ITERATIVE SOLUTION OF TWO MATRIX EQUATIONS

CHUN-HUA GUO AND PETER LANCASTER

Abstract. We study iterative methods for finding the maximal Hermitian positive definite solutions of the matrix equations $X + A^*X^{-1}A = Q$ and $X - A^*X^{-1}A = Q$, where Q is Hermitian positive definite. General convergence results are given for the basic fixed point iteration for both equations. Newton's method and inversion free variants of the basic fixed point iteration are discussed in some detail for the first equation. Numerical results are reported to illustrate the convergence behaviour of various algorithms.

1. Introduction

In this paper, we are concerned with the iterative solution of the matrix equations

$$X + A^*X^{-1}A = Q, \tag{1.1}$$

and

$$X - A^*X^{-1}A = Q. \tag{1.2}$$

In both cases, the matrix Q is $m \times m$ Hermitian positive definite and Hermitian positive definite solutions are required. These two equations have been studied recently by several authors (see [1], [2], [3], [4], [16], [17]). For the application areas in which the equations arise, see the references given in [1] and [4].

For Hermitian matrices X and Y, we write $X \geq Y (X > Y)$ if $X - Y$ is positive semidefinite (definite). A Hermitian solution X_+ of a matrix equation is called maximal if $X_+ \geq X$ for any Hermitian solution X of the matrix equation. A minimal solution can be defined similarly.

It is proved in [3] that if (1.1) has a positive definite solution, then it has a maximal Hermitian solution X_+ and a minimal Hermitian solution X_-. Indeed, we have $0 < X_- \leq X \leq X_+$ for any Hermitian solution X of (1.1). Moreover, we have $\rho(X_+^{-1}A) \leq 1$ (see, e.g., [16]), where $\rho(\cdot)$ is the spectral radius.

When the matrix A is nonsingular, the minimal positive definite solution of (1.1) can be found via the maximal solution of another equation of the same type (cf. [3, Thm. 3.3]). In [3], an algorithm was presented to find the minimal solution of the equation (1.1) for the case where A is singular. The algorithm was based on a recursive reduction process. The reduction process is useful in showing that the minimal positive definite solution of (1.1) exists even if the matrix A is singular. However, it is usually impossible to find the minimal solution using that algorithm. The reason is that the minimal solution, as a function of (A, Q), is generally not continuous at a singular matrix A for fixed Q. The situation is already clear for

1991 *Mathematics Subject Classification.* Primary 15A24, 65F10; Secondary 65H10, 93B40.

Key words and phrases. matrix equations, positive definite solution, fixed point iteration, Newton's method, convergence rate, matrix pencils .

the scalar equation $X + \epsilon^2 X^{-1} = 1$. The minimal solution of this equation is not continuous at $\epsilon = 0$. We will therefore limit our discussion to the maximal solution.

The equation (1.2) is quite different. As the scalar case suggests, it always has a unique positive definite solution, which is the maximal solution (see [4]).

In Section 2, we discuss the convergence behaviour of the basic fixed point iteration for the maximal solution of (1.1) and (1.2). In Section 3, we study the convergence behaviour of inversion free variants of the basic fixed point iteration. In general, these algorithms are linearly convergent and do not perform well when there are eigenvalues of $X_+^{-1}A$ on, or near, the unit circle. However, Newton's method can be applied. This has a global convergence property when applied to (1.1) and, although it is step-wise expensive, offers considerable advantages in this situation. Sections 4 and 5 are devoted to properties of the Newton iteration. In Section 6, we give matrix pencil descriptions for the eigenvalues of $X_+^{-1}A$. This admits the computation of these eigenvalues without prior knowledge of X_+. Some numerical examples are reported in Section 7.

Throughout the paper, $\|\cdot\|$ will be the spectral norm for square matrices unless otherwise noted.

2. Basic fixed point iteration

The maximal solution X_+ of (1.1) can be found by the following basic fixed point iteration:

Algorithm 2.1.

$$X_0 = Q,$$
$$X_{n+1} = Q - A^* X_n^{-1} A, \quad n = 0, 1, \ldots.$$

For Algorithm 2.1, we have $X_0 \geq X_1 \geq \cdots$, and $\lim_{n\to\infty} X_n = X_+$ (see, e.g., [3]).

The following result is given in [16].

Theorem 2.2. *For any $\epsilon > 0$,*

$$\|X_{n+1} - X_+\| \leq (\|X_+^{-1}A\| + \epsilon)^2 \|X_n - X_+\|$$

for all n large enough.

We now show that the above result can be improved considerably.

Theorem 2.3. *For all $n \geq 0$,*

$$\|X_{n+1} - X_+\| \leq \|X_+^{-1}A\|^2 \|X_n - X_+\|.$$

Moreover,

$$\limsup_{n\to\infty} \sqrt[n]{\|X_n - X_+\|} \leq (\rho(X_+^{-1}A))^2.$$

Proof. Since $X_{n+1} = Q - A^* X_n^{-1} A$ and $X_+ = Q - A^* X_+^{-1} A$, we have

$$\begin{aligned} X_{n+1} - X_+ =& A^*(X_+^{-1} - X_n^{-1})A \\ =& A^*(X_+^{-1} + X_n^{-1} - X_+^{-1})(X_n - X_+)X_+^{-1}A \\ =& A^* X_+^{-1}(X_n - X_+)X_+^{-1}A \\ & - A^* X_+^{-1}(X_n - X_+)X_n^{-1}(X_n - X_+)X_+^{-1}A. \end{aligned}$$

Thus,

$$0 \leq X_{n+1} - X_+ \leq A^* X_+^{-1}(X_n - X_+)X_+^{-1}A. \tag{2.1}$$

Therefore,

$$\|X_{n+1} - X_+\| \le \|A^* X_+^{-1}(X_n - X_+)X_+^{-1}A\| \le \|X_+^{-1}A\|^2 \|X_n - X_+\|.$$

By repeated application of (2.1), we get

$$0 \le X_n - X_+ \le (A^* X_+^{-1})^n (X_0 - X_+)(X_+^{-1}A)^n.$$

Hence,

$$\|X_n - X_+\| \le \|(X_+^{-1}A)^n\|^2 \|X_0 - X_+\|,$$

and

$$\begin{aligned}\limsup_{n\to\infty} \sqrt[n]{\|X_n - X_+\|} &\le \lim_{n\to\infty} \sqrt[n]{\|(X_+^{-1}A)^n\|^2 \|X_0 - X_+\|} \\ &= (\rho(X_+^{-1}A))^2.\end{aligned}$$

In the last equality, we have used the fact that $\lim \|B^n\|^{1/n} = \rho(B)$ for any square matrix B and any norm. □

We mentioned earlier that $\rho(X_+^{-1}A) \le 1$ is always true. From the second part of the above result, we know that the convergence of the fixed point iteration is R-linear whenever $\rho(X_+^{-1}A) < 1$. For detailed definitions of the rates of convergence, see [13]. Zhan asked in [16] whether $\rho(X_+^{-1}A) \le 1$ implies $\|X_+^{-1}A\| \le 1$. This is not the case and, in fact, it is possible to have $\|X_+^{-1}A\| > 1$ when $\rho(X_+^{-1}A) < 1$. If $\rho(X_+^{-1}A) = 1$, the convergence of the fixed point iteration is typically sublinear.

Example 2.1. We consider the scalar case of (1.1) with $A = \frac{1}{2}$ and $Q = 1$, i.e.,

$$X + \frac{1}{4}X^{-1} = 1.$$

Clearly, $X_+ = \frac{1}{2}$ and $\rho(X_+^{-1}A) = 1$. For the fixed point iteration

$$\begin{aligned} X_0 &= 1, \\ X_{n+1} &= 1 - \frac{1}{4}X_n^{-1}, \quad n = 0, 1, \ldots, \end{aligned}$$

we have $X_0 > X_1 > \cdots$, and $\lim_{n\to\infty} X_n = \frac{1}{2}$. Note that

$$X_{n+1} - \frac{1}{2} = X_n - \frac{1}{2} - (X_n - \frac{1}{2})^2 X_n^{-1}.$$

Thus,

$$\lim_{n\to\infty} \sqrt[n]{X_n - \frac{1}{2}} = \lim_{n\to\infty} \frac{X_{n+1} - \frac{1}{2}}{X_n - \frac{1}{2}} = 1,$$

i.e., the convergence is sublinear.

For the matrix equation (1.2), the maximal solution X_+ can also be found by a fixed point iteration similar to that of Algorithm 2.1. Thus, we consider:

Algorithm 2.4.

$$\begin{aligned} X_0 &= Q, \\ X_{n+1} &= Q + A^* X_n^{-1} A, \quad n = 0, 1, \ldots. \end{aligned}$$

For Algorithm 2.4, we have $X_0 \le X_2 \le X_4 \le \cdots$, $X_1 \ge X_3 \ge X_5 \ge \cdots$, and $\lim_{n\to\infty} X_n = X_+$ (see [4]).

The following result is immediate (cf. Theorem 2.2).

Theorem 2.5. *For Algorithm* 2.4 *and any* $\epsilon > 0$,

$$\|X_{n+1} - X_+\| \le (\|X_+^{-1}A\|^2 + \epsilon)\|X_n - X_+\|$$

for all n *large enough.*

However, more can be said about Algorithm 2.4.

Theorem 2.6. *For Algorithm* 2.4, *we have*

$$\|X_{2k} - X_+\| \le \|X_+^{-1}A\|^2\|X_{2k-1} - X_+\| \tag{2.2}$$

for all $k \ge 1$, *and*

$$\limsup_{n\to\infty} \sqrt[n]{\|X_n - X_+\|} \le (\rho(X_+^{-1}A))^2 < 1.$$

Proof. Observe that

$$\begin{aligned} X_{n+1} - X_+ =& - A^*X_n^{-1}(X_n - X_+)X_+^{-1}A \\ =& - A^*X_+^{-1}(X_n - X_+)X_+^{-1}A \\ &+ A^*X_+^{-1}(X_n - X_+)X_n^{-1}(X_n - X_+)X_+^{-1}A. \end{aligned}$$

We have

$$0 \le X_+ - X_{2k} \le A^*X_+^{-1}(X_{2k-1} - X_+)X_+^{-1}A, \tag{2.3}$$

from which (2.2) follows. We also have

$$\begin{aligned} X_{2k+1} - X_+ =& A^*X_+^{-1}(X_+ - X_{2k})X_+^{-1}A \\ &+ A^*X_+^{-1}(X_+ - X_{2k})X_{2k}^{-1}(X_+ - X_{2k})X_+^{-1}A \\ =& A^*X_+^{-1}(X_+ - X_{2k})^{1/2} \\ &[I + (X_+ - X_{2k})^{1/2}X_{2k}^{-1}(X_+ - X_{2k})^{1/2}](X_+ - X_{2k})^{1/2}X_+^{-1}A. \end{aligned}$$

For any $\epsilon > 0$, there exists a k_0 such that

$$(X_+ - X_{2k})^{1/2}X_{2k}^{-1}(X_+ - X_{2k})^{1/2} \le \epsilon I$$

for all $k \ge k_0$. Therefore,

$$0 \le X_{2k+1} - X_+ \le (1+\epsilon)A^*X_+^{-1}(X_+ - X_{2k})X_+^{-1}A \tag{2.4}$$

for all $k \ge k_0$. Combining (2.3) and (2.4), we obtain for $k \ge k_0$

$$0 \le X_{2k+1} - X_+ \le (1+\epsilon)^{k-k_0+1}(A^*X_+^{-1})^{2(k-k_0)+1}(X_+ - X_{2k_0})(X_+^{-1}A)^{2(k-k_0)+1}.$$

Thus,

$$\|X_{2k+1} - X_+\| \le (1+\epsilon)^{k-k_0+1}\|(X_+^{-1}A)^{2(k-k_0)+1}\|^2\|X_+ - X_{2k_0}\|.$$

Similarly,

$$\|X_{2k+2} - X_+\| \le (1+\epsilon)^{k-k_0+1}\|(X_+^{-1}A)^{2(k-k_0)+2}\|^2\|X_+ - X_{2k_0}\|.$$

Therefore,

$$\limsup_{n\to\infty} \sqrt[n]{\|X_n - X_+\|} \le \sqrt{1+\epsilon}\,(\rho(X_+^{-1}A))^2.$$

Since $\epsilon > 0$ is arbitrary, we have

$$\limsup_{n\to\infty} \sqrt[n]{\|X_n - X_+\|} \le (\rho(X_+^{-1}A))^2.$$

Since

$$X_+ - (X_+^{-1}A)^*X_+(X_+^{-1}A) = Q,$$

with $Q > 0$ and $X_+ > 0$, we have $\rho(X_+^{-1}A) < 1$ (see, e.g., [12, p. 451]). □

3. Inversion free variant of the basic fixed point iteration

In [16], Zhan proposed an inversion free variant of the basic fixed point iteration for the maximal solution of (1.1) when $Q = I$. For general positive definite Q, Zhan's algorithm takes the following form:

Algorithm 3.1. Take $X_0 = Q, Y_0 = Q^{-1}$. For $n = 0, 1, \ldots,$ compute

$$\begin{aligned} X_{n+1} &= Q - A^*Y_nA, \\ Y_{n+1} &= Y_n(2I - X_nY_n). \end{aligned}$$

The convergence of Algorithm 3.1 was established in [16] for $Q = I$. Zhan's result can easily be transplanted and we have

Theorem 3.2. *If* (1.1) *has a positive definite solution then, for Algorithm* 3.1, $X_0 \geq X_1 \geq \cdots$, $Y_0 \leq Y_1 \leq \cdots$, *and* $\lim_{n\to\infty} X_n = X_+$, $\lim_{n\to\infty} Y_n = X_+^{-1}$.

The problem of convergence rate for Algorithm 3.1 was not solved in [16]. We now establish the following result:

Theorem 3.3. *For any* $\epsilon > 0$, *we have*

$$\|Y_{n+1} - X_+^{-1}\| \leq (\|AX_+^{-1}\| + \epsilon)^2\|Y_{n-1} - X_+^{-1}\| \tag{3.1}$$

and

$$\|X_{n+1} - X_+\| \leq \|A\|^2\|Y_n - X_+^{-1}\| \tag{3.2}$$

for all n *large enough. If* A *is nonsingular, we also have*

$$\|X_{n+1} - X_+\| \leq (\|X_+^{-1}A\| + \epsilon)^2\|X_{n-1} - X_+\| \tag{3.3}$$

for all n *large enough.*

Proof. We have from Algorithm 3.1

$$\begin{aligned} Y_{n+1} &= Y_n(2I - (Q - A^*Y_{n-1}A)Y_n) \\ &= 2Y_n - Y_nQY_n + Y_nA^*(X_+^{-1} + Y_{n-1} - X_+^{-1})AY_n \\ &= 2Y_n - Y_nX_+Y_n + Y_nA^*(Y_{n-1} - X_+^{-1})AY_n. \end{aligned}$$

Thus,

$$\begin{aligned} X_+^{-1} - Y_{n+1} &= X_+^{-1} - Y_n + Y_nX_+Y_n - Y_n + Y_nA^*(X_+^{-1} - Y_{n-1})AY_n \\ &= (X_+^{-1} - Y_n)X_+(X_+^{-1} - Y_n) + Y_nA^*(X_+^{-1} - Y_{n-1})AY_n. \end{aligned} \tag{3.4}$$

The inequality (3.1) follows since $\|Y_n - X_+^{-1}\| \leq \|Y_{n-1} - X_+^{-1}\|$ and $\lim Y_n = X_+^{-1}$. The inequality (3.2) is true since

$$X_{n+1} - X_+ = A^*(X_+^{-1} - Y_n)A. \tag{3.5}$$

If A is nonsingular, we have by (3.4) and (3.5)

$$X_{n+1} - X_+ = (X_n - X_+)A^{-1}X_+(X_+^{-1} - Y_{n-1})A + A^*Y_{n-1}(X_{n-1} - X_+)Y_{n-1}A.$$

Therefore, since $\|X_n - X_+\| \leq \|X_{n-1} - X_+\|$, (3.3) is true for all n large enough. □

The above proof shows that Algorithm 3.1 should be modified as follows to improve the convergence properties:

Algorithm 3.4. Take $X_0 = Q$, $0 < Y_0 \le Q^{-1}$. For $n = 0, 1, \dots,$ compute

$$Y_{n+1} = Y_n(2I - X_n Y_n),$$
$$X_{n+1} = Q - A^* Y_{n+1} A.$$

Note that one convenient choice of Y_0 is $Y_0 = I/\|Q\|_\infty$. We can also use this choice of Y_0 in Algorithm 3.1. Theorems 3.2 and 3.3 remain true for any Y_0 such that $0 < Y_0 \le Q^{-1}$.

Lemma 3.5 (cf. [16])**.** *If C and P are Hermitian matrices of the same order with $P > 0$, then $CPC + P^{-1} \ge 2C$.*

Theorem 3.6. *If* (1.1) *has a positive definite solution and $\{X_n\}, \{Y_n\}$ are determined by Algorithm* 3.4*, then $X_0 \ge X_1 \ge \cdots, \lim_{n\to\infty} X_n = X_+; Y_0 \le Y_1 \le \cdots, \lim_{n\to\infty} Y_n = X_+^{-1}$.*

Proof. It is clear that

$$X_0 \ge X_1 \ge \cdots \ge X_n \ge X_+, \quad Y_0 \le Y_1 \le \cdots \le Y_n \le X_+^{-1} \tag{3.6}$$

is true for $n = 1$. Assume (3.6) is true for $n = k$. We have by Lemma 3.5

$$Y_{k+1} = 2Y_k - Y_k X_k Y_k \le X_k^{-1} \le X_+^{-1}.$$

Therefore,

$$X_{k+1} = Q - A^* Y_{k+1} A \ge Q - A^* X_+^{-1} A = X_+.$$

Since $Y_k \le X_{k-1}^{-1} \le X_k^{-1}$, we have $Y_k^{-1} \ge X_k$. Thus,

$$Y_{k+1} - Y_k = Y_k(Y_k^{-1} - X_k)Y_k \ge 0,$$

and

$$X_{k+1} - X_k = -A^*(Y_{k+1} - Y_k)A \le 0.$$

We have now proved (3.6) for $n = k + 1$. Therefore, (3.6) is true for all n, and the limits $\lim_{n\to\infty} X_n$ and $\lim_{n\to\infty} Y_n$ exist. As in [16], we have $\lim X_n = X_+$, and $\lim Y_n = X_+^{-1}$. □

Theorem 3.7. *For Algorithm* 3.4 *and any $\epsilon > 0$, we have*

$$\|Y_{n+1} - X_+^{-1}\| \le (\|AX_+^{-1}\| + \epsilon)^2 \|Y_n - X_+^{-1}\|$$

and

$$\|X_n - X_+\| \le \|A\|^2 \|Y_n - X_+^{-1}\| \tag{3.7}$$

for all n large enough. If A is nonsingular, we also have

$$\|X_{n+1} - X_+\| \le (\|X_+^{-1}A\| + \epsilon)^2 \|X_n - X_+\|$$

for all n large enough.

Proof. The proof is very similar to that of Theorem 3.3. □

We see from the estimates in Theorem 3.3 and Theorem 3.7 that Algorithm 3.4 can be faster than Algorithm 3.1 by a factor of 2. Compared with Algorithm 2.1, Algorithm 3.4 needs more computational work per iteration. However, Algorithm 3.4 has better numerical properties since matrix inversions have been avoided. Algorithm 3.4 is particularly useful on a parallel computing system, since matrix-matrix multiplication can be carried out in parallel very efficiently (see, e.g., [6]).

For Algorithm 3.4, R-linear convergence can be guaranteed whenever $\rho(X_+^{-1}A) < 1$. This will be a consequence of the following general result.

Theorem 3.8 (cf. [10, p. 21])**.** *Let T be a (nonlinear) operator from a Banach space E into itself and $x^* \in E$ be a solution of $x = Tx$. If T is Fréchet differentiable at x^* with $\rho(T'_{x^*}) < 1$, then the iterates $x_{n+1} = Tx_n$ $(n = 0, 1, \ldots)$ converge to x^*, provided that x_0 is sufficiently close to x^*. Moreover, for any $\epsilon > 0$,*

$$\|x_n - x^*\| \le c(x_0; \epsilon)(\rho(T'_{x^*}) + \epsilon)^n,$$

where $\|\cdot\|$ is the norm in E and $c(x_0; \epsilon)$ is a constant independent of n.

Corollary 3.9. *For Algorithm* 3.4*, we have*

$$\limsup_{n\to\infty} \sqrt[n]{\|X_n - X_+\|} \le (\rho(X_+^{-1}A))^2. \tag{3.8}$$

Proof. For Algorithm 3.4, we have

$$Y_{n+1} = T(Y_n), \quad n = 0, 1, \ldots,$$

where the operator T is defined on $\mathbf{C}^{m\times m}$ (m is the order of Q) by

$$T(Y) = 2Y - YQY + YA^*YAY.$$

It is found that the Fréchet derivative $T'_Y : \mathbf{C}^{m\times m} \to \mathbf{C}^{m\times m}$ is given by

$$T'_Y(Z) = 2Z - ZQY - YQZ + ZA^*YAY + YA^*YAZ + YA^*ZAY.$$

Therefore,

$$T'_{X_+^{-1}}(Z) = X_+^{-1}A^*ZAX_+^{-1}.$$

The spectrum of $T'_{X_+^{-1}}$ consists of eigenvalues only. If λ is an eigenvalue of $T'_{X_+^{-1}}$, we have

$$X_+^{-1}A^*ZAX_+^{-1} = \lambda Z \tag{3.9}$$

for some $Z \ne 0$. If $|\lambda| > (\rho(X_+^{-1}A))^2$, the equation (3.9) would have zero as the only solution (see, e.g. [11, p. 100]). Therefore, $\rho(T'_{X_+^{-1}}) \le (\rho(X_+^{-1}A))^2$. In fact, we have $\rho(T'_{X_+^{-1}}) = (\rho(X_+^{-1}A))^2$, since (3.9) has a nonzero solution for $\lambda = (\rho(X_+^{-1}A))^2$.

By Theorem 3.8, we have

$$\limsup_{n\to\infty} \sqrt[n]{\|Y_n - X_+^{-1}\|} \le \rho(T'_{X_+^{-1}}) = (\rho(X_+^{-1}A))^2.$$

In view of (3.7), we also have (3.8). □

For equation (1.2), we can also have an algorithm similar to Algorithm 3.4. However, the algorithm is not always convergent.

4. Preliminaries on Newton's method

For equations (1.1) and (1.2), the convergence of the algorithms in the above two sections may be very slow when $X_+^{-1}A$ has eigenvalues close to (or even on) the unit circle. In these situations, Newton's method can be recommended.

The equation (1.1) is a special discrete algebraic Riccati equation (DARE), if we are willing to relax certain restrictions normally imposed on such equations for the purpose of analysis. Therefore, we will start with a review of some previous results on Newton's method for DAREs.

We consider a DARE of the form

$$-X + A^*XA + Q - (C + B^*XA)^*(R + B^*XB)^{-1}(C + B^*XA) = 0, \tag{4.1}$$

where $A, Q \in \mathbf{C}^{n\times n}, B \in \mathbf{C}^{n\times m}, C \in \mathbf{C}^{m\times n}, R \in \mathbf{C}^{m\times m}$, and $Q^* = Q, R^* = R$. We denote by $\mathcal{R}(X)$ the left-hand side of (4.1).

Let $\mathcal{H}$ be the set of Hermitian matrices in $\mathbf{C}^{n\times n}$ and let $\mathcal{D} = \{X \in \mathcal{H} \mid R + B^*XB$ is invertible$\}$. We have $\mathcal{R} : \mathcal{D} \to \mathcal{H}$. It is assumed throughout that $\mathcal{D}$ is nonempty and that there is an $X \in \mathcal{D}$ such that $R + B^*XB > 0$. The first Fréchet derivative of $\mathcal{R}$ at a matrix $X \in \mathcal{D}$ is a linear map $\mathcal{R}'_X : \mathcal{H} \to \mathcal{H}$ given by

$$\mathcal{R}'_X(H) = -H + \hat{A}^*H\hat{A}, \tag{4.2}$$

where $\hat{A} = A - B(R + B^*XB)^{-1}(C + B^*XA)$.

For $A \in \mathbf{C}^{n\times n}$ and $B \in \mathbf{C}^{n\times m}$, the pair (A, B) is said to be d-stabilizable if there is a $K \in \mathbf{C}^{m\times n}$ such that $A - BK$ is d-stable, i.e., all its eigenvalues are in the open unit disk. The following result is a modification of Theorem 13.1.1 in [11]. It has been noted in [7] that the matrix R does not need to be invertible.

Theorem 4.1. *Let (A, B) be a d-stabilizable pair and assume that there is a Hermitian solution $\tilde{X}$ of the inequality $\mathcal{R}(X) \geq 0$ for which $R+B^*\tilde{X}B > 0$. Then there exists a maximal Hermitian solution X_+ of $\mathcal{R}(X) = 0$. Moreover, $R+B^*X_+B > 0$ and all the eigenvalues of $A - B(R + B^*X_+B)^{-1}(C + B^*X_+A)$ lie in the closed unit disk.*

The Newton method for the solution of (4.1) is

$$X_i = X_{i-1} - (\mathcal{R}'_{X_{i-1}})^{-1}\mathcal{R}(X_{i-1}), \quad i = 1, 2, \ldots, \tag{4.3}$$

given that the maps $\mathcal{R}'_{X_i}(i = 0, 1, \ldots)$ are all invertible.

When we apply Newton's method to the DARE (4.1) with (A, B) d-stabilizable, the initial matrix X_0 is chosen so that $A - B(R + B^*X_0B)^{-1}(C + B^*X_0A)$ is d-stable. The usual way to generate such an X_0 is as follows. We choose $L_0 \in \mathbf{C}^{m\times n}$ such that $A_0 = A - BL_0$ is d-stable, and then take X_0 to be the unique solution of the Stein equation

$$X_0 - A_0^*X_0A_0 = Q + L_0^*RL_0 - C^*L_0 - L_0^*C. \tag{4.4}$$

In view of (4.2), the Newton iteration (4.3) can be rewritten as

$$X_i - A_i^*X_iA_i = Q + L_i^*RL_i - C^*L_i - L_i^*C, \quad i = 1, 2, \ldots, \tag{4.5}$$

where

$$L_i = (R + B^*X_{i-1}B)^{-1}(C + B^*X_{i-1}A)$$

and

$$A_i = A - BL_i.$$

Theorem 4.2. *Under the conditions of Theorem* 4.1 *and for any $L_0 \in \mathbf{C}^{m\times n}$ such that $A_0 = A-BL_0$ is d-stable, starting with the Hermitian matrix X_0 determined by* (4.4)*, the recursion* (4.5) *determines a sequence of Hermitian matrices $\{X_i\}_{i=0}^{\infty}$ for which $A-B(R+B^*X_iB)^{-1}(C+B^*X_iA)$ is d-stable for $i = 0, 1, \ldots$, $X_0 \geq X_1 \geq \cdots$, and $\lim_{i\to\infty} X_i = X_+$.*

An important feature of Newton's method applied to the Riccati equation is that the convergence is not local. The application of Newton's method to the Riccati equation was initiated in [8] under some conditions which, with the wisdom of hind-sight, are seen to be restrictive. Similarly, Theorem 4.2 was established in the proof of [14, Thm. 3.1] under the additional condition that $R > 0$. The positive definiteness of R was replaced by the invertibility of R in the proof of [11, Thm. 13.1.1]. It has been noted in [7] that the invertibility of R is also unnecessary. It is

the removal of this restriction that will allow its application to the matrix equation (1.1).

Theorem 4.3 (cf. [7])**.** *If $A - B(R + B^*X_+B)^{-1}(C + B^*X_+A)$ is d-stable in Theorem 4.2, then the sequence $\{X_i\}_{i=0}^{\infty}$ converges to X_+ quadratically.*

Theorem 4.4 (cf. [7])**.** *Under the conditions in Theorem 4.1 and assuming that all eigenvalues of $A-B(R+B^*X_+B)^{-1}(C+B^*X_+A)$ on the unit circle are semisimple, the Newton sequence $\{X_i\}$ converges to X_+ either quadratically, or linearly with rate $1/2$.*

5. Applications of Newton's method

We now let $m = n$ in DARE (4.1), and take $A = 0, R = 0, B = I$. The equation becomes $X + C^*X^{-1}C = Q$, which has the same form as (1.1), and the hypotheses of Theorem 4.1 are trivially satisfied whenever it has a positive definite solution. We can then apply the results we have just reviewed to the equation (1.1) (the matrix A in (1.1) has taken the place of the matrix C in (4.1)).

The next result is an immediate consequence of Theorem 4.1. The first conclusion has been proved in [3]. The second conclusion has been noted in [16].

Theorem 5.1. *If (1.1) has a positive definite solution, then it has a maximal positive definite solution X_+ and $\rho(X_+^{-1}A) \le 1$.*

By taking $L_0 = 0$ in (4.4), we obtain $A_0 = 0$ (which is certainly d-stable) and the following algorithm for equation (1.1):

Algorithm 5.2 (Newton's method for (1.1))**.** Take $X_0 = Q$. For $i = 1, 2, \ldots,$ compute $L_i = X_{i-1}^{-1}A$, and solve

$$X_i - L_i^*X_iL_i = Q - 2L_i^*A. \tag{5.1}$$

Note that the Stein equation (5.1) is uniquely solvable when $\rho(L_i) < 1$. From Theorems 4.2, 4.3 and 4.4 we have:

Theorem 5.3. *If (1.1) has a positive definite solution, then Algorithm 5.2 determines a sequence of Hermitian matrices $\{X_i\}_{i=0}^{\infty}$ for which $\rho(L_i) < 1$ for $i = 0, 1, \ldots, X_0 \ge X_1 \ge \cdots$, and $\lim_{i\to\infty} X_i = X_+$. The convergence is quadratic if $\rho(X_+^{-1}A) < 1$. If $\rho(X_+^{-1}A) = 1$ and all eigenvalues of $X_+^{-1}A$ on the unit circle are semisimple, then the convergence is either quadratic or linear with rate $1/2$.*

Note also that if there are eigenvalues of $X_+^{-1}A$ on the unit circle and linear convergence is identified then, as shown in [7], a double-step modification of Newton's method can be used to great advantage.

We now turn our attention to the equation (1.2). In [4] it is shown that, if A is nonsingular, the maximal solution X_+ of (1.2) is also the maximal Hermitian solution of the DARE

$$X = Q + FXF^* - FX(R + X)^{-1}XF^*,$$

where $F = A^*A^{-1}, R = AQ^{-1}A^*$. The maximal solution X_+ can then be found by direct application of Newton's method for the DARE (4.1). However, the overhead costs of Newton's method are higher than for equation (1.1), and comparison with the basic fixed point iteration is less favorable.

If we apply Newton's method directly to equation (1.2), convergence cannot be guaranteed if the initial guess is not close to X_+. However, Newton's method still has local quadratic convergence and can be used as an efficient correction method.

Algorithm 5.4 (Newton correction for (1.2))**.** For given X_k sufficiently close to X_+ and $i = k+1, k+2, \ldots$, compute $L_i = X_{i-1}^{-1}A$, and solve

$$X_i + L_i^* X_i L_i = Q + 2L_i^* A. \tag{5.2}$$

The equations (5.1) and (5.2) can be solved by a complex version of the algorithm described in [5]. The computational work per iteration for Algorithm 5.2 or 5.4 is roughly $10 \sim 15$ times that for Algorithm 2.1 or 2.4.

In contrast to the equation (5.1), the equation (5.2) is not necessarily nearly singular when $X_+^{-1}A$ has eigenvalues very close to the unit circle. This makes the Newton correction even more attractive.

6. Matrix pencils

As we have seen in the previous sections, the convergence rates of various algorithms for equation (1.1) or (1.2) are dependent on the eigenvalues of $X_+^{-1}A$, where X_+ is the solution of (1.1) or (1.2) that we seek. In this section, we will relate the eigenvalues of $X_+^{-1}A$ to the eigenvalues of a matrix pencil which is independent of X_+.

As we have seen, the equation (1.1) is a special case of the DARE (4.1). For (4.1) we consider the matrix pencil $\lambda F_e - G_e$ with

$$F_e = \begin{pmatrix} I & 0 & 0 \\ 0 & A^* & 0 \\ 0 & -B^* & 0 \end{pmatrix}, \qquad G_e = \begin{pmatrix} A & 0 & B \\ -Q & I & -C^* \\ C & 0 & R \end{pmatrix}.$$

Matrix pencils of this type were first introduced in [15].

Theorem 6.1 (cf. [7])**.** *If* (4.1) *has a Hermitian solution X, then $\lambda F_e - G_e$ is a regular pencil. Moreover, α is an eigenvalue of*

$$A - B(R + B^*XB)^{-1}(C + B^*XA) \tag{6.1}$$

if and only if α and $\bar{\alpha}^{-1}$ are eigenvalues of $\lambda F_e - G_e$. If we assume further that (A, B) is d-stabilizable and $R + B^*XB > 0$, then a unimodular α is an eigenvalue of (6.1) *with partial multiplicity k if and only if it is an eigenvalue of $\lambda F_e - G_e$ with partial multiplicity $2k$.*

Corollary 6.2. *For equation* (1.1)*, the eigenvalues of $X_+^{-1}A$ are precisely the eigenvalues of the matrix pencil*

$$\lambda F_1 - G_1 \equiv \lambda \begin{pmatrix} -I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & I & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & -I \\ Q & -I & A^* \\ -A & 0 & 0 \end{pmatrix}$$

inside or on the unit circle, with half of the partial multiplicities for each eigenvalue on the unit circle.

According to [3, Thm. 2.1], the equation (1.1) has a positive definite solution if and only if the rational matrix-valued function $\psi(\lambda) = Q + \lambda A + \lambda^{-1}A^*$ is regular (i.e., $\det\psi(\lambda) \neq 0$ for some λ) and $\psi(\lambda) \geq 0$ for all λ on the unit circle. In particular, (1.1) has a positive definite solution if $\psi(\lambda) > 0$ for all λ on the unit circle.

Let $r(T)$ be the numerical radius of $T \in \mathbf{C}^{m\times m}$, defined by $r(T) = \max\{|x^*Tx| : x \in \mathbf{C}^m, x^*x = 1\}$. Note that $r(T) \leq \|T\| \leq 2r(T)$ (see [9], for example). The following lemma has been proved in [3].

Lemma 6.3. *$\psi(\lambda) > 0$ for all λ on the unit circle if and only if*

$$r(Q^{-1/2}AQ^{-1/2}) < \frac{1}{2}.$$

As we have seen in Theorem 5.3, the convergence of Algorithm 5.2 is quadratic if $\rho(X_+^{-1}A) < 1$. Our final theorem clarifies this condition.

Theorem 6.4. *For equation* (1.1)*, $\rho(X_+^{-1}A) < 1$ if and only if*

$$r(Q^{-1/2}AQ^{-1/2}) < \frac{1}{2}.$$

Proof. By Corollary 6.2, it is enough to show $r(Q^{-1/2}AQ^{-1/2}) < \frac{1}{2}$ if and only if the pencil $\lambda F_1 - G_1$ has no eigenvalues on the unit circle. By appropriate block elimination we find that

$$\begin{aligned}
\det(\lambda F_1 - G_1) &= \det \begin{pmatrix} -\lambda I & 0 & I \\ -Q & I & -A^* \\ A & \lambda I & 0 \end{pmatrix} \\
&= \det \begin{pmatrix} 0 & 0 & I \\ -Q - \lambda A^* & I & -A^* \\ A & \lambda I & 0 \end{pmatrix} \\
&= \det \begin{pmatrix} -Q - \lambda A^* & I \\ A & \lambda I \end{pmatrix} \\
&= \det \begin{pmatrix} -Q - \lambda A^* & I \\ A + \lambda(Q + \lambda A^*) & 0 \end{pmatrix} \\
&= (-1)^m \lambda^m \det(Q + \lambda^{-1}A + \lambda A^*).
\end{aligned}$$

Therefore, $\lambda F_1 - G_1$ has no eigenvalues on the unit circle if and only if $\psi(\lambda) > 0$ for all λ on the unit circle, the latter is equivalent to $r(Q^{-1/2}AQ^{-1/2}) < \frac{1}{2}$ by Lemma 6.3. □

We now turn our attention to the equation (1.2). In this case, the eigenvalues of $X_+^{-1}A$ are related to the eigenvalues of the matrix pencil $\lambda F_2 - G_2$ with

$$F_2 = \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -I & 0 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 0 & 0 & -I \\ Q & -I & -A^* \\ -A & 0 & 0 \end{pmatrix}.$$

Lemma 6.5. *If X is a solution of* (1.2)*, then*

$$(\lambda F_2 - G_2) \begin{pmatrix} I & 0 & 0 \\ X & I & 0 \\ -X^{-1}A & 0 & I \end{pmatrix} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & A^*X^{-1} \\ -X & 0 & I \end{pmatrix} (\lambda M - N),$$

where

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & A^*X^{-1} & 0 \\ 0 & -I & 0 \end{pmatrix}, \quad N = \begin{pmatrix} X^{-1}A & 0 & -I \\ 0 & -I & 0 \\ 0 & 0 & -X \end{pmatrix}.$$

Proof. The result is easily verified by direct computation. □

Corollary 6.6. *The eigenvalues of $X_+^{-1}A$ are precisely the eigenvalues of $\lambda F_2 - G_2$ inside the unit circle.*

Proof. Taking $X = X_+$ in the above Lemma, we get

$$\begin{aligned}\det(\lambda F_2 - G_2) &= \det(\lambda M - N)\\ &= \det(X_+)\det(\lambda I - X_+^{-1}A)\det(\lambda A^* X_+^{-1} + I).\end{aligned}$$

Since $\rho(X_+^{-1}A) < 1$, the zeros of $\det(\lambda A^* X_+^{-1} + I) = 0$ are outside the unit circle. The conclusion in the corollary follows readily. □

7. Numerical results

In this section, we give some examples to illustrate the convergence behaviour of various algorithms we have discussed. Double precision is used in all computations.

Example 7.1. Consider equation (1.1) with

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}, \quad Q = \begin{pmatrix} 6.0 & 5.0 \\ 5.0 & 8.6 \end{pmatrix}.$$

The maximal solution (with the first 9 digits) is found to be

$$X_+ = \begin{pmatrix} 3.88319247 & 2.40094202 \\ 2.40094202 & 4.34595701 \end{pmatrix}.$$

We compare the number of iterations required for Algorithms 2.1, 3.1 and 3.4 to get the first 6 correct digits.

Algorithm 2.1 needs 16 iterations with

$$X_{16} = \begin{pmatrix} 3.88319512 & 2.40094422 \\ 2.40094422 & 4.34595998 \end{pmatrix}.$$

Algorithm 3.1 needs 34 iterations with

$$X_{34} = \begin{pmatrix} 3.88319648 & 2.40094414 \\ 2.40094414 & 4.34595965 \end{pmatrix}.$$

Algorithm 3.4 needs 19 iterations with

$$X_{19} = \begin{pmatrix} 3.88319736 & 2.40094456 \\ 2.40094456 & 4.34595963 \end{pmatrix}.$$

We have used $Y_0 = I/\|Q\|_\infty$ for Algorithms 3.1 and 3.4. If we use $Y_0 = Q^{-1}$, the numbers of iterations are 32 and 17, respectively.

The convergence is linear for all three algorithms. The convergence of Algorithm 2.1 is slightly faster than that of Algorithm 3.4, while the convergence of Algorithm 3.4 is faster than that of Algorithm 3.1 by roughly a factor of 2. These are consistent with the convergence results in Section 2 and Section 3. For this example, we have $\rho(X_+^{-1}A) = 0.6708$, $\|X_+^{-1}A\| = 1.3829$, and $\|AX_+^{-1}\| = 0.8321$.

The next two examples will show that, for equation (1.1), Algorithm 5.2 can be much more efficient than Algorithm 2.1. Of course, for easy problems the basic fixed point iteration needs no more than 30 iterations to get a good approximate solution. In these cases we cannot expect Newton's method to perform better, since two or three iterations are usually necessary for the Newton iteration. For these two examples, we use the practical stopping criterion

$$\|X_n + A^* X_n^{-1} A - Q\|_\infty < \epsilon \tag{7.1}$$

for both Algorithm 5.2 and Algorithm 2.1, where ϵ is a prescribed tolerance.

Example 7.2. We consider the equation (1.1) with $Q = I$ and

$$A = \begin{pmatrix} 0.20 & 0.20 & 0.10 \\ 0.20 & 0.15 & 0.15 \\ 0.10 & 0.15 & 0.25 \end{pmatrix}.$$

For this example, A is Hermitian (and hence normal). The exact maximal solution can be found according to the formula

$$X_+ = \frac{1}{2}[I + (I - 4A^*A)^{1/2}],$$

which is valid for any normal matrix A with $\|A\| \leq 1/2$ (see [17]).

Since $r(A) = \|A\| = 1/2$ for this example, we have $\rho(X_+^{-1}A) = 1$ (cf. Thm. 6.4). The convergence of Algorithm 2.1 turns out to be sublinear. It needs 7071 iterations to satisfy (7.1) for $\epsilon = 10^{-8}$, with

$$X_{7071}^F = \begin{pmatrix} 0.82656902 & -0.16835309 & -0.15814522 \\ & 0.83167296 & -0.16324916 \\ symm. & & 0.82146509 \end{pmatrix}.$$

On the other hand, the convergence of Algorithm 5.2 is linear with rate 1/2 (cf. Thm. 5.3). The stopping criterion is satisfied after 12 iterations, with

$$X_{12}^N = \begin{pmatrix} 0.82656580 & -0.16835631 & -0.15814844 \\ & 0.83166974 & -0.16325238 \\ symm. & & 0.82146187 \end{pmatrix}.$$

We find that both X_{7071}^F and X_{12}^N have four correct digits, with X_{12}^N slightly better. If we use a double Newton step following X_{12}^N, the resulting approximate solution with the first 8 digits (without rounding) is

$$X_{13} = \begin{pmatrix} 0.82654545 & -0.16837666 & -0.15816879 \\ & 0.83164938 & -0.16327272 \\ symm. & & 0.82144151 \end{pmatrix}.$$

All digits are the same as in the exact solution. This example shows that Newton's method can be much more efficient than the basic fixed point iteration when $r(Q^{-1/2}AQ^{-1/2})$ is equal or very close to 1/2.

Example 7.3. We consider the equation (1.1) with

$$A = \begin{pmatrix} 0.37 & 0.13 & 0.12 \\ -0.30 & 0.34 & 0.12 \\ 0.11 & -0.17 & 0.29 \end{pmatrix}, \quad Q = \begin{pmatrix} 1.20 & -0.30 & 0.10 \\ -0.30 & 2.10 & 0.20 \\ 0.10 & 0.20 & 0.65 \end{pmatrix}.$$

For this example, $r(Q^{-1/2}AQ^{-1/2}) < 1/2$. Thus the Newton iteration converges quadratically to the maximal solution. It needs 8 iterations to satisfy the stopping criterion (7.1) for $\epsilon = 10^{-12}$. The computed maximal solution is

$$X_+ = \begin{pmatrix} 0.94632675 & -0.19866482 & -0.05960039 \\ & 1.86737567 & 0.32524233 \\ symm. & & 0.41582003 \end{pmatrix}.$$

The basic fixed point iteration needs 332 iterations to satisfy the same criterion. The convergence is linear since $\rho(X_+^{-1}A) < 1$. Note that $\|X_+^{-1}A\| > 1$ for this example.

The last example is devoted to equation (1.2).

Example 7.4. Consider equation (1.2) with

$$A = \begin{pmatrix} 50 & 20 \\ 10 & 60 \end{pmatrix}, \quad Q = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}.$$

The maximal solution can be found to be

$$X_+ = \begin{pmatrix} 51.7993723118 & 16.0998802679 \\ 16.0998802679 & 62.2516164469 \end{pmatrix}.$$

Using Algorithm 2.4, we get after 100 iterations

$$X_{100} = \begin{pmatrix} 51.4950332009 & 16.0137829200 \\ 16.0137829200 & 61.8891412657 \end{pmatrix}.$$

After 300 more iterations, we get

$$X_{400} = \begin{pmatrix} 51.7993723016 & 16.0998802648 \\ 16.0998802648 & 62.2516164347 \end{pmatrix}.$$

If we use Newton correction (Algorithm 5.4) starting with X_{100}, we get after two iterations

$$X_{102} = \begin{pmatrix} 51.7993723045 & 16.0998802666 \\ 16.0998802666 & 62.2516164389 \end{pmatrix},$$

which is already slightly better than X_{400} given above. The slow convergence of Algorithm 2.4 is consistent with the second conclusion in Theorem 2.6, since we have $\rho(X_+^{-1}A) = 0.9719$ for this example. For this example, we have $\|X_+^{-1}A\| > 1$. Therefore, Theorem 2.5 and the first conclusion in Theorem 2.6 are useless.

References

1. W. N. Anderson, Jr., T. D. Morley, and G. E. Trapp, *Positive solutions to $X = A - BX^{-1}B^*$*, Linear Algebra Appl. **134** (1990), 53–62.
2. J. C. Engwerda, *On the existence of a positive definite solution of the matrix equation $X + A^TX^{-1}A = I$*, Linear Algebra Appl. **194** (1993), 91–108.
3. J. C. Engwerda, A. C. M. Ran, and A. L. Rijkeboer, *Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^*X^{-1}A = Q$*, Linear Algebra Appl. **186** (1993), 255–275.
4. A. Ferrante and B. C. Levy, *Hermitian solutions of the equation $X = Q + NX^{-1}N^*$*, Linear Algebra Appl. **247** (1996), 359–373.
5. J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler, *Solution of the Sylvester matrix equation $AXB^T + CXD^T = E$*, ACM Trans. Math. Software **18** (1992), 223–231.
6. G. H. Golub and C. F. Van Loan, *Matrix computations*, Third edition, Johns Hopkins University Press, Baltimore, MD, 1996.
7. C.-H. Guo, *Newton's method for discrete algebraic Riccati equations when the closed-loop matrix has eigenvalues on the unit circle*, SIAM J. Matrix Anal. Appl., to appear.
8. G. A. Hewer, *An iterative technique for the computation of the steady-state gains for the discrete optimal regulator*, IEEE Trans. Autom. Control **16** (1971), 382–384.
9. R. A. Horn and C. R. Johnson, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1991.
10. M. A. Krasnoselskii, G. M. Vainikko, P. P. Zabreiko, Ya. B. Rutitskii, and V. Ya. Stetsenko, *Approximate solution of operator equations*, Wolters-Noordhoff Publishing, Groningen, 1972.
11. P. Lancaster and L. Rodman, *Algebraic Riccati equations*, Clarendon Press, Oxford, 1995.
12. P. Lancaster and M. Tismenetsky, *The theory of matrices*, Second edition, Academic Press, Orlando, FL, 1985.
13. J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.
14. A. C. M. Ran and R. Vreugdenhil, *Existence and comparison theorems for algebraic Riccati equations for continuous- and discrete-time systems*, Linear Algebra Appl. **99** (1988), 63–83.

15. P. Van Dooren, *A generalized eigenvalue approach for solving Riccati equations*, SIAM J. Sci. Comput. **2** (1981), 121–135.
16. X. Zhan, *Computing the extremal positive definite solutions of a matrix equation*, SIAM J. Sci. Comput. **17** (1996), 1167–1174.
17. X. Zhan and J. Xie, *On the matrix equation* $X + A^T X^{-1} A = I$, Linear Algebra Appl. **247** (1996), 337–345.

Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

Current address: Department of Computer Science, University of California, Davis, California 95616-8562

E-mail address: guo@cs.ucdavis.edu

Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

E-mail address: lancaste@ucalgary.ca