Journal of Computational Mathematics
Vol.29, No.3, 2011, 341–366.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1010-m3258

ON EIGENVALUE BOUNDS AND ITERATION METHODS FOR DISCRETE ALGEBRAIC RICCATI EQUATIONS*

Hua Dai
Department of Mathematics, Nanjing University of Aeronautics and Astronautics,
Nanjing 210016, China
Email: hdai@nuaa.edu.cn
Zhong-Zhi Bai
State Key Laboratory of Scientific/Engineering Computing, Institute of Computational Mathematics,
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
Email: bzz@lsec.cc.ac.cn

Abstract

We derive new and tight bounds about the eigenvalues and certain sums of the eigenvalues for the unique symmetric positive definite solutions of the discrete algebraic Riccati equations. These bounds considerably improve the existing ones and treat the cases that have been not discussed in the literature. Besides, they also result in completions for the available bounds about the extremal eigenvalues and the traces of the solutions of the discrete algebraic Riccati equations. We study the fixed-point iteration methods for computing the symmetric positive definite solutions of the discrete algebraic Riccati equations and establish their general convergence theory. By making use of the Schulz iteration to partially avoid computing the matrix inversions, we present effective variants of the fixed-point iterations, prove their monotone convergence and estimate their asymptotic convergence rates. Numerical results show that the modified fixed-point iteration methods are feasible and effective solvers for computing the symmetric positive definite solutions of the discrete algebraic Riccati equations.

Mathematics subject classification: 15A15, 15A18, 15A24, 15A48, 65F30.
Key words: Discrete algebraic Riccati equation, Symmetric positive definite solution, Eigenvalue bound, Fixed-point iteration, Convergence theory.

1. Introduction

Consider the *discrete algebraic Riccati equation* (**DARE**)

$$X = A^TXA - A^TXB(G + B^TXB)^{-1}B^TXA + C^TC, \tag{1.1}$$

where $A \in \mathbf{R}^{n\times n}$, $B \in \mathbf{R}^{n\times m}$, $C \in \mathbf{R}^{p\times n}$ and $G \in \mathbf{R}^{m\times m}$ are given matrices, and the matrix G is assumed to be symmetric and positive definite. Let

$$R = BG^{-1}B^T \quad \text{and} \quad Q = C^TC. \tag{1.2}$$

Then by applying the Sherman-Morrison-Woodbury formula [10, P. 50], the DARE (1.1) can be equivalently reformulated as

$$X = A^TX(I + RX)^{-1}A + Q,$$

* Received November 11, 2009 / Revised version received June 1, 2010 / Accepted September 14, 2010 / Published online February 28, 2011 /

where I represents the identity matrix of appropriate size, and R and Q are the matrices defined in (1.2) satisfying $R \succeq 0$ and $Q \succeq 0$. Here and in the sequel, for a square matrix W we say $W \succ 0$ (or $W \succeq 0$) if W is symmetric positive definite (or symmetric positive semidefinite).

Throughout the paper we assume that (A, B) is a stabilizable pair and (A, C) is a detectable pair [1)] . Then the DARE (1.1) has a unique symmetric positive definite solution X such that the matrix $(I + RX)^{-1}A$ is stable, i.e., every eigenvalue λ of the matrix $(I + RX)^{-1}A$ satisfies $|\lambda| < 1$; see, e.g., [24, 44]. Under this assumption, the DARE (1.1) can be further rewritten in the symmetric form as

$$X = A^T (X^{-1} + R)^{-1} A + Q. \tag{1.3}$$

In this paper, we will focus on discussions about the discrete algebraic Riccati equations of the form (1.3).

The discrete algebraic Riccati equation (1.1) arises in many areas of engineering applications such as the optimal control design [21] and the filter design [2]. One typical and important application about the DARE (1.1) is the discrete-time LQ-problem in optimal control. Under the assumption that the matrix pairs (A, B) is stabilizable and (A, C) is detectable, the discrete-time linear system

$$\begin{cases} x_{k+1} = Ax_k + Bu_k, & x_0 \text{ given}, \\ y_k = Cx_k, & k = 0, 1, 2, \cdots, \end{cases}$$

exists an optimal control u_k, which is the minimizer of the quadratic cost functional

$$J = \sum_{k=0}^{\infty} \left(x_k^T Q x_k + u_k^T G u_k \right).$$

Then u_k can be recovered via x_k by

$$u_k = -(G + B^T X B)^{-1} B^T X A x_k, \qquad k = 0, 1, 2, \cdots,$$

where X is the unique symmetric positive definite solution of the DARE (1.1). We remark that when the above-mentioned linear system is subjected to perturbations, uncertainties, additive/multiplicative noises or a time delay, the DARE (1.3) may be appropriately modified and is often impossible to be solved exactly.

An accurate estimate about the solution of the DARE (1.3) or, equivalently, the DARE (1.1), is theoretically important and practically useful when we treat some control problems such as the stabilized control design for time-delay systems [29], the stability analysis in the presentations of time delay and perturbations [43], and the state and error covariance estimation [20], as well as when we select feasible starting points for certain iteration methods employed to solve the discrete algebraic Riccati equations.

In fact, a bound for the solution X of the DARE (1.3) can be provided through a bound on the eigenvalues $\lambda_i(X)$ of X. Various bounds about the extreme eigenvalues [9], the partial sum and the partial product of eigenvalues [17, 19], the trace [13, 22, 38], and the determinant [42] of the solution X have been derived during the past three decades; see [23, 37] for excellent

[1)] For a complex constant λ and vector w, if $w^*B = 0$ and $w^*A = \lambda w^*$ imply either $|\lambda| < 1$ or $w = 0$, then the matrix pair (A, B) is called stabilizable. The matrix pair (A, C) is called detectable if (A^T, C^T) is stabilizable. Here, $(\cdot)^T$ and $(\cdot)^*$ denote the transpose and the conjugate transpose of either a complex vector or a complex matrix, respectively.

overviews. In addition, lower and upper bounds about the solution X of the DARE (1.3) have been presented in [18, 26–28, 30]. Recently, upper bounds about the solution X for the DARE (1.1) have been also derived under the condition that a matrix K exists such that $A+BK$ is stabilizable; see [8]. We should mention that most of these known results hold true only when at least one of the conditions $R \succ 0$ and $Q \succ 0$ is satisfied. These conditions are, however, very restrictive and often violated in many actual control problems, as where the number n of the state variables is usually greater than the number m of the inputs and the matrix Q is generally symmetric positive semidefinite.

Of course, the DARE (1.3) is, in general, a nonlinear matrix equation and can be also considered as a system of nonlinear equations. Hence, numerical methods such as the Newton method, the Schur method and its variants, the fixed-point iteration method, and the doubling algorithm, etc., can be adopted to effectively compute its solution. Among them the Newton method is the oldest and the best studied one; see [11, 24]. At each step, however, it requires to solve a discrete Lyapunov equation, which is quite costly in actual computations, especially when the sizes of the matrices involved are very large, though several efficient solvers for lower or mildly high order discrete Lyapunov equations are available; see [4, 15]. The Schur method initially given in [25] was further described in detail in [3, 36, 40]; this method consists of computing a stable deflating subspace of a $2n \times 2n$ symplectic matrix pencil and, therefore, possesses good numerical stability. It has been used in the MATLAB control toolbox, though it demands large storage and memory in actual implementations. We remark that some effective variants of the Schur method have been developed based on the structure-preserving factorization methods; see [6, 31, 33, 34]. The doubling algorithm was derived in [1, 14] as an acceleration scheme for the fixed-point iteration

$$X_{k+1} = A^T X_k (I + RX_k)^{-1} A + Q, \qquad k = 0, 1, 2, \cdots . \tag{1.4}$$

Note that the iteration sequence $\{X_k\}_{k=0}^{\infty}$ generated by the scheme (1.4) is numerically nonsymmetric and has only linear convergence rate. Therefore, we can turn to produce the quadratically convergent matrix sequence $\{X_{2^k}\}_{k=0}^{\infty}$ instead of $\{X_k\}_{k=0}^{\infty}$; this is the basic idea of the doubling algorithm. However, at each step the doubling algorithm requires to compute two matrix inversions and eight matrix-matrix products, which is quite costly in actual computations. We mention that recently a doubling algorithm directly applied to the DARE (1.1) has been derived and studied; see [12, 32]. To produce a symmetric iteration sequence approximating the exact solution X of the DARE (1.3), Komaroff [18] presented the fixed-point iteration

$$X_{k+1} = A^T (X_k^{-1} + R)^{-1} A + Q, \qquad k = 0, 1, 2, \cdots , \tag{1.5}$$

and proved its linear convergence property when A is nonsingular, $R \succ 0$ and $Q \succ 0$.

The purpose of this paper is three folds. The first is to derive tight bounds about partial sum and partial product about the eigenvalues $\lambda_i(X)$ of the solution X for the DARE (1.3) without imposing the restriction $R \succ 0$ and $Q \succ 0$. These present improvements and completions for the existing bounds. The second is to demonstrate the convergence of the fixed-point iteration (1.5) without assuming that the matrix A is nonsingular. And the third is to establish an economical and effective variant for the fixed-point iteration (1.5) by reducing the matrix inversions at each step through utilizing the Schulz iteration, prove its linear convergence property and estimate its asymptotic convergence rate.

The paper is organized as follows. In Section 2, we derive lower and upper bounds for partial sum, and upper bounds for partial product of the eigenvalues for the solution of the

DARE (1.3) without assuming $R \succ 0$ and $Q \succ 0$. In Section 3, we discuss the convergence property of the fixed-point iteration (1.5) without assuming that A is nonsingular or $R \succ 0$. An effective variant of this fixed-point iteration is established in Section 4, where its convergence property is discussed and the corresponding asymptotic convergence rate is estimated. Some numerical examples are used to examine the sharpness of the new bounds with respect to the eigenvalues about the solution of the DARE (1.3) and to show the effectiveness of the new fixed-point iteration in Section 5. Finally, in Section 6, we end this paper with several concluding remarks.

2. Lower and Upper Bounds for Eigenvalues

Denote by $\mathbf{S}^n$ the set of $n \times n$ symmetric matrices and $\mathbf{S}^n_+$ the convex cone of symmetric positive semidefinite matrices. Then $\mathbf{S}^n_+$ naturally induces the partial ordering "$\succeq$" on $\mathbf{S}^n$ as follows: for $S, T \in \mathbf{S}^n$, $S \succeq T$ if $S - T \in \mathbf{S}^n_+$. In addition, for $S, T \in \mathbf{S}^n$, we define $S \succ T$ if $S - T$ is symmetric positive definite. For a complex n-by-n matrix $A \in \mathbf{C}^{n\times n}$, we use $\det(A)$, $\mathrm{tr}(A)$, $\mathrm{rank}(A)$, $\rho(A)$ and $\|A\|_2$ to represent its determinant, trace, rank, spectral radius and the Euclidean norm, respectively. Its eigenvalues are denoted by $\lambda_i(A)$, $i = 1, \cdots, n$, and ordered such that their real parts are nonincreasing, i.e.,

$$\mathrm{Re}(\lambda_1(A)) \geq \mathrm{Re}(\lambda_2(A)) \geq \cdots \geq \mathrm{Re}(\lambda_n(A)).$$

The ordering for the singular values $\sigma_i(A)$, $i = 1, \cdots, n$, of the matrix A is defined in an analogous fashion. Besides, we also use $\lambda_i^{-1}(A)$ and $\sigma_i^{-1}(A)$, $i = 1, \cdots, n$, to represent briefly the quantities $[\lambda_i(A)]^{-1}$ and $[\sigma_i(A)]^{-1}$, respectively, when they are nonzero; and $\lambda_i^2(A)$ and $\sigma_i^2(A)$, $i = 1, \cdots, n$, to represent briefly the quantities $[\lambda_i(A)]^2$ and $[\sigma_i(A)]^2$, respectively.

With respect to the DARE (1.3), we always assume that the symmetric positive semidefinite matrix R satisfies $\mathrm{rank}(R) = r \leq \min\{m, n\}$ and arrange its eigenvalues as

$$\lambda_1(R) \geq \lambda_2(R) \geq \cdots \geq \lambda_r(R) > 0 = \lambda_{r+1}(R) = \cdots = \lambda_n(R).$$

Then it holds that $\mathrm{rank}(R) = \mathrm{rank}(B) = r$.

The following results about the eigenvalues of symmetric matrices are essential and useful for deriving bounds about the eigenvalues for the solution of the DARE (1.3).

Lemma 2.1. ([7,35]) *Let $S, T \in \mathbf{S}^n$ and $P \in \mathbf{R}^{n\times m}$. Then*

(i) $\lambda_k(S) \geq \lambda_k(T)$ $(k = 1, \cdots, n)$ *if* $S \succeq T$;

(ii) $\lambda_n(S)I \preceq S \preceq \lambda_1(S)I$;

(iii) $P^TSP \succeq P^TTP$ *if* $S \succeq T$;

(iv) $T^{-1} \succeq S^{-1}$ *if* $S \succeq T \succ 0$;

(v) $\sum_{i=1}^{k} \lambda_i(S+T) \leq \sum_{i=1}^{k} \lambda_i(S) + \sum_{i=1}^{k} \lambda_i(T), \quad k = 1, \cdots, n;$

(vi) $\sum_{i=1}^{k} \lambda_i(S+T) \geq \sum_{i=1}^{k} \lambda_i(S) + \sum_{i=1}^{k} \lambda_{n-i+1}(T), \quad k = 1, \cdots, n;$

(vii) $\|S\|_2 \geq \|T\|_2$ *if* $S \succeq T \succ 0$.

Moreover, the inequalities in (v) *and* (vi) *become equalities when* $k = n$.

Lemma 2.2. ([46]) *If $C, P \in \mathbf{S}^n$ and $P \succ 0$, then $CPC + P^{-1} \succeq 2C$.*

Lemma 2.3. ([35]) *If $S, T \in \mathbf{S}_+^n$, then*

$$\sum_{i=1}^{k} \lambda_i(S)\lambda_{n-i+1}(T) \leq \sum_{i=1}^{k} \lambda_i(ST) \leq \sum_{i=1}^{k} \lambda_i(S)\lambda_i(T), \qquad k = 1, \cdots, n.$$

Lemma 2.4. ([35]) *Let u_i, $i = 1, \cdots, n$, be nonnegative real numbers, and x_i and y_i, $i = 1, \cdots, n$, be real numbers arranged in nonincreasing ordering. If*

$$\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i, \qquad k = 1, \cdots, n,$$

then

$$\sum_{i=1}^{k} u_i x_i \leq \sum_{i=1}^{k} u_i y_i, \qquad k = 1, \cdots, n.$$

For any symmetric positive definite matrix X and symmetric positive semidefinite matrix R, Komaroff [17] gave upper bounds about partial sums of the eigenvalues of the matrix $(X^{-1} + R)^{-1}$ in 1992. These inequalities are precisely stated in the following lemma.

Lemma 2.5. ([17]) *For any $X \in \mathbf{S}^n$ satisfying $X \succ 0$ and any $R \in \mathbf{S}_+^n$, the following inequalities hold true:*

$$\sum_{i=1}^{k} \lambda_i\big((X^{-1} + R)^{-1}\big) \leq \sum_{i=1}^{k} \lambda_i(X)\big(1 + \lambda_i(X)\lambda_{n-i+1}(R)\big)^{-1}, \qquad k = 1, \cdots, n.$$

Based on Lemma 2.1, we can derive lower bounds about partial sums of the eigenvalues of the matrix $A^T(X^{-1} + R)^{-1}A$, which are indispensable for estimating new bounds for the solution of the DARE (1.3).

Lemma 2.6. *Let $A \in \mathbf{R}^{n \times n}$ be a given matrix. Then the inequalities*

$$\sum_{i=1}^{k} \lambda_i\big(A^T(X^{-1} + R)^{-1}A\big) \geq \sum_{i=1}^{k} \sigma_{n-i+1}^2(A)\lambda_i(X)\big(1 + \lambda_i(X)\lambda_1(R)\big)^{-1},$$
$$\sum_{i=1}^{k} \lambda_i\big(A^T(X^{-1} + R)^{-1}A\big) \geq \sum_{i=1}^{k} \sigma_i^2(A)\lambda_{n-i+1}(X)\big(1 + \lambda_{n-i+1}(X)\lambda_1(R)\big)^{-1}$$

hold for any $X \in \mathbf{S}^n$ satisfying $X \succ 0$ and any $R \in \mathbf{S}_+^n$.

Proof. Noticing that $\lambda_i(UV) = \lambda_i(VU)$ holds for any $U, V \in \mathbf{R}^{n \times n}$, from Lemma 2.3 we can obtain

$$\begin{aligned} \sum_{i=1}^{k} \lambda_i\big(A^T(X^{-1} + R)^{-1}A\big) &= \sum_{i=1}^{k} \lambda_i\big((X^{-1} + R)^{-1}AA^T\big) \\ &\geq \sum_{i=1}^{k} \lambda_i\big((X^{-1} + R)^{-1}\big)\sigma_{n-i+1}^2(A). \end{aligned} \tag{2.1}$$

As $\lambda_i(X^{-1}) = \lambda_{n-i+1}^{-1}(X)$ implies $\lambda_i((X^{-1}+R)^{-1}) = \lambda_{n-i+1}^{-1}(X^{-1}+R)$, it follows from $R \preceq \lambda_1(R)I$ and Lemma 2.1(i) that

$$\begin{aligned}\lambda_{n-i+1}^{-1}(X^{-1}+R) \geq & \lambda_{n-i+1}^{-1}(X^{-1}+\lambda_1(R)I) \\ = & \frac{1}{\lambda_{n-i+1}(X^{-1})+\lambda_1(R)} = \frac{\lambda_i(X)}{1+\lambda_i(X)\lambda_1(R)}.\end{aligned} \tag{2.2}$$

Substituting (2.2) into (2.1), we immediately obtain the first inequality what we were proving. The second inequality can be demonstrated in an analogous fashion. □

Based on the above preparation, we give lower bounds about partial sum of the eigenvalues for the solution of the DARE (1.3).

Theorem 2.1. *Let $X \in \mathbf{R}^{n\times n}$ be the symmetric positive definite solution of the DARE* (1.3). *Then*

$$\sum_{i=1}^k \lambda_i(X) \geq \sum_{i=1}^k \left(\frac{\sigma_{n-i+1}^2(A)\lambda_i(Q)}{\lambda_i(Q)\lambda_1(R)+1} + \lambda_{n-i+1}(Q) \right), \qquad k = 1, \cdots, n, \tag{2.3}$$

$$\sum_{i=1}^k \lambda_i(X) \geq \sum_{i=1}^k \left(\frac{\sigma_i^2(A)\lambda_{n-i+1}(Q)}{\lambda_{n-i+1}(Q)\lambda_1(R)+1} + \lambda_{n-i+1}(Q) \right), \qquad k = 1, \cdots, n. \tag{2.4}$$

In particular, if $\det(A) \neq 0$ *and* $Q \succ 0$, *then it holds that*

$$\sum_{i=1}^k \lambda_i(X) \geq \frac{\theta + \sqrt{\theta^2 + 4\lambda_1(R)\sum_{i=1}^k \lambda_{n-i+1}(Q)}}{2\lambda_1(R)}, \qquad k = 1, \cdots, n, \tag{2.5}$$

where $\theta = \sigma_n^2(A) + \lambda_1(R)\lambda_n(Q) - 1$.

Proof. Because X is the symmetric positive definite solution of the DARE (1.3), from Lemma 2.1(vi) we have

$$\begin{aligned}\sum_{i=1}^k \lambda_i(X) = & \sum_{i=1}^k \lambda_i\big(A^T(X^{-1}+R)^{-1}A + Q\big) \\ \geq & \sum_{i=1}^k \lambda_i\big(A^T(X^{-1}+R)^{-1}A\big) + \sum_{i=1}^k \lambda_{n-i+1}(Q).\end{aligned}$$

By making use of the first inequality of Lemma 2.6 we obtain

$$\sum_{i=1}^k \lambda_i(X) \geq \sum_{i=1}^k \sigma_{n-i+1}^2(A)\lambda_i(X)\big(1+\lambda_i(X)\lambda_1(R)\big)^{-1} + \sum_{i=1}^k \lambda_{n-i+1}(Q). \tag{2.6}$$

As $X \succeq Q$, from Lemma 2.1(i) we know that

$$\lambda_i(X) \geq \lambda_i(Q), \qquad i = 1, \cdots, n.$$

It then follows from the monotonically increasing property of the function $f(t) = t/(1+t)$ that the inequality (2.3) holds true.

The validity of the inequality (2.4) can be analogously demonstrated by utilizing the second inequality of Lemma 2.6.

In addition, by making use of Lemma 2.4, from (2.6) we obtain

$$\sum_{i=1}^{k}\lambda_i(X)\big(1+\lambda_1(R)\lambda_i(X)\big)\geq\sum_{i=1}^{k}\Big(\sigma_{n-i+1}^2(A)\lambda_i(X)+\lambda_{n-i+1}(Q)\big(1+\lambda_1(R)\lambda_i(X)\big)\Big).$$

Applying

$$\left(\sum_{i=1}^{k}\lambda_i(X)\right)^2\geq\sum_{i=1}^{k}\lambda_i^2(X)$$

to the above inequality, we further get

$$\lambda_1(R)\left(\sum_{i=1}^{k}\lambda_i(X)\right)^2-\Big(\sigma_n^2(A)+\lambda_1(R)\lambda_n(Q)-1\Big)\sum_{i=1}^{k}\lambda_i(X)-\sum_{i=1}^{k}\lambda_{n-i+1}(Q)\geq 0.$$

By straightforwardly solving this quadratic inequality with respect to $\sum_{i=1}^{k}\lambda_i(X)$ we then obtain the inequality (2.5). □

Theorem 2.1 directly leads to the following lower bounds about the maximal eigenvalue and the trace for the solution of the DARE (1.3).

Corollary 2.1. *Let $X\in\mathbf{R}^{n\times n}$ be the symmetric positive definite solution of the DARE* (1.3)*. Then $\lambda_1(X)\geq\omega$, where*

$$\omega=\max\left\{\lambda_1(Q),\quad\frac{\sigma_n^2(A)\lambda_1(Q)}{\lambda_1(Q)\lambda_1(R)+1}+\lambda_n(Q),\quad\frac{\sigma_1^2(A)\lambda_n(Q)}{\lambda_n(Q)\lambda_1(R)+1}+\lambda_n(Q)\right\}.$$

Moreover, it holds that

$$\operatorname{tr}(X)\geq\operatorname{tr}(Q)+\sum_{i=n-r_a+1}^{n}\frac{\sigma_{n-i+1}^2(A)\lambda_i(Q)}{\lambda_i(Q)\lambda_1(R)+1},\qquad\textit{with }\operatorname{rank}(A)=r_a,\tag{2.7a}$$

$$\operatorname{tr}(X)\geq\operatorname{tr}(Q)+\sum_{i=n-r_q+1}^{n}\frac{\lambda_{n-i+1}(Q)\sigma_i^2(A)}{\lambda_{n-i+1}(Q)\lambda_1(R)+1},\qquad\textit{with }\operatorname{rank}(Q)=r_q.\tag{2.7b}$$

In particular, if $\det(A)\neq 0$ *and* $Q\succ 0$*, then*

$$\operatorname{tr}(X)\geq\frac{\theta+\sqrt{\theta^2+4\lambda_1(R)\operatorname{tr}(Q)}}{2\lambda_1(R)},$$

where $\theta=\sigma_n^2(A)+\lambda_1(R)\lambda_n(Q)-1$.

We remark that the lower bound given in (2.7b) is sharper than that in [13].

Now, we give upper bounds about partial sum of the eigenvalues for the solution of the DARE (1.3).

Theorem 2.2. *Let $X\in\mathbf{R}^{n\times n}$ be the symmetric positive definite solution of the DARE* (1.3)*. If $\sigma_1(A)<1$, then*

$$\sum_{i=1}^{k}\lambda_i(X)\leq\frac{1}{1-\sigma_1^2(A)}\sum_{i=1}^{k}\lambda_i(Q),\qquad k=1,\cdots,n-r,\tag{2.8a}$$

$$\sum_{i=1}^{k}\lambda_i(X)\leq\frac{k[\Theta+\sqrt{\Theta^2+4\lambda_r(R)\eta/k}]}{2\lambda_r(R)},\qquad k=n-r+1,n-r+2,\cdots,n,\tag{2.8b}$$

where $\eta = \min\{\beta, \delta\}$ *and*

$$\begin{aligned}\Theta =&\sigma_1^2(A) + \lambda_1(Q)\lambda_r(R) - 1,\\ \beta =&\sum_{i=1}^{k} \lambda_i(Q) + (n-r)\lambda_r(R)\left(\frac{\sigma_1(A)\lambda_1(Q)}{1-\sigma_1^2(A)}\right)^2,\\ \delta =&\sum_{i=1}^{k} \lambda_i(Q) + \lambda_r(R)\left(\frac{\sigma_1(A)\sum_{i=1}^{n-r}\lambda_i(Q)}{1-\sigma_1^2(A)}\right)^2.\end{aligned}$$

Proof. From Lemmas 2.3 and 2.5 we have

$$\begin{aligned}\sum_{i=1}^{k} \lambda_i\big(A^T(X^{-1}+R)^{-1}A\big) =&\sum_{i=1}^{k} \lambda_i\big((X^{-1}+R)^{-1}AA^T\big)\\ \le&\sum_{i=1}^{k} \lambda_i\big((X^{-1}+R)^{-1}\big)\lambda_i(AA^T)\\ \le&\sum_{i=1}^{k} \sigma_i^2(A)\lambda_i(X)\big(1+\lambda_i(X)\lambda_{n-i+1}(R)\big)^{-1}.\end{aligned}$$

Because $X \in \mathbf{R}^{n\times n}$ is the symmetric positive definite solution of the DARE (1.3), by making use of Lemma 2.1(v) we obtain

$$\begin{aligned}\sum_{i=1}^{k} \lambda_i(X) =&\sum_{i=1}^{k} \lambda_i(A^T(X^{-1}+R)^{-1}A+Q)\\ \le&\sum_{i=1}^{k} \big(\lambda_i(A^T(X^{-1}+R)^{-1}A)+\lambda_i(Q)\big)\\ \le&\sum_{i=1}^{k} \Big(\sigma_i^2(A)\lambda_i(X)\big(1+\lambda_i(X)\lambda_{n-i+1}(R)\big)^{-1}+\lambda_i(Q)\Big).\end{aligned} \tag{2.9}$$

For $k = 1, \cdots, n-r$, as $\lambda_{n-i+1}(R) = 0$ when $1 \le i \le k$, the inequality (2.9) naturally reduces to

$$\begin{aligned}\sum_{i=1}^{k} \lambda_i(X) \le&\sum_{i=1}^{k} \big(\sigma_i^2(A)\lambda_i(X)+\lambda_i(Q)\big)\\ \le&\sigma_1^2(A)\sum_{i=1}^{k}\lambda_i(X)+\sum_{i=1}^{k}\lambda_i(Q).\end{aligned}$$

It then follows that (2.8a) is valid.

For $k = n-r+1, n-r+2, \cdots, n$, the inequality (2.9) can be rewritten as

$$\sum_{i=1}^{k} \lambda_i(X) \le\sum_{i=1}^{n-r} \Big(\sigma_i^2(A)\lambda_i(X)+\lambda_i(Q)\Big)+\sum_{i=n-r+1}^{k}\Big(\sigma_i^2(A)\lambda_i(X)\big(1+\lambda_i(X)\lambda_r(R)\big)^{-1}+\lambda_i(Q)\Big).$$

By making use of Lemma 2.4 we get

$$\begin{aligned}
&\sum_{i=1}^{k}\lambda_i(X)\Big(1+\lambda_i(X)\lambda_r(R)\Big)\\
\le&\sum_{i=1}^{k}\lambda_i(Q)\Big(1+\lambda_i(X)\lambda_r(R)\Big)+\sum_{i=1}^{n-r}\sigma_i^2(A)\lambda_i(X)\Big(1+\lambda_i(X)\lambda_r(R)\Big)+\sum_{i=n-r+1}^{k}\sigma_i^2(A)\lambda_i(X)\\
\le&\Big(\lambda_1(Q)\lambda_r(R)+\sigma_1^2(A)\Big)\sum_{i=1}^{k}\lambda_i(X)+\sum_{i=1}^{k}\lambda_i(Q)+\sigma_1^2(A)\lambda_r(R)\sum_{i=1}^{n-r}\lambda_i^2(X). \qquad (2.10)
\end{aligned}$$

Because

$$\sum_{i=1}^{n-r}\lambda_i^2(X)\le\left(\sum_{i=1}^{n-r}\lambda_i(X)\right)^2\le\frac{(n-r)\lambda_1^2(Q)}{(1-\sigma_1^2(A))^2}, \qquad (2.11)$$

$$\sum_{i=1}^{k}\lambda_i^2(X)\ge\frac{1}{k}\left(\sum_{i=1}^{k}\lambda_i(X)\right)^2, \qquad (2.12)$$

we see that (2.10) is valid if

$$\lambda_r(R)\left(\sum_{i=1}^{k}\lambda_i(X)\right)^2-k\Theta\sum_{i=1}^{k}\lambda_i(X)-k\beta\le 0$$

holds true. By directly solving this quadratic inequality with respect to $\sum_{i=1}^{k}\lambda_i(X)$ we obtain the estimate (2.8b) for the case $\eta=\beta$.

We now turn to verify the validity of (2.8b) for the case $\eta=\delta$. In fact, from (2.11) and (2.12) we know that (2.10) is valid if

$$\lambda_r(R)\left(\sum_{i=1}^{k}\lambda_i(X)\right)^2-k\Theta\sum_{i=1}^{k}\lambda_i(X)-k\delta\le 0$$

holds true. After directly solving this quadratic inequality with respect to $\sum_{i=1}^{k}\lambda_i(X)$ we obtain the estimate (2.8b) when $\eta=\delta$. □

Theorem 2.2 directly leads to the following upper bound about the trace for the solution of the DARE (1.3).

Corollary 2.2. *Let $X\in\mathbf{R}^{n\times n}$ be the symmetric positive definite solution of the DARE* (1.3). *If $\sigma_1(A)<1$, then*

$$\operatorname{tr}(X)\le\frac{n\big(\Theta+\sqrt{\Theta^2+4\lambda_r(R)\eta/n}\big)}{2\lambda_r(R)},$$

where the constants Θ and η are defined as in Theorem 2.2.

From (1.2) we see that if $B=0$, then $R=0$. Hence, the DARE (1.3) reduces to the discrete Lyapunov equation

$$X=A^TXA+Q. \qquad (2.13)$$

About partial sum of the eigenvalues for its solution, in accordance with Theorem 2.2 we immediately have the following upper bound, which was originally given in [16].

Corollary 2.3. *Let $X \in \mathbf{R}^{n\times n}$ be the symmetric positive definite solution of the discrete Lyapunov equation* (2.13). *Assume that $\sigma_1(A) < 1$. Then*

$$\sum_{i=1}^{k} \lambda_i(X) \leq \frac{1}{1-\sigma_1^2(A)} \sum_{i=1}^{k} \lambda_i(Q), \qquad k = 1, \cdots, n.$$

If R is nonsingular, then $r = n$. In accordance with Theorem 2.2 again, we have the following upper bound about partial sum of the eigenvalues for the solution of the DARE (1.3), which was originally obtained in [16, 45].

Corollary 2.4. *Let $X \in \mathbf{R}^{n\times n}$ be the symmetric positive definite solution of the DARE* (1.3). *If R is nonsingular and $\sigma_1(A) < 1$, then*

$$\sum_{i=1}^{k} \lambda_i(X) \leq \frac{k\big(\Theta + \sqrt{\Theta^2 + 4\lambda_n(R)\beta/k}\big)}{2\lambda_n(R)}, \qquad k = 1, \cdots, n,$$

where

$$\Theta = \sigma_1^2(A) + \lambda_1(Q)\lambda_n(R) - 1, \qquad \beta = \sum_{i=1}^{k} \lambda_i(Q).$$

In particular, it holds that

$$\lambda_1(X) \leq \frac{\Theta + \sqrt{\Theta^2 + 4\lambda_n(R)\lambda_1(R)}}{2\lambda_n(R)}.$$

When the arithmetic-geometric mean inequality [35] is applied to the estimates in Theorem 2.2, we can obtain upper bound about partial product of the eigenvalues for the solution of the DARE (1.3).

Theorem 2.3. *Let $X \in \mathbf{R}^{n\times n}$ be the symmetric positive definite solution of the DARE* (1.3). *Assume that $\sigma_1(A) < 1$. Then it holds that*

$$\begin{aligned}
&\prod_{i=1}^{k} \lambda_i(X) \leq \left(\frac{1}{k(1-\sigma_1^2(A))} \sum_{i=1}^{k} \lambda_i(Q) \right)^{k}, && k = 1, \cdots, n-r, \\
&\prod_{i=1}^{k} \lambda_i(X) \leq \left(\frac{\Theta + \sqrt{\Theta^2 + 4\lambda_r(R)\eta/k}}{2\lambda_r(R)} \right)^{k}, && k = n-r+1, n-r+2, \cdots, n,
\end{aligned}$$

where the constants Θ and η are defined as in Theorem 2.2.

Theorem 2.3 directly leads to the following upper bound about the determinant of the solution for the DARE (1.3).

Corollary 2.5. *Let $X \in \mathbf{R}^{n\times n}$ be the symmetric positive definite solution of the DARE* (1.3). *Assume that $\sigma_1(A) < 1$. Then it holds that*

$$\det(X) \leq \left(\frac{\Theta + \sqrt{\Theta^2 + 4\lambda_r(R)\eta/n}}{2\lambda_r(R)} \right)^{n},$$

where the constants Θ and η are defined as in Theorem 2.2.

3. The Convergence of Fixed-Point Iteration (1.5)

The symmetric positive definite solution $X_\star$ of the DARE (1.3) can be computed by the fixed-point iteration scheme (1.5). When the matrices A is nonsingular and Q is symmetric positive definite, Komaroff proved in [18] that the matrix sequence $\{X_k\}_{k=0}^{\infty}$ generated by the iteration scheme (1.5), starting from $X_0 = Q$, is monotonically increasing and convergent to $X_\star$. We point out that the requirement of the nonsingularity of the matrix A is not necessary. Moreover, we establish the following theorem to precisely describe the linear convergence and the asymptotic convergence rate of the fixed-point iteration (1.5).

Theorem 3.1. *Consider the DARE* (1.3) *and let $X_\star$ be its unique symmetric positive definite solution. Let $Q \succ 0$ and $\{X_k\}_{k=0}^{\infty}$ be the iteration sequence generated by the fixed-point iteration* (1.5) *starting from $X_0 = Q$. Then*

(i) *the iteration sequence $\{X_k\}_{k=0}^{\infty}$ converges monotonically increasingly to $X_\star$, i.e., it holds that*

$$Q \preceq X_k \preceq X_{k+1} \preceq X_\star, \qquad k = 0, 1, 2, \cdots,$$

and $\lim_{k\to\infty} X_k = X_\star$;

(ii) *when $æ := \|(I + RX_\star)^{-1}A\|_2 < 1$ and $\epsilon := \|X_\star - X_0\|_2 < \frac{1-æ^2}{\gamma æ^2}$, with $\gamma = \|R\|_2$, the convergence rate of the iteration sequence $\{X_k\}_{k=0}^{\infty}$ is at least linear, i.e., it holds that*

$$\|X_\star - X_{k+1}\|_2 \le æ^2(1 + \gamma\epsilon\varrho^k)\|X_\star - X_k\|_2, \qquad k = 0, 1, 2, \cdots,$$

where $\varrho = æ^2(1 + \gamma\epsilon)$;

(iii) *the R-convergence factor of the iteration sequence $\{X_k\}_{k=0}^{\infty}$ is at most $æ_\star^2$, with $æ_\star = \rho((I + RX_\star)^{-1}A)$, i.e., it holds that*

$$\lim_{k\to\infty} \sup \sqrt[k]{\|X_\star - X_k\|_2} \le æ_\star^2.$$

Proof. (i) follows from slight and technical modifications of the corresponding proof in [18]. We now turn to demonstrate the validity of (ii). We first assert that for $X \in \mathbf{S}^n$ and $R \in \mathbf{S}_+^n$ such that $X \succ 0$, the matrix defined by

$$W = R(I + XR)^{-1}$$

satisfies

$$W \in \mathbf{S}_+^n \quad \text{and} \quad \|W\|_2 \le \|R\|_2. \tag{3.1}$$

In fact, if we write $Y = X^{1/2}RX^{1/2}$, then the matrix $Y \in \mathbf{S}_+^n$. Through Y we can easily rewrite W as

$$W = X^{-1/2}Y(I + Y)^{-1}X^{-1/2}.$$

Hence, $W \in \mathbf{S}_+^n$. Moreover, it holds that

$$\begin{aligned} W^2 =& (I + RX)^{-1}R^2(I + XR)^{-1} \\ \preceq& \rho(R)^2(I + RX)^{-1}(I + XR)^{-1} \\ =& \rho(R)^2[(I + XR)(I + RX)]^{-1} \\ =& \rho(R)^2X^{-1/2}(I + Y)^{-1}X(I + Y)^{-1}X^{-1/2} \\ \preceq& \rho(R)^2 I. \end{aligned}$$

Therefore, we can obtain $\|W\|_2 \leq \|R\|_2$.

As $X_\star$ is the unique symmetric positive definite solution of the DARE (1.3), it must satisfy the matrix equality

$$X_\star = A^T(X_\star^{-1} + R)^{-1}A + Q.$$

After subtracting

$$X_{k+1} = A^T(X_k^{-1} + R)^{-1}A + Q$$

from the above matrix equality we have

$$\begin{aligned} X_\star - X_{k+1} =& A^T[(X_\star^{-1} + R)^{-1} - (X_k^{-1} + R)^{-1}]A \\ =& A^T(I + X_kR)^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A \\ =& A^T(I + X_\star R)^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A \\ &+ A^T(I + X_\star R)^{-1}(X_\star - X_k)R(I + X_kR)^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A. \end{aligned}$$

By making use of (3.1), we know that

$$\begin{aligned} 0 \preceq& X_\star - X_{k+1} \\ \preceq& A^T(I + X_\star R)^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A \\ &+ \|R\|_2 A^T(I + X_\star R)^{-1}(X_\star - X_k)^2(I + RX_\star)^{-1}A. \end{aligned} \tag{3.2}$$

Hence, it follows from Lemma 2.1(vii) that

$$\|X_\star - X_{k+1}\|_2 \leq \|(I + RX_\star)^{-1}A\|_2^2(1 + \|R\|_2\|X_\star - X_k\|_2)\|X_\star - X_k\|_2. \tag{3.3}$$

For notational simplicity, we denote by

$$\epsilon_k = \|X_\star - X_k\|_2, \quad k = 0, 1, 2, \cdots .$$

So, the error relationship (3.3) can be equivalently expressed in the following brief form:

$$\epsilon_{k+1} \leq æ^2(1 + \gamma\epsilon_k)\epsilon_k, \quad k = 0, 1, 2, \cdots . \tag{3.4}$$

It easily follows from (i) that

$$X_\star - X_k \succeq X_\star - X_{k+1} \succeq 0, \quad k = 0, 1, 2, \cdots .$$

Hence, by making use of Lemma 2.1(vii) again we can get

$$\|X_\star - X_{k+1}\|_2 \leq \|X_\star - X_k\|_2 \leq \cdots \leq \|X_\star - X_0\|_2,$$

or, equivalently,

$$\epsilon_{k+1} \leq \epsilon_k \leq \epsilon, \qquad k = 0, 1, 2, \cdots .$$

By applying this estimate to (3.4) we obtain

$$\epsilon_k \leq æ^2(1 + \gamma\epsilon)\epsilon_{k-1} = \varrho\epsilon_{k-1} \leq \varrho^k\epsilon.$$

Therefore, after substituting this estimate into (3.4) again we have

$$\epsilon_{k+1} \leq æ^2(1 + \gamma\epsilon\varrho^k)\epsilon_k.$$

This demonstrates the validity of (ii).

The verification of (iii) is based on the matrix estimate (3.2). From (3.1) we have

$$R(I + X_k R)^{-1} \preceq \|R\|_2 I.$$

Recalling that the matrix sequence $\{X_k\}_{k=0}^{\infty}$ is convergent to $X_\star$, we know that for any $\epsilon_o > 0$ there exists a positive integer k_o such that

$$0 \preceq X_\star - X_k \preceq \epsilon_o I$$

holds for all $k = k_o, k_o + 1, \cdots$. Hence, we have

$$(X_\star - X_k)^{1/2} R(I + X_k R)^{-1} (X_\star - X_k)^{1/2} \preceq \epsilon_o \|R\|_2 I, \qquad k \geq k_o.$$

After substituting this estimate into (3.2), we immediately obtain that

$$\begin{aligned} 0 \preceq & X_\star - X_{k+1} \\ \preceq & \big(1 + \epsilon_o \|R\|_2\big) A^T (I + X_\star R)^{-1} (X_\star - X_k)(I + R X_\star)^{-1} A \\ \preceq & \big(1 + \epsilon_o \|R\|_2\big)^{k-k_o+1} \big(A^T (I + X_\star R)^{-1}\big)^{k-k_o+1} (X_\star - X_{k_o}) \big((I + R X_\star)^{-1} A\big)^{k-k_o+1} \end{aligned}$$

hold for all $k = k_o, k_o + 1, \cdots$. It follows from Lemma 2.1(vii) that

$$\|X_\star - X_k\|_2 \leq \big(1 + \epsilon_o \|R\|_2\big)^{k-k_o} \big\| \big((I + R X_\star)^{-1} A\big)^{k-k_o} \big\|_2^2 \, \|X_\star - X_{k_o}\|_2, \qquad k \geq k_o,$$

and, therefore,

$$\lim_{k \to \infty} \sup \sqrt[k]{\|X_\star - X_k\|_2} \leq \Big(\rho\big((I + R X_\star)^{-1} A\big) \Big)^2.$$

Here, we have used the facts that ϵ_o is an arbitrary small positive quantity and that $\lim_{k\to\infty} \|Z^k\|^{1/k} = \rho(Z)$ holds for any square matrix Z in any matrix norm. □

Evidently, it always holds that $\rho((I + R X_\star)^{-1} A) < 1$. We remark that the definition of the asymptotic convergence rate for an iteration sequence used here is the same as that given in [39].

A practical implementation strategy about the fixed-point iteration (1.5) may be as follows:

1. compute the Cholesky factorization of X_k to obtain $X_k = L_k L_k^T$ by employing a numerically stable algorithm; see, e.g., [10];
2. compute $\widehat{R}_k = I + L_k^T R L_k$ and $\widehat{A}_k = L_k^T A$;
3. compute the Cholesky factorization of $\widehat{R}_k$ to obtain $\widehat{R}_k = \widehat{L}_k \widehat{L}_k^T$ by employing, again, the numerically stable algorithm;
4. solve the lower-triangular system of linear equations $\widehat{L}_k \widehat{B}_k = \widehat{A}_k$ to obtain $\widehat{B}_k$; and
5. compute $X_{k+1} = \widehat{B}_k^T \widehat{B}_k + Q$ by only computing the entries of the upper- or the lower-triangular part of the matrix $\widehat{B}_k^T \widehat{B}_k + Q$ according to its symmetry.

When the matrix Q is singular, according to its spectral decomposition

$$Q = U\,\mathrm{diag}\big(\lambda_1(Q), \lambda_2(Q), \cdots, \lambda_{r_q}(Q), 0_{n-r_q}\big)\,U^T,$$

with $r_q = \mathrm{rank}(Q)$ and U an orthogonal matrix, we define

$$\widehat{Q} = U\,\mathrm{diag}\big(\lambda_1(Q), \lambda_2(Q), \cdots, \lambda_q(Q), \tau I\big)\,U^T,$$

where

$$\tau = \frac{\sigma_n^2(A) - 1 + \left|\sigma_n^2(A) - 1\right|}{2\lambda_1(R)} + \widehat{\varepsilon},$$

with $\widehat{\varepsilon} > 0$ a sufficiently small constant. Then we choose the starting matrix X_0 in the fixed-point iteration (1.5) as $X_0 = \widehat{Q}$. From [17] we know that $\tau \le \lambda_n(X_\star)$. Hence, it follows from $Q \preceq X_\star$ that $X_0 = \widehat{Q} \preceq X_\star$.

When the matrix R is symmetric positive definite, the symmetric positive definite solution $X_\star$ of the DARE (1.3) satisfies $X_\star \preceq A^T R^{-1} A + Q$; see [18]. Similarly, it can be shown that the iteration sequence $\{X_k\}_{k=0}^{\infty}$ generated by the fixed-point iteration (1.5), starting from $X_0 = A^T R^{-1} A + Q$, is monotonically decreasing and convergent to $X_\star$ R-linearly.

4. The Modified Fixed-Point Iteration

The fixed-point iteration (1.5) requires to compute two matrix inversions X_k^{-1} and $(X_k^{-1} + R)^{-1}$. In this section, we propose an effective variant that avoids computing the inverse of the matrix $X_k^{-1} + R$. The basic idea is to replace the matrix $(X_k^{-1} + R)^{-1}$ by only one-step approximation produced by the Schulz iteration [41]. This yields the following modified fixed-point iteration:

$$\begin{cases} X_{k+1} = A^T Y_k A + Q, \\ Y_{k+1} = Y_k\big(2I - (X_{k+1}^{-1} + R)Y_k\big), \end{cases} \qquad k = 0, 1, 2, \cdots, \tag{4.1}$$

where $Y_0 = (Q^{-1} + R)^{-1}$. To analyze the convergence of the iteration scheme (4.1), we introduce the following necessary notations:

$$\begin{aligned} &\text{æ} = \|(I + RX_\star)^{-1}A\|_2, \qquad \varkappa = \|(I + X_\star R)^{-1}\|_2, \qquad \varphi = \|(I + RQ)^{-1}\|_2, \\ &\gamma = \|R\|_2, \qquad \tau = \|Q^{-1}\|_2, \qquad \tau_\star = \|X_\star^{-1}\|_2, \qquad \omega = \|A\|_2. \end{aligned}$$

Based on them, we further define the following quantities:

$$\begin{aligned} \delta_x &= \max\left\{0, \quad \frac{1}{2}\left(\sqrt{[\omega^2(\tau_\star + \gamma) - \tau \text{æ}^2]^2 + \omega^2\tau^2(\omega\tau_\star\upsilon + 2\text{æ})^2} - [\omega^2(\tau_\star + \gamma) + \tau\text{æ}^2]\right)\right\}, \\ \delta_y &= \max\left\{0, \quad \frac{1}{2}\left(\sqrt{(\tau_\star + \gamma - \tau\varkappa^2)^2 + \tau^2(\tau_\star\upsilon + 2\varkappa)^2} - (\tau_\star + \gamma + \tau\varkappa^2)\right)\right\}, \end{aligned}$$

and

$$\begin{aligned} &\alpha_x = \omega^2(\tau_\star + \gamma) + \delta_x, \qquad &&\beta_x = \tau\text{æ}^2 + \delta_x, \\ &\alpha_y = \tau_\star + \gamma + \delta_y, &&\beta_y = \tau\varkappa^2 + \delta_y, \\ &\alpha = \max\{\alpha_x, \alpha_y\}, &&\beta = \max\{\beta_x, \beta_y\}, \qquad \psi = \max\{\text{æ}, \varkappa\}. \end{aligned}$$

The following theorem precisely describes the monotone convergence and the corresponding asymptotic convergence rate of the modified fixed-point iteration (4.1).

Theorem 4.1. *Consider the DARE* (1.3) *and let $X_\star$ be its unique symmetric positive definite solution. Let $Q \succ 0$ and $\{X_k\}_{k=0}^\infty$ be the iteration sequence generated by the modified fixed-point iteration* (4.1) *starting from $Y_0 = (Q^{-1} + R)^{-1}$. Let $X_0 = Q$ and denote by $Y_\star = (X_\star^{-1} + R)^{-1}$. Then*

(i) *the iteration sequence $\{X_k\}_{k=0}^\infty$ converges monotonically increasingly to $X_\star$, i.e., it holds that*

$$\begin{cases} Q \preceq X_k \preceq X_{k+1} \preceq X_\star, \\ Y_0 \preceq Y_k \preceq Y_{k+1} \preceq Y_\star, \end{cases} \qquad k = 0, 1, 2, \cdots, \tag{4.2}$$

and $\lim_{k\to\infty} X_k = X_\star$, $\lim_{k\to\infty} Y_k = Y_\star$;

(ii) *when* $(\alpha + \beta)\psi^2 < 1$ *and*

$$\epsilon_0 := \|X_\star - X_0\|_2 \le \epsilon < \frac{1 - (\alpha+\beta)\psi^2}{(\alpha+\beta)(\varkappa\varphi + \beta)},$$
$$\upsilon_0 := \|Y_\star - Y_0\|_2 \le \upsilon < \frac{1 - (\alpha+\beta)\psi^2}{(\alpha+\beta)(\alpha + \beta\omega^2)},$$

the convergence rate of the iteration sequence $\{X_k\}_{k=0}^\infty$ is at least linear, i.e., it holds that

$$\|X_\star - X_{k+1}\|_2 \le \varrho^{2(k-1)}(\alpha + \beta\omega^2)^2\upsilon^2 + \ae^2\|X_\star - X_k\|_2, \qquad k = 1, 2, \cdots, \tag{4.3}$$

where $\varrho = \left(1 + \frac{\alpha}{\beta}\right)\left((\alpha\upsilon + \beta\epsilon)\beta + \psi^2\right)$;

(iii) *the R-convergence factor of the iteration sequence $\{X_k\}_{k=0}^\infty$ is at most $\ae_\star^2$, with $\ae_\star = \rho((I + RX_\star)^{-1}A)$, i.e., it holds that*

$$\lim_{k\to\infty} \sup \sqrt[k]{\|X_\star - X_k\|_2} \le \ae_\star^2.$$

Proof. We first use induction to demonstrate the validity of (4.2). From Theorem 3.1(i) we have $X_\star \succeq X_1 \succeq X_0 = Q$. Hence, it follows from Lemma 2.1(iii)-(iv) that

$$\begin{aligned} Y_1 =& 2Y_0 - Y_0(X_1^{-1} + R)Y_0 \\ \succeq& 2Y_0 - Y_0(Q^{-1} + R)Y_0 = Y_0 \end{aligned}$$

and

$$\begin{aligned} Y_1 =& 2Y_0 - Y_0(X_1^{-1} + R)Y_0 \\ \preceq& (X_1^{-1} + R)^{-1} \\ \preceq& (X_\star^{-1} + R)^{-1} = Y_\star, \end{aligned}$$

where the inequality

$$2Y_0 - Y_0(X_1^{-1} + R)Y_0 \preceq (X_1^{-1} + R)^{-1}$$

is obtained by making use of Lemma 2.2. This shows the validity of the inequality (4.2) for $k = 0$.

Assume that the inequality (4.2) holds for $k = \ell$, i.e.,

$$\begin{cases} Q \preceq X_\ell \preceq X_{\ell+1} \preceq X_\star, \\ Y_0 \preceq Y_\ell \preceq Y_{\ell+1} \preceq Y_\star. \end{cases}$$

Then we need to verify that it is valid for $k = \ell + 1$, too. In fact, as

$$Y_\ell \preceq Y_{\ell+1} \preceq Y_\star,$$

by Lemma 2.1(iii) we have

$$\begin{aligned} X_{\ell+1} &= A^T Y_\ell A + Q \preceq A^T Y_{\ell+1} A + Q = X_{\ell+2}, \\ X_{\ell+2} &= A^T Y_{\ell+1} A + Q \preceq A^T Y_\star A + Q = A^T (X_\star^{-1} + R)^{-1} A + Q = X_\star. \end{aligned}$$

In addition, by making use of Lemma 2.2 we easily see that

$$Y_{\ell+1} = 2Y_\ell - Y_\ell (X_{\ell+1}^{-1} + R) Y_\ell \preceq (X_{\ell+1}^{-1} + R)^{-1}.$$

It then follows from this inequality and $X_{\ell+1} \preceq X_{\ell+2}$ that

$$Y_{\ell+1}^{-1} \succeq X_{\ell+1}^{-1} + R \succeq X_{\ell+2}^{-1} + R.$$

Hence,

$$Y_{\ell+2} = Y_{\ell+1} + Y_{\ell+1}\big(Y_{\ell+1}^{-1} - (X_{\ell+2}^{-1} + R)\big) Y_{\ell+1} \succeq Y_{\ell+1}.$$

Besides, as $X_{\ell+2} \preceq X_\star$, by making use of Lemma 2.2 and Lemma 2.1(iii) we obtain

$$\begin{aligned} Y_{\ell+2} =& 2Y_{\ell+1} - Y_{\ell+1}(X_{\ell+2}^{-1} + R) Y_{\ell+1} \\ \preceq& (X_{\ell+2}^{-1} + R)^{-1} \\ \preceq& (X_\star^{-1} + R)^{-1} = Y_\star. \end{aligned}$$

The above demonstration shows the validity of the inequality (4.2) for $k = \ell + 1$.

By induction, the inequality (4.2) is true for all nonnegative integers $k = 0, 1, 2, \cdots$.

According to the monotonicity, the matrix sequences $\{X_k\}_{k=0}^\infty$ and $\{Y_k\}_{k=0}^\infty$ are convergent. Through taking limits on both equalities in (4.1) and considering the uniqueness of the solution of the DARE (1.3), we know that

$$\lim_{k\to\infty} X_k = X_\star \quad \text{and} \quad \lim_{k\to\infty} Y_k = Y_\star.$$

Now, we turn to prove (ii). It follows from direct operations that

$$\begin{aligned} I - Y_k Y_\star^{-1} =& (I - Y_{k-1} Y_\star^{-1})^2 + Y_{k-1}[(X_k^{-1} + R) - (X_\star^{-1} + R)] Y_{k-1} Y_\star^{-1} \\ =& (I - Y_{k-1} Y_\star^{-1})^2 + Y_{k-1}(X_k^{-1} - X_\star^{-1}) Y_{k-1} Y_\star^{-1} \end{aligned}$$

and, hence,

$$\begin{aligned} Y_\star - Y_k =& (I - Y_{k-1} Y_\star^{-1})^2 Y_\star + Y_{k-1}(X_k^{-1} - X_\star^{-1}) Y_{k-1} \\ =& (Y_\star - Y_{k-1})(X_\star^{-1} + R)(Y_\star - Y_{k-1}) + Y_{k-1}(X_k^{-1} - X_\star^{-1}) Y_{k-1} \\ =& (Y_\star - Y_{k-1})(X_\star^{-1} + R)(Y_\star - Y_{k-1}) + Y_\star(X_k^{-1} - X_\star^{-1}) Y_\star + E_k, \end{aligned} \tag{4.4}$$

where

$$\begin{aligned} E_k =& Y_{k-1}(X_k^{-1} - X_\star^{-1}) Y_{k-1} - Y_\star (X_k^{-1} - X_\star^{-1}) Y_\star \\ =& (Y_{k-1} - Y_\star)(X_k^{-1} - X_\star^{-1})(Y_{k-1} - Y_\star) \\ & + (Y_{k-1} - Y_\star) X_k^{-1}(X_\star - X_k) X_\star^{-1} Y_\star \\ & + Y_\star X_\star^{-1}(X_\star - X_k) X_k^{-1}(Y_{k-1} - Y_\star) \\ =& (Y_{k-1} - Y_\star) X_k^{-1}(X_\star - X_k) X_\star^{-1}(Y_{k-1} - Y_\star) \\ & + (Y_{k-1} - Y_\star) X_k^{-1}(X_\star - X_k)(I + R X_\star)^{-1} \\ & + (I + X_\star R)^{-1}(X_\star - X_k) X_k^{-1}(Y_{k-1} - Y_\star). \end{aligned} \tag{4.5}$$

Evidently, E_k is a symmetric matrix. Note that

$$0 \prec X_\star^{-1} \preceq X_k^{-1} \preceq Q^{-1} \tag{4.6}$$

holds true due to $0 \prec Q \preceq X_k \preceq X_\star$; see (4.2). Then from direct operations we obtain

$$\begin{aligned} Y_\star(X_k^{-1} - X_\star^{-1})Y_\star =&Y_\star X_\star^{-1}(X_\star - X_k)X_\star^{-1}Y_\star - Y_\star X_\star^{-1}(I - X_\star X_k^{-1})(X_\star - X_k)X_\star^{-1}Y_\star \\ =&Y_\star X_\star^{-1}(X_\star - X_k)X_\star^{-1}Y_\star + Y_\star X_\star^{-1}(X_\star - X_k)X_k^{-1}(X_\star - X_k)X_\star^{-1}Y_\star \\ \preceq&Y_\star X_\star^{-1}(X_\star - X_k)X_\star^{-1}Y_\star + Y_\star X_\star^{-1}(X_\star - X_k)Q^{-1}(X_\star - X_k)X_\star^{-1}Y_\star. \end{aligned}$$

After substituting this inequality into (4.4) we get

$$\begin{aligned} Y_\star - Y_k \preceq&(Y_\star - Y_{k-1})(X_\star^{-1} + R)(Y_\star - Y_{k-1}) \\ &+ Y_\star X_\star^{-1}(X_\star - X_k)Q^{-1}(X_\star - X_k)X_\star^{-1}Y_\star \\ &+ Y_\star X_\star^{-1}(X_\star - X_k)X_\star^{-1}Y_\star + E_k, \end{aligned}$$

or equivalently,

$$\begin{aligned} Y_\star - Y_k \preceq&(Y_\star - Y_{k-1})(X_\star^{-1} + R)(Y_\star - Y_{k-1}) \\ &+ (I + X_\star R)^{-1}(X_\star - X_k)Q^{-1}(X_\star - X_k)(I + RX_\star)^{-1} \\ &+ (I + X_\star R)^{-1}(X_\star - X_k)(I + RX_\star)^{-1} + E_k. \end{aligned} \tag{4.7}$$

In addition, by applying (4.1) to (4.7) we get

$$\begin{aligned} X_\star - X_{k+1} =&A^T(Y_\star - Y_k)A \\ \preceq&A^T(Y_\star - Y_{k-1})(X_\star^{-1} + R)(Y_\star - Y_{k-1})A \\ &+ A^T(I + X_\star R)^{-1}(X_\star - X_k)Q^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A \\ &+ A^T(I + X_\star R)^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A + \widetilde{E}_k, \end{aligned} \tag{4.8}$$

where

$$\begin{aligned} \widetilde{E}_k =&A^T E_k A \\ =&A^T(Y_{k-1} - Y_\star)X_k^{-1}(X_\star - X_k)X_\star^{-1}(Y_{k-1} - Y_\star)A \\ &+ A^T(Y_{k-1} - Y_\star)X_k^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A \\ &+ A^T(I + X_\star R)^{-1}(X_\star - X_k)X_k^{-1}(Y_{k-1} - Y_\star)A. \end{aligned} \tag{4.9}$$

Evidently, $\widetilde{E}_k$ is a symmetric matrix due to the symmetry of E_k. For notational convenience, we denote by

$$\epsilon_k = \|X_\star - X_k\|_2 \quad \text{and} \quad \upsilon_k = \|Y_\star - Y_k\|_2, \qquad k = 0, 1, 2, \cdots.$$

As

$$\begin{aligned} Y_\star - Y_0 =&(X_\star^{-1} + R)^{-1} - (Q^{-1} + R)^{-1} \\ =&(X_\star^{-1} + R)^{-1}(Q^{-1} - X_\star^{-1})(Q^{-1} + R)^{-1} \\ =&(I + X_\star R)^{-1}(X_\star - Q)(I + RQ)^{-1}, \end{aligned}$$

it holds that

$$\begin{aligned}\upsilon_0 =&\|Y_\star - Y_0\|_2\\ \leq&\|(I + X_\star R)^{-1}\|_2\|X_\star - Q\|_2\|(I + RQ)^{-1}\|_2\\ =&\varkappa\varphi\epsilon = \upsilon,\end{aligned}$$

where $\upsilon = \varkappa\varphi\epsilon$. Therefore, from (4.2) we can easily obtain the relationships

$$\begin{aligned}&0 \preceq X_\star - X_{k+1} \preceq X_\star - X_k \preceq \cdots \preceq X_\star - X_0 \preceq \epsilon I,\\ &0 \preceq Y_\star - Y_{k+1} \preceq Y_\star - Y_k \preceq \cdots \preceq Y_\star - Y_0 \preceq \upsilon I.\end{aligned}$$

By making use of Lemma 2.1(vii) we can further get

$$\begin{aligned}&\|X_\star - X_{k+1}\|_2 \leq \|X_\star - X_k\|_2 \leq \cdots \leq \|X_\star - X_0\|_2 \leq \epsilon,\\ &\|Y_\star - Y_{k+1}\|_2 \leq \|Y_\star - Y_k\|_2 \leq \cdots \leq \|Y_\star - Y_0\|_2 \leq \upsilon,\end{aligned}$$

or equivalently,

$$\epsilon_{k+1} \leq \epsilon_k \leq \epsilon \quad \text{and} \quad \upsilon_{k+1} \leq \upsilon_k \leq \upsilon, \qquad k = 0, 1, 2, \cdots . \tag{4.10}$$

It follows from (4.5), (4.9) and (4.10) that

$$\begin{aligned}\|E_k\|_2 \leq&\|Y_{k-1} - Y_\star\|_2^2\ \|X_k^{-1}\|_2\|X_\star - X_k\|_2\ \|X_\star^{-1}\|_2\\ &+ 2\|Y_{k-1} - Y_\star\|_2\ \|X_k^{-1}\|_2\ \|X_\star - X_k\|_2\ \|(I + RX_\star)^{-1}\|_2\\ \leq&\|Q^{-1}\|_2\|X_\star^{-1}\|_2\ \|Y_{k-1} - Y_\star\|_2^2\ \|X_\star - X_k\|_2\\ &+ 2\|Q^{-1}\|_2\ \|(I + RX_\star)^{-1}\|_2\ \|Y_{k-1} - Y_\star\|_2\ \|X_\star - X_k\|_2\\ =&\tau(\tau_\star\upsilon_{k-1} + 2\varkappa)\upsilon_{k-1}\epsilon_k \leq \tau(\tau_\star\upsilon + 2\varkappa)\upsilon_{k-1}\epsilon_k\end{aligned} \tag{4.11}$$

and

$$\begin{aligned}\|\widetilde{E}_k\|_2 \leq&\|A\|_2^2\ \|Q^{-1}\|_2\ \|X_\star^{-1}\|_2\|Y_\star - Y_{k-1}\|_2^2\ \|X_\star - X_k\|_2\\ &+ 2\ \|Q^{-1}\|_2\ \|A\|_2\ \|(I + RX_\star)^{-1}A\|_2\|Y_\star - Y_{k-1}\|_2\ \|X_\star - X_k\|_2\\ \leq&\omega\tau(\omega\tau_\star\upsilon_{k-1} + 2\text{æ})\upsilon_{k-1}\epsilon_k \leq \omega\tau(\omega\tau_\star\upsilon + 2\text{æ})\upsilon_{k-1}\epsilon_k.\end{aligned} \tag{4.12}$$

Here we have used the estimate (4.6) and Lemma 2.1(vii). In addition, from (4.7) and (4.11) as well as (4.8) and (4.12), by making use of Lemma 2.1(vii) again we obtain

$$\begin{aligned}\|Y_\star - Y_k\|_2 \leq&\|(Y_\star - Y_{k-1})(X_\star^{-1} + R)(Y_\star - Y_{k-1})\|_2\\ &+ \|(I + X_\star R)^{-1}(X_\star - X_k)Q^{-1}(X_\star - X_k)(I + RX_\star)^{-1}\|_2\\ &+ \|(I + X_\star R)^{-1}(X_\star - X_k)(I + RX_\star)^{-1}\|_2 + \|E_k\|_2\\ \leq&(\|X_\star^{-1}\|_2 + \|R\|_2)\|Y_\star - Y_{k-1}\|_2^2\\ &+ \|Q^{-1}\|_2\|(I + X_\star R)^{-1}\|_2^2\|X_\star - X_k\|_2^2\\ &+ \|(I + X_\star R)^{-1}\|_2^2\|X_\star - X_k\|_2 + \|E_k\|_2\\ \leq&(\tau_\star + \gamma)\upsilon_{k-1}^2 + \varkappa^2(\tau\epsilon_k + 1)\epsilon_k + \tau(\tau_\star\upsilon + 2\varkappa)\upsilon_{k-1}\epsilon_k\\ \leq&(\alpha_y\upsilon_{k-1} + \beta_y\epsilon_k)^2 + \varkappa^2\epsilon_k \leq (\alpha\upsilon_{k-1} + \beta\epsilon_k)^2 + \psi^2\epsilon_k,\end{aligned}$$

as well as

$$\begin{aligned}
\|X_\star - X_{k+1}\|_2 \leq & \|A^T(Y_\star - Y_{k-1})(X_\star^{-1} + R)(Y_\star - Y_{k-1})A\|_2 \\
& \quad + \|A^T(I + X_\star R)^{-1}(X_\star - X_k)Q^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A\|_2 \\
& \quad + \|A^T(I + X_\star R)^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A\|_2 + \|\widetilde{E}_k\|_2 \\
\leq & \|A\|_2^2(\|X_\star^{-1}\|_2 + \|R\|_2)\|Y_\star - Y_{k-1}\|_2^2 \\
& \quad + \|Q^{-1}\|_2\|(I + RX_\star)^{-1}A\|_2^2\|X_\star - X_k\|_2^2 \\
& \quad + \|(I + RX_\star)^{-1}A\|_2^2\|X_\star - X_k\|_2 + \|\widetilde{E}_k\|_2 \\
\leq & \omega^2(\tau_\star + \gamma)\upsilon_{k-1}^2 + \text{æ}^2(\tau\epsilon_k + 1)\epsilon_k + \omega\tau(\omega\tau_\star\upsilon + 2\text{æ})\upsilon_{k-1}\epsilon_k \\
\leq & (\alpha_x\upsilon_{k-1} + \beta_x\epsilon_k)^2 + \text{æ}^2\epsilon_k \leq (\alpha\upsilon_{k-1} + \beta\epsilon_k)^2 + \psi^2\epsilon_k.
\end{aligned}$$

That is to say, it holds that

$$\epsilon_{k+1} \leq (\alpha\upsilon_{k-1} + \beta\epsilon_k)^2 + \text{æ}^2\epsilon_k \leq (\alpha\upsilon_{k-1} + \beta\epsilon_k)^2 + \psi^2\epsilon_k, \tag{4.13a}$$

$$\upsilon_k \leq (\alpha\upsilon_{k-1} + \beta\epsilon_k)^2 + \varkappa^2\epsilon_k \leq (\alpha\upsilon_{k-1} + \beta\epsilon_k)^2 + \psi^2\epsilon_k. \tag{4.13b}$$

Define the difference sequence $\{\chi_k\}_{k=1}^{\infty}$ as

$$\chi_k = \alpha\upsilon_{k-1} + \beta\epsilon_k, \quad k = 1, 2, \cdots.$$

Then we have

$$\chi_1 = \alpha\upsilon_0 + \beta\epsilon_1 \leq (\alpha + \beta\omega^2)\upsilon, \tag{4.14}$$

where we have used the estimate

$$\epsilon_1 = \|X_\star - X_1\|_2 = \|A^T(Y_\star - Y_0)A\|_2 \leq \|A\|_2^2\|Y_\star - Y_0\|_2 = \omega^2\upsilon.$$

In addition, it easily follows from (4.13) and (4.14) that the sequence $\{\chi_k\}_{k=1}^{\infty}$ is monotonically decreasing and satisfies

$$\left\{\begin{array}{l} \chi_1 \leq (\alpha + \beta\omega^2)\upsilon, \\ \chi_{k+1} \leq \left(1 + \frac{\alpha}{\beta}\right)(\beta\chi_k + \psi^2)\chi_k, \end{array}\right. \qquad k = 1, 2, \cdots.$$

With successive recursion, by making use of (4.10) we have $\chi_k \leq \alpha\upsilon + \beta\epsilon$, and

$$\begin{aligned}
\chi_{k+1} \leq & \left(1 + \frac{\alpha}{\beta}\right)(\beta\chi_k + \psi^2)\chi_k \\
\leq & \left(1 + \frac{\alpha}{\beta}\right)\big((\alpha\upsilon + \beta\epsilon)\beta + \psi^2\big)\chi_k \\
= & \varrho\chi_k \leq \varrho^{k-1}\chi_1 \leq \varrho^{k-1}(\alpha + \beta\omega^2)\upsilon, \qquad k = 1, 2, \cdots.
\end{aligned}$$

So, by making use of (4.13) again we then obtain the estimate (4.3).

The verification of (iii) is based on the matrix estimate (4.8) and (4.9). By rewriting (4.8) we get

$$\begin{aligned}
X_\star - X_{k+1} \preceq & A^T(I + X_\star R)^{-1}(X_\star - X_k)Q^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A \\
& + A^T(I + X_\star R)^{-1}(X_\star - X_k)(I + RX_\star)^{-1}A + \widehat{E}_k,
\end{aligned} \tag{4.15}$$

where

$$\widehat{E}_k = \widetilde{E}_k + A^T(Y_\star - Y_{k-1})(X_\star^{-1} + R)(Y_\star - Y_{k-1})A,$$

with $\widetilde{E}_k$ being defined by (4.9). Evidently, $\widehat{E}_k$ is a symmetric matrix due to the symmetry of $\widetilde{E}_k$. It follows from (4.12) that

$$\|\widehat{E}_k\|_2 \le \omega\tau(\omega\tau_\star\upsilon + 2æ)\upsilon_{k-1}\epsilon_k + \omega^2(\tau_\star + \gamma)\upsilon_{k-1}^2.$$

Recalling that the matrix sequences $\{X_k\}_{k=0}^\infty$ and $\{Y_k\}_{k=0}^\infty$ are convergent to $X_\star$ and $Y_\star$, respectively, we know that for any $\epsilon_o > 0$ there exists a positive integer k_o such that

$$0 \preceq X_\star - X_k \preceq \epsilon_o I \quad \text{and} \quad 0 \preceq Y_\star - Y_{k-1} \preceq \epsilon_o I$$

hold for all $k = k_o, k_o + 1, \cdots$. Hence, we have

$$(X_\star - X_k)^{1/2} Q^{-1} (X_\star - X_k)^{1/2} \preceq \epsilon_o \|Q^{-1}\|_2 I = \epsilon_o \tau I, \qquad k = k_o, k_o + 1, \cdots, \tag{4.16}$$

$$\widehat{E}_k \preceq \big(\omega\tau(\omega\tau_\star\upsilon + 2æ) + \omega^2(\tau_\star + \gamma)\big)\epsilon_o^2 I = c_o\epsilon_o^2 I, \tag{4.17}$$

where $c_o := \omega\tau(\omega\tau_\star\upsilon + 2æ) + \omega^2(\tau_\star + \gamma)$. After substituting (4.16) and (4.17) into (4.15), we immediately obtain that

$$\begin{aligned} 0 \preceq & X_\star - X_{k+1} \\ \preceq & (1 + \epsilon_o\tau) A^T (I + X_\star R)^{-1} (X_\star - X_k)(I + RX_\star)^{-1} A + c_o\epsilon_o^2 I \\ \preceq & (1 + \epsilon_o\tau)^{k-k_o+1} (A^T (I + X_\star R)^{-1})^{k-k_o+1} (X_\star - X_{k_o}) \big((I + RX_\star)^{-1} A\big)^{k-k_o+1} + c_o\epsilon_o^2 M_k \end{aligned}$$

hold for all $k = k_o, k_o + 1, \cdots$, where

$$M_k := \sum_{j=0}^{k-k_o} (1 + \epsilon_o\tau)^j \big(A^T (I + X_\star R)^{-1}\big)^j \big((I + RX_\star)^{-1} A\big)^j.$$

Because $\rho((I + RX_\star)^{-1} A) < 1$, from [39] we know that there exists a compatible matrix norm, say $|||\cdot|||$, such that

$$\widehat{æ} := \max\Big\{ \big|\big|\big|(I + RX_\star)^{-1} A\big|\big|\big|, \quad \big|\big|\big|(A^T (I + X_\star R)^{-1})\big|\big|\big| \Big\} < 1$$

holds true. Hence,

$$\begin{aligned} \big|\big|\big|M_k\big|\big|\big| &\le \sum_{j=0}^{k-k_o} (1 + \epsilon_o\tau)^j \big|\big|\big|A^T (I + X_\star R)^{-1}\big|\big|\big|^j \big|\big|\big|(I + RX_\star)^{-1} A\big|\big|\big|^j \\ &\le \sum_{j=0}^{k-k_o} (1 + \epsilon_o\tau)^j \widehat{æ}^{2j}. \end{aligned}$$

If we further restrict ϵ_o so small that $(1 + \epsilon_o\tau)\widehat{æ}^2 < 1$, then it holds that

$$\big|\big|\big|M_k\big|\big|\big| \le \frac{1}{1 - (1 + \epsilon_o\tau)\widehat{æ}^2},$$

which implies that the matrix sequence $\{M_k\}_{k=k_o}^\infty$ is uniformly bounded with respect to k. So, according to the equivalence of the matrix norms there exists a positive constant μ such that

$$\|M_k\|_2 \le \mu, \quad k = k_o, k_o + 1, \cdots.$$

It follows from Lemma 2.1(vii) that

$$\begin{aligned}\|X_\star - X_k\|_2 \le& (1+\epsilon_o\tau)^{k-k_o}\|(A^T(I+X_\star R)^{-1})^{k-k_o}\|_2\ \|X_\star - X_{k_o}\|_2\ \|((I+RX_\star)^{-1}A)^{k-k_o}\|_2 + c_o\mu\epsilon_o^2 \\ \le& (1+\epsilon_o\tau)^{k-k_o}\|((I+RX_\star)^{-1}A)^{k-k_o}\|_2^2\ \|X_\star - X_{k_o}\|_2 + c_o\mu\epsilon_o^2\end{aligned}$$

hold for $k = k_o, k_o+1, \cdots$, and, therefore,

$$\lim_{k\to\infty} \sup \sqrt[k]{\|X_\star - X_k\|} \le (\rho((I+RX_\star)^{-1}A))^2.$$

Here, we have also used the facts that ϵ_o is an arbitrary small positive quantity and that $\lim_{k\to\infty} \|Z^k\|^{1/k} = \rho(Z)$ holds for any square matrix Z in any matrix norm. □

A practical implementation strategy about the modified fixed-point iteration (4.1) may be as follows:

1. compute the Cholesky factorization of Q to obtain $Q = L_Q L_Q^T$ by employing a numerically stable algorithm (see, e.g., [10]), form the matrix $\widehat{R}_Q = I + L_Q^T R L_Q$, compute the Cholesky factorization of $\widehat{R}_Q$ to obtain $\widehat{R}_Q = \widehat{L}_R \widehat{L}_R^T$, solve $\widehat{Y}_0$ from $\widehat{L}_R \widehat{Y}_0 = L_Q$, and form the matrix $Y_0 = \widehat{Y}_0^T \widehat{Y}_0$;
2. compute $X_{k+1} = A^T Y_k A + Q$;
3. compute the Cholesky factorization of X_{k+1} to obtain $X_{k+1} = L_{k+1} L_{k+1}^T$ by employing a numerically stable algorithm;
4. form the matrix $\widehat{R}_{k+1} = I + L_{k+1}^T R L_{k+1}$;
5. solve $\widehat{Y}_k$ from $L_{k+1}\widehat{Y}_k = Y_k$;
6. compute $Y_{k+1} = 2Y_k - \widehat{Y}_k^T \widehat{R}_{k+1} \widehat{Y}_k$ by only computing the entries of the upper- or the lower-triangular part of the right-hand-side matrix according to its symmetry.

When the matrix Q is singular, similarly to the treatment to the fixed-point iteration (1.5), we can define a symmetric positive definite matrix $\widehat{Q}$ based on the spectral decomposition of Q and choose the starting matrix $Y_0 = (\widehat{Q}^{-1} + R)^{-1}$. Then, it follows from $\widehat{Q} \preceq X_\star$ that $Y_0 \preceq (X_\star^{-1} + R)^{-1}$.

When the matrix R is symmetric positive definite, we can analogously show that the iteration sequence $\{X_k\}_{k=0}^{\infty}$ generated by the modified fixed-point iteration (4.1), starting from $Y_0 = R^{-1}$, is monotonically decreasing and convergent to $X_\star$ R-linearly.

5. Numerical Examples

In this section, we use several examples to examine the accuracy of the eigenvalue bounds about the symmetric positive definite solution of the DARE (1.3) given in Section 2, and also show the effectiveness of the modified fixed-point iteration (4.1).

In actual implementations, we terminate all iteration schemes once the current iterates X_k satisfy

$$\|X_\star - X_k\|_\infty \le 10^{-8}$$

if the exact solution $X_\star$ of the DARE (1.3) is known, or

$$\|X_{k+1} - X_k\|_\infty \leq 10^{-8}$$

otherwise. All codes were written in Fortran 90 with double precision and run on a Pentium IV personal computer.

Example 5.1. Consider the DARE (1.1) with

$$A = \begin{pmatrix} 0 & 0 \\ 0.5 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 \end{pmatrix} \quad \text{and} \quad G = I_2,$$

where I_2 is the identity matrix. Then its unique symmetric positive definite solution is

$$X_\star = \begin{pmatrix} 0.25 & 0 \\ 0 & 1 \end{pmatrix}.$$

By straightforward computations, we obtain $R = \mathrm{diag}(0.25, 0)$ and $Q = \mathrm{diag}(0, 1)$. Evidently, both R and Q are singular matrices. In addition, we have

$$\lambda_1(X_\star) = 1, \quad \lambda_2(X_\star) = 0.25, \quad \mathrm{tr}(X_\star) = 1.25 \quad \text{and} \quad \det(X_\star) = 0.25.$$

By applying Theorems 2.1 and 2.2, and Corollaries 2.1, 2.2 and 2.5, we know that

$$1 \leq \lambda_1(X_\star) \leq 1.3333, \quad 1 \leq \mathrm{tr}(X_\star) \leq 1.5901 \quad \text{and} \quad \det(X_\star) \leq 0.2650.$$

Obviously, these estimated bounds are very close to the actual ones.

Because the matrices Q and R are singular, the lower and the upper bounds in [9] about the maximal eigenvalue, the lower bound in [9] about partial sum of the eigenvalues, and the upper bounds in [17] about individual eigenvalues, partial sum and partial product of the eigenvalues, and the lower bound in [42] about determinant of the solution for the DARE (1.3) are failed. In addition, the upper bounds about the solution given in [18,26–28,30] cannot be applied, too. We have noticed that the lower bounds in [19, 22] about trace of the solution are 0.5 and 0, respectively, which are much rougher than ours.

As the matrix Q is singular, we define $\widehat{Q} = \mathrm{diag}(10^{-5}, 1)$. Then we choose the starting matrices X_0 and Y_0 in the iterations (1.5) and (4.1) as

$$X_0 = \widehat{Q} \quad \text{and} \quad Y_0 = (\widehat{Q}^{-1} + R)^{-1},$$

respectively. After 2 steps of the iterations (1.5) and (4.1), we obtain $X_2 = \mathrm{diag}(0.25, 1)$. If we define

$$\widehat{Q} = \mathrm{diag}(10^{-4}, 1), \quad \mathrm{diag}(10^{-3}, 1), \quad \mathrm{diag}(10^{-2}, 1) \quad \text{and} \quad \mathrm{diag}(10^{-1}, 1),$$

respectively, then the iterations (1.5) and (4.1) also yield the same result after 2 steps.

We remark that $\rho((I + RX_\star)^{-1}A) = 0$ holds for this example. Hence, both iterations (1.5) and (4.1) converge to the exact solution of the DARE (1.3) R-superlinearly.

Example 5.2. ([27]) Consider the DARE (1.3) with

$$A = \begin{pmatrix} 0.4 & 0.2 & 0.2 \\ -0.6 & 0 & 0.1 \\ 0 & 0 & 0.1 \end{pmatrix}, \quad Q = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Then its unique symmetric positive definite solution is

$$X_\star = \begin{pmatrix} 3.6590085409 & 1.0407861936 & 0.9379715209 \\ 1.0407861936 & 2.0480405499 & 0.0439300472 \\ 0.9379715209 & 0.0439300472 & 2.0623919675 \end{pmatrix}.$$

Evidently, R is a singular matrix. In addition, we have

$$\lambda_1(X_\star) = 4.4824, \quad \lambda_2(X_\star) = 2.0124, \quad \lambda_3(X_\star) = 1.2747 \quad \text{and} \quad \mathrm{tr}(X_\star) = 7.7694.$$

By applying Theorems 2.1 and 2.2, and Corollaries 2.1 and 2.2, we know that

$$4 \le \lambda_1(X_\star) \le 8.5903, \quad 3.2110 \le \lambda_1(X_\star) + \lambda_2(X_\star) \le 12.8855, \quad 7.2125 \le \mathrm{tr}(X_\star) \le 59.2791.$$

Obviously, these estimated bounds are very close to the actual ones.

Because the matrix R is singular, the lower and the upper bounds in [9] about the maximal eigenvalue, and the upper bounds in [17] about individual eigenvalues, partial sum and partial product of the eigenvalues of the solution for the DARE (1.3) are failed. In addition, the upper bounds about the solution given in [18,26,30] cannot be applied, too. We have noticed that the lower bounds in [19, 22] about trace of the solution are 6.5338 and 3.435, respectively, which are much rougher than ours.

As the matrix Q is nonsingular, we choose the starting matrices X_0 and Y_0 in the iterations (1.5) and (4.1) as $X_0 = Q$ and $Y_0 = (Q^{-1} + R)^{-1}$, respectively. After 8 steps of the iterations (1.5) and (4.1), we obtain

$$X_8 = \begin{pmatrix} 3.65900854086 & 1.04078619363 & 0.93797152094 \\ 1.04078619363 & 2.04804054987 & 0.04393004718 \\ 0.93797152094 & 0.04393004718 & 2.06239196746 \end{pmatrix},$$

$$X_8 = \begin{pmatrix} 3.65900854028 & 1.04078619344 & 0.93797152087 \\ 1.04078619344 & 2.04804054979 & 0.04393004713 \\ 0.93797152087 & 0.04393004713 & 2.06239196743 \end{pmatrix},$$

respectively.

We remark that $\rho((I + RX_\star)^{-1}A) = 0.2321$ holds for this example. Hence, both iterations (1.5) and (4.1) converge to the exact solution of the DARE (1.3) R-linearly.

Example 5.3. ([5, 40]) Consider the DARE (1.3) with

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & & & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad R = BB^T \quad \text{and} \quad Q = I_n,$$

where I_n is the identity matrix. Then its stabilized symmetric positive definite solution is

$$X_\star = \mathrm{diag}(1, 2, \cdots, n).$$

As the matrix Q is nonsingular, we choose the starting matrices X_0 and Y_0 in the iterations (1.5) and (4.1) as $X_0 = Q$ and $Y_0 = (Q^{-1} + R)^{-1}$, respectively. For $n = 100$, after 100 and

106 steps of the iterations (1.5) and (4.1), we obtain the exact solution $X_\star$ of the DARE (1.3), respectively. And for $n = 1000$, after 1000 and 1011 steps of the iterations (1.5) and (4.1), we obtain the exact solution $X_\star$, respectively.

We remark that $\rho((I + RX_\star)^{-1}A) = 0$ holds for this example. Hence, both iterations (1.5) and (4.1) converge to the exact solution of the DARE (1.3) R-superlinearly.

6. Concluding Remarks

The lower and the upper bounds for partial sum about the eigenvalues, the upper bounds for partial product about the eigenvalues, and the upper bounds for trace and determinant of the solution for the DARE (1.3) are derived. Examples show that these bounds are much sharper than the existing ones. However, the condition $\sigma_1(A) < 1$ may restrict applications of these upper bounds. A possible remedy is to use a nonsingular and symmetric matrix D to congruently transform the DARE (1.3), so that the obtained discrete algebraic Riccati equation

$$\tilde{X} = \tilde{A}^T(\tilde{X}^{-1} + \tilde{R})^{-1}\tilde{A} + \tilde{Q},$$

with

$$\tilde{X} = DAD, \quad \tilde{A} = D^{-1}AD, \quad \tilde{Q} = DQD \quad \text{and} \quad \tilde{R} = D^{-1}RD^{-1},$$

satisfies $\sigma_1(\tilde{A}) < 1$. For example, in Example 5.3 if we choose $D = \operatorname{diag}\left(1, 2^{-1}, \cdots, 2^{1-n}\right)$, then it holds that $\sigma_1(\tilde{A}) = 0.5$.

By technically incorporating the Schulz iteration into the fixed-point iteration (1.5), we have established a modified fixed-point iteration (4.1) for computing the unique symmetric positive definite solution of the DARE (1.3). This iteration scheme is convergent monotonically and R-linearly. Of course, the Schulz iteration may be applied to the modified fixed-point iteration (4.1) again so that an inversion-free variant of the iteration scheme (1.5) can be established, which completely avoids computing the matrix inversion. The establishment of such a fixed-point iteration scheme and the analysis of its monotone convergence property should be an interesting problem to be discussed in future.

Acknowledgments. This work was started when the first author was visiting State Key Laboratory of Scientific/Engineering Computing, Chinese Academy of Sciences, during March-May in 2008. The support and hospitality from LSEC are very much appreciated. Supported by The National Basic Research Program (No. 2005CB321702), The China Outstanding Young Scientist Foundation (No. 10525102), and The National Natural Science Foundation for Innovative Research Groups (No. 11021101), P.R. China.

References

[1] B.D.O. Anderson, Second-order convergent algorithms for the steady-state Riccati equation, *Intern. J. Control*, 28 (1978), pp. 295-306.

[2] B.D.O. Anderson and J.B. Moore, Optimal Filtering, *Prentice-Hall*, Englewood Cliffs, New Jersey, 1979.

[3] W.F. Arnold III and A.J. Laub, Generalized eigenproblem algorithms and software for algebraic Riccati equations, *Proc. IEEE*, 72 (1984), pp. 1746-1754.

[4] A.Y. Barraud, A numerical algorithm to solve $A^TXA - X = Q$, *IEEE Trans. Automat. Control*, AC-22 (1977), pp. 883-885.

[5] P. Benner, A.J. Laub and V. Mehrmann, A Collection of Benchmark Examples for the Numerical Solution of Algebraic Riccati Equations II: Discrete-Time Case, *Technical Report SPC 95-23*, Faculty of Mathematics, TU Chemnitz-Zwickau, D-09107 Chemnitz, December 1995.

[6] P. Benner, V. Mehrmann and H.-G. Xu, A numerically stable, structure preserving method for computing the eigenvalues of real Hamiltonian or symplectic pencils, *Numer. Math.*, 78 (1998), pp. 329-358.

[7] H. Dai, The Theory of Matrices, Science Press, Beijing, 2001.

[8] R. Davies, P. Shi and R. Wiltshire, New upper solution bounds of the discrete algebraic Riccati matrix equation, *J. Comput. Appl. Math.*, 213 (2008), pp. 307-315.

[9] J. Garloff, Bounds for the eigenvalues of the solution of the discrete Riccati and Lyapunov equations and the continuous Lyapunov equation, *Intern. J. Control*, 43 (1986), pp. 423-431.

[10] G.H. Golub and C.F. Van Loan, Matrix Computations, 3rd Edition, The Johns Hopkins University Press, Baltimore and London, MD, 1996.

[11] 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.*, 20 (1998), pp. 279-294.

[12] T.-M. Hwang, E.K.-W. Chu and W.-W. Lin, A generalized structure-preserving doubling algorithm for generalized discrete-time algebraic Riccati equations, *Intern. J. Control*, 78 (2005), pp. 1063-1075.

[13] S.W. Kim, P. Park and W.H. Kwon, Lower bounds for the trace of the solution of the discrete algebraic Riccati equation, *IEEE Trans. Automat. Control*, 38 (1993), pp. 312-314.

[14] M. Kimura, Convergence of the doubling algorithm for the discrete-time algebraic Riccati equation, *Intern. J. Systems Sci.*, 19 (1988), pp. 701-711.

[15] G. Kitagawa, An algorithm for solving the matrix equation $X = FXF^T + S$, *Intern. J. Control*, 25 (1977), pp. 745-753.

[16] N. Komaroff, Upper bounds for the eigenvalues of the solution of the Lyapunov matrix equation, *IEEE Trans. Automat. Control*, 35 (1990), pp. 737-739.

[17] N. Komaroff, Upper bounds for the solution of the discrete Riccati equation, *IEEE Trans. Automat. Control*, 37 (1992), pp. 1370-1372.

[18] N. Komaroff, Iterative matrix bounds and computational solutions to the discrete algebraic Riccati equation, *IEEE Trans. Automat. Control*, 39 (1994), pp. 1676-1678.

[19] N. Komaroff and B. Shahian, Lower summation bounds for the discrete Riccati and Lyapunov equations, *IEEE Trans. Automat. Control*, 37 (1992), pp. 1078-1080.

[20] V.S. Kouikoglou and Y.A. Phillis, Trace bounds on the covariances of continuous-time systems with multiplicative noise, *IEEE Trans. Automat. Control*, 38 (1993), pp. 138-142.

[21] H. Kwakernaak and R. Sivan, Linear Optimal Control Systems, John Wiley & Sons, New York, 1972.

[22] B.H. Kwon, M.J. Youn and Z. Bien, On bounds of the Riccati and Lyapunov matrix equations, *IEEE Trans. Automat. Control*, AC-30 (1985), pp. 1134-1135.

[23] W.H. Kwon, Y.S. Moon and S.C. Ahn, Bounds in algebraic Riccati and Lyapunov equations: a survey and some new results, *Intern. J. Control*, 64 (1996), pp. 377-389.

[24] P. Lancaster and L. Rodman, Algebraic Riccati Equations, The Clarendon Press, Oxford and New York, 1995.

[25] A.J. Laub, A Schur method for solving algebraic Riccati equations, *IEEE Trans. Automat. Control*, AC-24 (1979), 913-921.

[26] C.-H. Lee, On the matrix bounds for the solution matrix of the discrete algebraic Riccati equation, *IEEE Trans. Circuits Systems-I: Fundamental Theory Appl.*, 43 (1996), pp. 402-407.

[27] C.-H. Lee, Upper matrix bound of the solution for the discrete Riccati equation, *IEEE Trans. Automat. Control*, 42 (1997), pp. 840-842.

[28] C.-H. Lee, Upper and lower bounds of the solutions of the discrete algebraic Riccati and Lyapunov matrix equations, *Intern. J. Control*, 68 (1997), pp. 579-598.

[29] C.-H. Lee, Simple stabilizability criteria and memoryless state feedback control design for time-delay systems with time-varying perturbations, *IEEE Trans. Circuits Systems-I: Fundamental Theory Appl.*, 45 (1998), pp. 1211-1215.

[30] C.-H. Lee, Matrix bounds of the solutions of the continuous and discrete Riccati equations - a unified approach, *Intern. J. Control*, 76 (2003), pp. 635-642.

[31] W.-W. Lin and C.-S. Wang, On computing stable Lagrangian subspaces of Hamiltonian matrices and symplectic pencils, *SIAM J. Matrix Anal. Appl.*, 18 (1997), pp. 590-614.

[32] W.-W. Lin and S.-F. Xu, Convergence analysis of structure-preserving doubling algorithms for Riccati-type matrix equations, *SIAM J. Matrix Anal. Appl.*, 28 (2006), pp. 26-39.

[33] L.-Z. Lu and W.-W. Lin, An iterative algorithm for the solution of the discrete-time algebraic Riccati equation, *Linear Algebra Appl.*, 188/189 (1993), pp. 465-488.

[34] L.-Z. Lu, W.-W. Lin, and C.E.M. Pearce, An efficient algorithm for the discrete-time algebraic Riccati equation, *IEEE Trans. Automat. Control*, 44 (1999), pp. 1216-1220.

[35] A.W. Marshall and I. Olkin, Inequalities: Theory of Majorization and Its Applications, Academic Press, New York and London, 1979.

[36] V. Mehrmann, The Autonomous Linear Quadratic Control Problem: Theory and Numerical Solution, in Lecture Notes in Control and Information Sciences 163, Springer-Verlag, Berlin, 1991.

[37] T. Mori and I.A. Derese, A brief summary of the bounds on the solution of the algebraic matrix equations in control theory, *Intern. J. Control*, 39 (1984), pp. 247-256.

[38] T. Mori, N. Fukuma and M. Kuwahara, On the discrete Riccati equation, *IEEE Trans. Automat. Control*, AC-32 (1987), pp. 828-829.

[39] J.M. Ortega and W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, SIAM, Philadelphia, PA, 2000.

[40] T. Pappas, A.J. Laub and N.R. Sandell, On the numerical solution of the discrete-time algebraic Riccati equation, *IEEE Trans. Automat. Control*, AC-25 (1980), pp. 631-641.

[41] G. Schulz, Iterative berechnung der reziproken matrix, *Z. Angew. Math. Mech.*, 13 (1933), pp. 57-59.

[42] M.T. Tran and M.E. Sawan, On the discrete Riccati matrix equation, *SIAM J. Alg. Disc. Meth.*, 6 (1985), pp. 107-108.

[43] S.-S. Wang, B.-S. Chen and T.-P. Lin, Robust stability of uncertain time-delay systems, *Intern. J. Control*, 46 (1987), pp. 963-976.

[44] W.M. Wonham, Linear Multivariable Control: A Geometric Approach, 2nd Edition, Springer-Verlag, New York and Berlin, 1979.

[45] K. Yasuda and K. Hirai, Upper and lower bounds on the solution of the algebraic Riccati equation, *IEEE Trans. Automat. Control*, AC-24 (1979), pp. 483-487.

[46] X.-Z. Zhan, Computing the extremal positive definite solutions of a matrix equation, *SIAM J. Sci. Comput.*, 17 (1996), pp. 1167-1174.