МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2008
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2008
Proceedings of the Thirty Seventh Spring Conference of
the Union of Bulgarian Mathematicians
Borovetz, April 2–6, 2008

HIGHER ORDER FRÉCHET DERIVATIVES OF MATRIX POWER FUNCTIONS*

Mihail Mihaylov Konstantinov, Juliana Kostadinova Boneva, Petko Hristov Petkov, Vladimir Todorov Todorov

We study higher order Fréchet derivatives of matrix power functions $X \mapsto X^p$, $p \in \mathbb{Q}$, $X \in \mathbb{C}^{n\times n}$. The results obtained may be applied to the accuracy estimation of Taylor approximations of matrix power functions as well as to the perturbation analysis of non–linear matrix equations.

Introduction and notations. In a recent paper [4] we have studied some properties of first order Frechét derivatives of the function $X \mapsto f_p(X) = X^p$ for rational p. In this paper we consider second and higher order Fréchet derivatives of f_p for rational p, different from 0 and 1. Higher order Fréchet derivatives when p is a positive integer or the reciprocal of such an integer have been considered in [1].

The results obtained below are applicable to the analysis of fractional-affine matrix equations as well as to the estimation of the accuracy of the Taylor approximations of $(A+H)^p$ for small H.

We use the following notations: $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ – the sets of natural, integer, rational, real and complex numbers, respectively; $\mathrm{i} = \sqrt{-1}$; $|z|$ and $\arg z \in (-\pi, \pi]$ – the module and the principal argument of $0 \neq z \in \mathbb{C}$ such that $z = |z|\mathrm{e}^{\mathrm{i}\arg z}$; $\ln z = \ln|z| + \mathrm{i}\arg z$ – the principal natural logarithm of $z \neq 0$. For any $z \neq 0$ and $c \in \mathbb{C}$ we define the principal c–th power of z as $z^c = \mathrm{e}^{c\ln z}$.

For matrices the notations are as follows: $\mathbb{C}^{n\times n}$ – the space of $n\times n$ matrices over $\mathbb{C}$; I_n – the identity $n \times n$ matrix; $\mathbb{C}_*^{n\times n} \subset \mathbb{C}^{n\times n}$ – the set of matrices X with no negative eigenvalues, i.e. $\mathrm{spect}(X) \cap (-\infty, 0] = \emptyset$; $\mathcal{L}(n)$ – the space of linear matrix operators $\mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$.

If $\mathcal{F} : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ is a differentiable operator we denote by $\mathcal{F}^{(m)}$ the m–th iteration of $\mathcal{F}$, while $\mathcal{F}^m(H) = (\mathcal{F}(H))^m$. The m–th Fréchet derivative of $\mathcal{F}$ at a given point is an m–linear form denoted as $\mathrm{D}^m\mathcal{F}(H_1, H_2, \ldots, H_m)$. When $H_1 = H_2 = \cdots = H_m = H$, this derivative is written as $\mathrm{D}^m\mathcal{F}(H)$ and we have $\|\mathrm{D}^m\mathcal{F}(H)\| = \mathrm{O}(\|H\|^m)$.

The abbreviation ":=" stands for "equal by definition".

Fréchet derivatives of functions $\mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$. Let $f : \mathcal{D} \to \mathbb{C}^{n\times n}$ be a holomorphic function defined in the open domain $\mathcal{D} \subset \mathbb{C}^{n\times n}$. For $A \in \mathcal{D}$ and $H \in \mathbb{C}^{n\times n}$

---

*2000 **Mathematics Subject Classification:** 15A57, 15A24.
Key words: Matrix power functions, higher order Fréchet derivatives.

sufficiently small we have

$$f(A+H)=\sum_{m=0}^{\infty}\frac{\mathrm{D}^m f(A)(H)}{m!},$$

where the m–th derivative $\mathrm{D}^m f(A) : (\mathbb{C}^{n\times n})^m \to \mathbb{C}^{n\times n}$ is an m–linear function. Further we shall use the following result.

Proposition 1. *The value $\mathcal{F}(H)$ of the m–linear form $\mathcal{F} : (\mathbb{C}^{n\times n})^m \to \mathbb{C}^{n\times n}$ ($m \geq 1$) may be represented as a finite sum of matrix products*

$$\sum_{k=1}^{n_m} A_{m,1,k}HA_{m,2,k}\cdots A_{m,m,k}HA_{m,m+1,k},$$

where $A_{m,j,k}$, $j=1,2,\ldots,m+1$, $k=1,2,\ldots,n_m$, are given constant matrices.

Proof. Let $X=[x_{i,j}]\in\mathbb{C}^{n\times n}$ and $E_{i,j}=e_i^{\mathrm{T}}e_j$, where e_i is the i–th column of I_n. Then, the matrix

$$YXE_{i,i_1}XE_{j_1,i_2}X\cdots XE_{j_{m-1},i_m}XE_{j_m,j} \tag{1}$$

is equal to $x_{i_1,j_1}x_{i_2,j_2}\cdots x_{i_m,j_m}E_{i,j}$. It suffices to observe that each m–linear form $\mathcal{F}(H)$ may be represented as a finite sum of matrices of the form (1). $\square$

Definition of f_p and its derivatives. When defining the matrix X^p, we impose no restrictions on $X\in\mathbb{C}^{n\times n}$ if $p\in\mathbb{N}$ and we suppose that $X\in\mathbb{C}_*^{n\times n}$ if $p\notin\mathbb{N}$.

In what follows $A\in\mathbb{C}^{n\times n}$ is a fixed matrix and $H\in\mathbb{C}^{n\times n}$ is an increment of A with 2–norm denoted as $\eta:=\|H\|_2$. It is supposed that asymptotics of the form $\mathrm{O}(\eta^m)$, $m\geq 1$, is valid for $\eta\to 0$. If $p\in\mathbb{N}$, then the matrices A and H may be arbitrary. Otherwise, both A and $A+H$ have to be from $\mathbb{C}_*^{n\times n}$. The latter is achieved if $A\in\mathbb{C}_*^{n\times n}$ and η is small enough.

The function $f_p:\mathbb{C}_*^{n\times n}\to\mathbb{C}_*^{n\times n}$ and its Fréchet derivatives $\mathrm{D}^m f_p(A) : (\mathbb{C}^{n\times n})^m \to \mathbb{C}^{n\times n}$ may be correctly defined using either the Jordan (or Schur) form of A, or contour Cauchy type integrals. Suppose, in particular, that the eigenvalues of the matrix A are pairwise distinct and let $A=USU^{-1}$, where the matrix $U\in\mathbb{C}^{n\times n}$ is invertible and $S=\mathrm{diag}(\mathrm{s}_1,\mathrm{s}_2,\ldots,\mathrm{s}_n)$, is the diagonal Jordan form of A. Then, we may set $A^p=US^pU^{-1}$, where $S^p=\mathrm{diag}\,(\mathrm{s}_1^p,\mathrm{s}_2^p,\ldots,\mathrm{s}_n^p)$ and s_k^p is the principal p–th power of the eigenvalue s_k of A.

The matrix A^p may also be defined for any $p\in\mathbb{R}$ for matrices $A\in\mathbb{C}_*^{n\times n}$ which are not diagonalizable. In this case the Jordan form of A may be used. The more general case when A is only invertible and/or the degree p is complex will be considered elsewhere.

Another approach to define the function f_p and its derivatives is by contour integrals. Denote by $C:=\{z\in\mathbb{C}:|z-a|=\rho\}\subset\mathbb{C}$ a contour encircling the spectrum of $A\in\mathbb{C}_*^{n\times n}$ and let $\mathcal{B}\subset\mathbb{C}_*^{n\times n}$ be the set of matrices X such that $\|aI_n-X\|_2<\rho$. Let $f:\mathbb{C}\to\mathbb{C}$ be a function which is holomorphic in the interior C^o of C. Then, the function f is defined on the spectrum of each $A\in\mathcal{B}$ and we may define the matrix counterpart $f:\mathcal{B}\to\mathbb{C}_*^{n\times n}$ of the scalar function f by the contour integral

$$f(A)=\frac{1}{2\pi\mathrm{i}}\oint_C f(z)R_A(z)\mathrm{d}z,\ A\in\mathcal{B}, \tag{2}$$

where $R_A(z):=(zI_n-A)^{-1}$, $z\notin\mathrm{spect}(A)$, is the resolvent for A. In our case we may take $f(z) = z^p$, which gives

$$A^p = \frac{1}{2\pi\mathrm{i}} \oint_C z^p\, R_A(z)\mathrm{d}z.$$

Using (2) we can define the Fréchet derivative of f at a given point A as follows. For a fixed $A \in \mathcal{B}$ choose $H \in \mathbb{C}^{n\times n}$ with η sufficiently small. Then, the matrix $zI_n - A - H$ is invertible for $|z - a| = \rho$ and we have

$$\begin{aligned} R_{A+H}(z) &= (zI_n - A - H)^{-1} = \big((zI_n - A)\left(I_n - (zI_n - A)^{-1}H\right)\big)^{-1} \\ &= (I_n - R_A(z)H)^{-1}\, R_A(z) = R_A(z) + R_A(z)HR_A(z) + \mathrm{O}(\eta^2). \end{aligned} \tag{3}$$

On the other hand,

$$f(A + H) = \frac{1}{2\pi\mathrm{i}} \oint_C f(z)R_{A+H}(z)\mathrm{d}z = f(A) + \mathrm{D}f(A)(H) + \mathrm{O}(\eta^2). \tag{4}$$

Substituting (3) in (4), we obtain

$$\mathrm{D}f(A)(H) = \frac{1}{2\pi\mathrm{i}} \oint_C f(z)R_A(z)HR_A(z)\mathrm{d}z. \tag{5}$$

The techniques of contour integrals may be used in order to define the m–th Fréchet derivative $\mathrm{D}^m f_p(A)$ of f_p at the point A as an m–linear form. Indeed, we have

$$R_{A+H}(z) = \sum_{k=0}^{m} (R_A(z)H)^k R_A(z) + \mathrm{O}(\eta^{m+1}),$$

which yields

$$\mathrm{D}^m f_p(A)(H) = \frac{1}{2\pi\mathrm{i}} \int_C z^p (R_A(z)H)^m R_A(z)\mathrm{d}z.$$

In particular, the second Fréchet derivative is given by

$$\mathrm{D}^2 f_p(A)(H) = \frac{1}{2\pi\mathrm{i}} \int_C z^p R_A(z)HR_A(z)HR_A(z)\mathrm{d}z.$$

We point out that the norms of the operators $\mathrm{D}^m f_p(A)$ are studied in [2].

For $A \in \mathbb{C}_*^{n\times n}$, $\eta < \eta_0$ and $\eta_0 > 0$ sufficiently small the function $H \mapsto (A + H)^p$ is holomorphic or real analytical in the real case. Hence,

$$(A + H)^p = \sum_{m=0}^{\infty} \mathcal{F}_m(p, A)(H), \tag{6}$$

where $\mathcal{F}_m(p, A)(H) = \mathrm{D}^m f_p(A)(H)/m! = \mathrm{O}(\eta^m)$ are m–linear forms in H. In particular, we have $\mathcal{F}_0(p, A)(H) = A^p$ and $\mathcal{F}_1(p, A) = \mathrm{D}f_p(A) \in \mathcal{L}(n)$ is the Fréchet derivative of f_p at the point A.

Higher order Fréchet derivatives of f_p. The first derivative $\mathcal{F}_1(p, A)$ is well studied for $p \in \mathbb{Z}$ and $1/p \in \mathbb{N}$, see [1], and for $p \in \mathbb{Q}$, see [4]. The more general case $p \in \mathbb{R}$ has been considered in [3] for some particular structures of A. In what follows we describe the expressions $\mathcal{F}_m(p, A)$ for some values of p. When $p < 0$, we suppose that $\eta < \|A^{-1}\|_2^{-1}$.

The case $p \in \mathbb{N}$. Here the expressions for $\mathcal{F}_m(p, A)$ are well known and we give them only for completeness.

Theorem 2. *We have*

$$\mathcal{F}_m(p,A)(H) = \sum_{|\Sigma_{m+1}|=p-m} A^{\sigma_1}HA^{\sigma_2}\cdots A^{\sigma_m}HA^{\sigma_{m+1}},\ m \le p,$$

and $\mathcal{F}_m(r,A) = 0$ *for* $m > p$*, where* $\Sigma_{m+1} = (\sigma_1,\sigma_2,\ldots,\sigma_{m+1}) \in \mathbb{Z}^{m+1}$*,* $0 \le \sigma_i \le p-m$ *and* $|\Sigma_{m+1}| := \sigma_1+\sigma_2+\cdots+\sigma_{m+1}$*.*

In particular,

$$\mathcal{F}_1(p,A)(H) = \sum_{|\Sigma_2|=p-1} A^{\sigma_1}HA^{\sigma_2},\ \mathcal{F}_2(p,A)(H) = \sum_{|\Sigma_3|=p-2} A^{\sigma_1}HA^{\sigma_2}HA^{\sigma_3}.$$

The case $p=-1$**.** Here the result follows immediately from the expansion

$$(A+H)^{-1} = \sum_{m=0}^{\infty}(-1)^m A^{-1}(HA^{-1})^m = \sum_{m=0}^{\infty}(-1)^m(A^{-1}H)^m A^{-1}.$$

Theorem 3. *We have the following result,*

$$\mathcal{F}_m(-1,A)(H) = (-1)^m A^{-1}(HA^{-1})^m = (-1)^m(A^{-1}H)^m A^{-1}.$$

In particular,

$$\mathcal{F}_1(-1,A)(H) = -A^{-1}HA^{-1},\ \mathcal{F}_2(-1,A)(H) = A^{-1}HA^{-1}HA^{-1}.$$

The case $p=-r$**,** $r \in \mathbb{N}$**.** Here we have

$$\begin{aligned}(A+H)^{-r} &= \left(\sum_{m=0}^{\infty}\mathcal{F}_m(-1,A)(H)\right)^r\\ &= \sum_{m=0}^{\infty}\sum_{|\Sigma_r|=m}\mathcal{F}_{\sigma_1}(-1,A)(H)\mathcal{F}_{\sigma_2}(-1,A)(H)\cdots\mathcal{F}_{\sigma_r}(-1,A)(H).\end{aligned}$$

Therefore,

$$\mathcal{F}_m(-r,A) = \sum_{|\Sigma_r|=m}\mathcal{F}_{\sigma_1}(-1,A)(H)\mathcal{F}_{\sigma_2}(-1,A)(H)\cdots\mathcal{F}_{\sigma_r}(-1,A)(H).$$

After some standard but cumbersome calculations, we obtain an explicit expression for $\mathcal{F}_m(-r,A)$ as follows.

Theorem 4. *It is fulfilled that*

$$\mathcal{F}_m(-r,A)(H) = (-1)^m \sum_{|\Sigma_{m+1}|=m+r} A^{-\sigma_1}HA^{-\sigma_2}\cdots A^{-\sigma_m}HA^{-\sigma_{m+1}}. \tag{7}$$

In particular we have

$$\begin{aligned}
\mathcal{F}_1(-2,A)(H) &= -H_{2,1}-H_{1,2},\\
\mathcal{F}_2(-2,A)(H) &= H_{1,1,2}+H_{1,2,1}+H_{2,1,1},\\
\mathcal{F}_3(-2,A)(H) &= -H_{1,1,1,2}-H_{1,1,2,1}-H_{1,2,1,1}-H_{2,1,1,1}\\
\mathcal{F}_1(-3,A)(H) &= -H_{1,3}-H_{2,2}-H_{3,1},\\
\mathcal{F}_2(-3,A)(H) &= H_{1,1,3}+H_{1,3,1}+H_{3,1,1}+H_{1,2,2}+H_{2,1,2}+H_{2,2,1},\\
\mathcal{F}_3(-3,A)(H) &= -H_{1,1,1,3}-H_{1,1,3,1}-H_{1,3,1,1}-H_{3,1,1,1}-H_{1,1,2,2}\\
&\quad -H_{1,2,1,2}-H_{1,2,2,1}-H_{2,1,1,2}-H_{2,1,2,1}-H_{2,2,1,1},
\end{aligned}$$

where

$$H_{\sigma_1,\sigma_2,\ldots,\sigma_{m+1}} := A^{-\sigma_1}HA^{-\sigma_2}\cdots A^{-\sigma_m}HA^{-\sigma_{m+1}}.$$

The case $p=1/s$, $s\in\mathbb{N}$. Here we have

$$A+H=\left(\sum_{m=0}^{\infty}\mathcal{F}_m(1/s,A)(H)\right)^s.$$

Based on this identity, we may prove the following result.

Theorem 5. *Let the linear operator $\mathcal{L}_s := \mathcal{F}_1(s,A^{1/s})$ be invertible. Then: (i) $\mathcal{F}_1(1/s,A)=\mathcal{L}_s^{-1}$; (ii) the spectrum of $\mathcal{L}_s$ consists of the numbers*

$$\sum_{k=0}^{s-1}\mathrm{s}_i^{k/s}\mathrm{s}_j^{(s-k-1)/s},\ i,j=1,2,\ldots,n,$$

where $\{\mathrm{s}_1,\mathrm{s}_2,\ldots,\mathrm{s}_n\}$ is the full spectrum of A.

The determination of $\mathcal{F}_m(1/s,A)$ for $m\geq 2$ is more subtle. We give below some particular results using the abbreviation

$$F_{m,s} := \mathcal{F}_m(1/s,A)(H).$$

Theorem 6. *The following representations are valid for $p=1/2$ and $p=1/3$:*

$$\begin{aligned}
F_{m,2} &= \mathcal{L}_2^{-1}\left(\sum_{i=1}^{m-1}F_{i,2}F_{m-i,2}\right),\ m\in\mathbb{N},\\
F_{2,3} &= -\mathcal{L}_3^{-1}\left(A^{1/3}F_{1,3}^2+F_{1,3}A^{1/3}F_{1,3}+F_{1,3}^2A^{1/3}\right)\\
&= -\mathcal{L}_3^{-1}\circ\mathcal{F}_1(3,A^{1/3})(F_{1,3}),\\
F_{3,3} &= -\mathcal{L}_3^{-1}\Big(F_{1,3}^3+A^{1/3}F_{1,3}F_{2,3}+F_{1,3}A^{1/3}F_{2,3}+F_{1,3}F_{2,3}A^{1/3}\\
&\quad +F_{2,3}F_{1,3}A^{1/3}+F_{2,3}A^{1/3}F_{1,3}+A^{1/3}F_{2,3}F_{1,3}\Big),
\end{aligned}$$

where the invertibility of the operator $\mathcal{L}_2$ and/or $\mathcal{L}_3$ is presupposed.

To illustrate the above results let us consider a simple example.

Example 7. Let $n=2$, $H=[h_{i,j}]$ and $A=\mathrm{diag}(\mathrm{s}_1,\mathrm{s}_2)$, where $\mathrm{s}_1\mathrm{s}_2\neq 0$. Then, the elements $\varphi_{i,j}$ of the matrix $F_{2,2} = \mathcal{F}_2(1/2, A)(H) \in \mathbb{C}^{2\times 2}$ are given by

$$\begin{aligned}
\varphi_{i,i} &= \frac{1}{2\sqrt{\mathrm{s}_i}}\left(\frac{h_{i,i}^2}{4\mathrm{s}_i} + \frac{h_{1,2}h_{2,1}}{(\sqrt{\mathrm{s}_1}+\sqrt{\mathrm{s}_2})^2}\right),\\
\varphi_{i,j} &= \frac{h_{i,j}}{2(\sqrt{\mathrm{s}_1}+\sqrt{\mathrm{s}_2})^2}\left(\frac{h_{1,1}}{\sqrt{\mathrm{s}_1}} + \frac{h_{2,2}}{\sqrt{\mathrm{s}_2}}\right),\ i \neq j.
\end{aligned}$$

Conclusions. In this paper we have derived explicit expressions for the m–th order Fréchet derivatives of matrix power functions $X \mapsto X^p$ for $p \in \mathbb{Z} \cup \{1/2\}$ and $m \in \mathbb{N}$ as well as for $p = 1/3$ and $m = 2, 3$. The investigation of the general cases $p \in \mathbb{Q}$ (and even for $p \in \mathbb{R}$), $m \in \mathbb{N}$, remains an open problem.

REFERENCES

[1] R. Bhatia, D. Singh, K. Sinha. Differentiation of operator functions and perturbation bounds. *Commun. Math. Phys.*, **191** (1998), 603–611, ISSN 0010–3616.

[2] R. Bhatia, J. Holbrook. Fréchet derivatives of the power function. *Indiana Univ. Math. J.* **49** (2000), 1155–1173, MR1803224.

[3] J. Boneva, M. Konstantinov, P. Petkov. Perturbation analysis for the complex matrix equation $Q \pm A^{\mathrm{H}}X^pA - X = 0$. *Surveys Math. Appl.*, **2** (2007), 29–41, ISSN 1842–6298.

[4] M. Konstantinov, J. Boneva, P. Petkov. Fréchet derivatives of rational power matrix functions. *Math. and Education in Math.*, **35** (2006), 169–174.

M. Konstantinov, J. Boneva, V. Todorov
UACEG
1046 Sofia, Bulgaria
e-mail: mmk_fte@uacg.bg
boneva_fte@uacg.bg
vttp@yahoo.com

P. Petkov
Department of Automatics
TU – Sofia
1756 Sofia, Bulgaria
e-mail: php@tu-sofia.bg

ПРОИЗВОДНИ ПО ФРЕШЕ ОТ ПО–ВИСОК РЕД НА МАТРИЧНИ СТЕПЕННИ ФУНКЦИИ

Михаил Михайлов Константинов, Юлияна Костадинова Бонева, Петко Христов Петков, Владимир Тодоров Тодоров

Изучени са производните по Фреше от по–висок ред на матричните степенни функции $X \mapsto X^p$, $p \in \mathbb{Q}$, $X \in \mathbb{C}^{n\times n}$. Получените резултати могат да се приложат при анализа на точността на апроксимацията по Тейлър на тези функции и при пертурбационния анализ на нелинейни матрични уравнения.