Chapter 6

Determinants

In this chapter all vector spaces are defined over an arbitrary field K. For the sake of concreteness, the reader may safely assume that $K = \mathbb{R}$.

6.1 Permutations, Signature of a Permutation

This chapter contains a review of determinants and their use in linear algebra. We begin with permutations and the signature of a permutation. Next we define multilinear maps and alternating multilinear maps. Determinants are introduced as alternating multilinear maps taking the value 1 on the unit matrix (following Emil Artin). It is then shown how to compute a determinant using the Laplace expansion formula, and the connection with the usual definition is made. It is shown how determinants can be used to invert matrices and to solve (at least in theory!) systems of linear equations (the Cramer formulae). The determinant of a linear map is defined. We conclude by defining the characteristic polynomial of a matrix (and of a linear map) and by proving the celebrated Cayley–Hamilton theorem which states that every matrix is a "zero" of its characteristic polynomial (we give two proofs; one computational, the other one more conceptual).

Determinants can be defined in several ways. For example, determinants can be defined in a fancy way in terms of the exterior algebra (or alternating algebra) of a vector space. We will follow a more algorithmic approach due to Emil Artin. No matter which approach is followed, we need a few preliminaries about permutations on a finite set. We need to show that every permutation on n elements is a product of transpositions and that the parity of the number of transpositions involved is an invariant of the permutation. Let $[n] = \{1, 2 \ldots, n\}$, where $n \in \mathbb{N}$, and $n > 0$.

Definition 6.1. A *permutation on n elements* is a bijection $\pi \colon [n] \to [n]$. When $n = 1$, the only function from [1] to [1] is the constant map: $1 \mapsto 1$. Thus, we will assume that $n \geq 2$. A *transposition* is a permutation $\tau \colon [n] \to [n]$ such that, for some $i < j$ (with $1 \leq i < j \leq n$), $\tau(i) = j$, $\tau(j) = i$, and $\tau(k) = k$, for all $k \in [n] - \{i, j\}$. In other words, a transposition exchanges two distinct elements $i, j \in [n]$.

If τ is a transposition, clearly, $\tau \circ \tau = \mathrm{id}$. We will also use the terminology product of permutations (or transpositions) as a synonym for composition of permutations.

A permutation σ on n elements, say $\sigma(i) = k_i$ for $i = 1, \ldots, n$, can be represented in functional notation by the $2 \times n$ array

$$\begin{pmatrix} 1 & \cdots & i & \cdots & n \\ k_1 & \cdots & k_i & \cdots & k_n \end{pmatrix}$$

known as *Cauchy two-line notation*. For example, we have the permutation σ denoted by

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 3 & 6 & 5 & 1 \end{pmatrix}.$$

A more concise notation often used in computer science and in combinatorics is to represent a permutation by its image, namely by the sequence

$$\sigma(1)\ \sigma(2)\ \cdots\ \sigma(n)$$

written as a row vector without commas separating the entries. The above is known as the *one-line notation*. For example, in the one-line notation, our previous permutation σ is represented by

$$2\ 4\ 3\ 6\ 5\ 1.$$

The reason for not enclosing the above sequence within parentheses is avoid confusion with the notation for cycles, for which is it customary to include parentheses.

Clearly, the composition of two permutations is a permutation and every permutation has an inverse which is also a permutation. Therefore, the set of permutations on $[n]$ is a *group* often denoted $\mathfrak{S}_n$ and called the *symmetric group* on n elements.

It is easy to show by induction that the group $\mathfrak{S}_n$ has $n!$ elements. The following proposition shows the importance of transpositions.

Proposition 6.1. *For every $n \geq 2$, every permutation $\pi \colon [n] \to [n]$ can be written as a nonempty composition of transpositions.*

Proof. We proceed by induction on n. If $n = 2$, there are exactly two permutations on $[2]$, the transposition τ exchanging 1 and 2, and the identity. However, $\mathrm{id}_2 = \tau^2$. Now let $n \geq 3$. If $\pi(n) = n$, since by the induction hypothesis, the restriction of π to $[n-1]$ can be written as a product of transpositions, π itself can be written as a product of transpositions. If $\pi(n) = k \neq n$, letting τ be the transposition such that $\tau(n) = k$ and $\tau(k) = n$, it is clear that $\tau \circ \pi$ leaves n invariant, and by the induction hypothesis, we have $\tau \circ \pi = \tau_m \circ \ldots \circ \tau_1$ for some transpositions, and thus

$$\pi = \tau \circ \tau_m \circ \ldots \circ \tau_1,$$

a product of transpositions (since $\tau \circ \tau = \mathrm{id}_n$). □

Remark: When $\pi = \mathrm{id}_n$ is the identity permutation, we can agree that the composition of 0 transpositions is the identity. Proposition 6.1 shows that the transpositions generate the group of permutations $\mathfrak{S}_n$.

A transposition τ that exchanges two consecutive elements k and $k+1$ of $[n]$ ($1 \leq k \leq n-1$) may be called a *basic* transposition. We leave it as a simple exercise to prove that every transposition can be written as a product of basic transpositions. In fact, the transposition that exchanges k and $k+p$ ($1 \leq p \leq n-k$) can be realized using $2p-1$ basic transpositions. Therefore, the group of permutations $\mathfrak{S}_n$ is also generated by the basic transpositions.

Given a permutation written as a product of transpositions, we now show that the parity of the number of transpositions is an invariant. For this, we introduce the following function.

Definition 6.2. For every $n \geq 2$, let $\Delta\colon \mathbb{Z}^n \to \mathbb{Z}$ be the function given by

$$\Delta(x_1, \ldots, x_n) = \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

More generally, for any permutation $\sigma \in \mathfrak{S}_n$, define $\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ by

$$\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = \prod_{1 \leq i < j \leq n} (x_{\sigma(i)} - x_{\sigma(j)}).$$

The expression $\Delta(x_1, \ldots, x_n)$ is often called the *discriminant* of $(x_1, \ldots, x_n)$.

$\Delta(x_1, \ldots, x_n) \neq 0$. The discriminant consists of $\binom{n}{2}$ factors. When $n = 3$,

$$\Delta(x_1, x_2, x_3) = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3).$$

If σ is the permutation

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \end{pmatrix},$$

then

$$\begin{aligned} \Delta(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}) &= (x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)}) \\ &= (x_2 - x_3)(x_2 - x_1)(x_3 - x_1). \end{aligned}$$

Observe that

$$\Delta(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}) = (-1)^2 \Delta(x_1, x_2, x_3),$$

since two transpositions applied to the identity permutation 123 (written in one-line notation) give rise to 2 3 1. This result regarding the parity of $\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ is generalized by the following proposition.

Proposition 6.2. *For every basic transposition τ of $[n]$ ($n \geq 2$), we have*

$$\Delta(x_{\tau(1)}, \ldots, x_{\tau(n)}) = -\Delta(x_1, \ldots, x_n).$$

The above also holds for every transposition, and more generally, for every composition of transpositions $\sigma = \tau_p \circ \cdots \circ \tau_1$, we have

$$\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = (-1)^p \Delta(x_1, \ldots, x_n).$$

Consequently, for every permutation σ of $[n]$, the parity of the number p of transpositions involved in any decomposition of σ as $\sigma = \tau_p \circ \cdots \circ \tau_1$ is an invariant (only depends on σ).

Proof. Suppose τ exchanges x_k and x_{k+1}. The terms $x_i - x_j$ that are affected correspond to $i = k$, or $i = k+1$, or $j = k$, or $j = k+1$. The contribution of these terms in $\Delta(x_1, \ldots, x_n)$ is

$$\begin{aligned}(x_k - x_{k+1})[(x_k - x_{k+2}) \cdots (x_k - x_n)][(x_{k+1} - x_{k+2}) \cdots (x_{k+1} - x_n)] \\ [(x_1 - x_k) \cdots (x_{k-1} - x_k)][(x_1 - x_{k+1}) \cdots (x_{k-1} - x_{k+1})].\end{aligned}$$

When we exchange x_k and x_{k+1}, the first factor is multiplied by -1, the second and the third factor are exchanged, and the fourth and the fifth factor are exchanged, so the whole product $\Delta(x_1, \ldots, x_n)$ is is indeed multiplied by -1, that is,

$$\Delta(x_{\tau(1)}, \ldots, x_{\tau(n)}) = -\Delta(x_1, \ldots, x_n).$$

For the second statement, first we observe that since every transposition τ can be written as the composition of an odd number of basic transpositions (see the the remark following Proposition 6.1), we also have

$$\Delta(x_{\tau(1)}, \ldots, x_{\tau(n)}) = -\Delta(x_1, \ldots, x_n).$$

Next we proceed by induction on the number p of transpositions involved in the decomposition of a permutation σ.

The base case $p = 1$ has just been proven. If $p \geq 2$, if we write $\omega = \tau_{p-1} \circ \cdots \circ \tau_1$, then $\sigma = \tau_p \circ \omega$ and

$$\begin{aligned}\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) &= \Delta(x_{\tau_p(\omega(1))}, \ldots, x_{\tau_p(\omega(n))}) \\ &= -\Delta(x_{\omega(1)}, \ldots, x_{\omega(n)}) \\ &= -(-1)^{p-1} \Delta(x_1, \ldots, x_n) \\ &= (-1)^p \Delta(x_1, \ldots, x_n),\end{aligned}$$

where we used the induction hypothesis from the second to the third line, establishing the induction hypothesis. Since $\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ only depends on σ, the equation

$$\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = (-1)^p \Delta(x_1, \ldots, x_n).$$

shows that the parity $(-1)^p$ of the number of transpositions in any decomposition of σ is an invariant. □

In view of Proposition 6.2, the following definition makes sense:

Definition 6.3. For every permutation σ of $[n]$, the parity $\epsilon(\sigma)$ (or $\mathrm{sgn}(\sigma)$) of the number of transpositions involved in any decomposition of σ is called the *signature* (or *sign*) of σ.

Obviously $\epsilon(\tau) = -1$ for every transposition τ (since $(-1)^1 = -1$).

A simple way to compute the signature of a permutation is to count its number of inversions.

Definition 6.4. Given any permutation σ on n elements, we say that a pair (i, j) of indices $i, j \in \{1, \ldots, n\}$ such that $i < j$ and $\sigma(i) > \sigma(j)$ is an *inversion* of the permutation σ.

For example, the permutation σ given by

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 3 & 6 & 5 & 1 \end{pmatrix}$$

has seven inversions

$$(1,6),\ (2,3),\ (2,6),\ (3,6),\ (4,5),\ (4,6),\ (5,6).$$

Proposition 6.3. *The signature $\epsilon(\sigma)$ of any permutation σ is equal to the parity $(-1)^{I(\sigma)}$ of the number $I(\sigma)$ of inversions in σ.*

Proof. In the product

$$\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = \prod_{1 \leq i < j \leq n} (x_{\sigma(i)} - x_{\sigma(j)}),$$

the terms $x_{\sigma(i)} - x_{\sigma(j)}$ for which $\sigma(i) < \sigma(j)$ occur in $\Delta(x_1, \ldots, x_n)$, whereas the terms $x_{\sigma(i)} - x_{\sigma(j)}$ for which $\sigma(i) > \sigma(j)$ occur in $\Delta(x_1, \ldots, x_n)$ with a minus sign. Therefore, the number ν of terms in $\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ whose sign is the opposite of a term in $\Delta(x_1, \ldots, x_n)$, is equal to the number $I(\sigma)$ of inversions in σ, which implies that

$$\Delta(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = (-1)^{I(\sigma)} \Delta(x_1, \ldots, x_n).$$

By Proposition 6.2, the sign of $(-1)^{I(\sigma)}$ is equal to the signature of σ. □

For example, the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 3 & 6 & 5 & 1 \end{pmatrix}$$

has odd signature since it has seven inversions and $(-1)^7 = -1$.

Remark: When $\pi = \mathrm{id}_n$ is the identity permutation, since we agreed that the composition of 0 transpositions is the identity, it it still correct that $(-1)^0 = \epsilon(\mathrm{id}) = +1$. From Proposition 6.2, it is immediate that $\epsilon(\pi' \circ \pi) = \epsilon(\pi')\epsilon(\pi)$. In particular, since $\pi^{-1} \circ \pi = \mathrm{id}_n$, we get $\epsilon(\pi^{-1}) = \epsilon(\pi)$.

We can now proceed with the definition of determinants.

6.2 Alternating Multilinear Maps

First we define multilinear maps, symmetric multilinear maps, and alternating multilinear maps.

Remark: Most of the definitions and results presented in this section also hold when K is a commutative ring and when we consider modules over K (free modules, when bases are needed).

Let $E_1, \ldots, E_n$, and F, be vector spaces over a field K, where $n \geq 1$.

Definition 6.5. A function $f\colon E_1 \times \ldots \times E_n \to F$ is a *multilinear map (or an n-linear map)* if it is linear in each argument, holding the others fixed. More explicitly, for every i, $1 \leq i \leq n$, for all $x_1 \in E_1, \ldots, x_{i-1} \in E_{i-1}$, $x_{i+1} \in E_{i+1}, \ldots, x_n \in E_n$, for all $x, y \in E_i$, for all $\lambda \in K$,

$$\begin{aligned} f(x_1, \ldots, x_{i-1}, x + y, x_{i+1}, \ldots, x_n) &= f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) \\ &\quad + f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n), \\ f(x_1, \ldots, x_{i-1}, \lambda x, x_{i+1}, \ldots, x_n) &= \lambda f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n). \end{aligned}$$

When $F = K$, we call f an *n-linear form (or multilinear form)*. If $n \geq 2$ and $E_1 = E_2 = \ldots = E_n$, an n-linear map $f\colon E \times \ldots \times E \to F$ is called *symmetric*, if $f(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)})$ for every permutation π on $\{1, \ldots, n\}$. An n-linear map $f\colon E \times \ldots \times E \to F$ is called *alternating*, if $f(x_1, \ldots, x_n) = 0$ whenever $x_i = x_{i+1}$ for some i, $1 \leq i \leq n-1$ (in other words, when two adjacent arguments are equal). It does no harm to agree that when $n = 1$, a linear map is considered to be both symmetric and alternating, and we will do so.

When $n = 2$, a 2-linear map $f\colon E_1 \times E_2 \to F$ is called a *bilinear map*. We have already seen several examples of bilinear maps. Multiplication $\cdot\colon K \times K \to K$ is a bilinear map, treating K as a vector space over itself.

The operation $\langle -, - \rangle\colon E^* \times E \to K$ applying a linear form to a vector is a bilinear map.

Symmetric bilinear maps (and multilinear maps) play an important role in geometry (inner products, quadratic forms) and in differential calculus (partial derivatives).

A bilinear map is symmetric if $f(u, v) = f(v, u)$, for all $u, v \in E$.

Alternating multilinear maps satisfy the following simple but crucial properties.

Proposition 6.4. *Let $f\colon E \times \ldots \times E \to F$ be an n-linear alternating map, with $n \geq 2$. The following properties hold:*

(1)

$$f(\ldots, x_i, x_{i+1}, \ldots) = -f(\ldots, x_{i+1}, x_i, \ldots)$$

(2)

$$f(\ldots, x_i, \ldots, x_j, \ldots) = 0,$$

where $x_i = x_j$, and $1 \leq i < j \leq n$.

(3)

$$f(\ldots, x_i, \ldots, x_j, \ldots) = -f(\ldots, x_j, \ldots, x_i, \ldots),$$

where $1 \leq i < j \leq n$.

(4)

$$f(\ldots, x_i, \ldots) = f(\ldots, x_i + \lambda x_j, \ldots),$$

for any $\lambda \in K$, *and where* $i \neq j$.

Proof. (1) By multilinearity applied twice, we have

$$f(\ldots, x_i + x_{i+1}, x_i + x_{i+1}, \ldots) = f(\ldots, x_i, x_i, \ldots) + f(\ldots, x_i, x_{i+1}, \ldots) + f(\ldots, x_{i+1}, x_i, \ldots) + f(\ldots, x_{i+1}, x_{i+1}, \ldots),$$

and since f is alternating, this yields

$$0 = f(\ldots, x_i, x_{i+1}, \ldots) + f(\ldots, x_{i+1}, x_i, \ldots),$$

that is, $f(\ldots, x_i, x_{i+1}, \ldots) = -f(\ldots, x_{i+1}, x_i, \ldots)$.

(2) If $x_i = x_j$ and i and j are not adjacent, we can interchange x_i and x_{i+1}, and then x_i and x_{i+2}, etc, until x_i and x_j become adjacent. By (1),

$$f(\ldots, x_i, \ldots, x_j, \ldots) = \epsilon f(\ldots, x_i, x_j, \ldots),$$

where $\epsilon = +1$ or -1, but $f(\ldots, x_i, x_j, \ldots) = 0$, since $x_i = x_j$, and (2) holds.

(3) follows from (2) as in (1). (4) is an immediate consequence of (2). □

Proposition 6.4 will now be used to show a fundamental property of alternating multilinear maps. First we need to extend the matrix notation a little bit. Let E be a vector space over K. Given an $n \times n$ matrix $A = (a_{i\,j})$ over K, we can define a map $L(A)\colon E^n \to E^n$ as follows:

$$\begin{aligned} L(A)_1(u) &= a_{1\,1}u_1 + \cdots + a_{1\,n}u_n, \\ &\cdots \\ L(A)_n(u) &= a_{n\,1}u_1 + \cdots + a_{n\,n}u_n, \end{aligned}$$

for all $u_1, \ldots, u_n \in E$ and with $u = (u_1, \ldots, u_n)$. It is immediately verified that $L(A)$ is linear. Then given two $n \times n$ matrices $A = (a_{i\,j})$ and $B = (b_{i\,j})$, by repeating the calculations establishing the product of matrices (just before Definition 2.14), we can show that

$$L(AB) = L(A) \circ L(B).$$

It is then convenient to use the matrix notation to describe the effect of the linear map $L(A)$, as

$$\begin{pmatrix} L(A)_1(u) \\ L(A)_2(u) \\ \vdots \\ L(A)_n(u) \end{pmatrix} = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n\,1} & a_{n\,2} & \ldots & a_{n\,n} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Lemma 6.5. *Let $f\colon E \times \ldots \times E \to F$ be an n-linear alternating map. Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be two families of n vectors, such that,*

$$\begin{aligned} v_1 &= a_{1\,1}u_1 + \cdots + a_{n\,1}u_n, \\ &\cdots \\ v_n &= a_{1\,n}u_1 + \cdots + a_{n\,n}u_n. \end{aligned}$$

Equivalently, letting

$$A = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n\,1} & a_{n\,2} & \ldots & a_{n\,n} \end{pmatrix},$$

assume that we have

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = A^\top \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Then,

$$f(v_1, \ldots, v_n) = \Big(\sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi) a_{\pi(1)\,1} \cdots a_{\pi(n)\,n} \Big) f(u_1, \ldots, u_n),$$

where the sum ranges over all permutations π on $\{1, \ldots, n\}$.

Proof. Expanding $f(v_1, \ldots, v_n)$ by multilinearity, we get a sum of terms of the form

$$a_{\pi(1)\,1} \cdots a_{\pi(n)\,n} f(u_{\pi(1)}, \ldots, u_{\pi(n)}),$$

for all possible functions $\pi\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$. However, because f is alternating, only the terms for which π is a permutation are nonzero. By Proposition 6.1, every permutation π is a product of transpositions, and by Proposition 6.2, the parity $\epsilon(\pi)$ of the number of transpositions only depends on π. Then applying Proposition 6.4 (3) to each transposition in π, we get

$$a_{\pi(1)\,1} \cdots a_{\pi(n)\,n} f(u_{\pi(1)}, \ldots, u_{\pi(n)}) = \epsilon(\pi) a_{\pi(1)\,1} \cdots a_{\pi(n)\,n} f(u_1, \ldots, u_n).$$

Thus, we get the expression of the lemma. □

For the case of $n = 2$, the proof details of Lemma 6.5 become

$$\begin{aligned}
f(v_1, v_2) &= f(a_{11}u_1 + a_{21}u_2, a_{12}u_1 + a_{22}u_2) \\
&= f(a_{11}u_1 + a_{21}u_2, a_{12}u_1) + f(a_{11}u_1 + a_{21}u_2, a_{22}u_2) \\
&= f(a_{11}u_1, a_{12}u_1) + f(a_{21}u_2, a_{12}u_1) \\
&\quad + f(a_{11}u_a, a_{22}u_2) + f(a_{21}u_2, a_{22}u_2) \\
&= a_{11}a_{12}f(u_1, u_1) + a_{21}a_{12}f(u_2, u_1) + a_{11}a_{22}f(u_1, u_2) \\
&\quad + a_{21}a_{22}f(u_2, u_2) \\
&= a_{21}a_{12}f(u_2, u_1)a_{11}a_{22}f(u_1, u_2) \\
&= (a_{11}a_{22} - a_{12}a_{22})\, f(u_1, u_2).
\end{aligned}$$

Hopefully the reader will recognize the quantity $a_{11}a_{22} - a_{12}a_{22}$. It is the determinant of the 2×2 matrix

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

This is no accident. The quantity

$$\det(A) = \sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi) a_{\pi(1)\,1} \cdots a_{\pi(n)\,n}$$

is in fact the value of the determinant of A (which, as we shall see shortly, is also equal to the determinant of $A^\top$). However, working directly with the above definition is quite awkward, and we will proceed via a slightly indirect route

Remark: The reader might have been puzzled by the fact that it is the transpose matrix $A^\top$ rather than A itself that appears in Lemma 6.5. The reason is that if we want the generic term in the determinant to be

$$\epsilon(\pi) a_{\pi(1)\,1} \cdots a_{\pi(n)\,n},$$

where the permutation applies to the first index, then we have to express the v_js in terms of the u_is in terms of $A^\top$ as we did. Furthermore, since

$$v_j = a_{1\,j}u_1 + \cdots + a_{i\,j}u_i + \cdots + a_{n\,j}u_n,$$

we see that v_j corresponds to the jth column of the matrix A, and so the determinant is viewed as a function of the *columns* of A.

The literature is split on this point. Some authors prefer to define a determinant as we did. Others use A itself, which amounts to viewing det as a function of the rows, in which case we get the expression

$$\sum_{\sigma \in \mathfrak{S}_n} \epsilon(\sigma) a_{1\,\sigma(1)} \cdots a_{n\,\sigma(n)}.$$

Corollary 6.8 show that these two expressions are equal, so it doesn't matter which is chosen. This is a matter of taste.

6.3 Definition of a Determinant

Recall that the set of all square $n \times n$-matrices with coefficients in a field K is denoted by $\mathrm{M}_n(K)$.

Definition 6.6. A *determinant* is defined as any map

$$D\colon \mathrm{M}_n(K) \to K,$$

which, when viewed as a map on $(K^n)^n$, i.e., a map of the n columns of a matrix, is n-linear alternating and such that $D(I_n) = 1$ for the identity matrix I_n. Equivalently, we can consider a vector space E of dimension n, some fixed basis $(e_1, \ldots, e_n)$, and define

$$D\colon E^n \to K$$

as an n-linear alternating map such that $D(e_1, \ldots, e_n) = 1$.

First we will show that such maps D exist, using an inductive definition that also gives a recursive method for computing determinants. Actually, we will define a family $(\mathcal{D}_n)_{n\geq 1}$ of (finite) sets of maps $D\colon \mathrm{M}_n(K) \to K$. Second we will show that determinants are in fact uniquely defined, that is, we will show that each $\mathcal{D}_n$ consists of a *single map*. This will show the equivalence of the direct definition $\det(A)$ of Lemma 6.5 with the inductive definition $D(A)$. Finally, we will prove some basic properties of determinants, using the uniqueness theorem.

Given a matrix $A \in \mathrm{M}_n(K)$, we denote its n columns by $A^1, \ldots, A^n$. In order to describe the recursive process to define a determinant we need the notion of a minor.

Definition 6.7. Given any $n \times n$ matrix with $n \geq 2$, for any two indices i, j with $1 \leq i, j \leq n$, let A_{ij} be the $(n-1) \times (n-1)$ matrix obtained by deleting Row i and Column j from A and called a *minor*:

$$A_{ij} = \begin{pmatrix} & & & & \times & & \\ & & & & \times & & \\ \times & \times & \times & \times & \times & \times & \times \\ & & & & \times & & \\ & & & & \times & & \\ & & & & \times & & \\ & & & & \times & & \end{pmatrix}.$$

For example, if

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

then

$$A_{2\,3} = \begin{pmatrix} 2 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}.$$

Definition 6.8. For every $n \geq 1$, we define a finite set $\mathcal{D}_n$ of maps $D\colon \mathrm{M}_n(K) \to K$ inductively as follows:

When $n = 1$, $\mathcal{D}_1$ consists of the single map D such that, $D(A) = a$, where $A = (a)$, with $a \in K$.

Assume that $\mathcal{D}_{n-1}$ has been defined, where $n \geq 2$. Then $\mathcal{D}_n$ consists of all the maps D such that, for some i, $1 \leq i \leq n$,

$$D(A) = (-1)^{i+1} a_{i\,1} D(A_{i\,1}) + \cdots + (-1)^{i+n} a_{i\,n} D(A_{i\,n}),$$

where for every j, $1 \leq j \leq n$, $D(A_{i\,j})$ is the result of applying any D in $\mathcal{D}_{n-1}$ to the minor $A_{i\,j}$.

We confess that the use of the same letter D for the member of $\mathcal{D}_n$ being defined, and for members of $\mathcal{D}_{n-1}$, may be slightly confusing. We considered using subscripts to distinguish, but this seems to complicate things unnecessarily. One should not worry too much anyway, since it will turn out that each $\mathcal{D}_n$ contains just one map.

Each $(-1)^{i+j} D(A_{i\,j})$ is called the *cofactor* of $a_{i\,j}$, and the inductive expression for $D(A)$ is called a *Laplace expansion of D according to the i-th Row*. Given a matrix $A \in \mathrm{M}_n(K)$, each $D(A)$ is called a *determinant* of A.

We can think of each member of $\mathcal{D}_n$ as an *algorithm* to evaluate "the" determinant of A. The main point is that these algorithms, which recursively evaluate a determinant using all possible Laplace row expansions, all yield the *same result*, $\det(A)$.

We will prove shortly that $D(A)$ is uniquely defined (at the moment, it is not clear that $\mathcal{D}_n$ consists of a single map). Assuming this fact, given a $n \times n$-matrix $A = (a_{i\,j})$,

$$A = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \dots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \dots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n\,1} & a_{n\,2} & \dots & a_{n\,n} \end{pmatrix},$$

its determinant is denoted by $D(A)$ or $\det(A)$, or more explicitly by

$$\det(A) = \begin{vmatrix} a_{1\,1} & a_{1\,2} & \dots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \dots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n\,1} & a_{n\,2} & \dots & a_{n\,n} \end{vmatrix}.$$

Let us first consider some examples.

Example 6.1.

1. When $n = 2$, if
$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$
then by expanding according to any row, we have
$$D(A) = ad - bc.$$

2. When $n = 3$, if
$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$
then by expanding according to the first row, we have
$$D(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix},$$
that is,
$$D(A) = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}),$$
which gives the explicit formula
$$D(A) = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13}.$$

We now show that each $D \in \mathcal{D}_n$ is a determinant (map).

Lemma 6.6. *For every $n \geq 1$, for every $D \in \mathcal{D}_n$ as defined in Definition 6.8, D is an alternating multilinear map such that $D(I_n) = 1$.*

Proof. By induction on n, it is obvious that $D(I_n) = 1$. Let us now prove that D is multilinear. Let us show that D is linear in each column. Consider any Column k. Since
$$D(A) = (-1)^{i+1}a_{i1}D(A_{i1}) + \cdots + (-1)^{i+j}a_{ij}D(A_{ij}) + \cdots + (-1)^{i+n}a_{in}D(A_{in}),$$
if $j \neq k$, then by induction, $D(A_{ij})$ is linear in Column k, and a_{ij} does not belong to Column k, so $(-1)^{i+j}a_{ij}D(A_{ij})$ is linear in Column k. If $j = k$, then $D(A_{ij})$ does not depend on Column $k = j$, since A_{ij} is obtained from A by deleting Row i and Column $j = k$, and a_{ij}

belongs to Column $j = k$. Thus, $(-1)^{i+j}a_{i\,j}D(A_{i\,j})$ is linear in Column k. Consequently, in all cases, $(-1)^{i+j}a_{i\,j}D(A_{i\,j})$ is linear in Column k, and thus, $D(A)$ is linear in Column k.

Let us now prove that D is alternating. Assume that two adjacent columns of A are equal, say $A^k = A^{k+1}$. Assume that $j \neq k$ and $j \neq k+1$. Then the matrix $A_{i\,j}$ has two identical adjacent columns, and by the induction hypothesis, $D(A_{i\,j}) = 0$. The remaining terms of $D(A)$ are

$$(-1)^{i+k}a_{i\,k}D(A_{i\,k}) + (-1)^{i+k+1}a_{i\,k+1}D(A_{i\,k+1}).$$

However, the two matrices $A_{i\,k}$ and $A_{i\,k+1}$ are equal, since we are assuming that Columns k and $k+1$ of A are identical and $A_{i\,k}$ is obtained from A by deleting Row i and Column k while $A_{i\,k+1}$ is obtained from A by deleting Row i and Column $k+1$. Similarly, $a_{i\,k} = a_{i\,k+1}$, since Columns k and $k+1$ of A are equal. But then,

$$\begin{aligned}(-1)^{i+k}a_{i\,k}D(A_{i\,k}) + (-1)^{i+k+1}a_{i\,k+1}D(A_{i\,k+1})&\\ = (-1)^{i+k}a_{i\,k}D(A_{i\,k}) - (-1)^{i+k}a_{i\,k}D(A_{i\,k}) = 0.&\end{aligned}$$

This shows that D is alternating and completes the proof. □

Lemma 6.6 shows the existence of determinants. We now prove their uniqueness.

Theorem 6.7. *For every $n \geq 1$, for every $D \in \mathcal{D}_n$, for every matrix $A \in \mathrm{M}_n(K)$, we have*

$$D(A) = \sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi)a_{\pi(1)\,1} \cdots a_{\pi(n)\,n},$$

where the sum ranges over all permutations π on $\{1, \ldots, n\}$. As a consequence, $\mathcal{D}_n$ consists of a single map for every $n \geq 1$, and this map is given by the above explicit formula.

Proof. Consider the standard basis $(e_1, \ldots, e_n)$ of K^n, where $(e_i)_i = 1$ and $(e_i)_j = 0$, for $j \neq i$. Then each column A^j of A corresponds to a vector v_j whose coordinates over the basis $(e_1, \ldots, e_n)$ are the components of A^j, that is, we can write

$$\begin{aligned} v_1 &= a_{1\,1}e_1 + \cdots + a_{n\,1}e_n,\\ &\cdots\\ v_n &= a_{1\,n}e_1 + \cdots + a_{n\,n}e_n. \end{aligned}$$

Since by Lemma 6.6, each D is a multilinear alternating map, by applying Lemma 6.5, we get

$$D(A) = D(v_1, \ldots, v_n) = \Big(\sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi)a_{\pi(1)\,1} \cdots a_{\pi(n)\,n}\Big)D(e_1, \ldots, e_n),$$

where the sum ranges over all permutations π on $\{1, \ldots, n\}$. But $D(e_1, \ldots, e_n) = D(I_n)$, and by Lemma 6.6, we have $D(I_n) = 1$. Thus,

$$D(A) = \sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi)a_{\pi(1)\,1} \cdots a_{\pi(n)\,n},$$

where the sum ranges over all permutations π on $\{1, \ldots, n\}$. □

From now on we will favor the notation $\det(A)$ over $D(A)$ for the determinant of a square matrix.

Remark: There is a geometric interpretation of determinants which we find quite illuminating. Given n linearly independent vectors $(u_1, \ldots, u_n)$ in $\mathbb{R}^n$, the set

$$P_n = \{\lambda_1 u_1 + \cdots + \lambda_n u_n \mid 0 \leq \lambda_i \leq 1,\ 1 \leq i \leq n\}$$

is called a *parallelotope*. If $n = 2$, then P_2 is a *parallelogram* and if $n = 3$, then P_3 is a *parallelepiped*, a skew box having u_1, u_2, u_3 as three of its corner sides. See Figures 6.1 and 6.2.

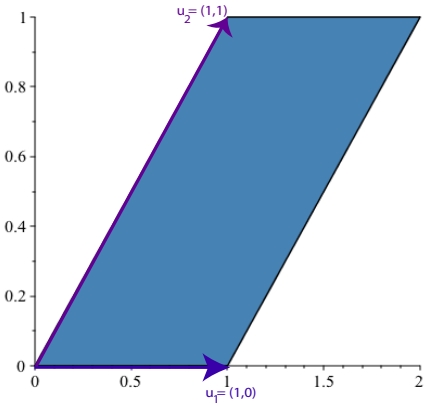

Figure 6.1: The parallelogram in $\mathbb{R}^w$ spanned by the vectors $u_1 = (1, 0)$ and $u_2 = (1, 1)$.

Then it turns out that $\det(u_1, \ldots, u_n)$ is the *signed volume* of the parallelotope P_n (where volume means n-dimensional volume). The sign of this volume accounts for the orientation of P_n in $\mathbb{R}^n$.

We can now prove some properties of determinants.

Corollary 6.8. *For every matrix $A \in \mathrm{M}_n(K)$, we have $\det(A) = \det(A^\top)$.*

Proof. By Theorem 6.7, we have

$$\det(A) = \sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi) a_{\pi(1)\,1} \cdots a_{\pi(n)\,n},$$

where the sum ranges over all permutations π on $\{1, \ldots, n\}$. Since a permutation is invertible, every product

$$a_{\pi(1)\,1} \cdots a_{\pi(n)\,n}$$

can be rewritten as

$$a_{1\,\pi^{-1}(1)} \cdots a_{n\,\pi^{-1}(n)},$$

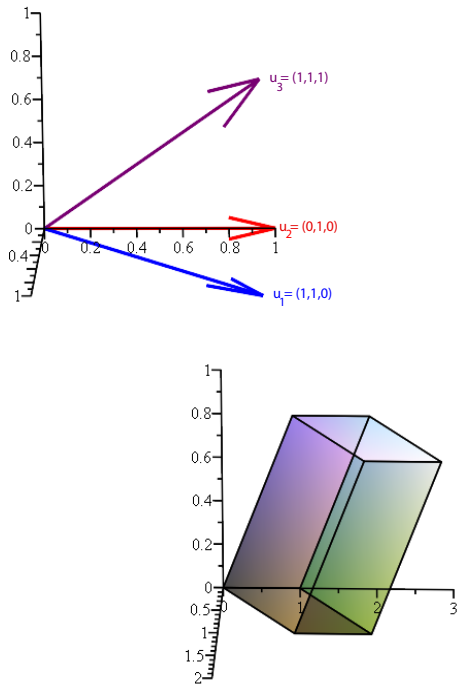

Figure 6.2: The parallelepiped in $\mathbb{R}^3$ spanned by the vectors $u_1 = (1, 1, 0)$, $u_2 = (0, 1, 0)$, and $u_3 = (0, 0, 1)$.

and since $\epsilon(\pi^{-1}) = \epsilon(\pi)$ and the sum is taken over all permutations on $\{1, \ldots, n\}$, we have

$$\sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi) a_{\pi(1)\,1} \cdots a_{\pi(n)\,n} = \sum_{\sigma \in \mathfrak{S}_n} \epsilon(\sigma) a_{1\,\sigma(1)} \cdots a_{n\,\sigma(n)},$$

where π and σ range over all permutations. But it is immediately verified that

$$\det(A^\top) = \sum_{\sigma \in \mathfrak{S}_n} \epsilon(\sigma) a_{1\,\sigma(1)} \cdots a_{n\,\sigma(n)}. \qquad \square$$

A useful consequence of Corollary 6.8 is that the determinant of a matrix is also a multilinear alternating map of its *rows*. This fact, combined with the fact that the determinant of a matrix is a multilinear alternating map of its columns, is often useful for finding short-cuts in computing determinants. We illustrate this point on the following example which shows up in polynomial interpolation.

Example 6.2. Consider the so-called *Vandermonde determinant*

$$V(x_1, \ldots, x_n) = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ x_1^2 & x_2^2 & \ldots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \ldots & x_n^{n-1} \end{vmatrix}.$$

We claim that

$$V(x_1, \ldots, x_n) = \prod_{1 \leq i < j \leq n} (x_j - x_i),$$

with $V(x_1, \ldots, x_n) = 1$, when $n = 1$. We prove it by induction on $n \geq 1$. The case $n = 1$ is obvious. Assume $n \geq 2$. We proceed as follows: multiply Row $n-1$ by x_1 and subtract it from Row n (the last row), then multiply Row $n-2$ by x_1 and subtract it from Row $n-1$, etc, multiply Row $i-1$ by x_1 and subtract it from row i, until we reach Row 1. We obtain the following determinant:

$$V(x_1, \ldots, x_n) = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ 0 & x_2 - x_1 & \ldots & x_n - x_1 \\ 0 & x_2(x_2 - x_1) & \ldots & x_n(x_n - x_1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_2^{n-2}(x_2 - x_1) & \ldots & x_n^{n-2}(x_n - x_1) \end{vmatrix}.$$

Now expanding this determinant according to the first column and using multilinearity, we can factor $(x_i - x_1)$ from the column of index $i-1$ of the matrix obtained by deleting the first row and the first column, and thus

$$V(x_1, \ldots, x_n) = (x_2 - x_1)(x_3 - x_1) \cdots (x_n - x_1) V(x_2, \ldots, x_n),$$

which establishes the induction step.

Remark: Observe that

$$\Delta(x_1, \ldots, x_n) = V(x_n, \ldots, x_1) = (-1)^{\binom{n}{2}} V(x_1, \ldots x_n),$$

where $\Delta(x_1, \ldots, x_n)$ is the discriminant of $(x_1, \ldots, x_n)$ introduced in Definition 6.2.

Lemma 6.5 can be reformulated nicely as follows.

Proposition 6.9. *Let $f \colon E \times \ldots \times E \to F$ be an n-linear alternating map. Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be two families of n vectors, such that*

$$\begin{aligned} v_1 &= a_{1\,1} u_1 + \cdots + a_{1\,n} u_n, \\ &\cdots \\ v_n &= a_{n\,1} u_1 + \cdots + a_{n\,n} u_n. \end{aligned}$$

Equivalently, letting

$$A = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n\,1} & a_{n\,2} & \ldots & a_{n\,n} \end{pmatrix},$$

assume that we have

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = A \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Then,

$$f(v_1, \ldots, v_n) = \det(A) f(u_1, \ldots, u_n).$$

Proof. The only difference with Lemma 6.5 is that here we are using $A^\top$ instead of A. Thus, by Lemma 6.5 and Corollary 6.8, we get the desired result. □

As a consequence, we get the very useful property that the determinant of a product of matrices is the product of the determinants of these matrices.

Proposition 6.10. *For any two $n \times n$-matrices A and B, we have* $\det(AB) = \det(A)\det(B)$.

Proof. We use Proposition 6.9 as follows: let $(e_1, \ldots, e_n)$ be the standard basis of K^n, and let

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = AB \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}.$$

Then we get

$$\det(w_1, \ldots, w_n) = \det(AB)\det(e_1, \ldots, e_n) = \det(AB),$$

since $\det(e_1, \ldots, e_n) = 1$. Now letting

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = B \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix},$$

we get

$$\det(v_1, \ldots, v_n) = \det(B),$$

and since

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = A \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

we get

$$\det(w_1, \ldots, w_n) = \det(A)\det(v_1, \ldots, v_n) = \det(A)\det(B).$$ □

It should be noted that all the results of this section, up to now, also hold when K is a commutative ring and not necessarily a field. We can now characterize when an $n \times n$-matrix A is invertible in terms of its determinant $\det(A)$.

6.4 Inverse Matrices and Determinants

In the next two sections, K is a commutative ring and when needed a field.

Definition 6.9. Let K be a commutative ring. Given a matrix $A \in \mathrm{M}_n(K)$, let $\widetilde{A} = (b_{ij})$ be the matrix defined such that

$$b_{ij} = (-1)^{i+j} \det(A_{ji}),$$

the cofactor of a_{ji}. The matrix $\widetilde{A}$ is called the *adjugate* of A, and each matrix A_{ji} is called a *minor* of the matrix A.

For example, if

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -2 & -2 \\ 3 & 3 & -3 \end{pmatrix},$$

we have

$$b_{11} = \det(A_{11}) = \begin{vmatrix} -2 & -2 \\ 3 & -3 \end{vmatrix} = 12 \qquad b_{12} = -\det(A_{21}) = -\begin{vmatrix} 1 & 1 \\ 3 & -3 \end{vmatrix} = 6$$

$$b_{13} = \det(A_{31}) = \begin{vmatrix} 1 & 1 \\ -2 & -2 \end{vmatrix} = 0 \qquad b_{21} = -\det(A_{12}) = -\begin{vmatrix} 2 & -2 \\ 3 & -3 \end{vmatrix} = 0$$

$$b_{22} = \det(A_{22}) = \begin{vmatrix} 1 & 1 \\ 3 & -3 \end{vmatrix} = -6 \qquad b_{23} = -\det(A_{32}) = -\begin{vmatrix} 1 & 1 \\ 2 & -2 \end{vmatrix} = 4$$

$$b_{31} = \det(A_{13}) = \begin{vmatrix} 2 & -2 \\ 3 & 3 \end{vmatrix} = 12 \qquad b_{32} = -\det(A_{23}) = -\begin{vmatrix} 1 & 1 \\ 3 & 3 \end{vmatrix} = 0$$

$$b_{33} = \det(A_{33}) = \begin{vmatrix} 1 & 1 \\ 2 & -2 \end{vmatrix} = -4,$$

we find that

$$\widetilde{A} = \begin{pmatrix} 12 & 6 & 0 \\ 0 & -6 & 4 \\ 12 & 0 & -4 \end{pmatrix}.$$

Note the reversal of the indices in

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

Thus, $\widetilde{A}$ is the *transpose* of the matrix of cofactors of elements of A.

We have the following proposition.

Proposition 6.11. *Let K be a commutative ring. For every matrix $A \in \mathrm{M}_n(K)$, we have*

$$A\widetilde{A} = \widetilde{A}A = \det(A) I_n.$$

As a consequence, A is invertible iff $\det(A)$ is invertible, and if so, $A^{-1} = (\det(A))^{-1}\widetilde{A}$.

Proof. If $\widetilde{A} = (b_{i\,j})$ and $A\widetilde{A} = (c_{i\,j})$, we know that the entry $c_{i\,j}$ in row i and column j of $A\widetilde{A}$ is

$$c_{i\,j} = a_{i\,1}b_{1\,j} + \cdots + a_{i\,k}b_{k\,j} + \cdots + a_{i\,n}b_{n\,j},$$

which is equal to

$$a_{i\,1}(-1)^{j+1}\det(A_{j\,1}) + \cdots + a_{i\,n}(-1)^{j+n}\det(A_{j\,n}).$$

If $j = i$, then we recognize the expression of the expansion of $\det(A)$ according to the i-th row:

$$c_{i\,i} = \det(A) = a_{i\,1}(-1)^{i+1}\det(A_{i\,1}) + \cdots + a_{i\,n}(-1)^{i+n}\det(A_{i\,n}).$$

If $j \neq i$, we can form the matrix A' by replacing the j-th row of A by the i-th row of A. Now the matrix $A_{j\,k}$ obtained by deleting row j and column k from A is equal to the matrix $A'_{j\,k}$ obtained by deleting row j and column k from A', since A and A' only differ by the j-th row. Thus,

$$\det(A_{j\,k}) = \det(A'_{j\,k}),$$

and we have

$$c_{i\,j} = a_{i\,1}(-1)^{j+1}\det(A'_{j\,1}) + \cdots + a_{i\,n}(-1)^{j+n}\det(A'_{j\,n}).$$

However, this is the expansion of $\det(A')$ according to the j-th row, since the j-th row of A' is equal to the i-th row of A. Furthermore, since A' has two identical rows i and j, because det is an alternating map of the rows (see an earlier remark), we have $\det(A') = 0$. Thus, we have shown that $c_{i\,i} = \det(A)$, and $c_{i\,j} = 0$, when $j \neq i$, and so

$$A\widetilde{A} = \det(A)I_n.$$

It is also obvious from the definition of $\widetilde{A}$, that

$$\widetilde{A}^{\top} = \widetilde{A^{\top}}.$$

Then applying the first part of the argument to $A^{\top}$, we have

$$A^{\top}\widetilde{A^{\top}} = \det(A^{\top})I_n,$$

and since $\det(A^{\top}) = \det(A)$, $\widetilde{A}^{\top} = \widetilde{A^{\top}}$, and $(\widetilde{A}A)^{\top} = A^{\top}\widetilde{A}^{\top}$, we get

$$\det(A)I_n = A^{\top}\widetilde{A^{\top}} = A^{\top}\widetilde{A}^{\top} = (\widetilde{A}A)^{\top},$$

that is,

$$(\widetilde{A}A)^{\top} = \det(A)I_n,$$

which yields

$$\widetilde{A}A = \det(A)I_n,$$

since $I_n^\top = I_n$. This proves that

$$A\widetilde{A} = \widetilde{A}A = \det(A)I_n.$$

As a consequence, if $\det(A)$ is invertible, we have $A^{-1} = (\det(A))^{-1}\widetilde{A}$. Conversely, if A is invertible, from $AA^{-1} = I_n$, by Proposition 6.10, we have $\det(A)\det(A^{-1}) = 1$, and $\det(A)$ is invertible. □

For example, we saw earlier that

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -2 & -2 \\ 3 & 3 & -3 \end{pmatrix} \quad \text{and} \quad \widetilde{A} = \begin{pmatrix} 12 & 6 & 0 \\ 0 & -6 & 4 \\ 12 & 0 & -4 \end{pmatrix},$$

and we have

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & -2 & -2 \\ 3 & 3 & -3 \end{pmatrix} \begin{pmatrix} 12 & 6 & 0 \\ 0 & -6 & 4 \\ 12 & 0 & -4 \end{pmatrix} = 24 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with $\det(A) = 24$.

When K is a field, an element $a \in K$ is invertible iff $a \neq 0$. In this case, the second part of the proposition can be stated as A is invertible iff $\det(A) \neq 0$. Note in passing that this method of computing the inverse of a matrix is usually not practical.

6.5 Systems of Linear Equations and Determinants

We now consider some applications of determinants to linear independence and to solving systems of linear equations. Although these results hold for matrices over certain rings, their proofs require more sophisticated methods. Therefore, we assume again that K is a field (usually, $K = \mathbb{R}$ or $K = \mathbb{C}$).

Let A be an $n \times n$-matrix, x a column vectors of variables, and b another column vector, and let $A^1, \ldots, A^n$ denote the columns of A. Observe that the system of equations $Ax = b$,

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n\,1} & a_{n\,2} & \ldots & a_{n\,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

is equivalent to

$$x_1 A^1 + \cdots + x_j A^j + \cdots + x_n A^n = b,$$

since the equation corresponding to the i-th row is in both cases

$$a_{i\,1}x_1 + \cdots + a_{i\,j}x_j + \cdots + a_{i\,n}x_n = b_i.$$

First we characterize linear independence of the column vectors of a matrix A in terms of its determinant.

Proposition 6.12. *Given an $n \times n$-matrix A over a field K, the columns $A^1, \ldots, A^n$ of A are linearly dependent iff $\det(A) = \det(A^1, \ldots, A^n) = 0$. Equivalently, A has rank n iff $\det(A) \neq 0$.*

Proof. First assume that the columns $A^1, \ldots, A^n$ of A are linearly dependent. Then there are $x_1, \ldots, x_n \in K$, such that

$$x_1 A^1 + \cdots + x_j A^j + \cdots + x_n A^n = 0,$$

where $x_j \neq 0$ for some j. If we compute

$$\det(A^1, \ldots, x_1 A^1 + \cdots + x_j A^j + \cdots + x_n A^n, \ldots, A^n) \\ = \det(A^1, \ldots, 0, \ldots, A^n) = 0,$$

where 0 occurs in the j-th position. By multilinearity, all terms containing two identical columns A^k for $k \neq j$ vanish, and we get

$$\det(A^1, \ldots, x_1 A^1 + \cdots + x_j A^j + \cdots + x_n A^n, \ldots, A^n) = x_j \det(A^1, \ldots, A^n) = 0.$$

Since $x_j \neq 0$ and K is a field, we must have $\det(A^1, \ldots, A^n) = 0$.

Conversely, we show that if the columns $A^1, \ldots, A^n$ of A are linearly independent, then $\det(A^1, \ldots, A^n) \neq 0$. If the columns $A^1, \ldots, A^n$ of A are linearly independent, then they form a basis of K^n, and we can express the standard basis $(e_1, \ldots, e_n)$ of K^n in terms of $A^1, \ldots, A^n$. Thus, we have

$$\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} b_{1\,1} & b_{1\,2} & \ldots & b_{1\,n} \\ b_{2\,1} & b_{2\,2} & \ldots & b_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n\,1} & b_{n\,2} & \ldots & b_{n\,n} \end{pmatrix} \begin{pmatrix} A^1 \\ A^2 \\ \vdots \\ A^n \end{pmatrix},$$

for some matrix $B = (b_{i\,j})$, and by Proposition 6.9, we get

$$\det(e_1, \ldots, e_n) = \det(B) \det(A^1, \ldots, A^n),$$

and since $\det(e_1, \ldots, e_n) = 1$, this implies that $\det(A^1, \ldots, A^n) \neq 0$ (and $\det(B) \neq 0$). For the second assertion, recall that the rank of a matrix is equal to the maximum number of linearly independent columns, and the conclusion is clear. □

We now characterize when a system of linear equations of the form $Ax = b$ has a unique solution.

Proposition 6.13. *Given an $n \times n$-matrix A over a field K, the following properties hold:*

(1) For every column vector b, there is a unique column vector x such that $Ax = b$ iff the only solution to $Ax = 0$ is the trivial vector $x = 0$, iff $\det(A) \neq 0$.

(2) If $\det(A) \neq 0$*, the unique solution of* $Ax = b$ *is given by the expressions*

$$x_j = \frac{\det(A^1, \ldots, A^{j-1}, b, A^{j+1}, \ldots, A^n)}{\det(A^1, \ldots, A^{j-1}, A^j, A^{j+1}, \ldots, A^n)},$$

known as Cramer's rules.

(3) The system of linear equations $Ax = 0$ *has a nonzero solution iff* $\det(A) = 0$.

Proof. (1) Assume that $Ax = b$ has a single solution x_0, and assume that $Ay = 0$ with $y \neq 0$. Then,

$$A(x_0 + y) = Ax_0 + Ay = Ax_0 + 0 = b,$$

and $x_0 + y \neq x_0$ is another solution of $Ax = b$, contradicting the hypothesis that $Ax = b$ has a single solution x_0. Thus, $Ax = 0$ only has the trivial solution. Now assume that $Ax = 0$ only has the trivial solution. This means that the columns $A^1, \ldots, A^n$ of A are linearly independent, and by Proposition 6.12, we have $\det(A) \neq 0$. Finally, if $\det(A) \neq 0$, by Proposition 6.11, this means that A is invertible, and then for every b, $Ax = b$ is equivalent to $x = A^{-1}b$, which shows that $Ax = b$ has a single solution.

(2) Assume that $Ax = b$. If we compute

$$\det(A^1, \ldots, x_1A^1 + \cdots + x_jA^j + \cdots + x_nA^n, \ldots, A^n) = \det(A^1, \ldots, b, \ldots, A^n),$$

where b occurs in the j-th position, by multilinearity, all terms containing two identical columns A^k for $k \neq j$ vanish, and we get

$$x_j \det(A^1, \ldots, A^n) = \det(A^1, \ldots, A^{j-1}, b, A^{j+1}, \ldots, A^n),$$

for every j, $1 \leq j \leq n$. Since we assumed that $\det(A) = \det(A^1, \ldots, A^n) \neq 0$, we get the desired expression.

(3) Note that $Ax = 0$ has a nonzero solution iff $A^1, \ldots, A^n$ are linearly dependent (as observed in the proof of Proposition 6.12), which, by Proposition 6.12, is equivalent to $\det(A) = 0$. □

As pleasing as Cramer's rules are, it is usually impractical to solve systems of linear equations using the above expressions. However, these formula imply an interesting fact, which is that the solution of the system $Ax = b$ are continuous in A and b. If we assume that the entries in A are continuous functions $a_{ij}(t)$ and the entries in b are are also continuous functions $b_j(t)$ of a real parameter t, since determinants are polynomial functions of their entries, the expressions

$$x_j(t) = \frac{\det(A^1, \ldots, A^{j-1}, b, A^{j+1}, \ldots, A^n)}{\det(A^1, \ldots, A^{j-1}, A^j, A^{j+1}, \ldots, A^n)}$$

are ratios of polynomials, and thus are also continuous as long as $\det(A(t))$ is nonzero. Similarly, if the functions $a_{ij}(t)$ and $b_j(t)$ are differentiable, so are the $x_j(t)$.

6.6 Determinant of a Linear Map

Given a vector space E of finite dimension n, given a basis $(u_1, \ldots, u_n)$ of E, for every linear map $f\colon E \to E$, if $M(f)$ is the matrix of f w.r.t. the basis $(u_1, \ldots, u_n)$, we can define $\det(f) = \det(M(f))$. If $(v_1, \ldots, v_n)$ is any other basis of E, and if P is the change of basis matrix, by Corollary 3.5, the matrix of f with respect to the basis $(v_1, \ldots, v_n)$ is $P^{-1}M(f)P$. By Proposition 6.10, we have

$$\det(P^{-1}M(f)P) = \det(P^{-1})\det(M(f))\det(P) = \det(P^{-1})\det(P)\det(M(f)) = \det(M(f)).$$

Thus, $\det(f)$ is indeed independent of the basis of E.

Definition 6.10. Given a vector space E of finite dimension, for any linear map $f\colon E \to E$, we define the *determinant* $\det(f)$ *of* f as the determinant $\det(M(f))$ of the matrix of f in any basis (since, from the discussion just before this definition, this determinant does not depend on the basis).

Then we have the following proposition.

Proposition 6.14. *Given any vector space E of finite dimension n, a linear map $f\colon E \to E$ is invertible iff* $\det(f) \neq 0$.

Proof. The linear map $f\colon E \to E$ is invertible iff its matrix $M(f)$ in any basis is invertible (by Proposition 3.2), iff $\det(M(f)) \neq 0$, by Proposition 6.11. □

Given a vector space of finite dimension n, it is easily seen that the set of bijective linear maps $f\colon E \to E$ such that $\det(f) = 1$ is a group under composition. This group is a subgroup of the general linear group $\mathbf{GL}(E)$. It is called the *special linear group (of E)*, and it is denoted by $\mathbf{SL}(E)$, or when $E = K^n$, by $\mathbf{SL}(n, K)$, or even by $\mathbf{SL}(n)$.

6.7 The Cayley–Hamilton Theorem

We next discuss an interesting and important application of Proposition 6.11, the *Cayley–Hamilton theorem*. The results of this section apply to matrices over any commutative ring K. First we need the concept of the characteristic polynomial of a matrix.

Definition 6.11. If K is any commutative ring, for every $n \times n$ matrix $A \in \mathrm{M}_n(K)$, the *characteristic polynomial* $P_A(X)$ of A is the determinant

$$P_A(X) = \det(XI - A).$$

The characteristic polynomial $P_A(X)$ is a polynomial in $K[X]$, the ring of polynomials in the indeterminate X with coefficients in the ring K. For example, when $n = 2$, if

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

then

$$P_A(X) = \begin{vmatrix} X-a & -b \\ -c & X-d \end{vmatrix} = X^2 - (a+d)X + ad - bc.$$

We can substitute the matrix A for the variable X in the polynomial $P_A(X)$, obtaining a *matrix* P_A. If we write

$$P_A(X) = X^n + c_1 X^{n-1} + \cdots + c_n,$$

then

$$P_A = A^n + c_1 A^{n-1} + \cdots + c_n I.$$

We have the following remarkable theorem.

Theorem 6.15. *(Cayley–Hamilton) If K is any commutative ring, for every $n \times n$ matrix $A \in \mathrm{M}_n(K)$, if we let*

$$P_A(X) = X^n + c_1 X^{n-1} + \cdots + c_n$$

be the characteristic polynomial of A, then

$$P_A = A^n + c_1 A^{n-1} + \cdots + c_n I = 0.$$

Proof. We can view the matrix $B = XI - A$ as a matrix with coefficients in the polynomial ring $K[X]$, and then we can form the matrix $\widetilde{B}$ which is the transpose of the matrix of cofactors of elements of B. Each entry in $\widetilde{B}$ is an $(n-1) \times (n-1)$ determinant, and thus a polynomial of degree a most $n-1$, so we can write $\widetilde{B}$ as

$$\widetilde{B} = X^{n-1} B_0 + X^{n-2} B_1 + \cdots + B_{n-1},$$

for some $n \times n$ matrices $B_0, \ldots, B_{n-1}$ with coefficients in K. For example, when $n = 2$, we have

$$B = \begin{pmatrix} X-a & -b \\ -c & X-d \end{pmatrix}, \quad \widetilde{B} = \begin{pmatrix} X-d & b \\ c & X-a \end{pmatrix} = X \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -d & b \\ c & -a \end{pmatrix}.$$

By Proposition 6.11, we have

$$B\widetilde{B} = \det(B) I = P_A(X) I.$$

On the other hand, we have

$$B\widetilde{B} = (XI - A)(X^{n-1} B_0 + X^{n-2} B_1 + \cdots + X^{n-j-1} B_j + \cdots + B_{n-1}),$$

and by multiplying out the right-hand side, we get

$$B\widetilde{B} = X^n D_0 + X^{n-1} D_1 + \cdots + X^{n-j} D_j + \cdots + D_n,$$

with

$$\begin{aligned} D_0 &= B_0 \\ D_1 &= B_1 - AB_0 \\ &\vdots \\ D_j &= B_j - AB_{j-1} \\ &\vdots \\ D_{n-1} &= B_{n-1} - AB_{n-2} \\ D_n &= -AB_{n-1}. \end{aligned}$$

Since

$$P_A(X)I = (X^n + c_1X^{n-1} + \cdots + c_n)I,$$

the equality

$$X^nD_0 + X^{n-1}D_1 + \cdots + D_n = (X^n + c_1X^{n-1} + \cdots + c_n)I$$

is an equality between two matrices, so it requires that all corresponding entries are equal, and since these are polynomials, the coefficients of these polynomials must be identical, which is equivalent to the set of equations

$$\begin{aligned} I &= B_0 \\ c_1I &= B_1 - AB_0 \\ &\vdots \\ c_jI &= B_j - AB_{j-1} \\ &\vdots \\ c_{n-1}I &= B_{n-1} - AB_{n-2} \\ c_nI &= -AB_{n-1}, \end{aligned}$$

for all j, with $1 \leq j \leq n-1$. If, as in the table below,

$$\begin{aligned} A^n &= A^nB_0 \\ c_1A^{n-1} &= A^{n-1}(B_1 - AB_0) \\ &\vdots \\ c_jA^{n-j} &= A^{n-j}(B_j - AB_{j-1}) \\ &\vdots \\ c_{n-1}A &= A(B_{n-1} - AB_{n-2}) \\ c_nI &= -AB_{n-1}, \end{aligned}$$

we multiply the first equation by A^n, the last by I, and generally the $(j+1)$th by A^{n-j}, when we add up all these new equations, we see that the right-hand side adds up to 0, and we get our desired equation

$$A^n + c_1 A^{n-1} + \cdots + c_n I = 0,$$

as claimed. □

As a concrete example, when $n = 2$, the matrix

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

satisfies the equation

$$A^2 - (a+d)A + (ad - bc)I = 0.$$

Most readers will probably find the proof of Theorem 6.15 rather clever but very mysterious and unmotivated. The conceptual difficulty is that we really need to understand how polynomials in one variable "act" on vectors in terms of the matrix A. This can be done and yields a more "natural" proof. Actually, the reasoning is simpler and more general if we free ourselves from matrices and instead consider a finite-dimensional vector space E and some given linear map $f\colon E \to E$. Given any polynomial $p(X) = a_0 X^n + a_1 X^{n-1} + \cdots + a_n$ with coefficients in the field K, we define the *linear map* $p(f)\colon E \to E$ by

$$p(f) = a_0 f^n + a_1 f^{n-1} + \cdots + a_n \mathrm{id},$$

where $f^k = f \circ \cdots \circ f$, the k-fold composition of f with itself. Note that

$$p(f)(u) = a_0 f^n(u) + a_1 f^{n-1}(u) + \cdots + a_n u,$$

for every vector $u \in E$. Then we define a new kind of scalar multiplication $\cdot\colon K[X] \times E \to E$ by polynomials as follows: for every polynomial $p(X) \in K[X]$, for every $u \in E$,

$$p(X) \cdot u = p(f)(u).$$

It is easy to verify that this is a "good action," which means that

$$\begin{aligned} p \cdot (u + v) &= p \cdot u + p \cdot v \\ (p + q) \cdot u &= p \cdot u + q \cdot u \\ (pq) \cdot u &= p \cdot (q \cdot u) \\ 1 \cdot u &= u, \end{aligned}$$

for all $p, q \in K[X]$ and all $u, v \in E$. With this new scalar multiplication, E is a $K[X]$-module.

If $p = \lambda$ is just a scalar in K (a polynomial of degree 0), then

$$\lambda \cdot u = (\lambda \mathrm{id})(u) = \lambda u,$$

which means that K acts on E by scalar multiplication as before. If $p(X) = X$ (the monomial X), then

$$X \cdot u = f(u).$$

Now if we pick a basis $(e_1, \ldots, e_n)$ of E, if a polynomial $p(X) \in K[X]$ has the property that

$$p(X) \cdot e_i = 0, \quad i = 1, \ldots, n,$$

then this means that $p(f)(e_i) = 0$ for $i = 1, \ldots, n$, which means that the linear map $p(f)$ vanishes on E. We can also check, as we did in Section 6.2, that if A and B are two $n \times n$ matrices and if $(u_1, \ldots, u_n)$ are any n vectors, then

$$A \cdot \left(B \cdot \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \right) = (AB) \cdot \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}.$$

This suggests the plan of attack for our second proof of the Cayley–Hamilton theorem. For simplicity, we prove the theorem for vector spaces over a field. The proof goes through for a free module over a commutative ring.

Theorem 6.16. *(Cayley–Hamilton) For every finite-dimensional vector space over a field K, for every linear map $f\colon E \to E$, for every basis $(e_1, \ldots, e_n)$, if A is the matrix over f over the basis $(e_1, \ldots, e_n)$ and if*

$$P_A(X) = X^n + c_1 X^{n-1} + \cdots + c_n$$

is the characteristic polynomial of A, then

$$P_A(f) = f^n + c_1 f^{n-1} + \cdots + c_n \mathrm{id} = 0.$$

Proof. Since the columns of A consist of the vector $f(e_j)$ expressed over the basis $(e_1, \ldots, e_n)$, we have

$$f(e_j) = \sum_{i=1}^{n} a_{i\,j} e_i, \quad 1 \leq j \leq n.$$

Using our action of $K[X]$ on E, the above equations can be expressed as

$$X \cdot e_j = \sum_{i=1}^{n} a_{i\,j} \cdot e_i, \quad 1 \leq j \leq n,$$

which yields

$$\sum_{i=1}^{j-1} -a_{i\,j} \cdot e_i + (X - a_{j\,j}) \cdot e_j + \sum_{i=j+1}^{n} -a_{i\,j} \cdot e_i = 0, \qquad 1 \leq j \leq n.$$

Observe that the transpose of the characteristic polynomial shows up, so the above system can be written as

$$\begin{pmatrix} X - a_{1\,1} & -a_{2\,1} & \cdots & -a_{n\,1} \\ -a_{1\,2} & X - a_{2\,2} & \cdots & -a_{n\,2} \\ \vdots & \vdots & \vdots & \vdots \\ -a_{1\,n} & -a_{2\,n} & \cdots & X - a_{n\,n} \end{pmatrix} \cdot \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

If we let $B = XI - A^\top$, then as in the previous proof, if $\widetilde{B}$ is the transpose of the matrix of cofactors of B, we have

$$\widetilde{B}B = \det(B)I = \det(XI - A^\top)I = \det(XI - A)I = P_A I.$$

But since

$$B \cdot \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and since $\widetilde{B}$ is matrix whose entries are polynomials in $K[X]$, it makes sense to multiply on the left by $\widetilde{B}$ and we get

$$\widetilde{B} \cdot B \cdot \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = (\widetilde{B}B) \cdot \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = P_A I \cdot \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \widetilde{B} \cdot \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix};$$

that is,

$$P_A \cdot e_j = 0, \quad j = 1, \ldots, n,$$

which proves that $P_A(f) = 0$, as claimed. □

If K is a field, then the characteristic polynomial of a linear map $f\colon E \to E$ is independent of the basis $(e_1, \ldots, e_n)$ chosen in E. To prove this, observe that the matrix of f over another basis will be of the form $P^{-1}AP$, for some inverible matrix P, and then

$$\begin{aligned} \det(XI - P^{-1}AP) &= \det(XP^{-1}IP - P^{-1}AP) \\ &= \det(P^{-1}(XI - A)P) \\ &= \det(P^{-1})\det(XI - A)\det(P) \\ &= \det(XI - A). \end{aligned}$$

Therefore, the characteristic polynomial of a linear map is intrinsic to f, and it is denoted by P_f.

The zeros (roots) of the characteristic polynomial of a linear map f are called the *eigenvalues* of f. They play an important role in theory and applications. We will come back to this topic later on.

6.8 Permanents

Recall that the explicit formula for the determinant of an $n \times n$ matrix is

$$\det(A) = \sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi) a_{\pi(1)\,1} \cdots a_{\pi(n)\,n}.$$

If we drop the sign $\epsilon(\pi)$ of every permutation from the above formula, we obtain a quantity known as the *permanent*:

$$\mathrm{per}(A) = \sum_{\pi \in \mathfrak{S}_n} a_{\pi(1)\,1} \cdots a_{\pi(n)\,n}.$$

Permanents and determinants were investigated as early as 1812 by Cauchy. It is clear from the above definition that the permanent is a multilinear symmetric form. We also have

$$\mathrm{per}(A) = \mathrm{per}(A^\top),$$

and the following unsigned version of the Laplace expansion formula:

$$\mathrm{per}(A) = a_{i\,1}\mathrm{per}(A_{i\,1}) + \cdots + a_{i\,j}\mathrm{per}(A_{i\,j}) + \cdots + a_{i\,n}\mathrm{per}(A_{i\,n}),$$

for $i = 1, \ldots, n$. However, unlike determinants which have a clear geometric interpretation as signed volumes, permanents do not have any natural geometric interpretation. Furthermore, determinants can be evaluated efficiently, for example using the conversion to row reduced echelon form, but computing the permanent is hard.

Permanents turn out to have various combinatorial interpretations. One of these is in terms of perfect matchings of bipartite graphs which we now discuss.

See Definition 18.5 for the definition of an undirected graph. A *bipartite* (undirected) graph $G = (V, E)$ is a graph whose set of nodes V can be partitioned into two nonempty disjoint subsets V_1 and V_2, such that every edge $e \in E$ has one endpoint in V_1 and one endpoint in V_2.

An example of a bipartite graph with 14 nodes is shown in Figure 6.3; its nodes are partitioned into the two sets $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ and $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$.

A *matching* in a graph $G = (V, E)$ (bipartite or not) is a set M of pairwise non-adjacent edges, which means that no two edges in M share a common vertex. A *perfect matching* is a matching such that every node in V is incident to some edge in the matching M (every node in V is an endpoint of some edge in M). Figure 6.4 shows a perfect matching (in red) in the bipartite graph G.

Obviously, a perfect matching in a bipartite graph can exist only if its set of nodes has a partition in two blocks of equal size, say $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_m\}$. Then there is a bijection between perfect matchings and bijections $\pi\colon \{x_1, \ldots, x_m\} \to \{y_1, \ldots, y_m\}$ such that $\pi(x_i) = y_j$ iff there is an edge between x_i and y_j.

Now every bipartite graph G with a partition of its nodes into two sets of equal size as above is represented by an $m \times m$ matrix $A = (a_{ij})$ such that $a_{ij} = 1$ iff there is an edge between x_i and y_j, and $a_{ij} = 0$ otherwise. Using the interpretation of perfect matchings as

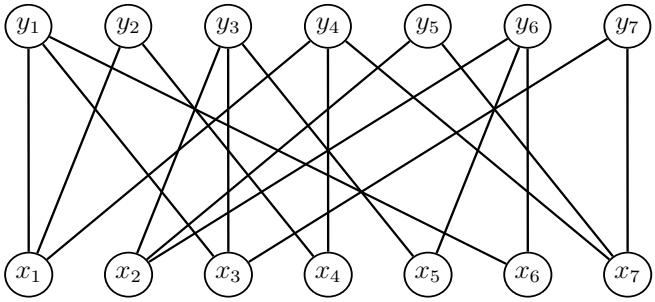

Figure 6.3: A bipartite graph G.

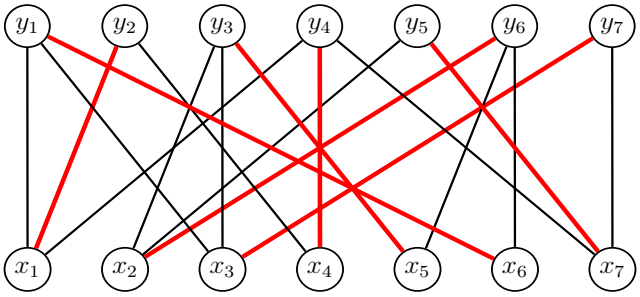

Figure 6.4: A perfect matching in the bipartite graph G.

bijections $\pi\colon \{x_1, \ldots, x_m\} \to \{y_1, \ldots, y_m\}$, we see that *the permanent* $\mathrm{per}(A)$ *of the* $(0,1)$*-matrix* A *representing the bipartite graph* G *counts the number of perfect matchings in* G.

In a famous paper published in 1979, Leslie Valiant proves that computing the permanent is a #P-complete problem. Such problems are suspected to be intractable. It is known that if a polynomial-time algorithm existed to solve a #P-complete problem, then we would have $P = NP$, which is believed to be very unlikely.

Another combinatorial interpretation of the permanent can be given in terms of systems of distinct representatives. Given a finite set S, let $(A_1, \ldots, A_n)$ be any sequence of nonempty subsets of S (not necessarily distinct). A *system of distinct representatives* (for short *SDR*) of the sets $A_1, \ldots, A_n$ is a sequence of n distinct elements $(a_1, \ldots, a_n)$, with $a_i \in A_i$ for $i = 1, \ldots, n$. The number of SDR's of a sequence of sets plays an important role in combinatorics. Now, if $S = \{1, 2, \ldots, n\}$ and if we associate to any sequence $(A_1, \ldots, A_n)$ of nonempty subsets of S the matrix $A = (a_{ij})$ defined such that $a_{ij} = 1$ if $j \in A_i$ and $a_{ij} = 0$ otherwise, then *the permanent* $\mathrm{per}(A)$ *counts the number of SDR's of the sets* $A_1, \ldots, A_n$.

This interpretation of permanents in terms of SDR's can be used to prove bounds for the permanents of various classes of matrices. Interested readers are referred to van Lint and Wilson [71] (Chapters 11 and 12). In particular, a proof of a theorem known as *Van der Waerden conjecture* is given in Chapter 12. This theorem states that for any $n \times n$ matrix A with nonnegative entries in which all row-sums and column-sums are 1 (doubly stochastic matrices), we have

$$\mathrm{per}(A) \geq \frac{n!}{n^n},$$

with equality for the matrix in which all entries are equal to $1/n$.

6.9 Summary

The main concepts and results of this chapter are listed below:

- *Permutations*, *transpositions*, *basics transpositions*.
- Every permutation can be written as a composition of permutations.
- The *parity* of the number of transpositions involved in any decomposition of a permutation σ is an invariant; it is the *signature* $\epsilon(\sigma)$ of the permutation σ.
- *Multilinear maps* (also called *n-linear maps*); *bilinear maps*.
- *Symmetric* and *alternating* multilinear maps.
- A basic property of alternating multilinear maps (Lemma 6.5) and the introduction of the formula expressing a determinant.
- Definition of a *determinant* as a multlinear alternating map $D\colon \mathrm{M}_n(K) \to K$ such that $D(I) = 1$.

- We define the set of algorithms $\mathcal{D}_n$, to compute the determinant of an $n \times n$ matrix.
- *Laplace expansion according to the ith row*; *cofactors*.
- We prove that the algorithms in $\mathcal{D}_n$ compute determinants (Lemma 6.6).
- We prove that all algorithms in $\mathcal{D}_n$ compute the same determinant (Theorem 6.7).
- We give an interpretation of determinants as *signed volumes*.
- We prove that $\det(A) = \det(A^\top)$.
- We prove that $\det(AB) = \det(A)\det(B)$.
- The *adjugate* $\widetilde{A}$ of a matrix A.
- Formula for the inverse in terms of the adjugate.
- A matrix A is invertible iff $\det(A) \neq 0$.
- Solving linear equations using *Cramer's rules*.
- Determinant of a linear map.
- The *characteristic polynomial* of a matrix.
- The *Cayley–Hamilton theorem*.
- The action of the polynomial ring induced by a linear map on a vector space.
- *Permanents*.
- Permanents count the number of perfect matchings in bipartite graphs.
- Computing the permanent is a #P-perfect problem (L. Valiant).
- Permanents count the number of SDRs of sequences of subsets of a given set.

6.10 Further Readings

Thorough expositions of the material covered in Chapter 2–5 and 6 can be found in Strang [64, 63], Lax [44], Lang [41], Artin [3], Mac Lane and Birkhoff [46], Hoffman and Kunze [38], Dummit and Foote [19], Bourbaki [8, 9], Van Der Waerden [70], Serre [57], Horn and Johnson [36], and Bertin [7]. These notions of linear algebra are nicely put to use in classical geometry, see Berger [5, 6], Tisseron [67] and Dieudonné [17].

6.11 Problems

Problem 6.1. Prove that every transposition can be written as a product of basic transpositions.

Problem 6.2. (1) Given two vectors in $\mathbb{R}^2$ of coordinates (c_1-a_1, c_2-a_2) and (b_1-a_1, b_2-a_2), prove that they are linearly dependent iff

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

(2) Given three vectors in $\mathbb{R}^3$ of coordinates $(d_1-a_1, d_2-a_2, d_3-a_3)$, $(c_1-a_1, c_2-a_2, c_3-a_3)$, and $(b_1-a_1, b_2-a_2, b_3-a_3)$, prove that they are linearly dependent iff

$$\begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ 1 & 1 & 1 & 1 \end{vmatrix} = 0.$$

Problem 6.3. Let A be the $(m+n)\times(m+n)$ block matrix (over any field K) given by

$$A = \begin{pmatrix} A_1 & A_2 \\ 0 & A_4 \end{pmatrix},$$

where A_1 is an $m \times m$ matrix, A_2 is an $m \times n$ matrix, and A_4 is an $n \times n$ matrix. Prove that $\det(A) = \det(A_1)\det(A_4)$.

Use the above result to prove that if A is an upper triangular $n \times n$ matrix, then $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

Problem 6.4. Prove that if $n \geq 3$, then

$$\det \begin{pmatrix} 1+x_1y_1 & 1+x_1y_2 & \cdots & 1+x_1y_n \\ 1+x_2y_1 & 1+x_2y_2 & \cdots & 1+x_2y_n \\ \vdots & \vdots & \vdots & \vdots \\ 1+x_ny_1 & 1+x_ny_2 & \cdots & 1+x_ny_n \end{pmatrix} = 0.$$

Problem 6.5. Prove that

$$\begin{vmatrix} 1 & 4 & 9 & 16 \\ 4 & 9 & 16 & 25 \\ 9 & 16 & 25 & 36 \\ 16 & 25 & 36 & 49 \end{vmatrix} = 0.$$

Problem 6.6. Consider the $n \times n$ symmetric matrix

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 & \dots & 0 & 0 \\ 2 & 5 & 2 & 0 & \dots & 0 & 0 \\ 0 & 2 & 5 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 2 & 5 & 2 & 0 \\ 0 & 0 & \dots & 0 & 2 & 5 & 2 \\ 0 & 0 & \dots & 0 & 0 & 2 & 5 \end{pmatrix}.$$

(1) Find an upper-triangular matrix R such that $A = R^\top R$.
(2) Prove that $\det(A) = 1$.
(3) Consider the sequence

$$\begin{aligned} p_0(\lambda) &= 1 \\ p_1(\lambda) &= 1 - \lambda \\ p_k(\lambda) &= (5 - \lambda)p_{k-1}(\lambda) - 4p_{k-2}(\lambda) \quad 2 \leq k \leq n. \end{aligned}$$

Prove that

$$\det(A - \lambda I) = p_n(\lambda).$$

Remark: It can be shown that $p_n(\lambda)$ has n distinct (real) roots and that the roots of $p_k(\lambda)$ separate the roots of $p_{k+1}(\lambda)$.

Problem 6.7. Let B be the $n \times n$ matrix ($n \geq 3$) given by

$$B = \begin{pmatrix} 1 & -1 & -1 & -1 & \cdots & -1 & -1 \\ 1 & -1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & -1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & -1 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & -1 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & -1 \end{pmatrix}.$$

Prove that

$$\det(B) = (-1)^n(n-2)2^{n-1}.$$

Problem 6.8. Given a field K (say $K = \mathbb{R}$ or $K = \mathbb{C}$), given any two polynomials $p(X), q(X) \in K[X]$, we says that $q(X)$ *divides* $p(X)$ (and that $p(X)$ *is a multiple of* $q(X)$) iff there is some polynomial $s(X) \in K[X]$ such that

$$p(X) = q(X)s(X).$$

In this case we say that $q(X)$ *is a factor of* $p(X)$, and if $q(X)$ has degree at least one, we say that $q(X)$ *is a nontrivial factor of* $p(X)$.

Let $f(X)$ and $g(X)$ be two polynomials in $K[X]$ with

$$f(X) = a_0X^m + a_1X^{m-1} + \cdots + a_m$$

of degree $m \geq 1$ and

$$g(X) = b_0X^n + b_1X^{n-1} + \cdots + b_n$$

of degree $n \geq 1$ (with $a_0, b_0 \neq 0$).

You will need the following result which you need not prove:

Two polynomials $f(X)$ *and* $g(X)$ *with* $\deg(f) = m \geq 1$ *and* $\deg(g) = n \geq 1$ *have some common nontrivial factor iff there exist two nonzero polynomials* $p(X)$ *and* $q(X)$ *such that*

$$fp = gq,$$

with $\deg(p) \leq n - 1$ *and* $\deg(q) \leq m - 1$.

(1) Let $\mathcal{P}_m$ denote the vector space of all polynomials in $K[X]$ of degree at most $m - 1$, and let $T\colon \mathcal{P}_n \times \mathcal{P}_m \to \mathcal{P}_{m+n}$ be the map given by

$$T(p, q) = fp + gq, \quad p \in \mathcal{P}_n,\ q \in \mathcal{P}_m,$$

where f and g are some fixed polynomials of degree $m \geq 1$ and $n \geq 1$.

Prove that the map T is linear.

(2) Prove that T is not injective iff f and g have a common nontrivial factor.

(3) Prove that f and g have a nontrivial common factor iff $R(f, g) = 0$, where $R(f, g)$ is the determinant given by

$$R(f,g) = \begin{vmatrix} a_0 & a_1 & \cdots & \cdots & a_m & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & a_0 & a_1 & \cdots & \cdots & a_m & 0 & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & a_0 & a_1 & \cdots & \cdots & a_m \\ b_0 & b_1 & \cdots & \cdots & \cdots & \cdots & \cdots & b_n & 0 & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & \cdots & \cdots & \cdots & \cdots & b_n & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & b_0 & b_1 & \cdots & \cdots & \cdots & \cdots & \cdots & b_n \end{vmatrix}.$$

The above determinant is called the *resultant of* f *and* g.

Note that the matrix of the resultant is an $(n + m) \times (n + m)$ matrix, with the first row (involving the a_is) occurring n times, each time shifted over to the right by one column, and the $(n + 1)$th row (involving the b_js) occurring m times, each time shifted over to the right by one column.

Hint. Express the matrix of T over some suitable basis.

(4) Compute the resultant in the following three cases:

(a) $m = n = 1$, and write $f(X) = aX + b$ and $g(X) = cX + d$.

(b) $m = 1$ and $n \geq 2$ arbitrary.

(c) $f(X) = aX^2 + bX + c$ and $g(X) = 2aX + b$.

(5) Compute the resultant of $f(X) = X^3 + pX + q$ and $g(X) = 3X^2 + p$, and

$$f(X) = a_0X^2 + a_1X + a_2$$
$$g(X) = b_0X^2 + b_1X + b_2.$$

In the second case, you should get

$$4R(f, g) = (2a_0b_2 - a_1b_1 + 2a_2b_0)^2 - (4a_0a_2 - a_1^2)(4b_0b_2 - b_1^2).$$

Problem 6.9. Let A, B, C, D be $n \times n$ real or complex matrices.

(1) Prove that if A is invertible and if $AC = CA$, then

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(AD - CB).$$

(2) Prove that if H is an $n \times n$ Hadamard matrix ($n \geq 2$), then $|\det(H)| = n^{n/2}$.

(3) Prove that if H is an $n \times n$ Hadamard matrix ($n \geq 2$), then

$$\det \begin{pmatrix} H & H \\ H & -H \end{pmatrix} = (2n)^n.$$

Problem 6.10. Compute the product of the following determinants

$$\begin{vmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{vmatrix} \begin{vmatrix} x & -y & -z & -t \\ y & x & -t & z \\ z & t & x & -y \\ t & -z & y & x \end{vmatrix}$$

to prove the following identity (due to Euler):

$$\begin{aligned}(a^2 + b^2 + c^2 + d^2)(x^2 + y^2 + z^2 + t^2) \\ = (ax + by + cz + dt)^2 + (ay - bx + ct - dz)^2 \\ + (az - bt - cx + dy)^2 + (at + bz - cy + dx)^2.\end{aligned}$$

Problem 6.11. Let A be an $n \times n$ matrix with integer entries. Prove that A^{-1} exists and has integer entries if and only if $\det(A) = \pm 1$.

Problem 6.12. Let A be an $n \times n$ real or complex matrix.

(1) Prove that if $A^\top = -A$ (A is *skew-symmetric*) and if n is odd, then $\det(A) = 0$.

(2) Prove that

$$\begin{vmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{vmatrix} = (af - be + dc)^2.$$

Problem 6.13. A *Cauchy matrix* is a matrix of the form

$$\begin{pmatrix} \dfrac{1}{\lambda_1-\sigma_1} & \dfrac{1}{\lambda_1-\sigma_2} & \cdots & \dfrac{1}{\lambda_1-\sigma_n} \\ \dfrac{1}{\lambda_2-\sigma_1} & \dfrac{1}{\lambda_2-\sigma_2} & \cdots & \dfrac{1}{\lambda_2-\sigma_n} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{1}{\lambda_n-\sigma_1} & \dfrac{1}{\lambda_n-\sigma_2} & \cdots & \dfrac{1}{\lambda_n-\sigma_n} \end{pmatrix}$$

where $\lambda_i \neq \sigma_j$, for all i, j, with $1 \leq i, j \leq n$. Prove that the determinant C_n of a Cauchy matrix as above is given by

$$C_n = \frac{\prod_{i=2}^{n}\prod_{j=1}^{i-1}(\lambda_i-\lambda_j)(\sigma_j-\sigma_i)}{\prod_{i=1}^{n}\prod_{j=1}^{n}(\lambda_i-\sigma_j)}.$$

Problem 6.14. Let $(\alpha_1, \ldots, \alpha_{m+1})$ be a sequence of pairwise distinct scalars in $\mathbb{R}$ and let $(\beta_1, \ldots, \beta_{m+1})$ be any sequence of scalars in $\mathbb{R}$, not necessarily distinct.

(1) Prove that there is a unique polynomial P of degree at most m such that

$$P(\alpha_i) = \beta_i, \quad 1 \leq i \leq m+1.$$

Hint. Remember Vandermonde!

(2) Let $L_i(X)$ be the polynomial of degree m given by

$$L_i(X) = \frac{(X-\alpha_1)\cdots(X-\alpha_{i-1})(X-\alpha_{i+1})\cdots(X-\alpha_{m+1})}{(\alpha_i-\alpha_1)\cdots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\cdots(\alpha_i-\alpha_{m+1})},$$

$$1 \leq i \leq m+1.$$

The polynomials $L_i(X)$ are known as *Lagrange polynomial interpolants*. Prove that

$$L_i(\alpha_j) = \delta_{i\,j} \quad 1 \leq i, j \leq m+1.$$

Prove that

$$P(X) = \beta_1 L_1(X) + \cdots + \beta_{m+1} L_{m+1}(X)$$

is the unique polynomial of degree at most m such that

$$P(\alpha_i) = \beta_i, \quad 1 \leq i \leq m+1.$$

(3) Prove that $L_1(X), \ldots, L_{m+1}(X)$ are linearly independent, and that they form a basis of all polynomials of degree at most m.

How is 1 (the constant polynomial 1) expressed over the basis $(L_1(X), \ldots, L_{m+1}(X))$?

Give the expression of every polynomial $P(X)$ of degree at most m over the basis $(L_1(X), \ldots, L_{m+1}(X))$.

(4) Prove that the dual basis $(L_1^*, \ldots, L_{m+1}^*)$ of the basis $(L_1(X), \ldots, L_{m+1}(X))$ consists of the linear forms L_i^* given by

$$L_i^*(P) = P(\alpha_i),$$

for every polynomial P of degree at most m; this is simply *evaluation at* α_i.