Permutations, the Parity Theorem, and Determinants

John A. Gubner
Department of Electrical and Computer Engineering
University of Wisconsin–Madison

Abstract

The *Parity Theorem* says that whenever an even (resp. odd) permutation is expressed as a composition of transpositions, the number of transpositions must be even (resp. odd). The purpose of this article is to give a simple definition of when a permutation is even or odd, and develop just enough background to prove the parity theorem. Several examples are included to illustrate the use of the notation and concepts as they are introduced. We then define the determinant in terms of the parity of permutations. We establish basic properties of the determinant. In particular, we show that $\det BA = \det B \det A$, and we show that A is nonsingular if and only if $\det A \neq 0$. The characteristic polynomial is introduced and simple properties of its coefficients derived. The formula for the directional derivative of the determinant is also established.

If you find this writeup useful, or if you find typos or mistakes, please let me know at John.Gubner@wisc.edu

Contents

1. What is a Permutation?

A **permutation** is an invertible function that maps a finite set to itself.[1] If we specify an order for the elements in the finite set and apply a given permutation to each point in order, then the function values we generate simply list all the points of the set in a new order. In this way, a permutation specifies a reordering of the elements of a finite set. Without loss of generality, it suffices to take as our finite set $\{1,\ldots,n\}$ for some positive, finite integer n.

A permutation φ on $\{1,\ldots,n\}$ can be described explicitly with the $2\times n$ matrix

$$\begin{pmatrix} 1 & \cdots & n \\ \varphi 1 & \cdots & \varphi n \end{pmatrix}.$$

The top row lists the first n integers in their usual order, and the bottom row lists them in a new order.

Example 1. On $\{1,2,3,4,5\}$,

$$\varphi := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 5 & 1 & 2 \end{pmatrix}$$

is the permutation such that

$$\varphi 1 = 4, \quad \varphi 2 = 3, \quad \varphi 3 = 5, \quad \varphi 4 = 1, \quad \text{and} \quad \varphi 5 = 2.$$

Similarly,

$$\psi := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 3 & 1 & 2 \end{pmatrix}$$

is the permutation such that

$$\psi 1 = 4, \quad \psi 2 = 5, \quad \psi 3 = 3, \quad \psi 4 = 1, \quad \text{and} \quad \psi 5 = 2.$$

2. Cycles

A cycle is an especially simple kind of permutation. Given k *distinct* elements in $\{1,\ldots,n\}$, say $x_1,\ldots,x_k$, we write $\varphi = (x_1,\ldots,x_k)$ if φ takes x_1 to x_2, x_2 to x_3, $\ldots$, x_{k-1} to x_k, and x_k to x_1, while leaving all other inputs unchanged. In other words,

$$(x_1,\ldots,x_i,\ldots,x_k)x_i = x_{i+1}, \quad \text{where } x_{k+1} := x_1, \tag{1}$$

[1] To say that a function is invertible means that it is both **one-to-one** and **onto**. One-to-one means that no pair of points can map to a common destination point. Onto means that every point is the image of some point.

and
$$(x_1,\dots,x_i,\dots,x_k)x = x, \quad \text{if } x \notin \{x_1,\dots,x_k\}. \tag{2}$$
Such a permutation is called a ***k*-cycle**.

The **support** of a k-cycle $\varphi = (x_1,\dots,x_k)$ is $\operatorname{supp}\varphi := \{x_1,\dots,x_k\}$. Formula (2) says that all other values of x are **fixed points** of the cycle.

Remarks. (i) From (1), it is apparent that for $x \in \operatorname{supp}\varphi$, $\varphi^k x = x$. Of course, for $x \notin \operatorname{supp}\varphi$, (2) also implies $\varphi^k x = x$. Hence, for a k-cycle φ, we always have that φ^k is the **identity**, which we denote by $\mathscr{I}$. In particular, for $k = 2$, $(x_1,x_2)(x_1,x_2) = \mathscr{I}$, which says that a 2-cycle is its own inverse.

(ii) The cycle notation is not unique in the sense that any circular shift of the sequence $x_1,\dots,x_k$ yields the same permutation; i.e., $(x_1,\dots,x_k)$, $(x_2,\dots,x_k,x_1)$, $\dots$, and $(x_k,x_1,\dots,x_{k-1})$ all define the same permutation. In particular, note that $(x_1,x_2) = (x_2,x_1)$.

Example 2. On $\{1,2,3,4,5,6,7,8,9\}$, consider the 3-cycle $\psi := (1,2,6)$ and the 4-cycle $\varphi := (3,5,7,6)$. The following calculations illustrate how to compute with cycle notation:
$$\begin{aligned}
\varphi 7 &= (3,5,7,6)7 = 6\\
\varphi 6 &= (3,5,7,6)6 = 3\\
\psi 6 &= (1,2,6)6 = 1\\
\psi\varphi 6 &= (1,2,6)(3,5,7,6)6 = (1,2,6)3 = 3\\
\varphi\psi 6 &= (3,5,7,6)(1,2,6)6 = (3,5,7,6)1 = 1\\
\varphi 4 &= (3,5,7,6)4 = 4\\
\psi 4 &= (1,2,6)4 = 4\\
\psi\varphi 4 &= (1,2,6)(3,5,7,6)4 = (1,2,6)4 = 4\\
\varphi\psi 4 &= (3,5,7,6)(1,2,6)4 = (3,5,7,6)4 = 4.
\end{aligned}$$
In particular, notice that $\psi\varphi 6 = 3 \neq 1 = \varphi\psi 6$. So in general, cycles do not commute.

Proposition 3. *If cycles $\psi_1,\dots,\psi_m$ have pairwise disjoint supports, then*
$$\psi_1\cdots\psi_m x = \begin{cases} \psi_i x, & x \in \operatorname{supp}\psi_i,\\[1ex] x & x \notin \displaystyle\bigcup_{i=1}^m \operatorname{supp}\psi_i. \end{cases}$$
Furthermore, the ψ_i commute and can be applied in any order.

Proof. Fix a value of i in the range from 1 to m. If $x \in \operatorname{supp} \psi_i$, then $\psi_i x$ also belongs to $\operatorname{supp} \psi_i$. Therefore, x and $\psi_i x$ do *not* belong to the supports of the other cycles; i.e., x and $\psi_i x$ are fixed points of the other cycles. Hence, we can write

$$\begin{aligned}
\psi_1 \cdots \psi_m x &= \psi_1 \cdots \psi_{m-1} x, \quad \text{since } x \text{ is a fixed point of } \psi_m, \\
&= \psi_1 \cdots \psi_{m-2} x, \quad \text{since } x \text{ is a fixed point of } \psi_{m-1}, \\
&= \vdots \\
&= \psi_1 \cdots \psi_{i-1} \psi_i x \\
&= \psi_1 \cdots \psi_{i-2} \psi_i x, \quad \text{since } \psi_i x \text{ is a fixed point of } \psi_{i-1}, \\
&= \psi_1 \cdots \psi_{i-3} \psi_i x, \quad \text{since } \psi_i x \text{ is a fixed point of } \psi_{i-2}, \\
&= \vdots \\
&= \psi_1 \psi_i x \\
&= \psi_i x.
\end{aligned}$$

We can similarly argue that $\psi_m \cdots \psi_1 x = \psi_i x$. In fact, applying the cycles in any order to $x \in \operatorname{supp} \psi_i$ always results in $\psi_i x$.

Now suppose x does not belong to the support of any ψ_i. Then x is a fixed point of every ψ_i and we can write $\psi_1 \cdots \psi_m x = \psi_1 \cdots \psi_{m-1} x = \cdots = \psi_1 x = x$. Again, we can apply the cycles in any order; e.g., $\psi_m \cdots \psi_1 x = x$. □

Proposition 3 is important, because we will see later that every permutation can be decomposed into a composition of cycles with pairwise disjoint supports.

2.1. Transpositions

A **transposition** is a 2-cycle such as (x,y), where $x \neq y$. Thus, $(x,y)x = y$ and $(x,y)y = x$, while for all $z \neq x,y$, we have $(x,y)z = z$. As mentioned in the Remarks above, a transposition is its own inverse, and $(x,y) = (y,x)$.

Lemma 4. *Let ψ be a k-cycle, and let τ be a transposition whose support is a subset of the support of ψ. Then $\psi\tau$ is equal to the composition of two cycles with disjoint supports.*

Proof. Suppose $\psi = (x_1, \ldots, x_i, \ldots, x_j, \ldots, x_k)$ and $\tau = (x_i, x_j)$. It is easy to check that

$$\psi\tau = (x_1, \ldots, x_i, x_{j+1}, \ldots, x_k)(x_{i+1}, \ldots, x_j),$$

which is the composition of two cycles with disjoint supports. □

Lemma 5. *Let ψ and η be cycles with disjoint supports, and let τ be a transposition whose support intersects the supports of both ψ and η. Then $\psi\eta\tau$ is a single cycle whose support is* $\operatorname{supp}\psi \cup \operatorname{supp}\eta$.

Proof. Suppose $\psi = (x_1,\ldots,x_i,\ldots,x_k)$, $\eta = (y_1,\ldots,y_j,\ldots,y_m)$, and $\tau = (x_i,y_j)$. It is easy to check that

$$\psi\eta\tau = (x_1,\ldots,x_i,y_{j+1},\ldots,y_m,y_1,\ldots,y_j,x_{i+1},\ldots,x_k),$$

which is a single cycle whose support is $\operatorname{supp}\psi \cup \operatorname{supp}\eta$. □

3. Orbits

Let φ be a permutation on $\{1,\ldots,n\}$. Given $x \in \{1,\ldots,n\}$, consider the sequence $x, \varphi x, \varphi^2 x, \ldots, \varphi^n x$. These $n+1$ values all belong to $\{1,\ldots,n\}$. So there must be at least two values that are the same, say $\varphi^k x = \varphi^m x$ for some $0 \le k < m \le n$. Now apply φ^{-1} to both sides k times to get $x = \varphi^{m-k}x$. Hence, for any x, there is a *smallest* positive integer ℓ (depending on x) for which $\varphi^\ell x = x$. The **orbit** of x (under φ) is the set $O_x := \{x, \varphi x, \ldots, \varphi^{\ell-1}x\}$. If $\ell = 1$, the orbit is just the singleton set $\{x\}$. A singleton orbit is a fixed point.

Example 6. Consider the permutation φ of Example 1. We see that $O_1 = \{1,4\}$ and $O_2 = \{2,3,5\}$. However, $O_3 = O_2$, $O_4 = O_1$, and $O_5 = O_2$. The orbits of the permutation ψ of Example 1 are $O_1 = \{1,4\}$, $O_2 = \{2,5\}$, and $O_3 = \{3\}$, while $O_4 = O_1$ and $O_5 = O_2$.

For each $x \in \{1,\ldots,n\}$ we can determine its orbit O_x, and since each x belongs to its own orbit; i.e., $x \in O_x$, we can write

$$\{1,\ldots,n\} = \bigcup_{x=1}^{n} O_x.$$

Consider two orbits O_y and O_z for $y \ne z$. We show below that if they are not disjoint, then they are the same. If $O_y = O_z$, then the above union can be simplified to

$$\{1,\ldots,n\} = \bigcup_{\substack{x=1 \\ x\ne z}}^{n} O_x.$$

Proceeding in this way, after a finite number of steps, we obtain a sequence of distinct points $y_1,\ldots,y_{n^*}$ with

$$\{1,\ldots,n\} = \bigcup_{i=1}^{n^*} O_{y_i}, \tag{3}$$

where $n^* \le n$ and the orbits $O_{y_1}, \dots, O_{y_{n^*}}$ are pairwise disjoint. If $n - n^*$ is even, we say that the permutation φ is **even**, and we write $\operatorname{sgn}\varphi = 1$. If $n - n^*$ is odd, we say that the permutation φ is **odd**, and we write $\operatorname{sgn}\varphi = -1$. The quantity $\operatorname{sgn}\varphi$ is called the **sign**, **signature**, or **parity** of the permutation φ.

Example 7. Consider the **identity permutation**, which we denote by $\mathscr{I}$. For each x, $\mathscr{I}x = x$, and so the orbit of x under $\mathscr{I}$ is $\{x\}$. Hence, the number of disjoint orbits of $\mathscr{I}$ is $n^* = n$. Since $n - n^* = 0$ is even, the identity is an even permutation, and $\operatorname{sgn}\mathscr{I} = 1$.

Example 8. Let us determine the disjoint orbits of a k-cycle $\varphi = (x_1, \dots, x_k)$. If we start with x_1 and apply φ over and over using (1), we find that $O_{x_1} = \{x_1, \dots, x_k\} = \operatorname{supp}\varphi$. If fact, starting with x_2 or x_3, …, or x_k, we find they all have the same orbit, $\operatorname{supp}\varphi$. On the other hand if we start with an $x \notin \operatorname{supp}\varphi$, we have by (2) that $\varphi x = x$. Hence, the orbit of such an x is the singleton set $\{x\}$. We have thus shown that for a k-cycle, the number of disjoint orbits is $n^* = 1 + (n-k)$. Since this formula is equivalent to $n - n^* = k - 1$, the sign of a k-cycle is 1 if k is odd and -1 if k is even.

We now show that if two orbits are not disjoint, they are the same. Suppose $O_y \cap O_z \ne \varnothing$, and let x denote a point in $O_y \cap O_z$. Since $x \in O_y$, we must have $x = \varphi^r y$ for some nonnegative integer r. Since $x \in O_z$, we must have $x = \varphi^s z$ for some nonnegative integer s. Hence, $\varphi^r y = \varphi^s z$, which implies $y = \varphi^{s-r} z$. This further implies $\varphi^m y = \varphi^{m+s-r} z \in O_z$ for all m; hence, $O_y \subset O_z$. Similarly, writing $z = \varphi^{r-s} y$ implies $O_z \subset O_y$.

4. The Parity Theorem

It is easy to write a k-cycle as a composition of transpositions. Consider the formula

$$(x_1, \dots, x_k) = (x_1, x_k)(x_1, x_{k-1}) \cdots (x_1, x_3)(x_1, x_2). \tag{4}$$

Notice that each of $x_2, \dots, x_k$ appears in only one factor, while x_1 appears in every factor. If we apply the right-hand side to x_1, we get

$$\begin{aligned} (x_1, x_k)(x_1, x_{k-1}) \cdots (x_1, x_3)(x_1, x_2)x_1 &= (x_1, x_k)(x_1, x_{k-1}) \cdots (x_1, x_3)x_2 \\ &= x_2, \end{aligned}$$

since x_2 is not in the supports of any of the remaining transpositions. If we start with x_2, we get

$$\begin{aligned} (x_1, x_k)(x_1, x_{k-1}) \cdots (x_1, x_4)(x_1, x_3)(x_1, x_2)x_2 &= (x_1, x_k)(x_1, x_{k-1}) \cdots (x_1, x_4)(x_1, x_3)x_1 \\ &= (x_1, x_k)(x_1, x_{k-1}) \cdots (x_1, x_4)x_3 \\ &= x_3. \end{aligned}$$

Continuing in this way, we see that (4) holds.

Notice that in (4), there are $k-1$ transpositions. We saw earlier that a k-cycle φ has $n^* = 1+(n-k)$ orbits so that $k-1 = n-n^*$ determines the sign of φ. Hence, it is possible to write a k-cycle φ as a composition of transpositions such that the number of transpositions is even if $\operatorname{sgn}\varphi = 1$, and the number of transpositions is odd if $\operatorname{sgn}\varphi = -1$. However, there are many ways to write a k-cycle as compositions of different numbers of transpositions.

Example 9. On $\{1,2,3,4,5\}$, we can use (4) to write the 3-cycle $(1,2,3)$ as $(1,2,3) = (1,3)(1,2)$. Since $k=3$ is odd, $k-1=2$ is even. Hence, it is no surprise that we can write a 3-cycle as the composition of 2 (an even number) transpositions. However, since $(4,5)(4,5) = \mathscr{I}$, we can also write $(1,2,3) = (1,3)(1,2)(4,5)(4,5)$, which is another way to write this 3 cycle as an even number of transpositions.

4.1. Decomposition of Permutations into Cycles with Disjoint Supports

For an arbitrary permutation φ, once we have identified its disjoint orbits we can associate each orbit with a cycle in the following way. The disjoint orbit decomposition (3) suggests that we put

$$\psi_i x := \begin{cases} \varphi x, & x \in O_{y_i}, \\ x, & \text{otherwise.} \end{cases} \tag{5}$$

Note that the ψ_i are cycles and have disjoint supports. We can further write

$$\varphi = \psi_1 \cdots \psi_{n^*}, \tag{6}$$

which expresses φ as the composition of n^* cycles with disjoint supports. The fact that (6) holds follows from Proposition 3 and the definition (5).

Parity Theorem. *Whenever an even (resp. odd) permutation is expressed as a composition of transpositions, the number of transpositions must be even (resp. odd).*

Proof. Consider a permutation φ with n^* disjoint orbits and corresponding representation as cycles with disjoint supports as in (6). Suppose also that $\varphi = \tau_1 \cdots \tau_m$, where each τ_i is a transposition. Recalling that a transposition is its own inverse, write

$$\mathscr{I} = \varphi\varphi^{-1} = (\psi_1 \cdots \psi_{n^*})(\tau_1 \cdots \tau_m)^{-1} = (\psi_1 \cdots \psi_{n^*})(\tau_m \cdots \tau_1) = \psi_1 \cdots \psi_{n^*}\tau_m \cdots \tau_1.$$

Consider the expression $\psi_1 \cdots \psi_{n^*}\tau_m$. Since the supports of the ψ_i partition $\{1,\ldots,n\}$, the two points in the support of τ_m must belong to the supports of the ψ_i. If $\tau_m = (u,v)$, there are two cases to consider. First, there is the case that u and v both belong to the support of a single ψ_i. Second, u belongs to the support of some ψ_i, while v belongs to the support of some other ψ_j with $j > i$.[2] In the first case, we use Proposition 3 and Lemma 4 to write

$$\psi_1 \cdots \psi_{n^*} \tau_m = \psi_1 \cdots \psi_{i-1} \psi_{i+1} \cdots \psi_{n^*} (\psi_i \tau_m),$$

where $\psi_i \tau_m$ is equal to the composition of two cycles with disjoint supports created from $\operatorname{supp} \psi_i$. Hence, the above expression is a permutation with $n^* + 1$ pairwise disjoint orbits. In the second case, we use Proposition 3 and Lemma 5 to write

$$\psi_1 \cdots \psi_{n^*} \tau_m = \psi_1 \cdots \psi_{i-1} \psi_{i+1} \cdots \psi_j \psi_{j+1} \cdots \psi_{n^*} (\psi_i \psi_j \tau_m),$$

where $\psi_i \psi_j \tau_m$ is equal to a single cycle whose support is the union of the supports of ψ_i and ψ_j. Hence, the above expression is a permutation with $n^* - 1$ pairwise disjoint orbits. Since we do not know which of the two cases τ_m falls into, let us denote the new number of disjoint orbits by $n^* + \sigma_m$, where $\sigma_m = \pm 1$.

Now that we have determined the number of disjoint orbits of $\psi_1 \cdots \psi_{n^*} \tau_m$, we can determine the number of disjoint orbits of $\psi_1 \cdots \psi_{n^*} \tau_m \tau_{m-1}$ as $n^* + \sigma_m + \sigma_{m-1}$, and so on. The number of disjoint orbits of $\mathscr{I} = \psi_1 \cdots \psi_{n^*} \tau_m \cdots \tau_1$ can be written as

$$n^* + \sum_{k=1}^{m} \sigma_k,$$

where each $\sigma_k = \pm 1$. The sign of $\mathscr{I} = \psi_1 \cdots \psi_{n^*} \tau_m \cdots \tau_1$ is determined by whether

$$n - \left(n^* + \sum_{k=1}^{m} \sigma_k \right) = (n - n^*) - \sum_{k=1}^{m} \sigma_k$$

is even or odd. However, we know from Example 7 that the identity is even. This means that $(n - n^*)$ minus the above sum has to be even. Therefore, if $(n - n^*)$ is even, the above sum must be even, while if $(n - n^*)$ is odd, the above sum must be odd. Now observe that if m is even, the possible values for the above sum are $m, m-2, \ldots, 4, 2, 0, -2, -4, \ldots, -m$, which are all even, while if m is odd the possible values are $m, m-2, \ldots, 3, 1, -1, -3, \ldots, -m$, which are all odd. Hence, if $(n - n^*)$ is even, i.e., if the original permutation φ is even, then m must be even, while if φ is odd, then m must be odd. $\square$

Corollary 10. *If φ and ψ are permutations, then $\operatorname{sgn}(\varphi\psi) = \operatorname{sgn}(\varphi)\operatorname{sgn}(\psi)$, and is therefore equal to $\operatorname{sgn}(\psi\varphi)$. In particular, if ψ is itself a transposition (so that $\operatorname{sgn}(\psi) = -1$), $\operatorname{sgn}(\varphi\psi) = -\operatorname{sgn}(\varphi)$.*

[2] There is no loss of generality in assuming $j > i$ because (u, v) is equal to (v, u).

Proof. We only prove $\operatorname{sgn}(\varphi\psi) = \operatorname{sgn}(\varphi)\operatorname{sgn}(\psi)$ since the other parts of the corollary follow immediately from this. Suppose $\varphi = \tau_1 \cdots \tau_m$ and $\psi = \theta_1 \cdots \theta_k$, where the τ_i and θ_j are transpositions. By the Parity Theorem, $\operatorname{sgn}(\varphi\psi) = \pm$ according to whether $m+k$ is even or odd. Similarly for $\operatorname{sgn}(\varphi)$ and m and $\operatorname{sgn}(\psi)$ and k. It is easy to verify $\operatorname{sgn}(\varphi\psi) = \operatorname{sgn}(\varphi)\operatorname{sgn}(\psi)$ for each of the four possibilities of m and k being even/odd. $\square$

5. Determinants

If A is an $n \times n$ matrix with columns $a(1), \ldots, a(n)$, then the ith row of the column vector $a(j)$, denoted by $a_i(j)$, is the i, j entry of A. The **determinant** of A is

$$\det A := \sum_{\varphi} \operatorname{sgn}(\varphi) a_1(\varphi 1) \cdots a_n(\varphi n), \tag{7}$$

where the sum is over all possible $n!$ permutations φ of the n integers $\{1, \ldots, n\}$.

5.1. Simple Properties

If every entry in A is multiplied by a constant c, then every term in (7) will have a factor of c^n. Thus, $\det(cA) = c^n \det A$. In particular, $\det(-A) = (-1)^n \det A$.

We show that the determinant of a **diagonal matrix** is the product of its diagonal entries. To be precise, a matrix A is diagonal if $a_i(j) = 0$ whenever $j \neq i$. Consider a typical term in (7). If $\varphi \neq \mathscr{I}$, then for some i, $\varphi i \neq i$, and so the factor $a_i(\varphi i) = 0$ when A is diagonal. Hence, the only term in (7) that is not zero is the term with $\varphi = \mathscr{I}$, and that term is equal to $a_1(1) \cdots a_n(n)$ since $\operatorname{sgn}(\mathscr{I}) = 1$.

More generally, we have the same result for **triangular matrices**.

Proposition 11. *The determinant of a triangular matrix is the product of its diagonal entries.*

Proof. To begin, recall that A is upper (resp. lower) triangular if the elements below (resp. above) the main diagonal are zero. Without loss of generality, assume A is upper triangular so that $a_i(j) = 0$ for $i > j$. The result will follow if we can show that all terms in (7) with $\varphi \neq \mathscr{I}$ have zero as a factor. If $\varphi \neq \mathscr{I}$, then for some i, we must have $i > \varphi i$.[3] and for such i, since A is upper triangular, $a_i(\varphi i) = 0$. $\square$

Recall that the **transpose** of A, denoted by A^T, is defined by $a_i^\mathsf{T}(j) := a_j(i)$. Thus,

$$\det(A^\mathsf{T}) = \sum_{\varphi} \operatorname{sgn}(\varphi) a_{\varphi 1}(1) \cdots a_{\varphi n}(n),$$

[3] To see this, consider the decomposition (6). If $\varphi \neq \mathscr{I}$, some ψ_i must be a k-cycle with $k \geq 2$, say $\psi_i = (x_1, \ldots, x_k)$. These x_j are all distinct, and one is the largest, say x_l. Then $\varphi x_l = \psi_i x_l < x_l$.

Proposition 12. *The determinant of a matrix is equal to the determinant of its transpose.*

Proof. First write

$$\det A = \sum_{\varphi} \operatorname{sgn}(\varphi) \prod_{i=1}^{n} a_i(\varphi i)$$

and make the change of variable $i = \varphi^{-1} j$ to get

$$\det A = \sum_{\varphi} \operatorname{sgn}(\varphi) \prod_{j=1}^{n} a_{\varphi^{-1} j}(j).$$

Now replace φ by ψ^{-1} and note the $\operatorname{sgn}(\psi^{-1}) = \operatorname{sgn}(\psi)$. Thus,

$$\det A = \sum_{\psi} \operatorname{sgn}(\psi) \prod_{j=1}^{n} a_{\psi j}(j) = \sum_{\psi} \operatorname{sgn}(\psi) a_{\psi 1}(1) \cdots a_{\psi n}(n). \qquad \square$$

5.2. Determinant of a Product

Fix a particular column index u, and consider a typical term in (7). This term contains the factor $a_i(u)$ for some i,[4] and no other factor in that term involves an entry from the column $a(u)$. Hence, when all the columns of A are fixed except for column u, $\det A$ is linear in column u.

Let u and v be two distinct integers from $\{1, \ldots, n\}$, and let $\tau := (u, v)$ be the transposition that interchanges u and v. Let B denote the matrix with columns $b(j) := a(\tau j)$. Then

$$b(j) := a(\tau j) = \begin{cases} a(j), & j \notin \{u, v\}, \\ a(v), & j = u, \\ a(u), & j = v. \end{cases}$$

In other words, B is obtained from A by interchanging columns u and v. We claim that $\det B = -\det A$. To see this, write

$$\begin{aligned} \det B &= \sum_{\varphi} \operatorname{sgn}(\varphi) b_1(\varphi 1) \cdots b_n(\varphi n) \\ &= \sum_{\varphi} \operatorname{sgn}(\varphi) a_1(\tau\varphi 1) \cdots a_n(\tau\varphi n) \\ &= \sum_{\psi} \operatorname{sgn}(\tau^{-1}\psi) a_1(\tau\tau^{-1}\psi) \cdots a_n(\tau\tau^{-1}\psi) \\ &= \sum_{\psi} \operatorname{sgn}(\tau^{-1}\psi) a_1(\psi) \cdots a_n(\psi) \end{aligned}$$

[4] The value of i is $\varphi^{-1} u$.

$$
\begin{aligned}
&= \operatorname{sgn}(\tau^{-1}) \sum_{\psi} \operatorname{sgn}(\psi) a_1(\psi) \cdots a_n(\psi) \\
&= -\sum_{\psi} \operatorname{sgn}(\psi) a_1(\psi) \cdots a_n(\psi) \\
&= -\det A,
\end{aligned}
$$

where the third equality follows by substituting $\varphi = \tau^{-1}\psi$; the fifth equality follows because the sign of a composition of permutations is the product of their signs by Corollary 10; and the sixth equality follows because transpositions are odd.

Now suppose A has two equal columns, and B is obtained by interchanging those two columns. Then by the preceding paragraph, $\det B = -\det A$. But since $B = A$, $\det B = \det A$. Therefore, $\det A = 0$.

We showed above that if $b(j) = a(\tau j)$ for some transposition τ, then $\det B = -\det A$. What if $b(j) = a(\varphi j)$ for some arbitrary permutation φ? Writing φ in terms of transpositions, say $\varphi = \tau_1 \cdots \tau_k$, we see that $\det B = (-1)^k \det A = \operatorname{sgn}(\varphi) \det A$.

Theorem 13. $\det BA = \det B \det A$.

Proof. It is convenient to write $\det A$ in terms of its columns. We use the notation $\Delta(a(1), \ldots, a(n)) = \det A$. If B is another matrix with columns $b(1), \ldots, b(n)$, then the columns of BA are $Ba(1), \ldots, Ba(n)$, and

$$
\det BA = \Delta(Ba(1), \ldots, Ba(n)).
$$

Now recall that

$$
a(j) = \sum_{i=1}^{n} a_i(j) e(i), \quad j = 1, \ldots, n.
$$

Write

$$
\begin{aligned}
\det BA &= \Delta\left(\left[B \sum_{i=1}^{n} a_i(1) e(i) \right], Ba(2), \ldots, Ba(n) \right) \\
&= \sum_{i=1}^{n} a_i(1) \Delta(Be(i), Ba(2), \ldots, Ba(n)).
\end{aligned}
$$

Repeating this calculation for $a(2), \ldots, a(n)$ yields

$$
\begin{aligned}
\det BA &= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1}(1) \cdots a_{i_n}(n) \Delta(Be(i_1), \ldots, Be(i_n)) \\
&= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1}(1) \cdots a_{i_n}(n) \Delta(b(i_1), \ldots, b(i_n)).
\end{aligned}
$$

Among the n-tuples $(i_1,\ldots,i_n)$, if the i_k are not distinct, then $\Delta(b(i_1),\ldots,b(i_n))$ is the determinant of a matrix with two or more equal columns and is therefore zero. Otherwise, $(i_1,\ldots,i_n)=(\varphi 1,\ldots,\varphi n)$ for some permutation φ. Hence,

$$\begin{aligned}
\det BA &= \sum_\varphi a_{\varphi 1}(1)\cdots a_{\varphi n}(n)\Delta(b(\varphi 1),\ldots,b(\varphi n))\\
&= \sum_\varphi a_{\varphi 1}(1)\cdots a_{\varphi n}(n)\operatorname{sgn}(\varphi)\det B\\
&= \det B\sum_\varphi \operatorname{sgn}(\varphi)a_1(\varphi^{-1}1)\cdots a_n(\varphi^{-1}n)\\
&= \det B\sum_\psi \operatorname{sgn}(\psi^{-1})a_1(\psi 1)\cdots a_n(\psi n)\\
&= \det B\sum_\psi \operatorname{sgn}(\psi)a_1(\psi 1)\cdots a_n(\psi n)\\
&= \det B\,\det A.
\end{aligned}$$

□

Corollary 14. $\det A = 0$ *if and only if A is nonsingular.*

Proof. If A is nonsingular, then $1=\det\mathbf{I}=\det(A^{-1}A)=(\det A^{-1})(\det A)$ implies $\det A\neq 0$.

Now suppose A is singular. Then the columns of A are linearly dependent. Without loss of generality, suppose $a(n)+\sum_{j=1}^{n-1}c_j a(j)=0$ for some coefficients c_j. Hence,

$$\begin{aligned}
0 &= \Delta\big(a(1),\ldots,a(n-1),0\big)\\
&= \Delta\left(a(1),\ldots,a(n-1),a(n)+\sum_{j=1}^{n-1}c_j a(j)\right)\\
&= \Delta\big(a(1),\ldots,a(n-1),a(n)\big)+\sum_{j=1}^{n-1}c_j\Delta\big(a(1),\ldots,a(n-1),a(j)\big)\\
&= \det A + 0,
\end{aligned}$$

since in the sum over j, the jth term involves the determinant of a matrix whose column j and column n are the same. Such a determinant is zero. □

5.3. Characteristic Polynomial

The **characteristic polynomial** of a matrix A is

$$\xi(\lambda) := \det(\lambda\mathbf{I}-A),$$

where $\mathbf{I}$ denotes the identity matrix; that is the diagonal matrix with ones along the diagonal. If we put $M := \lambda\mathbf{I} - A$, then $M_i(i) = \lambda - a_i(i)$, while for $j \neq i$, we have $M_i(j) = -a_i(j)$. Thus, the term in (7) with $\varphi = \mathscr{I}$ is

$$(\lambda - a_1(1)) \cdots (\lambda - a_n(n)), \tag{8}$$

which is a polynomial in λ of degree n. We show below that for all other φ, the corresponding term in (7) is a polynomial of degree at most $n-2$. Hence, the coefficient of λ^{n-1} in the polynomial $\xi(\lambda)$ is the coefficient of λ^{n-1} in (8). The coefficient of λ^{n-1} in (8) is seen to be $-(a_1(1) + \cdots + a_n(n))$. The sum of the diagonal elements of a matrix is called the **trace**, and is denote by

$$\operatorname{tr}(A) := \sum_{i=1}^{n} a_i(i).$$

Thus,

$$\det(\lambda\mathbf{I} - A) = \lambda^n - \operatorname{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det A,$$

where the formula for the last term follows by observing that $\xi(0) = \det(-A) = (-1)^n \det A$.

For all the $\varphi \neq \mathscr{I}$ in (7), there is some i with $\varphi i \neq i$; but since φ is onto, there must be some $j \neq i$ with $\varphi j = i$. In other words, such terms in (7) must include at least two factors off the diagonal and result in a polynomial of degree at most $n-2$.

5.3.1. Differentiation of the Determinant

Since

$$\det(\mathbf{I} - tA) = t^n \xi(1/t) = 1 - \operatorname{tr}(A)t + \cdots + (-t)^n \det A,$$

we see that

$$p_A(t) := \det(\mathbf{I} + tA) = 1 + t\operatorname{tr}(A) + \cdots + t^n \det A.$$

It follows that

$$\lim_{t \to 0} \frac{p_A(t) - p_A(0)}{t} = \operatorname{tr}(A).$$

For convenience of notation, put $f(B) := \det B$. If B is invertible, then

$$f(B + tA) = \det(B + tA) = \det(B)\det(\mathbf{I} + tB^{-1}A) = \det(B)p_{B^{-1}A}(t).$$

The **directional derivative**, or **Gâteaux derivative**, of f at B in the direction A is

$$(Df)(B,A) := \lim_{t \to 0} \frac{f(B + tA) - f(B)}{t} = \det(B)\operatorname{tr}(B^{-1}A).$$

With a bit more work, we can show that the **Fréchet derivative** of f at B applied to A is equal to $\det(B)\,\mathrm{tr}(B^{-1}A)$. To prove this, it again suffices to treat the case $B = \mathbf{I}$. We must show that given $\varepsilon > 0$, for sufficiently small $\|A\|$, we have

$$|f(\mathbf{I}+A) - f(\mathbf{I}) - \mathrm{tr}(A)| \leq \varepsilon \|A\|. \tag{9}$$

Without loss of generality, we use the infinity norm on A; i.e., $\|A\| = \max_{i,j} |A_{ij}|$. Since

$$\begin{aligned} f(\mathbf{I}+A) &= (1+A_{1,1})\cdots(1+A_{n,n}) + \sum_{\sigma\neq\mathscr{I}} \mathrm{sgn}(\sigma)(\mathbf{I}+A)_1(\sigma 1)\cdots(\mathbf{I}+A)_n(\sigma n) \\ &= 1 + \mathrm{tr}(A) + \text{other terms}, \end{aligned}$$

where every term in "other terms" includes at least two factors of the form A_{ij} with $i \neq j$. Hence,

$$\begin{aligned} |f(\mathbf{I}+A) - f(\mathbf{I}) - \mathrm{tr}(A)| &\leq K_2\|A\|^2 + \cdots + K_n\|A\|^n + \sum_{\sigma\neq\mathscr{I}} (1+\|A\|)^{n-2}\|A\|^2 \\ &= \left[K_2\|A\| + \cdots + K_n\|A\|^{n-1} + \sum_{\sigma\neq\mathscr{I}} (1+\|A\|)^{n-2}\|A\|\right]\|A\|. \end{aligned}$$

For sufficiently small $\|A\|$, the sum in brackets is less than ε, and so (9) follows.

References

[1] T. L. Barlow, "An historical note on the parity of permutations," *The American Mathematical Monthly*, vol. 79, no. 7, pp. 766–769, 1972.

[2] W. Phillips, "On the definition of even and odd permutations," *The American Mathematical Monthly*, vol. 74, no. 10, pp. 1249–1251, 1967.

[3] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.

Index