A USER-FRIENDLY INTRODUCTION TO THE THEORY OF DETERMINANTS

NAT KUHN

> *"Can we all agree that the determinant is one of the most difficult things we try to teach to innocent unsuspecting students?"*
>
> Deane Yang, Courant Institute

1. Introduction

Mathematics is the realm of the necessary. Every mathematical truth, founded as it is on the laws of logic, *has* to be true. But is doesn't always feel that way. When you see how the mathematical pieces fit together—especially when you have more than one way of approaching a subject or a proof—it does generally feel like "that's just how it has to be." Occasionally, though, a mathematical result shows up unexpectedly out of left field, seeming either miraculous, or annoying, or both. For me, and I think for many others, the determinant can fall into the "left field" group.

Students who continue with mathematics beyond calculus typically take a course in linear algebra next. As it is taught today, linear algebra is generally developed very smoothly, starting with vector spaces, bases, linear transformations, and on to matrices, elementary row operations, and Gaussian elimination. It all makes sense. It all feels necessary. And then: determinants. It may start with a thicket of notation involving indices on your indices, such as $a_{1i_1}a_{2i_2}\cdots a_{ni_n}$, which I suspect typically stuns students into mute submission. Or it may start with a gentler discussion, producing a claimed invariant without a lot of understanding of why it is in fact an invariant.

The existence of the determinant is a deep mathematical fact that most working mathematicians take for granted, in part because it is introduced so early in the curriculum. By the same token, many mathematicians wind up having to teach it to undergraduates, which can be a stumbling block. When a group of mathematicians starts talking about how to teach determinants, the discussion gets surprisingly lively.

This article grew out of a couple of such discussions. In one of them someone pointed out that the determinant is characterized by the fact that it is the unique alternating multilinear form in the rows of an $n \times n$ matrix whose value on the identity matrix is equal to 1; that was the starting point of this paper. Section 2 starts with some geometric motivation for why those particular characteristics might be

Date: January 2018.

Many thanks to Matt Grayson for making the figures, and for feedback; and thanks to Ben Kuhn for feedback.

worth studying. In Section 3, we formally introduce these features of the determinant, which are referred to throughout the article as Properties 1 (multilinear), 2 (alternating), and 3 (normalized).

One of the best known mathematical strategies enables us to prove that an object with certain properties doesn't exist: assume that it does, and then deduce a contradiction. There is a less celebrated but still important strategy for proving that an object with certain properties *does* exist: assume that it does, deduce exactly what it has to be, and then show that your answer has the requisite properties. That is what we will do here.

Starting with the assumption that a determinant function satisfying Properties 1-3 exists, some basic linear algebra leads naturally to the main results in the field, including the formula for what the determinant *must be*. Once we have the formula, we can show that it does in fact give a determinant satisfying our conditions (Sections 4, 5, and 7). There is still a whiff of the contingent, though: the existence of the determinant depends on the sign of a permutation being well-defined (Section 6).

The determinant is often taught in a way that is heavy on formulas, notation, and algebraic manipulation. A certain amount of that is inescapable, at least if you don't approach the subject from the standpoint of exterior algebra. The approach taken here does not eliminate these difficulties, but it does lighten them somewhat, because approaching the determinant as a multilinear form allows you to treat them as a *part* of linear algebra, rather than an equation-bomb that is dropped *on top of* linear algebra.

We build up to multilinearity somewhat gradually. Before Section 5 takes us into the realm of multilinearity, we see in Section 4 that you can get surprisingly far just applying basic knowledge of linearity to a single row of your matrix. An introductory linear algebra course does not typically include the word "multilinear," but it does typically spend some time on bilinear forms, and Section 5 works to leverage that into the general case. For that reason, it would be better to attempt this approach to determinants after introducing bilinear forms, which is typically not its usual place in the syllabus.

The connection between determinants and geometry (in particular, volume) is often an afterthought, mentioned only in passing. Here, we take it up as motivation right at the start, in Section 2. We return to it for a more detailed, rigorous treatment at the end (Sections 9 and 10). It is not necessary to cover these last two sections, but people who really believe in geometry will take some comfort in knowing that they are there. In terms of prerequisites, Section 9 does make some use of orthogonal projection and will make more sense to students who have been through the Gram-Schmidt process elsewhere.

With or without the last two sections, the student who has followed this approach to determinants will be much better prepared to make the transition to differential forms, a subject which for the last 50 years or so has been poised to take its rightful place earlier in the advanced calculus curriculum at any moment. To help bridge the chasm which still exists, I have inserted a few remarks connecting the material in this article with exterior algebra.

Many of the details of what follows are left as exercises. Presenting the material in this way enhances learning by calling on the reader to be an active participant; it also helps to clarify which parts of the exposition reflect essential ideas versus relatively routine mathematical reasoning. Exercises that are part of the essential understanding of the subject are marked with an asterisk. One result of this approach is that the exercises are of varying levels of difficulty. With perhaps one exception, they should all be within the range of capabilities of a bright, motivated math major. Most students of linear algebra do not fit that full description—for example, many will be engineers who are very bright but motivated by things other than proving theorems. As a result, a teacher incorporating this material in lecture course will undoubtedly have to supply additional details.

2. Invitation

In plane geometry, the area of a rectangle is defined to be its base times its height. One of the most basic theorems of plane geometry (Euclid, Book I, Proposition 35, more or less) is that this same formula applies to any parallelogram. [The most obvious demonstration of this fact involves a cut-and-paste argument, as depicted in Figure 1: lop a triangle off of one end and bring it around to the other to produce a rectangle on the same base, with the same height. Euclid is too canny to fall for this, because the argument fails if no part of the top lies directly above the base.]

In a more modern context, any two vectors in the plane span a parallelogram. Can we look at area from the standpoint of linear algebra, in a way that might generalize to volume in three and even more dimensions? Let's start with a function $A(\mathbf{u}, \mathbf{v})$ which gives us the area of a parallelogram bounded by the vectors $\mathbf{u}$ and $\mathbf{v}$. We'll make $\mathbf{u}$ the base of our parallelogram, and for convenience we'll draw it horizontally pointing to the right (facing eastward). The fact that the area depends only on the height means that $A(\mathbf{u}, \mathbf{v}) = A(\mathbf{u}, \mathbf{v} + t\mathbf{u})$ for any $t \in \mathbb{R}$, as shown in Figure 2.

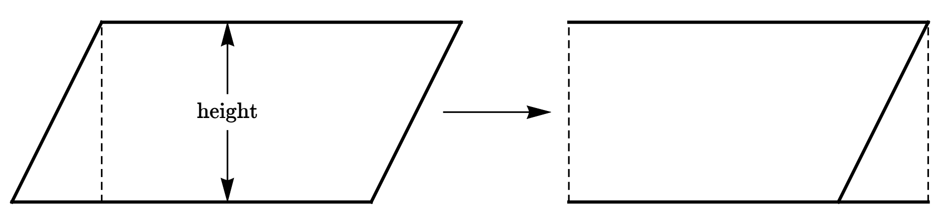

Figure 1. area = base × height: a tempting argument

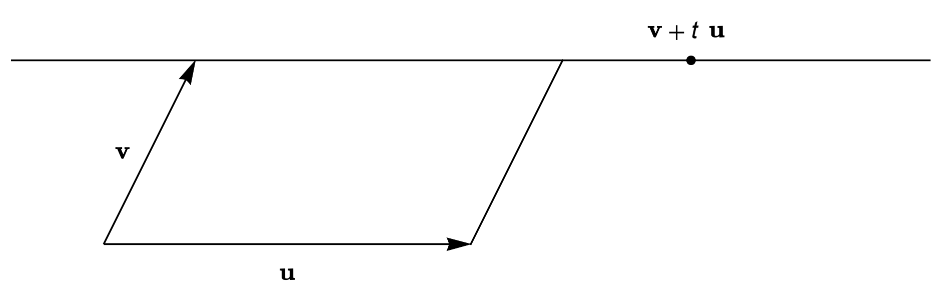

Figure 2. The vector $\mathbf{v} + t\mathbf{u}$ for $t \in \mathbb{R}$ spans (along with $\mathbf{u}$) a parallelogram of the same height, and hence the same area

There is more that we can say from the standpoint of linear algebra. If the base $\mathbf{u}$ is held constant, the height of a vector $\mathbf{v}$ above the base is a linear functional of $\mathbf{v}$, we could call it $h(\mathbf{v})$. As seen in Figure 3 the heights add; multiplying by the base shows that

$$A(\mathbf{u}, \mathbf{v} + \mathbf{w}) = A(\mathbf{u}, \mathbf{v}) + A(\mathbf{u}, \mathbf{w}) \tag{1}$$

In fact, this formula works for all $\mathbf{v}$ and $\mathbf{w}$, as long as we have some mathematical courage: when $\mathbf{v}$ is below the “horizontal,” our height functional h has a negative value, so multiplying by the base (the length of $\mathbf{u}$), we get a negative area. Figure 4 depicts this situation. The absolute value of this negative area is the length of the base time the absolute value of the height functional, which is the actual geometric height, so even when A is negative its absolute value gives us the true geometric area.

Equation 1 tells us that A is additive in its second variable when its first variable is held constant. It also respects scalar multiplication in the second variable, since h respects scalar multiplication. That is,

$$A(\mathbf{u}, c\mathbf{v}) = cA(\mathbf{u}, \mathbf{v}) \tag{2}$$

Geometrically, stretching one side of a parallelogram by a factor c multiplies the area by c. Putting this together, A is linear in the second variable when the first variable is held constant.

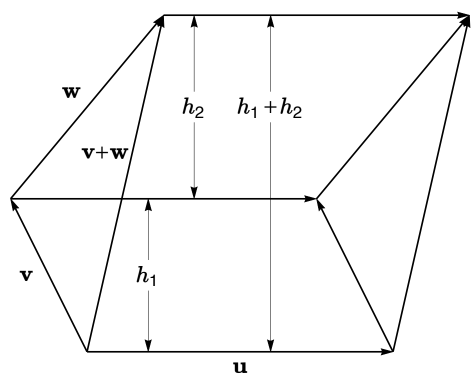

FIGURE 3. Heights add

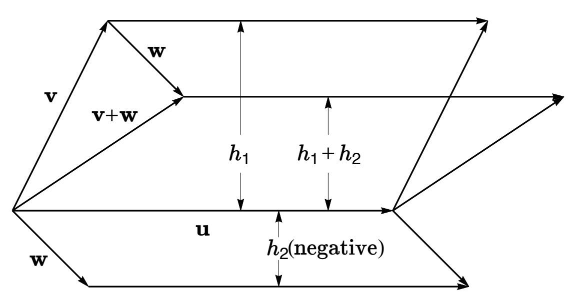

FIGURE 4. Parallelograms below the horizontal have negative values for height and area

The fact that some areas can be negative can take a little getting used to. We say that A gives us the **signed area** of a parallelogram But—like negative numbers, fractions, and complex numbers—the notion of negative area extends our scope in a consistent and powerful way. Furthermore, "negative area" is actually quite a familiar concept from first-year calculus: the area "under" a negative-valued function is negative. In that situation, the line between positive and negative area is the x axis, which is fixed. In our case, the criterion is mobile: consider an observer standing on the vector $\mathbf{u}$ with her toes aligned with $\mathbf{u}$'s arrow. If $\mathbf{v}$ points leftward, then $A(\mathbf{u}, \mathbf{v})$ is positive; if rightward, the area is negative. Obviously, there is a quite a bit of hand-waving here, potentially of a more literal kind than usual. But, patience! We will make all of this rigorous.

What other properties does A have? Well, obviously the area of the parallelogram must not change if you interchange the order of the vectors. Except, oops... The actual area—which is the absolute value of A—doesn't change, but its sign can, and in fact it does. If $\mathbf{v}$ points to your left when your toes point along $\mathbf{u}$, then $\mathbf{u}$ will point to your right when your toes are aligned along $\mathbf{v}$ (Figure 5), so:

$$A(\mathbf{u}, \mathbf{v}) = -A(\mathbf{v}, \mathbf{u}) \tag{3}$$

In other words, A is *antisymmetric*. It is an easy consequence of antisymmetry that A is also linear in its *first* variable when the *second* variable is held constant. That is, A is a bilinear form. [For readers not familiar with bilinear forms, there is more information in the next section.]

Any scalar multiple of the area satisfies equations 1–3. To uniquely determine the area, we need to *normalize* it so that the area of a unit square is equal to 1. That is, we need

$$A(\mathbf{e}_1, \mathbf{e}_2) = 1 \tag{4}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are unit vectors (horizontal and vertical, respectively).

Well, that looks reasonably promising from a linear algebra perspective. What about the hope of generalizing to 3-dimensional volume and higher dimensions? The key fact that we have used is that the area of a parallelogram is the length of its base times its height. What about the three-dimensional analogue, a parallelepiped? It is in fact true that its volume is equal to the area of its base times its height. [One way to see this is to recall from first-year calculus that volume depends only on the

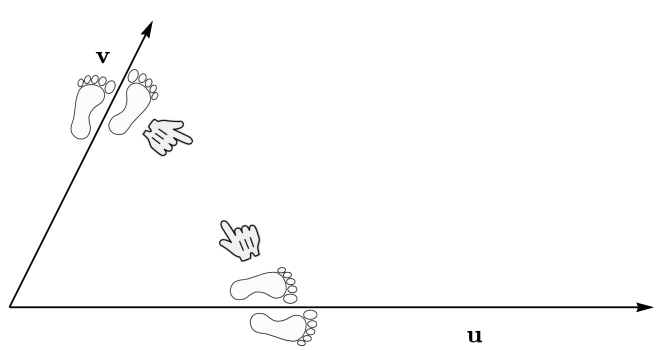

FIGURE 5. Reversing the arguments changes the sign of A

area of horizontal slices, not on their relative arrangement.] So there is a reasonable hope that that part of our argument could be extended to higher dimensions. The more difficult is whether it is possible to come up with some consistent analogue of "above the horizontal" and below it. It can be done, but it is non-trivial, and that will be a significant part of our story.

We will work primarily in the n-dimensional vector space E_n, thought of as "row vectors" of length n:

$$[\ a_1\ a_2 \cdots a_n\ ]$$

where the a_i are scalars; when we are thinking about geometric volume, our scalars will be real numbers. In fancier language, our field of scalars will be $\mathbb{R}$. However, everything we do from the perspective of linear algebra will work over any field as long as the field does not have characteristic 2. If you are not familiar with the term "field," just think "$\mathbb{R}$" (the real numbers) or "$\mathbb{C}$" (the complex numbers). Saying that a field has characteristic 2 essentially says that in that field $2 = 0$, so we know that $\mathbb{R}$ and $\mathbb{C}$ do not have characteristic 2! [In fact, except where noted otherwise, our arguments will apply not just to a field but to any commutative ring with unit in which 2 is not a divisor of zero.]

We will use $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ to denote the standard basis of E_n, in which $\mathbf{e}_i$ is the row vector of length n with a 1 in the i-th column and zeros elsewhere.

One focus of our work will be to producing a generalization of our function A to n dimensions, where it is called the determinant (det). It is a function of n vector variables $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in E_n$ so that $\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is the signed n-dimensional volume of an n-dimensional parallelepiped spanned by those vectors. This is not the standard point of view; the determinant is generally thought of a function of a single $n \times n$ matrix A. The two formulations are entirely equivalent: starting with vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, we get an $n \times n$ matrix by stacking the $\mathbf{v}_i$ on top of one another, in order. Starting from a matrix $[a_{ij}]$, we "unstack" the rows so that $\mathbf{v}_i$ is the i-th row of A. In other words:

$$\mathbf{v}_i = [\ a_{i1}\ a_{i2} \cdots a_{in}\ ]$$

We will treat the determinant as a function of a single matrix or n vectors quite interchangeably, trusting the reader to sort out the context.

Exercise 1. *Develop a formula for*

$$A([\ a\ b\ ], [\ c\ d\ ]) = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)$$

when $a, b, c, d > 0$. Hint: Draw the parallelogram spanned by the two vectors. It sits inside a large rectangle whose area is $(a + c)(b + d)$. What do you need to subtract to get the area of the parallelogram? Two small rectangles, and two pairs of triangles. Considering each pair of triangles together converts their area into another rectangle. Note that for the signs to be correct, the vector $[\ c\ d\ ]$ must be "to the left" of $[\ a\ b\ ]$ in your picture, i.e. $[\ a\ b\ ]$ must lie closer to the horizontal $\mathbf{e}_1$ axis.

3. Statement of Our Main Results

It turns out that a small number of properties adopted from our area function are enough to uniquely characterize the determinant. In this section we will give a rigorous statement of the situation. In subsequent sections we will use these properties to deduce all of the other usual properties of the determinant, including the usual formula for it. Once we know the formula, we will use it to *define* the determinant, and then prove that the determinant defined in this way does in fact have the requisite properties.

Our main result is this:

Theorem 1. *There is a unique scalar-valued function* det *on $n \times n$ matrices over a field F (with characteristic not equal to 2) having the following properties:*

(1) det *is linear in each row when all the other rows are held constant.*
(2) Any of the following equivalent conditions hold:
   (a) If two rows of a matrix are equal, its determinant vanishes.
   (b) Adding an arbitrary multiple of one row to another does not change the determinant of a matrix.
   (c) Interchanging any two rows of a matrix changes the sign of its determinant.
(3) The determinant of the $n \times n$ identity matrix, $\det(I_n)$*, is equal to 1.*

We have already established these properties for our area function A from Section 2. Let's look at them in more depth.

Viewed as a function of the n rows, Property 1 says that det is an example of a multilinear form—specifically an n-linear form.

Definition. *A k-**linear form** on a vector space V is a scalar-valued function of k arguments, each of which lies in V, which is linear in each argument when the other arguments are held constant.*

In the language of physics, a k-linear form is a covariant tensor of rank k.

The simplest example of a multilinear form is a bilinear form... Oh, wait, the simplest example is a 1-linear form, which goes by the name "linear functional." The simplest *non-trivial* example is a bilinear form (which could also be called a "2-linear form"), such as an inner product, or our A function from Section 2. Saying that B is a bilinear form means that it is linear in the first variable, holding the second variable constant, i.e., $B(\mathbf{u}+\mathbf{v},\mathbf{w}) = B(\mathbf{u},\mathbf{w}) + B(\mathbf{v},\mathbf{w})$ and $B(c\mathbf{u},\mathbf{w}) = cB(\mathbf{u},\mathbf{w})$; and linear in the second variable, holding the first variable constant, i.e., $B(\mathbf{u},\mathbf{v}+\mathbf{w}) = B(\mathbf{u},\mathbf{v}) + B(\mathbf{u},\mathbf{w})$ and $B(\mathbf{u},c\mathbf{w}) = cB(\mathbf{u},\mathbf{w})$.

A multilinear form is said to be **alternating** if interchanging any two of its arguments changes its sign. Thus in the presence of property 1, Property 2c is equivalent to saying that det is an alternating multilinear form in the rows of our matrix. An alternating bilinear form is uh, alternatively known as antisymmetric; we have already argued is holds for A. Even before that, we suggested (Figure 2) that Property 2b applies to A.

Then Property 3 simply says that

$$\det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = 1$$

which was also a property of A (equation 4).

So if Theorem 1 is correct, and if you believe our assertions about A, then A must be equal to the 2×2 determinant.

What if we drop Property 3, which normalizes the determinant? This result will be useful:

Theorem 2. *If D is a function mapping $n \times n$ matrices over a field F to F itself and satisfying Properties 1 and 2 of Theorem 1, then D is* $\det$ *multiplied by the scalar $D(I_n)$.*

In other words, an alternating multilinear form in the rows of a matrix is a multiple of the determinant.

Exercise 2. *Show that Theorem 2 is equivalent to saying that for an alternating n-linear form D, if $D(I_n) = 0$ then D vanishes on all matrices. Hint: If $D(I_n) \neq 0$, show that Theorem 1 applies to*

$$\frac{1}{D(I_n)} D$$

[Note that this is one case in which we are using the properties of a field; later we will give another proof that works in a commutative ring.]

***Exercise 3.** *Develop formula for the determinant of a 1×1 matrix, and prove Theorems 1 and 2 in the case $n = 1$, where property 2 is vacuous.*

Finally, we close with this:

***Exercise 4.** *Expand the following sketch of a proof of the equivalence of properties 2a, 2b, and 2c into something you understand and believe.*

(2a)⇒(2b):

$$\begin{aligned} D(\mathbf{v}, \mathbf{w} + c\mathbf{v}, \dots) &= D(\mathbf{v}, \mathbf{w}, \dots) + D(\mathbf{v}, c\mathbf{v}, \dots) \\ &= D(\mathbf{v}, \mathbf{w}, \dots) + cD(\mathbf{v}, \mathbf{v}, \dots) \\ &= D(\mathbf{v}, \mathbf{w}, \dots) \end{aligned}$$

(2b)⇒(2a): $D(\mathbf{v}, \mathbf{w} + \mathbf{v}, \dots) = D(\mathbf{v}, \mathbf{w}, \dots) + D(\mathbf{v}, \mathbf{v}, \dots)$

(2a)⇒(2c):

$$\begin{aligned} 0 &= D(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}, \dots) \\ &= D(\mathbf{v}, \mathbf{v}, \dots) + D(\mathbf{v}, \mathbf{w}, \dots) + D(\mathbf{w}, \mathbf{v}, \dots) + D(\mathbf{w}, \mathbf{w}, \dots) \\ &= D(\mathbf{v}, \mathbf{w}, \dots) + D(\mathbf{w}, \mathbf{v}, \dots) \end{aligned}$$

(2c)⇒(2a) $D(\mathbf{v}, \mathbf{v}, \dots) = -D(\mathbf{v}, \mathbf{v}, \dots)$*. For those with some familiarity with abstract algebra, this is where we use the assumption that the characteristic of F is not 2.*

***Exercise 5.** *Assuming Theorem 1, show that if A is an $n \times n$ matrix and c is a scalar constant, then* $\det cA = c^n \det A$.

4. Warm-Up: Use Linearity

In this section, we will assume Theorems 1 and 2; we will see that a number of powerful results in the theory of determinants follow from our three simple properties.

To avoid getting bogged down in notation, let's start with a 4×4 matrix

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix}$$

By Property 1, det is a linear functional of the the first row, when all other columns are held constant. In other words,

$$\det A = A_1 a_1 + A_2 a_2 + A_3 a_3 + A_4 a_4 \tag{5}$$

where the "constants" A_i depend on values of the bs, cs, and ds.

What is the value of A_1? We can determine it by setting $a_1 = 1$ and $a_2 = a_3 = a_4 = 0$. In other words,

$$A_1 = \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix}$$

Now we can apply property 2b to subtract multiples of row 1 from the other rows to cancel out the rest of the first column, without changing the value of the determinant, or the other columns. As a result,

$$A_1 = \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & b_2 & b_3 & b_4 \\ 0 & c_2 & c_3 & c_4 \\ 0 & d_2 & d_3 & d_4 \end{bmatrix} \tag{6}$$

So in fact A_1 does not depend on b_1, c_1, and d_1, just on the other bs, cs, and ds, that is, the entries of the matrix

$$B = \begin{bmatrix} b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \\ b_4 & c_4 & d_4 \end{bmatrix}$$

How does the value of A_1 given by equation 6 depend on the entries of the matrix B? Well, based on Property 1 of 4×4 determinants, it is multilinear in rows of B. Similarly, based on Property 2 of 4×4 determinants, it is alternating. Finally, what is its value if B is the identity matrix I_3? It is

$$\det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \det I_4 = 1$$

It follows from Theorem 2 (or the uniqueness part of Theorem 1) applied to 3×3 matrices that $A_1 = \det B$.

How about the other A_i? Or what if we want to do the same computation based on a row other than the first one? It is helpful to introduce the following notation: let the matrix $A_{(ij)}$ denote the $n-1 \times n-1$ matrix obtained by omitting the i-th row and j-th column of the matrix A. In the argument above, $B = A_{(11)}$.

For example, what is the value of A_4 in our set-up above? The same argument shows that A_4 is a some multilinear form D in the entries of the 3×3 matrix $A_{(14)}$, and that D is alternating. As a result, Theorem 2 tells us that D is equal to $\det A_{(14)}$ multiplied by $D(I_3)$. What is $D(I_3)$? It is

$$\det \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{7}$$

We can compute this by successively interchanging rows of our matrix, changing the sign of the determinant each time:

$$\det \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = -\det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = -\det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

so the value of this determinant is -1. Thus Theorem 2 tells us that $A_4 = -\det A_{(14)}$. A similar argument shows that $A_2 = -\det A_{(12)}$ and $A_3 = \det A_{(13)}$. Then by Equation 5

$$\det A = a_1 \det A_{(11)} - a_2 \det A_{(12)} + a_3 \det A_{(13)} - a_4 \det A_{(14)}$$

This gives a method of computing determinants, known as the *Laplace expansion*, *expansion by minors*, or *expansion by cofactors*. It works, but it is so computationally inefficient for either people or computers that it is of primarily theoretical importance. For example, if you use it to compute the determinant of a 4×4 matrix on an exam, and your teacher is conscientious enough to check your work, they will be mad at you, trust me. In general, it is much more efficient to compute the determinant by row reduction, using Properties 1 and 2 to reduce the computation to an upper-triangular matrix and then apply Exercise 8.

Let's tackle the general case of the Laplace expansion. Suppose we want to compute our determinant by focusing on the i-th row, rather than the first one. As above, the determinant is a linear function of the i-th row, and we need to compute the coefficient of a_{ij} for each j. The same argument as above shows that the coefficient of a_{ij} is an alternating multilinear form in the columns of $A_{(ij)}$, and so by Theorem 2 it is a multiple of $\det A_{(ij)}$.

What multiple is it? It is the determinant of the matrix $n \times n$ matrix B whose (i, j) entry is 1, which has zeros in the rest of the i-th row and j-th column, and whose remaining entries give the identity matrix when the i-th row and j-th column are omitted, that is, $B_{(ij)} = I_{n-1}$.

***Exercise 6.** *With B as above, prove that* $\det B = (-1)^{i+j}$. *Hint: Suppose $i < j$. Show that, with a proper understanding of the "corner cases,"*

$$\det B = D(\mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \mathbf{e}_j, \mathbf{e}_{i+1}, \ldots, \mathbf{e}_{j-1}, \mathbf{e}_{j+1}, \ldots, \mathbf{e}_n)$$

Show that this expression can be reduced to $\det I_n$ *with $j - i$ row switches. Give a similar proof for $i > j$, and the easy case of $i = j$.*

Note: if you came up with a proof that seems much easier to you than that, make sure that you didn't use the fact that interchanging *columns* of a determinant also changes the sign. It's true, and we'll prove it later, but it's not one of our defining properties.

The consequence is that the coefficient of a_{ij} in the formula for the determinant of a is $(-1)^{i+j} \det A_{ij}$. As a result, we have:

Theorem 3 (Laplace expansion)**.** *With the notation as above, for any i with $1 \leq i \leq n$,*

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

The $\det A_{(ij)}$ are known as *minors*, while the $(-1)^{i+j}\det A_{(ij)}$ are known as *cofactors*.

***Exercise 7.** *Use Theorem 3 and Exercise 3 to prove a formula for the determinant of a 2×2 matrix. There are not very many formulas in mathematics that you should really memorize, but this is one of them. Do so, after checking that your answer agrees with the one from Exercise 1.*

***Exercise 8.** *Using Theorem 3, prove that the determinant of an upper-triangular matrix is the product of its diagonal entries. (An upper-triangular matrix is one whose entries below the diagonal are zero; equivalently it is a matrix $[a_{ij}]$ where $a_{ij} = 0$ if $i < j$.)*

Theorem 3 is pretty cool, but now an even cooler thing happens. Suppose we form a matrix $A' = [a'_{ij}]$ which is equal to A except that its i-th row is a duplicate of the k-th row of A where $k \neq i$. We know by Property 2a that $\det A' = 0$. Let's apply Theorem 3, expanding $\det A'$ on the i-th row. we have

$$0 = \det A' = \sum_{j=1}^{n}(-1)^{k+j}a'_{ij}\det A'_{(ij)} \quad k \neq j$$

Now $a'_{ij} = a_{kj}$ by construction, and $A'_{(ij)} = A_{(ij)}$ because once you omit the i-th row, A and A' are identical. So we have:

$$0 = \sum_{j=1}^{n}(-1)^{i+j}a_{kj}\det A_{(ij)} \quad k \neq i \tag{8}$$

Combining Theorem 3 and equation 8 gives us

$$\sum_{j=1}^{n}(-1)^{i+j}a_{kj}\det A_{(ij)} = \delta_{ki}\det A \tag{9}$$

where δ_{ki} denotes the Kronecker delta (1 if $k = i$, 0 if $k \neq i$).

We can form the *adjunct matrix* (or *adjugate matrix*, or *classical adjoint*) $\operatorname{adj} A$ whose (i, j) entry is the (j, i) cofactor $(-1)^{i+j}\det A_{(ji)}$ (note the reversal of indices i and j). If we use b_{ij} to denote this (i, j) entry of the adjunct matrix, equation 9 reads:

$$\sum_{j=1}^{n} a_{kj}b_{ji} = \delta_{ki}\det A$$

Which yields the following result:

Theorem 4.

$$A\operatorname{adj} A = (\det A)I_n$$

Corollary. *A matrix A is invertible if $\det A \neq 0$. (The converse of this fact is Exercise 30.)*

In fact, we now have a formula for the inverse of the matrix A, albeit a computationally inefficient one:

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A$$

[For readers with some knowledge of abstract algebra, it is worth noting that until we took the reciprocal of the determinant, everything we have done so far could be done in a commutative ring R (with unit). So in fact Theorem 4 shows that if A is an $n \times n$ matrix with coefficients in R and $\det A$ is a unit in R, then A has an inverse with entries in R.]

***Exercise 9.** *Compute the adjunct matrix of a* 2×2 *matrix; use it to give a formula for the inverse of a* 2×2 *matrix. Multiply your result by the original matrix to check your result.*

***Exercise 10.** *Prove* Cramer's rule: *Suppose that* $\mathbf{x}A = \mathbf{b}$, *with* A *an* $n\times n$ *matrix,* $\mathbf{x}, \mathbf{b} \in E_n$ *and*

$$\mathbf{x} = [x_1 \dots x_n]$$

Show that

$$x_i = \frac{\det A_i}{\det A}$$

where A_i *is the matrix obtained by replacing the* i*-th row of* A *by* $\mathbf{b}$.

Exercise 11. *Get any textbook on linear algebra. Find worked examples of computing determinants, or problems with answers. Compute the determinants on your own using row reduction and Properties 1 and 2 to reduce the matrix to a upper-triangular form and then apply Exercise 8. Use the Laplace expansion to compute the determinant of one of the* 4×4 *matrices.*

5. Next Step: Use Multilinearity

Having seen some consequences of Theorems 1 and 2, we turn to proving them in this and the following two sections. To do so, we need to understand more about multilinearity.

In the previous section, we used the fact that a linear functional L is completely determined by its value on the n basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$; by applying L to the basis vectors we can "read off" the coefficients used in evaluating the functional. This is not a deep theorem. It is a simple consequence of linearity; all the "depth" is in the definition of linearity. Specifically, if L is a linear functional on V and $\mathbf{v} = \sum_{i=1}^n a_j \mathbf{e}_i$, then

$$L(\mathbf{v}) = \sum_{i=1}^n a_i L(\mathbf{e}_i)$$

Similarly, a bilinear form is completely determined by its values on the n^2 pairs of basis vectors $(\mathbf{e}_i, \mathbf{e}_j)$, and evaluating it involves a sum over n^2 corresponding terms. If in addition to $\mathbf{v}$ as described above, we take $\mathbf{w} = \sum_{j=1}^n b_j \mathbf{e}_j$, then we can use linearity in each "slot" of B to evaluate $B(\mathbf{v}, \mathbf{w})$. Expanding the first slot,

$$B(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^n a_i B(\mathbf{e}_i, \mathbf{w})$$

Expanding each of these n terms on the second slot,

$$= \sum_{i=1}^n \sum_{j=1}^n a_i b_j B(\mathbf{e}_i, \mathbf{e}_j)$$

As a result,

$$B(\mathbf{v}, \mathbf{w}) = \sum a_i b_j B(\mathbf{e}_i, \mathbf{e}_j) \tag{10}$$

where the sum in equation 10 is over the n^2 pairs (i, j) with $1 \le i, j \le n$.

The generalization of equation 10 is not conceptually difficult, just notationally complex. To make that transition, let's continue with our bilinear form, but instead of looking at $\mathbf{v}$ and $\mathbf{w}$ let's look at $\mathbf{v}_1$ and $\mathbf{v}_2$ with $\mathbf{v}_i = \sum_{j=1}^{n} a_{ij}\mathbf{e}_j$, in this case for $i \in \{1, 2\}$. Then equation 10 becomes

$$B(\mathbf{v}_1, \mathbf{v}_2) = \sum a_{1i}a_{2j}B(\mathbf{e}_i, \mathbf{e}_j) \tag{11}$$

In the 3-linear case, we the sum would be over triples (i, j, k); for a k-linear form the sum is over k-tuples. But in that situation, we can't just keep on "grabbing letters." The usual solution is to use "indexed indices" $i_1, i_2, ..., i_k$ for the elements of our k-tuple, resulting in subscripts with subscripts and quite a notational thicket. An alternate approach to k-tuples of integers from 1 to n is to use functions mapping $\{1, 2, \ldots, k\}$ to $\{1, 2, \ldots, n\}$, which are in one-to-one correspondence with the k-tuples. If we define $\overline{n} = \{1, 2, \ldots, n\}$, the set-theoretic notation for the set of such functions is simply $\overline{n}^{\overline{k}}$, which, not coincidentally, has n^k elements. In this set-up, equation 11 becomes

$$B(\mathbf{v}_1, \mathbf{v}_2) = \sum_{f \in \overline{n}^{\overline{2}}} a_{1f(1)}a_{2f(2)}B(\mathbf{e}_{f(1)}, \mathbf{e}_{f(2)}) \tag{12}$$

Although the general case may look intimidating, the reader should be able to see that it is a direct generalization of equation 12, which is itself just a restatement of equation 10:

Proposition 5. *Suppose that M is a k-linear form on E_n, and $\mathbf{v}_i = \sum_{j=1}^{n} a_{ij}\mathbf{e}_j$ for $1 \le i \le k$. Then*

$$M(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \sum_{f \in \overline{n}^{\overline{k}}} a_{1f(1)}a_{2f(2)} \cdots a_{kf(k)}M(\mathbf{e}_{f(1)}, \mathbf{e}_{f(2)}, \ldots, \mathbf{e}_{f(k)})$$

***Exercise 12** (optional)**.** *If you feel that you need a proof of Proposition 5, give one. Hint: Use induction on k. It will be easiest to expand M on the last slot first, and then use the induction hypothesis on each ($k-1$)-linear form $M(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{e}_i)$ where i runs from 1 to n.*

In the case of Theorems 1 and 2, we are specifically interested in the case where k and n are equal, and our form, let's call it D, is alternating. Then Proposition 5 gives us

$$D(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \sum_{f \in \overline{n}^{\overline{n}}} a_{1f(1)}a_{2f(2)} \cdots a_{nf(n)}D(\mathbf{e}_{f(1)}, \mathbf{e}_{f(2)}, \ldots, \mathbf{e}_{f(n)}) \tag{13}$$

Note that the n-tuple of basis vectors $(\mathbf{e}_{f(1)}, \mathbf{e}_{f(2)}, \cdots, \mathbf{e}_{f(n)})$ in equation 13 corresponds to an $n \times n$ matrix with a single 1 in each row (in the $f(i)$-th column of the i-th row), and zeros everywhere else.

Equation 13 is what we get by fully exploiting the fact that D is multilinear (Property 1). What do we get from the fact that D is alternating (Property 2)? If $f(i) = f(j)$ for any $i \ne j$ then the i-th and the j-th rows are equal, and Property 2b says that the corresponding term vanishes. This means we can ignore any f in the expansion which is not injective.

An injective map of $\overline{n} = \{1, 2, \ldots, n\}$ to itself is actually bijective, since the set $\overline{n}$ is finite. Such a map is called a **permutation**, and there are $n!$ of them. This means that we can reduce the n^n terms of Equation 13 to $n!$ terms, which is significant pruning. Plus, as soon as something is a permutation, we can stop calling it f, which seems boring, and instead call it π, which has a bit more pizazz.

Exercise 13. *Use Theorem 3 to give an alternate proof that the determinant of an $n \times n$ matrix can be written as a sum of $n!$ terms.*

Exercise 14.** *Show that the matrix corresponding to the n-tuple of basis vectors $(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)})$ has $\delta_{j\pi(i)}$ as its (i, j) entry. Such matrices are known as* ***permutation matrices*. Show that the permutation matrices are exactly those matrices with a single 1 in every row, a single 1 in every column, and zeros everywhere else.*

The set of all permutations of $\{1, 2, \ldots, n\}$ is denoted by S_n. Being functions, permutations have an important feature: you can *compose* two of of them. For $\pi, \pi' \in S_n$, their composition $\pi \circ \pi'$, defined by $(\pi \circ \pi')(k) = \pi(\pi'(k))$, is also a permutation. Furthermore, the identity function $1_{\overline{n}}$ is a permutation, and $\pi \circ 1_{\overline{n}} = 1_{\overline{n}} \circ \pi = \pi$. Since a permutation π is bijective, it has an inverse function π^{-1}, which is also a permutation and has the property that $\pi \circ \pi^{-1} = \pi^{-1} \circ \pi = 1_{\overline{n}}$. [Readers with some familiarity with abstract algebra will note that these, combined with the fact that the composition of functions is associative, mean that S_n forms a *group*. The group-ness of S_n will be lurking in the background, but we won't assume any previous knowledge of abstract algebra.]

If $f \in \overline{n}^{\overline{n}}$ but $f \notin S_n$, its corresponding term in equation 13 vanishes; as a result, we can rewrite equation 13 as

$$D(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \sum_{\pi \in S_n} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} D(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)}) \tag{14}$$

Note that $D(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)})$ is D applied to the permutation matrix corresponding to π.

We can squeeze more out of the fact that det is alternating; to do that we need some additional properties of permutations, which will be covered in the next section. However, an enterprising reader could attempt Theorem 2 and the uniqueness part of Theorem 1 based on little more than what we know already.

Exercise 15. *Prove that a k-linear form on an n-dimensional vector space V must vanish if $k > n$.*

Exercise 16 (optional)**.** *As we saw in equation 11, we can we expand a bilinear form B in terms of the $B(\mathbf{e}_i, \mathbf{e}_j)$. The converse is also true: given n^2 arbitrarily chosen scalars b_{ij} there is a unique bilinear form B such that $B(\mathbf{e}_i, \mathbf{e}_j) = b_{ij}$, and its formula is*

$$B(\mathbf{v}_1, \mathbf{v}_2) = \sum a_{1i} a_{2j} b_{ij}$$

State and prove the generalization of this fact for k-linear forms. Hint: If M is a k-linear form and $f \in \overline{n}^{\overline{k}}$, let $m_f = M(\mathbf{e}_{f(1)}, \mathbf{e}_{f(2)}, \ldots, \mathbf{e}_{f(k)})$. Restate Proposition 5 in terms of the m_f. What if I give you an arbitrary collection of n^k scalars m_f?

6. Side Trip: Some Basics of Permutations

To recap our plot so far: equation 14 says that an alternating multilinear form D in the rows of an $n \times n$ matrix (such as the determinant, if Theorem 1 is true and it actually exists) can be calculated if you know the value of D on the permutation matrices, that is, $D(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)})$. Since D is alternating, that quantity changes sign when we transpose two slots. It's not hard to believe that by successively transposing slots, we can reduce it to $\pm D(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = \pm 1$, with a $+$ sign if there are an even number of transpositions and a $-$ sign for an odd number.

To proceed along these lines, we first need to show that any permutation can be represented as sequential transpositions.

Definition. *A **transposition** is a permutation which interchanges two numbers, i and j and leaves all the others fixed. It is denoted by $(i\ j)$.*

***Exercise 17.** *Prove that any permutation is a composition of transpositions. Hint: Use induction on n, and suppose π is a permutation of $\{1, 2, \ldots, n\}$. Let $j = pi(n)$, and let π' be the permutation $(j\ n) \circ \pi$. Show that $\pi'(n) = n$, so that π' gives a permutation of $\{1, 2, \ldots, n-1\}$. Then π' will be a composition of transpositions. Show that $\pi = (j\ n) \circ \pi'$; the result follows.*

In some sense, we have "factored" π into a product of transpositions. However, this factorization is not unique by any stretch. That is fine... *unless* it were possible to write π as *both* the composition of an even number of transpositions *and* an odd number. In that case $D(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)})$ would be equal to both $+D(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ and $-D(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$, and $D(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ would have to equal 0. In particular, we would have $\det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = 0$; since Property 3 says this quantity is equal to 1, Theorem 1 could not be true. Catastrophe! There would *be* no theory of determinants. Luckily, that can't happen, and we will spend the remainder of this section proving that it can't, primarily as a series of exercises.

We will use polynomials in the n variables $x_1, x_2, \ldots, x_n$ to accomplish that goal. Given such a polynomial $f(x_1, x_2, \ldots, x_n)$ and a permutation π, it is possible to form a new polynomial $S_\pi(f) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$; that is, $S_\pi(f)$ is the result of substituting $x_{\pi(i)}$ for each occurrence of x_i in f.

***Exercise 18.** *Show that composition corresponds to successive substitution, that is, $S_{\pi' \circ \pi}(f) = S_\pi(S_{\pi'}(f))$. (Note the reversal of the order of π and π'.)*

In particular, we will focus on the polynomial

$$P(x_1, x_2, \ldots, x_n) = \prod_{1 \le i < j \le n} (x_i - x_j)$$

of degree $n(n-1)/2$. $S_\pi(P)$ is a product of exactly the same terms, but some of them have their signs reversed. So $S_\pi(P) = \pm P$. We can now give the following

Definition. *The **sign** of the permutation π, $\operatorname{sgn}(\pi)$, is the number ($+1$ or -1) such that*

$$S_\pi(P) = \operatorname{sgn}(\pi) P$$

[Sticklers may wish to fuss about whether S_π respects sums and products, i.e., that it is a ring homomorphism. It is, and those who wish to fuss about it are free to do so. Elsewhere.]

***Exercise 19.** *Use Exercise 18 to prove that $\operatorname{sgn}(\pi' \circ \pi) = \operatorname{sgn}(\pi') \operatorname{sgn}(\pi)$. [In the language of abstract algebra, this shows that sgn is a homomorphism from S_n to $\{1, -1\}$, the multiplicative group of order two.]*

Exercise 20.** *The* ***inversion number *of a permutation π is the number of pairs (i, j) with $i < j$ and $\pi(i) > \pi(j)$. As examples, show that the inversion number of the identity permutation $1_{\overline{n}}$ is 0, and that the inversion number of $\pi(i) = n + 1 - i$ is $n(n-1)/2$. Finally, if k is the inversion number of an arbitrary permutation π, show that $\operatorname{sgn}(\pi) = (-1)^k$.*

***Exercise 21.** *Show that*

(1) $\operatorname{sgn}(1_{\overline{n}}) = 1$
(2) $\operatorname{sgn}(\pi) = \operatorname{sgn}(\pi^{-1})$ *for any permutation π*
(3) sgn *of a transposition $(i\ j)$ is -1. Hint: Show that the inversion number of a transposition is odd.*

We are finally in a position to prove that the catastrophe mentioned above does not come to pass.

***Exercise 22.** *Show that the sign of the composition of an even number of transpositions is $+1$, while the sign of the composition of an odd number of transpositions is -1. Thus, a permutation π cannot be written as the composition of both an even number and an odd number of transpositions.*

The following result is what we came here for; it allows us to further simplify Equation 14.

***Exercise 23.** *Let D be an alternating multilinear form on a vector space V and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$. Show that*

$$D(\mathbf{v}_{\pi(1)}, \mathbf{v}_{\pi(2)}, \ldots, \mathbf{v}_{\pi(n)}) = \operatorname{sgn}(\pi) D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$$

Hint: Express π as a composition of transpositions. Try giving an informal proof, and then a more formal one using induction on the number of transpositions.

We are now in a position to prove our main results.

7. Proof of Theorems 1 and 2

We start with the following lemma:

Proposition 6. *Let D be an alternating n-linear form on E_n, and let $\mathbf{v}_i = \sum_{j=1}^{n} a_{ij} \mathbf{e}_j$ for $1 \leq i \leq n$. Then*

$$D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = D(I_n) \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)}$$

Proof. Equation 14 says:

$$D(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \sum_{\pi \in S_n} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} D(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)})$$

Applying Exercise 23, we have:

$$D(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \sum_{\pi \in S_n} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} \operatorname{sgn}(\pi) D(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$$

The result follows, remembering that when we regard D as a scalar-valued function of matrices, $D(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = D(I_n)$. □

Now suppose that Theorem 1 holds, and det exists. Property 1 says it is multilinear. Property 2 says it is alternating, so Proposition 6 applies. Property 3 says that $\det(I_n) = 1$, so Proposition 6 tells us that the determinant of an $n \times n$ matrix must be given by the following formula:

$$\det([a_{ij}]) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} \tag{15}$$

Let's pause and admire our handiwork for a bit. Starting from three simple properties, and the assumption that something satisfies them, we have just written down a formula for exactly what that something has to be. What does this get us?

First, it tells us that det must be unique—as Theorem 1 claims—since its value must be the one given by this formula. Second, substituting equation 15 back into Proposition 6 gives us

$$D([a_{ij}]) = D(I_n) \det([a_{ij}])$$

That is, we have proved Theorem 2. [As promised in Exercise 2, this proof works over any commutative ring.]

All that is left is the existence part of Theorem 1. To prove it, we simply *define* det by equation 15. Then all we need to do is show that det defined in this way satisfies Properties 1, 2, and 3.

Proof that det*, as defined by equation 15, satisfies Properties 1, 2, and 3.*

Property 1 (det is multilinear): This is left as Exercise 24.

Property 2 (det is alternating): Let $B = [b_{ij}]$ be the matrix A with the k-th and ℓ-th rows interchanged, with $k < \ell$; we need to show that $\det B = -\det A$. Every term in the expansion of $\det B$ has the form

$$\operatorname{sgn}(\pi) b_{1\pi(1)} \cdots b_{k\pi(k)} \cdots b_{\ell\pi(\ell)} \cdots b_{n\pi(n)}$$

Substituting the corresponding entries of A, we have

$$\operatorname{sgn}(\pi) a_{1\pi(1)} \cdots a_{\ell\pi(k)} \cdots a_{k\pi(\ell)} \cdots a_{n\pi(n)}$$

where all the other terms in the product have the form $a_{i\pi(i)}$. Transposing the k and ℓ terms, so that the rows are back in ascending order by the first index, gives us

$$\operatorname{sgn}(\pi) a_{1\pi(1)} \cdots a_{k\pi(\ell)} \cdots a_{\ell\pi(k)} \cdots a_{n\pi(n)}$$

Letting $\pi' = \pi \circ (k\ \ell)$, this is

$$\operatorname{sgn}(\pi) a_{1\pi'(1)} \cdots a_{k\pi'(k)} \cdots a_{\ell\pi'(\ell)} \cdots a_{n\pi'(n)}$$

By Exercises 19 and 21, $sgn(\pi') = -sgn(\pi)$, so this term is

$$-\operatorname{sgn}(\pi') a_{1\pi'(1)} \cdots a_{k\pi'(k)} \cdots a_{\ell\pi'(\ell)} \cdots a_{n\pi'(n)} \tag{16}$$

These are simply the terms of $\det A$ with a minus sign out front, so (with one detail to be verified in Exercise 25), we have shown that $\det B = -\det A$.

Property 3 ($\det I_n = 1$): $I_n = [\delta_{ij}]$, so the terms of $\det I_n$ are:

$$\operatorname{sgn}(\pi) \delta_{\pi(1)1} \delta_{\pi(2)2} \cdots \delta_{\pi(n)n}$$

which vanishes unless $\pi(1) = 1, \pi(2) = 2, \ldots, \pi(n) = n$—in other words, $\pi = 1_{\overline{n}}$. Since $\operatorname{sgn}(1_{\overline{n}}) = 1$ by Exercise 21, the only non-vanishing term is a 1, and the proof is complete. □

***Exercise 24.** *Prove that* det*, as defined by equation 15, satisfies Property 1, i.e., is multilinear. If you were devoted enough to prove the optional Exercise 16, show that it follows from that. If not, prove it directly; it follows relatively straightforwardly from the fact that each of the n factors in each term of the determinant depend linearly on a single entry from row i.*

Exercise 25. *Complete the proof of Property 2 above by showing that*

$$\pi \leftrightarrow \pi' = \pi \circ (k\ \ell)$$

is a one-to-one correspondence, so that the sum over $\pi \in S_n$ of all the terms in equation 16 is in fact equal to $-\det A$.

Exercise 26. *Extend the proof of Property 3 given above to show that the determinant of a diagonal matrix is the product of the diagonal entries. Extend it even further to show the same thing for an upper-triangular matrix, the same result as Exercise 8.*

Exercise 27. *Give two proofs of the fact that the determinant of a permutation matrix (see Exercise 14) is the sign of the permutation. Hint: (a) Extend the proof of Property 3 above to show that the determinant of a permutation matrix has exactly one non-vanishing term, whose value is the sign of the permutation. (b) Use Exercise 23.*

8. Other Basic Properties of the Determinant

Starting from three simple properties of the determinant, we have proved a number of basic facts about it including the Laplace expansion and Cramer's rule; developed a formula for what it must be; and shown that that formula does in fact satisfy the three properties we started with. There are only a few more key properties left to establish, which turn out to be surprisingly easy.

Theorem 7. *The determinant is invariant under transposition. That is, if A is an $n \times n$ matrix,* $\det A = \det A^T$

***Exercise 28.** *Prove Theorem 7. Hint: Use commutativity of multiplication to show that*

$$a_{1\pi(1)}a_{2\pi(2)}\cdots a_{n\pi(n)} = a_{\pi^{-1}(1)1}a_{\pi^{-1}(2)2}\cdots a_{\pi^{-1}(n)n}$$

Complete the proof by showing that $\pi \leftrightarrow \pi^{-1}$ is a one-to-one correspondence, and using the fact that $\operatorname{sgn}(\pi) = \operatorname{sgn}(\pi^{-1})$*, from Exercise 21.*

It follows from Theorem 7 that everything we have said about determinants with respect to rows of a matrix is equally true for columns of a matrix.

Corollary. $\det A$ *is alternating and multilinear in the columns of A, in addition to the rows of A.*

***Exercise 29.** *State and give whatever proof is necessary for a version of the Laplace expansion (Theorem 3) based on a column of a matrix A rather than a row.*

There are certain situations—such as the next proof—in which it is convenient to view $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ as columns of a matrix rather than as rows, particularly since linear transformations are almost universally thought of as acting on column vectors. For that reason, a reader might believe that it would have been better to center this article around column vectors rather than row vectors. In fact, I started it that way, but after a while it felt notationally offensive to write $\mathbf{v}_j = \sum_{i=1}^n a_{ij}\mathbf{e}_i$ instead of $\mathbf{v}_i = \sum_{j=1}^n a_{ij}\mathbf{e}_j$ and $a_{\pi(1)1}a_{\pi(2)2}\cdots a_{\pi(k)k}$ instead of $a_{1\pi(1)}a_{2\pi(2)}\cdots a_{k\pi(k)}$. So I switched to rows, and only occasionally look back.

Theorem 8. *If A and B are $n \times n$ matrices,* $\det AB = \det A \cdot \det B$

Proof. Let $D(M) = \det AM$ for all $n \times n$ matrices M. By the definition of matrix multiplication, the j-th column of AM is A multiplied by the j-th column of M. It follows that D is multilinear and alternating in the columns of M. By Theorem 2,

$$D(B) = D(I_n) \cdot \det B = \det AI_n \cdot \det B = \det A \cdot \det B$$

Since $D(B)$ is defined to be $\det AB$, the result follows. □

***Exercise 30.** *Show that if A is an invertible $n \times n$ matrix, then $\det A \neq 0$ and that*

$$\det A^{-1} = \frac{1}{\det A}$$

***Exercise 31.** *Show that if A is an orthogonal matrix (i.e., $A^{-1} = A^T$), that* $\det A = \pm 1$.

***Exercise 32.** *Show that* $\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 0 \Leftrightarrow$ *the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly dependent. Hint: Look at the matrix whose columns (or rows) are the the $\mathbf{v}_j$; it is invertible if an only if the $\mathbf{v}_j$ are linearly independent. Then use Theorem 4 and Exercise 30. While you're at it, give an alternate proof of $\Leftarrow$ (the easy part), using the fact that if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly dependent, one of the $\mathbf{v}_i$ can be written as a linear combination of the others.*

***Exercise 33.** *If T is a linear transformation of an n-dimensional vector space V to itself, show that the notion of the determinant of T is well-defined. Hint: Use the fact that if A is the matrix representation of T in one choice of basis, the representation of T in another basis is $B^{-1}AB$ for some $n \times n$ matrix B.*

Exercise 34. *Extend Theorem 4 to show that*

$$(\operatorname{adj} A)A = A \operatorname{adj} A = (\det A)I_n$$

Hint: What happens if you apply Theorem 4 to A^T?

Exercise 35. *Readers familiar with with exterior algebra may enjoy trying to understand Theorem 4 in coordinate-free language. Hint: (1) The matrix* $(\operatorname{adj} A)A$ *actually represents a bilinear form, not a linear transformation, and (2) you will need to understand the relationship between the matrix of cofactors and the Hodge dual of $\bigwedge^{n-1} A$.*

Exercise 36. *Give a third proof of the proposition stated in Exercise 27. Hint: Show that the product of permutation matrices is the permutation matrix of their composition. Show that the permutation matrix of a transposition is the identity matrix with two rows interchanged and therefore by Property 2 has determinant -1. Use Theorem 8 and the fact that any permutation is a composition of transpositions.*

9. The Determinant and Unsigned Volume

In Section 2, we introduced determinants and suggested that the determinant of a matrix could be interpreted geometrically as the signed volume of the parallelepiped spanned by its rows. (We now know by Theorem 7 that we can equally well look at the signed volume of the parallelepiped spanned by its columns.)

The most elegant ending to our determinant story would be to give a purely geometric definition of the signed volume of a parallelepiped, and then show that it satisfies Properties 1-3. Theorem 1 would then tell us that the signed volume is equal to the determinant. In fact, in Section 2 we argued that the two-dimensional area does does satisfy Properties 1-3. The problem is that there is no simple geometric way to arrive at the correct sign for the n-dimensional volume independent of the determinant. In my view it is likely to be possible, but unlikely to be worth the trouble. Readers may recall that even in the two-dimensional case we had to resort to some "hand-waving" to define the sign geometrically.

Although we will have to take a less slick approach to prove an actual theorem, we can get some motivation by taking an informal look at Properties 1-3 in relation to geometrically-defined volume. The arguments we gave in Section 2 for linearity (Property 1) still work if you are willing to accept that the volume of a parallelepiped is the $n-1$-dimensional volume of its base multiplied by its height; we shall see below that this last point is closely related to Property 2b. If two edges of a parallelepiped are the same, its volume is 0 because it lies in an $n-1$-dimensional subspace; therefore Property 2a holds. Finally, Property 3 just says that a unit n-cube has volume 1.

In proving our results, we will use $\mathrm{vol}(A)$ to denote the volume of a set $A \subset E_n$.

Definition. *The **parallelepiped** spanned by the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is the set of points*

$$P(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \{t_1\mathbf{v}_1 + t_1\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k \mid 0 \le t_i \le 1\}$$

*It is known as a **rectangular parallelepiped** if the $\mathbf{v}_i$ are all mutually perpendicular.*

A rectangular parallelepiped is simply the n-dimensional generalization of a rectangle.

Theorem 9. *If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are vectors in E_n,*

$$\mathrm{vol}(P(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)) = |\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)|$$

Our proof of Theorem 9 will not include a definition of volume. Readers familiar with Lebesgue measure may use that definition, although it is massive overkill for our purposes. We will assume that our function vol that measures n-dimensional volume is defined on some collection of subset of E_n; we won't be too fussy about what those subsets are, but you will see from what follows that they are not very complicated. Let's assume that our collection of subsets is closed under finite union, intersection, and difference, and that all of our sets are bounded. Let's further suppose that vol satisfies the following basic properties for sets on which it is defined:

(V1) $\mathrm{vol}(A) \ge 0$

(V2) $\mathrm{vol}(A \cup B) = \mathrm{vol}(A) + \mathrm{vol}(B) - \mathrm{vol}(A \cap B)$

(V3) vol is unchanged by translation or rotation (application of an orthogonal linear transformation)

(V4) If a set A lies in an $n-1$ dimensional subspace of E_n, then $\mathrm{vol}(A) = 0$
(V5) vol of a rectangular parallelepiped (n-dimensional rectangle) is the product of the lengths of the edges

In approaching Theorem 9, note first that if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly dependent, Theorem 9 holds because the determinant is 0 (by Exercise 32) and the volume is also 0 (by property V4). Thus we can assume that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent.

We start with an analog of Property 2b.

Proposition 10. $\mathrm{vol}(P(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)) = \mathrm{vol}(P(\mathbf{v}_1, \mathbf{v}_2 + t\mathbf{v}_1, \ldots, \mathbf{v}_n))$ *for all* $t \in \mathbb{R}$

Note that P does not depend on the order of the $\mathbf{v}_i$, so that Proposition 10 actually applies to any pair of the $\mathbf{v}_i$, not just $\mathbf{v}_1$ and $\mathbf{v}_2$.

To prove Proposition 10, we first prove it for any t in the closed interval $[-1, 1]$; the full result follows by repeated applications of this restricted case.

First consider the case $n = 2$. Figure 6 illustrates a cut-and-paste argument for the value $t = -0.4$. A similar diagram works for values of $t \in [0, 1]$. The argument can be extended to $n > 2$ as illustrated in Figure 7.

Exercise 37. *Complete the proof of Proposition 10 sketched above. The validity of your cut-and-paste argument can be demonstrated using properties V2-V4 of the volume function.*

Proof of Theorem 9. If P is the parallelepiped $P(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$, we need to show that its volume is equal to the absolute value of the determinant of the matrix M whose columns are equal to the $\mathbf{v}_i$, under the assumption that the $\mathbf{v}_i$ are linearly independent.

Lets look at the special in which P is a rectangular parallelepiped with $\mathbf{v}_i$ a scalar multiple of $\mathbf{e}_i$, say $c_i\mathbf{e}_i$. Then M is a diagonal matrix with c_i along the diagonal,

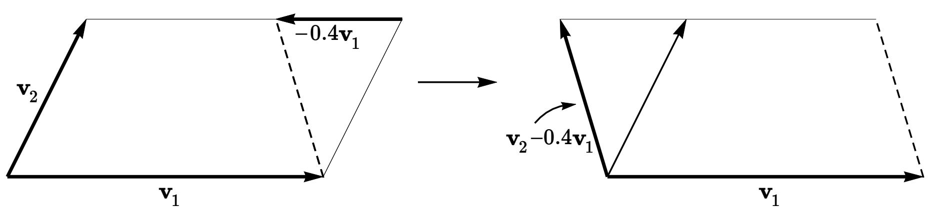

FIGURE 6. Proposition 10 in dimension $n = 2$, with $t = -0.4$

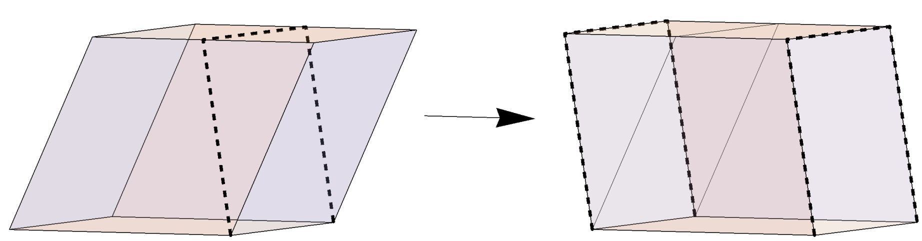

FIGURE 7. Proposition 10 in dimension $n = 3$

and $\det M$ is the product of the c_i by Exercise 8 (or Exercise 26). $|\det M|$ is the product of the $|c_i|$. But $|c_i|$ is the length of the i-th edge of our parallelepiped, and so our the result follows by property V5.

To reduce an arbitrary rectangular parallelepiped P to the previous case, find an orthogonal matrix O which rotates the each $\mathbf{v}_i$ to be a multiple of some basis vector. Note that $|\det M|$ does not depend on the order of the $\mathbf{v}_i$, so we can reorder the $\mathbf{v}_i$ so that $O\mathbf{v}_i = c_i\mathbf{e}_i$. By property V3, transformation by O does not change the volume of P, so

$$\mathrm{vol}(P) = \mathrm{vol}(O(P)) = |\det OM| = |\det O|\,|\det M|$$

The result follows by Exercise 31.

We can reduce the general case of linearly independent $\mathbf{v}_i$ to the rectangular case by applying a process of orthogonalization. The usual Gram-Schmidt orthonormalization process works fine if you leave out the normalization part.

We start by letting $\mathbf{v}'_1 = \mathbf{v}_1$. By orthogonal projection (see Figure 8), we can write $\mathbf{v}_2$ as the sum of a vector $\mathbf{v}'_2$ which is perpendicular to $\mathbf{v}'_1$ and some multiple of $\mathbf{v}'_1$, say $a_1\mathbf{v}'_1$. Then $\mathbf{v}'_2 = \mathbf{v}_2 - a_1\mathbf{v}'_1$.

To move from $\mathbf{v}_j$ to $\mathbf{v}'_j$, refer to Figure 9, depicting the case $j = 3$. Again by orthogonal projection we can write $\mathbf{v}_j$ as a vector $\mathbf{v}'_j$ (perpendicular to all the $\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_{j-1}$) and some linear combination of the $\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_{j-1}$, call it $\sum_{i=1}^{j-1} a_i\mathbf{v}'_i$. So

$$\mathbf{v}'_j = \mathbf{v}_j - \sum_{i=1}^{j-1} a_i\mathbf{v}'_i$$

When we replace $\mathbf{v}_j$ with $\mathbf{v}'_j$, we modify $\mathbf{v}_j$ by adding scalar multiples of other vectors $\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_{j-1}$. By Proposition 10 this does not change the volume of the parallelepiped, and by Property 2b it does not change the value (and hence the absolute value) of the determinant. At the end of the process, the $\mathbf{v}'_i$ are mutually perpendicular, and we have reduced the general case to the previous rectangular case, proving our theorem. □

Exercise 38. *Show that the volume of a parallelepiped is the volume of its base times its height. Hint: You can extend the argument given in the proof of Theorem 9; if the vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$ *span the base, the vector* $\mathbf{v}'_n$ *represents the height. Alternatively, now that we proved Theorem 9 you can use determinants! It's probably worth understanding both.*

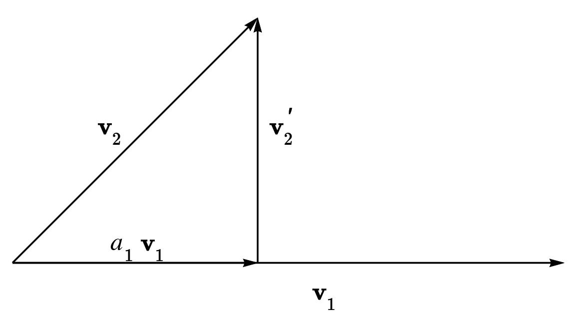

FIGURE 8. The first (non-trivial) step of orthogonalization

Exercise 39. *Show that property V5 of* vol *can be deduced from the weaker assumption that the volume of the unit n-dimensional cube* $\mathrm{vol}(P(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)) = 1$*, and that property V4 can be eliminated entirely. Hint: Use a cut-and-paste argument to deduce V5 in the case where the lengths of the edges are rational numbers. Show that if $A \subset B$, then* $\mathrm{vol}(A) \leq vol(B)$*, and use this to prove the result for arbitrary real edge lengths. Use a similar argument to show that the volume of a bounded subset of an $n-1$-dimensional space is 0.*

Exercise 40. *Use a row-reduction argument to show that we do not need to assume that* vol *is unchanged by orthogonal transformations.*

10. Signed Volume and Orientation

Now that we have dealt with unsigned volume, we will close our treatment of determinants with a brief discussion of signed volume and orientation.

As discussed in the previous section, the most straightforward approach to signed volume is to simply define the signed volume of the parallelepiped in E_n spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ to be $\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. Obviously this depends on the order of the $\mathbf{v}_i$; Property 2 tells us that transposing two of them changes the sign. However, the absolute value remains unchanged, and Theorem 9 tells us that the absolute value is equal to the geometric volume of the parallelepiped $P(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$.

What about volumes in an arbitrary n-dimensional vector space V over the real numbers? The situation in E_n is special because not only does it have a preferred basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ which we have used to normalize volume, that basis has a preferred order, which determines the sign of the volume. When we drop the normalization condition (Property 3) on the determinant, we are left with an alternating n-linear form. This suggests that the way to deal with volume and orientation on an arbitrary n-dimensional vector space is to provide a non-zero alternating n-linear form Ω, which we'll call a **volume form**. Given a choice of Ω, we define the signed volume of the parallelepiped spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ to be $\Omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$; the unsigned volume is its absolute value. Choosing Ω fixes the scale of volume, as the following exercise makes clear.

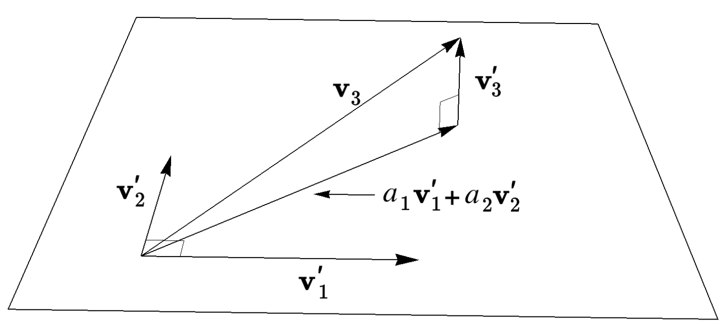

Figure 9. The next step of orthogonalization

Exercise 41. *If V is an n-dimensional vector space, show that a volume form (non-zero alternating n-linear form) Ω exists, and that any other alternating n-linear form on V is a scalar multiple of it. Hint: Here are three entirely equivalent approaches to this problem. They all start with choosing a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ for V. (1) You can repeat the Proofs of Theorems 1 and 2 to come up with an alternating n-linear form Ω such that $\Omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 1$. (2) Given n vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n \in V$, write $\mathbf{w}_i$ in terms of the $\mathbf{v}_j$ as $\mathbf{w}_i = \sum_{j=1}^{n} a_{ij}\mathbf{v}_j$. Then let*

$$\Omega(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n) = \det([a_{ij}])$$

(3) Use the basis to set up an isomorphism (non-singular linear transformation) $T : V \to E_n$. Let Ω be the "pullback" of det, *i.e.,*

$$\Omega(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n) = \det(T(\mathbf{w}_1), T(\mathbf{w}_2), \ldots, T(\mathbf{w}_n)) \tag{17}$$

Exercise 41 is valid over any field of scalars, and it demonstrates that the alternating n-linear forms on an n-dimensional vector space V themselves form a 1-dimensional vector space. [In the language of exterior algebra, this vectors space is known as $\bigwedge^n(V^*)$. $\Omega = T^*(\det)$ is a restatement of equation 17 using pullback operator T^*. Curious readers may be able to work out a definition of T^* from equation 17.]

Working over $\mathbb{R}$, the choice of a volume form Ω not only fixes a scale for volume on V, it allows us to talk about the orientation of bases of V. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis for V (actually, an ordered basis), we say that it is **positively oriented** if $\Omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) > 0$; we say that it is **negatively oriented** if $\Omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) < 0$. Thus by definition, the sign of the signed volume reflects the orientation of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ as a basis, while the absolute value represents the geometric (unsigned) volume of the parallelepiped spanned by the $\mathbf{v}_i$. There is one detail to clean up, in the following exercise.

Exercise 42. *If Ω is a volume form on an n-dimensional vector space V, show that $\Omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 0 \Leftrightarrow$ the $\mathbf{v}_i$ are linearly dependent. Hint: use Exercise 32.*

By Exercise 41, a different choice of volume form will be some scalar c times Ω. Unsigned volumes will differ by a factor of $|c|$. Furthermore, if $c > 0$, the two volume forms will identify the same bases as positively- and negatively-oriented; if $c < 0$, they will identify bases oppositely.

So while there is no a priori standard of volume on an arbitrary vector space V, any two volume forms will assign the same *ratio* of volumes to a pair of parallelepipeds. Similarly, while there is no a priori standard of orientation on an arbitrary vector space V, any two volume forms will agree on whether two bases have the same or the opposite orientation.

Exercise 43. *Formalize and prove the statements of the last paragraph.*

As a result, if T is a linear transformation from a vector space V to itself, the ratio of the volume of the image under T of a parallelepiped to the volume of the original parallelepiped is independent of any choice of volume form. Amazingly, this ratio is given by $\det T$.

Proposition 11. *Let T be a linear transformation of an n-dimensional vector space V to itself, let Ω be a volume form on V, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be vectors in V. Then*

$$\Omega(T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)) = \det T \cdot \Omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$$

[In the language of exterior algebra, the pullback operator T^* operating on the one-dimensional space of alternating n-linear forms on V acts as multiplication by the scalar $\det T$.]

***Exercise 44.** *Prove Proposition 11. Hint: Show that Ω' defined by*

$$\Omega'(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \Omega(T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n))$$

is an alternating n-linear form. By Exercise 41, Ω' is a scalar times Ω. Prove that the scalar is $\det T$. *[The pullback appears yet again!]*

It follows that $|\det T|$ tells us the factor by which T expands geometric volume for parallelepipeds. In the spirit of our earlier work with vol you can show that T expands volume for other sets by the same factor. This is why the absolute value of the determinant of the Jacobian matrix appears in the change-of-variable formula in multivariable calculus.

The sign of $\det T$ also has a geometric interpretation:

Exercise 45. *Let T be a linear transformation T of an n-dimensional vector space V to itself. Give reasonable definitions of what it means to say that T is* ***orientation-preserving*** *or that T is* ***orientation-reversing****, and show that your definitions are independent of the choice of volume form on V. Prove that T is orientation-preserving if* $\det T > 0$ *and orientation-reversing if* $\det T < 0$.

An equivalent way looking at orientation is this: bases for a vector space over $\mathbb{R}$ are naturally divided into two collections. An **orientation** for a vector space V is the choice of one of these collections as "positive" and the other as "negative." Thus, any volume form on V will give an orientation. A positive scalar multiple of the volume form will give the same orientation, and a negative scalar multiple will give the opposite orientation.

The notion of orientation is important in the study of n-manifolds, that is, n-dimensional surfaces. An orientation on a manifold is a choice of an orientation for each tangent space in a way that varies smoothly. Most well-known manifolds have two possible orientations, but there are manifolds such as the Möbius band and the Klein bottle for which no orientation exists.

In studying calculus on manifolds, such as Cartan's generalization of Stokes' Theorem, the geometrically natural thing to integrate on an (oriented) n-manifold is an "n-form," a smoothly varying choice of a volume form on the tangent space at each point of M. In introductory calculus, we form an integral by in some sense breaking the closed interval $[a, b]$ up into small segments, adding bits of area together, and taking the limit as the segment size goes to zero. The corresponding process on a manifold breaks the manifold into small (approximate) parallelepipeds, uses the volume form to assign them volume, adds them together, and takes the limit as their size goes to zero. In both cases there is a significant amount of work to show that this process results in a coherent answer, but the theory of alternating multilinear forms is a key element of that work in the n-dimensional case.