Math on a Desert Isle

Daniel Rui - December 30, 2019

This paper is dedicated to providing some of the motivation, or backdrop, to two particular areas of mathematics I long considered "black magic", namely vector multiplication and determinants. Of course, I can't possibly encompass all the necessary foundations of math in this short paper, but I'll try my best to be as descriptive and intuitive as possible. After all, the imagined situation is that you — a bright young mathematician with great passion for mathematics, but possessing only basic garden-variety textbook algebra knowledge — are stranded on a desert island trying to derive formulas for vector multiplication and determinants from scratch. In that case, I hope this paper will provide some semblance of guidance as to how you would accomplish such a task.

1 $\vec{a} \times \vec{b}$

Here, we have a nice little garden-variety vector:

$$\vec{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

We can define addition and scalar multiplication in the familiar and intuitive ways as

$$\vec{a} + \vec{b} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix} \quad \text{and} \quad c\vec{a} = \begin{bmatrix} ca_1 \\ \vdots \\ ca_n \end{bmatrix}$$

These two definitions are actually really good! Beyond their intuitiveness, they actually satisfy a very important condition: they are impervious to rotations: consider a rotation $\mathcal{M}$, say a 90 degree rotation clockwise. Notice that applying such a rotation to two individual vectors and adding them together results in the *same* vector as the rotation applied to the sum of the two vectors (picture this in your mind for maximum effect). Scalar multiplication is analogous. Generally, this is denoted

$$\mathcal{M}\vec{a} + \mathcal{M}\vec{b} = \mathcal{M}(\vec{a} + \vec{b}) \quad \text{and} \quad \mathcal{M}(c\vec{a}) = c\mathcal{M}\vec{a}$$

Sidenote: addition and scalar multiplication are actually impervious to *any* linear transformation, but later we'll see that vector multiplication is not consistent with "imperviousness to any linear transformation".

The failure to satisfy this property of "invariance under rotations" is why our intuitive definition of vector multiplication is insufficient. For example, if we define multiplication as

$$\vec{a}\vec{b} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 b_1 \\ \vdots \\ a_n b_n \end{bmatrix},$$

we should want to the following two multiplications to return the same vector, only rotated a bit:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix}$$

But of course, there is no rotation going from

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix}$$

Our task now is to find some operation (which I will denote as $\vec{a} \times \vec{b}$) that has some notion of multiplicativeness, satisfies the aforementioned property of invariance, and takes in two vectors and spits out a resulting vector. These "properties of multiplicativeness" (called bilinearity) are defined as such:

$$\begin{aligned} \lambda\vec{a} \times \vec{b} &= \lambda(\vec{a} \times \vec{b}) = \vec{a} \times \lambda\vec{b} \\ \vec{a} \times (\vec{b} + \vec{c}) &= \vec{a} \times \vec{b} + \vec{a} \times \vec{c} \\ (\vec{a} + \vec{b}) \times \vec{c} &= \vec{a} \times \vec{c} + \vec{b} \times \vec{c} \end{aligned}$$

Notice that these just deal with distributivity, not commutativity or associativity. Although it is not obvious at all why we isolated these properties, it turns out that these properties are quite fundamental to things we want to call "multiplication".

Just using these two properties of bilinearity and invariance, we will be able to find exactly one expression that is deserving to be called vector-to-vector multiplication. The first thing we should do to try to understand this operation, is of course to go to the standard basis vectors and see what the operation does to them: e.g. what is $\hat{\imath} \times \hat{\jmath}$? Well, consider the transformation $\mathcal{M}$ corresponding to a 180 degree rotation around 0 (or you could say a 180 degree rotation around the k-axis). Using our properties, we find that

$$\mathcal{M}\hat{\imath} \times \mathcal{M}\hat{\jmath} = -\hat{\imath} \times -\hat{\jmath} = \hat{\imath} \times \hat{\jmath} = \mathcal{M}(\hat{\imath} \times \hat{\jmath})$$

In 2 dimensions, there is *no* vector that satisfies the above equality (except $\vec{0}$, but we're ignoring that — we don't want two non-zero vectors to cross to $\vec{0}$). In 4 dimensions and above, there are simply way too many vectors that satisfy the equality (and hence our operation would not be uniquely defined). But in 3 dimensions, there is exactly one vector (or more precisely, one vector and its scalar multiples) that satisfy $\hat{\imath} \times \hat{\jmath} = \mathcal{M}(\hat{\imath} \times \hat{\jmath})$: $\hat{k}$. For the sake of ease, we'll pick just the unit vector $\hat{k}$, because 1 is a very natural constant to use when defining things arbitrarily.

At this point, it is clear that had we defined our operation to be invariant under all linear transformations, we would have gotten two contradictory answers: considering the linear transformation which stretches $\hat{\imath}$ to $2\hat{\imath}$, we have on one hand that $\mathcal{M}\hat{\imath} \times \mathcal{M}\hat{\jmath} = \mathcal{M}\hat{k} \implies 2\hat{\imath} \times \hat{\jmath} = \hat{k}$, but on the other hand (by bilinearity), $2\hat{\imath} \times \hat{\jmath} = 2\hat{k}$. For this reason, we must restrict ourselves to only invariance under rotations, which is "good enough" for us. I would also say it's a very intuitive property; you want the picture to look the same no matter how you tilt your head. To help you form a picture in your mind of what these vectors look like in 3 dimensional space, note that for any two vectors $\vec{a}$ and $\vec{b}$, the above 180 degree rotation that sends $\vec{a}$ and $\vec{b}$ to $-\vec{a}$ and $-\vec{b}$ respectively, satisfies $\vec{a} \times \vec{b} = \mathcal{M}(\vec{a} \times \vec{b})$, which implies that $\vec{a} \times \vec{b}$ must be perpendicular to both $\vec{a}$ and $\vec{b}$.

With this first identity $\hat{\imath} \times \hat{\jmath} = \hat{k}$, we can perform two more rotations: a 120 degree rotation around $\hat{\imath} + \hat{\jmath} + \hat{k}$ applied once and twice to the first identity (for a total of 3 identities), and a 180 degree rotation around $\hat{\imath} + \hat{\jmath}$ applied to each of the previous three identities, to yield 6 identities (grouped in the way I described them above):

$$\hat{\imath} \times \hat{\jmath} = \hat{k} \qquad \hat{\jmath} \times \hat{k} = \hat{\imath} \quad \hat{k} \times \hat{\imath} = \hat{\jmath} \qquad \hat{\jmath} \times \hat{\imath} = -\hat{k} \quad \hat{k} \times \hat{\jmath} = -\hat{\imath} \quad \hat{\imath} \times \hat{k} = -\hat{\jmath}$$

which you can verify by visualizing the transformations yourself. Finally, we need to determine the value of $\hat{\imath} \times \hat{\imath}$. Rotating 180 degrees around the $\hat{\jmath}$ axis yields $-\hat{\imath} \times -\hat{\imath} = (-1)^2 \hat{\imath} \times \hat{\imath} = \hat{\imath} \times \hat{\imath}$, which means that a rotation around the $\hat{\jmath}$ axis doesn't affect the result. Thus, the resulting vector must be some multiple of $\hat{\jmath}$. However, if we do the same rotation but around the $\hat{k}$ axis, we find that the resulting vector is some multiple of $\hat{k}$ too! The only vector that is both a multiple of $\hat{\jmath}$ and $\hat{k}$ is $\vec{0}$. Similarly, we find that

$$\hat{\imath} \times \hat{\imath} = \hat{\jmath} \times \hat{\jmath} = \hat{k} \times \hat{k} = \vec{0}$$

With these identities, we can calculate the cross product of any two vectors:

$$\vec{a} \times \vec{b} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = (a_1\hat{\imath} + a_2\hat{\jmath} + a_3\hat{k}) \times (b_1\hat{\imath} + b_2\hat{\jmath} + b_3\hat{k})$$

which when we distribute (by the properties of bilinearity) and simplify, yields

$$(a_2b_3 - a_3b_2)\hat{\imath} - (a_1b_3 - a_3b_1)\hat{\jmath} + (a_1b_2 - a_2b_1)\hat{k} \quad \square$$

2 $\vec{a} \cdot \vec{b}$

Now look above again at our "goals" in creating the cross product: we wanted a bilinear, invariant operation that returned a vector. We found that in 3 dimensions, we could uniquely define such an operation (up to a scalar multiple). However, if we tweak the goals to now return a *number* instead of a vector, we get some other notion of vector multiplication (that is perhaps even more interesting).

We have to tweak our definition of "invariance" if we focus on numbers only — this time it's not the picture that stays the same, but rather just the resulting number that should stay the same. In other words, if we notate this operation as $\vec{a} \cdot \vec{b}$, we want that for any rotation $\mathcal{M}$,

$$\mathcal{M}\vec{a} \cdot \mathcal{M}\vec{b} = \vec{a} \cdot \vec{b}$$

Spoiler alert: this operation can be defined in any dimension, so let's just say that we are working in $\mathbb{R}^n$ with the standard basis vectors

$$\hat{e}_1 = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} \quad \cdots \quad \hat{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$$

Now imagine we have $\hat{e}_i$ and $\hat{e}_j$. Consider the rotation $\mathcal{M}$ that we perform by rotating $\hat{e}_i$ to $-\hat{e}_i$ around the $\hat{e}_j$-axis. You can visualize this in three dimensions e.g. by rotating $\hat{\imath}$ to $-\hat{\imath}$ by spinning the xy-plane around the z-axis. With this $\mathcal{M}$, we have that

$$\hat{e}_i \cdot \hat{e}_j = \mathcal{M}\hat{e}_i \cdot \mathcal{M}\hat{e}_j = -\hat{e}_i \cdot e_j = -(\hat{e}_i \cdot e_j) \implies \hat{e}_i \cdot \hat{e}_j = 0$$

Furthermore, notice that we can always rotate any given $\hat{e}_i$ to any $\hat{e}_j$, and so for any i and j,

$$\hat{e}_i \cdot \hat{e}_i = \mathcal{M}\hat{e}_i \cdot \mathcal{M}\hat{e}_i = \hat{e}_j \cdot \hat{e}_j$$

Therefore,

$$\hat{e}_1 \cdot \hat{e}_1 = \hat{e}_2 \cdot \hat{e}_2 = \ldots = \hat{e}_n \cdot \hat{e}_n = c$$

Again like above, we choose 1 to be our arbitrary (but natural!) constant c. And so generally, we have that

$$\vec{a} \cdot \vec{b} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = (a_1\hat{e}_1 + \ldots + a_n\hat{e}_n)(b_1\hat{e}_1 + \ldots + b_n\hat{e}_n) = a_1b_1 + \ldots + a_nb_n \quad \square$$

3 $\det(\mathcal{M})$

Disclaimer: this is mostly taken from the resource cited in the "References" section (found below), but with some edits.

I'll start by giving the algebraic definition of the determinant (which is essentially just the formula for how to calculate it): given a matrix

$$\mathcal{M} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix},$$

the determinant of $\mathcal{M}$ is

$$\det(\mathcal{M}) = \sum_{\vec{p}} \sigma(\vec{p}) \cdot a_{1p_1} \cdot a_{2p_2} \cdots a_{np_n}$$

where $\vec{p} = (p_1, p_2, \cdots, p_n)$ is a rearrangement/permutation of the numbers $\{1, \dots, n\}$, and $\sigma(\vec{p})$ is the "signature" or "parity" (or "sign"!) of that arrangement, defined as $(-1)^k$, where k is the number of times that pairs of numbers in $\vec{p}$ have to be swapped in order to get to the arrangement $(1, 2, \cdots, n)$. (Quiz time: why is it impossible to find another sequence of swaps with a different parity? What is an algorithm to determine the smallest number of swaps necessary to go from any $\vec{p}$ to $(1, 2, \dots, n)$?)

The original resource had this example, so I will indulge the reader: if we e.g. have

$$\mathcal{M} = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} = \begin{pmatrix} 4 & 2 & 1 \\ 2 & 7 & 3 \\ 5 & 2 & 2 \end{pmatrix},$$

then

$$\begin{aligned} \det(\mathcal{M}) &= \sum_{\vec{p}} \sigma(\vec{p}) a_{1p_1} a_{2p_2} a_{3p_3} \\ &= \begin{cases} \sigma(1,2,3) a_{11} a_{22} a_{33} + \sigma(1,3,2) a_{11} a_{23} a_{32} + \sigma(2,1,3) a_{12} a_{21} a_{33} \\ + \sigma(2,3,1) a_{12} a_{23} a_{31} + \sigma(3,1,2) a_{13} a_{21} a_{32} + \sigma(3,2,1) a_{13} a_{22} a_{31} \end{cases} \\ &= a_{11} a_{22} a_{33} - a_{11} a_{23} a_{32} - a_{12} a_{21} a_{33} + a_{12} a_{23} a_{31} + a_{13} a_{21} a_{32} - a_{13} a_{22} a_{31} \\ &= 4 \cdot 7 \cdot 2 - 4 \cdot 2 \cdot 3 - 2 \cdot 2 \cdot 2 + 2 \cdot 2 \cdot 1 + 5 \cdot 2 \cdot 3 - 5 \cdot 7 \cdot 1 \\ &= 23 \end{aligned}$$

(More quiz time: verify for yourself that the other algebraic definition — of taking each element of the first row, multiplying it by the determinant of the submatrix, and alternatively adding and subtracting; see here — gives the formula above).

Now you know how to compute the determinant of any matrix. But look at the formula again — do you see anything? I certainly do not. As it is, it is just a computation; no intuition, no meaning, and to be frank, not beautiful at all. So let me define the determinant *geometrically*: for any $n \times n$-matrix $\mathcal{M}$, the determinant is the n-dimensional volume of the parallelpiped built by the vectors $\vec{v}_1, \dots, \vec{v}_n$ that are the columns of that matrix.

Why would we define something like this? Two hugely important reasons:

- The determinant is 0 if and only if the matrix, when viewed as a linear transformation (i.e. a linear transformation stretching n-dimensional space by stretching all the basis vectors $\hat{e}_i$ to the column vector $\vec{v}_i$), transforms n-dimensional space into a lower dimensional space. For example, the matrix

$$\mathcal{M} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

flattens the 2-dimensional plane onto a 1-dimensional line. If you've taken any linear algebra, you'll know that a linear transformation flattens n-dimensional space into some lower dimension if and only if the column vectors are linearly dependent. Furthermore, a linear transformation is invertible if and only if it keeps everything in n-dimensional space (intuitively if a transformation flattens to a lower dimension, many vectors map to the same vector, so there is no way to invert the transformation). In other words the determinant is an computational way to determine if a matrix/linear transformation is invertible.

Furthermore, a non-invertability (and hence a zero determinant) *also* means that a transformation sends some non-zero vector to 0, which paves the way to eigenvectors and eigenvalues.

- Secondly, again using the linear transformation perspective on matrices, the determinant is the volume of the transformed unit n-dimensional cube, i.e. the volume of the parellelpiped that the unit cube was stretched into by the transformation. And because the transformation is linear, if we know the factor by which the n-dimensional unit cube (with n-dimensional volume 1) is transformed, we know the factor by which *all shapes* in $\mathbb{R}^n$ are transformed, be it a sphere or a blob. (If you've taken multivariable calculus, you know that knowing the factor by which areas/volumes are scaled is extremely important when changing coordinates for higher-dimensional integration.)

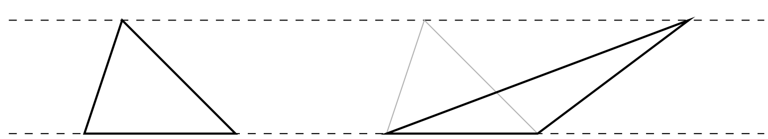
I advise you to check out 3Blue1Brown's excellent video on determinants for some visual intuition.

Let us denote $D(\vec{v}_1, \ldots, \vec{v}_n)$ to be the volume of the parellelpiped created by the vectors $\vec{v}_1, \ldots, \vec{v}_n$. We now list out some important properties of D that we can use to find a formula for it:

(i) $D(\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n) = 1$, because a $1 \times 1 \times \cdots 1$ box has a volume of 1.

(ii) $D(\vec{v}_1, \ldots, \vec{v}_n) = 0$ whenever $\{\vec{v}_1, \ldots, \vec{v}_n\}$ are linearly dependent (corresponding to when the parallelpiped is "flat").

(iii) If $\{\vec{v}_1, \ldots, \vec{v}_n\}$ are linearly independent and $\vec{w}$ is in the span of (i.e. is a linear combination of) $\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_n\}$ (i.e. all n vectors except for $\vec{v}_i$), then $D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n) = D(\vec{v}_1, \ldots, \vec{v}_i + \vec{w}, \ldots, \vec{v}_n)$. This is because moving the "top" of the parellelpiped in a direction parallel to the "base" does not change the volume, similar to how the two triangles below have the same area even though the top vertex of the second one is shifted right:



(iv) $D(\vec{v}_1, \ldots, a\vec{v}_i \ldots, \vec{v}_n) = aD(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n)$, which you can think of intuitively as saying that "doubling the length of one side doubles the volume".

Slight caveat: what if we have negative values for a? Then with the above property we would have something like $D(-\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n) = -D(\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n) = -1$, which doesn't make much intuitive sense (I mean, what is a "negative volume"?). Well, it turns out that the first and most intuitive idea you would think of for fixing the property, i.e. making it $D(\vec{v}_1, \ldots, a\vec{v}_i \ldots, \vec{v}_n) = |a| D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n)$ instead, is quite bad, for reasons I'll get to in a minute; and so I'm afraid this is how the property must go, and hence, we'll have to work out what negative volumes mean later.

(v) We now come to the most important property of D — linearity: $D(\vec{v}_1, \ldots, \vec{v}_i + \vec{w}, \ldots, \vec{v}_n) = D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n) + D(\vec{w}, \ldots, \vec{w}, \ldots, \vec{v}_n)$. This comes pretty easily from properties (iii) and (iv) and the fact that we can write $\vec{w} = a\vec{v}_i + \vec{u}$ where $a \in \mathbb{R}$ and $\vec{u}$ is in the span of $\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_n\}$:

$$\begin{aligned} D(\vec{v}_1, \ldots, \vec{v}_i + \vec{w}, \ldots, \vec{v}_n) &= D(\vec{v}_1, \ldots, \vec{v}_i + a\vec{v}_i + \vec{u}, \ldots, \vec{v}_n) \\ &= D(\vec{v}_1, \ldots, \vec{v}_i + a\vec{v}_i, \ldots, \vec{v}_n) = D(\vec{v}_1, \ldots, (1+a)\vec{v}_i, \ldots, \vec{v}_n) \\ &= (1+a) D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n) \\ &= D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n) + a D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n) \\ &= D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n) + D(\vec{v}_1, \ldots, a\vec{v}_i, \ldots, \vec{v}_n) \\ &= D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n) + D(\vec{v}_1, \ldots, a\vec{v}_i + \vec{u}, \ldots, \vec{v}_n) \\ &= D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_n) + D(\vec{v}_1, \ldots, \vec{w}, \ldots, \vec{v}_n) \end{aligned}$$

Back to the quick caveat from above — the reason why we won't use the absolute value is because if we do, then D fails to be linear; e.g. if we made D both obey the absolute value version of property (iv) *and* be linear, we would have that

$$\begin{aligned} D(\vec{v}_1, \ldots, \vec{v}_n) &= D(2\vec{v}_1 + 1(-\vec{v}_1), \ldots, \vec{v}_n) \\ &= D(2\vec{v}_1 + 1(-\vec{v}_1), \ldots, \vec{v}_n) \\ &= D(2\vec{v}_1, \ldots, \vec{v}_n) + D(-1\vec{v}_1, \ldots, \vec{v}_n) \\ &= 2D(\vec{v}_1, \ldots, \vec{v}_n) + D(\vec{v}_1, \ldots, \vec{v}_n) \\ &= 3D(\vec{v}_1, \ldots, \vec{v}_n) \end{aligned}$$

(vi) With all that in place, let's establish one final property — switching two vectors in the determinant swaps the sign:

$$\begin{aligned} & D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_j, \ldots, \vec{v}_n) & \\ &= D(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_j + \vec{v}_i, \ldots, \vec{v}_n) & \text{from (iii)} \\ &= D(\vec{v}_1, \ldots, \vec{v}_i - (\vec{v}_j + \vec{v}_i), \ldots, \vec{v}_j + \vec{v}_i, \ldots, \vec{v}_n) & \text{from (iii) again} \\ &= D(\vec{v}_1, \ldots, -\vec{v}_j, \ldots, \vec{v}_j + \vec{v}_i, \ldots, \vec{v}_n) & \\ &= -D(\vec{v}_1, \ldots, \vec{v}_j, \ldots, \vec{v}_j + \vec{v}_i, \ldots, \vec{v}_n) & \text{from (iv)} \\ &= -\big(D(\vec{v}_1, \ldots, \vec{v}_j, \ldots, \vec{v}_j, \ldots, \vec{v}_n) + D(\vec{v}_1, \ldots, \vec{v}_j, \ldots, \vec{v}_i, \ldots, \vec{v}_n)\big) & \text{from (v)} \\ &= -D(\vec{v}_1, \ldots, \vec{v}_j, \ldots, \vec{v}_i, \ldots, \vec{v}_n) & \text{from (ii)} \end{aligned}$$

Using the above properties, we can figure out an explicit formula for calculating $D(\vec{v}_1, \ldots, \vec{v}_n)$ by writing all $\vec{v}_i$ as $\vec{v}_i = a_{i1}\hat{e}_1 + \ldots + a_{in}\hat{e}_n$:

$$\begin{aligned} D(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) &= D(a_{11}\hat{e}_1 + \ldots + a_{1n}\hat{e}_n, \vec{v}_2, \ldots, \vec{v}_n) \\ &= D(a_{11}\hat{e}_1, \vec{v}_2, \ldots, \vec{v}_n) + \ldots + D(a_{1n}\hat{e}_n, \vec{v}_2, \ldots, \vec{v}_n) \\ &= a_{11}D(\hat{e}_1, \vec{v}_2, \cdots, \vec{v}_n) + \ldots + a_{1n}D(\hat{e}_n, \vec{v}_2, \ldots, \vec{v}_n) \\ &= \sum_{i=1}^{n} a_{1i} \cdot D(\hat{e}_i, \vec{v}_2, \ldots, \vec{v}_n) \end{aligned}$$

Similarly if we decompose $\vec{v}_2$, we get that

$$D(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{1i} \cdot a_{2j} \cdot D(\hat{e}_i, \hat{e}_j, \vec{v}_3, \ldots, \vec{v}_n)$$

The same thing can be done to the rest of the vectors; but rather than writing n different summations, we can instead write $D(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ as

$$\sum_{\vec{p}} a_{1p_1} \cdot a_{2p_2} \cdots a_{np_n} \cdot D(\hat{e}_{p_1}, \ldots, \hat{e}_{p_n}), \text{ summing over all } \vec{p} \in \{1, \ldots, n\}^n$$

But if $\vec{p}$ contains repeat numbers, then the D portion would be zero (by property (ii)), and so we only need to sum over all rearrangements/permutations of $\{1, \ldots, n\}$. And from property (vi), we know that $D(\hat{e}_{p_1}, \ldots, \hat{e}_{p_n})$ is simply $(-1)^k D(\hat{e}_1, \ldots, \hat{e}_n) = (-1)^k$, where k is the number of times you need to swap terms to get to $D(\hat{e}_1, \ldots, \hat{e}_n)$; in other words, $D(\hat{e}_{p_1}, \ldots, \hat{e}_{p_n}) = \sigma(\vec{p})$. And thus we have derived our formula from earlier:

$$\det(\mathcal{M}) = \sum_{\vec{p}} a_{1p_1} \cdot a_{2p_2} \cdots a_{np_n} \cdot \sigma(\vec{p}), \text{ summing over all permutations of } \{1, \ldots, n\}$$

References:

- Motivation for Cross Products, Math Stack Exchange
- Motivation for Determinants, Ask a Physicist
- Yet another perspective on dot products: Dot Products and Duality, 3Blue1Brown