Higher Order Logarithmic Derivatives of Matrices

Rajendra Bhatia
and
Ludwig Elsner

April 10, 2002
isid/ms/2002/03

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

Higher Order Logarithmic Derivatives of Matrices

Rajendra Bhatia
and
Ludwig Elsner

1. Introduction

Let A be an $n \times n$ complex matrix and $||A||$ its norm as a linear operator on the Euclidean space $\mathbb{C}^n$; i.e.,

$$||A|| = \sup\,\{||Ax||_2 : x \in \mathbb{C}^n,\ ||x||_2 = 1\}, \tag{1}$$

where $||x||_2$ is the Euclidean norm of the vector x . The initial–value problem

$$\dot{x}(t) = Ax(t),\ x(0) = x_0, \tag{2}$$

has the solution

$$x(t) = e^{tA}\,x_0. \tag{3}$$

For many purposes — such as error bounds — one needs upper bounds for the quantity $||e^{tA}||$. A very useful bound is given in terms of the **logarithmic derivative** of A defined as

$$\mu(A) = \lim_{h \to 0^+} \frac{||e^{hA}|| - 1}{h}. \tag{4}$$

We have

$$||e^{tA}|| \ \leq\ e^{\mu(A)t} \ \text{ for all } \ t \geq 0, \tag{5}$$

and $\mu(A)$ is the smallest number for which such an inequality holds. We know that

$$\mu(A) = \lambda_1\left(\frac{A + A^*}{2}\right), \tag{6}$$

where $\lambda_1(H)$ denotes the maximum eigenvalue of a Hermitian matrix A. See [1, 4].

In a recent paper [3], L. Kohaupt has studied the problem of finding the second logarithmic derivative, and solved it when the operator norm is induced not by the Euclidean norm as in our definition (1) but by the p–norm where $p = 1$ or ∞. In this note we resolve the problem for $p = 2$. The somewhat unexpected answer led us to investigate the third derivative as well. We prove the following

Theorem 1: Let $\varphi(t) = ||e^{tA}||,\ t \geq 0$, and let $\dot{\varphi}(0),\ \ddot{\varphi}(0),\ \dddot{\varphi}(0)$ denote the first three right derivatives of φ at 0. Then

$$\ddot{\varphi}(0) = \dot{\varphi}(0)^2. \tag{7}$$

Let $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of $A + A^*$ and $x_1, \ldots, x_n$ the corresponding eigenvectors.

Then

$$\dddot{\varphi}(0) = \dot{\varphi}(0)^3 - \frac{1}{4} \sum_{j=2}^{n} (\lambda_1 - \lambda_j)\ | < x_j, Ax_1 > |^2. \tag{8}$$

Note that $\dot{\varphi}(0)$ is just $\mu(A)$. The equality (7) is a little surprising and does not persist when we go to the third derivative. Our proof of (8) requires the assumption that the eigenvalue λ_1 is simple. It might be possible to drop this requirement.

2. Proofs

To handle higher order terms we need an extension of a standard perturbation result. The discussion in the next paragraph is modelled on that in [5, p. 69].

Consider the eigen equation

$$(A + \epsilon B + \epsilon^2 C)\ x_1(\epsilon) = \lambda_1(\epsilon)\ x_1(\epsilon), \tag{9}$$

where A, B, C are Hermitian, and $\lambda_1(0) = \lambda_1$ is a simple eigenvalue of A. Then we have a series expansion

$$\lambda_1(\epsilon) = \lambda_1 + \epsilon k_1 + \epsilon^2 k_2 + \cdots. \tag{10}$$

Let $x_1,\ x_2, \ldots, x_n$ be the eigenvectors of A corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. The vector $x_1(\epsilon)$ has a series expansion

$$x_1(\epsilon) = x_1 + (\epsilon t_{21} + \epsilon^2 t_{22} + \cdots)\ x_2 + \ldots + (\epsilon t_{n1} + \epsilon^2 t_{n2} + \cdots)\ x_n. \tag{11}$$

The coefficients k_1 and k_2 are found as follows. Combine (9), (10), (11) and equate the first order terms in ϵ to get

$$A(t_{21}x_2 + t_{31}x_3 + \cdots + t_{n1}x_n) + Bx_1 = \lambda_1(t_{21}x_2 + t_{31}x_3 + \cdots + t_{n1}x_n) + k_1x_1.$$

Taking inner product of both sides with x_1 , we get

$$< x_1, B\ x_1 > = k_1, \tag{12}$$

while taking inner products with $x_j,\ j \geq 2$, we get (using $Ax_j = \lambda_j x_j$)

$$t_{j1}\lambda_j + < x_j, Bx_1 > = \lambda_1 t_{j1}, \quad \text{or}$$

$$t_{j1} = \frac{< x_j, Bx_1 >}{\lambda_1 - \lambda_j},\ j \geq 2. \tag{13}$$

Again, using (9), (10), (11) and equating second order terms in ϵ , we get

$$A\left(\sum_{j=2}^{n} t_{j2}x_j\right) + B\left(\sum_{j=2}^{n} t_{j1}x_j\right) + C\ x_1 = \lambda_1\left(\sum_{j=2}^{n} t_{j2}x_j\right) + k_1\left(\sum_{j=2}^{n} t_{j1}x_j\right) + k_2x_1.$$

Taking inner product of both sides with x_1 , we get

$$\sum_{j=2}^{n} t_{j1}\ < x_1, Bx_j > + < x_1, Cx_1 > = k_2,$$

and then substituting for t_{j1} from (13)

$$\sum_{j=2}^{n} \frac{|< x_j, Bx_1 >|^2}{\lambda_1 - \lambda_j} + < x_1, Cx_1 > = k_2. \tag{14}$$

The information contained in (10) and (12) is often written as

$$\lambda_1(A + \epsilon B) = \lambda_1 + \epsilon < x_1, Bx_1 > + O(\epsilon^2), \tag{15}$$

where x_1 is the normalised eigenvector corresponding to the simple eigenvalue λ_1 of A . More generally, when λ_1 is not a simple eigenvalue, we have

$$\lambda_1(A + \epsilon B) = \lambda_1 + \epsilon \max_{x \in M,\ ||x||=1} < x, Bx > +O(\epsilon^2), \tag{16}$$

where M is the eigenspace corresponding to the eigenvalue λ_1 of A . See, e.g., equation (3.8) in [2].

Now for any matrix A consider the function

$$g(t) = \varphi(t)^2 = ||e^{tA}||^2 = \lambda_1(e^{tA}\ e^{tA^*}).$$

Then

$$\dot{g}(t) = \lim_{h\to 0} \frac{1}{h} \left[\lambda_1(e^{tA}(I+hA)\ (I+hA^*)\ e^{tA*}) - \lambda_1(e^{tA}\ e^{tA^*}\right]$$
$$= \lim_{h\to 0} \frac{1}{h} \left[\lambda_1(e^{tA}\ e^{tA^*} + h\ e^{tA}(A+A^*)\ e^{tA^*}) - \lambda_1(e^{tA}\ e^{tA^*})\right].$$

Using (16) we get

$$\dot{g}(t) = \max_{x\in M,\ ||x||=1} < x, e^{tA}(A+A^*)e^{tA^*}\ x >, \tag{17}$$

where M is the eigenspace of $e^{tA}\ e^{tA^*}$ corresponding to the largest eigenvalue. When $t=0$, this is the entire space $\mathbb{C}^n$, and hence

$$\dot{g}(0) = \max_{||x||=1} < x, (A+A^*)x > = \lambda_1(A+A^*). \tag{18}$$

Since

$$\dot{\varphi}(t) = \frac{\dot{g}(t)}{2\varphi(t)}, \tag{19}$$

this gives the known result

$$\mu(A) = \dot{\varphi}(0) = \lambda_1\left(\frac{A+A^*}{2}\right).$$

From (17) and (19) we have

$$\dot{\varphi}(t) = \frac{\varphi(t)\ \dot{g}(t)}{2\ \varphi(t)^2} = \varphi(t) \max_{x\in M,\ ||x||=1} \frac{< x, e^{tA}(A+A^*)e^{tA^*}x >}{2 < x, e^{tA}\ e^{tA^*}x >}$$

$$\leq \varphi(t)\ \lambda_1\left(\frac{A+A^*}{2}\right) = \varphi(t)\ \dot{\varphi}(0).$$

This implies that

$$\varphi(t) \leq e^{\dot{\varphi}(0)t} \text{ for all } t \geq 0,$$

which is the known result (5).

To calculate the second and the third derivatives we need the following

Lemma: Let x be a (normalised) eigenvector of $A+A^*$. Then

$$< x, A^*\ f(A+A^*)Ax > = < x, A\ f(A+A^*)A^*\ x >, \tag{20}$$

for every function f. In particular,

$$< x, A^*A\ x > = < x, AA^*\ x >, \tag{21}$$

(22) $$< x, A^*(A+A^*)A\ x > \; = \; < x, A(A+A^*)A^*\ x > .$$

Proof: Choose an orthonormal basis consisting of eigenvectors of $A+A^*$ starting with x. Let $A = B+iC$, where $B = \frac{1}{2}\,(A+A^*)$, $C = \frac{1}{2i}\,(A-A^*)$. In the basis we have chosen, let $b_1,\ldots,b_n$ be the diagonal entries of B, and let a_{ij} be the entries of A. The two sides of (20) are the $(1,1)$ entries of the matrices $A^*\ f(A+A^*)A$ and $A\ f(A+A^*)A^*$, respectively. A simple calculation shows each of them is equal to $\sum\limits_{j=1}^{n} f(2b_j)\ |a_{1j}|^2$.

∎

Proof of Theorem 1: We have

$$\begin{aligned} g(h) = \|e^{hA}\|^2 = \lambda_1(e^{hA}\ e^{hA^*}) &= \lambda_1\left[(I+hA+\frac{h^2}{2}\ A^2)\ (I+hA^*+\frac{h^2}{2}\ A^{*2})\right] + O(h^3) \\ &= 1 + h\ \lambda_1\left[(A+A^*) + \frac{h}{2}\ (2AA^* + A^2 + A^{*2})\right] + O(h^3). \end{aligned}$$

Using (16) we get from this

$$g(h) = 1 + h\ \lambda_1(A+A^*) + \frac{h^2}{2} \max_{x\in M,\ ||x||=1} < x, (2AA^* + A^2 + A^{*2})x > + O(h^3),$$

where M is the eigenspace of $A+A^*$ corresponding to its largest eigenvalue. Now using (21) we see that for every $x \in M$

$$< x, (2AA^* + A^2 + A^{*2})x > \; = \; < x, (A+A^*)^2 x > \; = \; \lambda_1^2(A+A^*).$$

This shows that

(23) $$\ddot{g}(0) = \lambda_1^2(A+A^*).$$

Since $\ddot{g}(t) = 2\ \dot{\varphi}(t)^2 + 2\varphi(t)\ \ddot{\varphi}(t)$, we have

$$\ddot{\varphi}(0) \;=\; \frac{\ddot{g}(0) - 2\dot{\varphi}(0)^2}{2\varphi(0)}.$$

Substituting the values of $\ddot{g}(0)$ and $\dot{\varphi}(0)$ from (23) and (6) here we get

$$\ddot{\varphi}(0) = \lambda_1^2\left(\frac{A+A^*}{2}\right) \;=\; \dot{\varphi}(0)^2.$$

This proves (7).

To study the third derivative write out the expansion of $g(h)$ as

$$g(h) = 1 + h\ \lambda_1(\widetilde{A} + h\widetilde{B} + h^2\ \widetilde{C}) + O(h^4),$$

where

$$\begin{aligned}
\widetilde{A} &= A + A^*, \\
\widetilde{B} &= \frac{1}{2}\ (A^2 + A^{*2} + 2AA^*), \\
\widetilde{C} &= \frac{1}{6}\ \{A^3 + A^{*3} + 3A(A+A^*)A^*\}.
\end{aligned} \tag{24}$$

From (9), (10), (12) and (14) we know that

$$\lambda_1(\widetilde{A} + h\widetilde{B} + h^2\widetilde{C}) = \lambda_1(\widetilde{A}) + h\ k_1 + h^2\ k_2 + O(h^3),$$

where

$$k_2 \ = < x_1, \widetilde{C}x_1 > + \sum_{j=2}^{n} \frac{|< x_j, \widetilde{B}x_1 >|^2}{\lambda_1 - \lambda_j}, \tag{25}$$

λ_j being the eigenvalues of $\widetilde{A} = A + A^*$, and x_j the corresponding eigenvectors. To calculate the second term in (25) note that

$$\begin{aligned}
< x_i, \widetilde{B}x_j > \ &= \ \frac{1}{2}\ < x_i, (A^2 + A^{*2} + 2AA^*)x_j > \\
&= \ \frac{1}{2}\ < x_i, [(A+A^*)^2 + A(A+A^*) - (A+A^*)A]x_j > \\
&= \ \frac{1}{2}\ \{\lambda_i^2\ \delta_{ij} + (\lambda_j - \lambda_i)\ \ < x_i, Ax_j >\}.
\end{aligned}$$

Hence,

$$\sum_{j=2}^{n} \frac{|< x_j, \widetilde{B}x_1 >|^2}{\lambda_1 - \lambda_j} \ = \ \frac{1}{4} \sum_{j=2}^{n} (\lambda_1 - \lambda_j)\ |< x_j, Ax_1 >|^2. \tag{26}$$

To calculate the first term in (25) note that

$$\begin{aligned}
6\ \widetilde{C} &= (A + A^*)^3 + A(AA^* - A^*A) + (AA^* - A^*A)A^* \\
&\quad + \ [A(A+A^*)A^* - A^*(A+A^*)A].
\end{aligned} \tag{27}$$

If W is the term inside the square brackets in (27), then by (22)

$$< x_1, W\ x_1 > = \ 0. \tag{28}$$

Further, note that

$$
\begin{aligned}
< x_1, A(AA^* - A^*A)x_1 > &= < A^*x_1, [(A+A^*)A^* - A^*(A+A^*)]x_1 > \\
&= < A^*x_1, (A + A^* - \lambda_1\ I)A^*x_1 > \\
&= < A^*x_1, \left(\sum_{j=2}^{n} (\lambda_j - \lambda_1) x_j x_j^* \right) A^*x_1 > \\
&= \sum_{j=2}^{n} (\lambda_j - \lambda_1)\ | < x_j, A^*x_1 > |^2 \\
&= \sum_{j=2}^{n} (\lambda_j - \lambda_1)\ | < x_j, Ax_1 > |^2.
\end{aligned} \tag{29}
$$

(In the last step we have used the fact that x_j are eigenvectors of $A + A^*$). This shows also that

$$
< x_1, (AA^* - A^*A)A^*x_1 > = \sum_{j=2}^{n} (\lambda_j - \lambda_1)\ | < x_j, Ax_1 > |^2. \tag{30}
$$

Equations (26) — (30) show that

$$
< x_1, \widetilde{C}x_1 > = \frac{1}{6} \left\{ \lambda_1^3 + 2 \sum_{j=2}^{n} (\lambda_j - \lambda_1)\ | < x_j, Ax_1 > |^2 \right\}. \tag{31}
$$

From (25), (26) and (31) we obtain

$$
6\ k_2 = \lambda_1^3 - \frac{1}{2} \sum_{j=2}^{n} (\lambda_1 - \lambda_j)\ | < x_j, Ax_1 > |^2. \tag{32}
$$

This is then the value of $\dddot{g}(0)$. Since

$$
\dddot{\varphi}(0) \ = \ \frac{\dddot{g}(0) - 6\ \dot{\varphi}(0)\ddot{\varphi}(0)}{2},
$$

we obtain the equality (8) from the expressions already derived for $\dot{\varphi}(0)$ and $\ddot{\varphi}(0)$. ∎

3. Remarks

1. We have proved (7) without the assumption that λ_1 is a simple eigenvalue of $A + A^*$. The proof is facilitated by the first order expansion (16). We do not know of an analogous second order expansion when λ_1 is a multiple eigenvalue. This compels us to assume λ_1 is simple while proving (8). We believe this assumption is not necessary.

2. The inequality (5) says
$$\varphi(t) \ \leq \ e^{\dot{\varphi}(0)t}.$$
Because of (6) and (7) we know that
$$e^{\dot{\varphi}(0)t} - \varphi(t) = O(t^3),$$
and (8) tells us no further improvement is possible in general.

3. When the maximum eigenvalue of $A + A^*$ is simple we can get conditions for equality in (5) using our result 8:

Proposition: Suppose $\lambda_1(A + A^*)$ is a simple eigenvalue of $A + A^*$. Then the following conditions are equivalent:

(i) $||e^{tA}|| = e^{\mu(A)t}$ for all $t \geq 0$.

(ii) $||e^{hA}|| = e^{\mu(A)h}$ for some $h > 0$.

(iii) The eigenvector x of $A+A^*$ corresponding to λ_1 is also an eigenvector of A .

Proof: Clearly (i) $\Rightarrow$ (ii). If (ii) holds for some $h > 0$, then for all natural numbers m
$$||e^{h/m \ A}|| = e^{\mu(A)h/m}$$
because of submultiplicativity of the norm. Since $\dot{\varphi}(0) = \mu(A), \ \ddot{\varphi}(0) = \mu(A)^2$, we have from (8)
$$\sum_{j=2}^{n} (\lambda_1 - \lambda_j) \ | < x_j, Ax_1 > |^2 = 0.$$
Since $\lambda_j \neq \lambda_1$ for $j \geq 2$, this implies $< x_j, Ax_1 > = 0$. Hence $A \, x_1$ is a multiple of x_1 . Thus the statement (iii) is true if (ii) is.

Now suppose (iii) holds. If $Ax_1 = \lambda x_1$, then $A^* x_1 = \bar{\lambda} x_1$ and $\lambda_1 = \lambda + \bar{\lambda}$. In the orthonormal basis $x_1, \ldots, x_n$ we can write
$$A \ = \ \begin{bmatrix} \lambda & 0 \\ 0 & A_1 \end{bmatrix}.$$
Note that $\mu(A_1) \leq \mu(A) = \lambda_1/2 = \text{Re} \ \lambda$. Hence
$$||e^{tA}|| = \ \max \ (|e^{t\lambda}|, \ ||e^{tA_1}||) = e^{\mu(A)t}.$$
Thus (i) is true if (iii) is. ■

The second author thanks the National Board for Higher Mathematics (India) and the Indian Statistical Institute for sponsoring a visit in February, 2001 when this work was done.

References

1. G. Dahlquist, *Stability and error bounds in the numerical integration of ordinary differential equations*, Trans. Royal Institute of Technology, No. 130, Stockholm, 1959.

2. J.–B. Hiriart–Urruty and D. Ye, *Sensitivity analysis of all eigenvalues of a symmetric matrix*, Numer. Math. 70, (1995) 45–72.

3. L. Kohaupt, *Second logarithmic derivative of a complex matrix in the Chebyshev norm*, SIAM J. Matrix Anal. Appl. 21, (1999) 382–389.

4. S.M. Lozinskii, *Error estimates for the numerical integration of ordinary differential equations*, I. Izv. Vyss, Ucebn. Zaved. Matematika, 5 (6), (1958) 52–90, (in Russian).

5. J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, 1965.

Indian Statistical Institute
New Delhi - 110016, India
rbh@isid.ac.in

Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
D-33501 Bielefeld
Germany
elsner@mathematik.uni-bielefeld.de