

PERTURBATION BOUNDS FOR THE MATRIX EQUATION
 $X^s \pm A^H X^t A = Q$

Mihail Konstantinov, Petko Petkov*, Ivan Popchev**,
Vera Angelova**

(Submitted on May 5, 2010)

Abstract

In this paper we consider the sensitivity of the nonlinear complex matrix equation $X^s \pm A^H X^t A = Q$, where the exponents s and t are real numbers. Local and nonlocal perturbation bounds for the perturbation in the solution X are obtained.

Key words: perturbation analysis, nonlinear matrix equations, perturbation bounds, Lyapunov majorants

2000 Mathematics Subject Classification: 15A24

1. Introduction and notations. We consider the nonlinear matrix equation

$$(1) \quad X^s \pm A^H X^t A = Q,$$

where the data matrices A , Q and the solution X are $n \times n$ complex matrices, Q is positive definite and s , t are real numbers. Equations of this type arise from the areas of ladder networks, dynamic programming, control theory, stochastic filtering and statistics, etc.

This research work was supported by the Bulgarian Academy of Sciences under Grant No 010093/2010.

Several authors have proposed perturbation bounds for the solution of particular classes of equation (1) with $s = 1$. The case $t = -1$ is considered in [1], $t = 1$ in [19], $t = -q$, $q \in (0, 1]$ in [2], $t = -1/2$ in [3], $t = \pm 1/p$, $p = 1, 2, \dots$ in [4], $t = -2$ in [5-7], $t = -r$, $r = 1, 2, \dots$ in [8,9]. In [20] perturbation bounds to the matrix equation $C + \sum_{i=1}^r A_i X B_i + D X^s E = 0$ with r and $s \geq 2$ – positive integers are proposed.

The sensitivity of the solution of the real version of equation (1), when s is a positive integer, Q is the identity matrix and t is negative integer is studied firstly in [10]. Then in [11] the sensitivity of equation (1) with $Q > 0$ and s, t – both nonnegative integers, is considered. Perturbation estimates of the positive definite solution to the complex matrix equation (1) with s -positive integer and t -negative integer are proposed in [12,13].

In this paper we give perturbation bounds for the solution of the complex matrix equation (1) in the most general case, where s and t are real numbers. Using the technique of Fréchet derivatives and applying the method of Lyapunov majorants and the Schauder fixed point principle, we obtain local and non-local perturbation bounds for the positive definite solution of equation (1). For the computation of the p -th root $X^{1/p}$ in this case the algorithms considered in [2,14,15] may be used.

Throughout the paper we use the notations \mathbb{N} – the set of natural numbers; \mathbb{R} – the set of real numbers; \mathbb{C} – the set of complex numbers; $\mathbb{S}^{n \times n}$ – the set of all $n \times n$ positive definite matrices; I_n – the identity $n \times n$ matrix; A^\top, \bar{A}, A^H stands for the transpose, complex conjugate and complex conjugate transpose of A ; $\text{vec}(A) = [a_1^\top, a_2^\top, \dots, a_n^\top]^\top \in \mathbb{C}^{n^2}$ – the column-wise vector representation of the matrix $A = [a_1, a_2, \dots, a_n] \in \mathbb{C}^{n \times n}$, $a_j \in \mathbb{C}^n$, where $\mathbb{C}^n = \mathbb{C}^{n \times 1}$; $\text{Mat}(\mathcal{L}) \in \mathbb{C}^{n^2 \times n^2}$ – the matrix representation of the linear matrix operator $\mathcal{L} : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$; $A \otimes B = [A(k,l)B]$ – the Kronecker product of the matrices $A = [A(k,l)]$ and B ; $\|\cdot\|$ – a vector or a matrix norm; $\|\cdot\|_2$ – the Euclidean vector or the spectral matrix norm; $\|\cdot\|_F$ – the Frobenius norm; The notation “:=” stands for “equal by definition”.

2. Existence of a solution. In this section we give a theorem on the existence of a positive definite solution of equation (1).

Theorem 1. Denote $Y := X^s$ and $\mathcal{G}(X) := \pm X^{t/s} : \mathbb{S}^{n \times n} \mapsto \mathbb{S}^{n \times n}$. If $\mathcal{G}(Y)$ is continuous and monotone, (i.e. $X \leq Y$ implies $\mathcal{G}(X) \leq \mathcal{G}(Y)$) and if $A^H \mathcal{G}(Q) A \leq Q$, then equation (1) has a positive definite solution. If $\mathcal{G}(Y)$ is continuous and anti-monotone, (i.e. $X \leq Y$ implies $\mathcal{G}(X) \geq \mathcal{G}(Y)$) and if there exists Y_0 such that $Q - A^H \mathcal{G}(Y_0) A \leq Y_0$, then equation (1) has a positive definite solution.

Proof. Using the notations of $Y := X^s$ and $G(X) := \pm X^{t/s} : \mathbb{S}^{n \times n} \mapsto \mathbb{S}^{n \times n}$,

we rewrite equation (1) in the equivalent form

$$(2) \quad Y + A^H \mathcal{G}(Y)A = 0.$$

Let $\mathcal{G}(Y) : \mathbb{S}^{n \times n} \mapsto \mathbb{S}^{n \times n}$ be continuous and monotone. Assume that $A^H \mathcal{G}(Q)A \leq Q$. Then according to Corollary 2.2 from [16] equation (2) has a solution in $[0, Q]$.

Let $\mathcal{G}(Y) : \mathbb{S}^{n \times n} \mapsto \mathbb{S}^{n \times n}$ be continuous and anti-monotone. If there exists Y_0 such that $Q - A^H \mathcal{G}(Y_0)A \leq Y_0$, then according to Corollary 2.2 from [16] equation (2) has a positive definite solution in $[Q, Y_0]$.

Thereby under the assumptions, equation (1) has a positive definite solution. \square

3. Perturbation bound. In this section we develop perturbation bounds for the solution of equation (1). We consider the perturbed matrix equation

$$(3) \quad F(X + \delta X, A + \delta A, Q + \delta Q) := (X + \delta X)^s \\ \pm (A + \delta A)^H (X + \delta X)^t (A + \delta A) - Q - \delta Q = 0,$$

where δA and δQ are perturbations in the data matrices A and Q of equation (1). We assume that the perturbations are sufficiently small so that the perturbed equation (3) has a solution $X + \delta X$ in the neighbourhood of the unperturbed solution X .

3.1. Local bounds. Rewrite the perturbed equation (3) in terms of Fréchet derivatives.

$$(4) \quad F(X + \delta X, A + \delta A, Q + \delta Q) \\ = F(X, A, Q) + F_X(X, A, Q)(\delta X) + F_A(X, A, Q)(\delta A) \\ + F_{\bar{A}}(X, A, Q)(\delta \bar{A}) + F_Q(X, A, Q)(\delta Q) \\ + G(X, A, Q)(\delta X, \delta A, \delta Q) = 0.$$

Here $F(X, A, Q) := X^s \pm A^H X^t A - Q = 0$ denotes the unperturbed terms, $G(X, A, Q)(\delta X, \delta A, \delta Q)$ contains second and higher order terms in $\delta X, \delta A, \delta Q$, $F_X(X, A, Q)(Z) := \mathcal{F}(s, X)(Z) \pm A^H \mathcal{F}(t, X)(Z)A$ and $F_Q(X, A, Q)(Z) := -Z$ are the partial Fréchet derivatives of (3) in X and Q , respectively, computed at the point (X, A, Q) . The terms $F_A(X, A, Q)(Z) := \pm A^H X^t Z$ and $F_{\bar{A}}(X, A, Q)(Z) := \pm Z^H X^t A$ are the partial Fréchet pseudo derivatives of (3) in A . The notation $\mathcal{F}(p, X) : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$ is for the Fréchet derivative of the function $X \rightarrow X^p$, $p \in \mathbb{R}$ at the point $X \in \mathbb{S}(n)$ and a given increment $E \in \mathbb{C}^{n \times n}$, with $\|E\|_2 < \lambda_{\min}(X)$

$$(X + E)^p = X^p + \mathcal{F}(p, X)(E) + O(\|E\|_2), \quad E \rightarrow 0.$$

Recall that for different value of $p \in \mathbb{R}$ the derivatives $\mathcal{F}(p, X)(\delta X)$ are given in [4].

Having in mind that $F(X, A, Q) = 0$ we obtain

$$\begin{aligned} F_X(X, A, Q)(\delta X) &= -F_A(X, A, Q)(\delta A) \\ &\quad -F_{\bar{A}}(X, A, Q)(\delta \bar{A}) - F_Q(X, A, Q)(\delta Q) \\ &\quad -G(X, A, Q)(\delta X, \delta A, \delta Q). \end{aligned}$$

Denote $L := \text{Mat}\{F_X(X, A, Q)\} = L_s \pm (A^\top \otimes A^H)L_t$, $L_Q := \text{Mat}\{F_Q(X, A, Q)\} = -I_n$, $L_A := \text{Mat}\{F_A(X, A, Q)\} = I \otimes A^H X^t$, $L_{\bar{A}} := \text{Mat}\{F_{\bar{A}}(X, A, Q)\} = ((X^t A^H)^\top \otimes I)\mathcal{P}_{n^2}$. Here $\mathcal{P}_{n^2} \in \mathbb{R}^{n^2 \times n^2}$ is the so-called vec-permutation matrix, such that $\text{vec}(Y^\top) = \mathcal{P}_{n^2} \text{vec}(Y)$ for each $Y \in \mathbb{C}^{n \times n}$, L_s and L_t are the matrices of the corresponding Fréchet derivatives $\mathcal{F}(s, X)$ and $\mathcal{F}(t, X)$ for $p = s, t \in \mathbb{R}$. Expressions of L_p for different values of $p \in \mathbb{R}$ are given in [4]. Supposing that the operator $F_X(X, A, Q)$ is invertible, i.e its matrix L is non-singular, one obtains

$$\delta X = -F_X^{-1} \circ F_Q(\delta Q) - F_X^{-1} \circ F_A(\delta A) - F_X^{-1} \circ F_{\bar{A}}(\delta \bar{A}) + O(\|\delta\|^2), \quad \delta \rightarrow 0,$$

or in a vector form after applying the vec operator

$$(5) \quad \text{vec}(\delta X) = W_Q \text{vec}(\delta Q) + (W_A + W_{\bar{A}}) \text{vec}(\delta A) + O(\|\delta X\|^2).$$

Here $W_Q := -L^{-1}L_Q$, $W_A = -L^{-1}L_A$, $W_{\bar{A}} = -L^{-1}L_{\bar{A}}$.

Denote $\delta_X := \|\delta X\|_F$, $\delta := [\delta_1 \quad \delta_2]^\top = [\|\delta A\|_F \quad \|\delta Q\|_F]^\top \in \mathbb{R}_+^2$. Relation (5) gives

$$\delta_X \leq \left\| \begin{bmatrix} W_Q^{\mathcal{R}} & M_A \end{bmatrix} \right\|_2 \|\delta\|_2 + O((\delta_X + \delta)^2), \quad \delta_x, +\delta \rightarrow 0,$$

$$\delta_X \leq \sqrt{\delta^\top R \delta} + O((\delta_X + \delta)^2), \quad \delta_x, +\delta \rightarrow 0,$$

where $W_Q^{\mathcal{R}} := \begin{bmatrix} W_{Q0} & -W_{Q1} \\ W_{Q1} & W_{Q0} \end{bmatrix} \in \mathbb{R}^{2n \times 2n}$ is the real representation of the matrix

$$W_Q = W_{Q0} + iW_{Q1} \in \mathbb{C}^{n \times n},$$

$$M_A := M(W_A, W_{\bar{A}}) = \begin{bmatrix} W_{A0} + W_{\bar{A}0} & W_{\bar{A}1} - W_{A1} \\ W_{A1} + W_{\bar{A}1} & W_{A0} - W_{\bar{A}0} \end{bmatrix} \in \mathbb{R}^{2n \times 2n}$$

and

$$W_A = W_{A0} + iW_{A1}, \quad W_{\bar{A}} = W_{\bar{A}0} + iW_{\bar{A}1} \in \mathbb{C}^{n \times n}$$

with $A_{A0}, A_{A1}, A_{\bar{A}0}, A_{\bar{A}1} \in \mathbb{R}^{n \times n}$. The matrix R is given by

$$R = \begin{bmatrix} \|(W_Q^{\mathcal{R}})^{\top} W_Q^{\mathcal{R}}\|_2 & \|(W_Q^{\mathcal{R}})^{\top} M_A\|_2 \\ \|M_A^{\top} W_Q^{\mathcal{R}}\|_2 & \|M_A^{\top} M_A\|_2 \end{bmatrix} \in \mathcal{R}_+^2.$$

Hence, we formulate the following local perturbation bound for the perturbation δX in the solution X of equation (1)

$$(6) \quad \delta X \leq \text{est}(\delta)$$

$$:= \min\{\| [W_Q^{\mathcal{R}} \quad M_A] \|_2 \|\delta\|_2, \sqrt{\delta^{\top} R \delta}\} + O((\delta_X + \delta)^2), \delta_x, +\delta \rightarrow 0,$$

3.2. Non-local bound. In this section we present a nonlocal perturbation bound within second order terms in δX for the solution of equation (1). We rewrite the perturbed equation (3) in the form

$$F_X(\delta X) = -\Phi_0(\delta A, \delta Q) - \Phi_1(\delta X, \delta A, \delta Q) - \Phi_2(\delta X, \delta A, \delta Q) + O(\|\delta X\|^3),$$

with

$$\begin{aligned} \Phi_0(\delta A, \delta Q) &:= F_Q(X, A, Q)(\delta Q) + F_A(X, A, Q)(\delta A) + F_{\bar{A}}(X, A, Q)(\delta \bar{A}) \\ &\quad + \delta A^{\text{H}} X^t \delta A, \end{aligned}$$

$$\begin{aligned} \Phi_1(\delta X, \delta A, \delta Q) &:= \pm(A^{\text{H}} \mathcal{F}(t, X)(\delta X) \delta A \\ &\quad \pm \delta A^{\text{H}} \mathcal{F}(t, X)(\delta X) A \pm \delta A^{\text{H}} \mathcal{F}(t, X)(\delta X) \delta A), \end{aligned}$$

$$\Phi_2(\delta X, \delta A, \delta Q) := \pm(A + \delta A)^{\text{H}} \mathcal{F}_2(t, X)(\delta X)(A + \delta A),$$

where the expression $\mathcal{F}_2(t, X)(\delta X) := (X + \delta X)^t - X^t - \mathcal{F}(t, X)(\delta X) = O(\|\delta X\|^2)$ contains the terms of second and higher order in δX .

Assuming the invertibility of the operator $F_X(X, A, Q)$ we obtain the following vector operator equation within the terms of second order

$$\text{vec}(\delta X) = \Theta(\delta X, \delta A, \delta Q),$$

$$\begin{aligned} \Theta(\delta X, \delta A, \delta Q) &:= -L^{-1} \text{vec}(\Phi_0(\delta A, \delta Q)) - L^{-1} \text{vec}(\Phi_1(\delta X, \delta A, \delta Q)) \\ &\quad - L^{-1} \text{vec}(\Phi_2(\delta X, \delta A, \delta Q)). \end{aligned}$$

Let $\delta_X \leq \rho$, where ρ is some positive quantity. The Frobenius norm of $\Theta(\delta X, \delta A, \delta Q)$ may be estimated as

$$\|\Theta(\delta X, \delta A, \delta Q)\|_F \leq a_2(\delta)\rho^2 + a_1(\delta)\rho + a_0(\delta),$$

where

$$(7) \quad \begin{aligned} a_0(\delta) &:= \text{est}(\delta) + \|L^{-1}\|_2 \|X^t\|_2 \delta_A^2, \\ a_1(\delta) &:= \|L^{-1}\|_2 \|L_t\|_2 (2\|A\|_2 \delta_A + \delta_A^2), \\ a_2(\delta) &:= \|L^{-1}\|_2 (\|A\|_2 + \delta_A)^2 \varphi(t, X). \end{aligned}$$

Here $\text{est}(\delta)$ is the local bound (6), L is the matrix representation of the operator $F_X(X, A, Q)$ and $\varphi(t, X)$ denotes $\|\text{Mat}(\mathcal{F}_2(t, X))\|_2 \leq \varphi(t, X)\delta_X^2 + O(\|\delta X\|^3)$.

The Lyapunov majorant [17,18] for the vector operator equation $\text{vec}(\delta X) = \Theta(\delta X, \delta A, \delta Q)$ such that $\|\Theta(\delta X, \delta A, \delta Q)\|_F \leq h(\rho, \delta)$ is a quadratic function $h(\rho, \delta) = a_2(\delta)\rho^2 + a_1(\delta)\rho + a_0(\delta)$.

Suppose that $\delta \in \Omega$, where

$$(8) \quad \Omega := \left\{ \delta \geq 0 : a_1(\delta) + 2\sqrt{a_0(\delta)a_2(\delta)} \leq 1 \right\} \subset \mathbb{R}_+^{k+1}.$$

The inclusion $\delta \in \Omega$ guarantees that the equation $h(\rho, \delta) = \rho$ has a root

$$(9) \quad \rho = f(\delta) := \frac{2a_0(\delta)}{1 - a_1(\delta) + \sqrt{(1 - a_1(\delta))^2 - 4a_0(\delta)a_2(\delta)}}.$$

Hence, the operator $\Theta(\cdot, \delta A, \delta Q)$ maps the closed convex ball $\mathcal{B}(\delta) := \{H \in \mathbb{C}^{n \times n} : \|H\|_F \leq f(\delta), \delta \in \Omega\} \subset \mathbb{C}^{n \times n}$ into itself. According to the Schauder fixed point principle there exists a solution $\delta X \in \mathcal{B}(\delta)$ of equation $h(\rho, \delta) = \rho$, or, equivalently, $a_2(\delta)\rho^2 - (1 - a_1(\delta))\rho + a_0(\delta) = 0$, for which $\delta_X = \|\delta X\|_F \leq f(\delta), \delta \in \Omega$. Hence, we formulate a non-local perturbation bound for the solution of equation (1).

Theorem 2. *Let $\delta \in \Omega$, where Ω is given in (8). Then within second order terms in δX the non-local perturbation bound $\delta_X \leq f(\delta)$ is valid for equation (1), where $f(\delta)$ is determined by relations (9) and (7).*

The expressions of L_t and bounds for $\varphi(t, X)$ for different values of $t \in \mathbb{R}$: $t = \pm r$, $t = \pm 1/2$, $t = 1/3$, $t = \pm 1/p$, $t = \pm r/p$, $r, p = 1, 2, \dots$, are listed in Table 1.

T a b l e 1

t	L_t	$\varphi(t, X)$
r	$\sum_{k=0}^{r-1} (X^k)^\top \otimes X^{r-1-k}$	$\sum_{j=2}^r \left\ \sum_{i=j}^r (X^{r-i})^\top \otimes X^{i-j} \right\ \left\ X^{j-2} \right\ _2$ $\leq (r-1 + \binom{r}{2}) \ X\ _2^{\binom{r-2}{2}}$
$-r$	$-\sum_{k=0}^{r-1} (X^{k-r})^\top \otimes X^{-(1+k)}$	$\sum_{j=1}^r \left\ \sum_{i=1}^{r-j+1} (X^{-r+i+j-2})^\top \otimes X^{-i} \right\ \left\ X^{-j} \right\ _2$ $\leq (2r-1 + \binom{r}{2}) \ X^{-1}\ _2^{\binom{r+2}{2}}$
$\frac{1}{2}$	$I \otimes X + X^\top \otimes I$	$2\ L_{\frac{1}{2}}\ _2^3 / (1 - 2\ L_{\frac{1}{2}}\ _2^2 \delta_X + \sqrt{1 - 4\ L_{\frac{1}{2}}\ _2^2 \delta_X})$
$-\frac{1}{2}$	$\left(-(X^{-1})^\top \otimes X^{-1/2} - (X^{-1/2})^\top \otimes X^{-1} \right)^{-1}$	$\frac{3}{8} \ (X^{-1})^\top \otimes X^{-1/2}\ _2 \ X^{-1}\ _2,$
$\frac{1}{3}$	$I \otimes X^2 + X^\top \otimes X + (X^2)^\top \otimes I$	$3\ X\ _2^{1/3} \ L_{1/3}\ _2^3$
$\frac{1}{p}$	$\left(\sum_{k=0}^{p-1} (X^{k/p})^\top \otimes X^{(p-1-k)/p} \right)^{-1}$	$(\ X\ _2^{(p-2)/p} \ L_{1/p}\ _2^3 p(p-1)) / 2$
$-\frac{1}{p}$	$\left(-\sum_{k=0}^{p-1} (X^{(k-p)/p})^\top \otimes X^{-(1+k)/p} \right)^{-1}$	--
$\frac{r}{p}$	$\left(\sum_{k=0}^{r-1} (X^{k/p})^\top \otimes X^{(r-1-k)/p} \right) L_{1/p}$	--
$-\frac{r}{p}$	$-\left((X^{-r/p})^\top \otimes X^{-r/p} \right) L_{r/p}$	--

REFERENCES

- [1] HASANOV V. I., I. G. IVANOV. Linear Algebra Appl., **413**, 2006, No 1, 81–92.
- [2] LI JING, YUHAI ZHANG. Linear Algebra Appl., **431**, 2009, No 9, 1489–1501.
- [3] ANGELOVA V. A., M. M. KONSTANTINOV, P. HR. PETKOV, I. P. POPCHEV. In: Proc. 16th IFAC World Congress, Praha, **16**, 2005, part 1.

- [4] BONEVA J. K., M. M. KONSTANTINOV, P. H. PETKOV. *Surveys Math. Appl.*, **2**, 2007, 29–41.
- [5] IVANOV I. G., V. HASANOV, B. MINCHEV. *Linear Algebra Appl.*, **326**, 2001, No 1–3, 27–44.
- [6] ANGELOVA V. *Ann. Inst. Arch. Genie Civil Geod.*, fasc II Math., **41**, 2003, 33–41.
- [7] CHEN M., S. XU. *Linear Algebra Appl.*, **394**, 2005, 39–51.
- [8] HASANOV V. I., I. G. IVANOV. *Appl. Math. Comput.*, **156**, 2004, No 2, 513–525.
- [9] IVANOV I. G. *Linear Algebra Appl.*, **395**, 2005, 313–331.
- [10] LIU X.-G., H. GAO. *Linear Algebra Appl.*, **368**, 2003, 83–97.
- [11] JIA Z., M. WEI. *Appl. Math. Comput.*, **209**, 2009, No 2, 230–237.
- [12] YIN X., S. LIU, L. FANG. *Linear Algebra Appl.*, **431**, 2009, No 9, 1409–1421.
- [13] YIN X., S. LIU. *Appl. Math. Comput.*, **216**, 2010, No 1, 27–34.
- [14] BINI D., N. HIGHAM, B. MEINI. *Numer. Alg.*, **39**, 2005, 349–378.
- [15] GUO CHUN-HUA. *Linear Algebra Appl.*, **432**, 2010, No 8, 905–1922.
- [16] RAN A., M. REURINGS. *Linear Algebra Appl.*, **346**, 2002, No 1–3, 15–26.
- [17] KONSTANTINOV M. M., D. W. GU, V. MEHRMANN, P. H. PETKOV. *Perturbation Theory for Matrix Equations*, Amsterdam, North-Holland, 2003, [ISBN 0-444-51315-9].
- [18] KONSTANTINOV M. M., P. HR. PETKOV. *Appl. Comput. Math.*, **7**, 2008, No 2, 141–161.
- [19] ANGELOVA V. A. *Compt. rend. Acad. bulg. Sci.*, **56**, 2003, No 12, 47–52.
- [20] KONSTANTINOV M. M., P. HR. PETKOV, I. P. POPCHEV, V. A. ANGELOVA. *Compt. rend. Acad. bulg. Sci.*, **61**, 2008, No 9, 1111–1120.

*University of Architecture,
Civil Engineering and Geodesy
1, Hr. Smirnenski Blvd
1046 Sofia, Bulgaria*

**Department of Automatics
Technical University of Sofia
8, St. Kl. Ohridski Blvd
1756 Sofia, Bulgaria*

***Institute of Information Technologies
Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 2
1113 Sofia, Bulgaria*