IV. Determinants

Having encountered linear functionals on $\mathbb{R}^n$, you may wonder about functions on $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ that are linear in each factor separately. For 2 factors, the dot product gives an example, though it's far from unique. On the other hand, if there are n factors and we *also* impose an antisymmetry condition, something very special happens.

IV.A. Alternating multilinear functionals

The determinant of a 2×2 matrix with entries in $\mathbb{R}$, such as

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc, \tag{IV.A.1}$$

is a familiar object. It has a couple of charming properties, the first of which is "linearity in each column [vector] separately":

$$\det \begin{pmatrix} \alpha_1 a_1 + \alpha_2 a_2 & b \\ \alpha_1 c_1 + \alpha_2 c_2 & d \end{pmatrix} = \alpha_1 \det \begin{pmatrix} a_1 & b \\ c_1 & d \end{pmatrix} + \alpha_2 \det \begin{pmatrix} a_2 & b \\ c_2 & d \end{pmatrix},$$

$$\det \begin{pmatrix} a & \beta_1 b_1 + \beta_2 b_2 \\ c & \beta_1 d_1 + \beta_2 d_2 \end{pmatrix} = \beta_1 \det \begin{pmatrix} a & b_1 \\ c & d_1 \end{pmatrix} + \beta_2 \det \begin{pmatrix} a & b_2 \\ c & d_2 \end{pmatrix},$$

which we shall call "bilinearity". The second is the antisymmetry relation

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

(Both relations also hold with respect to the rows.) What we will now show is that the determinant is in fact "uniquely determined" by these two properties and the fact that

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

Uniquely determined amongst what set of objects? Basically, all functions on two copies of $\mathbb{R}^2$ — that is, functions sending any two vectors to a number. (Before we proceed, I note that you may replace $\mathbb{R}$ in what follows by any field of characteristic different from 2.)

So define a *bi*linear functional (or form) on a vector space V to be a function

$$B : V \times V \to \mathbb{R}$$

satisfying (for arbitrary vectors in V and coefficients in $\mathbb{R}$)

$$B(a_1\vec{v}_1 + a_2\vec{v}_2, \vec{w}) = a_1 B(\vec{v}_1, \vec{w}) + a_2 B(\vec{v}_2, \vec{w}),$$

$$B(\vec{v}, c_1\vec{w}_1 + c_2\vec{w}_2) = c_1 B(\vec{v}, \vec{w}_1) + c_2 B(\vec{v}, \vec{w}_2).$$

Call it *alternating* (or antisymmetric) if it also satisfies

$$B(\vec{v}, \vec{w}) = -B(\vec{w}, \vec{v}).$$

for all $\vec{v}, \vec{w} \in V$. Notice that this implies

$$B(\vec{v}, \vec{v}) = -B(\vec{v}, \vec{v}) \quad \Longrightarrow \quad B(\vec{v}, \vec{v}) = 0.$$

Now assume $V = \mathbb{R}^2$, B alternating bilinear, and set $b := B(\hat{e}_1, \hat{e}_2)$. Note that $B(\hat{e}_2, \hat{e}_1) = -b$ and $B(\hat{e}_1, \hat{e}_1) = B(\hat{e}_2, \hat{e}_2) = 0$. Given

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad \text{set } \vec{v}_1 = a_{11}\hat{e}_1 + a_{21}\hat{e}_2,\ \vec{v}_2 = a_{12}\hat{e}_1 + a_{22}\hat{e}_2$$

so that $\vec{v}_1$ and $\vec{v}_2$ are just the columns of A. Then

$$\begin{aligned} B(\vec{v}_1, \vec{v}_2) &= B(a_{11}\hat{e}_1 + a_{21}\hat{e}_2, \vec{v}_2) = a_{11}B(\hat{e}_1, \vec{v}_2) + a_{21}B(\hat{e}_2, \vec{v}_2) \\ &= a_{11}\left(a_{12}B(\hat{e}_1, \hat{e}_1) + a_{22}B(\hat{e}_1, \hat{e}_2)\right) + a_{21}\left(a_{12}B(\hat{e}_2, \hat{e}_1) + a_{22}B(\hat{e}_2, \hat{e}_2)\right) \\ &= a_{11}\left(a_{12}0 + a_{22}b\right) + a_{21}\left(a_{12}(-b) + a_{22}0\right) = b(a_{11}a_{22} - a_{21}a_{12})\,; \end{aligned}$$

that is,

$$B(\vec{v}_1, \vec{v}_2) = b \cdot \det(A).$$

Conclusion: Any *alternating multilinear functional* $B : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is determined uniquely by the number $b = B(\hat{e}_1, \hat{e}_2)$. If $b = 1$, allowing B to act on the column vectors of a 2×2 matrix A yields its determinant, $\det A$.

* * *

Now we proceed to do the same for $n \times n$ matrices, by defining "alternating multilinear functionals"

$$\mathcal{A}: \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ copies}} \to \mathbb{R}$$

and showing that with the additional "normalization" requirement that $\mathcal{A}(\hat{e}_1, \ldots, \hat{e}_n) = 1$, there is *only one*. The only difference is that this unique $\mathcal{A}$, applied to the column vectors of a matrix A, will then be our *definition* of its determinant.

To establish precisely the kind of antisymmetry we need here it is convenient to introduce the "symmetric group" $\mathcal{S}_n$ and sign function *sgn*. First, let a *multi-index* be any function from the set $\{1, \ldots, n\}$ into itself; they will be written $\boldsymbol{\imath} = (i_1, \ldots, i_n)$. A 1-to-1 multi-index is called a *permutation*; they get their own notation σ: $\boldsymbol{\imath} = \sigma$ means simply $(i_1, \ldots, i_n) = (\sigma(1), \ldots, \sigma(n))$. The permutation sends $1 \mapsto \sigma(1)$, $2 \mapsto \sigma(2)$, etc., and we may compose permutations as we would any functions. For instance, let $\sigma = (32)$ (resp. $'\sigma = (13)$) be the permutation swapping 3 and 2 (resp. 1 and 3). Then the composition $(32)(13) = \sigma \circ {}'\sigma$ looks like $1,2,3 \to 3,2,1 \to 2,3,1$.

The set of all permutations σ is called $\mathcal{S}_n$; it contains $n!$ elements. For instance, the elements of $\mathcal{S}_3$ are the permutations mapping

$$1,2,3 \to 1,2,3\ ;\ 1,3,2\ ;\ 2,1,3\ ;\ 2,3,1\ ;\ 3,1,2\ ;\ \text{and}\ 3,2,1.$$

Composition of permutations gives a product $\mathcal{S}_n \times \mathcal{S}_n \to \mathcal{S}_n$, with identity simply the permutation that doesn't do anything. Since every element $\sigma \in \mathcal{S}_n$ (being a 1-to-1 function) has an inverse, $\mathcal{S}_n$ is a group. It is a fact that every permutation is a composition of swaps of two elements (like (13) and (32)). While there are different sequences of swaps one can follow to produce a permutation σ, one thing is the same no matter what route you follow: the parity of the number of swaps (that is, whether the number is even or odd).

Therefore the function

$$sgn : \mathcal{S}_n \to \{\pm 1\}$$

given by

$$\sigma \mapsto (-1)^{\{\text{\# of swaps in } \sigma\}} =: sgn(\sigma)$$

is well-defined. The sgn functions of the six permutations in $\mathcal{S}_3$ above are (respectively) $+, -, -, +, +, -$.

Now we are ready to state our three requirements on $\mathcal{A}$. All $\vec{v}_i$ are arbitrary vectors $\in \mathbb{R}^n$.

• **Multilinearity:**

(IV.A.2)

$$\mathcal{A}(a\vec{v}_1 + 'a\,'\vec{v}_1,\, \vec{v}_2, \ldots, \vec{v}_n) = a\mathcal{A}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) + 'a\mathcal{A}('\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$$

and the same in the remanining $n-1$ entries.

• **Antisymmetry:**

(IV.A.3) $$\mathcal{A}(\vec{v}_{\sigma(1)}, \vec{v}_{\sigma(2)}, \ldots, \vec{v}_{\sigma(n)}) = sgn(\sigma) \cdot \mathcal{A}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)\,;$$

this has the implication that $\mathcal{A}(\vec{v}_1, \ldots, \vec{v}_n) = 0$ if $\vec{v}_i = \vec{v}_j$ for $i \neq j$ (two distinct entries the same). To see this, simply let σ be the swap $(i\,j)$. (We emphasize that the antisymmetry law must hold for *any* n *vectors* $\{\vec{v}_i\}$ and *all* $\sigma \in \mathcal{S}_n$, for $\mathcal{A}$ to be considered antisymmetric or "alternating".)

• **Normalization:**

(IV.A.4) $$\mathcal{A}(\hat{e}_1, \ldots, \hat{e}_n) = 1.$$

Combined with antisymmetry this gives that

$$\mathcal{A}(\hat{e}_{i_1}, \ldots, \hat{e}_{i_n}) = 0 \quad \text{unless } \boldsymbol{\imath} \text{ is a permutation } \sigma$$

in which case

$$\mathcal{A}(\hat{e}_{\sigma(1)}, \ldots, \hat{e}_{\sigma(n)}) = sgn(\sigma) \cdot 1 = sgn(\sigma).$$

We are now ready to calculate $\mathcal{A}$ on the columns $\vec{v}_j = \sum_i a_{ij}\hat{e}_i$ of

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

The result will show $\mathcal{A}$ is uniquely determined, and will also give our definition of $\det(A)$. We start by using multilinearity to obtain

$$\begin{aligned} \mathcal{A}(\vec{v}_1, \ldots, \vec{v}_n) &= \mathcal{A}\left(\sum_{i_1} a_{i_1 1}\hat{e}_{i_1}, \ldots, \sum_{i_n} a_{i_n n}\hat{e}_{i_n}\right) \\ &= \sum_{\boldsymbol{\imath}=(i_1,\ldots,i_n)} a_{i_1 1} \cdot \cdots \cdot a_{i_n n} \cdot \mathcal{A}(\hat{e}_{i_1}, \ldots, \hat{e}_{i_n}). \end{aligned}$$

If this is not completely clear, you may wish to check it step by step (using linearity in each entry separately) for $n = 3$. The next thing is to notice that we can throw out the terms for which $\boldsymbol{\imath}$ is not a permutation σ, since then $\mathcal{A}(\hat{e}_{i_1}, \ldots, \hat{e}_{i_n}) = 0$. So we end up with

$$\sum_{\sigma \in \mathcal{S}_n} a_{\sigma(1)1} \cdot \cdots \cdot a_{\sigma(n)n} \cdot \mathcal{A}(\hat{e}_{\sigma(1)}, \ldots, \hat{e}_{\sigma(n)}) \;=$$

(IV.A.5) $$\boxed{\sum_{\sigma \in \mathcal{S}_n} sgn(\sigma) \cdot a_{\sigma(1)1} \cdot \cdots \cdot a_{\sigma(n)n} \;=:\; \det(A)}$$

generalizing (IV.A.1), and it's no more complicated than that!

Interpretation: You should think of the series of entries

$$\{(\sigma(1), 1), \ldots, (\sigma(2), 2)\}$$

in the matrix A as a graph of a discrete 1-to-1 function from columns to rows (for each column j, choose the $\sigma(j)^{\text{th}}$ entry). The determinant is then the sum *over all such graphs*, of the products of the matrix entries "lying on the graph" (namely $a_{\sigma(1)1}$ thru $a_{\sigma(n)n}$) times an appropriate sign. For 2×2 matrices, there are only two such "graphs": the diagonal and the anti-diagonal

$$\begin{pmatrix} \circ & \\ & \circ \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} & \circ \\ \circ & \end{pmatrix},$$

while for 3×3's you already have 6 (with the same series of signs $+, -, -, +, +, -$ as given above):

$$\text{(IV.A.6)} \qquad \begin{pmatrix} \circ & & \\ & \circ & \\ & & \circ \end{pmatrix}, \begin{pmatrix} \circ & & \\ & & \circ \\ & \circ & \end{pmatrix}, \begin{pmatrix} & \circ & \\ \circ & & \\ & & \circ \end{pmatrix}, \begin{pmatrix} & & \circ \\ \circ & & \\ & \circ & \end{pmatrix}, \begin{pmatrix} & \circ & \\ & & \circ \\ \circ & & \end{pmatrix}, \begin{pmatrix} & & \circ \\ & \circ & \\ \circ & & \end{pmatrix}$$

For $n \times n$'s there are $n!$ of these "graphs", since of course they are just permutations. The way I wrote the 3×3's should already suggest to you the Laplace expansion in the first column (see the next section).

Exercises

(1) For an $n \times n$ matrix A, find the k for which $\det(rA) = r^k \det(A)$ holds. [Hint: use (IV.A.2).]

(2) Prove that the determinant of the Vandermonde matrix

$$\begin{pmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{pmatrix}$$

is $(b-a)(c-a)(c-b)$. [Hint: use (IV.A.5) and (IV.A.6).]

(3) An $n \times n$ matrix A is called triangular if $A_{ij} = 0$ whenever $i > j$ or if $A_{ij} = 0$ whenever $i < j$. Using (IV.A.5), prove that the determinant of a triangular matrix is the product $A_{11}A_{22} \cdots A_{nn}$ of its diagonal entries.

(4) Let A be a 3×3 matrix over the field of complex numbers. We form the matrix $x\mathbb{I}_3 - A$ with polynomial entries, the (i,j) entry of this matrix being the polynomial $x\delta_{ij} - A_{ij}$. If $f = \det(x\mathbb{I} - A)$, show that f is a monic polynomial of degree 3. ("Monic" means that the coefficient of the highest power of x is 1.) If we write $f = (x - c_1)(x - c_2)(x - c_3)$ with $c_i \in \mathbb{C}$, prove that $\sum c_i = \operatorname{tr}(A)$ and $c_1c_2c_3 = \det(A)$.