Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 413 (2006) 474–494
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

On the solution of algebraic Riccati equations arising in fluid queues

Dario A. Bini [a,*], Bruno Iannazzo [a], Guy Latouche [b], Beatrice Meini [a]

[a] *Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo 5, 56127 Pisa, Italy*
[b] *Départment d'Informatique, U.L.B., C.P. 212, Boulevard du Triomphe, 1050 Bruxelles, Belgium*

Received 21 December 2004; accepted 19 April 2005
Available online 1 July 2005
Submitted by G. de Oliveira

Abstract

New algorithms for solving algebraic Riccati equations (ARE) which arise in fluid queues models are introduced. They are based on reducing the ARE to a unilateral quadratic matrix equation of the kind $AX^2 + BX + C = 0$ and on applying the Cayley transform in order to arrive at a suitable spectral splitting of the associated matrix polynomial. A shifting technique for removing unwanted eigenvalues of modulus 1 is complemented with a suitable parametrization of the matrix equation in order to arrive at fast and numerically reliable solvers based on quadratically convergent iterations like logarithmic reduction and cyclic reduction. Numerical experiments confirm the very good performance of these algorithms.

Keywords: Algebraic Riccati equations; Quadratic matrix equations; Fluid queues; Cyclic reduction; Cayley transform

1. Introduction

Let m, n be positive integers and let $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{n\times n}$, be such that the matrix

* Corresponding author.
E-mail address: bini@dm.unipi.it (D.A. Bini).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2005.04.019

$$M = \begin{bmatrix} D & -C \\ -B & A \end{bmatrix} \tag{1}$$

is an irreducible singular M-matrix. This means that $M = \alpha I - N$ where N has nonnegative entries and $\alpha = \rho(N)$, where $\rho(N)$ is the spectral radius of N. Under this assumption, D and A are M-matrices and C and B are nonnegative matrices.

In the analysis of two-dimensional continuous-time Markov processes, called fluid queues, a crucial step is to compute the element-wise minimal nonnegative solution $S \in \mathbb{R}^{m\times n}$ of the nonsymmetric algebraic Riccati matrix equation

$$XCX - AX - XD + B = 0. \tag{2}$$

This solution provides important information about the state probabilities of the model. To this regard, we refer the reader to the papers [1–4]; the existence of the minimal solution S is established in [5] and it is shown in [6] that $S > 0$ under the irreducibility assumption on the matrix M.

Several methods have been designed for computing S, in particular in the papers [7,5,8,4]. In [4] Ramaswami has observed that, under suitable conditions, S can be obtained as a submatrix of the minimal nonnegative solution of the unilateral quadratic matrix equation

$$A_2Y^2 + A_1Y + A_0 = 0, \tag{3}$$

where, for a suitable parameter t,

$$A_2 = \begin{bmatrix} \frac{1}{2}I & 0 \\ 0 & 0 \end{bmatrix}, \quad A_1 = \begin{bmatrix} -\frac{1}{2}(I + tA) & 0 \\ tC & -I \end{bmatrix}, \quad A_0 = \begin{bmatrix} 0 & \frac{1}{2}tB \\ 0 & -tD + I \end{bmatrix}.$$

In this way, one can use well known iterative techniques specially designed for (3).

The reduction proposed by Ramaswami is not the only possible one, but it has the advantage of splitting into two sets the roots of the matrix polynomial

$$A(z) = A_2z^2 + A_1z + A_0$$

associated with the matrix equation (3): a set made up by $n + m$ roots inside the closed unit disk of which m are equal to zero, and a set of $n + m$ roots outside the open unit disk of which n are equal to infinity. We refer the reader to [9] concerning properties of matrix polynomials. In the case where the $2(m + n)$ roots of the matrix polynomial $A(z)$ can be split into two sets, a set made up by $m + n$ roots inside the closed unit disk and a set of $m + n$ roots outside the open unit disk, we say that $A(z)$ has a *splitting with respect to the unit circle*. This property allows one to solve the unilateral matrix equation (3) by means of the most efficient algorithms like logarithmic reduction (LR) [10] and cyclic reduction (CR) [11]. The convergence of these two algorithms is generally quadratic. More precisely, denoting by $\xi \leqslant 1$ the maximum ratio of the moduli of the $m + n$ roots in the closed unit disk and the $m + n$ roots outside the open unit disk (*isolation ratio*), the approximation error after i steps

is bounded by ξ^{2^i} if $\xi < 1$. If $\xi = 1$ the convergence is expected to be linear with rate 1/2 [12].

Concerning spectral properties, Guo [5] has shown that the eigenvalues of the matrix

$$H = \begin{bmatrix} -D & C \\ -B & A \end{bmatrix}, \tag{4}$$

are *split with respect to the imaginary axis*, i.e., m eigenvalues have nonnegative real part, and n eigenvalues have nonpositive real part. The matrix H is very important in the classical theory of algebraic Riccati equations [13], in fact all the solutions of the equation (2) can, in principle, be recovered from the invariant subspaces of H.

In this paper we present a different reduction of the nonsymmetric algebraic Riccati equation to a unilateral quadratic equation, whose associated matrix polynomial has a set of roots which is the union of the set of the eigenvalues of H, and of the set made up by m null roots and by n roots at infinity. Indeed, the roots of this matrix polynomial different from zero and infinity have a splitting with respect to the imaginary axis. To achieve the splitting with respect to the unit circle needed for the convergence of LR and CR, we apply the Cayley transform to our unilateral matrix equation. After this transformation, the roots at 0 and at infinity of the original matrix polynomial are mapped to 1 and to -1, respectively, so that the isolation ratio of this splitting is $\xi = 1$. For reducing the isolation ratio, we apply the technique of the selective shift of the spectrum introduced in [14] and generalized in [15,16], in order to remove the two unwanted multiple roots from 1 and -1 to zero and to infinity, respectively. In this way, the new equation obtained after the shift has roots with isolation ratio $\xi < 1$ provided that the matrix H has no pure imaginary eigenvalues so that the application of cyclic reduction or of logarithmic reduction yields a quadratically convergent algorithm. The combination of the Cayley transform with the shift technique is complemented with the use of a suitable parameter introduced for increasing the numerical performance of the computation.

In certain problems typically encountered in Markov chains, the matrix H has pure imaginary eigenvalues so that the matrix polynomial obtained after this transformation has still one or more roots on the unit circle. In this case it is possible to apply a second shift stage in order to improve the isolation ratio. The use of this *double shift technique* provides quadratic convergence of CR and LR even in the critical cases where Newton's iteration converges linearly.

The algorithm has been implemented in the CR version both with the simple shift and with the double shift technique, and compared with Newton's iteration of [8] and with the algorithm of [4] on the set of test problems of [5] and [1]. The computational cost per step amounts to $34n^3 + \mathrm{O}(n^2)$ arithmetic operations (ops) where we assume for simplicity $m = n$ whereas the cost of Newton's iteration amounts to $62n^3$ ops [8]. The number of steps needed by our algorithm is generally inferior to the number of steps required by Newton's iteration. In the null recurrent problems, where H has two null eigenvalues, Newton's iteration has a linear convergence while CR still keeps

a quadratic convergence if applied with the double shift. The approximation errors provided by our method are generally inferior to the errors of the approximation provided by Newton's iteration.

The paper is organized as follows. After introducing some preliminary tools in Section 2, we describe in Section 3 the reduction of an ARE to a unilateral quadratic matrix equation. In Section 4 we introduce the main ideas, that is, the Cayley transform and the shift technique, in order to arrive at a splitting with respect to the unit circle. In Section 5 we deal with the computational details, in particular we recall the algorithm of cyclic reduction, introduce the parametrization technique (scaling of the equation) analyze the structure of the matrices generated by CR, describe the algorithm and perform its complexity analysis. In Section 6 we report the results of our numerical experiments.

2. Preliminaries

We denote by $\mathscr{D}_<$ [resp. $\mathscr{D}_\leqslant$] the open [resp. closed] unit disk, and by $\mathscr{D}_>$ [resp. $\mathscr{D}_\geqslant$] the set of complex numbers outside the closed [resp. open] unit disk and the point at infinity. Similarly we denote by $\mathbb{C}_<$ [resp. $\mathbb{C}_>$] the open left [resp. right] half plane of the complex plane and by $\mathbb{C}_\leqslant$, $\mathbb{C}_\geqslant$ the closed ones, where we assume that the point at infinity belongs both to $\mathbb{C}_\leqslant$ and $\mathbb{C}_\geqslant$. We also denote by $\sigma(A)$ the set of the eigenvalues of the square matrix A and by $\boldsymbol{e} = (1, \ldots, 1)^{\mathrm{T}}$ the vector with all components 1. For a matrix polynomial $A(z)$ we call *roots* of $A(z)$ the zeros of $\det A(z)$.

We are looking for a nonnegative solution S of the nonsymmetric algebraic Riccati equation (2) that is minimal, in the sense of element-wise order. Any solution of this kind is called *minimal solution*, whereas we use the expression *spectral minimal solution* to denote a solution having minimal spectral radius.

In the case where the matrix M of (1) is an irreducible singular M-matrix, it was shown by Guo [5, Theorem 3.1] that the algebraic Riccati equation has a nonnegative solution S such that $D - CS$ is an M-matrix, and then $\sigma(-D + CS) \subset \mathbb{C}_\leqslant$. Recall that an M-matrix has only eigenvalues in $\mathbb{C}_\geqslant$. [17]. It is noted in [5] that this solution S is minimal.

An almost general situation encountered in fluid queues models is the case where $M\boldsymbol{e} = 0$, in fact the infinitesimal generator of a continuous-time Markov chain has this property [4,2].

In a more general setting the following result holds (see the discussions in [5], particularly Theorem 4.7 and [6, Theorem 5]).

Theorem 1. *For the irreducible singular M-matrix M, the matrix H has two real eigenvalues 0 and λ^*, along with $n - 1$ eigenvalues in $\mathbb{C}_<$ and $m - 1$ eigenvalues in $\mathbb{C}_>$. Moreover, $D - CS$ is an irreducible M-matrix which is singular if $\lambda^* \geqslant 0$ and is nonsingular if $\lambda^* < 0$.*

3. Reduction to a quadratic unilateral equation

Given Eq. (2) define the matrices

$$M_0 = \begin{bmatrix} 0 & B \\ 0 & -D \end{bmatrix}, \quad M_1 = \begin{bmatrix} -A & 0 \\ C & -I \end{bmatrix}, \quad M_2 = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \tag{5}$$

and consider the matrix polynomial $F(z) = z^2 M_2 + z M_1 + M_0$.

Theorem 2. *The matrix polynomial $F(z)$ has m roots equal to 0, n roots at infinity, and $m + n$ roots which are the eigenvalues of the matrix H defined in (4).*

Proof. Since

$$F(z) = \begin{bmatrix} z^2 I - zA & B \\ zC & -zI - D \end{bmatrix} = \begin{bmatrix} -zI + A & B \\ -C & -zI - D \end{bmatrix} \begin{bmatrix} -zI & 0 \\ 0 & I \end{bmatrix}$$

then $\det F(z) = (-z)^m \det(K - zI)$ where $K = \begin{bmatrix} A & B \\ -C & -D \end{bmatrix}$. The proof follows since K is similar to H, and the dimension of $\ker(M_2)$ is n. □

The location in the complex plane of the eigenvalues of H is fully characterized by Theorem 1. In particular, we may split the roots of $F(z)$ into the two sets $\Lambda_- \subset \mathbb{C}_< \cup \{0\}$ and $\Lambda_+ \subset \mathbb{C}_> \cup \{0, \infty\}$, both of cardinality $m + n$, given by:

(1) if $\lambda^* < 0$

$\Lambda_- = \{\lambda \in \sigma(H) \cap \mathbb{C}_<\} \cup \{0 \text{ with multiplicity } m\} \cup \{\lambda^*\}$,
$\Lambda_+ = \{\lambda \in \sigma(H) \cap \mathbb{C}_>\} \cup \{\infty \text{ with multiplicity } n\} \cup \{0\}$;

(2) if $\lambda^* \geqslant 0$

$\Lambda_- = \{\lambda \in \sigma(H) \cap \mathbb{C}_<\} \cup \{0 \text{ with multiplicity } m + 1\}$,
$\Lambda_+ = \{\lambda \in \sigma(H) \cap \mathbb{C}_>\} \cup \{\infty \text{ with multiplicity } n\} \cup \{\lambda^*\}$.

We may easily verify that if S is a solution of (2), then the matrix

$$G = \begin{bmatrix} 0 & S \\ 0 & V \end{bmatrix}, \tag{6}$$

where $V = -D + CS$, solves the unilateral equation

$$M_2 X^2 + M_1 X + M_0 = 0. \tag{7}$$

Therefore, from any solution S of the algebraic Riccati equation (2) we may construct a solution G of the unilateral equation. Conversely, if G is a solution of (7) having the block structure (6), then S solves the algebraic Riccati equation (2). Moreover, for the block upper triangular structure of G, G has as eigenvalues zero, with multiplicity m, and the n eigenvalues of V. Under our assumptions, if S is the minimal nonnegative solution of (2), then by Theorem 1 the matrix $-V = D - CS$ is an irreducible M-matrix which is singular if $\lambda^* \geqslant 0$, nonsingular if $\lambda^* < 0$. Since the eigenvalues

of a (singular irreducible) M-matrix belong to $\mathbb{C}_>(\mathbb{C}_> \cup \{0\})$ [17], and since the eigenvalues of G must be roots of $F(z)$ [9], we conclude that $\sigma(G) = \Lambda_-$, i.e., G is the solution whose eigenvalues are the "leftmost" roots of $F(z)$ in the complex plane. Therefore any algorithm for computing the solution G of (7) with "leftmost" eigenvalues, provides the minimal nonnegative solution S of (2).

It is useful to remark that the equation

$$XC_tX - A_tX - XD_t + B_t = 0, \tag{8}$$

where $C_t = tC$, $A_t = tA + I$, $D_t = tD - I$, $B_t = tB$, for $t \neq 0$, shares its solutions with the original equation (2). Applying the same technique we may reduce (8) to the unilateral equation

$$M_2X^2 + M_1(t)X + M_0(t) = 0, \tag{9}$$

where

$$M_0(t) = \begin{bmatrix} 0 & B_t \\ 0 & -D_t \end{bmatrix}, \quad M_1(t) = \begin{bmatrix} -A_t & 0 \\ C_t & -I \end{bmatrix}, \quad M_2 = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \tag{10}$$

whose solutions are

$$G_t = \begin{bmatrix} 0 & S \\ 0 & V_t \end{bmatrix},$$

where $V_t = -D_t + C_tS = I + t(-D + CS)$. Premultiplying (9) by the matrix

$$\begin{bmatrix} \frac{1}{2}I & 0 \\ 0 & I \end{bmatrix}$$

yields Eq. (3) found by Ramaswami [4]. In the case where $M = (m_{i,j})$ is an infinitesimal generator of a Markov chain, by using probabilistic arguments Ramaswami has shown that for $t \leqslant 1/\mu$, where $\mu = \max_{1\leqslant i\leqslant m+n} m_{ii}$ the matrix polynomial

$$F_t(z) = M_0(t) + zM_1(t) + z^2M_2 = \begin{bmatrix} z^2I - zA_t & B_t \\ zC_t & -zI - D_t \end{bmatrix}$$

has $m + n$ roots in $\mathscr{D}_\leqslant$ and $m + n$ roots in $\mathscr{D}_\geqslant$. In particular, for $t \leqslant 1/\mu$, V_t has eigenvalues inside the unit disk, and G_t is the solution with smallest spectral radius of (9). In the following remark we show that this property can be proved also by means of purely algebraic arguments.

Remark 3. Let $M = (m_{ij})$ be an irreducible singular M-matrix such that $Me = 0$, i.e., such that $\sum_j m_{ij} = 0$ for any i. Therefore $\sum_{j\neq i} |m_{ij}| = m_{ii}$. This means that the Gershgorin disks for the matrix M are all centered in m_{ii} and have radius m_{ii}, moreover the one corresponding to $\mu = \max_i m_{ii}$ contains all the remaining disks and therefore all the eigenvalues. By following the same arguments of the proof of Theorem 2, we may verify that $F_t(z)$ has m roots equal to 0, n roots at infinity, and $m + n$ roots which are the eigenvalues of the matrix $H_t = \begin{bmatrix} -D_t & C_t \\ -B_t & A_t \end{bmatrix}$. We observe

that the Gershgorin disks for the matrix H are the same as the ones of M or they are the disks of M reflected with respect to the imaginary axis, so the whole spectrum of H is contained in the two disks of center respectively μ and $-\mu$ and radius μ. Now, to transform this splitting of the eigenvalues of H with respect to the imaginary axis into a splitting of the eigenvalues of H_t with respect to the unit circle, it is enough to shrink the disks and translate the point 0 to 1. This can be obtained by choosing $t \leqslant 1/\mu$; in particular, with this choice of t, the condition $\sigma(-D_t + C_t S) \subset \mathscr{D}_{\leqslant}$ holds, therefore G_t is the solution with smallest spectral radius.

As pointed out in [1], the splitting of the roots of $F_t(z)$ with respect to the unit circle is a very important property, since it allows to apply efficient algorithms, like cyclic reduction (CR) or logarithmic reduction (LR), for computing the minimal solution G_t. For details on cyclic reduction (CR) or logarithmic reduction (LR) applied to solve quadratic matrix equations we refer the reader to [10,11,16].

Unfortunately, the Ramaswami trick to achieve the splitting with respect to the unit circle can be applied only if M is the infinitesimal generator of a Markov chain, i.e., M is a singular M-matrix such that $M\boldsymbol{e} = 0$. In the next section we will introduce a technique to transform the unilateral equation (7) into a new univariate equation, whose minimal spectral solution provides the desired matrix S, which works for a general singular irreducible M-matrix.

4. Transforming the unilateral quadratic equation

In this section, first we apply the Cayley transform to the matrix polynomial $F(z)$ and obtain a new polynomial $R(z)$ having a splitting of the roots with respect to the unit circle. Then we introduce the shift technique in order to shift to zero and to infinity the roots of $R(z)$ which belong to the unit circle. After this manipulation we obtain a new matrix polynomial $\psi(z)$ with $m + n$ roots of modulus less than 1 and $m + n$ roots with modulus greater than 1. The matrix equation associated with this polynomial has a spectral minimal solution which provides the minimal solution of the original algebraic Riccati equation. Due to the splitting of the spectrum, one can apply fast algorithms like cyclic reduction or logarithmic reduction for computing the solution of this univariate quadratic equation.

4.1. Cayley transform

The Cayley transform

$$z \to \frac{1+z}{1-z}, \tag{11}$$

defined on the Riemann sphere, maps the open left [resp. right] half plane inside [resp. outside] the unit disk and maps the imaginary axis into the unit circle. Its

inverse is given by $\frac{z-1}{z+1}$. This transform can be applied formally to matrices and to matrix polynomials. In particular, for

$$R(z) = (1+z)^2 F\left(\frac{z-1}{z+1}\right),$$

we have $R(z) = R_0 + zR_1 + z^2R_2$ where

$$\begin{aligned} R_0 &= M_0 - M_1 + M_2, \\ R_1 &= 2(M_0 - M_2), \\ R_2 &= M_0 + M_1 + M_2. \end{aligned}$$

We have the following.

Theorem 4. *The matrix polynomial $R(z)$ has m roots equal to 1, n roots equal to -1 and $m+n$ roots equal to $\mu_i = (1+\lambda_i)/(1-\lambda_i)$, for $i = 1, \ldots, m+n$, where λ_i are the eigenvalues of the matrix H of (4), and where we assume $\mu_i = \infty$ if $\lambda_i = 1$ for some i. Moreover, the set $\mathscr{K}$ of the roots of* $\det((1+z)H + (1-z)I)$ *has $n-1$ elements in $\mathscr{D}_<$, $m-1$ elements in $\mathscr{D}_>$, one element equal to 1 and one element equal to $\mu^* = (1+\lambda^*)(1-\lambda^*)^{-1}$, where λ^* is defined in Theorem 1.*

Proof. The result follows from the proof of Theorem 2 and from the relation $\det R(z) = (1+z)^{2(m+n)} \det F\left(\frac{z-1}{z+1}\right)$. □

From the above theorem we deduce that the roots of $R(z)$ can be split into the two sets $\Gamma_- \subset \mathscr{D}_< \cup \{1\}$ and $\Gamma_+ \subset \mathscr{D}_> \cup \{-1, 1\}$, both of cardinality $m+n$, given by

(1) if $\lambda^* < 0$,
$$\begin{aligned} \Gamma_- &= \{\lambda \in \mathscr{K} \cap \mathscr{D}_<\} \cup \{1 \text{ with multiplicity } m\} \cup \{\mu^*\}, \\ \Gamma_+ &= \{\lambda \in \mathscr{K} \cap \mathscr{D}_>\} \cup \{-1 \text{ with multiplicity } n\} \cup \{1\}; \end{aligned}$$
(2) if $\lambda^* \geqslant 0$,
$$\begin{aligned} \Gamma_- &= \{\lambda \in \mathscr{K} \cap \mathscr{D}_<\} \cup \{1 \text{ with multiplicity } m+1\}, \\ \Gamma_+ &= \{\lambda \in \mathscr{K} \cap \mathscr{D}_>\} \cup \{-1 \text{ with multiplicity } n\} \cup \{\mu^*\}. \end{aligned}$$

We may easily verify that the matrix

$$T = (I+G)(I-G)^{-1} = \begin{bmatrix} I & 2S(I-V)^{-1} \\ 0 & (I+V)(I-V)^{-1} \end{bmatrix} \tag{12}$$

obtained by applying the Cayley tranform (11) to the matrix G of (6), solves the unilateral equation

$$R_0 + R_1X + R_2X^2 = 0 \tag{13}$$

associated with the matrix polynomial $R(z)$. Moreover, the $m+n$ eigenvalues of T are given by 1, counted with multiplicity m, and by the n eigenvalues of

$(I+V)(I-V)^{-1}$. Assuming that S is the minimal nonnegative solution of (2), the eigenvalues of $(I+V)(I-V)^{-1}$ belong to the open [resp. closed] unit disk if the eigenvalues of V belong to $\mathbb{C}_<$ [resp. $\mathbb{C}_< \cup \{0\}$]. In light of the results of Section 3, the eigenvalues of V belong to $\mathbb{C}_<$ if $\lambda^* < 0$, they belong to $\mathbb{C}_< \cup \{0\}$ if $\lambda^* \geqslant 0$. In both cases, since the eigenvalues of T must be roots of $R(z)$ [9], we conclude that $\sigma(T) = \Gamma_-$, i.e., T is a solution whose eigenvalues are the smallest moduli roots of $R(z)$. In particular, from the matrix T we may recover the matrix V and the matrix S which define G.

In this way we have reduced the problem of computing G to the problem of computing T. Unfortunately, the matrix T is not necessarily the unique solution of (13) with smallest spectral radius since the matrix polynomial $R(z)$ has more than $m+n$ roots in the closed unit disk and $m+n$ of them have modulus 1.

For computational reasons it would be important to transform the matrix polynomial $R(z)$ into a new one for which there exists unique the solution of the associated matrix equation having the smallest spectral radius. This will be achieved in the next subsection, by moving the undesired roots equal to 1 and to -1 by means of the shift technique to zero and to infinity, respectively.

4.2. *Shifting techniques*

In this section we apply the shift technique of [14–16] in order to construct a new matrix polynomial $\psi(z)$ having the same roots of $R(z)$ except for m roots equal to 1 and n roots equal to -1 which are moved to zero and to the infinity, respectively.

If $\lambda^* \neq 0$, the quadratic matrix equation associated with $\psi(z)$ has a unique minimal spectral solution, from which we may recover the matrix G of (6). If $\lambda^* = 0$ the matrix polynomial $\psi(z)$ still has a root equal to 1 so that we have to apply once again the shift technique in order to arrive at a matrix equation having a unique minimal spectral solution. We briefly recall the shift technique, then we apply it for achieving the desired transformation. For more details we refer the reader to [14–16].

Consider the quadratic matrix polynomial $\alpha(z) = z^2A_2 + zA_1 + A_0$, let $\lambda \neq 0$ be a simple root of $\alpha(z)$ and let $\boldsymbol{v}$ be a nonzero vector such that $\alpha(\lambda)\boldsymbol{v} = 0$. Choose any vector $\boldsymbol{u}$ such that $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 1$, and define the "shifted" function $\beta(z) = \alpha(z)(I - z^{-1}\lambda\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}})^{-1}$. Then $\beta(z)$ is the quadratic matrix polynomial $\beta(z) = z^2B_2 + zB_1 + B_0$, where

$$\begin{cases} B_2 = A_2, \\ B_1 = A_1 + \lambda A_2\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}, \\ B_0 = A_0 - A_0\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}. \end{cases} \tag{14}$$

Moreover, the following properties hold:

(1) if W is a solution of the equation $A_2X^2 + A_1X + A_0 = 0$ such that $W\boldsymbol{v} = \lambda\boldsymbol{v}$, then $Y = W - \lambda\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}$ is a solution of the equation $B_2X^2 + B_1X + B_0 = 0$, and $Y\boldsymbol{v} = 0$;

(2) the roots of $\beta(z)$ concide with the roots of $\alpha(z)$, except for λ, which is shifted to zero;
(3) if $\alpha(z)$ can be factored as $\alpha(z) = (I - zR)U(zI - W)$, then $\beta(z)$ can be factored as $\beta(z) = (I - zR)U(zI - Y)$; in other words, if R solves the matrix equation $X^2A_0 + XA_1 + A_2 = 0$, then R solves also the matrix equation $X^2B_0 + XB_1 + B_2 = 0$.

If we wish to shift the root λ to infinity, instead of to zero, it is sufficient to consider the reversed matrix polynomial $\alpha_r(z) = A_2 + zA_1 + z^2A_0$ whose roots are the reciprocal of the roots of $\alpha(z)$ (where we assume that $1/\infty = 0$ and $1/0 = \infty$), so that λ^{-1} is a simple root of $\alpha_r(z)$. We perform a shift of the root λ^{-1} to 0 in the polynomial $\alpha_r(z)$, thus obtaining the "shifted" polynomial $\beta_r(z) = \alpha_r(z)(I - z^{-1}\lambda^{-1}\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}})^{-1}$. By reverting again the polynomial $\beta_r(z)$ we obtain the polynomial $\beta(z) = \alpha(z)(I - z\lambda^{-1}\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}})^{-1}$ such that the roots of $\beta(z)$ concide with the roots of $\alpha(z)$, except for λ, which is shifted to infinity.

Now, we apply the above shift technique first to move to zero m roots of $R(z)$ equal to 1, and then to move to infinity n roots of $R(z)$ equal to -1.

We observe that 1 is a root of $R(z)$ of multiplicity m and that the dimension of the null space of $R(1)$ is m; more precisely since

$$R(1) = \begin{bmatrix} 0 & 4B \\ 0 & -4D \end{bmatrix},$$

the kernel of $R(1)$ is generated by the columns of the matrix $\begin{bmatrix} I \\ 0 \end{bmatrix}$, where I is the $m \times m$ identity matrix. By proceeding as in [14–16] we may easily verify that

$$\begin{aligned} \widehat{R}(z) &= R(z)\left(I - z^{-1}\begin{bmatrix} I \\ 0 \end{bmatrix}\begin{bmatrix} I & 0 \end{bmatrix}\right)^{-1} \\ &= \begin{bmatrix} 0 & B \\ 0 & I - D \end{bmatrix} + z\begin{bmatrix} -I - A & 2B \\ C & -2D \end{bmatrix} + z^2\begin{bmatrix} I - A & B \\ C & -I - D \end{bmatrix} \end{aligned}$$

and that the matrix

$$\widehat{T} = T - \begin{bmatrix} I \\ 0 \end{bmatrix}\begin{bmatrix} I & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2S(I - V)^{-1} \\ 0 & (I + V)(I - V)^{-1} \end{bmatrix}$$

solves the equation $\widehat{R}_0 + \widehat{R}_1\widehat{T} + \widehat{R}_2\widehat{T}^2 = 0$ where $\widehat{R}(z) = \widehat{R}_0 + \widehat{R}_1 z + \widehat{R}_2 z^2$.

Recall that $\widehat{R}(z)$ has the same roots of $R(z)$ except for m roots equal to 1 which are replaced by 0 in $\widehat{R}(z)$, therefore $\widehat{R}(z)$ still has n roots equal to -1. Now we want to apply the shift technique in order to arrive at a quadratic matrix polynomial $\varphi(z)$ which shares with $\widehat{R}(z)$ all the roots, except for the roots equal to -1, which are moved to infinity.

Differently from the previous shift, where the solution T of the matrix equation associated with $R(z)$ is transformed into the matrix $\widehat{T}$ which solves the matrix equation

associated with $\widehat{R}(z)$, now we want to perform a shift such that $\widehat{T}$ is still a solution of the matrix equation associated with $\varphi(z)$. In order to obtain this, in light of the properties summarized at the beginning of this section, it is sufficient to multiply $\widehat{R}(z)$ on the left, instead of on the right, by the suitable matrix function which performs the shift of the roots. More specifically, we observe that

$$\widehat{R}(-1) = \begin{bmatrix} 2I & 0 \\ 0 & 0 \end{bmatrix}$$

so that the rows of the matrix $[0\ I]$, where I is the identity matrix of size n, span the left null space of $\widehat{R}(-1)$. Therefore, by proceeding as in [14–16] we may easily verify that

$$\begin{aligned} \varphi(z) &= \left(I + z \begin{bmatrix} 0 \\ I \end{bmatrix} \begin{bmatrix} 0 & I \end{bmatrix} \right)^{-1} \widehat{R}(z) \\ &= \begin{bmatrix} 0 & B \\ 0 & I - D \end{bmatrix} + z \begin{bmatrix} -I - A & 2B \\ C & -I - D \end{bmatrix} + z^2 \begin{bmatrix} I - A & B \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{15}$$

has $m + n - 1$ roots in $\mathscr{D}_<$ (of which m are zero), $m + n - 1$ in $\mathscr{D}_>$ (of which n are at infinity), one root equal to 1 and the root $\mu^* = (1 + \lambda^*)/(1 - \lambda^*)$.

We have three cases:

- μ^* belongs to $\mathscr{D}_<$ and then the matrix polynomial $\varphi(z)$ has $m + n$ roots in $\mathscr{D}_<$ (which are the eigenvalues of the matrix $\widehat{T}$) and $m + n$ eigenvalues in $\mathscr{D}_> \cup \{1\}$; therefore, the matrix $\widehat{T}$ is the unique minimal spectral solution of the equation associated with $\varphi(z)$.
- μ^* belongs to $\mathscr{D}_>$ and then $\varphi(z)$ has $m + n$ roots in $\mathscr{D}_< \cup \{1\}$ (which are the eigenvalues of the matrix $\widehat{T}$) and $m + n$ eigenvalues in $\mathscr{D}_>$; therefore, the matrix $\widehat{T}$ is the unique minimal spectral solution of the equation associated with $\varphi(z)$.
- $\mu^* = 1$ and then $\varphi(z)$ has $m + n$ roots in $\mathscr{D}_< \cup \{1\}$ (which are the eigenvalues of the matrix $\widehat{T}$) and $m + n$ eigenvalues in $\mathscr{D}_> \cup \{1\}$; therefore, the matrix $\widehat{T}$ is a minimal spectral solution of the equation associated with $\varphi(z)$.

From the above properties, we deduce that, if $\mu^* \neq 1$ then $\widehat{T}$ is the unique minimal spectral solution of the matrix equation. Therefore we may apply fast algorithms, like cyclic reduction or logarithmic reduction, in order to compute $\widehat{T}$. Once we have computed the matrix $\widehat{T}$, which has the structure

$$\widehat{T} = \begin{bmatrix} 0 & X \\ 0 & Y \end{bmatrix},$$

from the equations $X = 2S(I - V)^{-1}$ and $Y = (I + V)(I - V)^{-1}$, we may recover $S = X(I + Y)^{-1}$.

In the critical case where $\mu^* = 1$ we have to apply a further shift in order to move one of the two unit roots of $\varphi(z)$ to zero. For simplicity, we consider the case encountered in practice where M is a singular M-matrix such that $M\boldsymbol{e} = 0$. In this

case Guo [5] has shown that, if $Me = 0$ and $\mu^* = 1$, then $Se = e$. Moreover, it is easy to verify that the vector $v = \begin{bmatrix} e & \frac{1}{2}e \end{bmatrix}^{\mathrm{T}}$ is such that $\widehat{T}v = v$. Therefore, choosing $u^{\mathrm{T}} = \begin{bmatrix} 0 & \frac{2}{n}e^{\mathrm{T}} \end{bmatrix}$ we have $u^{\mathrm{T}}v = 1$, $[I \;\; 0]u = 0$, i.e., v is orthogonal to the null space of $R(1)$ so that we may apply the shift technique to move the root 1 to zero. In this way, denoting by $E_{nn} = eu^{\mathrm{T}}$, if $e \in \mathbb{R}^n$ and by $E_{mn} = eu^{\mathrm{T}}$, if $e \in \mathbb{R}^m$, we obtain the new matrix polynomial

$$\begin{aligned}\psi(z) = \varphi(z)\left(I - z^{-1}\begin{bmatrix} e \\ \frac{1}{2}e \end{bmatrix} u^{\mathrm{T}}\right)^{-1} &= z^2\begin{bmatrix} I - A & B \\ 0 & 0 \end{bmatrix} \\ &+ z\begin{bmatrix} -I - A & 2B + \frac{1}{n}(2E_{mn} - BE_{nn}) \\ C & -I - D \end{bmatrix} \\ &+ \begin{bmatrix} 0 & B(1 - \frac{1}{n}E_{nn}) \\ 0 & (I - D)(I - \frac{1}{n}E_{nn}) \end{bmatrix}, \end{aligned} \tag{16}$$

which has the same roots of $\varphi(z)$ except for the root 1 which is moved to 0. In this way the matrix

$$W = \widehat{T} - uv^{\mathrm{T}} = \begin{bmatrix} 0 & 2S(I - V)^{-1} - \frac{2}{n}ee^{\mathrm{T}} \\ 0 & (I + V)(I - V)^{-1} - \frac{1}{n}ee^{\mathrm{T}} \end{bmatrix} \tag{17}$$

is the minimal spectral solution of the matrix equation associated with $\psi(z)$.

Remark 5. It is important to remark that the same argument can be applied to the case $|\mu^*| > 1$, when $Me = 0$, which corresponds to a positive recurrent stochastic process. In this case the shift does not change the order of convergence, that is still quadratic, but reduces the number of steps necessary to approximate the solution to a specific accuracy.

5. Implementative details

In order to solve the unilateral quadratic equation in a fast and efficient way, we apply the logarithmic reduction and cyclic reduction algorithms, that converge to a solution having minimal spectral radius (if it exists). The convergence is quadratic if there is a gap between the modulus of the $(n + m)$th and the $(n + m + 1)$st root of the associated matrix polynomial, ordered by increasing modulus, i.e., if the isolation ratio ξ is less than 1. This gap is a measure of the speed of the convergence as stated by the following result which describes cyclic reduction algorithm [18].

Theorem 6. *Given the unilateral quadratic matrix equation* $N_2X^2 + N_1X + N_0 = 0$, *where* $N_2, N_1, N_0 \in \mathbb{R}^{h\times h}$, *set* $N_2^{(0)} = N_2$, $N_1^{(0)} = N_1$, $N_0^{(0)} = N_0$, $\widehat{N}_1 = N_1$, *and for* $i = 0, 1, \ldots,$ *define*

$$\text{(CR iteration)} \quad \begin{cases} N_1^{(i+1)} = N_1^{(i)} - N_2^{(i)}(N_1^{(i)})^{-1}N_0^{(i)} \\ \qquad\qquad -N_0^{(i)}(N_1^{(i)})^{-1}N_2^{(i)}, \\ N_2^{(i+1)} = -N_2^{(i)}(N_1^{(i)})^{-1}N_2^{(i)}, \\ N_0^{(i+1)} = -N_0^{(i)}(N_1^{(i)})^{-1}N_0^{(i)}, \\ \widehat{N}_1^{(i+1)} = \widehat{N}_1^{(i)} - N_2^{(i)}(N_1^{(i)})^{-1}N_0^{(i)}, \end{cases} \tag{18}$$

where we assume $\det(N_1^{(i)}) \neq 0$. *If the quadratic matrix equations* $N_2X^2 + N_1X + N_0 = 0$ *and* $N_0Y^2 + N_1Y + N_2 = 0$ *have spectral minimal solutions* X *and* Y *with* $\rho(X) < 1$ *and* $\rho(Y) < 1$, *then* X *is such that* $X = X^{(i)} + O(\tau^{2^i})$ *where* $X^{(i)} = -(\widehat{N}_1^{(i)})^{-1}N_0$, *and* τ *is any real number less than* 1 *and greater than* $|\lambda_h/\lambda_{h+1}|$ *where* $\lambda_1, \ldots, \lambda_{2h}$ *are the roots of* $N_2\lambda^2 + N_1\lambda + N_0$ *ordered by nondecreasing modulus, where we assume zeros at infinity if* N_2 *is rank deficient*.

Observe that the existence of the minimal spectral solutions X and Y in the above theorem implies that the roots of the matrix polynomials $N_2z^2 + N_1z + N_0$ and $N_0z^2 + N_1z + N_2$ have a splitting with respect to the unit circle. When $\tau = 1$ the convergence of $X^{(i)}$ to X turns to linear or in certain cases it does not occur, in these situations the shift technique provides an effective tool for removing this drawback. Let us rewrite the unilateral quadratic matrix equation that we obtain with the shift technique: in the case of a simple shift from (15) we have

$$\begin{bmatrix} I - A & B \\ 0 & 0 \end{bmatrix} Z^2 + \begin{bmatrix} -I - A & 2B \\ C & -I - D \end{bmatrix} Z + \begin{bmatrix} 0 & B \\ 0 & I - D \end{bmatrix} = 0, \tag{19}$$

whose minimal solution

$$Z = \begin{bmatrix} 0 & X \\ 0 & Y \end{bmatrix}$$

allows us to recover the minimal solution of the algebraic Riccati equation by the simple formula $S = X(I + Y)^{-1}$.

In the case of double shift, from (16) we obtain the matrix equation

$$\begin{bmatrix} I - A & B \\ 0 & 0 \end{bmatrix} Z^2 + \begin{bmatrix} -I - A & 2B + \frac{1}{n}(2E_{mn} - BE_{nn}) \\ C & -I - D \end{bmatrix} Z + \begin{bmatrix} 0 & B(1 - \frac{1}{n}E_{mn}) \\ 0 & (I - D)(I - \frac{1}{n}E_{nn}) \end{bmatrix} = 0 \tag{20}$$

and from its minimal spectral solution W in (17) we obtain

$$X = W_{1,2} + \frac{2}{n}\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}, \quad Y = W_{2,2} + \frac{1}{n}\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}, \tag{21}$$

where W is partitioned into the four blocks $W_{i,j}$, $i, j = 1, 2$, so that we may write $S = X(I + Y)^{-1}$.

5.1. Scaling

Instead of applying CR directly to Eqs. (19) and (20), it is more convenient to perform a simple formal manipulation which allows us to introduce a parameter. The choice of the parameter can be tuned in order to optimize the numerical performance of the algorithms that we obtain in this way. In fact, from the numerical point of view the two crucial points are the Cayley transform and the matrix inversions during the CR algorithms. The potential numerical instability of the Cayley transform may appear in the formula $S = X(I + Y)^{-1}$, if $I + Y$ is ill-conditioned. So, for the stability of the algorithm it is important that the matrix $I + Y$ is well conditioned. Since $I + Y = I + (I + V)(I - V)^{-1} = 2(I - V)^{-1}$, the matrix $I + Y$ is ill-conditioned if $I - V = I + D - CS$ is ill-conditioned. Recalling that $D - CS$ is an M-matrix so that its eigenvalues are in $\mathbb{C}_{\leqslant}$, one obtains that the matrix $I - V$ is always nonsingular and is well conditioned if $V \approx 0$.

Using the straightforward fact that if

$$G = \begin{bmatrix} 0 & S \\ 0 & V \end{bmatrix}$$

solves Eq. (5), namely $M_2G^2 + M_1G + M_0 = 0$, then $F = G/\gamma$ solves $\gamma^2 M_2F^2 + \gamma M_1F + M_0 = 0$, we obtain a slightly different equation

$$\begin{bmatrix} \gamma^2 I - \gamma A & B \\ 0 & 0 \end{bmatrix} Z^2 + \begin{bmatrix} -\gamma^2 I - \gamma A & 2B \\ \gamma C & -\gamma I - D \end{bmatrix} Z + \begin{bmatrix} 0 & B \\ 0 & \gamma I - D \end{bmatrix} = 0, \tag{22}$$

whose minimal solution is

$$Z = \begin{bmatrix} 0 & X_\gamma \\ 0 & Y_\gamma \end{bmatrix}$$

and the matrix S can be recovered by the formula $S = \gamma X_\gamma (I + Y_\gamma)^{-1}$, so that $I + Y_\gamma = (I - V/\gamma)^{-1}$. As γ tends to infinity, the condition number of $I + Y_\gamma$ tends to 1, so, in principle, choosing a large value for γ gives the better conditioning of the matrix. However, the drawback in choosing a large γ is that cyclic reduction slows down its convergence speed.

A heuristic value that in our numerical tests has given a good performance is $\bar{\gamma} = \max_{1 \leqslant i \leqslant n+m} m_{ii}$, where m_{ii} are the diagonal entries of the matrix M.

By different arguments one can obtain diverse parameterizations of Eq. (19). For instance, if we divide the algebraic Riccati equation by a parameter γ we obtain

$$\begin{bmatrix} I - \nu A & \nu B \\ 0 & 0 \end{bmatrix} Z^2 + \begin{bmatrix} -I - \nu A & 2\nu B \\ \nu C & -I - \nu D \end{bmatrix} Z + \begin{bmatrix} 0 & \nu B \\ 0 & I - \nu D \end{bmatrix} = 0, \tag{23}$$

with $\nu = 1/\gamma$ and $S = X_\nu (I + Y_\nu)^{-1}$. As in the previous parametrization, large values of γ reduce the ill conditioning of the matrix $I + Y$.

Similarly, we may parametrize Eq. (20) obtained by means of a double shift and arrive at the following equation:

$$\begin{bmatrix} I-\nu A & \nu B \\ 0 & 0 \end{bmatrix} Z^2 + \begin{bmatrix} -I-\nu A & 2\nu B + \frac{1}{n}(2E_{mn} - \nu B E_{nn}) \\ \nu C & -I-\nu D \end{bmatrix} Z + \begin{bmatrix} 0 & \nu B(1-\frac{1}{n}E_{mn}) \\ 0 & (I-\nu D)(I-\frac{1}{n}E_{nn}) \end{bmatrix} = 0. \tag{24}$$

5.2. Structure of the blocks

It is interesting to observe that both Eqs. (22) and (24) have matrix coefficients with a special structure which can be used for reducing the computational cost of cyclic reduction.

More precisely, since in the cyclic reduction algorithm we have

$$N_2^{(0)} = \begin{bmatrix} * & * \\ 0 & 0 \end{bmatrix}, \qquad N_1^{(0)} = \begin{bmatrix} * & * \\ * & * \end{bmatrix}, \qquad N_0^{(0)} = \begin{bmatrix} 0 & * \\ 0 & * \end{bmatrix},$$

then at each step all the matrices $N_2^{(i)}$, $N_1^{(i)}$ and $N_0^{(i)}$ share the same structure of $N_2^{(0)}$, $N_1^{(0)}$ and $N_0^{(0)}$, respectively. Let us describe a single step of CR. For the sake of notational simplicity let us denote by N_2, N_1, $\widehat{N}_1$ and N_0 the matrices obtained at a generic step of CR and by N_2', N_1', $\widehat{N}_1'$ and N_0' the matrices obtained after one step of CR applied to N_2, N_1, $\widehat{N}_1$ and N_0. Denote

$$N_2 = \begin{bmatrix} P_1 & P_2 \\ 0 & 0 \end{bmatrix}, \quad N_1 = \begin{bmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2} \end{bmatrix}, \quad \widehat{N}_1 = \begin{bmatrix} \widehat{M}_{1,1} & \widehat{M}_{1,2} \\ \widehat{M}_{2,1} & \widehat{M}_{2,2} \end{bmatrix},$$
$$N_0 = \begin{bmatrix} 0 & Q_1 \\ 0 & Q_2 \end{bmatrix}, \quad N_1^{-1} = \begin{bmatrix} U_{1,1} & U_{1,2} \\ U_{2,1} & U_{2,2} \end{bmatrix} \tag{25}$$

and similarly do for N_2', N_1', $\widehat{N}_1'$ and N_0'. Then, form Theorem 6 we obtain that

$$\begin{aligned}
&P_1' = -H_1 P_1, \quad P_2' = -H_1 P_2 \\
&Q_1' = -Q_1 K_1, \quad Q_2' = -Q_2 K_1 \\
&\widehat{M}_{1,1}' = \widehat{M}_{1,1} - H_1 Q_1 - L_1 Q_2, \quad \widehat{M}_{1,2}' = \widehat{M}_{1,2}, \\
&\widehat{M}_{2,1}' = \widehat{M}_{2,1}, \quad \widehat{M}_{2,2}' = \widehat{M}_{2,2}, \\
&M_{1,1}' = \widehat{M}_{1,1}' - R_1 P_1, \\
&M_{1,2}' = \widehat{M}_{1,2}' - R_1 P_2, \\
&M_{2,1}' = \widehat{M}_{2,1}' - S_1 P_1, \\
&M_{2,2}' = \widehat{M}_{2,2}' - S_1 P_2,
\end{aligned} \tag{26}$$

where

$$\begin{aligned} H_1 &= P_1U_{1,1} + P_2U_{2,1}, \\ L_1 &= P_1U_{1,2} + P_2U_{2,2}, \\ K_1 &= U_{2,1}Q_1 + U_{2,2}Q_2, \\ R_1 &= Q_1U_{2,1},\ S_1 = Q_2U_{2,1}. \end{aligned} \tag{27}$$

For $m = n$, the above relations allow one to implement the CR step with the computational cost of 18 multiplications of $n \times n$ matrices and with the inversion of a $2n \times 2n$ matrix. Since the cost of a matrix multiplication amounts to $2n^3 - n^2$ arithmetic operations (ops) and the inversion of a $2n \times 2n$ matrix amounts to $2(2n)^3$ ops, we arrive at the overal cost of $34n^3 + \mathrm{O}(n^2)$ ops for a single step of CR. This complexity bound compares favourably with the complexity of Newton's iteration which is $62n^3$ ops [8].

Algorithm 1. CR for solving nonsymmetric algebraic Riccati equations.

- Input: the coefficients A, B, C, D of the algebraic Riccati equation, a parameter $\gamma > 0$, a small positive number ε.
- Consider the matrix equation (22) and from A, B, C, D construct its matrix coefficients M_2, M_1 and M_0. Initialize cyclic reduction with $N_2^{(0)} = M_2$, $N_1^{(0)} = M_1$, $N_0^{(0)} = M_0$ and $\widehat{N}_1^{(0)} = M_1$.
- While $\|N_2^{(i)}\| \geqslant \varepsilon$ and $\|N_0^{(i)}\| \geqslant \varepsilon$ perform a step of cyclic reduction, as described in general form in Eqs. (26) and (27).
- Compute $T = -(\widehat{N}_1^{(i+1)})^{-1}N_0^{(0)}$.
- Output: $S = \gamma T_{12}(I + T_{22})^{-1}$ the minimal solution of the algebraic Riccati equation, where $T_{12} = [T]_{i=1,m;\, j=m+1,m+n}$ and $T_{22} = [T]_{i,\, j=m+1,m+n}$.

We recall that one can consider Eq. (23) and apply to it the cyclic reduction algorithm. With this parametrization the required solution is recovered by the formula $S = T_{12}(I + T_{22})^{-1}$.

In the case of null recurrent stochastic processes, where linear convergence occurs, or in the case of positive recurrent proceses, to increase convergence speed it is convenient to consider Eq. (16) instead of (19) in order to apply the double shift. Observe that the algorithm is essentially the same except for the initial conditions, because the unilateral equation has the same structure with zero blocks. In practice it is preferable to consider the parametrized equation (24). The only difference, besides the initial assignment for $N_2^{(0)}$, $N_1^{(0)}$, $\widehat{N}_1^{(0)}$, and $N_0^{(0)}$, is the way of recovering the solution at the end of the CR stage, by means of the expression $T = -(\widehat{N}_1^{(i+1)})^{-1}K_0$, where

$$K_0 = \begin{bmatrix} 0 & \nu B \\ 0 & I - \nu D \end{bmatrix}$$

is the last term of the equation with the simple shift. In this case the solution can be written as $S = T_{12}(I + T_{22})^{-1}$, as before. The validity of this formula can be derived from [16]. We denote by Algorithm 2 the modification of Algorithm 1 performed along this lines for implementing the double shift.

6. Numerical experiments

We have tested our algorithms and compared them with some other methods using the test problems of [5] and [1]. The numerical experiments have been performed in Matlab 6.0, the code is available from the authors upon request. The residual error that we report in the tables is given by

$$\text{res} = \frac{\|XCX - XD - AX + B\|}{\|XCX\| + \|XD\| + \|AX\| + \|B\|},$$

where X is the computed approximation to a solution. Here and throughout this section $\|\cdot\|$ denotes the 1-norm.

The first test is taken from [5].

Test 1 [5, Example 6.1]. Consider a pseudo random $n \times n$ matrix R with no zero elements and set $W = \text{diag}(Re) - R$ so that W is an irreducible singular M-matrix. Introduce a real parameter α and let

$$M = \alpha I + W = \begin{bmatrix} D & -C \\ -B & A \end{bmatrix}$$

be the matrix defining the coefficients of the algebraic Riccati equation. The existence of a positive solution of (2) is guaranteed for $\alpha \geqslant 0$. For $\alpha = 0$ and for different values of n, Table 1 reports the number of steps and the relative residual of our algorithm 1 in the case where scaling is applied with the values $\gamma = \max m_{ii}$ and $\gamma = 1$ (no scaling), we make a comparison with cyclic reduction applied to the equation derived by Ramaswami (CRR) with the choice $t = 1/\max_i m_{i,i}$ [4]. To reduce effect of

Table 1
Number of steps and relative residual for test 1

n	$\gamma = 1$		$\gamma = \max m_{ii}$		CRR	
	iter	res	iter	res	iter	res
10	8	3.1×10^{-13}	10	2.0×10^{-16}	11	1.4×10^{-16}
20	9	3.1×10^{-12}	11	3.1×10^{-16}	12	1.9×10^{-16}
50	11	2.2×10^{-10}	12	4.4×10^{-16}	13	2.5×10^{-16}
100	12	5.4×10^{-8}	12	8.6×10^{-16}	13	3.0×10^{-16}

randomness, we run our algorithms 10 times and put on the table the worst residual and the rounded average number of steps. Cyclic reduction with scaling shows a good accuracy and convergence rate.

The next four tests are taken from [1] where Newton's iteration is compared with several fixed point methods (FP) and with the algorithm CRR and LRR obtained by applying cyclic reduction and logarithmic reduction to the unilateral equation (3) derived by Ramaswami [4]. The performances of CRR and LRR are very similar therefore we report only the ones of CRR.

Test 2 [1, Example 1]. Consider the algebraic Riccati equation associated with the M-matrix

$$M = \left[\begin{array}{cc|cc} 0.003 & -0.001 & -0.001 & -0.001 \\ -0.001 & 0.003 & -0.001 & -0.001 \\ \hline -0.001 & -0.001 & 0.003 & -0.001 \\ -0.001 & -0.001 & -0.001 & 0.003 \end{array}\right].$$

It easy to check that the matrix $S = \frac{1}{2}\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}$ is a solution of the algebraic Riccati equation, and with a bit more work it can be proved that it is the minimal nonnegative solution.

In this case the associated stochastic process is null-recurrent and the FP methods considered in [1] require so many iterations that it is impractical to use them for their sublinear convergence [1]. Both Newton's iteration and our algorithm 1 have a linear convergence and need a large number of steps in order to provide a reasonable approximation. Because of the shift technique, our algorithm 2 has a quadratic convergence. Table 2 reports the number of step for Newton's method, algorithm 1 and algorithm 2, together with the residual error and the relative error $\|\widehat{S} - S\|/\|S\|$, where $\widehat{S}$ is the computed approximation to the solution S. For all the methods we stopped the iteration when the residual ceased to decrease. In both Algorithms 1 and 2, we have chosen for the parameter γ the value $\max_i m_{ii}$. It is important to point out that our algorithm 2 (with the double shift) provides an approximation with full accuracy in just one step.

Table 2
Comparison of methods for test 2

	Newton	Alg 1	Alg 2	CRR
iter	25	25	1	29
res	8.7×10^{-17}	4.3×10^{-17}	8.7×10^{-17}	4.3×10^{-17}
err	3.0×10^{-8}	2.3×10^{-8}	1.7×10^{-16}	1.2×10^{-8}

Table 3
Comparison of methods for test 3

	Newton	Alg 1	Alg 2	CRR
iter	5	18	1	19
res	4.6×10^{-13}	1.4×10^{-12}	8.7×10^{-17}	3.1×10^{-13}
err	5.0×10^{-13}	1.2×10^{-13}	1.9×10^{-15}	7.0×10^{-13}

Test 3 [1, Example 2]. For the Riccati equation associated with the M-matrix

$$M = \left[\begin{array}{cc|cc} 0.003 & -0.001 & -0.001 & -0.001 \\ -0.001 & 0.003 & -0.001 & -0.001 \\ \hline -0.001 & -0.001 & 100.002 & -100 \\ -0.001 & -0.001 & -100 & 100.002 \end{array}\right],$$

the results are similar to the previous example. In fact, the associated stochastic process is still null-recurrent. Our algorithm 1 requires 34 iterations to converge, more than in the previous case, Newton's method requires 18 iteration, but algorithm 2 converges again in 1 step with a relative error of 1.4×10^{-16}, much less than the other methods.

Test 4 [1, Example 3]. Let $m = 2$, $n = 18$, and let all the off-diagonal entries of A be zero, all entries of B and C be equal to -0.001 and all off-diagonal entries of A be equal to 10. The remaining entries are such that $Me = e$. The minimal solution is a rectangular $m \times n$ matrix with all entries equal to 1/18. This example models a strongly positive recurrent process for which there is no need of performing a complete shift, but by Remark 5 it can be performed to improve the convergence, as reported in Table 3. As before we stopped the algorithms when residual ceased to decrease.

Test 5 [1, Example 4]. Let $m = n = 2$ and

$$A = \begin{bmatrix} 0.003 & -0.0001 \\ -0.0001 & 0.003 \end{bmatrix}, \quad D = \begin{bmatrix} 0.003 & 0 \\ 0 & 0.003 \end{bmatrix},$$
$$C = \begin{bmatrix} 0.0019 & 0.001 \\ 0.0019 & 0.001 \end{bmatrix}, \quad B = \begin{bmatrix} 0.0015 & 0.0015 \\ 0.0029 & 0.0001 \end{bmatrix}.$$

In this case the associated stochastic process is weakly transient and the minimal nonnegative solution is $S = \begin{bmatrix} \frac{19}{30} & \frac{1}{3} \\ \frac{19}{30} & \frac{1}{3} \end{bmatrix}$.

For this test the FP methods still require a large number of iterations, whereas Newton's iteration and CRR require about 10 steps for a residual error of the order 10^{-17}. A residual of the same order is obtained by Algorithm 1 in 9 steps. Applying Algorithm 2 with the complete shift provides convergence in one step to a stochastic solution different from the wanted solution S that is substochastic. In fact, the process

is transient and shifting the root 1 to zero alters the balancing of the eigenvalues with respect the unit circle.

Appendix A. The symmetric equation

It is worth to point out that the above reduction of a Riccati equation to a unilateral equation applies in any case, even in the very important case of the symmetric algebraic Riccati equations that arise in control theory [13]

$$XDX + A^{\mathrm{T}}X + XA - C = 0, \tag{A.1}$$

where the matrices D and C are real symmetric and each matrix in the equation is square. In the applications the required solution is the stabilizing one, i.e., the solution X such that $\sigma(A + DX) \subset \mathbb{C}_{<}$.

For symmetric algebraic Riccati equations the matrix H is the well-known Hamiltonian matrix

$$H = \begin{bmatrix} A & D \\ C & -A^{\mathrm{T}} \end{bmatrix},$$

which in most applications has no purely imaginary eigenvalues and satisfies the property of splitting with respect to the imaginary axis, in fact half of its eigenvalues are in $\mathbb{C}_{<}$ and the other half in $\mathbb{C}_{>}$. This allows one to apply the same reduction that we used for the nonsymmetric equations.

The main difference is that after the simple shift we obtain an equation with a pure splitting with respect to the unit disk and then the convergence is quadratic.

The equation after the shifts is

$$\begin{bmatrix} \gamma^2 I + \gamma A^{\mathrm{T}} & -C \\ 0 & 0 \end{bmatrix} Z^2 + \begin{bmatrix} -\gamma^2 I + \gamma A^{\mathrm{T}} & -2C \\ \gamma D & -\gamma I + A \end{bmatrix} Z + \begin{bmatrix} 0 & -C \\ 0 & \gamma I + A \end{bmatrix} = 0, \tag{A.2}$$

which can be solved with the algorithm described in Section 5. A good choice for the parameter is $\gamma = |\det(H)|^{\frac{1}{2n}}$.

The resulting algorithm provided a good accuracy in many tests, even if its computational cost is greater than the one of other solvers for symmetric algebraic Riccati equations. This topic is still under investigation.

Acknowledgments

The authors wish to thank an anonymous referee for providing many detailed remarks and suggestions which greatly improved the overall presentation of this paper.

References

[1] N.G. Bean, M.M. O'Reilly, P.G. Taylor, Algorithms for return probabilities for stochastic fluid flows, Stoch. Models 21 (1) (2005) 149–184.
[2] L.C.G. Rogers, Fluid models in queueing theory and Wiener–Hopf factorization of Markov chains, Ann. Appl. Probab. 4 (2) (1994) 390–413.
[3] D. Williams, A "potential-theoretic" note on the quadratic Wiener–Hopf equation for Q-matrices, in: Seminar on Probability, XVI, Lecture Notes in Math., 920, Springer, Berlin, 1982, pp. 91–94.
[4] V. Ramaswami, Matrix analytic methods for stochastic fluid flows, in: Proceedings of the 16th International Teletraffic Congress, Edimburgh, 7–11 June 1999, pp. 1019–1030.
[5] C.-H. Guo, Nonsymmetric algebraic Riccati equations and Wiener–Hopf factorization for M-matrices, SIAM J. Matrix Anal. Appl. 23 (1) (2001) 225–242 (electronic).
[6] C.-H. Guo, A note on the minimal nonnegative solution of a nonsymmetric algebraic Riccati equation, Linear Algebra Appl. 357 (2002) 299–302.
[7] S. Asmussen, Stationary distributions for fluid flow models with or without Brownian noise, Comm. Statist. Stochastic Models 11 (1) (1995) 21–49.
[8] C.-H. Guo, A.J. Laub, On the iterative solution of a class of nonsymmetric algebraic Riccati equations, SIAM J. Matrix Anal. Appl. 22 (2) (2000) 376–391 (electronic).
[9] I. Gohberg, P. Lancaster, L. Hodman, Matrix Polynomials, Comp. Sci. Appl. Math., Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982.
[10] G. Latouche, V. Ramaswami, A logarithmic reduction algorithm for quasi-birth–death processes, J. Appl. Probab. 30 (3) (1993) 650–674.
[11] D. Bini, B. Meini, On the solution of a nonlinear matrix equation arising in queueing problems, SIAM J. Matrix Anal. Appl. 17 (4) (1996) 906–926.
[12] C.-H. Guo, Convergence analysis of the Latouche–Ramaswami algorithm for null recurrent quasi-birth–death processes, SIAM J. Matrix Anal. Appl. 23 (3) (2001/02) 744–760 (electronic).
[13] P. Lancaster, L. Rodman, Algebraic Riccati Equations, Oxford Science Publications, The Clarendon Press Oxford University Press, New York, 1995.
[14] C. He, B. Meini, N.H. Rhee, A shifted cyclic reduction algorithm for quasi-birth–death problems, SIAM J. Matrix Anal. Appl. 23 (3) (2001/02) 673–691 (electronic).
[15] D.A. Bini, L. Gemignani, B. Meini, Solving certain matrix equations by means of Toeplitz computations: algorithms and applications, in: Fast Algorithms for Structured Matrices: Theory and Applications (South Hadley, MA, 2001), Contemp. Math., 323, Amer. Math. Soc., Providence, RI, 2003, pp. 151–167.
[16] D.A. Bini, G. Latouche, B. Meini, Numerical Methods for Structured Markov Chains, Oxford University Press, Oxford, 2005.
[17] R.S. Varga, Matrix Iterative Analysis, Springer Ser. Comput. Math., Springer-Verlag, Berlin, 2000.
[18] D.A. Bini, L. Gemignani, B. Meini, Computations with infinite Toeplitz matrices and polynomials, Structured and infinite systems of linear equations, Linear Algebra Appl. 343/344 (2002) 21–61 (special issue).