Long Live Determinants!

Will Dana

July 17, 2019

1 Determinant as Volume

We consider a linear map $f : V \to V$ from an n-dimensional vector space to itself. The base field of the vector space will not be all that important, although for this section we'll take it to be $\mathbb{R}$. In future sections, we'll default to $\mathbb{C}$, which has the feature of being algebraically closed; however, in most cases this won't matter, and we'll mention explicitly when it does.

If we choose a basis of V, we can identify it with $\mathbb{R}^n$ and describe f by an $n \times n$ matrix. The determinant is a quantity we will associate to that matrix. To start motivating the definition, though, we'll focus on f as a transformation applied to a space.

One way we could extract simple information from the map f is by asking how much it stretches or shrinks space. Of course, a linear transformation might stretch space in one direction and squeeze it in another. But if we want to sum up the entire transformation in one number, the following definition is a good place to start.

Slightly Wrong Definition. *Let $C := [0,1]^n \subset \mathbb{R}^n$ be the unit hypercube in $\mathbb{R}^n$. The determinant* $\det f$ *of a linear map $f : \mathbb{R}^n \to \mathbb{R}^n$ is (not quite) the hypervolume of $f(C)$.*

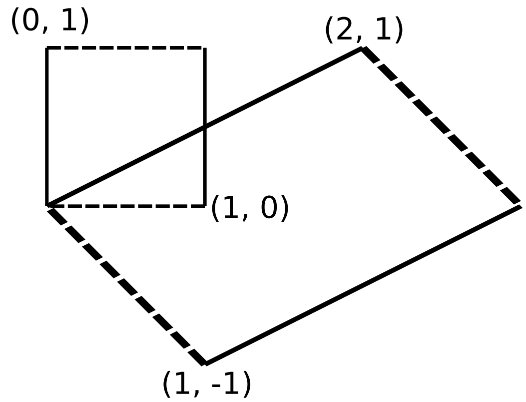
The reason this definition is slightly wrong is that it only produces nonnegative numbers, and as long as we're trying to cram information about a matrix into a single number, we might as well use its sign. We'll soon convey a notion of orientation using the sign, but for the moment, let's just think about this volume, which is the absolute value of the determinant, $|\det f|$.

Let's look at an example of this that I can actually put on paper.

Example. Consider the matrix

$$\begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}$$

Then we apply this to the unit square in the plane. The result:



The x-axis splits this parallelogram into two triangles of base 3 and height 1. Thus the parallelogram has volume 3, and that is the absolute value of the determinant of this matrix.

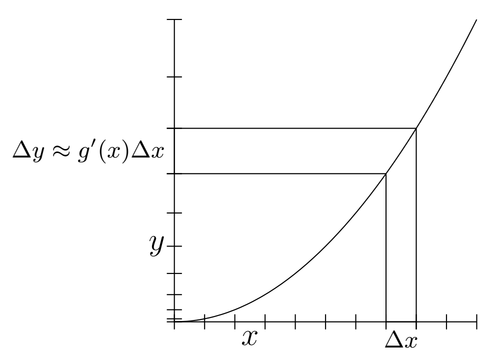
This example also illustrates a slightly different perspective on the volume interpretation: we're looking at the parallelogram with edges given by the two columns of the matrix, since these columns are where the standard basis vectors get sent.

1.1 Application: Change of Variables in Multiple Integrals

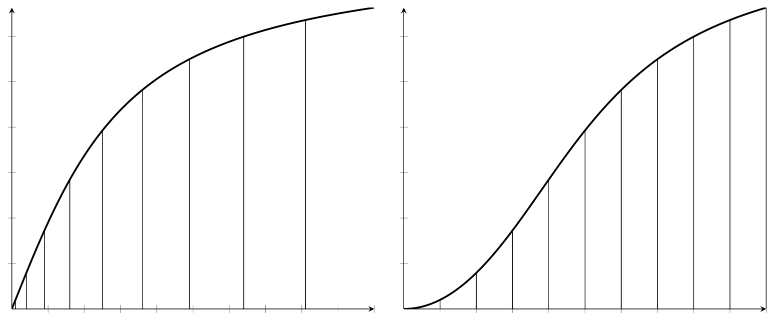
The single-variable substitution rule for integrals can be phrased like this:

$$\int_{g(a)}^{g(b)} f(u)\,du = \int_a^b f(g(x))g'(x)\,dx$$

The $g'(x)$ factor tells us, at each point x, how much our transformation g is stretching the real line:



If we imagine the integral as approximated by a Riemann sum, this stretching stretches the bases of our rectangles, which is where the $g'(x)$ comes from.



Now, if we have a function $f(x_1, \ldots, x_n)$ on $\mathbb{R}^n$, we can integrate it over a region S of n-dimensional space. We can transform the region of integration by applying a map $g = (g_1, \ldots, g_n) : \mathbb{R}^n \to \mathbb{R}^n$. Then we'd expect some sort of formula like

$$\int_{g(S)} f(u_1, \ldots, u_n)\, du_1 \cdots du_n = \int_S f(g(x_1, \ldots, x_n)) \cdot ?\, dx_1 \cdots dx_n$$

as above. What should this question mark be?

Well, for our function g to differentiable means that on sufficiently small regions of its domain, it appears linear. Specifically, if we're at a point $(z_1, \ldots, z_n)$ and increment z_i by a tiny amount Δx_i, then the jth coordinate of $g(z_1, \ldots, z_n)$ changes by approximately $\frac{\partial g_j}{\partial x_i}(z_1, \ldots, z_n) \cdot \Delta x_i$. So if we imagine a tiny hypercube based at the point $(z_1, \ldots, z_n)$, then the tiny edges in each of the coordinate directions will be sent to the vectors

$$\begin{pmatrix} (\partial g_1/\partial x_1)(z_1, \ldots, z_n) \\ (\partial g_2/\partial x_1)(z_1, \ldots, z_n) \\ \vdots \\ (\partial g_n/\partial x_1)(z_1, \ldots, z_n) \end{pmatrix}, \begin{pmatrix} (\partial g_1/\partial x_2)(z_1, \ldots, z_n) \\ (\partial g_2/\partial x_2)(z_1, \ldots, z_n) \\ \vdots \\ (\partial g_n/\partial x_2)(z_1, \ldots, z_n) \end{pmatrix}, \ldots, \begin{pmatrix} (\partial g_1/\partial x_n)(z_1, \ldots, z_n) \\ (\partial g_2/\partial x_n)(z_1, \ldots, z_n) \\ \vdots \\ (\partial g_n/\partial x_n)(z_1, \ldots, z_n) \end{pmatrix}$$

So the stretching factor that will occur near this point of the space is given (up to sign) by the determinant of the **Jacobian matrix**

$$\begin{pmatrix} \partial g_1/\partial x_1 & \partial g_1/\partial x_2 & \cdots & \partial g_1/\partial x_n \\ \partial g_2/\partial x_1 & \partial g_2/\partial x_2 & \cdots & \partial g_2/\partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial g_n/\partial x_1 & \partial g_n/\partial x_2 & \cdots & \partial g_n/\partial x_n \end{pmatrix} (z_1, \ldots, z_n)$$

and that's what works in the question mark in the formula above!

For example, we convert from polar to rectangular coordinates by the formula $x = r\cos\theta, y = r\sin\theta$. Define our map g by $g(r, \theta) = (r\cos\theta, r\sin\theta)$: in particular, if T is a region described in rectangular coordinates, then $g^{-1}(T)$ gives the corresponding inequalities defining it in polar coordinates. The Jacobian matrix associated to these coordinates is

$$\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

As we'll see soon (or as you can calculate if you already know how to take a 2×2 determinant, or as the following exercise will show), the determinant of this matrix is r.

Exercise 1. *Using the definition of determinant as volume given above, verify that the determinant of this Jacobian matrix is r.*

And indeed, the formula for changing a double integral to polar coordinates says

$$\iint_T f(x, y)\, dx\, dy = \iint_{g^{-1}(T)} f(r\cos\theta, r\sin\theta)\, r\, dr\, d\theta.$$

1.2 Linear Dependence and Collapsing

This volume perspective also illustrates one of the most important facts about determinants.

Proposition 1. *The determinant of a matrix is 0 if and only if it does not have full rank n.*

Proof. For our matrix to have full rank means that its columns span the entire space. If they don't, then the image of the matrix lies inside a subspace of smaller dimension. But then the hypervolume of the image of the unit cube will be 0.

Conversely, if the matrix does have full rank, then the images of the standard basis vectors span the entire space. If we tile the space with copies of the unit hypercube, then apply our matrix, the resulting sets tile the entire space. That is, we get a countable collection of sets, all with the same volume, whose union is the entire space; the volume of these sets can't then be 0. □

Of course, the key theorems of linear algebra give us many other ways of describing what it means for the determinant to vanish.

Corollary 1. *The determinant of a matrix is 0 if and only if:*

- *The columns are linearly dependent.*
- *The rows are linearly dependent.*
- *The matrix is not invertible.*

In particular, keep the first of these in mind for the next section.

2 Determinant as a Multilinear Function

This idea of determinant as volume is excellent for baseline intuition, but it's not very useful for what's to come:

- Actually computing any determinants this way, especially in higher dimensions, is a mess.
- To reason formally about the properties of the determinant, we need to be able to reason formally about volume. It's often easier to do the former first and then use it to help with the latter.
- We've said the sign of the determinant should capture an idea of "orientation", but we haven't defined what this actually means.

So now, inspired by volume, we'll develop a more algebraic definition of the determinant. We'll define it as a function $\det : (\mathbb{R}^n)^n \to \mathbb{R}$ on an n-tuple of n-dimensional vectors, and then say that the determinant of a matrix is given by applying this function to its columns. If we want the determinant to describe volume, what properties should this function have?

Observation 1. *For any constant a, we should have*

$$\det(v_1, \ldots, av_i, \ldots, v_n) = a \det(v_1, \ldots, v_i, \ldots, v_n)$$

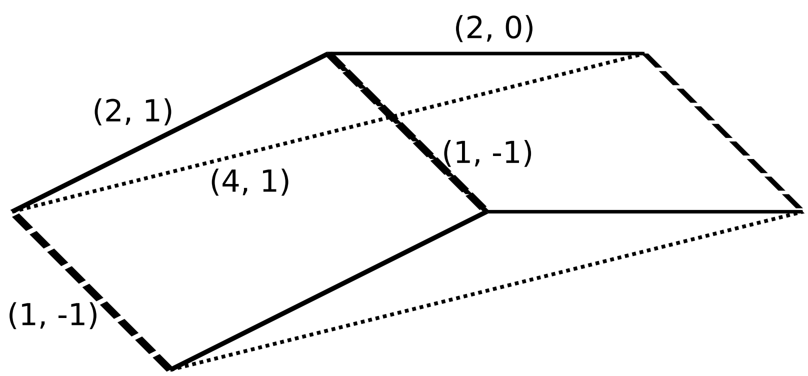
Indeed, this just corresponds to stretching the image of the unit hypercube in the x_i direction by a factor of a, which should multiply the volume by a. More subtle is the following:

Observation 2. *For any two vectors v_i and w_i, we should have*

$$\det(v_1, \ldots, v_i + w_i, \ldots, v_n) = \det(v_1, \ldots, v_i, \ldots, v_n) + \det(v_1, \ldots, w_i, \ldots, v_n)$$

Let's illustrate why we want this to hold with an example, showing

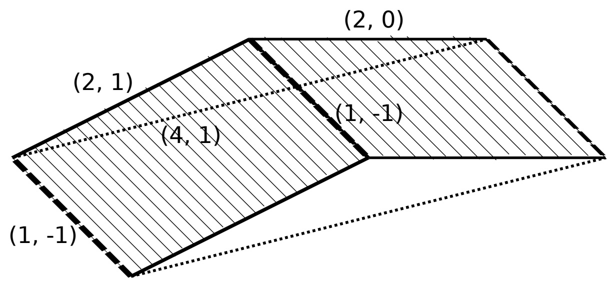
$$\det \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} + \det \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \det \begin{pmatrix} 1 & 4 \\ -1 & 1 \end{pmatrix}$$



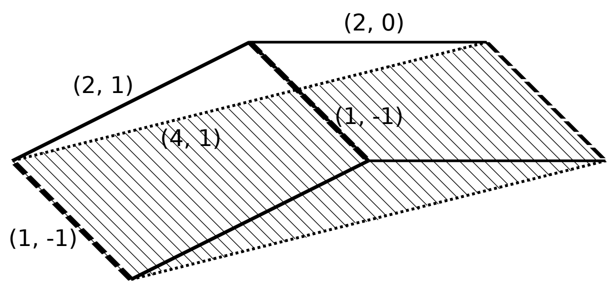
In the diagram above, the first summand is the area of the parallelogram with solid edges on the left. The second summand is the area of the parallelogram with solid edges on the right. The determinant on the right side is the area of the long parallelogram with dotted edges.

We've set up the diagram using the standard method of depicting the sum of two vectors: joining $(2,1)$ and $(2,0)$ end to end and drawing the vector with coordinates $(4,1)$ between the endpoints. We see that the combined areas of the first two parallelograms can be converted into the area of the third by cutting a triangle off the top and reattaching it on the bottom, analogously to how one computes the area of a single parallelogram.

Another way of looking at this is that we can break down each of the smaller parallelograms as a union of many thin slices, each with the magnitude and direction of the vector $(1,-1)$:



Then we can slide all of those slices down in the $(1,-1)$ direction such that they cover the long parallelogram instead:



Exercise 2. *Find some vectors v_1, w_1, v_2 such that, when drawn in this way, the parallelograms given by v_1, v_2 and w_1, v_2 overlap. What statement can we then make about the area of the parallelogram given by $v_1 + w_1, v_2$?*

Together, the previous two observations say that the determinant should be **multilinear**: if we fix all but one of the inputs, then the determinant should be a linear map on the remaining input.

Our next couple of observations come from considering the circumstances under which the determinant vanishes. In general, applying the determinant to a linearly dependent collection of vectors should give 0. Of course, a very particular kind of linear dependence is when two of the vectors are the same.

Observation 3. *If $v_i = v_j$ for some pair $i \neq j$, we should have*

$$\det(v_1, \ldots, v_n) = 0.$$

Exercise 3. *Prove directly (without using the later results in this section) that if a map $f : (\mathbb{R}^n)^n \to \mathbb{R}$ is multilinear, and is 0 whenever two of its arguments are the same, then it will actually be 0 whenever the arguments are linearly dependent. (So this observation actually covers all the cases in which the vectors are linearly dependent.)*

So far, the observations we have made come from thinking about volumes. But they have an elementary yet surprising consequence which forces how the sign of the determinant must behave.

Lemma 1. *Let $B : V \times V \to \mathbb{R}$ be any bilinear map on pairs of vectors in a vector space V, such that $B(v, v) = 0$ for any $v \in V$. Then*

$$B(w, v) = -B(v, w).$$

Proof.

$$\begin{aligned} 0 &= B(v + w, v + w) && \text{(by assumption)} \\ &= B(v, v + w) + B(w, v + w) && \text{(linearity in the first argument)} \\ &= B(v, v) + B(v, w) + B(w, v) + B(w, w) && \text{(linearity in the second argument)} \\ &= B(v, w) + B(w, v) && \text{(by assumption).} \end{aligned}$$

□

Observation 4 (Corollary to Observation 3)**.** *For any pair of indices $i \neq j$,*

$$\det(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -\det(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n)$$

In general, we say that a map with this property (that switching any two of its arguments reverses the sign) is **alternating**.

So we've made some observations about how determinants relate to each other. There's also one particular value of the determinant we can be pretty sure about:

Observation 5. *Let $e_1, \ldots, e_n$ be the standard basis vectors. Then*

$$\det(e_1, \ldots, e_n) = 1$$

After all, the identity matrix shouldn't change the volume of a unit hypercube.

This last observation may seem a bit trivial. Why did we put it in? Because together, the observations we've made about the determinant are actually enough to uniquely characterize it!

Theorem 1. *The determinant is the only multilinear, alternating map $(\mathbb{R}^n)^n \to \mathbb{R}$ which is 1 when given the standard basis vectors in order.*

Proof. We'll prove this theorem by assuming that f is such a map and showing that three properties listed are already enough to compute all of its values.

Given any input to the map, we can write each vector in terms of the standard basis vectors:

$$\begin{aligned} f(v_1, \ldots, v_n) = f(&a_{11}e_1 + \ldots + a_{1n}e_n, \\ &a_{21}e_1 + \ldots + a_{2n}e_n, \\ &\ldots, \\ &a_{n1}e_1 + \ldots + a_{nn}e_n) \end{aligned}$$

Then we can apply linearity at each of the inputs in turn. For example, we can write

$$\begin{aligned} &f(a_{11}e_1 + \ldots + a_{1n}e_n, v_2, \ldots, v_n) \\ &= a_{11}f(e_1, v_2, \ldots, v_n) + \ldots + a_{1n}f(e_n, v_2, \ldots, v_n) \end{aligned}$$

and then, in each of the resulting terms, expand out v_2 as a combination of the standard basis vectors, and so forth. The end result is that we can describe $f(v_1, \ldots, v_n)$ as a (very long) linear combination of the values of f with various combinations of e_i as inputs.

Thus it suffices to know the value of $f(e_{i_1}, e_{i_2}, \ldots, e_{i_n})$ for indices $i_1, \ldots, i_n$. But if any two of the indices are equal, we know f must be 0. If the indices are all distinct, then they are a permutation of 1 through n. We know that, by swapping any two of the inputs, we get the same value but with the sign flipped. So by an appropriate sequence of swaps, we can rearrange the inputs until we get $f(e_1, \ldots, e_n)$, which we know to be 1. Backtracking the reductions we did to get here gives the value of $f(v_1, \ldots, v_n)$. □

This theorem allows us to give an elegant, if abstract, definition of the determinant:

Definition. *The determinant is the unique map specified by the above theorem.*

Now we can say a bit about what the sign of the determinant should be. As we've seen, it's essentially forced by the definition. But what geometric significance does the sign have in lower dimensions?

In two dimensions, note that in the unit square, e_2 lies $90°$ counterclockwise[1] from e_1. In particular, it's less than $180°$ counterclockwise from e_1. If we exchange e_1 and e_2, the volume remains the same, but this relationships is swapped, and now e_2 is more than $180°$ clockwise from e_1 (and thus closer on the clockwise side).

In general, the sign of the determinant will be positive if the image of e_2 is less than $180°$ degrees counterclockwise away from the image of e_1, and negative otherwise. This is backed up by what happens when the image of e_2 is exactly $180°$ away from the image of e_1: they are linearly dependent, and the determinant is 0.

In three dimensions, the essence of orientation is captured by how e_1 and e_2 are positioned rotationally *relative to* e_3. This is typically summed up by the

[1] That is to say, widdershins.

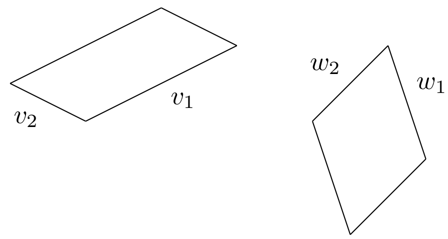


Figure 1: The sign of the determinant represented by the left parallelogram is positive. The sign of the determinant represented by the right parallelogram is negative. If we switch the order of the vectors in either case, the sign flips.

"right-hand rule". Stick out the fingers of your right hand in the direction of the image of e_1, and then curl them roughly in the direction of the image of e_2. If the direction your thumb is pointing is roughly in the direction of the image of e_3, then the determinant is positive; otherwise, it's negative[2].

3 Determinant as a Polynomial

Looking at things from the algebraic point of view, we've just shown that if there is an alternating multilinear map which is 1 on the identity matrix, it is unique. But we haven't actually shown such a map must exist[3]! And if we ever want to compute anything with the determinant (and we do!) then we'll definitely want it to exist. Ideally, we should provide an explicit formula for it.

What features should such a formula have?

- It should be a polynomial in the entries of the matrix. This isn't obvious, but it's what we'd hope for: if we view the determinant as the volume of a parallelotope, it shouldn't require anything more than multiplying and adding the entries of the matrix together.
- In each term of the polynomial, we should have one entry from each column. This is because, if we scale any one of the columns by a factor of a, the entire determinant scales by a factor of a as well. The easiest way for this to happen is for each term of the determinant to get scaled by a, and the easiest way for this to happen is for exactly one "representative" of each column to appear in each term.
- Less obviously, in each term of the polynomial, we should have one entry from each row. This is because, if we scale any one of the rows by a factor

[2]If you have some bizarre mutation that allows your fingers to bend equally well backwards and forwards, this might be difficult to use.

[3]While our proof of uniqueness laid out a strategy for computing a map with these properties, it didn't actually prove that this map is consistently multilinear and alternating.

of a, the determinant again scales by a factor of a. Geometrically, this is because scaling the ith row by a factor of a amounts to stretching the image of the unit hypercube by a in the x_i direction.

Exercise 4. *Prove, using only the properties of multilinearity and alternation, that scaling rows causes this to happen.*

- Finally, if we swap two of the vectors, or columns, being fed into the determinant, the sign should swap. The easiest way to realize this is for all of the terms of the polynomial to still be present after the swap, but with exchanged signs.

Keep these stipulations in mind when considering the definition to come. Before we describe the determinant in full, we need one smaller definition. In the proof of uniqueness above, we made brief reference to unscrambling a permutation of the standard basis vectors by repeatedly swapping elements, and keeping track of the sign changes that occur when we do this. That information is captured in the idea of sign:

Definition. *The* ***sign*** *of a permutation of n letters $\sigma \in S_n$, $\operatorname{sign}(\sigma)$, is -1 if σ can be written as the composition of an odd number of transpositions, and 1 if it can be written as the composition of an even number of transpositions.*

Exercise 5. *Prove this is well-defined. (There are a few different ways of seeing this, so if you've already seen a proof of this, try to come up with another one!)*

Definition. *Let M be an $n \times n$ matrix with entries m_{ij}. The determinant of M is given by*

$$\det(M) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} m_{i\sigma(i)}$$

For example, consider a 3×3 matrix. There is one permutation of $\{1, 2, 3\}$ requiring no transpositions (the identity), three requiring one transposition (the transpositions), and the remaining two permutations, which cycle through the numbers and can be done with two transpositions (check this). This gives the determinant

$$m_{11}m_{22}m_{33} - m_{12}m_{21}m_{33} - m_{11}m_{23}m_{32} - m_{13}m_{22}m_{31} + m_{12}m_{23}m_{31} + m_{13}m_{21}m_{32}$$

This is disgusting, but let's focus on the positives for now. First, this shows that the determinant exists, as we now check.

Theorem 2. *The determinant, as defined here, is multilinear, alternating, and 1 on the identity matrix.*

Proof. The multilinearity comes from the fact that for a fixed column of the matrix, each term contains exactly one entry from that column, as suggested above. (The details of this are best checked on your own.)

As for the the alternation, consider applying a transposition τ to the columns of the matrix. Then the expression becomes

$$\sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} m_{i\tau(\sigma(i))} = \sum_{\sigma' \in S_n} \operatorname{sign}(\tau^{-1}\sigma') \prod_{i=1}^{n} m_{i\sigma'(i)}$$

$$= \sum_{\sigma' \in S_n} -\operatorname{sign}(\sigma') \prod_{i=1}^{n} m_{i\sigma'(i)} = -\det(M).$$

Here we reindex using the fact that multiplying each element of S_n on the left by a transposition just permutes the elements of S_n.

Finally, note that if we plug in the identity matrix, the only nonzero term is the one corresponding to the identity permutation. (This is because, for any $i \neq j$, $m_{ij} = 0$.) This term is 1. □

Secondly, as with any good new definition, this shows us another property of the determinant that was not at all obvious from the definitions we presented above.

Proposition 2.

$$\det(M^T) = \det(M)$$

Proof. The sign of a permutation is the same as the sign of its inverse (since you can obtain the inverse by just applying all of the transpostions in reverse order). So we can reindex the sum over permutations as a sum over their inverses like so:

$$\sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} m_{i\sigma(i)} = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma^{-1}) \prod_{i=1}^{n} m_{\sigma^{-1}(i)i}$$

$$= \sum_{\sigma' \in S_n} \operatorname{sign}(\sigma') \prod_{i=1}^{n} m_{\sigma'(i)i} = \det(M^T)$$

□

Exercise 6. *An $n \times n$ matrix $M = (m_{ij})$ is called skew-symmetric if $m_{ij} = -m_{ji}$ for all indices i, j. Show that if n is odd, a skew-symmetric matrix cannot be invertible. Show that if n is even, it can be.*

Finally, this expression of the determinant as a sum over permutations allows for some intriguing links to combinatorics, which we may look at later in the class.

Exercise 7. *Using the determinant, show that the number of permutations with sign $+1$ and the number of permutations with sign -1 are equal.*

4 An Excursion: Determinant as a Functor

Mathematicians don't like tying themselves down to bases of vector spaces, and we prefer whenever possible to define vector space operations in a way that doesn't depend on a choice of basis. There are a couple of different reasons for this:

- If a concept doesn't depend on the bases we choose, defining it in a way that uses a choice of basis muddies what's actually important about the concept. We initially defined the determinant using a unit hypercube, but there's nothing intrinsically cubey about the determinant, so this isn't the cleanest definition.

- Although the examples of vector spaces in introductory linear algebra are frequently given with bases, vector spaces encountered in the wild may not have an obvious choice of basis. For instance, consider the tangent plane to a sphere at a particular point. There's no obvious way to choose a basis, and so having to pick one in order to implement a definition adds an unnecessary layer of complication.

However, all 3 different ways of defining the determinant above at some point involved picking a basis. In this section, we'll introduce a definition of the determinant very similar to the above, but which purposefully avoids talking about basis; as a result, one very important fact about determinants will fall out with very little effort.

Given a vector space V, we define a new vector space $\wedge^k V$, called the **kth exterior power**, as follows.

Intuitively, we want $\wedge^k V$ to be spanned by symbols $v_1 \wedge \ldots \wedge v_k$, where $v_1, \ldots, v_k \in V$. We additionally want the $\wedge$ symbol to satisfy some rules: we want the map

$$
\begin{aligned}
V^k &\to \wedge^k V \\
(v_1, \ldots, v_k) &\mapsto v_1 \wedge \ldots \wedge v_k
\end{aligned}
$$

to be k-linear (i.e., multilinear on k variables) and alternating. A more formal definition follows.

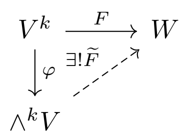
Start with the space W spanned by all formal symbols of the form $v_1 \wedge v_2 \wedge \cdots \wedge v_k$, where $v_1, \ldots, v_k \in V$. This is an infinite-dimensional (in fact, uncountably dimensional) space. But now we quotient by (the subspace spanned by) the following elements, for all choices of $v_i, w_i \in V$ and $a \in \mathbb{R}$:

$$
\begin{aligned}
&(v_1 \wedge \ldots \wedge (v_i + w_i) \wedge \ldots \wedge v_k) - (v_1 \wedge \ldots \wedge v_i \wedge \ldots \wedge v_k) - (v_1 \wedge \ldots \wedge w_i \wedge \ldots \wedge v_k) \\
&(v_1 \wedge \ldots \wedge (a v_i) \wedge \ldots \wedge v_k) - a(v_1 \wedge \ldots \wedge v_i \wedge \ldots \wedge v_k) \\
&(v_1 \wedge \ldots \wedge v_i \wedge \ldots \wedge v_j \wedge \ldots \wedge v_k) + (v_1 \wedge \ldots \wedge v_j \wedge \ldots \wedge v_i \wedge \ldots \wedge v_k)
\end{aligned}
$$

and we define this quotient to be the space $\wedge^k V$. These relations look like a mess by themselves, but quotienting out by the subspace spanned by them amounts to setting them all to 0, which is the same as enforcing the conditions of multilinearity and alternation.

Because the exterior power has the alternating k-linear property baked into it, it satisfies an important universal property: k-linear alternating maps on V^k are "the same as" ordinary linear maps on $\wedge^k V$. More precisely:

Theorem 3 (The Universal Property of Exterior Power)**.** *Let $\varphi : V^k \to \wedge^k V$ be the alternating multilinear map $(v_1, \ldots, v_k) \mapsto v_1 \wedge \ldots \wedge v_k$. Let $F : V^k \to W$ be any alternating k-linear map. Then there is a unique linear map $\widetilde{F} : \wedge^k V \to W$ such that $F = \widetilde{F} \circ \varphi$. This can be visualized in the following diagram.*



Exercise 8. *Prove this. (It's mostly a matter of chasing definitions, but it's good practice for getting familiar with those definitions.)*

What makes this universal property nice is that it gives us a way to describe k-linear maps (which, a priori, we don't understand) in terms of ordinary linear maps (which we do). As such, linear maps between exterior powers are a useful thing to consider. Here's a way to come up with a bunch of them.

Exercise 9. *Let $A : V \to V$ be a linear map from a vector space to itself. Show that there is a well-defined map*

$$\wedge^k A : \wedge^k V \to \wedge^k V$$

given by

$$(\wedge^k A)(v_1 \wedge \ldots \wedge v_k) = (Av_1) \wedge \ldots \wedge (Av_k)$$

Suppose V is n-dimensional. Then interesting things happen when we consider the vector space $\wedge^n V$.

Proposition 3. *$\wedge^n V$ is 1-dimensional.*

Proof. Our most effective way to extract information about $\wedge^n V$ is using its universal property. In particular, the space of all linear maps $\wedge^n V \to \mathbb{R}$ corresponds to the space of all alternating n-linear maps $V^n \to \mathbb{R}$. But we proved over the last couple of sections that this is a one-dimensional space: it's given by multiples of the determinant.

If $\wedge^n V$ has basis $e_1, \ldots, e_d$, then each basis vector induces a map $\wedge^n V \to \mathbb{R}$ sending that vector to 1 and all others to 0. These maps are all linearly independent. On the other hand, we just showed the space of linear maps $\wedge^n V \to \mathbb{R}$ is 1-dimensional, so $\wedge^n V$ can be at most 1-dimensional. (It can't be 0-dimensional, or it wouldn't admit any nontrivial maps to $\mathbb{R}$.) □

So for any linear map $A : V \to V$, what is the linear map $\wedge^n A : \wedge^n V \to \wedge^n V$ it induces? Well, as a map from a 1-dimensional vector space to itself, this must be multiplication by a scalar, and that scalar is exactly the determinant of A!

Exercise 10. *Check this, by using the definition of the determinant as an alternating multilinear map.*

This might seem like a rather complicated way of getting at the determinant, but it gives a nice fact which is intuitively clear from our geometric definition but fiddly to actually prove with the other algebraic ones:

Theorem 4.

$$\det(AB) = \det(A)\det(B)$$

Proof. Essentially by definition, $\wedge^k(AB) = \wedge^k A \circ \wedge^k B$ for any k. □

5 Actually Computing the Determinant

The definitions of the determinant given above, while varied, were definitely from a mathematician's perspective. We had a geometric intuition that doesn't give us any useful formulas, a definition which was contingent on proving the determinant actually exists, a polynomial with a number of terms which grows as the factorial of the size of the matrix, and a definition which explicitly avoids choosing a basis (the main tool for all computation in linear algebra). None of this is actually helpful for computing the thing.

However, because the determinant has such nice properties, there are a lot of tricks we can exploit to compute it. Specifically, we know how the elementary row (and column) operations affect the determinant.

Proposition 4.

- *Adding a multiple of a column to another column does not change the determinant.*
- *Scaling a column by a scalar scales the determinant by the same amount.*
- *Swapping two columns flips the sign of the determinant.*

Proof. The second and third points are just two of the defining properties of the determinant. As for the first, we have

$$\det(v_1, v_2 + av_1, v_3, \ldots, v_n) = \det(v_1, v_2, v_3, \ldots, v_n) + \det(v_1, av_1, v_3, \ldots, v_n)$$

and this second summand is 0 by linear dependence.

To get the same results for rows, apply the proposition to the determinant of the transpose. □

This is great, because it means we can calculate the determinant purely through Gaussian elimination. A handy shortcut is:

Proposition 5. *The determinant of an upper (or lower) triangular matrix is the product of its diagonal entries.*

Proof. We can reduce an upper triangular matrix to a diagonal matrix by subtracting off multiples of the first column to zero out the first row in the following columns, then doing the same with the second column to zero out the second row in the following columns, and so on; this does not change the determinant, and leaves the diagonal entries intact. Then the determinant is just the product of the scale factors applied to each column, which are the diagonal entries.

The corresponding result for lower triangular matrices follows from doing the same for rows. □

Example.

$$\det \begin{pmatrix} -1 & 3 & 2 \\ -3 & 4 & -1 \\ 1 & 4 & 0 \end{pmatrix} = \det \begin{pmatrix} -1 & 3 & 2 \\ 0 & -5 & -7 \\ 0 & 7 & 2 \end{pmatrix} = \det \begin{pmatrix} -1 & 3 & 2 \\ 0 & -5 & -7 \\ 0 & 0 & -\frac{39}{5} \end{pmatrix}$$
$$= (-1)(-5)(-\frac{39}{5}) = -39$$

Exercise 11. *Show that a block upper triangular matrix of the form*

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

with A, B, C square matrices has determinant $\det(A)\det(C)$.

Exercise 12. *By contrast, show that in general*

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} \neq \det(A)\det(D) - \det(B)\det(C)$$

5.1 Cofactor Expansion (Is Bad)

One way of computing the determinant is the recursive formula by cofactors. This formula is typically bad, but it's a way in which the determinant is often introduced, so it's worth seeing where it comes from. Consider an $n \times n$ matrix $A = (a_{ij})$. Using linearity in the first column, we can write

$$\det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$
$$= a_{11} \det \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{pmatrix} + \ldots + a_{n1} \det \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

Then, by subtracting multiples of the first column from the other ones, we can clear out the associated row in each matrix, to get

$$a_{11}\det\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & a_{22} & \cdots & a_{2n}\\ 0 & a_{32} & \cdots & a_{3n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & a_{n2} & \cdots & a_{nn}\end{pmatrix} + a_{21}\det\begin{pmatrix}0 & a_{12} & \cdots & a_{1n}\\ 1 & 0 & \cdots & 0\\ 0 & a_{32} & \cdots & a_{3n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & a_{n2} & \cdots & a_{nn}\end{pmatrix} + \ldots$$
$$+ a_{n1}\det\begin{pmatrix}0 & a_{12} & \cdots & a_{1n}\\ 0 & a_{22} & \cdots & a_{2n}\\ 0 & a_{32} & \cdots & a_{3n}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 0 & \cdots & 0\end{pmatrix}$$

Now in each of the determinants, we swap the row with a single 1 with the rows above it in succession. Each time we do this, the sign flips, so the sum turns into an alternating sum:

$$a_{11}\det\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & a_{22} & \cdots & a_{2n}\\ 0 & a_{32} & \cdots & a_{3n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & a_{n2} & \cdots & a_{nn}\end{pmatrix} - a_{21}\det\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & a_{12} & \cdots & a_{1n}\\ 0 & a_{32} & \cdots & a_{3n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & a_{n2} & \cdots & a_{nn}\end{pmatrix} + \ldots$$
$$+ (-1)^{n+1}a_{n1}\det\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & a_{12} & \cdots & a_{1n}\\ 0 & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & a_{(n-1)2} & \cdots & a_{(n-1)n}\end{pmatrix}$$

Finally, by Exercise 11 (or just by a careful examination of the formula for the determinant) we can ignore the 1 in the upper right corner and equate each determinant with the determinant of the remaining $(n-1)\times(n-1)$ submatrix.

At this point a little more notation comes in handy. Let A_{ij} be the $(n-1)\times(n-1)$ matrix obtained by crossing out the ith row and jth column from A.[4] Then what we've just shown is that

$$\det(A) = a_{11}\det(A_{11}) - a_{21}\det(A_{21}) + \ldots + (-1)^{n+1}a_{n1}\det(A_{n1})$$

Of course, there's nothing special about expanding the determinant using multilinearity in the first column, as we did here. We could have used the second column. This involves an additional swap, in order to move the emptied-out second column into the first one, and so there is an additional change of

[4]This notation is not standard, but it's also vague enough that I feel justified in using it this way.

sign:

$$\det(A) = -a_{12}\det(A_{12}) + a_{22}\det(A_{22}) + \ldots + (-1)^{n+2}a_{n2}\det(A_{n2})$$

And in general, for some column index m, we can say

$$\det(A) = (-1)^{1+m}a_{1m}\det(A_{1m}) + \ldots + (-1)^{n+m}a_{nm}\det(A_{nm})$$

In fact, we could just as well expand out along one of the rows instead, again using the invariance of the determinant under transpose.

$$\det(A) = (-1)^{m+1}a_{m1}\det(A_{m1}) + \ldots + (-1)^{m+n}a_{mn}\det(A_{mn})$$

The flexibility of choosing a row or column to expand along (for instance, we could choose one with mostly 0s) makes this recursive approach occasionally a handy way to compute determinants by hand. However, past size 3×3, it's still typically a mess. And this is ultimately just packaging the polynomial definition in a recursive form, so it's not good as a computer algorithm either: it runs in $O(n!)$, or "very not polynomial time"[5].

Still, cofactor expansion can have its niche uses.

Exercise 13. *Compute the determinant of every matrix of the form*

$$\begin{pmatrix} 3 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 3 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & 3 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 3 \end{pmatrix}$$

6 The Vandermonde Determinant

Here's a fact.

Theorem 5.

$$\det\begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \alpha_n^2 & \cdots & \alpha_n^{n-1} \end{pmatrix} = \prod_{1 \le i < j \le n} (\alpha_j - \alpha_i)$$

For example, evaluating this with $n = 3$ gives

$$(\alpha_2 - \alpha_1)(\alpha_3 - \alpha_1)(\alpha_3 - \alpha_2)$$

The matrix here is called the **Vandermonde matrix**, and the determinant is called the **Vandermonde determinant**. Why would we care about this matrix?

[5] This notation is not standard, but I wish it were.

Among other things, it describes the linear map which outputs a polynomial at a fixed set of points. From this interpretation, the Vandermonde determinant gives us a nice corollary, which is generally accepted but not trivial to prove:

Corollary 2. *Given any n points $(x_1, y_1), \ldots, (x_n, y_n)$ with distinct x-coordinates, there exists a unique polynomial p of degree $\leq n-1$ with $p(x_i) = y_i$ for all i.*

Proof. Let $M(\alpha_1, \ldots, \alpha_n)$ be the Vandermonde matrix. Then it is straightforward to check that for an arbitrary polynomial $p(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$,

$$M(\alpha_1, \ldots, \alpha_n) \cdot \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{pmatrix} = \begin{pmatrix} p(\alpha_1) \\ p(\alpha_2) \\ \vdots \\ p(\alpha_n) \end{pmatrix}$$

Thus finding a polynomial which passes through the given points is the same as solving the equation

$$M(x_1, \ldots, x_n) \cdot v = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Now by Theorem 5, since $x_1, \ldots, x_n$ are distinct, the determinant of $M(x_1, \ldots, x_n)$ is nonzero. Thus $M(x_1, \ldots, x_n)$ is invertible, and the system of equations has a unique solution. ☐

The Vandermonde matrix, and related matrices, show up in several other contexts. The formula for the determinant is also very pretty on its own merits. Let's prove it in three different ways.

6.1 Proof 1: Deft Column Reduction

If we attempt to directly apply Gaussian elimination to the Vandermonde determinant, we will eventually get an answer, but the computation will get very nasty. But we can be crafty about how we do it.

Start with the Vandermonde matrix, and, proceeding from right to left, subtract α_1 times the ith column from the $(i+1)$th column. (We do this in the order we do so that the column being subtracted is not changed by a previous step.)

This produces the matrix

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & \alpha_2 - \alpha_1 & \alpha_2^2 - \alpha_2\alpha_1 & \cdots & \alpha_2^{n-1} - \alpha_2^{n-2}\alpha_1 \\ 1 & \alpha_3 - \alpha_1 & \alpha_3^2 - \alpha_3\alpha_1 & \cdots & \alpha_3^{n-1} - \alpha_3^{n-2}\alpha_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n - \alpha_1 & \alpha_n^2 - \alpha_n\alpha_1 & \cdots & \alpha_n^{n-1} - \alpha_n^{n-2}\alpha_1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & \alpha_2 - \alpha_1 & \alpha_2(\alpha_2 - \alpha_1) & \cdots & \alpha_2^{n-2}(\alpha_2 - \alpha_1) \\ 1 & \alpha_3 - \alpha_1 & \alpha_3(\alpha_3 - \alpha_1) & \cdots & \alpha_3^{n-2}(\alpha_3 - \alpha_1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n - \alpha_1 & \alpha_n(\alpha_n - \alpha_1) & \cdots & \alpha_n^{n-2}(\alpha_n - \alpha_1) \end{pmatrix}$$

Then we do something similar: proceeding from right to left, subtract α_2 times the ith column from the $(i+1)$th column. This time, we stop one column shorter than last time, leaving the second column (as well as the first) untouched. This gives

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & \alpha_2 - \alpha_1 & 0 & \cdots & 0 \\ 1 & \alpha_3 - \alpha_1 & (\alpha_3 - \alpha_2)(\alpha_3 - \alpha_1) & \cdots & \alpha_3^{n-3}(\alpha_3 - \alpha_2)(\alpha_3 - \alpha_1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n - \alpha_1 & (\alpha_n - \alpha_2)(\alpha_n - \alpha_1) & \cdots & \alpha_n^{n-3}(\alpha_n - \alpha_2)(\alpha_n - \alpha_1) \end{pmatrix}$$

Then we do something similar again: proceeding from right to left, subtract α_3 times the ith column from the $(i+1)$th column. But again, stop one column shorter still, leaving the first 3 columns untouched.

Exercise 14. *Convince yourself that this eventually produces a lower triangular matrix with diagonal entries*

$$\begin{aligned} &1 \\ &\alpha_2 - \alpha_1 \\ &(\alpha_3 - \alpha_2)(\alpha_3 - \alpha_1) \\ &(\alpha_4 - \alpha_3)(\alpha_4 - \alpha_2)(\alpha_4 - \alpha_1) \\ &\vdots \\ &(\alpha_n - \alpha_{n-1})(\alpha_n - \alpha_{n-2}) \cdots (\alpha_n - \alpha_1) \end{aligned}$$

Deduce, by Proposition 5, the formula for the Vandermonde determinant.

6.2 Proof 2: Nice Properties of Polynomials

The determinant may be a very ugly polynomial, but merely the fact that it is a polynomial can open up some approaches, as we see here.

In particular, the Vandermonde determinant itself is a polynomial in the quantities $\alpha_1, \ldots, \alpha_n$. Then by the alternating property of the determinant, for any pair of distinct indices i and j, if $\alpha_i = \alpha_j$, the determinant is 0.

In particular, if we view the determinant as a polynomial in the variable α_i alone, then it is divisible by $(\alpha_i - \alpha_j)$ for all $j \neq i$. These polynomials are all relatively prime for different values of i and j (up to exchanging i and j), so since the Vandermonde determinant is divisible by all of them, it must actually be divisible by their product,

$$\prod_{1 \leq i < j \leq n} (\alpha_j - \alpha_i).$$

On the other hand, let's consider the degree of the determinant as a polynomial in $\alpha_1, \ldots, \alpha_n$. Just using the formula for the determinant, it's a sum of terms of the form

$$\alpha_1^{\sigma(1)-1} \alpha_2^{\sigma(2)-1} \cdots \alpha_n^{\sigma(n)-1}$$

for some permutation σ. The numbers $\sigma(i) - 1$ run through $0, \ldots, n-1$ in some order, so the total degree of this term is necessarily $0 + 1 + \ldots + (n-1) = \frac{n(n-1)}{2}$.

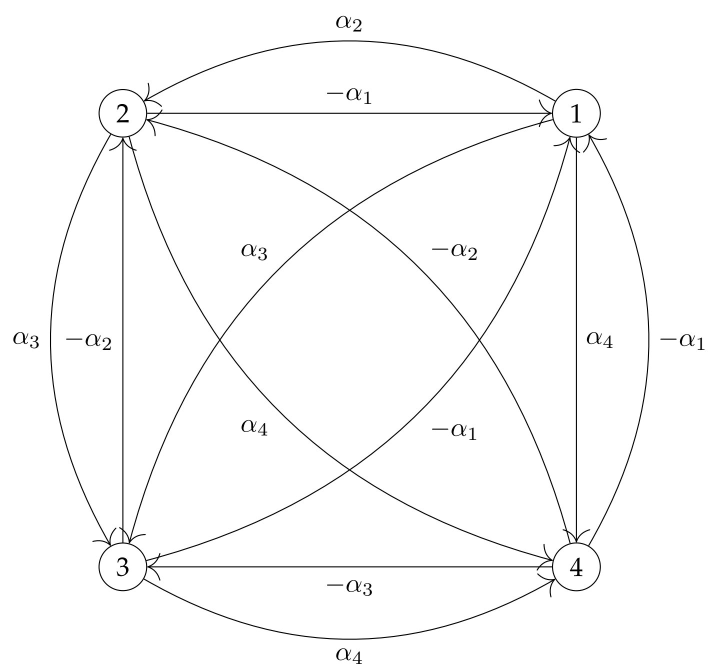
Back on the first hand, the degree of $\prod_{1 \leq i < j \leq n} (\alpha_j - \alpha_i)$ is also $\frac{n(n-1)}{2}$, since it's a product of that many linear terms. So the determinant we seek must actually be a scalar multiple of this product. It remains to verify that this scalar factor is 1, which we can do just be examining the coefficient $\alpha_2 \alpha_3^2 \cdots \alpha_n^{n-1}$-term on both sides. In the determinant, the coefficient is 1: this is the term in the formula for determinant associated to the identity permutation. On the other hand, it's straightforward to verify that the coefficient of this term in the product is also 1: this term comes from taking the first term in each $(\alpha_j - \alpha_i)$. This completes the proof.

6.3 Proof 3: Detournaments

While titling this section, I realized that, although "tournament" and the last three syllables of "determinant" contain the same three vowel sounds, the letters used to denote them are completely different. English spelling continues to be weird.

After the slickness of the previous two proofs, the final one we look at might seem a little overcomplicated. But it illustrates an important connection of determinants to combinatorics, by emphasizing the definition as a sum over permutations.

First, we define a directed graph on vertices labeled $1, \ldots, n$ as follows. Between each pair of vertices, add two edges, one in each direction. Then label each edge with a variable α_i corresponding to the vertex it points to. If the edge points from a higher number to a lower one, make the edge label negative. Otherwise, leave it postive. The diagram for $n = 4$ is shown below:



We can this of this diagram as representing a round-robin tournament (so every pair of players plays once) between n players. Suppose that, in any matchup, the player with the larger number is expected to win (but anything could happen).

Then picking an outcome for the tournament means, for every pair of vertices, choosing one of the directed edges between them, pointing to the winner of their match. Our labeling above means that each edge is labeled with a variable associated to the winner of the match, and the label is negative if the match was an upset (the favored player, with the larger number, didn't win). We will refer to such a choice of directed edges as just a **tournament**, and we will say that the **weight** of a tournament is the product of the edge labels.

Lemma 2. *The sum of all tournaments' weights is*

$$\prod_{1 \leq i < j \leq n} (\alpha_j - \alpha_i)$$

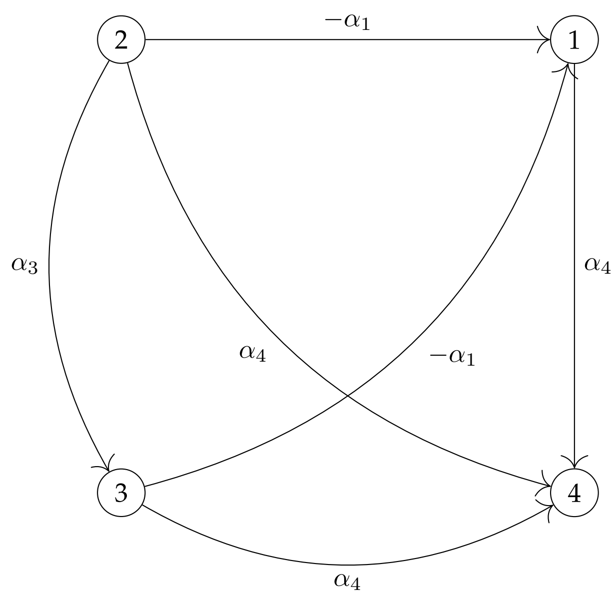
Proof. We can associate each of the $(\alpha_j - \alpha_i)$ terms with the match between i and j in the tournament. When we expand out this product, each term in the sum is given by choosing one of the two terms from each $(\alpha_j - \alpha_i)$ and multiplying them together. But each of these choices corresponds to choosing a winner of the match between i and j: if i wins, the match has weight $-\alpha_i$ (since j was favored) and if j wins, the match has weight α_j. So each term in the expansion corresponds to the weight of a particular tournament. □

Now we make a distinction between two types of tournaments.

Definition. *A tournament is **transitive** if there is no trio of players i, j, k such that i beats j, j beats k, and k beats i. Equivalently, the directed graph corresponding to the tournament contains no directed 3-cycles.*

Alternatively, we can phrase this as "if i beats j and j beats k, then i beats k", which is the standard idea of what it means to be transitive. In particular, a transitive tournament defines a total ordering of all the participants: the "i loses to j" relation is antisymmetric and transitive, and any two players are comparable in this way. Conversely, if we have a total ordering of the participants and declare that each player beats everyone below them in the order, we get a tournament.

Here is an example of such a tournament:



What total ordering does this tournament induce? Well, our unfortunate player 2 loses to everyone, so he comes first in the order. 3 doesn't do well against 1 and 4, but manages to beat 2, so she comes second. 1 beats 2 and 3, but still loses to 4, so she is the third player in order; meanwhile, 4 beats everyone and is the highest in the order. The result:

$$2 < 3 < 1 < 4$$

What is the weight associated to this transitive tournament? It's $\alpha_1^2\alpha_3\alpha_4^3$. Note that the exponent on each α_i is the number of matches player i won, while the sign is given by the number of upsets: 1 beating 2 and beating 3 were both upsets, but the other outcomes were expected, so we flip sign twice and end up with a positive sign.

This talk of ordering the players, along with the presence of a sign related to that ordering, and also the fact that this class is about determinants, should be leading you to think about determinants.

Lemma 3. *The Vandermonde determinant is the sum of the weights of all transitive tournaments.*

Proof. Writing out the Vandermonde determinant using the definition gives

$$\sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} \alpha_i^{\sigma(i)-1}.$$

We associate the term of a given permutation σ to the transitive tournament which induces the total ordering

$$\sigma^{-1}(1) < \sigma^{-1}(2) < \cdots < \sigma^{-1}(n)$$

Then in this tournament, 0 matches are won by player $\sigma^{-1}(1)$, 1 match is won by player $\sigma^{-1}(2)$,..., and $n-1$ matches are won by player $\sigma^{-1}(n)$. Thus the tournament has weight

$$\pm\alpha_{\sigma^{-1}(2)}\alpha^2_{\sigma^{-1}(3)} \cdots \alpha^{n-1}_{\sigma^{-1}(n)} = \pm\alpha_1^{\sigma(1)-1}\alpha_2^{\sigma(2)-1} \cdots \alpha_n^{\sigma(n)-1}$$

The sign of the tournament's weight is $+1$ if the number of upsets (i.e., pairs that are out of order in the permuted ordering) is even, and -1 if it is odd. This should look suspiciously similar to the definition of the sign of a permutation, and in fact it is the same (but this is left as an exercise):

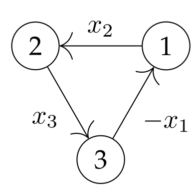
Exercise 15. *The sign of a permutation is $+1$ if there are an even number of pairs of elements out of order after the permutation is applied, and -1 if there are an odd number.*

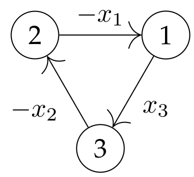
Once we know this, the result follows. □

The Vandermonde determinant formula has come down to showing that the sum over the weights of transitive tournaments (the determinant of the Vandermonde matrix) is equal to the sum over the weights of *all* tournaments (the expansion of $\prod_{1\le i<j\le n}(\alpha_j - \alpha_i)$). So it remains to prove the following:

Proposition 6. *The sum of the weights of all non-transitive tournaments is 0.*

Proof. By definition, a non-transitive tournament will contain a 3-cycle. Then we can transform it into a different non-transitive tournament by reversing the direction of the three edges constituting the cycle and leaving the others untouched. When we do this, the only effect it will have on the tournament's weight will be changing the sign, as illustrated by the example of a 3-vertex tournament below:





Our strategy, then, is to pair up the nontransitive tournaments in such a way that each pair of tournaments are related by reversing a 3-cycle. This is not trivial: if we want to come up with an algorithm that flips a 3-cycle in any tournament T to produce a tournament T' paired with it, we need to make sure our algorithm also sends T' back to T.

However, it's not too hard to prove that such a pairing exists. For this, we use the following lemma:

Lemma 4. *Suppose that T is a tournament, and that T' is obtained from T by reversing a directed 3-cycle. Then T and T' have the same number of directed 3-cycles.*

Proof. Suppose $i \to j \to k \to i$ is a directed 3-cycle in T, and let ℓ be any other vertex. Then we will show that the number of directed 3-cycles involving i, j, k, ℓ is the same in both T and T':

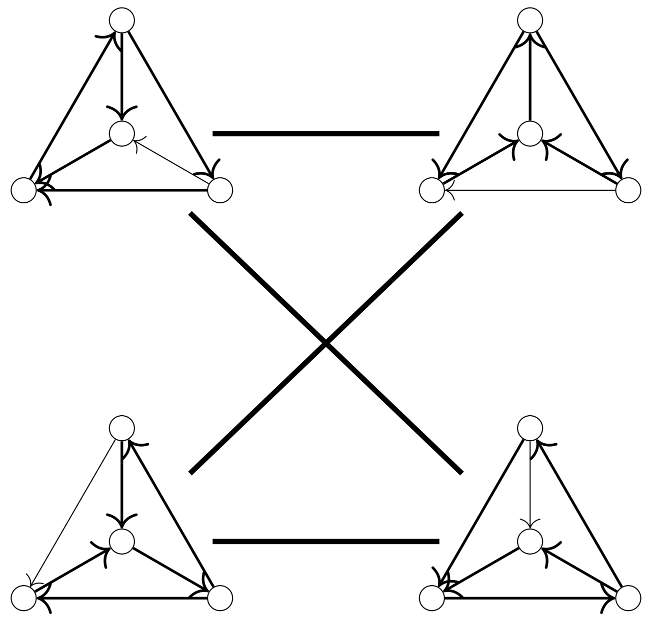
- If all of the edges between ℓ and i, j, k point away from or to ℓ, then there is only one directed 3-cycle in either case: the original $i \to j \to k \to i$ (and its reversed counterpart).

- If ℓ points toward one vertex (WLOG i), there are two directed 3-cycles: $i \to j \to k \to i$ and $i \to j \to \ell \to i$. Then in T', there are $i \to k \to j \to i$ and $i \to k \to \ell \to i$.

- If ℓ points toward two vertices, we can reduce to the previous case by reversing all of the arrows (which will not change the number of directed 3-cycles in any case).

□

Now for a nontransitive tournament T, with weight $w(T)$, we define an undirected graph G_T as follows: it has a vertex for every tournament obtained from T by reversing directed 3-cycles, and an edge between two tournaments when one is obtained from another by reversing a single directed 3-cycle. An example of this graph is shown below, where the directed 3-cycles are highlighted in bold.



Then by the above results, we know:

- We can divide the tournaments into two families, the tournaments with weight $w(T)$ and the tournaments with weight $-w(T)$.
- There is a number d such that each tournament is adjacent to exactly d tournaments in the other group. This is the number of directed 3-cycles in each tournament, which we know remains the same. Additionally, there are no edges within a family, because reversing a 3-cycle must flip the sign. If you're familiar with the language of graph theory, this is the same as saying our graph is d-regular and bipartite.

But from this second statement, we know that each family has the same number of members—otherwise, we'd have different numbers of outgoing edges from each side and they would not match up. Thus the sum over all of their weights cancels out. We repeat this process for all of the non-transitive tournaments, and deduce that all their weights sum to 0. □