Cayley-Hamilton Theorem for mixed discriminants

R. B. Bapat
Indian Statistical Institute, Delhi
7-SJSS Marg, New Delhi 110 016, India.
email: rbb@isid.ac.in

Souvik Roy
Indian Statistical Institute, Kolkata
203, B.T. Road
Kolkata 700 108, India.
email: souvik.2004@gmail.com

Abstract

The mixed discriminant of an n-tuple of $n \times n$ matrices $A_1, \ldots, A_n$ is defined as

$$\mathfrak{D}(A_1, A_2, \ldots, A_n) = \frac{1}{n!} \sum_{\sigma \in S(n)} \det(A^{(1)}_{\sigma(1)}, A^{(2)}_{\sigma(2)}, \ldots, A^{(n)}_{\sigma(n)}),$$

where $A^{(i)}$ denotes the ith column of the matrix A and $S(n)$ denotes the group of permutations of $1, 2, \ldots, n$. For n matrices $A_1, \ldots, A_n$ and indeterminates $\lambda_1, \ldots, \lambda_n$, set

$$\Phi_{\lambda_1, \ldots, \lambda_n}(A_1, \ldots, A_n) = \mathfrak{D}(\lambda_1 I - A_1, \ldots, \lambda_n I - A_n).$$

It is shown that $\Phi_{A_1, \ldots, A_n}(A_1, \ldots, A_n) = 0$.

Key words. Mixed discriminant, Cayley-Hamilton Theorem, Directed path

AMS Subject Classifications. 15A15, 05C05.

1 Introduction

We consider complex matrices. If A is an $n \times n$ matrix, then we denote its determinant by $\det(A)$. The identity matrix of the appropriate order is denoted by I. The characteristic polynomial of the $n \times n$ matrix A is given by $f_A(\lambda) = \det(\lambda I - A)$. The Cayley-Hamilton Theorem asserts that for any $n \times n$ matrix A, $f_A(A) = 0$. The purpose of this note is to prove an analogue of the Cayley-Hamilton Theorem for mixed discriminants (see Theorem 1.1).

To put our result in context, we begin by a brief account of the Cayley-Hamilton Theorem. Cayley proved the theorem for 2×2 matrices in 1858, where he also asserted that he had verified it for 3×3 matrices. In 1853, Hamilton [5] proved a result for matrices over quaternions, equivalent to the Cayley-Hamilton Theorem for 3×3 matrices. The result has been extended to $n \times n$ matrices by Zhang [12] using q-determinants.

A complete proof of the Cayley-Hamilton Theorem was given by Frobenius in 1878, using minimal polynomials.

The Cayley-Hamilton Theorem is usually stated for complex matrices but it holds for matrices over any field. In fact, the result is true for matrices over a commutative ring, as shown by A. Buchheim in 1884. For an outline of the proof, references to the original works of Cayley and Frobenius, and for additional remarks we refer to Horn and Johnson [6].

Straubing [10] gave a particularly instructive combinatorial proof of the Cayley-Hamilton Theorem. The main contribution of Straubing was to consider matrix A over a noncommutative ring, with the assumption that entries in distinct rows of A commute. For such a matrix, definition of the determinant does not pose a problem and Straubing showed that $f_A(A) = 0$. Moreover, the proof of Straubing is purely combinatorial, essentially showing that the Cayley-Hamilton Theorem can be regarded as a result about directed graphs. A readable exposition of the proof is given by Zeilberger [11].

For some other extensions of Cayley-Hamilton Theorem to more general algebraic structures, we refer to [4,8].

We now turn to the mixed discriminant. The mixed discriminant is an important matrix function that generalizes both the determinant and the permanent. Let $A_1, A_2, \ldots, A_n$ be an n-tuple of $n \times n$ matrices. The mixed discriminant of $A_1, A_2, \ldots, A_n$ is defined as

$$\mathfrak{D}(A_1, A_2, \ldots, A_n) = \frac{1}{n!} \sum_{\sigma \in S(n)} \det(A_{\sigma(1)}^{(1)}, A_{\sigma(2)}^{(2)}, \ldots, A_{\sigma(n)}^{(n)}),$$

where $A^{(i)}$ denotes the ith column of the matrix A and $S(n)$ denotes the group of permutations of $1, 2, \ldots, n$.

Example Consider the matrices

$$A_1 = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}.$$

Then

$$\mathfrak{D}(A_1, A_2) = \frac{1}{2} \times \left[\det\begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} + \det\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}\right] = -\frac{15}{2}.$$

Note that if $A_1 = A_2 = \cdots = A_n = A$, then

$$\mathfrak{D}(A_1, A_2, \ldots, A_n) = \det(A). \tag{1}$$

Let B be an $n \times n$ matrix. Recall that the permanent of B, denoted per (B) is defined as

$$\text{per}(B) = \sum_{\sigma \in S(n)} \prod_{i=1}^{n} b_{i\sigma(i)}.$$

Let A_i be the $n \times n$ diagonal matrix with diagonal elements $b_{1i}, \ldots, b_{ni}, i = 1, \ldots, n$. Then it can be seen that

$$\mathfrak{D}(A_1, A_2, \ldots, A_n) = \text{per}(B).$$

Thus the mixed discriminant generalizes the permanent.

We refer to [1] for basic properties of the mixed discriminant. The mixed discriminant is closely related to the mixed volume of convex bodies, see, for example Schneider [9]. In combinatorics, mixed discriminant can be used to refine some formulae which involve the determinant. An example is a formula for the number of spanning trees with color restrictions, see [2,3]. The present paper is in the same spirit.

For $n \times n$ matrices $A_1, ..., A_n$ and noncommuting indeterminates $\lambda_1, \ldots, \lambda_n$, we define $\Phi_{\lambda_1,\ldots,\lambda_n}(A_1, ..., A_n)$ as

$$\Phi_{\lambda_1,\ldots,\lambda_n}(A_1, ..., A_n) = \mathfrak{D}(\lambda_1 I - A_1, ..., \lambda_n I - A_n),$$

and it may be viewed as a generalized characteristic polynomial. It may be remarked that setting $\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$, the polynomial $\Phi_{\lambda,\ldots,\lambda}(A_1, ..., A_n)$ that we obtain has been called the *mixed characteristic polynomial* and is one of the tools in the recent celebrated proof of the Kadison-Singer problem [7].

The purpose of this paper is to prove the following.

Theorem 1.1 (Cayley-Hamilton Theorem for mixed discriminants) *Let $A_1, ..., A_n$ be $n \times n$ matrices. Then*

$$\Phi_{A_1,\ldots,A_n}(A_1, ..., A_n) = 0.$$

We illustrate the validity of Theorem 1.1 by an example.

Example Consider the matrices

$$A_1 = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}, \; A_2 = \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}.$$

Then we have

$$\begin{aligned} 2! \times \Phi_{\lambda_1,\lambda_2}(A_1, A_2) &= \mathfrak{D}(\lambda_1 I - A_1, \lambda_2 I - A_2) \\ &= \mathfrak{D}\left(\begin{pmatrix} \lambda_1 - 1 & -2 \\ -3 & \lambda_1 - 1 \end{pmatrix}, \begin{pmatrix} \lambda_2 - 1 & -4 \\ -3 & \lambda_2 - 2 \end{pmatrix} \right) \\ &= \lambda_1\lambda_2 + \lambda_2\lambda_1 - 3\lambda_1 - 2\lambda_2 - 15. \end{aligned}$$

Hence,

$$\begin{aligned} 2! \times \Phi_{A_1,A_2}(A_1, A_2) &= \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \\ &\quad -3\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} - 2\begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} - 15\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

2 Proof of the main result

We first prove a preliminary result.

Lemma 2.1 *For $n \times n$ matrices $A_1, \ldots, A_n$,*

$$\Phi_{\lambda_1,\ldots,\lambda_n}(A_1, \ldots, A_n) = \frac{1}{n!}\sum_{k=0}^{n} \sum_{(i_1,\ldots,i_k)} c_{i_1,\ldots,i_k}\lambda_{i_1}\ldots\lambda_{i_k}$$

where $c_{i_1,\ldots,i_k} = (-1)^{n-k}(n-k)!\sum_W \mathfrak{D}(A_{j_1}^W, \ldots, A_{j_{n-k}}^W)$, $\{j_1, \ldots, j_{n-k}\} = N\backslash\{i_1, \ldots, i_k\}$, and A_j^W is a $W \times W$ principal sub-matrix of A_j where W is a size $n-k$ subset of N.

Proof. Fix $i_1, \ldots, i_k$. Take a size $n-k$ subset $W = \{w_1, \ldots, w_{n-k}\}$ of N, where $w_1 < w_2 < \cdots < w_{n-k}$. Note that

$$\begin{aligned} \Phi_{\lambda_1,\ldots,\lambda_n}(A_1, \ldots, A_n) &= \mathfrak{D}(\lambda_1 I - A_1, \ldots, \lambda_n I - A_n) \\ &= \frac{1}{n!}\sum_{\sigma\in S(n)} \det(\lambda_{\sigma(1)}^{(1)} I - A_{\sigma(1)}^{(1)}, \ldots, \lambda_{\sigma(n)}^{(n)} I - A_{\sigma(n)}^{(n)}), \end{aligned}$$

where $\lambda^{(j)}$ denotes the column vector whose jth element is λ and all other elements are zero. Suppose $N \setminus W = \{w_{n-k+1}, \ldots, w_n\}$ where $w_{n-k+1} < \cdots < w_n$. Consider the set of permutations $\bar{S}(n) \subset S(n)$ such that for all $\sigma \in \bar{S}(n)$ we have $\sigma(w_m) = i_{m-n+k}$ for $m = n-k+1, \ldots, n$. Clearly there are $(n-k)!$ many such permutations. It follows from elementary linear algebra that for $\sigma \in \bar{S}(n)$, the coefficient of $\lambda_{i_1} \cdots \lambda_{i_k}$ in $\det(\lambda^{(1)}_{\sigma(1)} I - A^{(1)}_{\sigma(1)}, \ldots, \lambda^{(n)}_{\sigma(n)} I - A^{(n)}_{\sigma(n)})$ is $(-1)^{n-k} \det(A^{(w_1)}_{\sigma(w_1)}, \ldots, A^{(w_{n-k})}_{\sigma(w_{n-k})})$.

The coefficient of $\lambda_{i_1} \cdots \lambda_{i_k}$ in $\Phi_{\lambda_1,\ldots,\lambda_n}(A_1, \ldots, A_n)$ is given by

$$\frac{1}{n!} \sum (-1)^{n-k} \sum_{\sigma \in \bar{S}(n)} \det(A^{(w_1)}_{\sigma(w_1)}, \ldots, A^{(w_{n-k})}_{\sigma(w_{n-k})}), \tag{2}$$

where the first sum is taken over subsets $W = \{w_1, \ldots, w_{n-k}\}$ of N. It follows from the definition of the mixed discriminant that (2) equals

$$\frac{1}{n!}(-1)^{n-k}(n-k)! \sum_{W} \mathfrak{D}(A^W_{j_1}, \ldots, A^W_{j_{n-k}}),$$

and the proof is complete. ■

We introduce some terminology needed for the proof of the main result. Consider an $n \times n$ matrix $A_k = (a^k_{ij})_{n \times n}$ labeled by k. The graph of A_k is a k-colored directed weighted graph $G^k = (V, E^k, \mu)$ where $V = \{1, \ldots, n\}$, $E^k = V \times V \times \{k\}$, and $\mu((i,j,k)) = a^k_{ij}$ for all $i, j \in V$. The graph of a tuple of $n \times n$ matrices is the union of the graphs of the individual matrices. More formally, the graph of $A_1 = (a^1_{ij})_{n \times n}, A_2 = (a^2_{ij})_{n \times n}, \ldots, A_n = (a^n_{ij})_{n \times n}$ is the multi-edged directed weighted graph $G = (V, E, \mu)$ where $V = \{1, \ldots, n\}$, $E = \cup_{k=1}^n E^k$, $\mu((i,j,k)) = a^k_{ij}$ for all $i, j, k \in \{1, \ldots, n\}$. A path P of length m from a vertex i_1 to another vertex i_{m+1} in a graph G is a sequence of edges $((i_1, i_2, \cdot), (i_2, i_3, \cdot), \ldots, (i_m, i_{m+1}, \cdot))$. If $i_1 = i_{m+1}$ then the path is called a cycle of length m. Moreover, an empty sequence is called an empty path.

We follow the technique of Straubing[10], as explained in Zeilberger[11], with some modifications to prove the main result.

Proof of Theorem 1.1. Let G be the graph of $A_1, \ldots, A_n$. Fix i, j and let $\mathcal{A}(i,j)$ be the set of pairs (P, C) such that
(i) P is a path from i to j where P is possibly empty when $i = j$,
(ii) C is a disjoint (possibly empty) union of cycles, and
(iii) for each $A_k, k \in \{1, \ldots, n\}$, there is a unique edge $(\cdot, \cdot, k)$ in $P \cup C$.

Define the weight of (P, C), denoted by $\mu((P, C))$, to be $(-1)^{\#C}$ [product of the weights of the edges in (P,C)] where $\#C$ is the number of cycles in

C, and weight of $\mathcal{A}(i,j)$, denoted by $\mu(\mathcal{A}(i,j))$, is the total weight of all (P,C) in $\mathcal{A}(i,j)$.

CLAIM. The (i,j)-th entry in the left side of Theorem 1.1 is $\frac{1}{n!}(\mu(\mathcal{A}(i,j))$.

PROOF OF THE CLAIM. By using Lemma 2.1 it is enough to show that $\mu(P,C,L)$ is equal to the (i,j)-th entry of

$$\sum_{k=0}^{n} \sum_{(i_1,\dots,i_k)} (-1)^{n-k}(n-k)! \sum_{W} \mathfrak{D}(A_{j_1}^{W},\dots,A_{j_{n-k}}^{W})A_{i_1}\dots A_{i_k} \tag{3}$$

where $\{j_1,\dots,j_{n-k}\} = N\backslash\{i_1,\dots,i_k\}$, and A_j^W is a $W \times W$ principal submatrix of A_j where W is a size $n-k$ subset of N.

Fix k,$(i_1, i_2, \dots, i_k)$, and $W = \{w_1, w_2, \dots, w_{n-k}\}$ where $w_1 < w_2 < \dots < w_{n-k}$. The corresponding term in the expression (3) is

$$\begin{aligned} &(-1)^{n-k}(n-k)!\mathfrak{D}(A_{j_1}^{W},\dots,A_{j_{n-k}}^{W})A_{i_1}\dots A_{i_k} \\ &= (-1)^{n-k} \sum_{\sigma\in\bar{S}(n-k)} \det\left(A_{\sigma(1)}^{(w_1)},\dots,A_{\sigma(n-k)}^{(w_{n-k})}\right) A_{i_1}\dots A_{i_k} \end{aligned}$$

where $\bar{S}(n-k)$ is the set of bijections from the set $\{1,\dots,n-k\}$ to the set $\{j_1,\dots,j_{n-k}\}$. Now fix $\sigma \in \bar{S}(n-k)$ and suppose $\sigma(i) = j_i'$. Then the corresponding term is

$$\begin{aligned} &(-1)^{n-k}\det\left(A_{j_1'}^{(w_1)},\dots,A_{j_{n-k}'}^{(w_{n-k})}\right) A_{i_1}\dots A_{i_k} \\ &= (-1)^{n-k} \sum_{\pi\in S(n-k)} |\pi| \prod_{l=1}^{n-k} a_{w_l,w_{\pi(l)}}^{j_{\pi(l)}'} A_{i_1}\dots A_{i_k} \end{aligned}$$

where $S(n-k)$ is the group of permutations of the numbers $1,\dots,n-k$. Fix $\pi \in S(n-k)$ then the corresponding term is

$$\begin{aligned} &(-1)^{n-k}|\pi| \prod_{l=1}^{n-k} a_{w_l,w_{\pi(l)}}^{j_{\pi(l)}'} A_{i_1}\dots A_{i_k} \\ &= (-1)^{\#C} \prod_{l=1}^{n-k} a_{w_l,w_{\pi(l)}}^{j_{\pi(l)}'} A_{i_1}\dots A_{i_k}. \end{aligned}$$

Here the equality follows from the observation that

$$(-1)^{\#C} \prod_{l=1}^{n-k} a_{w_l,w_{\pi(l)}}^{j_{\pi(l)}'} = |\pi| \prod_{l=1}^{n-k} (-a_{w_l,w_{\pi(l)}}^{j_{\pi(l)}'})$$

where C is the disjoint union of cycles with $n-k$ edges of all colors $j_1, \ldots, j_{n-k}$ constituted by the sequence of edges $((w_l, w_{\pi(l)}, j'_{\pi(l)}); l = 1, \ldots n-k)$. Moreover, the (i,j)-th element in $A_{i_1} \ldots A_{i_k}$ is the sum of the weights of all paths P that start at i and ends at j and that has k many edges of all colors $i_1, \ldots, i_k$ (ordered). Note that if $k = 0$ then the term $(a_{w_l, w_{\pi(l)}}^{j'_{\pi(l)}}; l = 1, \ldots n)$ appears at the (i,j)-th entry of the expression (3) only if $i = j$. This term is represented as the sum of the weights of a pair (P, C) where P is empty. This shows that the (i,j)-th entry of the expression (3) is a sum of the weights of pairs (P, C) satisfying the conditions (i)-(iii) stated at the beginning of the proof. It is also clear from the construction of the pair (P, C) that every pair (P, C) comes exactly once in the expression (3). This completes the proof of the claim.

We return to proof of the theorem. We show that the weight of $\mathcal{A}(i,j)$ is zero. We consider an involution from $\mathcal{A}(i,j)$ into itself so that every pair (P, C) is mapped to some (P', C') such that $\mu(P, C) = -\mu(P', C')$. Consider a pair (P, C) such that P is non-empty. Note that as there are n many edges in $P \cup C$ and P is a non-empty path

(a) either there is a cycle in P that is disjoint from the cycles in C, if so then call the first (starting from i in P) such cycle $c(P)$, or

(b) P intersects some cycle in C, if so then call the first (starting from i in P) such cycle $c(C)$. Suppose P intersects $c(C)$ at vertex v.

In case of (a) we define $P' = P \setminus c(P)$ and $C' = C \cup c(P)$, and in case of (b) define $P' = P$ *with* $c(C)$ *inserted at* v, and $C' = C \setminus c(C)$.

Now consider a pair (P, C) where P is empty. As discussed before such a pair will occur for the (i,j)-th entry only if $i = j$. Let C be composed of disjoint cycles $C_1, \ldots, C_s$. Then i must be in exactly one of these cycles and suppose $i \in C_t$. Then we set $P' = C_t$ and $C' = C \setminus C_t$.

Note that (P', C') satisfies the conditions stated at the beginning of the proof. Moreover, $\mu(P, C) = -\mu(P', C')$ as they have the same set of edges and number of cycles in C and C' differ by exactly 1. The fact that the association of (P, C) with (P', C') is an involution is also clear from the construction. This proves that the weight of $\mathcal{A}(i,j)$ is zero. ■

Example This example illustrates the construction of the involution employed in the proof of Theorem 1.1. Suppose $n = 6, i = 1, j = 2$. Consider a pair (P, C) where $P = ((1,2,1), (2,3,2), (3,4,3), (4,3,4), (3,2,5))$ and $C = ((6,6,6))$. Note that this is an example of case (a) where the path P contains a cycle $((3,4,3),(4,3,4))$. So $c(P) = ((3,4,3),(4,3,4))$. Hence we have $P' = ((1,2,1),(2,3,2),(3,2,5))$, and $C' = ((3,4,3),(4,3,4)) \cup ((6,6,6))$.

Now consider a pair (P, C) where $P = ((1, 2, 1), (2, 3, 2), (3, 4, 3), (4, 2, 4))$ and $C = ((3, 5, 5), (5, 3, 6))$. Note that this is case (b) where the path P intersects the cycle $((3, 5, 5), (5, 3, 6))$ in C at the vertex 3. So $c(C) = ((3, 5, 5), (5, 3, 6))$. Hence we have $P' = ((1, 2, 1), (2, 3, 2), (3, 5, 5), (5, 3, 6), (3, 4, 3), (4, 2, 4))$, and $C' = \emptyset$.

Finally suppose $n = 6$ and $i = j = 4$. Consider the pair (P, C) where P is empty and $C = ((1, 2, 1), (2, 3, 2), (3, 1, 3)) \cup ((4, 5, 4), (5, 4, 5)) \cup ((6, 6, 6))$. Then $P' = ((4, 5, 4), (5, 4, 5))$ and $C' = ((1, 2, 1), (2, 3, 2), (3, 1, 3)) \cup ((6, 6, 6))$.

Observe that the image of (P', C') under the mapping we consider is (P, C) for all the above examples. This shows that the mapping is indeed an involution.

Note that by setting $A_1 = A_2 = \cdots = A_n = A$ in Theorem 1.1 we obtain the classical Cayley-Hamilton Theorem, in view of (1). We also note that although we considered complex matrices, the results are valid for matrices over a commutative ring with unity. Several proofs of the Cayley-Hamilton Theorem rely on first using a similarity transformation to change the matrix to a triangular, or a diagonal matrix. If $A_1, \ldots, A_n$ can be simultaneously changed to a triangular, or a diagonal form, then the proof of Theorem 1.1 can easily be given by first proving it for the special case. In general, however none of the proof techniques used for proving the Cayley-Hamilton Theorem seem to work for proving Theorem 1.1, except the combinatorial approach that we have used.

Acknowledgment. We sincerely thank the referee for several constructive comments. The first author acknowledges support from the JC Bose Fellowship, Department of Science and Technology, Government of India.

References

1. R.B. Bapat, Mixed discriminants of positive semidefinite matrices. Linear Algebra Appl. 126 (1989), 107–124.

2. R.B. Bapat, Mixed discriminants and spanning trees. Combinatorial mathematics and applications (Calcutta, 1988). Sankhya Ser. A 54 (1992), Special Issue, 49–55.

3. R.B. Bapat and G. Constantine, An enumerating function for spanning forests with color restrictions. Linear Algebra Appl. 173 (1992), 231–237.

4. Radu Grosu, The Cayley-Hamilton theorem for noncommutative semirings. Implementation and application of automata, 143–153, Lecture Notes in Comput. Sci., 6482, Springer, Berlin, 2011.

5. William R. Hamilton, Lectures on quaternions. Cambridge, MA, 1853.

6. Roger A. Horn and Charles R. Johnson, Matrix analysis. Second edition. Cambridge University Press, Cambridge, 2013.

7. A.W. Marcus, D.A. Spielman and N. Srivastava, Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. arXiv:1306.3969v4 [math.CO] 14 Apr 2014.

8. D.E. Rutherford, The Cayley-Hamilton theorem for semi-rings. Proc. Roy. Soc. Edinburgh Sect. A 66 1962/1964 (1965) 211–215.

9. Rolf Schneider, Convex bodies: the Brunn-Minkowski theory. Second expanded edition. Encyclopedia of Mathematics and its Applications, 151. Cambridge University Press, Cambridge, 2014.

10. Howard Straubing, A combinatorial proof of the Cayley-Hamilton theorem. Discrete Math. 43 (1983), no. 2-3, 273–279.

11. Doron Zeilberger, A combinatorial approach to matrix algebra. Discrete Math. 56 (1985), no. 1, 61–72.

12. Fuzhen Zhang, Quaternions and matrices of quaternions. Linear Algebra Appl., 251 (1997), 21–57.