isid/ms/2009/05
May 18, 2009
http://www.isid.ac.in/~statmath/eprints

Higher order derivatives and perturbation bounds for determinants

Rajendra Bhatia
Tanvi Jain

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Higher order derivatives and perturbation bounds for determinants

Rajendra Bhatia
rbh@isid.ac.in
and
Tanvi Jain
tanvij1705@rediffmail.com
Indian Statistical Institute, New Delhi 110016, India

May 18, 2009

Abstract : *The first derivative of the determinant function is given by the well-known Jacobi's formula. We obtain three different expressions for all higher order derivatives. Norms of these derivatives are then evaluated exactly.*

AMS Subject Classifications : 15A15, 15A60, 15A69.

Key words : determinant, derivative, norm, perturbation bound.

1 Introduction

Let $\det : \mathbb{M}(n) \to \mathbb{C}$ be the map taking an $n \times n$ complex matrix to its determinant. The Fréchet derivative of this at a point A is a linear map $D \det A$ from $\mathbb{M}(n)$ into $\mathbb{C}$; and for each X in $\mathbb{M}(n)$.

$$D \det A(X) = \left.\frac{d}{dt}\right|_{t=0} \det(A + tX). \tag{1}$$

The famous *Jacobi formula* says that

$$D \det A(X) = \operatorname{tr}(\operatorname{adj}(A)X), \tag{2}$$

where the symbol $\operatorname{adj}(A)$ stands for the *adjugate* (the *classical adjoint*) of A. The principal goal of this paper is to describe higher order derivatives of the determinant map.

Basic ideas of matrix differential calculus are summarised in Section X.4 of [2], and we follow the notations used there. A full length book on the subject by Magnus and Neudecker [9] has much to offer; the formula (2) is given there on page 149. The recent book by Higham [8] is devoted to various aspects of matrix functions, and the role of the Fréchet derivative in estimating condition numbers is emphasized at several places.

It would be convenient to have some equivalent descriptions of Jacobi's formula. For $1 \leq i, j \leq n$ let $A(i, j)$ be the $(n-1) \times (n-1)$ matrix obtained from A by deleting its ith row and jth column. Then (2) can be restated as

$$D \det A(X) = \sum_{i,j} (-1)^{i+j} \det A(i, j) x_{ij}. \tag{3}$$

For $1 \leq j \leq n$ let $A(j; X)$ be the matrix obtained from A by replacing the jth column of A by the jth column of X and keeping the rest of the columns unchanged. The relation (3) can be expressed also as

$$D \det A(X) = \sum_{j=1}^{n} \det A(j; X). \tag{4}$$

The kth derivative of det at a point A is a map $D^k \det A$ from the k-fold product $\mathbb{M}(n) \times \ldots \times \mathbb{M}(n)$ into $\mathbb{C}$. This map is linear in each of the k arguments and is symmetric in them. Its value at a point $(X^1, \ldots, X^k)$ is

$$D^k \det A(X^1, \ldots, X^k) = \left.\frac{\partial^k}{\partial t_1 \cdots \partial t_k}\right|_{t_1 = \cdots = t_k = 0} \det(A + t_1 X^1 + \cdots + t_k X^k). \tag{5}$$

We will give different formulas for this map that are visible generalisations of (2), (3) and (4). Some notation is needed first.

Let $Q_{k,n}$ be the collection of multiindices $\mathcal{I} = (i_1, \dots, i_k)$ in which $1 \leq i_1 < \cdots < i_k \leq n$. We use the symbol $|\mathcal{I}|$ for the sum $i_1 + \cdots + i_k$. Given two elements $\mathcal{I}$ and $\mathcal{J}$ of $Q_{k,n}$ let $A[\mathcal{I}|\mathcal{J}]$ be the $k \times k$ matrix obtained from A by picking its entries from the rows corresponding to $\mathcal{I}$ and the columns corresponding to $\mathcal{J}$ and let $A(\mathcal{I}|\mathcal{J})$ be the $(n-k) \times (n-k)$ matrix obtained from A by deleting these entries. We use the symbol $X_{[j]}$ to mean the jth column of the matrix X. Given $n \times n$ matrices $X^1, \dots, X^k$ and an element $\mathcal{J}$ of $Q_{k,n}$, the symbol $A(\mathcal{J}; X^1, \dots, X^k)$ will stand for the matrix Z that is obtained from A by replacing the j_pth column of A by the j_pth column of X^p for $1 \leq p \leq k$, and keeping the rest of the columns unchanged. In other words $Z_{[j_p]} = X^p_{[j_p]}$ for all $j_1, \dots, j_k$ in $\mathcal{J}$, and $Z_{[\ell]} = A_{[\ell]}$ if the index ℓ does not occur in $\mathcal{J}$. The symbol $Y_{[\mathcal{J}]}$ will stand for the $n \times n$ matrix which has $Y_{[j_p]} = X^p_{[j_p]}$ for $1 \leq p \leq k$ and the rest of whose columns are zero. Let S_k be the set of all permutations on k symbols and let σ be a typical element of this set. We write $Y^\sigma_{[\mathcal{J}]}$ for the $n \times n$ matrix whose columns are described as $Y^\sigma_{[j_p]} = X^{\sigma(p)}_{[j_p]}$ for $1 \leq p \leq k$, and the remaining $n-k$ columns are zero.

Theorem 1 *For $1 \leq k \leq n$ we have*

$$D^k \det A(X^1, \dots, X^k) = \sum_{\sigma \in S_k} \sum_{\mathcal{J} \in Q_{k,n}} \det A(\mathcal{J}; X^{\sigma(1)}, \dots, X^{\sigma(k)}). \tag{6}$$

In particular,

$$D^k \det A(X, \dots, X) = k! \sum_{\mathcal{J} \in Q_{k,n}} \det A(\mathcal{J}; X, \dots, X). \tag{7}$$

Theorem 2 *For $1 \leq k \leq n$ we have*

$$D^k \det A(X^1, \dots, X^k) = \sum_{\sigma \in S_k} \sum_{\mathcal{I}, \mathcal{J} \in Q_{k,n}} (-1)^{|\mathcal{I}|+|\mathcal{J}|} \det A(\mathcal{I}|\mathcal{J}) \det Y^\sigma_{[\mathcal{J}]}[\mathcal{I}|\mathcal{J}]. \tag{8}$$

In particular,

$$D^k \det A(X, \dots, X) = k! \sum_{\mathcal{I}, \mathcal{J} \in Q_{k,n}} (-1)^{|\mathcal{I}|+|\mathcal{J}|} \det A(\mathcal{I}|\mathcal{J}) \det X[\mathcal{I}|\mathcal{J}]. \tag{9}$$

Very clearly (6) and (7) are generalisations of the formula (4); and (8) and (9) generalise (3). To describe an analogue of Jacobi's formula (2) we need more of notation.

Let $\mathcal{H}$ be an n-dimensional Hilbert space, and let $\bigotimes^k \mathcal{H} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}$ be its k-fold tensor product. We denote by $\wedge^k \mathcal{H}$ the space of antisymmetric tensors. If e_i, $1 \leq i \leq n$, is the standard basis for $\mathcal{H}$, then the standard basis for $\wedge^k \mathcal{H}$ consists of the vectors $e_{\mathcal{I}} = e_{i_1}, \wedge \ldots \wedge e_{i_k}$, where the index set $Q_{k,n}$ is ordered lexicographically. (See [2] Chapter I.) Given an operator A on $\mathcal{H}$ we use the symbol $\bigotimes^k A$ for its k-fold tensor product. This operator on $\bigotimes^k \mathcal{H}$ leaves invariant the subspace $\wedge^k \mathcal{H}$, and its restriction to this space is called the kth *exterior power,* kth *antisymmetric tensor power*, or the kth *compound* of A. With respect to the standard basis the $(\mathcal{I}, \mathcal{J})$ entry of $\wedge^k A$ is $\det A[\mathcal{I}|\mathcal{J}]$, the $k \times k$ minor of A corresponding to the rows of A in the index set $\mathcal{I}$ and the columns in $\mathcal{J}$.

The matrix $\operatorname{adj}(A)$ is the transpose of the matrix whose entries are $(-1)^{i+j} \det A(i,j)$, and can be identified with an operator on the space $\wedge^{n-1}\mathcal{H}$. We call this operator $\widetilde{\wedge}^{n-1} A$. It is unitarily equivalent to the operator $\wedge^{n-1} A$. (Put in other words $\operatorname{adj}(A)$ is a matrix representation of the operator $\wedge^{n-1} A$ but not in the standard orthonormal basis.) Likewise the transpose of the matrix with entries $(-1)^{|\mathcal{I}|+|\mathcal{J}|} \det A(\mathcal{I}|\mathcal{J})$ can be identified with an operator on the space $\wedge^{n-k}\mathcal{H}$. We call this operator $\widetilde{\wedge}^{n-k} A$, and note that it is unitarily equivalent to $\wedge^{n-k} A$.

Given operators $X^1, \ldots, X^k$ on $\mathcal{H}$ consider the operator

$$\frac{1}{k!} \sum_{\sigma \in S_k} X^{\sigma(1)} \otimes X^{\sigma(2)} \otimes \cdots \otimes X^{\sigma(k)}, \tag{10}$$

on the space $\bigotimes^k \mathcal{H}$. One can check that this leaves invariant the space $\wedge^k \mathcal{H}$, and we use the notation

$$X^1 \wedge X^2 \wedge \ldots \wedge X^k, \tag{11}$$

for the restriction of the operator (10) to the subspace $\wedge^k \mathcal{H}$.

Jacobi's formula (2) can be written also as

$$D \det A(X) = \operatorname{tr} \left(\widetilde{\wedge}^{n-1} A \right) X. \tag{12}$$

The next theorem is an extension of this.

Theorem 3 *For $1 \leq k \leq n$, we have*

$$D^k \det A(X^1, \ldots, X^k) = k! \operatorname{tr} \left[(\widetilde{\wedge}^{n-k} A)(X^1 \wedge \ldots, \wedge X^k) \right]. \tag{13}$$

In particular,

$$D^k \det A(X, \ldots, X) = k! \operatorname{tr} \left[(\widetilde{\wedge}^{n-k} A)(\wedge^k X) \right]. \tag{14}$$

Let $s_1(A) \geq \cdots \geq s_n(A)$ be the singular values of A, and let $\|A\| := s_1(A)$ be the *operator norm* of A. The norm of the linear operator $D \det A$ is defined as

$$\|D \det A\| = \sup_{\|X\|=1} \|D \det A(X)\|. \tag{15}$$

For $1 \leq k \leq n$ let $p_k(x_1, \ldots, x_n)$ be the kth elementary symmetric polynomial in n variables. A simple consequence of (12) is the equality

$$\|D \det A\| = p_{n-1}(s_1(A), \ldots, s_n(A)), \tag{16}$$

which in turn leads to inequality

$$\|D \det A\| \leq n\|A\|^{n-1}. \tag{17}$$

In [5] Bhatia and Friedland proved a more general theorem giving exact expressions for $\|D \wedge^k A\|$ for all $1 \leq k \leq n$. The special case $k = n$ reduces to the determinant. Using the theorems stated above we can extend (16) to all derivatives. We have the following.

Theorem 4 *Let A be an $n \times n$ matrix with singular values $s_1(A), \ldots, s_n(A)$. Then for $k = 1, 2, \ldots, n$, we have*

$$\|D^k \det A\| = k!\; p_{n-k}(s_1(A), \ldots, s_n(A)). \tag{18}$$

As a corollary we have the following perturbation bound obtained in a recent paper of Ipsen and Rehman [10].

Corollary 5 *Let A and X be $n \times n$ matrices. Then*

$$|\det(A + X) - \det A| \leq \sum_{k=1}^{n} p_{n-k}\left(s_1(A), \ldots, s_n(A)\right) \|X\|^k. \tag{19}$$

2 Proofs

The easiest approach to the derivative formulas is to prove Theorem 1 first and to derive the others from it. The quantity $\det(A+tX)$ is a polynomial in t, and from (1) it is clear that $D \det A(X)$ is the coefficient of t in this polynomial. The determinant is a linear function of each of its columns. Using this we obtain the equality (4) at once.

The same idea can be carried further. It is clear from the definition (5) that $D^k \det A(X^1, \ldots, X^k)$ is the coefficient of the term involving $t_1 t_2 \cdots t_k$ in the expansion of $\det(A + t_1 X^1 + \cdots + t_k X^k)$. Again we can identify this coefficient using the linearity of the det function with respect to each of the columns. The reader can check that the result is the formula (6).

We should note here the special case

$$D^n \det A(X, \ldots, X) = n!\; \det X, \tag{20}$$

and the fact that for $k > n$

$$D^k \det A(X, \ldots, X) = 0. \tag{21}$$

The formula (9) can be obtained from (7) using the *Laplace expansion formula* [11]. If T is any $n \times n$ matrix, and $\mathcal{J}$ any element of $Q_{k,n}$, then Laplace's formula says

$$\det T = (-1)^{|\mathcal{J}|} \sum_{\mathcal{I} \in Q_{k,n}} (-1)^{|\mathcal{I}|} \det T(\mathcal{I}|\mathcal{J}) \det T[\mathcal{I}|\mathcal{J}]. \tag{22}$$

Using this we get from (7)

$$\begin{aligned} & D^k \det A(X, \ldots, X) \\ = \;& k! \sum_{\mathcal{J} \in Q_{k,n}} (-1)^{|\mathcal{J}|} \sum_{\mathcal{I} \in Q_{k,n}} (-1)^{|\mathcal{I}|} \det A(\mathcal{I}|\mathcal{J}) \det X[\mathcal{I}|\mathcal{J}] \\ = \;& k! \sum_{\mathcal{I},\mathcal{J} \in Q_{k,n}} (-1)^{|\mathcal{I}|+|\mathcal{J}|} \det A(\mathcal{I}|\mathcal{J}) \det X[\mathcal{I}|\mathcal{J}]. \end{aligned}$$

This is the formula (9). In the same way, one obtains the expression (8) from (6).

There is another way of writing these formulas in terms of *mixed discriminants* [1]. If $X^1, \ldots, X^n$ are $n \times n$ matrices, then their mixed discriminant is defined as

$$\Delta(X^1, \ldots, X^n) = \frac{1}{n!} \sum_{\sigma \in S_n} \det \left[X^{\sigma(1)}_{[1]}, \ldots, X^{\sigma(n)}_{[n]} \right]. \tag{23}$$

When all $X^j = X$ we have

$$\Delta(X, \ldots, X) = \det X. \tag{24}$$

With this notation the formula (8) can be rewritten as

$$\begin{aligned} & D^k \det A(X^1, \ldots, X^k) \\ = \;& k! \sum_{\mathcal{I},\mathcal{J} \in Q_{k,n}} (-1)^{|\mathcal{I}|+|\mathcal{J}|} \det A(\mathcal{I}|\mathcal{J}) \Delta \left(X^1[\mathcal{I}|\mathcal{J}], \ldots, X^k[\mathcal{I}|\mathcal{J}] \right), \end{aligned} \tag{25}$$

and (6) as

$$\begin{aligned} & D^k \det A(X^1, \ldots, X^k) \\ = \;& \frac{n!}{(n-k)!} \Delta(A, \ldots, A, X^1, \ldots, X^k). \end{aligned} \tag{26}$$

This connection between derivatives and mixed discriminants is not surprising. It is well-known [7] that

$$\Delta(X^1, \ldots, X^n) = \frac{1}{n!} \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \det(t_1 X^1 + \cdots + t_n X^n). \tag{27}$$

With some manipulations the formula (26) can be derived from (27).

The matrix of the operator defined in (10) and (11) with respect to the standard basis of $\wedge^k \mathcal{H}$ has as its $(\mathcal{I}, \mathcal{J})$ entry the mixed discriminant

$$\Delta(X^1[\mathcal{I}|\mathcal{J}], \ldots, X^k[\mathcal{I}|\mathcal{J}]) \tag{28}$$

Therefore, the formula (25) can be restated as (13).

We have finished the proof of Theorems 1-3. Now we turn to norms. The *trace norm* of A is defined as

$$\|A\|_1 = s_1(A) + \cdots + s_n(A). \tag{29}$$

This is the dual of the operator norm [2 Ch.IV] and we have

$$\|A\|_1 = \max_{\|X\|=1} |\mathrm{tr}\, AX| \tag{30}$$

The singular values of $\wedge^k A$ are the products $s_{i_1}(A)\cdots s_{i_k}(A)$, $1 \leq i_1 < \cdots < i_k \leq n$. Hence

$$\| \wedge^k A\| = s_1(A)\cdots s_k(A), \tag{31}$$

and

$$\begin{aligned} \| \wedge^k A\|_1 &= \sum_{1\leq i_1 < \cdots < i_k \leq n} s_{i_1}(A)\cdots s_{i_k}(A) \\ &= p_k(s_1(A), \ldots, s_n(A)), \end{aligned} \tag{32}$$

where p_k is the kth elementary symmetric polynomial. The Jacobi formula (12) and these considerations immediately lead us to (16).

We have, by definition

$$\|D^k \det A\| = \max_{\|X^1\|=\cdots=\|X^k\|=1} \|D^k \det A(X^1, \ldots, X^k)\|. \tag{33}$$

The relation

$$\|X^1 \otimes X^2 \otimes \cdots \otimes X^k\| = \|X^1\|\ \|X^2\| \cdots \|X^k\|$$

is well-known. From this it follows that

$$\|X^1 \wedge X^2 \wedge \cdots \wedge X^k\| \leq 1$$

if $\|X^j\| = 1$ for all j. Using the generalised Jacobi formula (13) and the relation (30) one easily obtains Theorem 4.

Corollary 5 is a consequence of Taylor's theorem. Let f be a $p+1$ times differentiable function from a normed linear space X into Y. We use the notation $x^{(m)}$ for the m-tuple $(x, x, \ldots, x)$. Then we have the Taylor expansion

$$f(a+x) = f(a) + \sum_{k=1}^{p} \frac{1}{k!}\ D^k f(a)(x^{(m)}) + O\left(\|x\|^{p+1}\right).$$

Applying this to the determinant we obtain the inequality (19) from (18). This inequality is sharp even in the simplest commutative case. Let $A = I$ and $X = xI$. Then

$$\det(A+X) - \det A = nx + \binom{n}{2} x^2 + \cdots + x^n.$$

The coefficient of the term x^k here is

$$\binom{n}{k} = p_{n-k}(1, \ldots, 1).$$

3 Remarks

1. By now there are several examples of situations where the norm of the derivative of a matrix function turns out to be the same as that of the corresponding scalar function. See e.g. [6] for a sampler.

2. The expression (10) makes it transparent that the operator $X^1 \wedge \ldots \wedge X^k$ is positive semidefinite (p.s.d.) if all X^j are. Therefore the mixed discriminant of p.s.d. matrices is nonnegative, a well-known fact [1].

3. Let A be p.s.d. Then the expression (13) shows that $D^k \det A(X^1, \ldots, X^k)$ is nonnegative whenever $X^1, \ldots, X^k$ are p.s.d. In other words $D^k \det A$ is a positive linear functional [4] and by well-known theorems $\|D^k \det A\| = \|D^k \det A(I, \ldots, I)\|$. That is the underlying reason for the disappearance of noncommutativity mentioned in Remark 1.

4. Hidden behind our proof of Theorem 1 is a formula for $\det(A + X)$. It says

$$\det(A+X) = \sum_{k=0}^{n} \sum_{\mathcal{J} \in Q_{k,n}} \det A \left(\mathcal{J}; \underbrace{X, \ldots, X}_{k \text{ times}} \right). \tag{34}$$

   i.e.,

$$\det(A+X) = \det A + c_1(A,X) + \cdots + c_{n-1}(A,X) + \det X, \tag{35}$$

   where $c_k(A,X)$ is a sum of $\binom{n}{k}$ determinants that result from A by replacing k of its columns by the corresponding columns of X.

5. Perturbation bounds for the determinant have been considered by Ipsen and Rehman [10] and there is considerable common ground between that paper and ours. Our main interest here has been finding formulas for the kth order Fréchet derivatives of the det function. Perturbation bounds follow as a consequence. The authors of [10] take a different approach. Using the singular value decomposition they assume A is diagonal, establish an expansion like (35) in this special case and use it to get the inequality (19). The introduction of derivatives may give added insight into these inequalities.

6. Perturbation bounds for eigenvalues can be obtained from those for the determinant. See [2,3,5] for some methods that are useful.

We thank Professor R. B. Bapat for pointing out the connection between our problem and mixed discriminants. The first author is supported by a J. C. Bose National Fellowship.

References

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

[2] R. Bhatia, *Matrix Analysis,* Springer, 1997.

[3] R. Bhatia, *Perturbation Bounds for Matrix Eigenvalues,* SIAM, 2007, (expanded reprint of 1987 edition).

[4] R. Bhatia, *Positive Definite Matrices,* Princeton University Press, 2007.

[5] R. Bhatia and S. Friedland, *Variation of Grassman powers and spectra,* Linear Algebra Appl., 40 (1981) 1-18.

[6] R. Bhatia, D. Singh, and K. B. Sinha *Differentiation of operator functions and perturbation bounds,* Commun. Math. Phys. 191 (1998) 603-611.

[7] L. Gurvits, *The Van der Waerden conjecture for mixed discrimants,* Adv. Math., 200 (2006) 435-454.

[8] N. J. Higham, *Functions of Matrices,* SIAM, 2008.

[9] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus,* Wiley, 1988.

[10] I.C.F. Ipsen and R. Rehman, *Perturbation bounds for determinants and characteristic polynomials,* SIAM J. Matrix Anal., Appl., 30 (2008) 762-776.

[11] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities,* Dover, 1992, (reprint of 1964 edition).