NORMS OF DERIVATIVES OF MATRIX FUNCTIONS

Rajendra Bhatia

X, Y Banach spaces.

$E \subset X$ open set

$f : E \to Y$ is differentiable at x_0

if there exists a linear map

$Df(x_0) : X \to Y$

such that

$$f(x_0 + h) = f(x_0) + Df(x_0)(h) + o(h).$$

The <u>derivative</u> $Df(x_0)$ is given by

$$Df(x_0) = \frac{d}{dt}\Big|_{t=0} f(x_0 + th).$$

$$\| Df(x_0) \| := \sup_{\|h\|=1} \| Df(x_0)(h) \| .$$

$$\| f(x_0 + h) - f(x_0) \| \leq \| Df(x_0) \| \, \|h\| + o(\|h\|).$$

(A "first order perturbation bound").

Higher derivatives. Taylor Expansion:

$$f(x_0 + h) = f(x_0) + Df(x_0)(h) + D^2 f(x_0)(h, h) + \cdots$$

X, Y spaces of matrices/operators

$\|\cdot\|$ operator norm / "spectral norm"

$\|\cdot\|_2$ Frobenius norm

$f(A)$ could be

polynomials exp

$\otimes^k A$ $\quad \wedge^k A$ $\quad \det(A)$

indirectly defined $A = UP$, QR, LR.

For this talk

$f : I \longrightarrow \mathbb{R}$ induced map

$f : \mathbb{H}_n(I) \longrightarrow \mathbb{H}_n$

$I = \mathbb{R}_+ = (0, \infty)$

$f : \mathbb{P}_n \longrightarrow \mathbb{P}_n$ (pos. def.)

(Special interest $f(t) = t^r$ $r \in \mathbb{R}$)

$\mathbb{P}_n$ is open set in $\mathbb{H}_n$

$Df(A) : \mathbb{H}_n \longrightarrow \mathbb{H}_n$.

f' (usual) derivative of the real fn. f

$$Df(A)(B) := \frac{d}{dt}\bigg|_{t=0} f(A+tB)$$

$$= f'(A)B \qquad \underline{\text{if } A, B \text{ commute}}$$

Every scientist should once in a while try a damned fool experiment.

(Littlewood's Miscellany).

Could we have

$$\| Df(A) \| = \| f'(A) \| \quad ? \tag{1}$$

i.e.,

$$\sup_{\|B\|=1} \| Df(A)(B) \| = \sup_{\substack{\|B\|=1 \\ AB=BA}} \| Df(A)(B) \| \quad ?$$

<u>Def.</u> $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$ is in class $\mathcal{D}$ if (1) is true for f.

Is $\mathcal{D}$ nonempty?

Interesting examples of $f \in \mathcal{D}$?

Characterisation of $\mathcal{D}$?

Daleckii - Krein Formula

Choose an o.n.b. in which

$$A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Let $f^{[1]}(A)$ be the matrix with i,j entry

$$\frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} \qquad i \neq j$$

$$f'(\lambda_i) \qquad i = j$$

Then

$$Df(A)(B) = f^{[1]}(A) \circ B \qquad \text{(Schur prod.)}$$

$$:= \left[\frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} \; b_{ij} \right] \qquad (2)$$

So for the Frobenius norm

$$\| Df(A)(B) \|_2 \leq \max_{i,j} \left| \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} \right| \; \|B\|_2$$

By the Mean Value Theorem

$$\| Df(A) \|_{2 \to 2} \leq \| f' \|_\infty .$$

(Not so simple for $\|\cdot\|$ — no good dependence on matrix entries.)

Notation: $X \geq 0$ positive (semidefinite)

Def. f operator monotone if

$$A \geq B \implies f(A) \geq f(B).$$

Loewner's Theorem

f oper. monotone $\iff f^{[1]}(A) \geq 0 \quad \forall A.$

Norm of $A \circ B$ complicated in general, but

Schur's Theorem If $A \geq 0$, then

$$\| A \circ B \| \leq \max_i a_{ii} \; \|B\|.$$

(Subsumed in Russo-Dye Theorem.
See R.B. Positive Definite Matrices.)

Combining Loewner's, Schur's and D.-K formula (2):

If f is operator monotone, then

$$\| Df(A) \| \leq \max_i |f'(\lambda_i)| = \|f'(A)\|$$

Thus f oper. monotone $\implies f \in \mathcal{D}$

Fact $f(t) = t^r$ op. mon. $\iff r \in [0,1]$.

$$f(A) = A^m \qquad m = 1, 2, 3, \dots$$

Matrix version of $(x^m)' = m x^{m-1}$

$$DA^m(B) = \sum_{\substack{j,k \geq 0 \\ j+k = m-1}} A^j B A^k$$

$$\rightarrow \quad \| DA^m \| \leq \sum_{j+k=m-1} \|A\|^j \|A\|^k = m \|A\|^{m-1}$$

$$= \| m A^{m-1} \| = \| f'(A) \|.$$

i.e. $f \in \mathcal{D}$.

Matrix version of $(e^x)' = e^x$

$$De^A(B) = \int_0^1 e^{(1-t)A} \, B \, e^{tA} \, dt$$

$$\rightarrow \quad \| De^A(B) \| \leq \int_0^1 \| e^{(1-t)A} \| \, \|B\| \| e^{tA} \| \, dt$$

$$= \int_0^1 e^{(1-t)\|A\|} \, e^{t\|A\|} \, dt \; \|B\|$$

$(A \geq 0)$

$$= e^{\|A\|} \, \|B\|$$

$$\rightarrow \quad \| De^A \| = e^{\|A\|} = \| e^A \| = \| \exp'(A) \|.$$

So $\exp \in \mathcal{D}$

Neither A^m $(m > 1)$ nor exp is op. monotone

$$f(A) = A^{-r} \qquad r > 0$$

Then $f \in \mathcal{D}$.

Idea of Proof:

$$A^{-r} = \frac{1}{\Gamma(r)} \int_0^\infty e^{-\lambda A} \lambda^{r-1} \, d\lambda$$

$$=: \int_0^\infty e^{-\lambda A} \, d\mu(\lambda)$$

Use the property of exp.

(Generally: if $f(t) = \int_0^\infty e^{-\lambda t} \, d\mu(\lambda)$, then $f \in \mathcal{D}$.)

Another argument using integrals shows

$$f(A) = A^{r} \qquad r \geq 2 \qquad \text{is in } \mathcal{D}.$$

$$\left(t^{r+2} = t^2 \int_0^\infty e^{-\lambda/t} \, d\mu(\lambda), \quad r > 0 \right)$$

Thus $f(t) = t^r$ is in $\mathcal{D}$ for

$$t \in (-\infty, 1] \cup [2, \infty)$$

Further analysis shows that for

$1 < r < \sqrt{2}$

$f(t) = t^r$ is <u>not</u> in $\mathfrak{D}$.

(The failure occurs for 2×2 matrices.)

More surprising (?):

For $r \in [\sqrt{2}, 2]$

$f(t) = t^r$ <u>is</u> in $\mathfrak{D}$.

Ideas involved in proofs:

$$\|A\|_S := \sup_{\|X\|=1} \|A \circ X\|$$

If $A \geqslant 0$ $\|A\|_S = \max_i a_{ii}$

In general $\|A\|_S$ is difficult;

expression due to Haagerup:

$$\|A\|_S = \min \{ \|X\|_{col} \|Y\|_{col} : A = X^* Y \}.$$

Special things happen for $1 < r < 2$.

2×2 case is difficult enough.

Let $A = \begin{bmatrix} 1 & 0 \\ 0 & p \end{bmatrix}$ $\qquad 0 < p < 1$

$f(t) = t^r \qquad r > 1$

$$\to \quad f'(A) = \begin{bmatrix} r & \\ & r p^{r-1} \end{bmatrix},$$

$$\|f'(A)\| = r$$

$$f^{[1]}(A) = \begin{bmatrix} r & \dfrac{1-p^r}{1-p} \\ \dfrac{1-p^r}{1-p} & r p^{r-1} \end{bmatrix} =: r S_r$$

Q: When is $\|S_r\|_S = 1$? (Then $f \in \mathcal{D}$)

Using Haagerup's Theorem one gets

<u>Thm</u> $f \in \mathcal{D}$ iff

$$\left(\frac{1}{r} \frac{1-p^r}{1-p} \right)^2 \leq \frac{1+p^{r-1}}{2}, \quad 0<p<1$$

(*)

1. Let $p \downarrow 0$ in $(*)$

$r \geq \sqrt{2}$ necessary for $f \in \mathcal{D}$

(So: $1 < r < \sqrt{2}$ $f(t) = t^r$ not in $\mathcal{D}$).

2. Cases

$r \geq 2$, $\frac{3}{2} \leq r \leq 2$

by fairly standard analysis

$f \in \mathcal{D}$ for these values.

Very indirect / unusual argument to show that for $\sqrt{2} < r < 3/2$

$$\left(\frac{1 - p^r}{r(1-p)} \right)^2 \leq \frac{1 + p^{r-1}}{2} \qquad \forall\, 0<p<1.$$

General $(n > 2)$ case:

Interesting Theorem

Loewner: For $0 < r < 1$

$$f^{[1]}(A) := \frac{\lambda_i^r - \lambda_j^r}{\lambda_i - \lambda_j}$$

is positive. (all eigenvalues positive)

For $1 < r < 2$ this matrix has one positive and $n-1$ negative eigenvalues.

For $n > 2$ this fact is used to find a Haagerup decomposition of $f^{[1]}(A)$ for evaluation of $\|f^{[1]}(A)\|_s$.

Problem of general interest

Inertia of the matrix

$$Q_r := \left[\frac{\lambda_i^r - \lambda_j^r}{\lambda_i - \lambda_j} \right]$$

$0 < r < 1$ $\qquad$ $\operatorname{In} Q_r = (n, 0, 0)$

$1 < r < 2$ $\qquad$ $\operatorname{In} Q_r = (1, 0, n-1)$

$2 < r < 3$ $\qquad$?

Higher derivatives : $\mathcal{D}_k$

$\mathcal{D}$ not closed under products

not closed under sums

($t + 1/t$ is not in $\mathcal{D}$.)