The determinant

J. Palmer

Abstract

A (mostly) self-contained introduction to the determinant aimed at advanced undergraduates, written for the Math 350H class at Rutgers spring semester of 2017. Most of the material is taken from [1, 2], but it has been substantially reorganized. The goal of this article is to present an organized and motivated discussion of the abstract definition of the determinant and its most basic properties.

1 Notation

For the duration of the article let F denote an arbitrary field, and we use $M_{m\times n}(F)$ to denote the collection of $m \times n$ matrices with entries in F. Given a function $\delta\colon M_{n\times n}(F) \to F$ we will occasionally write

$$\delta \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

where $\alpha_1, \dots, \alpha_n \in F^n$ to mean δ applied to the matrix with rows $\alpha_1, \dots, \alpha_n$. We use $I_n \in M_{n\times n}(F)$ to denote the $n \times n$ identity matrix, which is all ones on the diagonal and all zeros off the diagonal. We denote by $e_1, \dots, e_n$ the rows of I_n, so that $e_j \in F^n$ is a vector of all zeros except for 1 in the j^{th} position. For $A \in M_{n\times n}(F)$ we write A_{ij} for the element in the i^{th} row and j^{th} column of A.

For $A \in M_{n\times n}(F)$ we denote by $\widetilde{A_{ij}}$ the matrix in $M_{(n-1)\times(n-1)}(F)$ obtained by removing the i^{th} row and j^{th} column from A (Definition 5.2).

2 Motivation: signed area and its properties

2.1 Signed area of a parallelogram

Let $u, v \in \mathbb{R}^2$ and consider the set

$$P(u, v) = \{tu + sv \mid 0 \leqslant t \leqslant 1, 0 \leqslant s \leqslant 1\}.$$

This is the parallelogram in $\mathbb{R}^2$ spanned be the vectors u and v, which degenerates to a line if u and v are parallel (when there exists $\lambda \in \mathbb{R}$ such that $u = \lambda v$). Let $A(u, v)$ denote the area of $P(u, v)$, which is zero in the case that $u = \lambda v$ for some $\lambda \in \mathbb{R}$. Next, let $\mathcal{O}(u, v)$ denote the *orientation* of the pair (u, v), which means that if $\theta \in [0, 2\pi)$ is the angle from u to v then

$$\mathcal{O}(u, v) = \begin{cases} +1 & 0 < \theta < \pi \\ -1 & \pi < \theta < 2\pi \\ 0 & \text{otherwise.} \end{cases}$$

The cases in which u and v are parallel (so $\mathcal{O}(u, v) = 0$) will turn out to not be important, but we define $\mathcal{O}$ there as well for completeness.

Now we will consider the product of these two functions, the function from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ given by $(u, v) \mapsto \mathcal{O}(u, v)A(u, v)$, which we call the *signed area of the parallelogram spanned by u and v*. Notice that since the domain of this function is $\mathbb{R}^2 \times \mathbb{R}^2 = \mathbb{R}^4$ we may also consider it as a function from $M_{2\times 2}(\mathbb{R})$ to $\mathbb{R}$ where the two vectors u and v are the rows of the matrix.

Definition 2.1. Define the signed area function, denoted by δ_{area}, by

$$\begin{aligned} \delta_{\text{area}} \colon M_{2\times 2}(\mathbb{R}) &\to \mathbb{R} \\ \begin{pmatrix} u \\ v \end{pmatrix} &\mapsto \mathcal{O}(u,v)A(u,v) \end{aligned}$$

where $u, v \in \mathbb{R}^2$ are row vectors and $\begin{pmatrix} u \\ v \end{pmatrix}$ is the 2×2 matrix with first row u and second row v.

2.2 Properties of signed area

This function has three important properties:

Alternating: Notice that $\delta_{\text{area}} \begin{pmatrix} u \\ u \end{pmatrix} = 0$ for any $u \in \mathbb{R}^2$ because this is the case in which the parallelogram degenerates to a line.

Normalized: Notice that $\delta_{\text{area}}(I_2) = 1$ because the unit square has area 1 and the basis $(1, 0)$, $(0, 1)$ is positively oriented.

2-linear: First, notice that δ_{area} is *not* a linear function on $M_{2\times 2}(\mathbb{R})$. This is because, for instance

$$\delta_{\text{area}} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \delta_{\text{area}} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = 0 + 0 = 0$$

but

$$\delta_{\text{area}} \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right) = \delta_{\text{area}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1.$$

While it is not linear on $M_{2\times 2}(\mathbb{R})$, it does satisfy a related property. Thinking of δ_{area} as a function just on $\mathbb{R}^2$, by fixing one of the rows, it is linear. That is, for any $u, u', v, v' \in \mathbb{R}^2$ and $c \in \mathbb{R}$ it is true that

$$\delta_{\text{area}} \begin{pmatrix} u + cu' \\ v \end{pmatrix} = \delta_{\text{area}} \begin{pmatrix} u \\ v \end{pmatrix} + c\delta_{\text{area}} \begin{pmatrix} u' \\ v \end{pmatrix}$$

and

$$\delta_{\text{area}} \begin{pmatrix} u \\ v + cv' \end{pmatrix} = \delta_{\text{area}} \begin{pmatrix} u \\ v \end{pmatrix} + c\delta_{\text{area}} \begin{pmatrix} u \\ v' \end{pmatrix}.$$

I will not prove this here (since it is tedious), but it can be proven with basic geometry. See [1, pages 205-207].

Conclustion: We conclude that there exists an alternating, 2-linear function on $M_{2\times 2}(\mathbb{R})$ which sends I_2 to 1.

2.3 Implications of these properties

Suppose now that F is an arbitrary field and $\delta\colon M_{2\times 2}(F) \to F$ is a function which has these three properties: it is normalized, alternating, and 2-linear. By this I mean that $\delta(I_2) = 1$, $\delta(A) = 0$ if the two rows of A are equal, and that

$$\delta\begin{pmatrix} u + cu' \\ v \end{pmatrix} = \delta\begin{pmatrix} u \\ v \end{pmatrix} + c\delta\begin{pmatrix} u' \\ v \end{pmatrix} \text{ and } \delta\begin{pmatrix} u \\ v + cv' \end{pmatrix} = \delta\begin{pmatrix} u \\ v \end{pmatrix} + c\delta\begin{pmatrix} u \\ v' \end{pmatrix}$$

for all $u, u', v, v' \in \mathbb{R}^2$, $c \in \mathbb{R}$.

Let $u, v \in \mathbb{R}^2$. Then by the alternating and 2-linearity properties we see that

$$0 = \delta\begin{pmatrix} u+v \\ u+v \end{pmatrix} = \delta\begin{pmatrix} u \\ u \end{pmatrix} + \delta\begin{pmatrix} u \\ v \end{pmatrix} + \delta\begin{pmatrix} v \\ u \end{pmatrix} + \delta\begin{pmatrix} v \\ v \end{pmatrix} = 0 + \delta\begin{pmatrix} u \\ v \end{pmatrix} + \delta\begin{pmatrix} v \\ u \end{pmatrix}$$

so

$$\delta\begin{pmatrix} u \\ v \end{pmatrix} = -\delta\begin{pmatrix} v \\ u \end{pmatrix}$$

for any $u, v \in \mathbb{R}^2$. Thus, we conclude that

$$\delta\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -\delta\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = -1$$

using the normalization property. Thus, if we let $e_1 = (1, 0)$ and $e_2 = (0, 1)$ we have that

$$\delta\begin{pmatrix} e_1 \\ e_1 \end{pmatrix} = \delta\begin{pmatrix} e_2 \\ e_2 \end{pmatrix} = 0,\ \delta\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = 1,\ \text{and } \delta\begin{pmatrix} e_2 \\ e_1 \end{pmatrix} = -1. \tag{2.1}$$

Now, let $\begin{pmatrix} u \\ v \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$ be arbitrary, and let $u = (a, b)$, $v = (c, d)$, so that $u = ae_1 + be_2$ and $v = ce_1 + de_2$. By using *only* the properties of normalization, alternating, and 2-linear (and Equation (2.1) which we derived from these properties) we can deduce that

$$\begin{aligned}
\delta\begin{pmatrix} u \\ v \end{pmatrix} &= \delta\begin{pmatrix} ae_1 + be_2 \\ v \end{pmatrix} = a\delta\begin{pmatrix} e_1 \\ v \end{pmatrix} + b\delta\begin{pmatrix} e_2 \\ v \end{pmatrix} \\
&= a\left(\delta\begin{pmatrix} e_1 \\ ce_1 + de_2 \end{pmatrix}\right) + b\left(\delta\begin{pmatrix} e_2 \\ ce_1 + de_2 \end{pmatrix}\right) \\
&= a\left(c\delta\begin{pmatrix} e_1 \\ e_1 \end{pmatrix} + d\delta\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}\right) + b\left(c\delta\begin{pmatrix} e_2 \\ e_1 \end{pmatrix} + d\delta\begin{pmatrix} e_2 \\ e_2 \end{pmatrix}\right) \\
&= a(c(0) + d(1)) + b(c(-1) + d(0)) \\
&= ad - bc.
\end{aligned}$$

So we have proven the following statement: **If there exists a 2-linear, alternating, normalized function on** $M_{2\times2}(\mathbb{R})$, then it must be given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto ad - bc.$$

It is important to notice that in the discussion in this section we have not proved that such a function does exist, but only what the formula would have to be if it does exist. That is, it may be that no such function exists and that the formula we derived does not actually satisfy all of the desired properties. In Section 2.2 we showed that there is at least one function with all of these properties, namely δ_{area}, so we have the following:

Theorem 2.2. *There is precisely one function* $\delta\colon M_{2\times2}(\mathbb{R}) \to \mathbb{R}$ *which is* 2*-linear, alternating, and satisfies* $\delta(I_2) = 1$*. This function is given by*

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

Definition 2.3. Let $\det\colon M_{2\times2}(\mathbb{R}) \to \mathbb{R}$ denote the unique 2-linear, alternating function which sents I_2 to 1.

Thus, we have found a formula for the function δ_{area}. Since $\delta_{\text{area}}\colon M_{2\times2}(\mathbb{R}) \to \mathbb{R}$ is 2-linear, alternating, and satisfies $\delta_{\text{area}}(I_2) = 2$ by Theorem 2.2 we see that $\delta_{\text{area}} = \det$ so

$$\delta_{\text{area}} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Of course, this function is the usual determinant of a 2×2 matrix. In Section 5 we will follow this same method to define the determinant for $n \times n$ matrices, where $n \geqslant 1$.

3 An application for 2×2 matrices

Recall that a matrix $A \in M_{2\times2}(F)$ is invertible if there exists another matrix $B \in M_{2\times2}(F)$ such that $AB = BA = I_2$. We call B the inverse of A and use the notation $B = A^{-1}$.

Remark 3.1. The following theorem holds for any field F, but we state it just for $\mathbb{R}$ because so far we have only shown that the det function exists for real 2×2 matrices. ⊘

Theorem 3.2. *Let* $A \in M_{2\times2}(\mathbb{R})$*. Then* A *in invertible if and only if* $\det(A) \neq 0$*. Moreover, in the case that* $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ *is invertible*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

Proof. Supposing that $\det(A) \neq 0$ we may define

$$B = \frac{1}{\det(A)} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

and straightforward computation gives $AB = BA = I_2$, which means that A is invertible with the desired inverse.

Now suppose that A is invertible. This means that A has rank 2 so the columns of A are linearly independent and thus the vector $\begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix}$ is nonzero. This means either $A_{11} \neq 0$ or $A_{21} \neq 0$.

If $A_{11} \neq 0$ then A can be transformed into

$$A' = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} - \frac{A_{21}}{A_{11}} A_{12} \end{pmatrix}$$

by subtracting $\frac{A_{21}}{A_{11}}$ times the first row from the second. Such a transformation does not change the rank of a matrix, so A' also has rank 2. This means the rows of A' are linearly independent so $A_{22} - \frac{A_{21}}{A_{11}} A_{12} \neq 0$ which implies $A_{11}A_{22} - A_{21}A_{12} \neq 0$ so $\det(A) \neq 0$. The case in which $A_{21} \neq 0$ is similar. □

4 Background

4.1 Definitions

Definition 4.1. Let $\delta\colon M_{n\times n}(F) \to F$. Then:

1. δ is *n-linear* if it is linear in each row while the others are held fixed. That is, for all $i = 1, \ldots, n$, for all $\alpha_1, \ldots, \alpha_n, \beta \in F^n$, and for all $c \in F$,

$$\delta \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{i-1} \\ \alpha_i + c\beta \\ \alpha_{i+1} \\ \vdots \\ \alpha_n \end{pmatrix} = \delta \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{i-1} \\ \alpha_i \\ \alpha_{i+1} \\ \vdots \\ \alpha_n \end{pmatrix} + c\delta \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{i-1} \\ \beta \\ \alpha_{i+1} \\ \vdots \\ \alpha_n \end{pmatrix};$$

2. δ is *alternating* if $\delta(A) = 0$ whenever $A \in M_{n\times n}(F)$ has two equal rows;

3. δ is *normalized* if $\delta(I_n) = 1$.

4.2 Implication of the alternating property

The definition of alternating in Definition 4.1 looks like it is inappropriately named, since that definition does not have to do with anything alternating. The following result shows that this definition implies a condition which aligns better with the name (and which is equivalent in many cases).

Proposition 4.2. *Let $\delta\colon M_{n\times n}(F) \to F$ be n-linear and alternating. If $A, B \in M_{n\times n}(F)$ where B is obtained from A by switching two rows, then $\delta(A) = -\delta(B)$.*

Proof. For simplicity of notation assume that B is obtained from A by switching the first two rows. That is, let $\alpha_1, \dots, \alpha_n \in F^n$ be the rows of A so that

$$A = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix} \text{ and } B = \begin{pmatrix} \alpha_2 \\ \alpha_1 \\ \alpha 3 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Then, where the first equality is because the first two rows are equal and the others are by n-linearity, we have

$$\begin{aligned} 0 &= \delta \begin{pmatrix} \alpha_1+\alpha_2 \\ \alpha_1+\alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix} = \delta \begin{pmatrix} \alpha_1 \\ \alpha_1+\alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix} + \delta \begin{pmatrix} \alpha_2 \\ \alpha_1+\alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix} \\ &= \delta \begin{pmatrix} \alpha_1 \\ \alpha_1 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix} + \delta \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix} + \delta \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} + \delta \begin{pmatrix} \alpha_2 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix} \\ &= 0 + \delta A + \delta(B) + 0. \end{aligned}$$

The result follows, and the case of switching two arbitrary rows is nearly identical. □

Remark 4.3. Consider the case that $F = \mathbb{R}$. In this case any n-linear function $\delta\colon M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ for which $\delta(A) = -\delta(B)$ whenever B is obtained by switching two rows of A is alternating (can you see why?). Thus, if $F = \mathbb{R}$ then alternating can be defined as producing a minus sign whenever two rows are switched, but this is not true for general fields. It is true for most fields you can think of ($\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, and even $\mathbb{Z}_p$ where p is a prime and $p \neq 2$) but it does not hold for fields such as $\mathbb{Z}_2$, which have so-called *characteristic* equal to 2. The characteristic of a field is not really in the scope of these notes, so I won't say anything more about it. ⊘

4.3 Permutation groups

For any $n \in \mathbb{Z}$ with $n > 0$ define the set

$$S_n = \{\sigma\colon \{1,\ldots,n\} \to \{1,\ldots,n\} \mid \sigma \text{ is a bijection}\}.$$

Recall that a function σ is a bijection if and only if it is one-to-one and onto. The set S_n is actually a group (where the group operation is composition of functions - recall that the composition of two bijections is a bijection) and it is referred to as the *symmetric group* or the *symmetric group on n elements*. The elements of S_n are called *permutations* and can be thought of as reordering the elements $\{1,\ldots,n\}$. There is a rich theory of S_n and its subgroups, but for our purposes we will only really need the definition and one important fact.

For any $\sigma \in S_n$ define $N(\sigma)$ to be the number of pairs of elements which switch order between the original ordering of $\{1,\ldots,n\}$ and the new ordering determined by σ. That is, if

$$S(\sigma) = \{i, j \in \{1,\ldots,n\} \mid i < j \text{ and } \sigma(i) > \sigma(j)\}$$

then $N(\sigma)$ is the size (or *cardinality*) of the set $S(\sigma)$.

Definition 4.4. Let $\operatorname{sgn}\colon S_n \to \{-1,+1\}$ be given by $\operatorname{sgn}(\sigma) = (-1)^{N(\sigma)}$. We call $\operatorname{sgn}(\sigma)$ the *sign* of the permutation σ. We say that σ is *even* if $\operatorname{sgn}(\sigma) = 1$ and we say that σ is *odd* if $\operatorname{sgn}(\sigma) = -1$.

A *transposition* is a permutation that only switches two elements and leaves the rest alone. That is, $\sigma \in S_n$ is a transposition if there exists $i, j \in \{1,\ldots,n\}$ such that $\sigma(i) = j$, $\sigma(j) = i$, and $\sigma(k) = k$ for all $k \in \{1,\ldots,n\}$ with $k \neq j$ and $k \neq i$. If σ is a transposition that $\operatorname{sgn}(\sigma) = -1$ because $N(\sigma) = 1$. The following proposition is not hard to prove, but I will omit the proof anyway.

Proposition 4.5. *Every permutation can be expressed as a composition of transpositions. Moreoever, if $\sigma \in S_n$ is a composition of k permutations then* $\operatorname{sgn}(\sigma) = (-1)^k$.

The first part of Proposition 4.5 is obvious, while the idea of the proof of the second part of Proposition 4.5 that each additional transposition either increases or decreases $N(\sigma)$ by 1, and in either case changes $\operatorname{sgn}(\sigma)$.

4.4 Permutation groups and alternating n-linear functions

Let $e_j \in F^n$ be a vector of all zeros except for a 1 in the j^{th} position, so that

$$I_n = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}.$$

Let $\mathcal{F}_n$ be the collection of all functions from $\{1,\ldots,n\}$ to itself, so

$$\mathcal{F}_n = \{f\colon \{1,\ldots,n\} \to \{1,\ldots,n\}\}.$$

Notice now that $S_n \subset \mathcal{F}_n$ where S_n is those elements of $\mathcal{F}_n$ which are bijections.

Proposition 4.6. *Let $\delta\colon M_{n\times n}(F) \to F$ be an n-linear, alternating function. If $f \in \mathcal{F}_n$ then*

$$\delta \begin{pmatrix} e_{f(1)} \\ e_{f(2)} \\ \vdots \\ e_{f(n)} \end{pmatrix} = \begin{cases} 0 & \text{if } f \notin S_n \\ \operatorname{sgn}(f)\delta(I_n) & \text{if } f \in S_n. \end{cases}$$

Proof. Recall that $f \in \mathcal{F}_n$ is a injection if and only if it is a surjection if and only if it is a bijection because its domain and codomain are both finite with the same cardinality. Thus, if $f \notin S_n$ we know that f is not injective, so there exists some distinct $i, j \in \{1, \ldots, n\}$ such that $f(i) = f(j)$ which means that the matrix with rows $e_{f(1)}, \ldots, e_{f(n)}$ has two rows which are equal, so

$$\delta \begin{pmatrix} e_{f(1)} \\ \vdots \\ e_{f(n)} \end{pmatrix} = 0$$

by definition since f is alternating.

All that remains is the case that $f \in S_n$. Since any permutation can be expressed as a sequence of transpositions (recall this expression is not unique, but it does always exist) let $f = \sigma_k \circ \ldots \sigma_1$ where each $\sigma_i \in S_n$ is a transposition. The strategy now is to apply the transpositions one at a time to the identity matrix, and recall by Proposition 4.2 that each time two rows are switched it multiples δ by -1. So we have

$$\delta \begin{pmatrix} e_{\sigma_1(1)} \\ \vdots \\ e_{\sigma_1(n)} \end{pmatrix} = (-1)\delta(I_n),$$

$$\delta \begin{pmatrix} e_{\sigma_2\circ\sigma_1(1)} \\ \vdots \\ e_{\sigma_2\circ\sigma_1(n)} \end{pmatrix} = (-1)\delta \begin{pmatrix} e_{\sigma_1(1)} \\ \vdots \\ e_{\sigma_1(n)} \end{pmatrix} = (-1)(-1)\delta(I_n) = (-1)^2\delta(I_n)$$

$$\delta \begin{pmatrix} e_{\sigma_3\circ\sigma_2\circ\sigma_1(1)} \\ \vdots \\ e_{\sigma_3\circ\sigma_2\circ\sigma_1(n)} \end{pmatrix} = (-1)\delta \begin{pmatrix} e_{\sigma_2\circ\sigma_1(1)} \\ \vdots \\ e_{\sigma_2\circ\sigma_1(n)} \end{pmatrix} = (-1)^3\delta(I_n)$$

and continuing in this way leads to

$$\delta \begin{pmatrix} e_{\sigma_k\circ\ldots\circ\sigma_1(1)} \\ \vdots \\ e_{\sigma_k\circ\ldots\circ\sigma_1(n)} \end{pmatrix} = (-1)^k\delta(I_n)$$

(I have been slightly informal here, but it can be done formally with induction). Thus, since $f = \sigma_k \circ \ldots \sigma_1$ we have

$$\delta \begin{pmatrix} e_{f(1)} \\ \vdots \\ e_{f(n)} \end{pmatrix} = (-1)^k\delta(I_n)$$

and the result follows since by Proposition 4.5 we know that $\operatorname{sgn}(f) = (-1)^k$. ☐

5 Defining the determinant for $n \times n$ matrices

5.1 Uniqueness

In this section we **assume that an n-linear, alternating function from $M_{n\times n}(F)$ to F exists** and prove that it is unique and must be given by a specific formula. This is generalizing what we did in Section 2.3 for the 2 dimensional case.

Proposition 5.1. *If $\delta\colon M_{n\times n}(F) \to F$ is n-linear and alternating then*

$$\delta(A) = \left(\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i\,\sigma(i)} \right) \delta(I_n).$$

In particular, if $\delta(I_n) = 1$ then

$$\delta(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i\,\sigma(i)}$$

and thus there is at most one n-linear, alternating, normalized function from $M_{n\times n}(F)$ to F.

Proof. Let $A \in M_{n\times n}(F)$ with entries A_{ij} and let $\alpha_1, \ldots, \alpha_n \in F^n$ denote the rows of A. Then the i^{th} row of A is given by

$$\alpha_i = (A_{i1}, A_{i2}, \ldots, A_{in}) = \sum_{k_i=1}^{n} A_{ik_i} e_{k_i} \tag{5.1}$$

where we have used a different "dummy variable" for each sum to avoid confusion. Equation (5.1) and the n-linearity of δ imply that

$$\begin{aligned}
\delta(A) = \delta \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} &= \delta \begin{pmatrix} \sum_{k_1=1}^n A_{1k_1} e_{k_1} \\ \sum_{k_2=1}^n A_{2k_2} e_{k_2} \\ \vdots \\ \sum_{k_n=1}^n A_{nk_n} e_{k_n} \end{pmatrix} \\
&= \sum_{k_1=1}^n A_{1k_1} \delta \begin{pmatrix} e_{k_1} \\ \sum_{k_2=1}^n A_{2k_2} e_{k_2} \\ \vdots \\ \sum_{k_n=1}^n A_{nk_n} e_{k_n} \end{pmatrix} \\
&= \sum_{k_1=1}^n A_{1k_1} \left(\sum_{k_2=1}^n A_{2k_2} \delta \begin{pmatrix} e_{k_1} \\ e_{k_2} \\ \sum_{k_3=1}^n A_{3k_3} e_{k_3} \\ \vdots \\ \sum_{k_n=1}^n A_{nk_n} e_{k_n} \end{pmatrix} \right) \\
&= \sum_{k_1=1}^n \sum_{k_2=1}^n \cdots \sum_{k_n=1}^n \left(A_{1k_1} A_{2k_2} \cdots A_{nk_n} \delta \begin{pmatrix} e_{k_1} \\ e_{k_2} \\ \vdots \\ e_{k_n} \end{pmatrix} \right) \\
&= \sum_{k_1, \dots, k_n \in \{1, \dots, n\}} \prod_{i=1}^n A_{ik_i} \delta \begin{pmatrix} e_{k_1} \\ e_{k_2} \\ \vdots \\ e_{k_n} \end{pmatrix}.
\end{aligned}$$

Now, we can rewrite this sum be realizing that there is a bijection between the set $\mathcal{F}_n$ and the set $\{k_1, \dots, k_n \in \{1, \dots, n\}\}$ given by $f \mapsto (f(1), \dots, f(n))$ so we have

$$\delta(A) = \sum_{f \in \mathcal{F}_n} \prod_{i=1}^n A_{if(i)} \delta \begin{pmatrix} e_{f(1)} \\ e_{f(2)} \\ \vdots \\ e_{f(n)} \end{pmatrix}. \tag{5.2}$$

Now we apply Proposition 4.6 to Equation 5.2 to get

$$\begin{aligned}
\delta(A) &= \sum_{f\in\mathcal{F}_n}\left(\prod_{i=1}^{n} A_{if(i)}\right)\delta\begin{pmatrix} e_{f(1)} \\ e_{f(2)} \\ \vdots \\ e_{f(n)} \end{pmatrix} \\
&= \sum_{f\in S_n}\left(\prod_{i=1}^{n} A_{if(i)}\right)\delta\begin{pmatrix} e_{f(1)} \\ e_{f(2)} \\ \vdots \\ e_{f(n)} \end{pmatrix} \\
&= \sum_{f\in S_n} \operatorname{sgn}(f)\prod_{i=1}^{n} A_{if(i)}\delta(I_n)
\end{aligned}$$

as desired. □

5.2 Existence

In this section we prove that an n-linear, alternating, normalized function from $M_{n\times n}(F)$ to F exists using the "cofactor expansion" technique, which gives an alternate formula for the determinant.

Definition 5.2. If $A \in M_{n\times n}(F)$ then for each $i, j \in \{1, \ldots, n\}$ let $\widetilde{A_{ij}}$ denote the matrix in $M_{(n-1)\times(n-1)}(F)$ given by removing the i^{th} row and j^{th} column from A.

Proposition 5.3. *Let $\delta : M_{(n-1)\times(n-1)}(F) \to F$ be an $(n-1)$-linear alternating function. Then for any choice of $j \in \{1, \ldots, n\}$ the function $E_j : M_{n\times n}(F) \to F$ given by*

$$E_j(A) = \sum_{i=1}^{n}(-1)^{i+j}A_{ij}\delta(\widetilde{A_{ij}})$$

is an n-linear alternating function.

Proof. First fix some $i, j \in \{1, \ldots, n\}$ and consider the function $\Phi\colon M_{n\times n}(F) \to F$ given by $\Phi(A) = A_{ij}\delta(\widetilde{A_{ij}})$. Let $\alpha_1, \ldots, \alpha_n \in F^n$ denote the rows of A, and let B be the same matrix except that row k is replaced by $\alpha_k + c\beta$ for some $c \in F$, $\beta \in F^n$ where $\beta = (b_1, \ldots, b_n)$. Also, let C be the matrix which is equal to A and B for all rows except row k, where the k^{th} row of C is given by β. We claim that for any value of k we have that $\Phi(B) = \Phi(A) + \Phi(C)$, which means that Φ is n-linear.

If $k = i$ then $\widetilde{B_{ij}} = \widetilde{A_{ij}}$ because the only different row has been removed, and $B_{ij} = A_{ij} + cb_j$, so

$$\Phi(B) = (A_{ij} + cb_j)\delta(\widetilde{A_{ij}}) = \Phi(A) + \Phi(C)$$

as desired. If $k \neq i$ then $B_{ij} = A_{ij} = C_{ij}$ and δ is n-linear, so again we have that $\Phi(B) = \Phi(A) + \Phi(C)$. Thus Φ is n-linear and since a linear combination of n-linear functions is still n-linear we conclude that E_j is n-linear.

To complete the proof we must show that E_j is alternating. Suppose that $A \in M_{n\times n}(F)$ has rows $\alpha_1, \ldots, \alpha_n \in F^n$ and for some distinct $k_1, k_2 \in \{1, \ldots, n\}$ we have that $\alpha_{k_1} = \alpha_{k_2}$, where we may assume that $k_1 < k_2$. We must show that $E_j(A) = 0$. If $k_1, k_2 \neq i$ then $\delta(\widetilde{A_{ij}}) = 0$ because δ is alternating and $\widetilde{A_{ij}}$ has two equal rows. Thus,

$$E_j(A) = \sum_{i=1}^{j} (-1)^{i+j} A_{ij}\delta(\widetilde{A_{ij}}) = (-1)^{k_1+j} A_{k_1 j}\delta(\widetilde{A_{k_1 j}}) + (-1)^{k_2+j} A_{k_2 j}\delta(\widetilde{A_{k_2 j}})$$

because all other terms in the sum are zero. To show that $E_j(A) = 0$ it is sufficient to argue that

$$A_{k_1 j}\delta(\widetilde{A_{k_1 j}}) + (-1)^{k_2-k_1} A_{k_2 j}\delta(\widetilde{A_{k_2 j}}) = 0$$

and in fact since the k_1 and k_2 rows of A are equal we know that $A_{k_1 j} = A_{k_2 j}$ so it is sufficient to show

$$\delta(\widetilde{A_{k_1 j}}) = (-1)^{k_2-k_1-1}\delta(\widetilde{A_{k_2 j}}).$$

Now, this equation holds because since the rows k_1 and k_2 are equal these two matrices, $\widetilde{A_{k_1 j}}$ and $\widetilde{A_{k_2 j}}$, have the same rows in a different order (since each of these matrices has eliminated one of the pair of rows which are equal to each other). Specifically, $\widetilde{A_{k_1 j}}$ may be obtained from $\widetilde{A_{k_2 j}}$ by moving the $(k_1)^{\text{th}}$ row of $\widetilde{A_{k_2 j}}$ down to the $k_2 - 1$ position by switching positions with the adjacent row $k_2 - k_1 - 1$ times. By Proposition 4.2 this introduces a factor of -1 each time on the function δ, since δ is alternating. The result follows. □

Corollary 5.4. *For each $n \in \mathbb{Z}$, $n \geqslant 1$, there exists an n-linear, alternating, normalized function from $M_{n\times n}(F) \to F$.*

Proof. We proceed by induction on n. For the case of $n = 1$ consider the function $\delta(A) = A_{11}$ where $A = (A_{11})$ is a 1×1 matrix. It is clear that $\delta(I_1) = 1$ and $\delta(\alpha + c\beta) = \delta(\alpha) + c\delta(\beta)$ for all $\alpha, \beta, c \in F$. Furthermore, it is vacuously true that δ is alternating, so we conclude δ is the desired function.

Now suppose the claim holds for some fixed n, so there exists an n-linear, alternating $\delta\colon M_{n\times n}(F) \to F$ for which $\delta(I_n) = 1$. Then by Proposition 5.3 the function $E_1\colon M_{(n+1)\times(n+1)}(F) \to F$ given by

$$E_1(A) = \sum_{i=1}^{n} (-1)^{i+1} A_{i1}\delta(\widetilde{A_{i1}})$$

is also n-linear and alternating. Moreover, since the entry in the i^{th} row of the first column of I_{n+1} is nonzero if and only if $i = 1$ we see that

$$E_1(I_{n+1}) = \sum_{i=1}^{n} (-1)^{i+1} (I_{n+1})_{i1}\delta(\widetilde{(I_{n+1})_{i1}}) = (-1)^{1+1}(I_{n+1})_{11}\delta(\widetilde{(I_{n+1})_{11}}) = 1\delta(I_n) = 1$$

so E_1 is normalized as well. □

5.3 Conclusion: the existence of a unique determinant

Now we can summarize the results of this section, and finally define the determinant for an $n\times n$ matrix. The following theorem is implied by combining Proposition 5.1 and Corollary 5.4.

Theorem 5.5. *For each $n \in \mathbb{Z}$, $n > 0$, and each field F there exists a unique n-linear, alternating, normalized function from $M_{n\times n}(F)$ to F.*

In light of Theorem 5.5 we may make the following definition.

Definition 5.6. We denote by $\det\colon M_{n\times n}(F) \to F$ the unique n-linear, alternating, normalized function from $M_{n\times n}(F)$ to F.

We have derived two formula for an n-linear, alternating, normalized function from $M_{n\times n}(F)$ to F, in Proposition 5.1 and in Proposition 5.3. By Theorem 5.5 there is only one such function, the determinant, so these two formula must both be ways of computing the determinant. Thus, we have the following:

Theorem 5.7. *Let $A \in M_{n\times n}(F)$ with entries A_{ij}. Then*

$$\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i\,\sigma(i)} = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det(\widetilde{A_{ij}})$$

for any $j = 1, \ldots, n$.

Also, now that we have defined the determinant function we can state the following useful corollary to Proposition 5.1.

Corollary 5.8. *If $\delta\colon M_{n\times n}(F) \to F$ is n-linear and alternating then $\delta(A) = \det(A)\delta(I_n)$.*

6 Additional properties of the determinant

Here we prove several important properties of the determinant.

Theorem 6.1. *Let $A, B \in M_{n\times n}(F)$. Then $\det(AB) = \det(A)\det(B)$.*

Proof. Fix any $A, B \in M_{n\times n}(F)$ and define $\Phi\colon M_{n\times n}(F) \to F$ by $\Phi(C) = \det(CB)$. Then for $\alpha_1, \ldots, \alpha_n \in F^n$ we have

$$\Phi \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \det \begin{pmatrix} \alpha_1 B \\ \vdots \\ \alpha_n B \end{pmatrix} \tag{6.1}$$

where $\alpha_i B \in F^n$ denotes the matrix multiplication between α_i thought of as a $1\times n$ matrix and the $n \times n$ matrix B. Equation (6.1) immediately implies that Φ is n-linear and also we see that if two rows of C are equal then two rows of CB are equal so $\Phi(C) = \det(CB) = 0$, so Φ is alternating.

Since Φ is an n-linear, alternating function by Corollary 5.8 we know that $\Phi(C) = \det(C)\Phi(I_n)$ and $\Phi(I_n) = \det(I_n B) = \det(B)$ so we conclude for any matrix $C \in M_{n\times n}(F)$ we have $\Phi(C) = \det(C)\det(B)$. In particular, $\det(AB) = \Phi(A) = \det(A)\det(B)$. □

Remark 6.2. In the language of group theory, Theorem 6.1 states that the determinant function is a *homomorphism* from the group of matrices (with operation given by matrix multiplication) to the group F (with operation given by the field multiplication). ⊘

Theorem 6.3. *Let $A \in M_{n\times n}(F)$. Then A is invertible if and only if $\det(A) \neq 0$. Moreover, if A is invertible then $\det(A^{-1}) = (\det(A))^{-1}$.*

Proof. First suppose that A is invertible and let A^{-1} denote its inverse. Then, using Theorem 6.1, we have

$$1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$$

so $\det(A)$ is nonzero and $\det(A^{-1}) = (\det(A))^{-1}$.

Next suppose that A is not invertible. This means that the rows of A do not span all of F^n, so they must form a linearly dependent set. This means one of the rows of A, which we can assume without loss of generality is the first row, can be written as a linear combination of the other rows. Let $\alpha_1, \dots, \alpha_n \in F^n$ be the rows of A and let $a_2, \dots, a_n \in F$ be such that $\alpha_1 = \sum_{i=2}^n a_i\alpha_1$ so, keeping in mind that the determinant of any matrix with two equal rows is zero, we have

$$\det(A) = \det\begin{pmatrix} \sum_{i=2}^n a_i\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \sum_{i=2}^n a_i \det\begin{pmatrix} \alpha_i \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = 0,$$

as desired. □

Proposition 6.4. *If $A \in M_{n\times n}(F)$ is upper or lower trianglular then $\det(A)$ is the product of the diagonal entries of A.*

Proof. Suppose that A is upper triangular. Then

$$\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n A_{i\,\sigma(i)}$$

but if $\sigma \in S_n$ is not the identity then there exists some $i \in \{1, \dots, n\}$ such that $i > \sigma(i)$ in which case $A_{i\,\sigma(i)} = 0$. Thus every term in the sum except for the term corresponding to σ being the identity permutation, which we denote here by I, is zero, so we have

$$\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n A_{i\,\sigma(i)} = \operatorname{sgn}(I) \prod_{i=1}^n A_{i\,I(i)} = \prod_{i=1}^n A_{ii}$$

as desired. The case of a lower trianglar matrix is similar. □

Recall that A^t is the *transpose* of A, defined so that $(A^t)_{ij} = A_{ji}$. Notice that every element of S_n is invertible and has exactly one inverse in S_n. This means that $\{\sigma^{-1} \mid \sigma \in S_n\} = S_n$. We use this fact, and the fact that $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$, to prove the following:

Proposition 6.5. *Let* $A \in M_{n\times n}(F)$*. Then* $\det A = \det A^t$*.*

Proof. We rewrite the sum in the definition of determinant several times, reordering a sum or product when necessary:

$$\begin{aligned}
\det(A^t) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} (A^t)_{i\,\sigma(i)} \\
&= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{\sigma(i)\,i} \\
&= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i\,\sigma^{-1}(i)} \text{ (reordered the sum)} \\
&= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i\,\sigma(i)} \quad \text{(reordered the product)} \\
&= \det(A).
\end{aligned}$$

□

A corollary of Proposition 6.5 is another way to compute the determinant.

Corollary 6.6. *Let* $A \in M_{n\times n}(F)$*. Then*

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

for any $i = 1, \ldots, n$*.*

7 Final thoughts

7.1 Interpretation and uses of determinant

In this (informal) section, we give a little bit of context to the determinant by listing some applications.

There is a function λ known as the *Lebesgue measure* which takes as an input a subset of $\mathbb{R}^n$ and gives as an output the "measure" of that set[1]. That is, for any relatively nice (i.e. measurable) set $S \subset \mathbb{R}^n$, $\lambda(S)$ denotes the usual n-volume of S. Let $A \in M_{n\times n}(\mathbb{R})$. Then, as in the motivating example in Section 2, $|\det(A)| = \lambda(\mathcal{B})$ where $\mathcal{B}$ is the paralellpiped spanned by the rows of A. In fact, $\det(A)$ gives information about how A changes the volume of any set. Indeed, for any measurable set $S \subset \mathbb{R}^n$ we have that $\lambda(A(S)) = |\det(A)|\, \lambda(S)$. This is related to the use of determinant in vector calculus. You may recall that when changing coordinates to perform an integral you had to include in the integral the "Jacobian

[1]There is a technicality here that not every subset has a well-defined volume, so λ can only act on what are known as *measurable* sets. Luckily, probably every set you can imagine is measurable so this doesn't cause us too much trouble.

determinant", which is the determinant of the matrix of derivatives of the coordinate change map. This factor includes the information about how the change of coordinates map is transforming the volume of the region.

Also, given any basis $\beta = (v_1, \ldots, v_n)$ of $\mathbb{R}^n$ for the matrix A which has rows $v_1, \ldots, v_n$. Since β is linearly independent left action by A is a bijection, so $\det(A) \neq 0$. We can use the sign of $\det(A)$ to determine the *orientation* of the basis, either positive or negative.

Finally, by Theorem 6.3, to check if a matrix is invertible it suffices to check if its determinant is non-zero, which is often much easier to check (this will be useful when talking about eigenvalues).

7.2 Summary

Here we summarize the important results which are listed elsewhere in this document:

Theorem. There exists exactly one n-linear, alternating, normalized function from $M_{n\times n}(F)$ to F, which we denote by det. Let $A \in M_{n\times n}(F)$, then we may compute $\det(A)$ in several ways:

$$\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i\,\sigma(i)} = \sum_{i=1}^{n}(-1)^{i+k}A_{ik}\det(\widetilde{A_{ik}}) = \sum_{j=1}^{n}(-1)^{k+j}A_{kj}\det(\widetilde{A_{kj}})$$

for any choice of $k \in \{1, \ldots, n\}$.

Theorem. Let $A, B \in M_{n\times n}(F)$. Then:

1. $\det(AB) = \det(A)\det(B)$;
2. $\det(A^t) = \det(A)$;
3. $\det(A) \neq 0$ if and only if A is invertible;
4. if A is invertible then $\det(A^{-1}) = (\det(A))^{-1}$;
5. if B can be obtained from A by switching two rows then $\det(B) = -\det(A)$;
6. if B can be obtained from A by multiplying one row by $k \in F$ then $\det(B) = k\det(A)$.

References

[1] Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence, *Linear algebra*, third ed., Prentice Hall, Inc., Upper Saddle River, NJ, 1997. MR 1434064

[2] Kenneth Hoffman and Ray Kunze, *Linear algebra*, Second edition, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971. MR 0276251