrm{tr}\left(X(t)^{-1}AX\right) \quad \text{(by (23))}\\
\frac{d|X(t)|}{dt} &= |X(t)|\,\mathrm{tr}(A) \quad \text{(by the property of trace: } \mathrm{tr}(ABC) = \mathrm{tr}(CAB))
\end{aligned}
$$

Therefore, $|X(t)|$ satisfies the following linear *scalar differential equation*:

$$
\begin{aligned}
\frac{d|X(t)|}{dt} &= \mathrm{tr}(A)\,|X(t)| && (25)\\
|X(0)| &= |X0| && (26)
\end{aligned}
$$

This is much easier to solve, the solution is $|X(t)| = |X0|\,e^{\mathrm{tr}(A)\,t}$.

Question: What kind of systems (defined by A) leave the volumes unchanged?

Notice that if $\langle A\mathbf{x}, \mathbf{x}\rangle = 0$; that is, A is *anti-symmetric*, then:

$$
\left\langle \mathbf{x}(t), \frac{d\,\mathbf{x}(t)}{dt} \right\rangle = \langle \mathbf{x}(t), A\mathbf{x}(t)\rangle = 0 \tag{27}
$$

This says that an object's velocity is perpendicular to its position. In this case, if $\{\mathbf{e}_i\}_{i=1}^{n}$ is an orthonormal system of vectors then $\mathrm{tr(A)} = \sum_{i=1}^{n} \langle \mathbf{e}_i, A\mathbf{e}_i\rangle = \sum_{i=1}^{n} 0 = 0$. Therefore, when A is anti-symmetric, the determinant of the evolving vectors – their volume – under the dynamical system does not change.

exercise: In this case, what kind of matrix is $X(t)$?