Math 108a	Professor: Padraic Bartlett

Lecture 19: The Determinant

Week 10	*UCSB 2013*

In our last class, we talked about how to calculate **volume** in n-dimensions. Specifically, we defined a parallelotope:

Definition. Given n vectors $\vec{w_1}, \ldots \vec{w_n} \in \mathbb{R}^n$, the **parallelotope** spanned by these vectors is the collection

$$\{a_1\vec{w_1} + \ldots + a_n\vec{w_n} \mid 0 \leq a_i \leq 1, \forall i\}.$$

In the case where $n = 2$, this is a parallelogram!

From here, we discussed how to find the **volume** of a parallelotope. Specifically, given a parallelotope spanned by the vectors $\{\vec{w_1}, \ldots \vec{w_n}\}$, we constructed the following vectors:

- $\vec{u_1} = \vec{w_1}$.
- $\vec{u_2} = \vec{w_2} - \text{proj}(\vec{w_2} \text{ onto } \vec{u_1})$.
- $\vec{u_3} = \vec{w_3} - \text{proj}(\vec{w_3} \text{ onto } \vec{u_1}) - \text{proj}(\vec{w_3} \text{ onto } \vec{u_2})$.
- $\vec{u_4} = \vec{w_4} - \text{proj}(\vec{w_4} \text{ onto } \vec{u_1}) - \text{proj}(\vec{w_4} \text{ onto } \vec{u_2}) - \text{proj}(\vec{w_4} \text{ onto } \vec{u_3})$.

$\vdots$

- $\vec{u_n} = \vec{w_n} - \text{proj}(\vec{w_n} \text{ onto } \vec{u_1}) - \ldots - \text{proj}(\vec{w_n} \text{ onto } \vec{u_{n-1}})$.

We thought of each of these vectors as representing the "height" of each $\vec{w_i}$ over the previous $\vec{w_1}, \ldots \vec{w_{i-1}}$. With this idea in mind, we defined the **volume** of our parallelepiped as

$$\prod_{i=1}^{n} ||\vec{u_i}||,$$

i.e. the product of the lengths of the vectors $\vec{u_i}$.

In this class, we discuss why we care about volume in this class: because it lets us study the **determinant**!

1 The Positive Determinant

Consider the following pair of definitions:

Definition. For any n, we define the n**-dimensional unit cube** as the set of all points

$$\{(a_1, \ldots a_n) \mid 0 \leq a_i \leq 1, \forall i\}$$

You can think of this as the paralleletope spanned by the basis vectors $\{\vec{e_1}, \ldots \vec{e_n}\}$.

Definition. Take an $n \times n$ matrix A. Look at where A sends the unit cube. We know that by definition, A sends the first basis vector $\vec{e_1}$ to its first column $\vec{a_{c_1}}$, the second basis vector $\vec{e_2}$ to $\vec{a_{c_2}}$, and in general the i-th basis vector $\vec{e_i}$ to $\vec{a_{c_i}}$.

Therefore, this matrix sends the unit cube to the parallelotope spanned by the columns of A!

We define the **positive determinant of** A, written $\det^+(A)$, as the volume of the parallelotope spanned by the columns of A.

To illustrate, we quickly calculate an example here:

Example. Calculate the positive determinant of the matrix

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Answer. We are essentially looking for the volume of the parallelotope spanned by the vectors $(1,1,0),(1,0,1),(0,1,1)$. We do this by following the outline we discussed last week, for turning these three vectors into "heights." In particular, if we label these vectors as $\vec{w_1}, \vec{w_2}, \vec{w_3}$, we can just find the $\vec{u_1}, \vec{u_2}, \vec{u_3}$ vectors as described on the first page of these notes, and calculate the lengths of those vectors.

First, we note that the length of the first vector $(1,1,0)$ is just $\sqrt{2}$.

Then, we note that the "height" of the vector $(1,0,1)$ over the vector $(1,1,0)$ is just the length of

$$\begin{aligned} \vec{u_2} &= (1,0,1) - \text{proj}((1,0,1) \text{ onto } (1,1,0)) \\ &= (1,0,1) - \frac{(1,0,1)\cdot(1,1,0)}{(1,1,0)\cdot(1,1,0)}(1,1,0) \\ &= (1,1,1) - \frac{1}{2}(1,1,0) \\ &= \left(\frac{1}{2}, -\frac{1}{2}, 1\right). \end{aligned}$$

which is just $\sqrt{3/2}$.

Finally, we need the height of the vector $(0,1,1)$ over the base spanned by $(1,1,0)$, $\left(\frac{1}{2}, -\frac{1}{2}, 1\right)$. We do this using the description we came up with before:

$$\begin{aligned} \vec{u_3} &= (0,1,1) - \text{proj}((0,1,1) \text{ onto } (1,1,0)) - \text{proj}\left((0,1,1) \text{ onto } \left(\frac{1}{2}, -\frac{1}{2}, 1\right)\right) \\ &= (0,1,1) - \frac{(0,1,1)\cdot(1,1,0)}{(1,1,0)\cdot(1,1,0)}(1,1,0) - \frac{(0,1,1)\cdot\left(\frac{1}{2},-\frac{1}{2},1\right)}{\left(\frac{1}{2},-\frac{1}{2},1\right)\cdot\left(\frac{1}{2},-\frac{1}{2},1\right)}\left(\frac{1}{2}, -\frac{1}{2}, 1\right) \\ &= (0,1,1) - \frac{1}{2}(1,1,0) - \frac{1}{3}\left(\frac{1}{2}, -\frac{1}{2}, 1\right) \\ &= \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right) \end{aligned}$$

which is just $2/\sqrt{3}$.

If we take the product of these three heights, we get $\sqrt{2} \cdot \sqrt{3/2} \cdot (2/\sqrt{3}) = 2$. So the volume of this parallelepiped is 2, and therefore we have

$$\det{}^+(A) = 2.$$

.

2 The Positive Determinant and Elementary Matrices

The positive determinant has a number of remarkably nice interactions with elementary matrices! We study these interactions in this section.

Theorem. Take any matrix A. Look at the matrix $A \cdot E$, where E is an elementary matrix of the form

$$E_{\text{multiply entry } k \text{ by } \lambda}.$$

Then

$$\det{}^+(A \cdot E) = |\lambda| \cdot \det{}^+(A).$$

Proof. Make the following observations:

- The positive determinant of A is just the volume of the parallelotope spanned by the column vectors of A.
- The matrix $A \cdot E$ is just the matrix A with its k-th column multiplied by λ.
- Therefore, the positive determinant of $A \cdot E$ is the volumne of the parallelotope spanned by the column vectors of A, where one of them is multiplied by λ.
- If we pick the λ-multiple as our first vector when calculating the volume, it is clear that the length of the base is scaled by $|\lambda|$, and the length of any height vector is unchanged (as those are calculated by looking at things orthogonal to the base, and therefore do not care about the length of the base!)
- Therefore, the volume of $A \cdot E$ is just the volume of A scaled by $|\lambda|$.

Done! $\square$

Theorem. Take any matrix A. Look at the matrix $A \cdot E$, where E is an elementary matrix of the form

$$E_{\text{switch entry } k \text{ and entry } l}.$$

Then

$$\det{}^+(A \cdot E) = \det{}^+(A).$$

Proof. This is like the above proof, but even easier. First, notice that the matrix $A \cdot E$ is just the matrix A, but with two columns swapped. Therefore, the volume of the parallelepiped spanned by the columns of $A \cdot E$ is the same as the volume of the parallelepiped spanned by the columns of A, because they're both the volumes of the **same** paralellepiped! $\square$

Theorem. Take any matrix A. Look at the matrix $A \cdot E$, where E is an elementary matrix of the form

$$E_{\text{add } \lambda \text{ copies of entry } i \text{ to entry } j}.$$

Then

$$\det{}^{+}(A \cdot E) = \det{}^{+}(A).$$

Proof. This is also like the above proof, but slightly harder. Notice that the matrix $A \cdot E$ is just the matrix A with λ copies of its i-th column added to its j-th column.

We calculate the volume of the parallelepiped spanned by $A \cdot E$ simultaneously with the volume of the parallelepiped spanned by A's columns in the following way. For both parallelotopes, let the vector corresponding to the i-th column in both be the "first" vector we study (i.e. the $\vec{w_1}$ vector), and the vector corresponding to the j-th column in both be the "second" vector we study (i.e. the $\vec{w_2}$ vector). The order of the rest won't matter, so pick any order for the rest.

To find the volume, then, we just do the following:

1. In both cases, we start by finding the length of the i-th column, which is the same in both cases – it's the length of the i-th column of both matrices, which is the same in both cases.

2. Now, in both cases, we move to the "height" of the j-th column over this i-th column.

   Notice that because the j-th column vector of $A \cdot E$ is just the j-th column of A, plus λ copies of the k-th column, we have that

   $$\begin{aligned} &\text{orth}\Big((\text{the } j\text{-th column of } A \cdot E) \text{ onto } (\text{the } k\text{-th column of } A)\Big) \\ =&\text{orth}\Big((\text{the } j\text{-th column of } A) \text{ onto } (\text{the } k\text{-th column of } A)\Big). \end{aligned}$$

   This is because adding copies of the k-th column to a vector doesn't change the "amount" of that vector that is orthogonal to that k-th column! (Basically, imagine adding λ copies of a vector $\vec{w}$ to another vector $\vec{v}$. This directly increases the quantity proj($\vec{v}$ onto $\vec{w}$) by $\lambda\vec{v}$; therefore, when we form the vector orth($\vec{v}$ onto $\vec{w}$) $= \vec{v} -$ proj($\vec{v}$ onto $\vec{w}$), we subtract those copies off again!)

   Therefore, the height of the k-th column over our j-th column is the same in both cases!

3. Now, notice that the spans of the k, j-th columns in the matrices $A, A \cdot E$ are the same in both cases, as they both consist of all multiples of the k-th and j-th columns! Therefore, the "height" of any other vector over these two is unchanged, as well.

Consequently, because the lengths of the base and of the heights are unchanged at each step, these two paralleletopes have the same volume. $\square$

3 Elementary Matrices and The Positive Determinant: Why We Care

Elementary matrices are pretty cool in general; however, the main reason we care about them here is because they let us understand the determinant! In particular, they give us new results about the positive determinant:

Theorem. Take any two $n \times n$ matrices A, B. Then

$$\det{}^+(A \cdot B) = \det{}^+(A) \cdot \det{}^+(B).$$

Proof. First, write A,B as products of elementary matrices:

$$A = E_1 \cdot \ldots \cdot E_n, B = E_{n+1} \cdot \ldots \cdot E_{n+m}.$$

We know we can do this from our work in week 8, where we showed that any matrix can be written as a product of elementary matrices.

Let $\lambda_1, \ldots \lambda_k$ denote the coefficients corresponding to all of the "multiply an entry by λ" elementary matrices above in A, and $\lambda_{k+1}, \ldots \lambda_{k+l}$ denote those coefficients in B. Then, we have

$$\begin{aligned} \det{}^+(A) &= |\lambda_1 \cdot \ldots \cdot \lambda_k|, \\ \det{}^+(B) &= |\lambda_{k+1} \cdot \ldots \cdot \lambda_{k+l}|, \text{ and} \\ \det{}^+(A \cdot B) &= |\lambda_1 \cdot \ldots \cdot \lambda k + l| = \det{}^+(A) \det{}^+(B). \end{aligned}$$

This is because we've shown in our work with elementary matrices that

$$\det{}^+(E_1 \cdot \ldots \cdot E_n) = \det{}^+(E_1 \cdot \ldots \cdot E_{n-1}) \cdot \alpha,$$

where α is equal to 1 if E_n was a "swap" or" add some multiple of an entry to another entry" matrix, and λ if it was a a "multiply an entry by λ" matrix. Repeatedly using this observation on $A = E_1 \cdot \ldots \cdot E_n, B = E_{n+1} \cdot \ldots \cdot E_{n+m}, A \cdot B = E_1 \cdot \ldots \cdot E_{n+m}$ gives us the three results above.

Therefore, the positive determinant of the product of two matrices is the product of the positive determinants of these two matrices. $\square$

A useful result we used in the above proof is that it gives us a second way to calculate the positive determinant of various matrices — specifically, decomposing it into elementary matrices! We explicitly state this here:

Lemma 1. *Take any $n \times n$ matrix A. Write A as the product of elementary matrices: i.e.*

$$A = E_1 \cdot \ldots \cdot E_n.$$

Then

$$\det{}^+(A) = |\lambda_1 \cdot \ldots \cdot \lambda_k|.$$

We calculate an example determinant with this method here:

Example. Consider the matrix

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 4 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

Write A as a product of elementary matrices. Use this information to calculate the positive determinant of A.

Proof. We actually studied this matrix in week 8 in the notes online; there, we showed that

$$\overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}^{\substack{\text{switch rows} \\ r_3 \text{ and } r_2}} \cdot \overbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}^{\substack{\text{switch rows} \\ r_2 \text{ and } r_1}} \cdot \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}}^{\substack{\text{add 2 copies of} \\ r_1 \text{ to } r_3}} \cdot \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}}^{\substack{\text{add } -1 \text{ copies of} \\ r_2 \text{ to } r_3}}$$

$$\cdot \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}^{\substack{\text{multiply row } r_3 \\ \text{by } 0}} \overbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}^{\substack{\text{add 1 copies of} \\ r_3 \text{ to } r_1}} \cdot \overbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}^{\substack{\text{multiply row } r_1 \\ \text{by } 2}} \cdot \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}}^{\substack{\text{add 2 copies of} \\ r_3 \text{ to } r_2}} = A.$$

If we use our lemma above, then we have that the determinant of this matrix is just the product of the lambda values from the "multiply entry k by λ" matrices: in particular, it's $2 \cdot 0 = 0$.

This makes sense: if we look at the three columns of A, we can see that -1 copies of the first column and four copies of the second column combine together to make two copies of the third column. Therefore, these three vectors are linearly dependent! In particular, this means that the parallelepiped spanned by them is "flat": the "height" of the third column vector over the first two is 0, and thus has zero volume. ☐

4 The General Determinant

The other reason we care about the interaction of elementary matrices and the "positive" determinant is that it gives us a way to define the **actual** determinant!

Definition. The **determinant** (as opposed to the "positive determinant") of a matrix A is defined as follows:

1. Take A, and write it as the product $E_1 \cdot \ldots \cdot E_n$ of elementary matrices.
2. To find the determinant $\det(A)$ of A, look at these elementary matrices. Let $\lambda_1, \ldots \lambda_k$ denote the constants that show up in the "multiply an entry by λ_i" elementary matrices, and l denote the number of "swap" elementary matrices. Then

$$\det(A) = (-1)^l \cdot \lambda_1 \cdot \ldots \cdot \lambda_k$$

This object, in other words, is just the positive determinant from before, i.e. the volume, except multiplied by a factor of ± 1 depending on the signs of the constants λ_i and the number of swaps performed. This gives us the following observation for free:

Observation. For any matrix A, $|\det(A)| = \det^+(A)$.

By literally repeating the proof methods used earlier in this talk, we can prove the following theorem:

Theorem. Take any two $n \times n$ matrices A, B. Then

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

5 The General Determinant: Why We Care

The main reason we care about this "new" determinant, where it can be either positive or negative, is because it has an exciting new property that the positive determinant did not: n**-linearity!**

Definition. Let T be a map from $n \times n$ matrices of real numbers to $\mathbb{R}$. We say that T is **n-linear** if the following always holds:

- Take any matrix A, with columns $\vec{a_{c_1}}, \ldots \vec{a_{c_n}}$.
- Suppose that $\vec{a_{c_i}}$ is equal to some sum of vectors $\vec{x} + \vec{y}$.
- Then, consider the two matrices created by replacing this i-th column with the vectors $\vec{x}, \vec{y}$ respectively:

$$A_x = \begin{bmatrix} \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vec{a_{c_1}} & \ldots & \vec{a_{c_{i-1}}} & \vec{x} & \vec{a_{c_{i+1}}} & \ldots & \vec{a_{c_n}} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \end{bmatrix},$$
$$A_y = \begin{bmatrix} \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vec{a_{c_1}} & \ldots & \vec{a_{c_{i-1}}} & \vec{y} & \vec{a_{c_{i+1}}} & \ldots & \vec{a_{c_n}} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \end{bmatrix},$$

A map is called **n-linear** if

$$T(A) = T(A_x) + T(A_y),$$

for any column $\vec{a_{c_i}}$ and pair of vectors $\vec{x}, \vec{y}$ such that $\vec{x} + \vec{y} = \vec{a_{c_i}}$.

Theorem. The determinant is **n-linear**.

We first note a quick example that shows why the positive determinant is not n-linear: simply observe that the positive determinant of

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

is 0, as the parallelogram spanned by $(0,0),(0,1)$ has zero area. However, the parallelograms spanned by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

both have area 1! Therefore, because $0 \neq 1+1$, the positive determinant is not n-linear.

The proof that the determinant is n-linear is kind of awful. We postpone it to the end of this talk, in favor of showing you **why** you care about it: because it gives you the formula for the determinant you've used in the past!

Theorem. Let A be a $n \times n$ matrix. Given a row i and a column j, let A_{ij} denote the $n-1 \times n-1$ matrix formed by deleting the i-th row and j-th column of A.

Consider the following object:

$$a_{11} \cdot \det(A_{11}) - a_{21} \cdot \det(A_{21}) + a_{31} \cdot \det(A_{31}) \ldots + (-1)^{n-1} a_{n1} \cdot \det(A_{n1}).$$

For example, for a 3×3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, this is just

$$a_{11} \cdot \det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right) - a_{21} \cdot \det\left(\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix}\right) + a_{31} \cdot \det\left(\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}\right).$$

This object is the determinant. I.e. the determinant satisfies the following property:

$$\det(A) = a_{11} \cdot \det(A_{11}) - a_{21} \cdot \det(A_{21}) + a_{31} \cdot \det(A_{31}) \ldots + (-1)^{n-1} a_{n1} \cdot \det(A_{n1}).$$

Proof. This is a just a consequence of n-linearity. Take the matrix A, and write its first column $\vec{a_{c_1}} = (a_{11}, a_{21}, \ldots a_{n1})$ as the sum $a_{11}\vec{e_1} + \ldots + a_{n1}\vec{e_n}$.

Let A_i denote the matrix whose first column is replaced by the vector $a_{i1}\vec{e_i}$. Then, we can use n-linearity to write

$$\det(A) = \det(A_1) + \ldots + \det(A_n).$$

Now, notice that for each of these matrices A_i, we either have

1. $a_{i1} = 0$. In this case, the first column of A_i is all-zeroes, and therefore the volume of the parallelepiped spanned by the columns of A is 0.

2. $a_{i1} \neq 0$. In this case, by repeatedly multiplying A_i on the right by matrices of the form

$$E_{\text{add } -\frac{a_{ij}}{a_{i1}} \text{ copies of entry 1 to entry } j},$$

we can get a matrix whose i-th row is 0 apart from the entry a_{i1} in the first column, without changing the determinant.

Why do we care? Well: notice that the determinant of this matrix is just the determinant of the matrix A_{i1}, formed by deleting the i-th row and first column from A, scaled by a_{i1}! (Persuade yourself that this is true if you don't see why.)

In either case, we have that the determinant of these A_i matrices is the same thing as $a_{i1} \det(A_{i1})$, and therefore that our claim holds. $\square$

This technique is what many of you have used before, and perhaps seen, as a definition of the determinant! It's great for proofs, but it bears noting that computationally it's not the fastest thing out there. In particular, it will need something like $n!$ computations to find the determinant, as to calculate a determinant of a $n \times n$ matrix you need to

- calculate n determinants of $n-1 \times n-1$-matrices, or
- $n(n-1)$ determinants of $n-2 \times n-2$-matrices, or
- $n(n-1)(n-2)$ determinants of $n-3 \times n-3$-matrices, or

$\vdots$

- $n!$ determinants of 1×1 matrices.

This gets expensive **fast**. For example, suppose that we have access to the world's fastest supercomputer as of this summer, the Tianhe-2, which has a max speed of 33.86 petaflops — i.e. it can perform about $3.386 \cdot 10^{16}$ arithmetic steps per second. Suppose we wanted to use it to calculate the determinant using the above process, and we could do everything but the final addition for free — i.e. assume that to take the determinant of a $n \times n$ matrix, we just have to perform $n!$ addition steps!

So: suppose we wanted to calculate the determinant of a 25×25 matrix. How long would you think this takes?

If you said 14.5 years, you're correct! Lesson: don't use this "determinants of smaller matrices" method.

Conversely: to do the elementary matrix method, we just have to repeatedly solve n systems of linear equations, one system for each row of A. This isn't too hard to do: it takes us like n^2 steps to solve a set of n systems of linear equations, and we do n of these operations, which results in about n^3 steps, up to some constants. Our supercomputer can solve this in $1.4 \cdot 10^{-20}$ of a second. So, um, faster.

6 Appendix: Proving The Determinant Is n-Linear

We prove that the determinant is n-linear here.

Proof. To see that the determinant is n-linear: take any matrix A, any column $\vec{a_{c_i}}$, and any pair of vectors $\vec{x}, \vec{y}$ such that $\vec{x} + \vec{y} = \vec{a_{c_i}}$.

Write both of the vectors $\vec{x}, \vec{y}$ as linear combinations

$$\vec{x} = \text{proj}(\vec{x} \text{ onto columns of } A) + \text{orth}(\vec{x} \text{ onto columns of } A),$$
$$\vec{y} = \text{proj}(\vec{y} \text{ onto columns of } A) + \text{orth}(\vec{y} \text{ onto columns of } A).$$

Notice that because $\mathrm{x} + \mathrm{y} = \vec{a_{c_i}}$, we have

$$\text{orth}(\vec{x} + \vec{y} \text{ onto columns of } A) = \text{orth}(\vec{a_{c_i}} \text{ onto columns of } A) = \vec{0},$$

because $\vec{a_{c_i}}$ is itself a column of A. Consequently, we have

$$\text{orth}(\vec{x} \text{ onto columns of } A) = -\text{orth}(\vec{y} \text{ onto columns of } A).$$

Now, notice that for the matrix A_x, we have

$$\det(A_x) = \det(A_x \cdot E_{\text{add } \lambda \text{ copies of entry } k \text{ to entry } l},$$

for **any** λ, k, l such that $k \neq l$! In particular, if we write

$$\text{proj}(\vec{x} \text{ onto columns of } A) = x_1 \vec{a_{c_1}} + \ldots + x_n \vec{a_{c_n}},$$
$$\text{proj}(\vec{y} \text{ onto columns of } A) = y_1 \vec{a_{c_1}} + \ldots + y_n \vec{a_{c_n}},$$

we can use these matrices to see that

$$\det(A_x) = \det\left(A_x \cdot \overbrace{E_{\text{add } -x_1 \text{ copies of entry 1 to entry } i} \cdot \ldots \cdot E_{\text{add } -x_n \text{ copies of entry } n \text{ to entry } i}}^{\text{these matrices add } -x_j \text{ copies of each column } j \neq i \text{ to column } i} \right),$$

and similarly that

$$\det(A_y) = \det\left(A_y \cdot \overbrace{E_{\text{add } -y_1 \text{ copies of entry 1 to entry } i} \cdot \ldots \cdot E_{\text{add } -y_n \text{ copies of entry } n \text{ to entry } i}}^{\text{these matrices add } -y_j \text{ copies of each column } j \neq i \text{ to column } i} \right).$$

But

$$A_x \cdot \overbrace{E_{\text{add } -x_1 \text{ copies of entry 1 to entry } i} \cdot \ldots \cdot E_{\text{add } -x_n \text{ copies of entry } n \text{ to entry } i}}^{\text{these matrices add } -x_j \text{ copies of each column } j \neq i \text{ to column } i}$$
$$= \begin{bmatrix} \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vec{a_{c_1}} & \ldots & \vec{a_{c_{i-1}}} & \vec{x} - (\overbrace{x_1 \vec{a_{c_1}} + \ldots + x_n \vec{a_{c_n}}}^{\text{all entries except} x_i \vec{a_{c_i}}}) & \vec{a_{c_{i+1}}} & \ldots & \vec{a_{c_n}} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \end{bmatrix}$$
$$= \begin{bmatrix} \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vec{a_{c_1}} & \ldots & \vec{a_{c_{i-1}}} & \text{orth}(\vec{x} \text{ onto columns of } A) + x_i \vec{a_{c_i}} & \vec{a_{c_{i+1}}} & \ldots & \vec{a_{c_n}} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \end{bmatrix},$$

and similarly

$$
\begin{aligned}
&A_y \cdot \overbrace{E_{\text{add } -y_1 \text{ copies of entry 1 to entry } i} \cdot \ldots \cdot E_{\text{add } -y_n \text{ copies of entry } n \text{ to entry } i}}^{\text{these matrices add } -y_j \text{ copies of each column } j \neq i \text{ to column } i} \\
&= \begin{bmatrix} \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vec{a_{c_1}} & \ldots & \vec{a_{c_{i-1}}} & \text{orth}(\vec{y} \text{ onto columns of } A) + y_i \vec{a_{c_i}} & \vec{a_{c_{i+1}}} & \ldots & \vec{a_{c_n}} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \end{bmatrix}.
\end{aligned}
$$

Call these two matrices A'_x, A'_y.

There are now two possibilities.

1. The columns of A form a basis for $\mathbb{R}^n$. In this case, we have that the two orth components above are both zero, because there is nothing in $\mathbb{R}^n$ orthogonal to all of $\mathbb{R}^n$. Therefore, we have that

$$
\begin{aligned}
\det(A_x) + \det(A_y) =& \det(A \cdot E_{\text{multiply entry i by } x_i}) + \det(A \cdot E_{\text{multiply entry i by } y_i}) \\
=& x_i \det(A) + y_i \det(A) \\
=& (x_i + y_i) \det(A).
\end{aligned}
$$

   So: what is $x_i + y_i$? On one hand, we know that

$$\vec{x} + \vec{y} = \vec{a_{c_i}} = x_1 \vec{a_{c_1}} + \ldots + x_n \vec{a_{c_n}} + y_1 \vec{a_{c_1}} + \ldots + y_n \vec{a_{c_n}},$$

   and therefore that

$$\vec{0} = (x_1 + y_1)\vec{a_{c_1}} + \ldots + (x_i + y_i - 1)\vec{a_{c_i}} + \ldots + (x_n + y_n)\vec{a_{c_n}}.$$

   This is a nontrivial linear combination of elements in a basis that equals 0; therefore, all of the coefficients above must be 0! As a result, we must have $x_i + y_i = 1$. This gives us

$$\det(A_x) + \det(A_y) = (x_i + y_i) \det(A) = \det(A),$$

   as requested.

2. Otherwise, the columns of A do not form a basis for $\mathbb{R}^n$. In this case, the columns of A are linearly dependent! Take a combination

$$b_1 \vec{a_{c_1}} + \ldots + b_n \vec{a_{c_n}} = \vec{0}$$

   where not all of the b_k's are zero.

   If in this combination the coefficent b_i is zero, then there is a combination of the columns of A, not using the i-th column, that combines to zero! This means that for both of the matrices A_x, A_y, the columns of these matrices are also a linearly dependent set, because this combination does not use the i-th column. Therefore, we have that the determinants of these two matrices are zero, much like the determinant of A itself, because all three are matrices with linearly dependent columns (and therefore correspond to paralleletopes that live in a $n-1$ dimensional space, and thus have zero volume.)

Otherwise, in this combination the b_i coefficient is nonzero. This gives us a way to express the i-th column of A as a linear combination of the other columns of A!

Therefore, by using the $E_{\text{add } \lambda \text{ copies of entry } k \text{ to entry } l}$ in a similar way to before, we can subtract multiples of all of the other columns of A_x from the i-th column of A'_x, such that we get rid of the $x_i \vec{a_{ci}}$ part, without changing the determinant! We can also do the same trick to the A'_y matrix; this gives us that $\det(A_x)$ is the determinant of the matrix that you get by replacing the i-th column of A with $\text{orth}(\vec{x} \text{ onto columns of } A)$, and similarly for $\det(A_y)$!

But, because $\text{orth}(\vec{x} \text{ onto columns of } A) = -\text{orth}(\vec{y} \text{ onto columns of } A)$, we have that the determinants of these two matrices are the same, except one is the opposite sign of the other! Therefore, we have that $\det(A_x) = -\det(A_y)$, and thus

$$\det(A_x) + \det(A_y) = 0 = \det(A),$$

again because the columns of A are linearly dependent.

$\square$