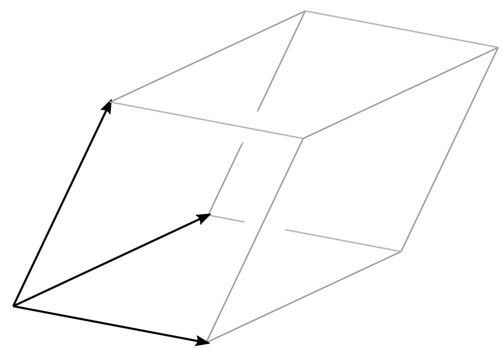
DETERMINANTS

DAN MARGALIT

Let's start by thinking about the relationship between invertibility and volumes. If we have a matrix A, we can look at the parallelepiped determined by the rows of A:



This has some n–dimensional volume. Let's call that the *volume* of A. What can we do with this? We have the following idea:

> *A is invertible if and only if the volume of A is nonzero.*

This is not hard to believe: the volume is zero exactly when the parallelepiped is flat, and this happens exactly when the rows are linearly dependent, and we already know that is equivalent to invertibility. You can really see this in the case when A is a 2×2 matrix. In this case the parallelepiped is just a parallelogram, and the area of the parallelogram is zero exactly when the rows of A are linearly dependent.

So what is a determinant? We'll see that a determinant is a way of assigning to each square matrix A a number $\det(A)$. Moreover, we'll see that $\det(A)$ is zero exactly when A is not invertible—just like volume! In fact we will see that the absolute value of the determinant exactly is the volume of the parallelepiped described above. We'll also describe a few different ways of computing the determinant, we'll use the determinant to give a formula for the inverse of a matrix, and we'll see how determinants can be thought of as stretch factors for linear transformations. That's a lot to do, so let's get going.

Written for Math 1553 at the Georgia Institute of Technology, 2015

1. THE COFACTOR EXPANSION

We are now going to give a formula for computing determinants. This formula is called the cofactor expansion of the determinant. We are not going to explain right away where this formula comes from or what it has to do with volume. That will come later.

Our formula for the determinant is going to be recursive, which means that the determinant of an $n \times n$ matrix is going to be defined in terms of the determinants of $(n-1)\times(n-1)$ matrices. So as long as there is some small n where you know how to compute the determinant of an $n \times n$ matrix, you will then be able to compute the determinants of all larger matrices.

Let's start with 1×1 matrices. Again, the idea of a determinant is to assign a real number to every matrix. Well, a 1×1 matrix already is a real number. So in this case the determinant is given by the formula:

$$\det(a_{11}) = a_{11}.$$

To state our first formula for the determinant of an $n \times n$ matrix we need some more notation. Let A be an $n \times n$ matrix and denote the ijth entry of A by a_{ij}. The ijth minor of A is the $(n-1) \times (n-1)$ matrix A_{ij} obtained from A by deleting the ith row and the jth column. Our first formula for the determinant of A is as follows:

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

If we write this out, it is:

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} - \cdots + (-1)^{n+1} a_{1n} \det A_{1n}.$$

Let's try this formula out, first for 2×2 matrices. Say that

$$A = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Since $n = 2$ we have

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12}.$$

We can see that $A_{11} = (a_{22})$ and $A_{12} = (a_{21})$ and so

$$\det A = a_{11} a_{22} - a_{12} a_{21}.$$

But that is the formula for the determinant of a 2×2 matrix that we already learned. And we already saw that A is invertible exactly when this number is nonzero. This is a good sign!

Let's write down the formula for 3×3 matrices. Say that

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Applying our formula directly we see that

$$\begin{aligned}
\det A &= a_{11}\det A_{11} - a_{12}\det A_{12} + a_{13}\det A_{13} \\
&= a_{11}\det\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12}\det\begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13}\det\begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \\
&= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})
\end{aligned}$$

A trick for 3×3 *matrices*. We can rearrange the formula for the determinant of a 3×3 matrix as follows:

$$\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

The first term is the product of the diagonal entries. The second term is the product of the three terms to the right of the diagonal. For this to make sense, we think of a_{31} as being to the right of a_{33} (think about the old-fashioned video games where a character moving off of the right-hand side of the screen reappears on the left). And the third term is the product of the three terms to the left of the diagonal. Similarly, the fourth term is the negation of the product of the three terms on the anti-diagonal, etc.

We won't write down the formula for $n \times n$ matrices with $n > 3$, but hopefully you are convinced that you could write it down if you wanted to.

Cofactors and other formulas for the determinant. The ijth cofactor of a matrix A is:

$$C_{ij} = (-1)^{i+j}\det A_{ij}.$$

With this notation, we can rewrite the above formula for $\det A$ in the following nice form:

$$\det A = \sum_{j=1}^{n} a_{1j}C_{1j}.$$

This is called the *cofactor expansion* of $\det A$ across the first row of A, since the a_{1j} entries are exactly the entries in the first row of A. It turns out that we can compute the determinant by doing a similar cofactor expansion across any row:

$$\det A = \sum_{j=1}^{n} a_{ij}C_{ij} \text{ for any fixed } i.$$

And what is more we can do analogous cofactor expansions down any column of A as well:

$$\det A = \sum_{i=1}^{n} a_{ij}C_{ij} \text{ for any fixed } j.$$

So there are $2n$ different formulas for $\det A$! You should check that the 6 cofactor expansions for the determinant of the above 3×3 matrix all give the same formula.

Now that we have all of these formulas for the determinant, we want to answer some questions:

- Where do the formulas come from?

- Why do they all give the same number?
- What does the determinant mean?

2. Determinants and row operations

The next step might seem out of the blue. Let's say that a *determinant* is a function

$$\det : \{\text{square matrices}\} \to \mathbf{R}$$

that satisfies four properties:

(1) if we perform a row replacement then det does not change
(2) if we perform a row swap then det changes by a factor of -1
(3) if we scale a row by k then det changes by a factor of k
(4) $\det(I_n) = 1$

Why would we do this? We are trying to mimic the idea of volume. Ignoring the second item for a moment, the first, third, and fourth properties exactly correspond to our volume. Indeed if we have a parallelepiped and scale one edge by a factor of k, the volume also scales by k. To see this, think of computing the volume of the parallelepiped as base times height, where the base is determined by the $n-1$ vectors that we are not scaling and the height is given by the vector that we are scaling. When we scale this vector by k we scale the height, hence the volume, by k. (Again, think about the 2–dimensional case.)

Similarly, if we do a row replacement the volume doesn't change. Why? Again think of the volume of the parallelepiped as base times height where the base is given by the vectors not being replaced and the height is given by the vector being replaced. When we do the row replacement, the height vector moves parallel to the base, so the height does not change. The base obviously does not change, and so as a result the volume does not change. (Again, think about the 2–dimensional case.)

The condition $\det(I_n) = 1$ is a way of calibrating our volume. We can think of this as saying that the volume of the unit cube is 1, which is of course what we want.

The second condition above is a little more subtle. It basically means that this function det (if it exists at all) should be thought of as a *signed* volume, not a volume. You saw signed areas in Calculus, and this is not so different. If you have a parallelogram, and the first and second row vectors are roughly arranged like the x- and y-axes, then we have positive area. If they are arranged in the other order, we will say that the parallelogram has negative area.

Existence and uniqueness. Here is the first theorem about determinants.

Theorem 1. There is a function

$$\det : \{\text{square matrices}\} \to \mathbf{R}$$

with the 4 properties listed above, and in fact there is only one such function.

Because of this theorem we are justified in referring to det as *the* determinant.

Theorem 1 is our main theorem about determinants—without it we would have nothing! We will prove it at the end of these notes. For now, let us just assume it is true.

Basic fact about the determinant. Assuming Theorem 1, how do we compute determinants? This does not seem so easy. For instance, what is the area of the parallelogram determined by $(7, 11)$ and $(-1, 5)$. This is not so obvious!

We'll build up a method for computing determinants in steps. First we have a couple of preliminary steps.

Fact 0. A square matrix with a zero row has determinant zero.

Let's see this by example. Say that

$$A = \det\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}.$$

Using the third property of determinants, we have:

$$2 \cdot \det A = 2 \cdot \det\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} = \det\begin{pmatrix} 3 & 0 \\ 2 \cdot 0 & 2 \cdot 0 \end{pmatrix} = \det\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} = \det A.$$

So

$$2 \cdot \det A = \det A.$$

It follows that $\det A = 0$, just like we wanted.

We have one more preliminary step. This next fact embodies the idea that we started with at the very beginning.

Fact 1. A square matrix is invertible if and only if its determinant is nonzero.

First suppose A is invertible. We know from the Invertible Matrix Theorem that A is row equivalent to the identity. Combining the first three properties of the determinant, we see that any sequence of row operations taking A to the identity matrix I will scale the determinant by a nonzero number. Since I has determinant 1, it follows that A has nonzero determinant.

If A is not invertible, then it is row equivalent to a matrix B with a zero row. By Fact 0, $\det B = 0$. Again, since the sequence of row operations taking A to B scales the determinant by a nonzero number, we must have that $\det A = 0$. That proves Fact 1!

Diagonal matrices and triangular matrices. Now let's move on to diagonal matrices.

> *Fact 2. The determinant of a diagonal matrix is the product of the diagonal entries.*

Let A be a diagonal matrix. First suppose that one of the diagonal entries of A is zero. This means that A has a zero row. And so by Fact 0, $\det A = 0$. But the product of the diagonal entries is also zero, so this proves Fact 2 in this case.

Now suppose that A is a diagonal matrix and that all of its diagonal entries are nonzero. To get from A to I by row operations, we just need to divide each row of A by whatever entry we see on the diagonal in that row. Fact 2 then follows from the third and fourth parts of the definition of the determinant.

For example:

$$\det\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} = 3\cdot\det\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = 6\cdot\det\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 6\cdot 1 = 6.$$

Now let's deal with triangular matrices.

Fact 3. The determinant of an upper (or lower) triangular matrix is the product of the diagonal entries.

Let A be an upper triangular matrix. If one of the diagonal entries of A is zero then A is not invertible (it will have a column without a pivot). By Fact 1, we have $\det(A)$ is equal to zero, which is exactly the product of the diagonal entries of A.

So let's now assume that all of the diagonal entries of A are nonzero. There is a sequence of row replacements that takes A to a diagonal matrix D that has the same diagonal entries as A. Row replacements do not affect the determinant, and so $\det(A) = \det(D)$. By Fact 1, $\det(D)$ is equal to the product of the diagonal entries of D, which are nothing other than the diagonal entries of A. Since $\det(A) = \det(D)$, this does it. Here is an example:

$$\det\begin{pmatrix} 3 & 2 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \det\begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \det\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 6.$$

Computing the determinant of any square matrix. We now have a method for computing the determinant of any square matrix: we just row reduce the matrix and keep track of any row scaling we do along the way. Here is an example:

$$\det\begin{pmatrix} 1 & 2 & 3 \\ 2 & -1 & 1 \\ 3 & 0 & -1 \end{pmatrix} = \det\begin{pmatrix} 1 & 2 & 3 \\ 0 & -5 & -5 \\ 0 & -6 & -10 \end{pmatrix} = -5\det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & -6 & -10 \end{pmatrix}$$
$$= -5\det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -4 \end{pmatrix} = -5(1\cdot 1\cdot -4) = 20.$$

The first step in the example is a pair of row replacements (which doesn't change the determinant), the second step is a row scaling (that's where we pick up the factor of

-5), and the third step is another row replacement (which again doesn't change the determinant). At the end we have an upper triangular matrix, and by Fact 3 we know its determinant is the product of the diagonal entries. That does it.

It is always possible to row reduce a matrix without scaling any rows along the way. If we only do row replacement and row swaps, then the determinant of the row reduced matrix, which is just the product of the diagonal entries, is equal to plus or minus the determinant of the original matrix. More precisely:

> *Fact 4. If we row reduce a matrix without row scaling then the determinant of the original matrix is*
>
> $$(-1)^{swaps}(\text{product of diagonal entries of REF}).$$

Let's do this for a general 2×2 matrix: For example:

$$\begin{aligned}
\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \frac{1}{a} \cdot \det\begin{pmatrix} a & b \\ ac & ad \end{pmatrix} \\
&= \frac{1}{a} \cdot \det\begin{pmatrix} a & b \\ 0 & ad - bc \end{pmatrix} \\
&= \frac{1}{a} \cdot a \cdot (ad - bc) \\
&= ad - bc.
\end{aligned}$$

This is exactly the formula we expected!

Determinants and products. Our next goal is to show the following very important formula.

Theorem 2. If A and B are square matrices then

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

We'll do this in a few steps.

Fact 5. If E is an elementary matrix and B is any matrix then $\det(EB) = \det(E)\det(B)$.

Why is this true? Let's first suppose that E is an elementary matrix corresponding to row replacement. Then E is triangular and has only 1's on the diagonal, and so $\det(E) = 1$. And so in this case what we are trying to show is $\det(EB) = \det(B)$. But EB differs from B by row replacement, so rule 1 of determinants says this is the case. What about the other two kinds of elementary matrices? The argument is basically the same, so I'll leave it to you.

We can now see:

Fact 6. If A is a product of elementary matrices $E_1 \cdots E_m$ then $\det(A) = \det(E_1) \cdots \det(E_m)$.

To verify this, just apply Fact 5 repeatedly!

But now we can easily prove Theorem 2. Suppose that A is a product of elementary matrices $E_1 \cdots E_m$ and B is a product of elementary matrices $E_1' \cdots E_n'$. Well then AB is the product of elementary matrices $E_1 \cdots E_m E_1' \cdots E_n'$. So applying Fact 6 twice, we have:

$$\begin{aligned}
\det(AB) &= \det(E_1 \cdots E_m E_1' \cdots E_n') \\
&= \det(E_1) \cdots \det(E_m) \det(E_1') \cdots \det(E_n') \\
&= \det(A) \det(B).
\end{aligned}$$

3. Cramer's rule and Inverses

You might remember that there is a formula for the inverse of a 2×2 matrix that uses the determinant:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

And you might wonder if there is an analogue of this formula for $n \times n$ matrices when $n \geq 3$. It turns out there is. Here is how it goes. Say that A is an $n \times n$ matrix. The cofactor matrix of A is the $n \times n$ matrix (C_{ij}) whose ij-entry is the ijth cofactor of A. Then the formula is:

Theorem 3. If A is an $n \times n$ matrix with nonzero determinant, then

$$A^{-1} = \frac{1}{\det A} (C_{ij})^T.$$

We can expand this out as:

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{pmatrix}.$$

Let's do a sanity check. The cofactor matrix for the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is just

$$\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

and the transpose of the cofactor matrix is then

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

If we divide this by the determinant $ad - bc$ we get the usual formula for the inverse of a 2×2 matrix!

Theorem 3 is not the fastest way to compute the inverse of a matrix, whether you are doing it by hand or by computer. To find the cofactor matrix you have to compute n^2 determinants. Our old method was row reduction, which is pretty fast. The main point is that it is an elegant formula and that it generalizes what we already knew for 2×2 matrices.

Let's explain why Theorem 3 is true. First, we need the following.

Fact 7 (Cramer's rule). Suppose that $x = (x_1, \ldots, x_n)$ is the solution to $Ax = b$ where A is an invertible $n \times n$ matrix. Let A_i be the matrix obtained from A by replacing the ith column of A by b. Then

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

Here is a nifty proof of Cramer's rule. The normal way to solve $Ax = b$ is to row reduce the augmented matrix $(A|b)$. Let's just check that the quotient $(\det A_i)/(\det A)$ never changes when we do a row operation on $(A|b)$. If we do a row replacement then neither the numerator nor the denominator change, so the quotient doesn't change. If we scale a row, then both change by the same factor, so the quotient doesn't change. And if we swap two rows then both change by a factor of -1 so again the quotient doesn't change. Since A is invertible it is row equivalent to the identity, so it now remains to check that Cramer's rule holds when A is the identity. But that's pretty easy to do, so that does it!

How can we use Cramer's rule to verify Theorem 3? We want to find the inverse of A. Let's be modest and just try to find the jth column c_j of A^{-1}. The vector c_j is the solution to

$$Ac_j = e_j$$

where e_j is the jth standard basis vector for $\mathbf{R}^n$ (can you see why?). Now let's be even more modest and just try to find the ith entry of c_j, in other words the ijth entry of A^{-1}. By Cramer's rule this is

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

where A_i is the matrix obtained from A by replacing the ith column by e_j. You can check that $\det A_i$ is exactly C_{ji}. For instance if

$$A = \begin{pmatrix} 7 & 2 & 1 \\ 0 & 3 & -1 \\ -3 & 4 & -2 \end{pmatrix}$$

and $j = 3$ and $i = 2$ then

$$A_i = \begin{pmatrix} 7 & 0 & 1 \\ 0 & 0 & -1 \\ -3 & 1 & -2 \end{pmatrix}.$$

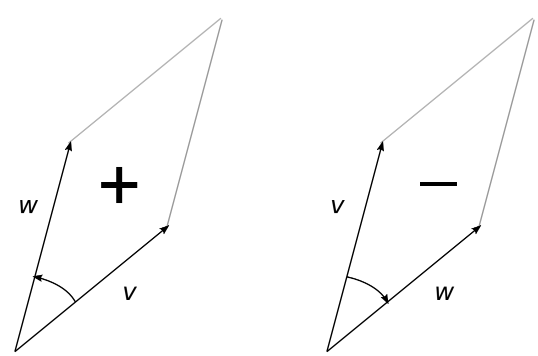
Expanding down the third column (the one we replaced), we see that the determinant of this matrix is $-1(7 \cdot -1 - 1 \cdot 0)$. Not only is this the cofactor C_{32} of A, but the calculation you do to find this determinant is exactly the same as the calculation you do to find the cofactor!

So in summary the ijth entry of A^{-1} is exactly $C_{ji}/(\det A)$, just like Theorem 3 says.

4. Determinants, volumes, and linear transformations

Our first task is to make good on our promise at the beginning that determinants should have something to do with volumes. You can readily think of examples of matrices that have negative determinants, and volumes usually aren't negative, so we first need to wrestle with that.

You've seen the idea of negative area and volume in Calculus. For instance, when you integrate $\sin(x)$ from 0 to π, you get 0, because there are two regions enclosed by $\sin(x)$ and the x–axis, one of positive area (above the x–axis) and one of negative area (below the x–axis). So we say that the integral computes *signed* area. Let's import this idea into linear algebra. Say that we have an *ordered* list of two vectors v, w. We can use v and w to make a parallelogram in the usual way. At the origin, the vectors v and w make two angles, one greater than or equal to π and one less than or equal to π. The parallelogram that lives in the part that is less than or equal to π. We will say that the parallelogram has positive area if within the parallelogram w is counterclockwise from v near the origin and it has negative area otherwise:



Here is another way to say it that is maybe more important: the area is positive if v and w are configured just like the x– and y–axes in $\mathbf{R}^2$ (in that order) and the area is negative otherwise. Look back at the picture to make sense of this. It turns out that you can generalize this idea to $\mathbf{R}^n$: if you have a list of n length n vectors $v_1, \dots, v_n$, then those vectors span what is called a parallelepiped in $\mathbf{R}^n$, and in a similar way we can say that the volume of this parallelepiped is positive or negative depending on whether $v_1, \dots, v_n$ is configured like $e_1, \dots, e_n$ or not. For $n = 3$, this is the same as the right-hand rule: the area of the parallelepiped spanned by v_1, v_2, v_3 is positive if when you stick your pointer finger on your right hand along v_1 and your middle finger along v_2, your thumb sticks along v_3 (check that this works for e_1, e_2, e_3).

Fact 8. If A is an $n \times n$ matrix has rows $v_1, \ldots, v_n$ (in order) then the signed volume of the parallelepiped spanned by $v_1, \ldots, v_n$ is $\det A$*.*

Why is this true? Well, all you have to do is convince yourself that signed volume satisfies the four properties of a determinant function. Once you've done that then you can apply Theorem 1, which says that there is only one determinant function. And so the signed volume function and the determinant function must be one and the same! Property 1 works since row replacement changes neither the base nor the height of the parallelepiped (thinking of the replaced row as giving the height). Property 2 is the trickiest, but our picture above clearly shows when $n = 2$ that swapping rows negates the signed volume. Property 3 is true because if you scale a row then, thinking of that row as being the one giving the height of the parallelepiped, the base doesn't change, but the height gets scaled by the same amount you scaled the row. And Property 4 is the easiest: the parallelepiped spanned by the standard basis vectors for $\mathbf{R}^n$ is the standard unit cube and hence has signed volume one (the volume is positive because the standard basis vectors are obviously configured like the standard basis vectors).

That's that. Now suppose that $T_A : \mathbf{R}^n \to \mathbf{R}^n$ is the linear transformation associated to an invertible $n \times n$ matrix A. We think of T_A as some operation that you *do* to $\mathbf{R}^n$. We've seen examples of this before, for instance when T_A is a reflection, a projection, a rotation, a stretch or a shear. Of course, in general, T_A can be some complicated combination of these.

So what can we say about what T_A does to $\mathbf{R}^n$? Well, we know that T_A takes the standard basis vectors $e_1, \ldots, e_n$ to the columns of A. It follows that the standard cube spanned by $e_1, \ldots, e_n$ gets taken to the parallelepiped spanned by the columns of A. What is the volume of this parallelepiped? Well, since $\det A = \det A^T$ (can you see why?), the volume of the parallelepiped is $\det A$ by Fact 8. In summary, T_A changes the volume of the standard cube by $\det A$. Now we know something about what T_A does.

Actually, we know more. If we take the cube $\varepsilon e_1, \ldots, \varepsilon e_n$ where ε is a small real number (wait, are we doing Calculus now?), then since T_A is linear, the image of this cube—which has signed volume ε^n—will have signed volume $\det A \varepsilon^n$. What does that buy us? Well now take any shape S in $\mathbf{R}^n$ that has finite volume. Fill it as best you can with ε-cubes. The total volume of the ε-cubes is approximately volume(S). Now apply T_A. What is the volume of $T_A(S)$? We just said that each of the little cubes gets its volume stretched by $\det A$. Since that's true, all of S gets its volume stretched by $\det A$. If we don't want to think about signed volume, and just want to worry about volume, then we can take the absolute value of $\det A$ and conclude the following fact.

Fact 9. If T_A is the linear transformation of $\mathbf{R}^n$ associated to an $n \times n$ matrix A and S is a region in $\mathbf{R}^n$, then

$$\mathit{volume}(T_A(S)) = |\det A| \mathit{volume}(S).$$

5. Linearity and the proof of existence and uniqueness of determinants

At long last we need to prove Theorem 1, which says that determinant functions exist and are unique. We also want to prove our cofactor formula for computing the determinant. We'll do both of these at the same time!

Here's where we pay the piper a little bit. If the rows of a matrix are $v_1, \ldots, v_n$, let's write the matrix as $(v_1, \ldots, v_n)$. We need to show that

Fact 7. Any determinant function det is linear in each row, that is:

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

For the sake of argument, let's assume that A is invertible so that the v_i span $\mathbf{R}^n$. We want to show that $\det(v_1 + w, \ldots, v_n)$ is equal to $\det(v_1, \ldots, v_n) + \det(w, v_2, \ldots, v_n)$. Why is this true? Since the v_i span (in fact form a basis for) $\mathbf{R}^n$, we can write w uniquely as $cv_1 + w'$ where w' is in the span of $v_2, \ldots, v_n$. So

$$\det(v_1 + w, \ldots, v_n) = \det(v_1 + cv_1 + w', v_2, \ldots, v_n)$$

But by doing row operations we can get rid of the w'!

$$\det(v_1 + cv_1 + w', v_2, \ldots, v_n) = \det(v_1 + cv_1, v_2, \ldots, v_n) = \det((c+1)v_1, v_2, \ldots, v_n)$$

But by the scaling rule this is

$$(c+1)\det(v_1, \ldots, v_n) = \det(v_1, \ldots, v_n) + c\det(v_1, \ldots, v_n)$$

But since w' is a linear combination of $v_2, \ldots, v_n$ we can use the same argument as before to show that $c\det(v_1, \ldots, v_n)$ is $\det(w, v_2, \ldots, v_n)$ (again use row operations to kill the w'). This exactly gives what we want!

Once we have this we can derive the cofactor formula for the determinant. Say $v_1 = c_1e_1 + \cdots c_ne_n$. (Note that the c_j are just a_{1j}.) Then by linearity we have that

$$det(v_1, \ldots, v_n) = \sum c_j \det(e_j, v_2, \ldots, v_n)$$

So it remains to find a formula for the latter. The claim is that the latter is equal to the cofactor C_{ij}, that is, $(-1)^{1+j}$ times the determinant of the ijth minor A_{ij}. This is not too hard using row swaps and row replacement.

We could have done this for any row. And so we have n formulas:

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

It is easy to check that these are all the same by using the row swap rule.

If you want a formula for columns, you need to show that $\det(A^T) = \det(A)$. But this follows easily from the above facts and the fact that it is true for invertible matrices. So in this entire discussion you can use columns instead of rows if you want. You therefore get n formulas for the determinant using columns.

All $2n$ numbers you get are the same by our reals, and clearly the function det exists because we just gave a recursive formula for it.