CHAPTER 1

The Theory of Permanents in the Historical Order of Development

1.1 Introduction

Modern mathematicians have a proclivity to invent flippant names for newly introduced mathematical entities and concepts. They delight in talking about mobs, radicals, derogatory matrices, osculating planes, improper ideals, etc. It may appear that the term "permanent" was also invented by a waggish algebraist. In fact, a few years ago a well-meaning referee admonished the author for daring to invent this ludicrous name for a function that Schur himself introduced without designating it by any specific term. The fact of the matter is that the permanent function was studied and called by that name before Schur was even born.

In his famous memoir of 1812, Cauchy [2] developed the theory of determinants as a special type of alternating symmetric functions, which he distinguished from the ordinary symmetric functions by calling the latter "fonctions symétriques permanentes." He also introduced a certain subclass of symmetric functions, which were later named *permanents* by Muir [14] and which are nowadays known by this name. These functions can be defined by means of matrices and modern notation as follows.

Let $A=(a_{ij})$ be an $m\times n$ matrix over any commutative ring, $m\leqslant n$. The *permanent* of A, written Per(A), or simply Per A, is defined by

$$\operatorname{Per}(A)=\sum_{\sigma} a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{m\sigma(m)}, \tag{1.1}$$

where the summation extends over all one-to-one functions from $\{1,\ldots,m\}$ to $\{1,\ldots,n\}$. The sequence $(a_{1\sigma(1)},\ldots,a_{m\sigma(m)})$ is called a *diagonal* of A, and the product $a_{1\sigma(1)}\cdots a_{m\sigma(m)}$ is a *diagonal product* of A. Thus the permanent of A is the sum of all diagonal products of A.

ENCYCLOPEDIA OF MATHEMATICS and Its Applications, Gian–Carlo Rota (ed.).
Vol. 6: Henryk Minc, Permanents

For example, if

$$A=[3\,2\,4],$$

$$B=\begin{bmatrix}3 & 2 & 4\\ 2 & 1 & 5\end{bmatrix},$$

$$C=\begin{bmatrix}3 & 2 & 4\\ 2 & 1 & 5\\ -1 & 2 & -2\end{bmatrix},$$

then Per $A=9$, Per $B=44$, and Per $C=18$.

The special case $m=n$ is of particular importance. We denote the permanent of a square matrix A by per(A) instead of Per(A). In fact, most writers restrict the designation "permanent" to the case of square matrices.

1.2 The Originators: Binet and Cauchy

Permanents were introduced in 1812 almost simultaneously by Binet [1] and Cauchy [2]. Binet in his memoir also gave formulas for computing the permanents of $m\times n$ matrices for $m\leqslant 4$.

The permanent of an $m\times n$ matrix A, $m\leqslant n$, is the sum of all the diagonal products of A. In other words, Per A is the sum of all products of m elements of A, no two in the same row or the same column. It follows that all the terms of Per A, and many other superfluous terms, are contained in the set of terms obtained by multiplying the row sums of A. For example, if A is a $2\times n$ matrix, then

$$\operatorname{Per} A=\sum_{s\neq t} a_{1s}a_{2t},$$

while the product of the row sums of A is

$$\prod_{i=1}^{2}\sum_{j=1}^{n} a_{ij}=\sum_{s,t=1}^{n} a_{1s}a_{2t}$$

$$=\sum_{s\neq t} a_{1s}a_{2t}+\sum_{s=1}^{n} a_{1s}a_{2s}.$$

Hence

$$\operatorname{Per} A=\prod_{i=1}^{2}\sum_{j=1}^{n} a_{ij}-\sum_{s=1}^{n} a_{1s}a_{2s}, \tag{2.1}$$

which is Binet's formula for $m=2$.

Let $Q_{t,k}$ denote the set of all strictly increasing sequences of integers $\omega=(\omega_1,\dots,\omega_t)$ satisfying $1\leqslant\omega_1<\omega_2<\cdots<\omega_t\leqslant k$. For an $m\times n$ matrix $A=(a_{ij})$ and a sequence $(i_1,\dots,i_s)\in Q_{s,m}$, define

$$r_{i_1*\cdots*i_s}=\sum_{j=1}^{n}a_{i_1j}a_{i_2j}\cdots a_{i_sj}.$$

In particular,

$$r_i=\sum_{j=1}^{n}a_{ij}$$

denotes the ith row sum of A. Then Binet's formula (2.1) can be written in the form

$$\operatorname{Per}A=r_1r_2-r_{1*2}.$$

Now consider a $3\times n$ matrix $A=(a_{ij})$, $n\geqslant 3$. The product $r_1r_2r_3$ contains all the terms of $\operatorname{Per}A$ and in addition $n^3-n(n-1)(n-2)$ "unwanted" terms such as $a_{11}a_{21}a_{3j},a_{12}a_{22}a_{3j},\dots,a_{11}a_{2j}a_{31},\dots,a_{1j}a_{21}a_{31},\dots,$etc., $j=1,\dots,n$—that is, the terms of $r_{1*2}r_3$, $r_{1*3}r_2$, and $r_{2*3}r_1$. It seems, therefore, that if we subtract $r_{1*2}r_3+r_{1*3}r_2+r_{2*3}r_1$ from $r_1r_2r_3$, we should be left with the terms of $\operatorname{Per}A$. Unfortunately, this is not the case. What happens is that, although we subtract all the "unwanted" terms, we subtract some of them more than once. To be precise, the terms $a_{1j}a_{2j}a_{3j},j=1,\dots,n$, appear in all three products $r_{1*2}r_3$, $r_{1*3}r_2$, and $r_{2*3}r_1$, and therefore each of them is subtracted three times instead of once. But r_{1*2*3} is the sum of all the $a_{1j}a_{2j}a_{3j}$. Thus we obtain Binet's second formula:

$$\operatorname{Per}A=r_1r_2r_3-(r_{1*2}r_3+r_{1*3}r_2+r_{2*3}r_1)+2r_{1*2*3}. \tag{2.2}$$

We now introduce the following simplifying notation: For $2\times n$ matrices,

$$S(1,1)=r_1r_2,$$
$$S(2)=r_{1*2};$$

for $3\times n$ matrices,

$$S(1,1,1)=r_1r_2r_3,$$
$$S(1,2)=r_1r_{2*3}+r_2r_{1*3}+r_3r_{1*2},$$
$$S(3)=r_{1*2*3};$$

for $4\times n$ matrices,

$$S(1,1,1,1)=r_1r_2r_3r_4,$$

$$S(1,1,2)=r_1r_2r_{3*4}+r_1r_3r_{2*4}+r_1r_4r_{2*3}+r_2r_3r_{1*4}+r_2r_4r_{1*3}+r_3r_4r_{1*2},$$

$$S(2,2)=r_{1*2}r_{3*4}+r_{1*3}r_{2*4}+r_{1*4}r_{2*3},$$

$$S(1,3)=r_1r_{2*3*4}+r_2r_{1*3*4}+r_3r_{1*2*4}+r_4r_{1*2*3},$$

$$S(4)=r_{1*2*3*4},$$

etc. In general, if A is an $m\times n$ matrix and $t_1,\ldots,t_k$ are integers, $1\leqslant t_1\leqslant\cdots\leqslant t_k, t_1+\cdots+t_k=m$, then $S(t_1,\ldots,t_k)$ is the symmetrized sum of all distinct products of the $r_{i_1*\cdots*i_s}, s=t_1,\ldots,t_k$, so that in each product the sequences $(i_1,\ldots,i_s)\in Q_{s,m}, s=t_1,\ldots,t_k$, partition the set $\{1,\ldots,m\}$.

Using this notation we can write equations (2.1) and (2.2) in the following form:

$$\operatorname{Per}A=S(1,1)-S(2), \qquad (2.1')$$

$$\operatorname{Per}A=S(1,1,1)-S(1,2)+2S(3). \qquad (2.2')$$

Both these formulas were proved by Binet by a very involved method. He also gave, without proof, a formula for the permanent of a $4\times n$ matrix A:

$$\operatorname{Per}A=S(1,1,1,1)-S(1,1,2)+S(2,2)+2S(1,3)-6S(4). \qquad (2.3)$$

This formula can be established by the use of the principle of inclusion and exclusion which we used to prove formulas (2.1) and (2.2).

Let A be an $4\times n$ matrix. The function $S(1,1,1,1)$ is the sum of all the terms of $\operatorname{Per}A$ and also some "superfluous" terms all of which are the terms of $S(1,1,2)$. We therefore subtract $S(1,1,2)$ from $S(1,1,1,1)$. However, we have "overreacted": Some of the terms, such as $a_{11}a_{21}a_{32}a_{42}$, $a_{11}a_{21}a_{31}a_{42}$, and $a_{11}a_{21}a_{31}a_{41}$, appear in $S(1,1,2)$ with multiplicity greater than 1. For example, $a_{11}a_{21}a_{32}a_{42}$ appears both in $r_{1*2}r_3r_4$ and in $r_1r_2r_{3*4}$. We compensate by adding $S(2,2)$. Now, we count the number of the terms of the form $a_{1i}a_{2j}a_{3j}a_{4j}, i\neq j$, in

$$S(1,1,1,1)-S(1,1,2)+S(2,2).$$

Each of them occurs once in $S(1,1,1,1)$, three times in $S(1,1,2)$ (for example, $a_{11}a_{22}a_{32}a_{42}$ appears once in each of $r_1r_2r_{3*4}$, $r_1r_3r_{2*4}$, and $r_1r_4r_{2*3}$) and does not appear in $S(2,2)$. Hence we compensate by adding twice $S(1,3)$:

$$S(1,1,1,1)-S(1,1,2)+S(2,2)+2S(1,3). \qquad (2.4)$$

It remains to account for the terms $a_{1j}a_{2j}a_{3j}a_{4j}, j=1,\ldots,n$. Each of them appears once in $S(1,1,1,1)$, six times in $S(1,1,2)$, three times in $S(2,2)$, and eight times in $2S(1,3)$. Therefore we must subtract $1-6+3+8=6$ times $S(4)$ from (2.4), and the formula (2.3) follows.

Example 2.1. Compute the permanent of matrix

$$A=\begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

We compute:

$$S(1,1,1,1)=144, \qquad S(1,1,2)=151,$$
$$S(2,2)=13, \qquad S(1,3)=13, \qquad S(4)=0.$$

Hence,

$$\operatorname{Per}(A)=144-151+13+2\times 13-6\times 0=32.$$

Binet did not explain how he derived the coefficients in (2.4), nor did he give a general formula for the permanent of an $m\times n$ matrix for $m>4$.

Recently Binet's formula was generalized to any $m\times n$ matrices, $m\leqslant n$ [301]. This formula and a more efficient formula due to Ryser [87] will be given in the chapter on the evaluation of permanents.

1.3 The Continuators: Borchardt, Cayley, and the Master from Edinburgh—Sir Thomas Muir

During the century that followed the appearance of the memoirs of Binet and Cauchy, some twenty papers on permanents were published. Most of them dealt with identities involving determinants and permanents. The results that created the most interest are identities of Borchardt [4], Cayley [6], and Muir [14]. All three are formulas for the product of the permanent and the determinant of a matrix.

Let $A=(a_{ij})$ be an $n\times n$ matrix. Then

$$\operatorname{per}(A)\det(A)=\left(\sum_{\sigma\in E}\prod_{i=1}^{n}a_{i\sigma(i)}\right)^2-\left(\sum_{\sigma\in F}\prod_{i=1}^{n}a_{i\sigma(i)}\right)^2, \tag{3.1}$$

where E and F are the sets of even and odd permutations, respectively. The problem is how to express the difference on the right in a more attractive form; for example, it is clearly equal to

$$\sum_{\sigma\in E}\prod_{i=1}^{n} a_{i\sigma(i)}^2-\sum_{\sigma\in F}\prod_{i=1}^{n} a_{i\sigma(i)}^2+f(A)=\det(A^{(2)})+f(A), \qquad (3.2)$$

where $A^{(2)}=A*A$ is the matrix whose (i,j) entry is a_{ij}^2, and $f(A)$ represents the remaining terms. If $n=2$, then actually $f(A)=0$. For $n=3$, Cayley expressed $f(A)$ in terms of the determinant of a related matrix.

THEOREM 3.1 (Cayley [6]). *Let $A=(a_{ij})$ be a 3×3 matrix, $a_{ij}\neq 0$, and let $A^{(-1)}$ be the 3×3 matrix whose (i,j) entry is a_{ij}^{-1}. Then*

$$\operatorname{per}(A)\det(A)=\det(A^{(2)})+2\left(\prod_{i,j} a_{ij}\right)\det(A^{(-1)}). \qquad (3.3)$$

The proof follows immediately from (3.1) and (3.2).

COROLLARY. *If $B=(b_{ij})$ is a singular matrix, $b_{ij}\neq 0$, then*

$$\operatorname{per}(B^{(-1)})\det(B^{(-1)})=\det(B^{(-2)}). \qquad (3.4)$$

Here $B^{(-2)}$ denotes the 3×3 matrix whose (i,j) entry is b_{ij}^{-2}.

Borchardt obtained a formula similar to (3.4) for any n, but only for a special type of matrix.

THEOREM 3.2 (Borchardt [4]). *Let A be an $n\times n$ matrix whose (i,j) entry is $(s_i-t_j)^{-1}$. Then*

$$\operatorname{per}(A)\det(A)=\det(A^{(2)}). \qquad (3.5)$$

We shall not offer a separate proof of Borchardt's result, since it is an immediate consequence of a generalization of Cayley's theorem by Carlitz and Levine [63], which we give at the end of this section.

Sir Thomas Muir occupies a unique position in the history of permanents, and even more so in the history of determinants. In his monumental *The Theory of Determinants in the Historical Order of Development* [25, 26, 35, 36, 40], he gives *inter alia* an abstract of every paper on permanents published before 1920, a third of which were his own contributions. Muir's papers deal mostly with expressions and identities involving permanents and determinants. Of these we give below one of the results in his first paper on permanents.

THEOREM 3.3 (Muir [14]). *Let $A=(a_{ij})$ and $X=(x_{ij})$ be n-square matrices. Then*

$$\operatorname{per}(A)\det(X)=\sum_{\sigma\in S_n}\varepsilon(\sigma)\det(A*X_\sigma), \qquad (3.6)$$

where X_σ *is the matrix whose* ith *row is the* $\sigma(i)$th *row of* X, $A*X_\sigma$ *is the Hadamard product, and* $\varepsilon(\sigma)$ *denotes the sign of* σ.

(The Hadamard product of two $n\times n$ matrices, $P=(p_{ij})$ and $Q=(q_{ij})$, is the $n\times n$ matrix whose (i,j) entry is $p_{ij}q_{ij}$.)

Each side of (3.6) contains $(n!)^2$ terms: To each pair of permutations (φ,ψ) corresponds a term. Thus to (φ,ψ) we can make correspond on the left-hand side the term

$$\varepsilon(\psi)\prod_{i=1}^{n} a_{i,\varphi(i)}\prod_{t=1}^{n} x_{t,\psi(t)},$$

and to (φ,σ) we can make correspond on the right-hand side the term

$$\varepsilon(\sigma)\left(\varepsilon(\varphi)\prod_{i=1}^{n}(A*X_\sigma)_{i,\varphi(i)}\right)=\varepsilon(\sigma\varphi)\prod_{i=1}^{n} a_{i,\varphi(i)}x_{\sigma(i),\varphi(i)}$$

$$=\varepsilon(\varphi\sigma^{-1})\prod_{i=1}^{n} a_{i,\varphi(i)}\prod_{t=1}^{n} x_{t,\varphi\sigma^{-1}(t)}.$$

Now, $\psi\leftrightarrow\varphi\sigma^{-1}$ establishes one-to-one correspondence between the equal terms on the left-hand side and those on the right-hand side.

One hundred years after the appearance of Cayley's paper, Levine [61] generalized the identity (3.4) to 4×4 matrices. A year later Carlitz and Levine [63] produced the following generalization to $n\times n$ matrices.

THEOREM 3.4. *If* $B=(b_{ij})$ *is an n-square matrix of rank* $\leqslant 2$ *and without zero entries, then*

$$\operatorname{per}(B^{(-1)})\det(B^{(-1)})=\det(B^{(-2)}). \tag{3.7}$$

(Recall that $B^{(-1)}$ and $B^{(-2)}$ denote $n\times n$ matrices whose (i,j) entries are b_{ij}^{-1} and b_{ij}^{-2}, respectively.)

Proof. In Muir's identity (3.6), replace both a_{ij} and x_{ij} by b_{ij}^{-1}. Then the left-hand side of (3.6) becomes the left-hand side of (3.7), and the term on the right corresponding to the identity permutation is the right-hand side of (3.7). It remains to show that all the terms corresponding to other permutations vanish. Consider a permutation σ that is factored into a product of disjoint cycles. If any of the cycles is a transposition, then the corresponding matrix on the right-hand side of (3.6) has two identical rows, and thus its determinant vanishes.

Assume that the rows of B are linear combinations of the first two rows and that the first cycle of σ is $(12\cdots h)$, where $3\leqslant h\leqslant n$.

Let $C=(c_{ij})$ be the $h\times n$ matrix consisting of the first h rows of the matrix $B^{(-1)}*B_\sigma^{(-1)}$. Then

$$c_{ij}=b_{ij}^{-1}b_{i'j}^{-1},$$

where $i'=i+1$ for $i=1,\ldots,h-1$, and $i'=1$ if $i=h$. Let $K=\mathrm{diag}(k_1,\ldots,k_n)$, where

$$k_j=\prod_{s=1}^{h}b_{sj},$$

$j=1,\ldots,n$. Then $D=(d_{ij})=CK$ is an $h\times n$ matrix whose (i,j) entry is

$$d_{ij}=\prod_{\substack{s=1\\ s\neq i,i'}}^{h}b_{sj},$$

$i=1,\ldots,h$; $j=1,\ldots,n$. Also, the rank of D is equal to that of C. But the rows of B are linear combinations of its first two rows; that is,

$$b_{sj}=\lambda_s b_{1j}+\mu_s b_{2j},$$

$s,j=1,\ldots,n$. Thus,

$$\begin{aligned}d_{ij}&=\prod_{\substack{s=1\\ s\neq i,i'}}^{h}(\lambda_s b_{1j}+\mu_s b_{2j})\\ &=\sum_{t=0}^{h-2}\nu_{it}b_{1j}^{h-2-t}b_{2j}^{t},\end{aligned}$$

so that the row space of D is spanned by $h-1$ vectors, and therefore the rank of C is less than h. It follows that the matrix corresponding to σ is singular. Hence the result. ■

Borchardt's identity (3.5) is an immediate corollary of Theorem 3.4, since the rank of $A^{(-1)}$ cannot exceed 2.

1.4 Renaissance of Permanents: Muirhead's Theorem, Pólya's Problem, Schur's Inequality and van der Waerden's Conjecture

The results on permanents of Borchardt, Cayley, Muir, and other writers in the nineteenth century consist of identities involving permanents and determinants. All these results are straightforward and elementary, although some of them are not easy to prove, owing to the complexity inherent in any multilinear function. The turning point in the history of permanents came at the beginning of this century with the appearance of a beautiful theorem of Muirhead. The other three outstanding events that ushered the new era were: a problem posed by Pólya, the concept of generalized matrix functions introduced by Schur, and a conjecture proposed by van der Waerden.

If $\alpha=(\alpha_1,\ldots,\alpha_n)$ is a real n-tuple, let $\alpha^*=(\alpha_1^*,\ldots,\alpha_n^*)$ denote the n-tuple α rearranged in nonincreasing order, $\alpha_1^*\geqslant\cdots\geqslant\alpha_n^*$.

DEFINITION 4.1. A nonnegative n-tuple $\alpha=(\alpha_1,\ldots,\alpha_n)$ is said to be *majorized* by a nonnegative n-tuple $\beta=(\beta_1,\ldots,\beta_n)$, written $\alpha\prec\beta$, if

$$\alpha_1^*+\cdots+\alpha_k^*\leqslant\beta_1^*+\cdots+\beta_k^*$$

for $k=1,\cdots,n-1$, and

$$\alpha_1+\cdots+\alpha_n=\beta_1+\cdots+\beta_n.$$

THEOREM 4.1 (Muirhead [24]). *Let $c=(c_1,\ldots,c_n)$ be a positive n-tuple, and let $\alpha=(\alpha_1,\ldots,\alpha_n)$ and $\beta=(\beta_1,\ldots,\beta_n)$ be n-tuples of nonnegative integers. Let A and B be $n\times n$ matrices whose (i,j) entries are $c_i^{\alpha_j}$ and $c_i^{\beta_j}$, respectively. A necessary and sufficient condition that*

$$\operatorname{per}(A)\leqslant\operatorname{per}(B) \tag{4.1}$$

is that

$$\alpha\prec\beta. \tag{4.2}$$

Equality holds in (4.1) *if and only if either $\alpha=\beta$ or $c_1=\cdots=c_n$.*

Hardy, Littlewood, and Pólya [44] extended Muirhead's theorem to any nonnegative n-tuples α and β, proved the extended theorem, and showed that (4.2) is equivalent to the following condition: *There exists a doubly stochastic $n\times n$ matrix S such that*

$$\alpha=S\beta. \tag{4.3}$$

For proofs of Muirhead's theorem, see [44]. An outline of a proof that (4.2) and (4.3) are equivalent may also be found in [92].

Apart from its intrinsic interest, Muirhead's elegant result and its Hardy–Littlewood–Pólya generalization have had numerous important applications in many branches of mathematics.

The next landmark in the history of permanents was a question asked by Pólya [29]. As we shall see in the next chapters, most problems involving permanents present considerably more difficulty than the corresponding problems for determinants. It would therefore be of interest to find a transformation that would convert permanents into determinants; specifically, given S, a set of $n \times n$ matrices, to find a linear transformation T on S such that

$$\operatorname{per}(T(A)) = \det(A) \tag{4.4}$$

for all $A \in S$. Pólya's problem concerns transformations that involve only a uniform affixing of a plus or a minus sign to each position in the matrix. For example, if S is M_2, the set of 2×2 matrices, and

$$T\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & -a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

then (4.4) holds for all matrices in M_2. Similarly, if S is the set of all 3×3 matrices with a zero in the $(1,3)$ position, define T by

$$T\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & 0 \\ -a_{21} & a_{22} & a_{23} \\ a_{31} & -a_{32} & a_{33} \end{bmatrix}.$$

Then again $\operatorname{per}(T(A)) = \det(A)$ for all A in S.

Pólya [29] asserted that, if S is the set of all $n \times n$ matrices, then for $n \geqslant 3$ there exists no transformation T involving uniform affixing of $\pm$ signs to entries of the matrices such that $\operatorname{per}(T(A)) = \det(A)$.

The proof is quite simple. First, observe that it suffices to prove the assertion for $n = 3$: If $n > 3$, consider the direct sums of 3×3 matrices and the identity matrix I_{n-3}. Let $n = 3$, and suppose that such a transformation exists. Then the number of minuses affixed to each diagonal corresponding to an even permutation would have to be even, and thus the total number of minuses in the transformation must be even. On the other hand, the number of minuses on each diagonal corresponding to an odd permutation would have to be odd, and hence the total number of minuses would have to be a sum of three odd numbers—that is, an odd number. Contradiction.

Pólya's result was substantially generalized by Marcus and Minc [70], who showed that for $n \geqslant 3$ there exists no linear transformation T on the set of $n \times n$ matrices such that $\operatorname{per}(T(A)) = \det(A)$. This result finally and completely closed the door on any hope that problems involving permanents can be solved easily via determinants.

As we saw in one of our examples above, in certain sets of matrices the permanent can be converted into the determinant by affixing in a uniform way $\pm$ signs to the matrix entries. In fact, Gibson [192] showed that, if an n-square $(0,1)$-matrix A has a positive permanent and if the permanent of A can be converted into a determinant by affixing $\pm$ signs to the entries of A, then A has at most $(n^2 + 3n - 2)/2$ ones.