Math 5327
Spring 2018

Determinants: Uniqueness and more

Uniqueness

The main theorem we are after:

Theorem 1 *The* ***determinant*** *of and $n \times n$ matrix A is the unique n-linear, alternating function from $F^{n\times n}$ to F that takes the identity to 1.*

This will follow if we can prove:

Theorem 2 *If $D : F^{n\times n} \to F$ is n-linear and alternating, then for all $n \times n$ matrices A, $D(A) = \det(A)D(I)$.*

So far, we have proved that determinants exist for all $n \times n$ matrices, but not the uniqueness part. The trick was to show that cofactor expansions define n-linear, alternating functions. Now we want to prove uniqueness by following something like what we did for $n = 2$:

Lemma 1 *The determinant,* $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$ *is the unique 2-linear, alternating function satisfying* $\det(I) = 1$. *Moreover, if D is any 2-linear, alternating function on 2×2 matrices, then $D(A) = \det(A)D(I)$ for all 2×2 matrices A.*

Proof: We do both parts at the same time. I will let you check that the usual determinant is, in fact, 2-linear and alternating. Suppose D is 2-linear and alternating. We use linearity in the first row, and then the second, using, for example, $(a, b) = a(1, 0) + b(0, 1)$. We have

$$\begin{aligned} D\begin{pmatrix} a & b \\ c & d \end{pmatrix} &= aD\begin{pmatrix} 1 & 0 \\ c & d \end{pmatrix} + bD\begin{pmatrix} 0 & 1 \\ c & d \end{pmatrix} \\ &= acD\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + adD\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + bcD\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + bdD\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

By alternating property (a), the first and last of the four terms are 0. By property (b), we can write

$$\begin{aligned} D\begin{pmatrix} a & b \\ c & d \end{pmatrix} &= adD\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + bcD\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= adD\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - bcD\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= (ad - bc)D\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = (ad - bc)D(I), \end{aligned}$$

as desired. ∎

Let's sketch out the proof in the case of 3×3 matrices. It goes like this: We ask that D be an n-linear, alternating function on 3×3 matrices, and see if this forces a formula out. The start would be

$$D\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aD\begin{pmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} + bD\begin{pmatrix} 0 & 1 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} + cD\begin{pmatrix} 0 & 0 & 1 \\ d & e & f \\ g & h & i \end{pmatrix},$$

where we have used linearity in the top row of the matrix. Next, we use linearity in the second row, and we get a total of 9 terms. Using linearity in the bottom row, we have a total of 27 terms in our sum. An example of one of the terms might be $bei\, D\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The pattern: given the coefficients we picked, there are 1's in their positions and 0's everywhere else. Since this matrix has two equal rows, it contributes 0 to the actual determinant. What **does** contribute to the determinant? Those terms where we pick one variable from each row and each column. We have

$$D\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei\, D\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + afh\, D\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} + bdi\, D\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$+ bfg\, D\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} + cdh\, D\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + ceg\, D\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Next, interchanging any two rows changes the sign of D. We can interchange two rows on each of the second, third and sixth matrices to convert them into I_3 For the fourth and fifth matrices, it takes two row interchanges to convert them into I, and for the first, no row interchanges. After doing these row interchanges, we can factor out $D(I)$ to get

$$D\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = (aei + bfg + cdh - afh - bdi - ceg)D\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This completes the proof of Theorem 2 when $n = 3$, and to get Theorem 1, we normalize $D(I) = 1$ and recover the usual formula for the determinant of a 3×3 matrix. Note this does not prove, technically, that the usual determinant is n-linear and alternating. Instead, what it showed is that if there **were** such a function, it would have to have the formula given (so there are either no such functions or exactly one of them). But from the first set of notes, we know determinants exist. This is what tells us that this formula does, in fact, give a determinant.

Now on to the $n \times n$ case. Let D be an n-linear, alternating function on $n \times n$ matrices. As with the 3×3 case, we can use linearity to expand across each of the n rows. This will result in n^n total terms. Using double subscript notation, each of these terms will have the form

$$a_{1,i_1} a_{2,i_2} \cdots a_{n,i_n}\, D(e_{i_1}, e_{i_2}, \ldots, e_{i_n})$$

for some collection of integers $i_1, \ldots, i_n$, all ranging from 1 to n. Here, e_k means the k'th standard basis vector, but in row form rather than the usual column form. This means we have a formula of sorts:

$$D(A) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} a_{1,i_1} a_{2,i_2} \cdots a_{n,i_n}\, D(e_{i_1}, e_{i_2}, \ldots, e_{i_n}).$$

However, most of the terms in the above sum are 0. In fact, if $n = 5$ the sum has 3125 terms, of which 3005 of them are 0. We try to weed these terms out. As with the 3×3 case, many terms will have equal rows. In fact, if any i_j equals an i_k then rows j and k will be the same and by the alternating property, the term must be 0. What remains is the case where every i_j is different from every other one. As we probably know, a collection of n integers between 1 and n is called a **permutation** of n. There are a total of $n!$ permutations of n, and this is a much smaller number than n^n. It is useful to think of a permutation as a function from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$. That is, the permutation of 4: (3, 2, 4, 1) can be though of as a function f with $f(1) = 3$, $f(2) = 2$, $f(3) = 4$, $f(4) = 1$. As we can see, to be a permutation, f must be one-to-one and onto. It is customary to use Greek letters like σ, τ, π to represent permutations. With such notation, we can write our determinant formula a bit more compactly:

$$D(A) = \sum_{\sigma} a_{1,\sigma(1)} a_{2,\sigma 2} \cdots a_{n,\sigma(n)}\, D(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}).$$

Here, it is understood that the sum ranges over all permutations σ. We can say more: Using the other property of alternating functions, we can re-order the standard basis vectors $e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}$ into the usual order $e_1, e_2, \ldots, e_n$. At this point, we will have $D(e_{i_1}, e_{i_2}, \ldots, e_{i_n}) \to D(I)$. This can be done with some number of row interchanges, and each interchange flips the sign of the term. This means that for each σ there is a sign associated with $D(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)})$, either positive if an even number of row interchanges brings the matrix to I. This sign is often denoted $\operatorname{sgn}(\sigma)$. Now our formula is

$$\begin{aligned}
D(A) &= \sum_{\sigma} a_{1,\sigma(1)} a_{2,\sigma 2} \cdots a_{n,\sigma(n)}\, D(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}) \\
&= \sum_{\sigma} a_{1,\sigma(1)} a_{2,\sigma 2} \cdots a_{n,\sigma(n)}\, \operatorname{sgn}(\sigma) D(I) \\
&= \left(\sum_{\sigma} a_{1,\sigma(1)} a_{2,\sigma 2} \cdots a_{n,\sigma(n)}\, \operatorname{sgn}(\sigma) \right) D(I).
\end{aligned}$$

We say the expression inside the parentheses is $\det(A)$ and we get that for any n-linear, alternating function D, that $D(A) = \det(A) D(I)$. That is, we have proved Theorems 1 and Theorem 2.

There are still some loose ends we can work on. The matrix with rows $e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}$ is called a permutation matrix and is denoted I_σ. For example, if σ is the permutation (5, 1, 3, 2, 4) then $I_\sigma = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$. One way to define $\operatorname{sgn}(\sigma)$ is to let it be $\det(I_\sigma)$. It will be $(-1)^k$ where k is the number of row interchanges needed to transform I_σ to the identity. The value of k is not unique, but it will always have the same parity. One simple way to calculate a k-value is the **inversion number** of the permutation. This is defined to be the smallest number of interchanges of adjacent entries to put the permutation in increasing order. An easy count to get the inversion number is for each number, to add up the number of things to its left that are larger than it. That is, for (5, 1, 3, 2, 4), 5 contributes no inversions, 1 contributes one (5 is left of 1 in σ), 3 contributes one inversion, 2 contributes two (5 and 3) and 4 contributes 1. This means the inversion number, denoted $\mathrm{inv}(\sigma)$ is $0 + 1 + 1 + 2 + 1 = 5$, and the sign of σ is $(-1)^5 = -1$. One commonly written formula for the determinant is

$$\det(A) = \sum_{\sigma} (-1)^{\mathrm{inv}(\sigma)} a_{1,\sigma(1)} a_{2,\sigma 2} \cdots a_{n,\sigma(n)}. \qquad (1)$$

You might want to check that this works in the 3×3 matrix case.

Properties of the determinant

Given our formula for the determinant, and the fact that it is unique, we have several consequences.

Corollary 1 *In the proof that determinants exist, Theorem 3 in the first set of notes, every E_j is the determinant. That is, $E_j(A) = E_k(A)$ for every j and k between 1 and n. Thus one can calculate the determinant by a cofactor expansion down any column.*

Corollary 2 *If A is upper triangular or lower triangular, then* $\det(A)$ *is the product of the elements on its diagonal.*

Proof: Let's give two proofs of this, both fairly similar. First, we can use Corollary 1 and do a cofactor expansion down the first column for upper triangular matrices, or down the last column for lower triangular matrices to reduce to the $(n-1) \times (n-1)$ case. The proof now follows by induction.

For a second proof, use the formula for the determinant. Let's focus on the case of upper triangular matrices. These matrices have the property that $a_{i,j} = 0$ for all $i > j$. We have terms of the form $a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$ to consider. We want terms that are nonzero, so none of the a's should be below the diagonal, meaning, again, that we want $i \leq \sigma(i)$ for all i. If $\sigma(n) \geq n$ we need $\sigma(n) = n$. Now we need $\sigma(n-1) \geq n-1$ but it can't equal n so we need $\sigma(n-1) = n-1$ and by induction, we can show that for all k, $\sigma(k) = k$. This means that there is only one permutation with all entries on or above the main diagonal and that is the permutation $\sigma(i) = i$ for all i so the sum reduces to a single term, $a_{1,1} a_{2,2} \cdots a_{n,n}$, the product of the things on the main diagonal. ■

Corollary 3 *Since the determinant is n-linear, one can use row reduction to calculate it.*

Proof: The catch is that we must pay attention to some of our steps. If we interchange two rows, we record a minus sign, if we divide a row by a constant, c, then we record that constant c as a multiplicative factor. But most interesting, adding a multiple of one row to another does not change the determinant. The reason for this: suppose we add $c(\text{row } i)$ to row j to get a new matrix, A'. We can use linearity in row j in A'. That row will be $(\text{row } j) + c(\text{row } i)$. This means $\det(A') = \det(A_1) + c\det(A_2)$. Now A_1 has row j (from A) as the j'th row, meaning $A_1 = A$. But A_2 has row i (from A) as its j'th row, and this means A_2 has two equal rows, row i and row j, so its determinant is 0, and $\det(A') = \det(A)$. ■

For example,

$$\begin{vmatrix} 4 & 25 & 49 \\ 2 & 5 & 7 \\ 8 & 125 & 343 \end{vmatrix} = -\begin{vmatrix} 2 & 5 & 7 \\ 4 & 25 & 49 \\ 8 & 125 & 343 \end{vmatrix} = -\begin{vmatrix} 2 & 5 & 7 \\ 0 & 15 & 35 \\ 0 & 105 & 315 \end{vmatrix} = -\begin{vmatrix} 2 & 5 & 7 \\ 0 & 15 & 35 \\ 0 & 0 & 70 \end{vmatrix} = -2100.$$

Corollary 4 *If A and B are $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$.*

Proof: Fix B and consider the function $D(A) = \det(AB)$. Since B acts linearly on the rows of A, D is an n-linear function. Moreover, if two rows of A are the same, then two rows of AB are the same so $\det(AB) = 0$ meaning that D is also alternating. By our theorem, $D(A) = \det(A)\,D(I) = \det(A)\det(B)$, as desired. ■

Corollary 5 *Consider the block upper triangular matrix $M = \left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right)$. If both A and C are square matrices then $\det(M) = \det(A)\det(C)$.*

Proof: Define a function $D(C)$ by $D(C) = \det\left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right)$, where A and B are fixed matrices. Then D is n-linear and alternating, where n is the size of C. As such, $D(C) = \det(C)D(I) = \det(C)\det\left(\begin{array}{c|c} A & B \\ \hline 0 & I \end{array}\right)$. Next, we can use row operations to use I to convert B to a matrix of all 0's. Finally, by cofactor expansions down the last n columns, we get that $\det\left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right) = \det(C)\det(A)$, as desired. ■

Note that we could not set $D(A) = \det(M)$ and used n-linearity with A because the matrix B is fixed, so only parts of a row of M would change as we changed rows of A. With C, the entries of M to the left of any row are all 0's, and things work. As an example, by this corollary,

$$\left(\begin{array}{cc|ccc} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 & 7 \\ \hline 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 3 & 5 \\ 0 & 0 & 4 & 9 & 4 \end{array}\right) = \det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \det\begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \\ 4 & 9 & 4 \end{pmatrix} = (-2)(-20) = 40.$$

The determinant of A^t.

We probably all know that a (square) matrix and its transpose have the same determinant. We now try to prove this. Recall that the transpose of a matrix interchanges the rows and columns. This means that the (i, j) entry in A^t is the (j, i) entry in A. If we use Formula (1) we can write

$$\det(A^t) = \sum_\sigma (-1)^{\mathrm{inv}(\sigma)} a_{\sigma(1),1} a_{\sigma 2,2} \cdots a_{\sigma(n),n} \tag{2}$$

We could rearrange the a's to put things in increasing order of the row index again. Doing so changes the permutation, however. In fact, the permutation that takes $\sigma(1)$ to 1, $\sigma(2)$ to 2 and so on is the **inverse permutation** of σ, written σ^{-1}. That is, consider the term $a_{1,4}a_{2,3}a_{3,5}a_{4,1}a_{5,2}$. The permutation here is σ defined by $\sigma(1) = 4,\ \sigma(2) = 3,\ \sigma(3) = 5,\ \sigma(4) = 1,\ \sigma(5) = 2$. When we take the transpose of A the term changes to $a_{4,1}a_{3,2}a_{5,3}a_{1,4}a_{2,5}$ which can be arranged to $a_{1,4}a_{2,5}a_{3,2}a_{4,1}a_{5,3}$. The new permutation, call it τ has $\tau(1) = 4,\ \tau(2) = 5,\ \tau(3) = 2,\ \tau(4) = 1,\ \tau(5) = 3$. This is the inverse permutation to σ: if $\sigma(i) = j$ then $\tau(j) = i$. Now summing over all permutations σ is the same as summing over all permutations σ^{-1}. But there is one catch. If we replace σ by σ^{-1} in (2) then

$$\begin{aligned}\det(A^t) &= \sum_\sigma (-1)^{\mathrm{inv}(\sigma)} a_{\sigma(1),1} a_{\sigma 2,2} \cdots a_{\sigma(n),n} \\ &= \sum_\sigma (-1)^{\mathrm{inv}(\sigma)} a_{1,\sigma^{-1}(1)} a_{2,\sigma^{-1}2} \cdots a_{n,\sigma^{-1}(n)} \\ &= \sum_\sigma (-1)^{\mathrm{inv}(\sigma^{-1})} a_{1,\sigma(1)} a_{2,\sigma 2} \cdots a_{n,\sigma(n)}.\end{aligned}$$

This is almost identical to Formula (1) but we have $\mathrm{inv}(\sigma^{-1})$ instead of $\mathrm{inv}(\sigma)$. We could get rid of the inversion number and write it as $\det(I_{\sigma^{-1}})$ instead. Does this help? I claim that $I_{\sigma^{-1}}$ is the inverse matrix of I_σ. If so, then

$$1 = \det(I) = \det(I_\sigma I_{\sigma^{-1}}) = \det(I_\sigma)\det(I_{\sigma^{-1}}).$$

Since each determinant is ± 1, they must be the same. Knowing this, we get that the inversion numbers for σ and σ^{-1} at least have the same parity and we can replace $\mathrm{inv}(\sigma^{-1})$ by $\mathrm{inv}(\sigma)$ and the final line in the calculation above becomes $\det(A)$ by formula (1), proving $\det(A^t) = \det(A)$.

We must still verify the claim, that $I_{\sigma^{-1}}$ is the inverse matrix of I_σ. It is pretty easy to see that the inverse of I_σ is its transpose. For example, with $\sigma = (5, 1, 3, 2, 4)$, used as an example before, we have

$$I_\sigma I_\sigma^t = \begin{pmatrix} 0&0&0&0&1\\ 1&0&0&0&0\\ 0&0&1&0&0\\ 0&1&0&0&0\\ 0&0&0&1&0 \end{pmatrix} \begin{pmatrix} 0&1&0&0&0\\ 0&0&0&1&0\\ 0&0&1&0&0\\ 0&0&0&0&1\\ 1&0&0&0&0 \end{pmatrix} = \begin{pmatrix} 1&0&0&0&0\\ 0&1&0&0&0\\ 0&0&1&0&0\\ 0&0&0&1&0\\ 0&0&0&0&1 \end{pmatrix}.$$

The reason for this is that column k of the transpose is row k of I_σ, and this has its 1 in exactly the same place so the (k,k) entry in the product will be 1, and all other entries in that column will be 0, meaning the product has all 1's on the diagonal and 0's everywhere else. Now we can write I_σ as the sum of basis matrices:

$$I_\sigma = E_{1,\sigma(1)} + E_{2,\sigma(2)} + \cdots + E_{n,\sigma(n)}.$$

Taking the transpose reverses the subscripts so

$$I_\sigma^t = E_{\sigma(1),1} + E_{\sigma(2),2} + \cdots + E_{\sigma(n),n},$$

but as we saw before, if we reorder these matrices by increasing row number, we get

$$E_{1,\sigma^{-1}(1)} + E_{2,\sigma^{-1}(2)} + \cdots + E_{n,\sigma^{-1}(n)} = I_{\sigma^{-1}},$$

so $I_\sigma^{-1} = I_\sigma^t = I_{\sigma^{-1}}$. Looking at the permutation $\sigma = (5,1,3,2,4)$ again, we see that the transpose matrix corresponds to the permutation $\tau = (2,4,3,5,1)$, the inverse permutation of σ. This is more obvious in the two-line formulation of permutations: $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 3 & 2 & 4 \end{pmatrix}$. The trick to getting the inverse is to interchange top and bottom rows, and sort the top row:

$$\sigma^{-1} = \begin{pmatrix} 5 & 1 & 3 & 2 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 3 & 5 & 1 \end{pmatrix} = \tau.$$

As a corollary, we can now calculate the determinant by a cofactor expansion across any row. Additionally, one can use column reduction (as opposed to row reduction) to calculate a determinant, along with any combination of row reduction, column reduction and cofactor expansion. For example,

$$\begin{vmatrix} 4 & 25 & 49 \\ 2 & 5 & 7 \\ 8 & 125 & 343 \end{vmatrix} = 2 \cdot 5 \cdot 7 \begin{vmatrix} 2 & 5 & 7 \\ 1 & 1 & 1 \\ 4 & 25 & 49 \end{vmatrix} = 70 \begin{vmatrix} 2 & 3 & 5 \\ 1 & 0 & 0 \\ 4 & 21 & 45 \end{vmatrix} = 70 \cdot 3 \cdot 5 \begin{vmatrix} 2 & 1 & 1 \\ 1 & 0 & 0 \\ 4 & 7 & 9 \end{vmatrix}$$

$$= -1050 \begin{vmatrix} 1 & 1 \\ 7 & 9 \end{vmatrix} = -1050 \cdot 2 = -2100.$$

Adjoints, inverses and Cramer's Rule

Given an $n \times n$ matrix A, we define the (i,j) cofactor of A to be $c_{i,j} = (-1)^{i+j} \det(A(i \,|\, j))$. The fact that the determinant can be calculated via a cofactor expansion across any row or down any column is expressed by the formulas

$$\det(A) = a_{i,1}c_{i,1} + a_{i,2}c_{i,2} + \cdots + a_{i,n}c_{i,n} \tag{3}$$
$$= a_{1,j}c_{1,j} + a_{2,j}c_{2,j} + \cdots + a_{n,j}c_{n,j}, \tag{4}$$

where the first line represents an expansion across row i and the second line is the expansion down column j. Given the double subscripts on $c_{i,j}$ it seems natural to gather all the c's into a matrix C, called the cofactor matrix of A. If $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 2 & 8 & 5 \end{pmatrix}$ then $C = \begin{pmatrix} -14 & 1 & 4 \\ 3 & 3 & -6 \\ 1 & -2 & 1 \end{pmatrix}$. Now formula (3) says we can calculate $\det(A)$ by taking any row of A, multiplying entries by the corresponding entries from that row of C and adding. For example, $\det(A) = 2 - 16 + 5 = -9$. Formula (4) says we could do the same for corresponding columns of A and C, for example, $\det(A) = 4 - 18 + 5 = -9$. It is curious that if we play this game, but use two different rows, say row 3 from A and row 1 from C then we get $-28 + 8 + 20 = 0$. Similarly, if we mis-match columns from A and C, the resulting calculation is 0. If this pattern is real, then consider AC^t. In this product, we multiply the ith row of A by the j'th column of C^t to get the (i, j) entry of the product. But the j'th column of C^t is the j'th row of C so when $i = j$ we get $\det(A)$ and when $i \neq j$ we get 0. As a consequence, $AC^t = \det(A)I$ or, when $\det(A) \neq 0$, $A\frac{1}{\det(A)}C^t = I$.

To prove that mismatched rows from A and C combine to get 0, we use a trick. Suppose we form a matrix B from A as follows: B is the same as A but we replace row j of A by row i. That is, B has two equal rows. Now do a cofactor expansion across row j of B. In this case, the cofactors will be the same as for A because for each k, $B(j \mid k) = A(j \mid k)$. What this means is $\det(B) = b_{j,1}c_{j,1} + b_{j,2}c_{j,2} + \cdots + b_{j,n}c_{j,n}$, where the cofactors are the same as the cofactors from A. But row j of B is row i from A so $b_{j,k} = a_{i,k}$ for each k. Since B has two equal rows, $\det(B) = 0 = a_{i,1}c_{j,1} + a_{i,2}c_{j,2} + \cdots + a_{i,n}c_{j,n}$, as desired. This is enough to justify $A\frac{1}{\det(A)}C^t = I$ so we have $A^{-1} = \dfrac{1}{\det(A)}C^t$. Interchanging A and A^{-1} tells us the columns also have the miss-match property. The transpose of the cofactor matrix is often called the **adjoint** of A. Some books say **adjugate** instead. The notation is $\operatorname{adj}(A) = C^t$.

Theorem 3 (Cramer's Rule) *Suppose A is an invertible $n \times n$ matrix. Then the solution to $A\vec{x} = \vec{b}$ is*

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad \ldots, \quad x_n = \frac{\det(A_n)}{\det(A)},$$

where A_k is the matrix obtained by replacing the k'th column of A by $\vec{b}$.

Proof: I'll only sketch the proof of this. We start by solving for $\vec{x}$, $\vec{x} = A^{-1}\vec{b} = \frac{1}{\det(A)}\operatorname{adj}(A)\vec{b}$. To get the k'th coordinate, x_k from $\vec{x}$ we multiply $\vec{b}$ by the k'th row of $\operatorname{adj}(A)$ and divide by the determinant of A. Now the k'th row of $\operatorname{adj}(A)$ is the k'th column of C, the cofactor matrix for A. Now form A_k by replacing the k'th column of A by $\vec{b}$ and do a cofactor expansion down that column. This will perform this calculation of multiplying entries of $\vec{b}$ by the appropriate cofactors. ∎