Chapter 5

Determinants

5.1 Permutations, Signature of a Permutation

We will follow an algorithmic approach due to Emil Artin. 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 5.1. 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.

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 and called the *symmetric group* on n elements.

It is easy to show by induction that the group $\mathfrak{S}_n$ has $n!$ elements.

Proposition 5.1. *For every $n \geq 2$, every permutation $\pi\colon [n] \to [n]$ can be written as a nonempty composition of transpositions.*

Remark: When $\pi = \mathrm{id}_n$ is the identity permutation, we can agree that the composition of 0 transpositions is the identity.

Proposition 5.1 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.

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.

Definition 5.2. 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

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

It is clear that if the x_i are pairwise distinct, then $\Delta(x_1, \ldots, x_n) \neq 0$.

Proposition 5.2. *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).

In view of Proposition 5.2, the following definition makes sense:

Definition 5.3. 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 $(-1)^1 = -1$).

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 5.2, it is immediate that

$$\epsilon(\pi' \circ \pi) = \epsilon(\pi')\epsilon(\pi).$$

In particular, since $\pi^{-1} \circ \pi = \mathrm{id}_n$, we get

$$\epsilon(\pi^{-1}) = \epsilon(\pi).$$

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

Definition 5.4. 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 5.3. *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.*

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 $(-1)^7 = -1$.

5.2 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.

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

Definition 5.5. 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 $\lambda \in K$,

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

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

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 not 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.

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 5.4. *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 $i \neq j$.

Proposition 5.4 will now be used to show a fundamental property of alternating multilinear maps.

First, we need to extend the matrix notation a little bit.

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 $n \times n$ matrice $A = (a_{i\,j})$ and $B = (b_{i\,j})$, by repeating the calculations establishing the product of matrices (just before Definition 1.6), 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 5.5. *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 $\{1, \ldots, n\}$.

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 ackward, 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 5.5.

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

5.3 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 5.6. 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 5.5 with the inductive definition $D(A)$.

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 5.7. 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 colummn 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 5.8. 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 $A_{i\,j}$.

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*, $\det(A)$.

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

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

Example 5.1.

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

2. When $n = 3$, if
$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$
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,
$$\begin{aligned} D(A) = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) \\ + \ a_{13}(a_{21}a_{32} - a_{31}a_{22}), \end{aligned}$$
which gives the explicit formula
$$\begin{aligned} D(A) = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} \\ - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13}. \end{aligned}$$

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

Lemma 5.6. *For every $n \geq 1$, for every $D \in \mathcal{D}_n$ as defined in Definition 5.8, D is an alternating multilinear map such that $D(I_n) = 1$.*

Lemma 5.6 shows the existence of determinants. We now prove their uniqueness.

Theorem 5.7. *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.

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.

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

A useful consequence of Corollary 5.8 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 5.2. 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 $n = 1$. This can be proved by induction on $n \geq 1$.

Example 5.3. If A is an $n \times n$ block matrix which is *block upper triangular*,

$$A = \begin{pmatrix} A_{11} & \times & \times & \cdots & \times \\ 0 & A_{22} & \times & \cdots & \times \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & A_{pp} \end{pmatrix},$$

where each A_{ii} is an $n_i \times n_i$ matrix, with $n_1 + \cdots + n_p = n$, then it can be shown by induction on $p \geq 1$ that

$$\det(A) = \det(A_{11}) \det(A_{22}) \cdots \det(A_{pp}).$$

Lemma 5.5 can be reformulated nicely as follows.

Proposition 5.9. *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*

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

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,

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

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 5.10. *For any two $n \times n$-matrices A and B, we have*

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

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 $\det(A)$.

5.4 Inverse Matrices and Determinants

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

Definition 5.9. Let K be a commutative ring. Given a matrix $A \in \mathrm{M}_n(K)$, let $\widetilde{A} = (b_{i\,j})$ be the matrix defined such that

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

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

Note the reversal of the indices in

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

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

Proposition 5.11. *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, $A^{-1} = (\det(A))^{-1}\widetilde{A}$.

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*.

We now consider some applications of determinants to linear independence and to solving systems of linear equations.

To avoid complications, 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 equation $Ax = b$,

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

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 5.12. *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* $\det(A) \neq 0$.

If we combine Proposition 5.12 with Proposition 9.12, we obtain the following criterion for finding the rank of a matrix.

Proposition 5.13. *Given any $m \times n$ matrix A over a field K (typically $K = \mathbb{R}$ or $K = \mathbb{C}$), the rank of A is the maximum natural number r such that there is an $r \times r$ submatrix B of A obtained by selecting r rows and r columns of A, and such that* $\det(B) \neq 0$.

5.5 Systems of Linear Equations and Determinants

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

Proposition 5.14. *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 $\det(A) = 0$.

As pleasing as Cramer's rules are, it is usually impractical to solve systems of linear equations using the above expressions.

5.6 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

$$\det(f) = \det(M(f)).$$

Using properties of determinants, it is not hard to show that $\det(f)$ is independent of the basis of E.

Definition 5.10. 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).

Proposition 5.15. *Given any vector space E of finite dimension n, a linear map $f\colon E \to E$ is invertible iff* $\det(f) \neq 0$.

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 $\mathbf{SL}(n)$.

5.7 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 5.11. 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 5.16. *(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

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

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 5.16 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 $(e_1, \ldots, e_n)$, 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.

This suggests the plan of attack for our second proof of the Cayley–Hamilton theorem.

For simplicity, we state the theorem for vector spaces over a field. The proof goes through for a free module over a commutative ring.

Theorem 5.17. *(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

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

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.

5.8 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 and 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.

Recall that a *bipartite* (undirected) graph $G = (V, E)$ is a graph whose set of nodes V can be partionned 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 bipatite graph with 14 nodes is shown in Figure 5.1; its nodes are partitioned into the two sets $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ and $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$.

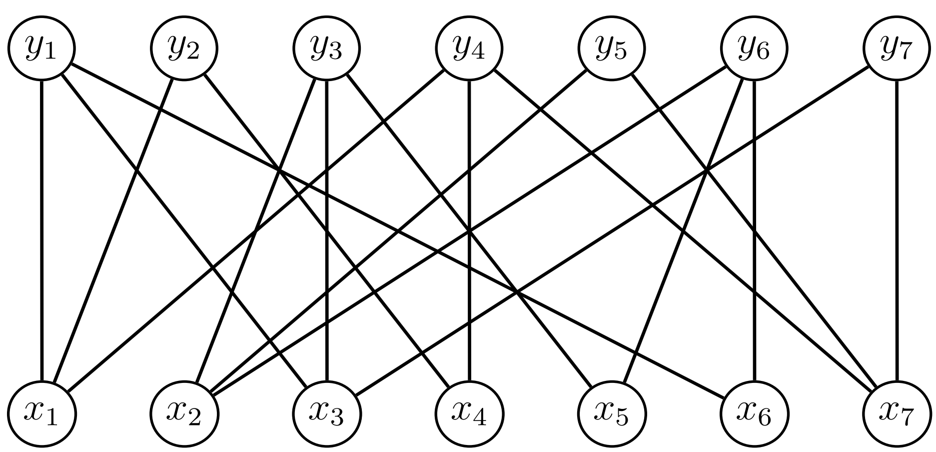

Figure 5.1: A bipartite graph G.

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 5.2 shows a perfect matching (in red) in the bipartite graph G.

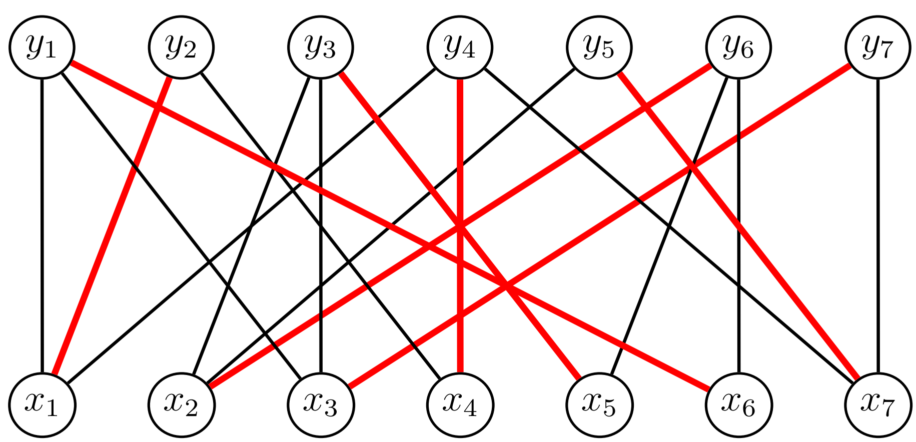

Figure 5.2: A perfect matching in the bipartite graph G.

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 machings as 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 set* $A_1, \ldots, A_n$.

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 [36] (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 $1/n$.

5.9 Further Readings

Thorough expositions of the material covered in Chapters 1–4 and 5 can be found in Strang [32, 31], Lax [25], Lang [23], Artin [1], Mac Lane and Birkhoff [26], Hoffman and Kunze [21], Bourbaki [5, 6], Van Der Waerden [35], Serre [29], Horn and Johnson [19], and Bertin [4]. These notions of linear algebra are nicely put to use in classical geometry, see Berger [2, 3], Tisseron [33] and Dieudonné [12].