



ITERATION WITH STEPSIZE PARAMETER AND CONDITION NUMBERS FOR A NONLINEAR MATRIX EQUATION*

SYED M. RAZA SHAH NAQVI[†], JIE MENG[†], AND HYUN-MIN KIM[†]

Abstract. In this paper, the nonlinear matrix equation $X^p + A^T X A = Q$, where p is a positive integer, A is an arbitrary $n \times n$ matrix, and Q is a symmetric positive definite matrix, is considered. A fixed-point iteration with stepsize parameter for obtaining the symmetric positive definite solution of the matrix equation is proposed. The explicit expressions of the normwise, mixed and componentwise condition numbers are derived. Several numerical examples are presented to show the efficiency of the proposed iterative method with proper stepsize parameter and the sharpness of the three kinds of condition numbers.

Key words. Matrix equation, Symmetric positive definite, Fixed-point iteration, Condition number, Mixed and componentwise.

AMS subject classifications. 15A24, 65F10, 65H10.

1. Introduction. We consider the nonlinear matrix equations

$$(1.1) \quad X^p + A^T X A = Q,$$

where p is a positive integer, $A \in \mathbb{R}^{n \times n}$ and Q is an $n \times n$ symmetric positive definite matrix.

This type of nonlinear matrix equation has been studied recently by several authors, see [6, 9–15]. For the case $p = 1$, equation (1.1) reduces to a symmetric linear matrix equation [16], which appears in each step of Newton's method for solving nonlinear matrix equations, such as the one appears in Chapter 7 of [19]. For the case $p > 1$, equation (1.1) is equivalent to $Y + A^T Y^{\frac{1}{p}} A = Q$ with $Y = X^p$. This is an example of the equation studied by El-Sayed and Ran [5]; see also [17].

Jia and Wei [10] studied the matrix equation $X^s + A^T X^t A = Q$, where s and t are both nonnegative integers, and they proved that a symmetric positive definite solution exists if $\lambda_{\max}(A^T A) \leq \lambda_{\min}(Q)(\lambda_{\max}(Q))^{-\frac{t}{s}}$. They have also showed that the positive definite solution could be unique under some certain condition. Meng and Kim [13] studied equation (1.1) and proposed two elegant estimates of the positive definite solution and three basic fixed-point iterations for obtaining the solution.

For the perturbation analysis of equation (1.1), Jia and Wei [10] investigated the algebraic perturbation analysis of the unique symmetric positive solution and they defined one normwise condition number. In [22], Wang, Yang and Li derived the explicit expressions of the normwise, mixed and componentwise condition numbers for the nonlinear matrix equation $X + A^* F(X) A = Q$. Motivated by this, we investigate the mixed and componentwise condition numbers of equation (1.1).

As a continuation of the previous work, we first propose a fixed-point iteration with stepsize parameter

*Received by the editors on November 30, 2016. Accepted for publication on October 16, 2017. Handling Editor: James G. Nagy. Corresponding Author: Hyun-Min Kim. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A3B04033516) and the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) (NRF-2017R1A5A1015722).

[†]Department of Mathematics, Pusan National University, Busan, 46241, Korea (raza@pusan.ac.kr, mengjiewh@163.com, hyunmin@pusan.ac.kr).

for obtaining the symmetric positive definite solution of equation (1.1). Based on the coupled fixed-point theory, we prove that the matrix sequences generated by the proposed iteration with stepsize parameter are convergent. We then investigate two kinds of normwise condition numbers, and derive the explicit expressions of mixed and componentwise condition numbers for equation (1.1).

This paper is organized as follows. In Section 2, we propose a fixed-point iteration with stepsize parameter for obtaining the symmetric positive definite solution of equation (1.1), and we prove the convergence of the matrix sequence. In Section 3, we derive the explicit expressions of the normwise, mixed and componentwise condition numbers. In Section 4, we give some numerical examples to show the efficiency of the proposed iterative method and the sharpness of the three kinds of condition numbers.

We begin with some notations used throughout this paper. $\mathbb{R}^{n \times n}$ stands for the set of $n \times n$ matrices with elements on field \mathbb{R} . $\|\cdot\|_2$, $\|\cdot\|_F$ and $\|\cdot\|_\infty$ are the spectral norm, Frobenius norm and matrix row norm, respectively. For a matrix $B = (b_{ij}) \in \mathbb{R}^{n \times n}$, $\|B\|_{\max}$ is the matrix norm given by $\|B\|_{\max} = \max_{i,j} |b_{ij}|$ and $|B|$ is the matrix whose elements are $|b_{ij}|$. The set of all $n \times n$ positive definite matrices is presented by $P(n)$. For a Hermitian matrix H , $\lambda_{\min}(H)$ and $\lambda_{\max}(H)$ denote the minimal eigenvalue and the maximal eigenvalue, respectively. Similarly, $\sigma_{\min}(H)$ and $\sigma_{\max}(H)$ denote the minimal and the maximal singular value, respectively. For Hermitian matrices X and Y , $X \geq Y$ ($X > Y$) means that $X - Y$ is positive semidefinite (definite), and $[\alpha I, \beta I]$ denotes the matrices set $\{X : X - \alpha I \geq 0 \text{ and } \beta I - X \geq 0\}$.

2. Iteration with stepsize parameter. In this section, we propose a fixed-point iteration with stepsize parameter for obtaining the symmetric positive definite solution of equation (1.1).

In [13], there was a basic fixed-point iteration as follows:

$$(2.2) \quad \begin{cases} X_0 = \tilde{\alpha} M_1, \\ X_{k+1} = (Q - A^T X_k A)^{\frac{1}{p}}, \end{cases}$$

where $\tilde{\alpha}$ is a positive solution of the following equations

$$\begin{cases} \tilde{\alpha} = \frac{\lambda_{\min}^{\frac{1}{p}}(Q - \tilde{\beta} A^T N_1 A)}{\lambda_{\min}^{\frac{1}{p}}(Q - A^T N_1 A)}, \\ \tilde{\beta} = \frac{\lambda_{\max}^{\frac{1}{p}}(Q - \tilde{\alpha} A^T M_1 A)}{\lambda_{\max}^{\frac{1}{p}}(Q)}, \end{cases}$$

with $M_1 = \lambda_{\min}^{\frac{1}{p}}(Q - \lambda_{\max}^{\frac{1}{p}}(Q) A^* A) I$ and $N_1 = \lambda_{\max}^{\frac{1}{p}}(Q) I$.

Based on the above iteration (2.2), we propose a fixed-point iteration with stepsize parameter as follows:

$$(2.3) \quad \begin{cases} X_0 = \sigma I, \sigma \in [a, b], \\ X_{k+1} = (1 - \alpha) X_k + \alpha (Q - A^T X_k A)^{\frac{1}{p}}, \end{cases}$$

where

$$\alpha \in (0, 1),$$

$$a = \left(\lambda_{\min}(Q) - \lambda_{\max}(A^T A) \lambda_{\max}^{\frac{1}{p}}(Q) \right)^{\frac{1}{p}},$$

$$b = \left(\lambda_{\max}(Q) \right)^{\frac{1}{p}}.$$

To prove the convergence of iteration (2.3), we introduce the following well-known results:

LEMMA 2.1. (Löwner-Heinz inequality, [23, Theorem 1.1]) *If $A \geq B \geq 0$ and $0 \leq r \leq 1$, then $A^r \geq B^r$.*

LEMMA 2.2. ([3]) *If $0 < \theta \leq 1$, and P and Q are positive definite matrices of the same order with $P, Q \geq bI > 0$, then $\|P^\theta - Q^\theta\| \leq \theta b^{\theta-1} \|P - Q\|$ and $\|P^{-\theta} - Q^{-\theta}\| \leq \theta b^{-(\theta+1)} \|P - Q\|$. Here, $\|\cdot\|$ stands for one kind of matrix norm.*

$F : X \times X \rightarrow X$ is a mixed monotone mapping if for any $x, y \in X$,

$$x_1, x_2 \in X, x_1 \leq x_2 \Rightarrow F(x_1, y) \leq F(x_2, y),$$

$$y_1, y_2 \in X, y_1 \leq y_2 \Rightarrow F(x, y_1) \geq F(x, y_2).$$

THEOREM 2.3. ([1, Theorem 3]) *Let (X, \leq) be a partially ordered set and suppose there is a metric d on X such that (X, d) is a complete metric space. Let $F : X \times X \rightarrow X$ be a mixed monotone mapping for which there exists a constant $k \in [0, 1)$ such that for each $x \geq u, y \leq v$,*

$$d(F(x, y), F(u, v)) + d(F(y, x), F(v, u)) \leq k[d(x, u) + d(y, v)].$$

If there exist $x_0, y_0 \in X$ such that

$$x_0 \leq F(x_0, y_0) \quad \text{and} \quad y_0 \geq F(y_0, x_0),$$

then there exist $\bar{x}, \bar{y} \in X$ such that

$$\bar{x} = F(\bar{x}, \bar{y}) \quad \text{and} \quad \bar{y} = F(\bar{y}, \bar{x}).$$

Moreover, the matrix sequences $\{x_k\}$ and $\{y_k\}$ generated by

$$\begin{cases} x_{k+1} = F(x_k, y_k), \\ y_{k+1} = F(y_k, x_k) \end{cases}$$

converge respectively to \bar{x} and \bar{y} . And if every pair $(x, y) \in X \times X$ has a lower bound or an upper bound, then

$$\bar{x} = \bar{y}.$$

Theorem 2.3 is a generalization of the coupled fixed-point theorem obtained by Bhaskar and Lakshminantham [2] and it will be used in our proof of the convergence of iteration (2.3).

THEOREM 2.4. *If $\lambda_{\min}(Q) > \lambda_{\max}(A^T A) \lambda_{\max}^{\frac{1}{p}}(Q)$, $\frac{a^{1-p} \|A\|_2^2}{p} < 1$, then the matrix sequence $\{X_k\}$ generated by iteration (2.3) is convergent to a symmetric positive definite solution of equation (1.1).*

Proof. Let $\Omega = \{X : X = X^T \text{ and } aI \leq X \leq bI\}$, under Frobenius norm, $(\Omega, \|\cdot\|_F)$ is a Banach space. Define F on $\Omega \times \Omega$ by

$$F(X, Y) = (1 - \alpha)X + \alpha(Q - A^T Y A)^{\frac{1}{p}}.$$

Clearly, F is a mixed monotone mapping. And $F : \Omega \times \Omega \rightarrow \Omega$. Indeed, for any $X, Y \in \Omega$, applying Lemma 2.1 yields

$$\begin{aligned} F(X, Y) &= (1 - \alpha)X + \alpha(Q - A^T Y A)^{\frac{1}{p}} \\ &\geq (1 - \alpha)aI + \alpha(Q - bA^T A)^{\frac{1}{p}} \\ &\geq (1 - \alpha)aI + \alpha aI \\ &= aI, \end{aligned}$$

and

$$\begin{aligned} F(X, Y) &\leq (1 - \alpha)bI + \alpha(Q - aA^T A)^{\frac{1}{p}} \\ &\leq (1 - \alpha)bI + \alpha Q^{\frac{1}{p}} \\ &\leq bI. \end{aligned}$$

Let $k = (1 - \alpha(1 - \frac{\|A\|_2^2}{pa^{p-1}}))$ and clearly $k \in (0, 1)$. For any $X, Y, U, V \in \Omega$ such that $X \geq U, Y \leq V$, note that

$$Q - A^*XA \geq a^p I \quad \text{and} \quad Q - A^*YA \geq a^p I,$$

by Lemma 2.2, we have

$$\begin{aligned} \|F(X, Y) - F(U, V)\|_F &\leq (1 - \alpha)\|X - U\|_F + \alpha\|(Q - A^T Y A)^{\frac{1}{p}} - (Q - A^T V A)^{\frac{1}{p}}\|_F \\ &\leq (1 - \alpha)\|X - U\|_F + \frac{\alpha a^{1-p}\|A\|_2^2}{p}\|V - Y\|_F, \end{aligned}$$

and similarly,

$$\|F(Y, X) - F(V, U)\|_F \leq (1 - \alpha)\|Y - V\|_F + \frac{\alpha a^{1-p}\|A\|_2^2}{p}\|X - U\|_F.$$

Then

$$\|F(X, Y) - F(U, V)\|_F + \|F(Y, X) - F(V, U)\|_F \leq k(\|X - U\|_F + \|Y - V\|_F).$$

Let $X_0 = aI, Y_0 = bI$, since $F(\Omega \times \Omega) \subseteq \Omega$, it is trivial that

$$X_0 \leq F(X_0, Y_0) \quad \text{and} \quad Y_0 \geq F(Y_0, X_0).$$

By Theorem 2.3, there are $\bar{X}, \bar{Y} \in \Omega$ such that

$$\bar{X} = F(\bar{X}, \bar{Y}) \quad \text{and} \quad \bar{Y} = F(\bar{Y}, \bar{X}).$$

Since every $(X, Y) \in \Omega \times \Omega$ has a lower bound or an upper bound, we can get

$$\bar{X} = \bar{Y}.$$

Define the matrix sequences $\{X'_k\}$ and $\{Y'_k\}$ by

$$\begin{cases} X'_0 = aI, \\ Y'_0 = bI, \\ X'_{k+1} = F(X'_k, Y'_k) = (1 - \alpha)X'_k + \alpha(Q - A^T Y'_k A)^{\frac{1}{p}}, \\ Y'_{k+1} = F(Y'_k, X'_k) = (1 - \alpha)Y'_k + \alpha(Q - A^T X'_k A)^{\frac{1}{p}}. \end{cases}$$

According to Theorem 2.3, we have

$$\lim_{k \rightarrow \infty} X'_k = \bar{X} \quad \text{and} \quad \lim_{k \rightarrow \infty} Y'_k = \bar{Y}.$$

To show the matrix sequence $\{X_k\}$ generated by (2.3) is convergent, we first show that $X'_k \leq X_k \leq Y'_k$ is true for $k = 0, 1, \dots$. Note that $X_{k+1} = F(X_k, X_k)$ and $X'_0 \leq X_0 \leq Y'_0$, by the mixed monotonicity of F , we have

$$F(X'_0, Y'_0) \leq F(X_0, X_0) \leq F(Y'_0, X'_0),$$

that is,

$$X'_1 \leq X_1 \leq Y'_1.$$

By mathematical induction, it is easy to get $X'_k \leq X_k \leq Y'_k$ for all $k = 0, 1, \dots$. Since

$$\lim_{k \rightarrow \infty} X'_k = \bar{X} = \bar{Y} = \lim_{k \rightarrow \infty} Y'_k,$$

it follows that $\{X_k\}$ converges to \bar{X} , we can see that \bar{X} satisfies equation (1.1). \square

THEOREM 2.5. *Let X_+ be a positive definite solution of equation (1.1). Consider the sequence $\{X_k\}$ generated by iteration (2.3). Assume that $X_k \leq X_+$ for some k , then $X_{k+1} \geq X_k$.*

Proof. Assume that $X_k \leq X_+$ holds for one integer k , then from Lemma 2.1 it follows that

$$\begin{aligned} X_{k+1} &= (1 - \alpha)X_k + \alpha(Q - A^T X_k A)^{\frac{1}{p}} \\ &\geq (1 - \alpha)X_k + \alpha(Q - A^T X_+ A)^{\frac{1}{p}} \\ &= (1 - \alpha)X_k + \alpha X_+ \\ &\geq X_k. \end{aligned} \quad \square$$

COROLLARY 2.6. *Let X_+ be a positive definite solution of equation (1.1). Consider the sequence $\{X_k\}$ generated by (2.3). Assume that $X_k \geq X_+$ for some k , then $X_{k+1} \leq X_k$.*

3. Normwise, mixed and componentwise condition numbers. In this section, we investigate the normwise, mixed and componentwise condition numbers of equation (1.1).

Consider the perturbed equation

$$(3.4) \quad (X + \Delta X)^p + (A + \Delta A)^T (X + \Delta X) (A + \Delta A) = Q + \Delta Q,$$

where $\Delta A, \Delta Q \in \mathbb{R}^{n \times n}$.

Subtracting (1.1) from (3.4) yields

$$(3.5) \quad (X + \Delta X)^p - X^p = \Delta Q - (A^T X \Delta A + \Delta A^T X A + A^T \Delta X A) - \Delta E,$$

where $\Delta E = \Delta A^T X \Delta A + A^T \Delta X \Delta A + \Delta A^T X \Delta A + \Delta A^T \Delta X \Delta A$.

For notational simplicity, we introduce a function $\phi : \bar{\mathbb{N}} \times \bar{\mathbb{N}} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ defined in [20] as

$$\begin{cases} \phi(i, 0)(X, Y) = X^i, & \phi(0, j)(X, Y) = Y^j, & i \in \bar{\mathbb{N}}, \\ \phi(i, j)(X, Y) = (X\phi(i-1, j) + Y\phi(i, j-1))(X, Y), & i, j \in \mathbb{N}^+, \end{cases}$$

where $\bar{\mathbb{N}}$ is the set of natural numbers and $\mathbb{N}^+ = \bar{\mathbb{N}} - \{0\}$. We can easily get

$$\phi(0, 0)(X, Y) = I_n,$$

and

$$\phi(n, 1)(X, Y) = \sum_{i=0}^n X^{n-i} Y X^i.$$

Using the notations given above, we can also have

$$(X + Y)^n = \sum_{i=0}^n \phi(n - i, i)(X, Y), \quad n \in \bar{\mathbb{N}}.$$

Then

$$\begin{aligned} (X + \Delta X)^p - X^p &= \sum_{i=0}^p \phi(p - i, i)(X, \Delta X) - X^p \\ &= \phi(p, 0)(X, \Delta X) + \phi(p - 1, 1)(X, \Delta X) + G(\Delta X) - X^p \\ (3.6) \quad &= \sum_{j=0}^{p-1} X^{p-1-j} \Delta X X^j + G(\Delta X), \end{aligned}$$

where $G(\Delta X) = \sum_{i=2}^p \phi(p - i, i)(X, \Delta X)$.

Combining (3.5) and (3.6) and applying the vec operator yields

$$\begin{aligned} &\left(\sum_{j=0}^{p-1} X^j \otimes X^{p-1-j} + A^T \otimes A^T \right) \text{vec}(\Delta X) \\ (3.7) \quad &= \text{vec}(\Delta Q) - \left((I \otimes A^T X) + ((A^T X) \otimes I) \Pi \right) \text{vec}(\Delta A) + O\left(\|\Delta Q, \Delta A\|_F^2\right), \end{aligned}$$

where $\Pi \in \mathbb{R}^{n^2 \times n^2}$ is the vec permutation satisfying $\Pi \text{vec}(A) = \text{vec}(A^T)$ and here it is defined by

$$\Pi = \sum_{i=1}^n \sum_{j=1}^n E_{ij}(n \times n) \otimes E_{ji}(n \times n),$$

with $E_{ij} = e_i^{(n)}(e_j^{(n)})^T \in \mathbb{R}^{n \times n}$ and $e_i^{(n)}$ is the i -th column of the identity matrix I . $O\left(\|\Delta Q, \Delta A\|_F^2\right)$ is the first order approximation of ΔX with respect to $(\Delta Q, \Delta A)$. According to the implicit function theorem, we can get $\Delta X \rightarrow 0$ as $(\Delta Q, \Delta A) \rightarrow 0$.

From Theorem 16.3.2 in [8], we can get $((A^T X) \otimes I) \Pi = \Pi(I \otimes (A^T X))$. Then omitting the higher order terms in (3.7) yields

$$\begin{aligned} &\left(\sum_{j=0}^{p-1} X^j \otimes X^{p-1-j} + A^T \otimes A^T \right) \text{vec}(\Delta X) \\ (3.8) \quad &= \text{vec}(\Delta Q) - \left((I_{n^2} + \Pi)(I \otimes A^T X) \right) \text{vec}(\Delta A). \end{aligned}$$

Define a map $\Phi : \mathbb{R}^{2n^2} \rightarrow \mathbb{R}^{n^2}$ by

$$(3.9) \quad \Phi : \omega = (\text{vec}(Q)^T, \text{vec}(A)^T)^T \rightarrow \text{vec}(X),$$

where X is the unique symmetric positive definite solution of equation (1.1).

3.1. Normwise condition number. According to Rice [18] and [21], we investigate two kinds of normwise condition numbers of map Φ , which are defined by

$$(3.10) \quad K_{rel}^i = \limsup_{\delta \rightarrow 0} \sup_{\Delta_i \leq \delta} \frac{\|\Delta X\|_F}{\delta \|X\|_F}, \quad i = 1, 2,$$

where

$$\Delta_1 = \frac{\|(\Delta Q, \Delta A)\|_F}{\|(Q, A)\|_F} \quad \text{and} \quad \Delta_2 = \left\| \left(\frac{\Delta Q}{\|Q\|_F}, \frac{\Delta A}{\|A\|_F} \right) \right\|_F.$$

If Φ is Fréchet differentiable at $\omega = (\text{vec}(Q), \text{vec}(A)^T)^T$, from Theorem 4 in [18], we have

$$(3.11) \quad K_{rel}^1(X) = \frac{\|\Phi'(\omega)\|_2 \|(Q, A)\|_F}{\|X\|_F},$$

where $\Phi'(\omega)$ is the Fréchet derivative of Φ at ω .

Let

$$\begin{aligned} S &= \sum_{j=0}^{p-1} X^j \otimes X^{p-1-j} + A^T \otimes A^T, \\ T &= \begin{pmatrix} I_{n^2} & - (I_{n^2} + \Pi)(I \otimes (A^T X)) \end{pmatrix}, \\ v &= (\text{vec}(\Delta Q)^T, \text{vec}(\Delta A)^T)^T. \end{aligned}$$

Then (3.8) can be written as

$$(3.12) \quad \text{Svec}(\Delta X) = Tv.$$

Note that if $\frac{a^{1-p}\|A\|_2^2}{p} < 1$, then

$$\sigma_{\min}(S) \geq \lambda_{\min} \left(\sum_{j=0}^{p-1} X^j \otimes X^{p-1-j} \right) - \|A\|_2^2 \geq pa^{p-1} - \|A\|_2^2 > 0,$$

which shows that S is invertible. Then, it follows from (3.12) that Φ is Fréchet differentiable and

$$(3.13) \quad \Phi'(\omega) = S^{-1}T.$$

Therefore, according to (3.11), the following theorem can be obtained.

THEOREM 3.1. *Let X be the positive definite solution of matrix equation (1.1). If $\frac{a^{1-p}\|A\|_2^2}{p} < 1$, then the relative condition number k_{rel}^1 of equation (1.1) defined in (3.10) is given by*

$$k_{rel}^1 = \frac{\|S^{-1}T\|_2 \|(Q, A)\|_F}{\|X\|_F}.$$

THEOREM 3.2. *Let X be the positive definite solution of matrix equation (1.1). If $\frac{a^{1-p}\|A\|_2^2}{p} < 1$, then the relative condition number k_{rel}^2 of equation (1.1) defined in (3.10) is given by*

$$(3.14) \quad k_{rel}^2 = \frac{\|[\|Q\|_F \Sigma_r - \|A\|_F \Omega_r]\|_2}{\|X\|_F},$$

where

$$\Sigma_r = S^{-1} \quad \text{and} \quad \Omega_r = S^{-1}(I \otimes (A^T X) + ((XA)^T \otimes I)\Pi).$$

Proof. Note that (3.12) can be rewritten as

$$(3.15) \quad \text{vec}(\Delta X) = S^{-1}T_1r_1,$$

where

$$T_1 = T \text{diag}(\|Q\|_F, \|A\|_F) \quad \text{and} \quad r_1 = \left(\frac{\text{vec}(\Delta Q)}{\|Q\|_F}, \frac{\text{vec}(A)}{\|A\|_F} \right)^T.$$

It follows from (3.15) that

$$(3.16) \quad \begin{aligned} \|\Delta X\|_F &= \|S^{-1}T_1r_1\|_2 \\ &\leq \|S^{-1}T_1\|_2 \|r_1\|_2 \\ &\leq \left\| \left[\|Q\|_F S^{-1}, -\|A\|_F S^{-1}(I_{n^2} + \Pi)(I \otimes (A^T X)) \right] \right\|_2 \|r_1\|_2. \end{aligned}$$

Since $\|r_1\|_2 = \Delta_2$, according to (3.10) (when $i = 2$) and inequality (3.16), we arrive at (3.14). \square

3.2. Mixed and componentwise condition numbers. We now consider the the mixed and componentwise condition numbers of matrix equation (1.1). We first introduce some definitions and useful results about these two condition numbers.

For any $a, b \in \mathbb{R}^n$, we define $a./b = [c_1, c_2, \dots, c_n]^T$ with

$$c_i = \begin{cases} a_i/b_i, & \text{if } b_i \neq 0, \\ 0, & \text{if } a_i = b_i = 0, \\ \infty, & \text{otherwise.} \end{cases}$$

Then we introduce one distance function

$$d(a, b) = \|(a - b)./b\|_\infty = \max_{i=1,2,\dots,n} \left\{ \left| \frac{a_i - b_i}{b_i} \right| \right\}.$$

In the rest of this paper we assume $d(a, b) < \infty$ for any pair (a, b) . And we extend the function d to matrices M and N by $d(M, N) = d(\text{vec}(M), \text{vec}(N))$. For $\epsilon > 0$, we denote $B^0(a, \epsilon) = \{x | d(x, a) \leq \epsilon\}$.

DEFINITION 3.3. ([7]) Let $F : \mathbb{R}^p \rightarrow \mathbb{R}^q$ be a continuous mapping defined on an open set $\text{Dom}(F) \subset \mathbb{R}^p$ such that $0 \notin \text{Dom}(F)$ and $F(a) \neq 0$ for a given $a \in \mathbb{R}^p$.

(1) The mixed condition number of F at a is defined by

$$m(F, a) = \lim_{\epsilon \rightarrow 0} \sup_{\substack{x \in B^0(a, \epsilon) \\ x \neq a}} \frac{\|F(x) - F(a)\|_\infty}{\|F(a)\|_\infty} \frac{1}{d(x, a)}.$$

(2) Suppose $F(a) = [f_1(a), f_2(a), \dots, f_q(a)]^T$ such that $f_j(a) \neq 0$ for $j = 1, 2, \dots, q$. The componentwise condition number of F at a is defined by

$$c(F, a) = \lim_{\epsilon \rightarrow 0} \sup_{\substack{x \in B^0(a, \epsilon) \\ x \neq a}} \frac{d(F(x), F(a))}{d(x, a)}.$$

From Gohberg and Koltracht [7] or Cucker, Diaio, and Wei [4], if F is Fréchet differentiable at a , the explicit expressions of the mixed and componentwise condition numbers of F at a are given by the following lemma.

LEMMA 3.4. Suppose F is Fréchet differentiable at a . We have

(1) if $F(a) \neq 0$, then

$$m(F, a) = \frac{\|F'(a)\text{diag}(a)\|_\infty}{\|F(a)\|_\infty} = \frac{\|F'(a)\|_a|a\|_\infty}{\|F(a)\|_\infty};$$

(2) if $F(a) = [f_1(a), f_2(a), \dots, f_q(a)]^T$ such that $f_j(a) \neq 0$ for $j = 1, 2, \dots, q$, then

$$c(F, a) = \|(\text{diag}(F(a)))^{-1}F'(a)\text{diag}(a)\|_\infty = \left\| \frac{F'(a)|a|}{|F(a)|} \right\|_\infty.$$

THEOREM 3.5. Let X be the symmetric positive definite solution of equation (1.1). If $\frac{a^{1-p}\|A\|_2^2}{p} < 1$, then the mixed and componentwise condition numbers of matrix equation (1.1) are given by

$$m(\Phi) = \frac{\| |S^{-1}|\text{vec}(|Q|) + |S^{-1}(I_{n^2} \otimes \Pi)(I \otimes (A^T X))|\text{vec}(|A|) \|_\infty}{\|X\|_{\max}},$$

and

$$c(\Phi) = \left\| \frac{|S^{-1}|\text{vec}(|Q|) + |S^{-1}(I_{n^2} + \Pi)(I \otimes (A^T X))|\text{vec}(|A|)}{\text{vec}(|X|)} \right\|_\infty.$$

Furthermore, we have two simple upper bounds for $m(\Phi)$ and $c(\Phi)$ as follows:

$$m_U(\Phi) := \frac{\|S^{-1}\|_\infty \| |Q| + |A^T X||A| + |A^T||AX| \|_{\max}}{\|X\|_{\max}} \gtrsim m(\varphi),$$

and

$$c_U(\Phi) := \| \text{diag}^{-1}(\text{vec}(X))S^{-1} \|_\infty = \| |Q| + |A^T X||A| + |A^T||XA| \|_{\max} \gtrsim c(\varphi).$$

Proof. From (3.13) we know $\Phi'(\omega) = S^{-1}T$, where $\omega = (\text{vec}(Q)^T, \text{vec}(A)^T)^T$, according to (1) of Lemma 3.4, we get

$$\begin{aligned} m(\Phi) &= \frac{\| |S^{-1}T||\omega| \|_\infty}{\|\text{vec}(X)\|_\infty} \\ &= \frac{\left\| \left[|S^{-1}|, |S^{-1}(I_{n^2} + \Pi)(I \otimes (A^T X))| \right] \begin{pmatrix} |\text{vec}(Q)| \\ |\text{vec}(A)| \end{pmatrix} \right\|_\infty}{\|X\|_{\max}} \\ &= \frac{\| |S^{-1}|\text{vec}(|Q|) + |S^{-1}(I_{n^2} + \Pi)(I \otimes (A^T X))|\text{vec}(|A|) \|_\infty}{\|X\|_{\max}}. \end{aligned}$$

Similarly, from (2) of Lemma 3.4, we get

$$\begin{aligned} c(\Phi) &= \left\| \frac{|S^{-1}T||\omega|}{|\text{vec}(X)|} \right\|_\infty \\ &= \left\| \frac{|S^{-1}|\text{vec}(|Q|) + |S^{-1}(I_{n^2} + \Pi)(I \otimes (A^T X))|\text{vec}(|A|)}{\text{vec}(|X|)} \right\|_\infty. \end{aligned}$$

Note that

$$\begin{aligned} \| |S^{-1}T||\omega| \|_\infty &\leq \| |S^{-1}T|T||\omega| \|_\infty \\ &\leq \|S^{-1}\|_\infty \|T||\omega| \|_\infty \\ &\leq \|S^{-1}\|_\infty \| \text{vec}(|Q|) + |(I_{n^2} + \Pi)(I \otimes (A^T X))|\text{vec}(|A|) \|_\infty \\ &\leq \|S^{-1}\|_\infty \| |Q| + |A^T X||A| + |A^T||AX| \|_{\max}, \end{aligned}$$

it follows that

$$\begin{aligned} m(\Phi) &\leq \frac{\|S^{-1}\|_{\infty} \|T\| \|\omega\|_{\infty}}{\|X\|_{\max}} \\ &= \frac{\|S^{-1}\|_{\infty} (\|Q\| + |A^T X| \|A\| + |A^T| \|X A\|)_{\max}}{\|X\|_{\max}}. \end{aligned}$$

Similarly, it holds that

$$\begin{aligned} c(\Phi) &= \|\text{diag}^{-1}(\text{vec}(X)) \|S^{-1} T\| \|\omega\|_{\infty} \\ &\leq \|\text{diag}^{-1}(\text{vec}(X)) S^{-1}\|_{\infty} \|T\| \|\omega\|_{\infty} \\ &\leq \|\text{diag}^{-1}(\text{vec}(X)) S^{-1}\|_{\infty} (\|Q\| + |A^T X| \|A\| + |A^T| \|X A\|)_{\max}. \end{aligned}$$

The proof is completed. □

4. Numerical examples. In this section, we give three examples to show the efficiency of the proposed iterative method with stepsize parameter, and the sharpness of the three kinds of condition numbers. The stepsize parameter is different case by case. Our experiments were done in MATLAB 7.10.0, which has the unit roundoff $\mu \approx 2.2 \cdot 10^{-16}$ and the iterations terminate if the relative residuals $\rho(X_k)$ satisfies

$$\rho(X_k) = \frac{\|f(X_k^p + A^T X_k A - Q)\|_F}{\|X_k\|_F^p + \|A^T\|_F \|X_k\|_F \|A\|_F + \|Q\|_F} \leq n\mu.$$

In [13], we also proposed the following two iterative methods which work better than iteration (2.2)

$$(4.1) \quad \begin{cases} X_0 = \gamma I, \\ X_{k+1} = (Q - A^T X_k A)^{\frac{1}{p}} \end{cases}$$

and

$$(4.2) \quad \begin{cases} X_0 = 0, \\ X_{k+1} = (Q - A^T (Q - A^T X_k A)^{\frac{1}{p}} A)^{\frac{1}{p}}, \end{cases}$$

where $\gamma > 0$ in (4.1) is a real number such that

$$\gamma^p + \gamma \lambda_{\max}(A^T A) = \lambda_{\min}(Q).$$

We will compare the proposed iteration (2.3) with iteration (2.2) and the above two iterations.

EXAMPLE 4.1. (Example 4.1, [13]) Let matrix $A = \text{rand}(10) \cdot 10^{-2}$, $Q = \text{eye}(10)$, and $p = 2, 3, 4, 5, 6, 7$. We apply the basic fixed-point iteration (2.2), (4.1), (4.2) and the iteration (2.3) with stepsize parameter on equation (1.1). Figure 1 shows the iterations of (2.3) with different values of α when $p = 3$ and 7, respectively. The iterations before convergence, the CPU time and the relative residuals are shown in 1 and Table 2.

From Figure 1, we can see that the stepsize parameter affects the performance of iteration (2.3) significantly and when it is close to 1 (in our example it is close to 0.9) iteration (2.3) works more efficiently. From Table 1 and Table 2, we can see that iteration (2.3) with proper stepsize parameter converges much faster and uses less time for obtaining the symmetric positive definite solution.

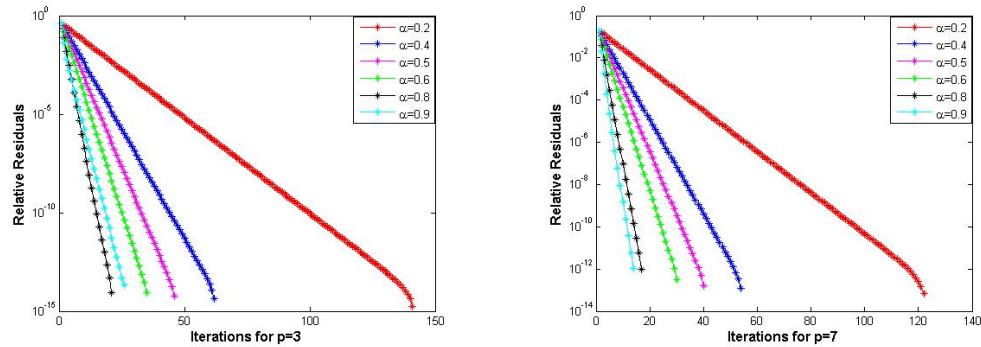


FIGURE 1. Iterations with different parameter α .

p	Iteration (2.3) with stepsize parameter				Iteration (2.2)		
	α	It(s)	CPU time	Relative res	It(s)	CPU time	relative res
2	0.79	22	0.01696	$1.27 \cdot 10^{-15}$	53	0.03924	$1.78 \cdot 10^{-15}$
3	0.82	19	0.01617	$1.93 \cdot 10^{-15}$	38	0.03234	$2.14 \cdot 10^{-15}$
4	0.83	18	0.01769	$4.54 \cdot 10^{-16}$	31	0.02978	$1.44 \cdot 10^{-15}$
5	0.86	16	0.01838	$8.89 \cdot 10^{-16}$	26	0.02930	$1.68 \cdot 10^{-15}$
6	0.88	14	0.01927	$1.96 \cdot 10^{-15}$	23	0.03121	$9.64 \cdot 10^{-16}$
7	0.89	13	0.02100	$2.02 \cdot 10^{-15}$	20	0.03203	$1.34 \cdot 10^{-15}$

TABLE 1

Comparison of iteration (2.3) with iteration (2.2).

EXAMPLE 4.2. Let

$$A = \begin{pmatrix} 0.1892 & 0.2406 & 0.1078 & 0.1682 \\ 0.0708 & 0.2020 & 0.0646 & 0.1774 \\ 0.1492 & 0.0138 & 0.2177 & 0.1643 \\ 0.0325 & 0.0228 & 0.0224 & 0.2160 \end{pmatrix},$$

$Q = \text{eye}(4)$ and $p = 3$. The perturbations in coefficient matrices are given by

$$\Delta A = (\text{rand}(4) \cdot 10^{-j}) \circ A \quad \text{and} \quad \Delta Q = (\text{rand}(4) \cdot 10^{-j}) \circ Q,$$

where j is a positive integer and \circ is the Hadamard product. Using iteration (2.3) with $X_0 = aI$, we can get the unique positive definite solution X of equation (1.1) and \tilde{X} of the corresponding perturbed equation. From Theorem 3.1 and Theorem 3.2, we get two local normwise perturbation bounds $\frac{\|\Delta X\|_F}{\|X\|_F} \leq k_{rel}^1 \Delta_1$ and $\frac{\|\Delta X\|_F}{\|X\|_F} \leq k_{rel}^2 \Delta_2$.

Let

$$\epsilon_0 = \min \{ \epsilon : |\Delta A| \leq \epsilon |A|, |\Delta Q| \leq \epsilon |Q| \}.$$

We obtain the local mixed and componentwise perturbation bounds

$$\|\Delta X\|_{\max} / \|X\|_{\max} \lesssim \epsilon_0 m_U(\varphi),$$

p	Iteration (2.3) with stepsize parameter			Iteration (4.1)		Iteration (4.2)	
	α	It(s)	CPU time	It(s)	CPU time	It(s)	CPU time
2	0.79	22	0.01696	48	0.0263	28	0.0276
3	0.82	19	0.01617	35	0.0211	20	0.0227
4	0.83	18	0.01769	28	0.0204	17	0.0233
5	0.86	16	0.01838	24	0.0221	14	0.0239
6	0.88	14	0.01927	21	0.0234	13	0.0279
7	0.89	13	0.02100	18	0.0247	11	0.0293

TABLE 2
 Comparisons of iteration (2.3) with iteration (4.1) and iteration (4.2).

j	3	4	5	6
$\frac{\ \Delta X\ _F}{\ X\ _F}$	2.8299e-004	2.3131e-005	1.4513e-006	2.0563e-007
$k_{rel}^1(\varphi)\Delta_1$	2.1257e-003	1.7298e-004	1.0497e-005	1.4720e-006
$k_{rel}^2(\varphi)\Delta_2$	3.1625e-004	2.8783e-005	1.8903e-006	2.6481e-007
$\frac{\ \Delta X\ _{\max}}{\ X\ _{\max}}$	3.2166e-004	2.9719e-005	1.8111e-006	2.8408e-007
$m_U(\varphi)\epsilon_0$	4.5666e-004	4.3999e-005	2.3324e-006	3.9336e-007
$\left\ \frac{\text{vec}(\Delta X)}{\text{vec}(X)} \right\ _{\infty}$	1.1047e-003	1.1322e-004	8.7107e-006	1.1544e-006
$c_U(\varphi)\epsilon_0$	3.0964e-002	29834e-003	1.5815e-004	2.6672e-005

TABLE 3
 Upper bounds given by normwise, mixed and componentwise condition numbers.

and

$$\|\text{vec}(\Delta X) ./ \text{vec}(X)\|_{\max} \lesssim \epsilon_0 c_U(\varphi).$$

Table 3 shows that the perturbation bounds given by the three kinds of condition numbers are very sharp.

EXAMPLE 4.3. Let

$$A = \begin{pmatrix} 2 & 0 & 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 & 3 & 0 \\ 0 & 1 & 0 & 0 & 1 & 2 \end{pmatrix},$$

$$Q = \begin{pmatrix} 1.6740 & 0.1069 & 0.2218 & 0.0033 & 0.0775 & 0.2502 \\ 0.1069 & 1.8446 & 0.2356 & 0.2854 & 0.2327 & 0.2553 \\ 0.2218 & 0.2356 & 1.7428 & 0.0088 & 0.2549 & 0.0884 \\ 0.0033 & 0.2854 & 0.0088 & 1.1526 & 0.1433 & 0.1666 \\ 0.0775 & 0.2327 & 0.2549 & 0.1433 & 1.6075 & 0.4349 \\ 0.2502 & 0.2553 & 0.0884 & 0.1666 & 0.4349 & 2.1978 \end{pmatrix},$$

j	7	9	11	13
$\frac{\ \Delta X\ _F}{\ X\ _F}$	1.7180e-09	1.7179e-11	1.7192e-13	1.6405e-15
$k_{rel}^1(\varphi)\Delta_1$	2.8176e-08	2.8176e-10	2.8176e-12	2.8176e-14
$k_{rel}^2(\varphi)\Delta_2$	4.5508e-09	4.5508e-11	4.5508e-13	4.5508e-15
$\frac{\ \Delta X\ _{\max}}{\ X\ _{\max}}$	1.8107e-09	1.8106e-11	1.8074e-13	1.9528e-15
$m_U(\varphi)\epsilon_0$	1.7763e-08	1.7763e-10	1.7763e-12	1.7763e-14
$\left\ \frac{\text{vec}(\Delta X)}{\text{vec}(X)} \right\ _{\infty}$	2.6329e-07	2.6315e-09	2.4086e-11	3.8679e-12
$c_U(\varphi)\epsilon_0$	6.9164e-05	6.9164e-07	6.9164e-09	6.9164e-11

TABLE 4
 Upper bounds given by normwise, mixed and componentwise condition numbers.

and $p = 100$. The perturbations in coefficient matrices are given by

$$\Delta A = (\text{rand}(6) \cdot 10^{-j}) \circ A \quad \text{and} \quad \Delta Q = (\text{rand}(6) \cdot 10^{-j}) \circ Q,$$

where j is a positive integer and \circ is the Hadamard product. This is another example. But in this example, we have a slightly greater order of matrix and also greater power of p . The results are shown in Table 4. And it can be easily deduced from the results that our estimations are almost tight.

5. Conclusion. In this paper, we propose an iteration with stepsize parameter for obtaining the symmetric positive definite solution (1.1). Different stepsize parameter affects the performance of the proposed iteration significantly, but it works efficiently with a proper stepsize parameter. We also investigate the normwise, mixed and componentwise condition numbers of equation (1.1).

Acknowledgment. The authors thank the valuable comments and suggestions of an anonymous referee and of the handing editor, Professor James Nagy, that improved the presentation of the paper.

REFERENCES

- [1] V. Berinde. Generalized coupled fixed point theorems for mixed monotone mappings in partially ordered metric spaces. *Nonlinear Analysis: Theory, Methods Applications*, 74:7347–7355, 2011.
- [2] T.G. Bhaskar and V. Lakshmikantham. Fixed point theorems in partially ordered metric spaces and applications. *Nonlinear Analysis: Theory, Methods, Applications*, 65:1379–1393, 2006.
- [3] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics, Vol. 169, Springer-Verlag, New York, 1997.
- [4] F. Cucker, H. Diao, and Y. Wei. On mixed and componentwise condition numbers fro Moore-Penrose inverse and linear least squares problems. *Mathematics of Computation*, 76:947–963, 2007.
- [5] S.M. El-Sayed and A.C.M. Ran. On an iteration method for solving a class of nonlinear matrix equations. *SIAM Journal on Matrix Analysis and Applications*, 23:632–645, 2001.
- [6] A. Ferrante and B.C. Levy. Hermitian solutions of the eqaution $X = Q + NX^{-1}N^*$. *Linear Algebra and its Applications*, 246:359–373, 1996.
- [7] A. Gohberg and I. Koltracht. Mixed, componentwise, and structed condition numbers. *SIAM Journal on Matrix Analysis and Applications*, 14:688–704, 1993.
- [8] D.A. Harville. *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, New York, 1997.
- [9] I.G. Ivanov, V.I. Hasanov, and F. Uhlig. Improved methods and starting values to solve the matrix equations $X \pm A^* X^{-1} A = I$ iteratively. *Mathematics of Computations*, 74:263–278, 2004.
- [10] Z. Jia and M. Wei. Solvability and sensitivity theory of polynomial matrix equation $X^s + A^T X^t A = Q$. *Applied Mathematics and Computation*, 209:230–237, 2009.

- [11] X.-G. Liu and H. Gao. On the positive definite solutions of the matrix equations $X^s \pm A^T X^{-t} A = I_n$. *Linear Algebra and Applications*, 368:83–97, 2003.
- [12] B. Meini. Efficient computation of the extreme solutions of $X + A^* X^{-1} A = Q$ and $X - A^* X^{-1} A = Q$. *Mathematics of Computation*, 71:1189–1204, 2002.
- [13] J. Meng and H.-M. Kim. The positive definite solution to a nonlinear matrix equation. *Linear and Multilinear Algebra*, 64:653–666, 2016.
- [14] A.C.M. Ran and M.C.B. Reurings. A nonlinear matrix equation connected to interpolation theory. *Linear Algebra and its Applications*, 379:289–302, 2004.
- [15] A.C.M. Ran and M.C.B. Reurings. On the nonlinear matrix equation $X + A^* \mathcal{F}(X) A = Q$: solution and perturbation theory. *Linear Algebra and its Applications*, 346:15–26, 2002.
- [16] A.C.M. Ran, M.C.B. Reurings. The symmetric linear matrix equation. *Electronic Journal of Linear Algebra*, 9:93–107, 2002.
- [17] M.C.B. Reurings. Contractive maps on normed linear spaces and their applications to nonlinear matrix equations. *Linear Algebra and its Applications*, 418:292–311, 2006.
- [18] J.R. Rice. A theory of condition. *SIAM Journal on Numerical Analysis*, 3:287–310, 1966.
- [19] L.A. Sakhnovich. *Interpolation Theory and Its Applications*. Mathematics and Its Applications, Vol. 428, Kluwer Academic Publishers, Dordrecht, 1997.
- [20] J.-H. Seo and H.-M. Kim. Convergence of pure and relaxed Newton methods for solving a matrix polynomial equation arising in stochastic models. *Linear Algebra and its Applications*, 440:34–49, 2014.
- [21] J. Sun. Condition numbers of algebraic Riccati equations in the Frobenius norm. *Linear Algebra and its Applications*, 350:237–261, 2002.
- [22] S. Wang, H. Yang, and H. Li. Condition numbers for the nonlinear matrix equation and their statistical estimation. *Linear Algebra and its Applications*, 482:221–240, 2015.
- [23] X. Zhan. *Matrix Inequalities*. Springer-Verlag, Berlin, 2002.