Linear Algebra and its Applications 434 (2011) 201–214

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Solution of a nonsymmetric algebraic Riccati equation from a two-dimensional transport model

Tiexiang Li [a], Eric King-wah Chu [b,*], Jong Juang [c], Wen-Wei Lin [d]

[a] Department of Mathematics, Southeast University, Nanjing 211189, People's Republic of China
[b] School of Mathematical Sciences, Building 28, Monash University, VIC 3800, Australia
[c] Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan
[d] CMMSC and NCTS, Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:
Received 18 January 2010
Accepted 25 August 2010

Submitted by V. Mehrmann

AMS classification:
15A24
45G10
45J05
65H10
65R20
65Z05
82D75

Keywords:
Algebraic Riccati equation
Doubling algorithm
Fixed-point iteration
Newton's method
Reflection kernel
Transport theory

ABSTRACT

For the steady-state solution of an integral–differential equation from a two-dimensional model in transport theory, we shall derive and study a nonsymmetric algebraic Riccati equation $B^- - XF^- - F^+X + XB^+X = 0$, where $F^\pm \equiv I - \hat{s}PD^\pm$, $B^- \equiv (\hat{b}I + \hat{s}P)D^-$ and $B^+ \equiv \hat{b}I + \hat{s}PD^+$ with a nonnegative matrix P, positive diagonal matrices $D^\pm$, and nonnegative parameters f, $\hat{b} \equiv b/(1-f)$ and $\hat{s} \equiv s/(1-f)$. We prove the existence of the minimal nonnegative solution X^* under the physically reasonable assumption $f + b + s\,\|P(D^+ + D^-)\|_\infty < 1$, and study its numerical computation by fixed-point iteration, Newton's method and doubling. We shall also study several special cases; e.g. when $\hat{b} = 0$ and P is low-ranked, then $X^* = \frac{\hat{s}}{2}UV$ is low-ranked and can be computed using more efficient iterative processes in U and V. Numerical examples will be given to illustrate our theoretical results.

* Corresponding author. Tel.: +61 3 99054480; fax: +61 3 99054403.
E-mail addresses: txli@seu.edu.cn (T. Li), eric.chu@monash.edu (E.K.-w. Chu), jjuang@math.nctu.edu.tw (J. Juang), wwlin@am.nctu.edu.tw (W.-W. Lin).

0024-3795/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2010.09.006

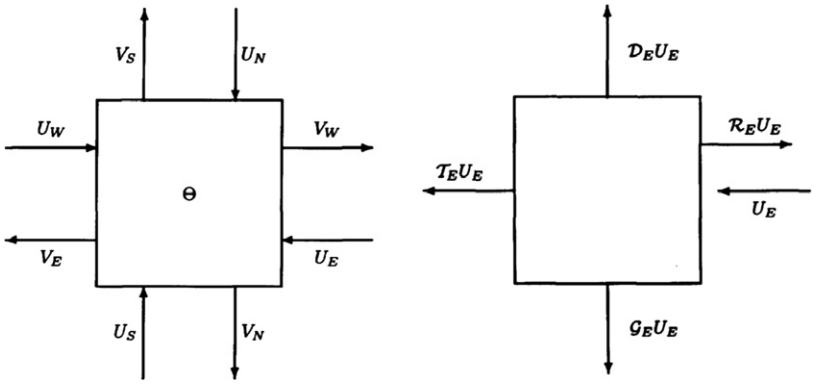

Fig. 1. The four-port-system.

1. Introduction

Transport theory has been an active area of research, associated with masters like Bellman and Chandrasekhar (see [2,13] and the references therein). A one-dimensional model was studied first in [13], stimulating a series of numerical studies, e.g., in [1,14,15,17], in the past 15 years. We shall study a more general two-dimensional model in this paper.

In [18,19], the transport of particles stemming from a rectangular beam bounded in the rectangle $[0, x] \times [0, y]$ and unbounded in the orthogonal z direction, incident upon a similar rectangular region, was considered. Assuming that the incident particle beam is from the East, it is important to determine the East reflection kernel, from which the corresponding transmission, left-turn and right-turn kernels can be deduced [18].

Consider the domain $\Theta = [0, \alpha] \times [0, \beta]$ in Fig. 1, with particle input U_d and output V_d from various directions $d = N, S, E, W$. From the input U_E from the East, we are interested in the corresponding transmission, left-turn, right-turn and reflection operators $\mathcal{T}_E$, $\mathcal{G}_E$, $\mathcal{D}_E$ and $\mathcal{R}_E$, respectively, producing outputs $\mathcal{T}_E U_E$, $\mathcal{G}_E U_E$, $\mathcal{D}_E U_E$ and $\mathcal{R}_E U_E$.

For example, as on the left of Fig. 2 for $\mathcal{T}_E$, the incident flux U_E comes in at $y = y_i$ and the resulting emerging flux emerges at $y = y_e$, produces an output in the form

$$[\mathcal{T}_E U_E](y_e) \equiv \int_0^\beta T_E(\alpha, \beta, y_e, y_i) U_E(y_i)\, dy_i,$$

where $T_E(x, y, y_e, y_i)$ is the corresponding transmission kernel. For the output flux V_E from the East, we have contributions from all four directions, summing to

$$V_E = \mathcal{T}_E U_E + \mathcal{D}_N U_N + \mathcal{R}_W U_W + \mathcal{G}_S U_S.$$

To understand the system, we need to determine all the kernels. Because of symmetry, we shall consider only the kernels corresponding to the Eastern direction.

From [18,19], the integral–differential equations for the kernels for $\mathcal{T}_E$, $\mathcal{G}_E$, $\mathcal{D}_E$, and $\mathcal{R}_E$ have been derived. For the kernel R_E of the Eastern reflection operator $\mathcal{R}_E$, we have

$$\begin{aligned}\frac{1}{\sigma}\frac{\partial R_E}{\partial x}(x, y, y_e, y_i) &= b\,\delta(y_e - y_i) + s\,p(y_e, y_i) + 2(f-1)\,R_E(x, y, y_e, y_i) \\ &\quad + s\int_0^y \left[p(y', y_i) R_E(x, y, y_e, y') + p(y_e, y') R_E(x, y, y', y_i)\right] dy' \\ &\quad + \int_0^y \left[b R_E(x, y, y_e, y') + s\int_0^y p(y'', y') R_E(x, y, y_e, y'')\, dy''\right] \\ &\quad \times R_E(x, y, y', y_i)\, dy', \end{aligned} \tag{1}$$

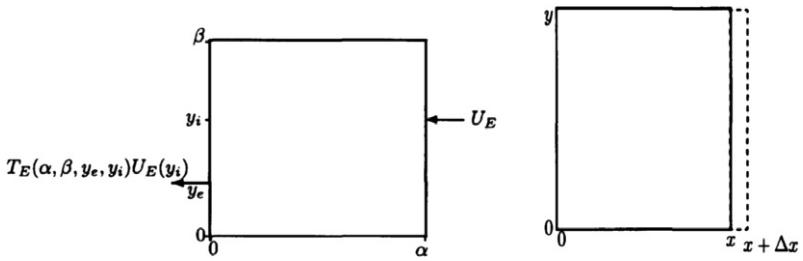

Fig. 2. The incremental layer and the transmission operator $\mathcal{T}_E$.

with initial/boundary conditions $R_E(0,\beta,y_e,y_i)=0$, where $\delta(\cdot)$ is the Kronecker function with $\delta(0)=1$ but vanishing at other arguments. For the parameters in (1), σ is the positive scattering cross-section, and f, b and s are nonnegative expected numbers of particles that emerged from a collision moving in, respectively, the same direction as the particle that engendered the collision, the opposite, or either of the two orthogonal directions. The nonnegative intensity kernel $p(\cdot,\cdot)$ defines the expected intensity operator $\mathcal{P}$:

$$(\mathcal{P}V)\left(y^{(k)}\right)=\int_0^y p\left(y^{(k)},y'\right)V(y')\,dy', \tag{2}$$

on some function V.

One important problem in transport theory is to understand the behaviour of the four-port model in Figs. 1 and 2, through the operators $\mathcal{T}_E$, $\mathcal{G}_E$, $\mathcal{D}_E$ and $\mathcal{R}_E$, whose kernels can be obtained by solving their corresponding integral–differential equations like (1). One important feature is that the equations for the kernels T_E, G_E and D_E are decoupled from each other, but dependent on the reflection kernel R_E; details can be found in [18,19]. In this paper, we shall concentrate on the steady-state solution to (1) for the reflection kernel R_E. The global solution of the integral–differential equation (1) can be computed as a fixed point of a positive operator, with the steady-state solution as its natural upper bound. Moreover, these global solutions converge to the steady-state solution under favourable conditions. Note that R_E is differentiated with respect to the spatial variable x in (1) and the steady-state solution has to be interpreted as the scattering behaviour stabilizing further away from the source, rather than against time.

Quoting from [18,19], we shall attempt to pass on the essence of the approach of "invariant imbedding", which produces (1).

To the subregion $[0,x]\times[0,y]$ (on the right in Fig. 2) of Θ (in Fig. 1) with $x\in(0,\alpha)$ and $y\in(0,\beta)$, imbed an additional strip $[x,x+\Delta x]\times[0,y]$ on the right. Particles enter the subregion from the right are reflected back from several possibilities. First, the probability of a particle having a collision inside the strip is $\sigma\Delta x$, of which $f\sigma\Delta x$ accounts for it continuing forward, $b\sigma\Delta x$ backward, and $s\sigma\Delta x$ left or right. When a particle enters the strip at $y=y_i$, collides and transverses up (or down) the strip, it produces a source and an emerging flux out of the strip to the left or right at $y=y_e$ with probability $s\sigma p(y_i,y_e)\Delta x$. Also, the probability of any particle going through the strip is $[1+\sigma\Delta x(f-1)]$, being the sum of $1-\sigma\Delta x$ (no collision) and $f\sigma\Delta x$ (going forward after collision). Let $R_E(x,y,y_e,y_i)$ and $R_E(x+\Delta x,y,y_e,y_i)$ represent the amounts of reflection, with incident particles at $y=y_i$ and emerging particles at $y=y_e$, in the subregion and the augmented region, respectively. Showing only the effects of at most two collisions in the strip, we have

$$\begin{aligned}R_E(x+\Delta x,y,y_e,y_i)&=[1+\sigma\Delta x(f-1)]R_E(x,y,y_e,y_i)[1+\sigma\Delta x(f-1)]\\&\quad+b\sigma\delta(y_e-y_i)\Delta x+s\sigma p(y_e,y_i)\Delta x+\cdots\\&=R_E(x,y,y_e,y_i)+2(f-1)\sigma R_E(x,y,y_e,y_i)\Delta x+b\sigma\delta(y_e-y_i)\Delta x\\&\quad+s\sigma p(y_e,y_i)\Delta x+\cdots.\end{aligned} \tag{3}$$

The first term in between the equality signs in (3) accounts for the particle going through the strip into the subregion, turning after reflection and then going through the strip again and emerge. The second

term accounts for the particle bouncing back in the strip, and the third term for the particle to collide in the strip, transverse and exit. Rearrange and let $\Delta x \to 0$, (3) implies

$$\frac{1}{\sigma}\frac{\partial R_E}{\partial x}(x, y, y_e, y_i) = b\,\delta(y_e - y_i) + s\,p(y_e, y_i) + 2(f-1)\,R_E(x, y, y_e, y_i) + \cdots,$$

a truncated version of (1). We now abbreviate the discussion and "explain" the remaining terms in (1). For the first integral, the particle collides and transverses in the strip, enters the subregion (at $y = y'$) and reflects, or reflects first before colliding (at $y = y'$) and transversing in the strip before emerging. The second integral accounts for the particle going through the strip, reflecting in the subregion, bouncing back to the strip (at $y = y'$), reflecting again and emerging to the right. For the last (double) integral, the particle goes through the strip, reflects, collides in the strip (at $y = y'$) and transverses in the strip, re-enters into the subregion (at $y = y''$), reflects and emerges to the right. The integrals sum all the possibilities in y' and y''. Many other paths for the particle are obviously possible but they contribute towards higher order terms in Δx and disappear when $\Delta x \to 0$.

From obvious probabilistic reasons in the above transport model, we assume the inequalities

$$b + f + 2s \leqslant 1, \tag{4}$$

$$\psi(\mathcal{P}) \equiv \max_{y\in[0,1],\|V\|_\infty=1} \left| \int_0^1 p(y, y')V(y')\,dy' \right| \leqslant 1, \tag{5}$$

with b, f, s and $p(y, y')$ being nonnegative. When (4) is satisfied with equality, our system and the resulting nonsymmetric algebraic Riccati equation (6) are described as *critical*.

For the steady-state solution, we have the right-hand-side of the integral–differential equation (1) equals to zero, yielding an integral equation in R_E. Assume without loss of generality that $y = 1$ and apply numerical quadrature with n positive weights $\{d_k^{\pm}\}$ and nodes $\{y^{(k)}\}$, we have the approximated equation at $y_i = y^{(m)}$ and $y_e = y^{(l)}$:

$$\begin{aligned} &b\delta_{lm} + sp_{lm} + 2(f-1)r_{lm} \\ &\quad + s\sum_k \left(r_{lk}d_k^- p_{km} + p_{lk}d_k^+ r_{km} \right) + b\sum_k r_{lk}d_k^- r_{km} + s\sum_{k,t} r_{lk}d_k^- p_{kt}d_t^+ r_{tm} = 0 \end{aligned}$$

with $p_{lm} \equiv p(y^{(l)}, y^{(m)})$, $r_{lm} \equiv R_E(x, y, y^{(l)}, y^{(m)})$ and $D^{\pm} \equiv \mathrm{diag}\{d_k^{\pm}\}$. Note that $d_k^{\pm}$ are of $O(n^{-1})$ for many standard numerical quadratures, and we allow the flexibility of different weights d_k^+ and d_k^-, thus different accuracies, for the numerical integration} respect to y' and y'' in the double integral in (1). In matrix form with $\hat{b} \equiv b/(1-f)$, $\hat{s} \equiv s/(1-f)$ and using the convention $M = [m_{ij}]$ (with capital letters denoting matrices and the corresponding lower-case letters with indices for their elements), we have the nonsymmetric algebraic Riccati equation (NARE):

$$\left(\hat{b}I + \hat{s}P\right) - 2R + \hat{s}(RD^-P + PD^+R) + \hat{b}RD^-R + \hat{s}RD^-PD^+R = 0.$$

Post-multiply the above equation by D^-, the NARE now reads:

$$B^- - XF^- - F^+X + XB^+X = 0 \tag{6}$$

with the $n \times n$ matrices $X \equiv RD^-$ and

$$F^{\pm} \equiv I - \hat{s}PD^{\pm}, \quad B^- \equiv \left(\hat{b}I + \hat{s}P\right)D^-, \quad B^+ \equiv \hat{b}I + \hat{s}PD^+. \tag{7}$$

Remark 1.1. The more useful variable is $X = RD^-$, not R or R_E, because ultimately we are interested in integrals like

$$(\mathcal{R}_E V)\left(y'\right) = \int_0^y R_E\left(x, y, y', y''\right) V\left(y''\right) dy'' \approx (RD^-v)_k = (Xv)_k \tag{8}$$

for some function V with the corresponding function values in $v = [V(y^{(1)}), \ldots, V(y^{(n)})]^\top$ and $y' \approx y^{(k)}$.

The "convergence" of the solution X of (6) to $\mathcal{R}_E$ for the original integral equation is an interesting problem on its own and will be pursued elsewhere. At the moment, we shall assume that the solution X is an accurate approximation to $\mathcal{R}_E$ for large enough values of n, in the sense that the difference between the left- and right-hand-sides in (8) diminishes to zero as $n \to \infty$.

In addition to $\mathcal{R}_E(V)$ in (8), we may also be interested in

$$\left(\mathcal{R}_E^{\star} V\right)\left(y''\right) \equiv \int_0^y V\left(y'\right) R_E\left(x, y, y', y''\right)\, dy' \approx \left(v^{\top} D^{+} R\right)_k = \left(v^{\top} Y\right)_k$$

with $Y = D^{+}R = D^{+}X(D^{-})^{-1}$. Analogously, we can derive an NARE for Y (whose existence requires an assumption on $D^{\pm}P$ in 1-norm, similar to that on $PD^{\pm}$ in (11)) but it may be simpler to solve (6) for X and then retrieve Y afterward.

2. Existence of solution

Some relevant definitions are as follows. For any matrices $\widehat{A}, \widehat{B} \in \mathbb{R}^{m\times n}$, we write $\widehat{A} \geqslant \widehat{B}$ ($\widehat{A} > \widehat{B}$) if their elements satisfy $\hat{a}_{ij} \geqslant \hat{b}_{ij}$ ($\hat{a}_{ij} > \hat{b}_{ij}$) for all i, j. A real square matrix $\widehat{A}$ is called a Z-matrix if all its off-diagonal elements are nonpositive. It is clear that any Z-matrix $\widehat{A}$ can be written as $sI - \widehat{B}$ with $\widehat{B} \geqslant 0$. A Z-matrix $\widehat{A}$ is called an M-matrix if $s \geqslant \rho(\widehat{B})$, where $\rho(\cdot)$ is the spectral radius; it is a singular M-matrix if $s = \rho(\widehat{B})$ and a nonsingular M-matrix if $s > \rho(\widehat{B})$. We have the following useful results from [3] and [8, Theorem 1.1]:

Lemma 2.1. *For a Z-matrix $\widehat{A}$, the following three statements are equivalent*:

(*a*) $\widehat{A}$ *is a nonsingular M-matrix*.
(*b*) $\widehat{A}^{-1} \geqslant 0$.
(*c*) $\widehat{A}v > 0$ *for some vector* $v > 0$.

Theorem 2.2. *For the NARE*

$$X\widehat{C}X - X\widehat{D} - \widehat{A}X + \widehat{B} = 0 \tag{9}$$

where $\widehat{A}, \widehat{B}, \widehat{C}$ and $\widehat{D}$ are real matrices of sizes $m \times m$, $m \times n$, $n \times m$, $n \times n$, respectively. Assume that

$$M = \begin{bmatrix} \widehat{D} & -\widehat{C} \\ -\widehat{B} & \widehat{A} \end{bmatrix} \tag{10}$$

is a nonsingular M-matrix or an irreducible singular M-matrix. Then the NARE has a minimal nonnegative solution S. If M is irreducible, then $S > 0$ and $\widehat{A} - S\widehat{C}$ and $\widehat{D} - \widehat{C}S$ are irreducible M-matrices. If M is a nonsingular M-matrix, then $\widehat{A} - S\widehat{C}$ and $\widehat{D} - \widehat{C}S$ are nonsingular M-matrices. If M is an irreducible singular M-matrix with positive left and right null vectors $[u_1^{\top}, u_2^{\top}]^{\top}$ and $[v_1^{\top}, v_2^{\top}]^{\top}$ (where $u_1, v_1 \in \mathbb{R}^n$ and $u_2, v_2 \in \mathbb{R}^m$) satisfying

$$u_1^{\top} v_1 \neq u_2^{\top} v_2,$$

then

$$M_S = I_n \otimes (\widehat{A} - S\widehat{C}) + (\widehat{D} - \widehat{C}S)^{\top} \otimes I_m$$

is a nonsingular M-matrix. If M is an irreducible singular M-matrix with $u_1^{\top} v_1 = u_2^{\top} v_2$, then M_S is an irreducible singular M-matrix.

Applying Lemma 2.1 and Theorem 2.2, we have the following existence result:

Theorem 2.3. *Under the assumption that*

$$b + f + s\, \|P(D^{+} + D^{-})\|_{\infty} < 1, \tag{11}$$

with b, f, s and P being nonnegative, the unique minimal nonnegative solution X^{*} of (6) exists.

Proof. Applying Theorem 2.2, we need to show that the Z-matrix

$$M = \begin{bmatrix} I - \hat{s}PD^{-} & -\left(\hat{b}I + \hat{s}PD^{+}\right) \\ -\left(\hat{b}I + \hat{s}P\right)D^{-} & I - \hat{s}PD^{+} \end{bmatrix} = I - \hat{b}\begin{bmatrix} D^{-} & I \end{bmatrix} - \hat{s}\begin{bmatrix} P \\ P \end{bmatrix}\left[D^{-},\ D^{+}\right] \tag{12}$$

is a nonsingular M-matrix or irreducible singular M-matrix. For the former, applying Lemma 2.1 to M requires a vector $v > 0$ such that $Mv > 0$. Let $v = e$ (the vector of all ones) and we need

$$\hat{b} + \hat{s}\ \|P(D^{+} + D^{-})\|_{\infty} < 1, \tag{13}$$

which is equivalent to our assumption (11). □

For the rest of the paper, any matrix norm will be the ∞-norm unless otherwise stated. Many other useful results on more general NAREs can be found in [8].

Remark 2.1. For the numerical quadrature chosen in deriving (6) from (1), we shall assume that it is exact for some interpolating function of appropriate smoothness. Note that most numerical quadratures can be derived through exact integration of such interpolating functions V, and different Vs yield different quadratures or weights. In other words, for some $v = [V(y^{(1)}), \ldots, V(y^{(n)})]^{\top}$ and its interpolating function V, we have

$$(\mathcal{P}V)(y^{(l)}) = \int_0^1 p(y^{(l)}, y')V(y')\,dy' = \sum_k p_{lk}d_k^{\pm}v_k = (PD^{\pm}v)_l.$$

Together with (5) and for some v with $\|v\| = 1$ and its interpolating function V, it implies that

$$\|PD^{\pm}\| = \|PD^{\pm}v\| = \max_l \left|(\mathcal{P}V)(y^{(l)})\right| \leqslant \psi(\mathcal{P}) \leqslant 1. \tag{14}$$

With (4) and (14) not *both* satisfied with equality, the sufficient condition in (13) or assumption (11) in Theorem 2.3 are satisfied. Consequently, the critical case (with equality in (4)) does not satisfy (14) only if (5) is also satisfied with equality.

We shall consider the super-critical case, when both (4) and (5) are satisfied with equality, later in the next sub-section. In some applications, P is of low rank. We shall consider this special case in Section 3.

2.1. NARE as an eigenvalue problem

The NARE (6) can be reformulated as the following eigenvalue problem

$$H\begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix} S, \quad H \equiv \begin{bmatrix} -F^{-} & B^{+} \\ -B^{-} & F^{+} \end{bmatrix} = \begin{bmatrix} -I & \hat{b}I \\ -\hat{b}D^{-} & I \end{bmatrix} + \hat{s}\begin{bmatrix} I \\ -I \end{bmatrix} P\left[D^{-},\ D^{+}\right]. \tag{15}$$

From (11), it is easy to see that the eigenvalues of H are shifted from ± 1, splitting equally on opposite sides of the imaginary axis. Using the Gerschgorin Theorem and denote $\mathcal{D}(a, r) \equiv \{x \in \mathbb{C} : |x - a| \leqslant r\}$, the eigenvalues are in the regions $\mathcal{D}(-1, \alpha) \cup \mathcal{D}(1, \alpha)$ on opposite sides of the imaginary axis, with $\alpha \equiv \hat{b} + \hat{s}\ \|P(D^{+} + D^{-})\| < 1$.

Remark 2.2. With $\alpha = 1$ in the super-critical case, a simple application of the Gerschgorin Theorem implies that H in (15) and M in (12) may be singular. However, this potential singularity may be detected or excluded, by applying the extensions of the Gerschgorin Theorem in [12, Section 6.2]. Consider all the Gerschgorin disks of H containing the origin, at least one of the corresponding inequalities should not be satisfied with equality. In other words, we may detect or exclude this super-critical case that *all* the first n rows have their row sums equal to zero.

Note that even if H or M are singular, the existence result in Theorem 2.2 still holds provided that M is irreducible. With the additional requirement for the null vectors as in Theorem 2.2, the Newton's method in Section 4.2 will be convergent quadratically.

3. Special cases

In this section, we shall consider several special cases. When $b = 0$, the NARE simplifies to (17). If, in addition, P is low-ranked, the NARE simplifies further to (18) and (19), yielding a solution of low-rank which can be solved efficiently via the nonlinear equations in (20) in Section 3.2. Other special cases are less interesting – when $s = \hat{s} = 0$ and $b \neq 0$, the NARE degenerates to a simple quadratic (see Section 3.3); and when $b = s = 0$, the problem becomes trivial, with the NARE degenerates into the $0 = 0$ situation. Finally, when $f = 0$, the NARE remains qualitatively the same as (6).

3.1. The $b = 0$ case

When $b = 0$, the NARE (6) then reads

$$\hat{s}PD^- - X(I - \hat{s}PD^-) - (I - \hat{s}PD^+)X + \hat{s}XPD^+X = 0. \tag{16}$$

Equivalently, we have

$$X = \phi(X) \equiv \frac{\hat{s}}{2}\left(PD^- + XPD^- + PD^+X + XPD^+X\right) = \frac{\hat{s}}{2}(I + X)P(D^- + D^+X), \tag{17}$$

implying that X is low-ranked when P is.

We have the following special case of [6, Theorem 2.3] for X^* and the iteration $X^{(k+1)} = \phi(X^{(k)})$:

Theorem 3.1. *Let $X^{(0)} = 0$ and $X^{(k+1)} = \phi(X^{(k)})$; i.e., the fixed-point iteration for the NARE in (17) when $b = 0$. Then under assumption (11) in Theorem 2.3, we have*

(i) the iterates satisfy $X^* \geqslant X^{(k+1)} \geqslant X^{(k)} \geqslant \frac{\hat{s}}{2}PD^- \geqslant 0$, and
(ii) $X^{(k)} \to X^*$ as $k \to \infty$.

We can apply Newton's method and doubling [4,11] to solve (16), as in the general case in Sections 4.2 and 4.3.

3.2. Low-ranked P when $b = 0$

When $P = P_1P_2$ is of rank r (with $P_1, P_2^\top$ being $n \times r$, $r < n$) and $b = 0$, (17) implies

$$X = \frac{\hat{s}}{2}UV \tag{18}$$

with the auxiliary variables

$$U = (I + X)P_1, \quad V = P_2(D^- + D^+X), \tag{19}$$

where $U, V^\top$ are $n \times r$. Substituting X in (18) into (19), we have $2rn$ nonlinear equations for the $2rn$ unknowns in U and V:

$$U = \left(I + \frac{\hat{s}}{2}UV\right)P_1, \quad V = P_2\left(D^- + \frac{\hat{s}}{2}D^+UV\right). \tag{20}$$

Convergence of various iterative schemes (e.g., Newton's method, generalized nonlinear Jacobi and Gauss–Seidel methods, as in [1,6–11,14,15]) for the above set of nonlinear equations (20) can be shown, similar to the proof in Theorem 2.3 (or techniques in the respective references).

Consider the following iterative schemes, all starting from $U_{\mathcal{I}}^{(0)}, V_{\mathcal{I}}^{(0)} = 0$ for $\mathcal{I} = \mathcal{S}, \mathcal{M}, \mathcal{J}$ and $\mathcal{G}$:

(I) Simple Iteration (SI):

$$U_{\mathcal{S}}^{(k+1)} = \left[I + \frac{\hat{s}}{2}U_{\mathcal{S}}^{(k)}V_{\mathcal{S}}^{(k)}\right]P_1, \quad V_{\mathcal{S}}^{(k+1)} = P_2\left[D^- + \frac{\hat{s}}{2}D^+U_{\mathcal{S}}^{(k)}V_{\mathcal{S}}^{(k)}\right].$$

(II) Modified Simple Iteration (MSI):

$$U_{\mathcal{M}}^{(k+1)} = \left[I + \frac{\hat{s}}{2} U_{\mathcal{M}}^{(k)} V_{\mathcal{M}}^{(k)}\right] P_1, \quad V_{\mathcal{M}}^{(k+1)} = P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{M}}^{(k+1)} V_{\mathcal{M}}^{(k)}\right].$$

(III) Nonlinear Block Jacobi Method (NBJ):

$$U_{\mathcal{J}}^{(k+1)} = \left[I + \frac{\hat{s}}{2} U_{\mathcal{J}}^{(k+1)} V_{\mathcal{J}}^{(k)}\right] P_1, \quad V_{\mathcal{J}}^{(k+1)} = P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{J}}^{(k)} V_{\mathcal{J}}^{(k+1)}\right].$$

(IV) Nonlinear Block Gauss–Seidel Method (NBGS):

$$U_{\mathcal{G}}^{(k+1)} = \left[I + \frac{\hat{s}}{2} U_{\mathcal{G}}^{(k+1)} V_{\mathcal{G}}^{(k)}\right] P_1, \quad V_{\mathcal{G}}^{(k+1)} = P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{G}}^{(k+1)} V_{\mathcal{G}}^{(k+1)}\right].$$

We have the following results for various iterates, similar to those in [10]:

Theorem 3.2. *Assume that (4) holds with $b = 0$ and the splitting $P = P_1 P_2$, with $P_1, P_2 \geqslant 0$, satisfies*

$$\|P_1\| = 1, \quad 0 < \|P_2 D^{\pm}\| \leqslant 1. \tag{21}$$

Ignoring the subscripts when the result holds for all the four methods, we have

(*i*) *the iterates satisfy $U^* \geqslant U^{(k+1)} \geqslant U^{(k)} \geqslant U^{(1)} \geqslant 0$, $V^* \geqslant V^{(k+1)} \geqslant V^{(k)} \geqslant V^{(1)} \geqslant 0$, for $k = 0, 1, \ldots$; with $U^* \equiv (I + X^*)P_1$ and $V^* \equiv P_2(D^- + D^+ X^*)$;*
(*ii*) *$U^{(k)} \to U^*$, $V^{(k)} \to V^*$ as $k \to \infty$;*
(*iii*) *for each k, $0 \leqslant U_{\mathcal{S}}^{(k)} \leqslant U_{\mathcal{M}}^{(k)} \leqslant U_{\mathcal{G}}^{(k)}$ and $0 \leqslant V_{\mathcal{S}}^{(k)} \leqslant V_{\mathcal{M}}^{(k)} \leqslant V_{\mathcal{G}}^{(k)}$; and*
(*iv*) *for each k, $0 \leqslant U_{\mathcal{S}}^{(k)} \leqslant U_{\mathcal{J}}^{(k)} \leqslant U_{\mathcal{G}}^{(k)}$ and $0 \leqslant V_{\mathcal{S}}^{(k)} \leqslant V_{\mathcal{J}}^{(k)} \leqslant V_{\mathcal{G}}^{(k)}$.*

Proof. First we prove the iterates are well-defined. For SI and MSI, the issue is trivial. For NBJ and NBGS, the formulae imply

$$U_{\mathcal{J}}^{(k+1)} = P_1 \left[I - \frac{\hat{s}}{2} V_{\mathcal{J}}^{(k)} P_1\right]^{-1}, \quad V_{\mathcal{J}}^{(k+1)} = \left[I - \frac{\hat{s}}{2} P_2 D^+ U_{\mathcal{J}}^{(k)}\right]^{-1} P_2 D^- \tag{22}$$

and

$$U_{\mathcal{G}}^{(k+1)} = P_1 \left[I - \frac{\hat{s}}{2} V_{\mathcal{G}}^{(k)} P_1\right]^{-1}, \quad V_{\mathcal{G}}^{(k+1)} = \left[I - \frac{\hat{s}}{2} P_2 D^+ U_{\mathcal{G}}^{(k+1)}\right]^{-1} P_2 D^-. \tag{23}$$

The matrices inside the square brackets in (22) and (23) can be proved to be nonsingular M-matrices in the form $I - K$ by induction, with nonnegative inverses. In particular, we need to show that $\|K\| < \frac{1}{2}$. Note that (4) implies that $\hat{s} \leqslant \frac{1}{2}$ when $b = 0$.

For $k = 0$, $U_{\mathcal{J}}^{(1)}, V_{\mathcal{J}}^{(1)}$ and $U_{\mathcal{G}}^{(1)}$ are obviously well-defined, as $K = 0$ in their respective formulae. For NBGS with $K \equiv \frac{\hat{s}}{2} P_2 D^+ U_{\mathcal{G}}^{(1)}$, (21) and (23) imply

$$\|U_{\mathcal{G}}^{(1)}\| = \|P_1\| = 1, \quad \|K\| \leqslant \frac{\hat{s}}{2} \leqslant \frac{1}{4} < \frac{1}{2}$$

so $V_{\mathcal{G}}^{(1)} = (I - K)^{-1} P_2 D^-$ is well-defined as $I - K$ is a nonsingular M-matrix. In addition, from the second formula in (23), we have

$$\left\|V_{\mathcal{G}}^{(1)}\right\| = \left\|(I + K + K^2 + \cdots) P_2 D^-\right\| \leqslant \left\|P_2 D^-\right\| \left(1 + \frac{1}{4} + \frac{1}{4^2} + \cdots\right) \leqslant \frac{4}{3} < 2. \tag{24}$$

For the induction step, assume that $U_{\mathcal{I}}^{(k)}$ and $V_{\mathcal{I}}^{(k)}$ are well-defined, with $\|U_{\mathcal{I}}^{(k)}\|, \|V_{\mathcal{I}}^{(k)}\| < 2$ ($\mathcal{I} = \mathcal{J}, \mathcal{G}$). For NBJ and NBGS, we obviously have $\|K\| < \frac{1}{2}$, and $\|U_{\mathcal{I}}^{(k+1)}\|, \|V_{\mathcal{J}}^{(k+1)}\| < 2$ ($\mathcal{I} = \mathcal{J}, \mathcal{G}$). For

$V_{\mathcal{G}}^{(k+1)}$, applying a similar argument as in (24) with $K = \frac{\hat{s}}{2} P_2 D^+ U_{\mathcal{G}}^{(k+1)}$, we complete the induction step with

$$\left\| V_{\mathcal{G}}^{(k+1)} \right\| = \left\| (I + K + K^2 + \cdots) P_2 D^- \right\| < \left\| P_2 D^- \right\| \left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots \right) \leqslant 2.$$

Consequently, (i) and (ii) can be proved similarly as in Theorem 3.1.

For (iii) and (iv), again by induction, we have

$$U_{\mathcal{I}}^{(1)} = P_1 \ \ (\forall \mathcal{I}), \quad V_{\mathcal{I}}^{(1)} = P_2 D^- \ \ (\mathcal{I} = \mathcal{S}, \mathcal{M}, \mathcal{J}).$$

For $V_{\mathcal{G}}^{(1)}$ with $K = \frac{\hat{s}}{2} P_2 D^+ U_{\mathcal{G}}^{(1)}$, we have

$$V_{\mathcal{G}}^{(1)} = (I - K)^{-1} P_2 D^- = (I + K + K^2 + \cdots) P_2 D^- \geqslant P_2 D^- = V_{\mathcal{I}}^{(1)} \ \ (\mathcal{I} = \mathcal{S}, \mathcal{M}, \mathcal{J}).$$

For the induction step, assume that (iii) and (iv) hold for some arbitrary value of k. We then have

$$\begin{aligned}
U_{\mathcal{S}}^{(k+1)} &= \left[I + \frac{\hat{s}}{2} U_{\mathcal{S}}^{(k)} V_{\mathcal{S}}^{(k)} \right] P_1 \leqslant \left[I + \frac{\hat{s}}{2} U_{\mathcal{M}}^{(k)} V_{\mathcal{M}}^{(k)} \right] P_1 = U_{\mathcal{M}}^{(k+1)}, \\
V_{\mathcal{S}}^{(k+1)} &= P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{S}}^{(k)} V_{\mathcal{S}}^{(k)} \right] \leqslant P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{M}}^{(k)} V_{\mathcal{M}}^{(k)} \right] \\
&\leqslant P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{M}}^{(k+1)} V_{\mathcal{M}}^{(k)} \right] = V_{\mathcal{M}}^{(k+1)}; \\
U_{\mathcal{S}}^{(k+1)} &= \left[I + \frac{\hat{s}}{2} U_{\mathcal{S}}^{(k)} V_{\mathcal{S}}^{(k)} \right] P_1 \leqslant \left[I + \frac{\hat{s}}{2} U_{\mathcal{J}}^{(k)} V_{\mathcal{J}}^{(k)} \right] P_1 \\
&\leqslant \left[I + \frac{\hat{s}}{2} U_{\mathcal{J}}^{(k+1)} V_{\mathcal{J}}^{(k)} \right] P_1 = U_{\mathcal{J}}^{(k+1)}, \\
V_{\mathcal{S}}^{(k+1)} &= P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{S}}^{(k)} V_{\mathcal{S}}^{(k)} \right] \leqslant P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{J}}^{(k)} V_{\mathcal{J}}^{(k+1)} \right] = V_{\mathcal{J}}^{(k+1)}; \\
U_{\mathcal{M}}^{(k+1)} &= \left[I + \frac{\hat{s}}{2} U_{\mathcal{M}}^{(k)} V_{\mathcal{M}}^{(k)} \right] P_1 \leqslant \left[I + \frac{\hat{s}}{2} U_{\mathcal{G}}^{(k+1)} V_{\mathcal{G}}^{(k)} \right] P_1 = U_{\mathcal{G}}^{(k+1)}, \\
V_{\mathcal{M}}^{(k+1)} &= P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{M}}^{(k+1)} V_{\mathcal{M}}^{(k)} \right] \leqslant P_2 \left[D^- + \frac{\hat{s}}{2} D^+ U_{\mathcal{G}}^{(k+1)} V_{\mathcal{G}}^{(k+1)} \right] = V_{\mathcal{G}}^{(k+1)};
\end{aligned}$$

and

$$\begin{aligned}
U_{\mathcal{J}}^{(k+1)} &= P_1 \left[I - \frac{\hat{s}}{2} V_{\mathcal{J}}^{(k)} P_1 \right]^{-1} \leqslant P_1 \left[I - \frac{\hat{s}}{2} V_{\mathcal{G}}^{(k)} P_1 \right]^{-1} = U_{\mathcal{G}}^{(k+1)}, \\
V_{\mathcal{J}}^{(k+1)} &= \left[I - \frac{\hat{s}}{2} P_2 D^+ U_{\mathcal{J}}^{(k)} \right]^{-1} P_2 D^- \leqslant \left[I - \frac{\hat{s}}{2} P_2 D^+ U_{\mathcal{G}}^{(k+1)} \right]^{-1} P_2 D^- = V_{\mathcal{G}}^{(k+1)}.
\end{aligned}$$

Note that the iterates $U^{(k)}$ and $V^{(k)}$ are increasing towards their respective limits U^* and V^*, and the right-most inequalities will be obvious from $(I - K)^{-1} = I + K + K^2 + \cdots$ The induction is complete and (iii) and (iv) are proved. □

After U and V are obtained, X can be retrieved from (18). As for the simpler equation in [1,10,13,14,15,17], the speed of convergence for various iterative schemes is reflected by the rates of increase in the iterates. Consequently, NBGS is the fastest method, as proven more elaborately in [10]. Note that these iterative schemes are of $O(r^2 n)$ complexity per iteration, with the inexpensive inversion of

$I - K \in \mathbb{R}^{r\times r}$ in NBJ and NBGS. (Only when $X = \frac{\hat{s}}{2}UV$ is formed, $O(rn^2)$ flops are involved.) However, all these are possible only for the special case $b = 0$ and $s \neq 0$, so we shall not attempt similar analysis as in [10]. Note that the iterative schemes for the simpler equations previously studied in [10] are of $O(n^2)$ complexity.

3.3. Explicit solution for special case when $s = 0$

When $s = 0$, the NARE (6) becomes the quadratic

$$\hat{b}D^- - 2X + \hat{b}X^2 = 0. \tag{25}$$

With the minimal nonnegative solution guaranteed to exist by Theorem 2.3, we may consider various iterative processes for solving (25). However, the fixed-point iteration $X \leftarrow \hat{b}(D^- + X^2)/2$ implies that $X^* = \mathrm{diag}\{x_i\}$ is diagonal and (25) degenerates to n scalar quadratics, implying that

$$X^* = \hat{b}^{-1}\left(I - \sqrt{I - \hat{b}^2D^-}\right) = \hat{b}D^-\left(I + \sqrt{I - \hat{b}^2D^-}\right)^{-1}.$$

As $D^{\pm} = O(1/n)$, a good approximation to X^* is given by $\hat{b}D^-/2$, identical to $X^{(1)}$ from the fixed-point iteration with $X^{(0)} = 0$ or solving (25) by ignoring the X^2 term.

4. The general case

For the general NARE (6):

$$B^- - XF^- - F^+X + XB^+X = 0$$

with $F^{\pm} \equiv I - \hat{s}PD^{\pm}$, $B^- \equiv (\hat{b}I + \hat{s}P)D^-$ and $B^+ \equiv \hat{b}I + \hat{s}PD^+$ in (7) for $b \neq 0$, we can apply fixed-point iteration, Newton's method [8,14,15] or doubling [4,11]. The existence of the unique minimal nonnegative solution X^* has been proved in Theorem 2.3. Similar results as in Theorems 3.1 and 3.2 can be proved. For $P = P_1P_2$ of rank r ($< n$), the additional structure can be exploited for lower operation counts.

4.1. Fixed-point iteration

There are many different versions of fixed-point iterations for (6). One obvious way, extending (17) in Section 3.1, is

$$X \leftarrow F(X) \equiv \frac{\hat{s}}{2}(I + X)P(D^- + D^+X) + \frac{\hat{b}}{2}(D^- + X^2), \quad X^{(0)} = 0. \tag{26}$$

Note that we have written $F(X)$ as the sum of the right-hand-side of (17) associated with $\hat{s}$ and the left-over $\hat{b}$ term, requiring one less matrix–matrix multiplication than the obvious

$$\widetilde{F}(X) = \frac{1}{2}\left[B^- - \hat{s}(XPD^- + PD^+X) + XB^+X\right].$$

For $P = P_1P_2$ of rank r ($< n$), only $[(6r + 4)n^2 + 2n^3]$ flops are required per iteration (see Table 5.1 for other operation counts).

Similar to Theorem 3.1, we have the following special case of [6, Theorem 2.3]:

Theorem 4.1. *Under assumption (11), for the fixed-point iteration (26), we have*

(*i*) *the iterates satisfy* $X^* \geqslant X^{(k+1)} \geqslant X^{(k)} \geqslant X^{(1)} = \frac{1}{2}B^- \geqslant 0$ $(k = 0, 1, \ldots)$; *and*
(*ii*) $X^{(k)} \to X^*$ *as* $k \to \infty$.

For the rate of convergence, please consult [6, Theorem 2.5].

4.2. Newton's method

From the NARE (6), let $R(X)$ denote the left-hand-side of the equation, in the computationally efficient form

$$R(X) = \hat{s}(I + X)P(D^- + D^+X) + \hat{b}(D^- + X^2) - 2X,$$

with further saving when $P = P_1P_2$ is low-ranked.

At the $(k+1)$th iteration with $X^{(k)}$ being an approximate solution and $X^{(k+1)} = X^{(k)} + \delta X^{(k)}$, Newton's method requires the solution of the Sylvester equation

$$\left(F^+ - X^{(k)}B^+\right)\delta X^{(k)} + \delta X^{(k)}\left(F^- - B^+X^{(k)}\right) = R\left(X^{(k)}\right). \tag{27}$$

Convergence of Newton's method follows readily.

Theorem 4.2. *Let X^* be the minimal nonnegative solution of (6). Then under assumption (11), for the Newton iteration (27) with $X^{(0)} = 0$, the sequence $\{X^{(k)}\}$ is well-defined, $X^{(k)} \leqslant X^{(k+1)} \leqslant X^*$ for all $k \geqslant 0$, and $\lim_{i\to\infty} X^{(k)} = X^*$.*

The proof makes use of selected results from Theorem 2.2. In particular when vectorized, the above Sylvester operator can be written as the matrix operator M_{X^*} (with $m = n$).

4.3. Doubling

We shall quote the doubling algorithm for the general NARE (9), with the matrix M in (10) being a nonsingular M-matrix, from [11]. Note that the doubling algorithm is at approximately three times faster than Newton's method, as concluded in [7,11] and Table 5.1; please consult the details in the respective references.

For the general NARE:

$$X\widehat{C}X - X\widehat{D} - \widehat{A}X + \widehat{B} = 0$$

with the corresponding matrix M in (10) being a nonsingular M-matrix, we first transform $\widehat{A}, \widehat{B}, \widehat{C}$ and $\widehat{D}$ to

$$E_\gamma = I - 2\gamma V_\gamma^{-1},\ G_\gamma = 2\gamma D_\gamma^{-1}\widehat{C}W_\gamma^{-1},\ F_\gamma = I - 2\gamma W_\gamma^{-1},\ H_\gamma = 2\gamma W_\gamma^{-1}\widehat{B}D_\gamma^{-1}$$

with the parameter $\gamma \geqslant \max_{i,j}\{\hat{a}_{ii}, \hat{b}_{jj}\}$ and

$$A_\gamma = \widehat{A} + \gamma I,\ D_\gamma = \widehat{D} + \gamma I,\ W_\gamma = A_\gamma - \widehat{B}D_\gamma^{-1}\widehat{C},\ V_\gamma = D_\gamma - \widehat{C}A_\gamma^{-1}\widehat{B}.$$

The doubling algorithm can then be summarized as:

$$\begin{aligned}
&E_0 = W_\gamma,\ F_0 = F_\gamma,\ G_0 = G_\gamma,\ H_0 = H_\gamma,\\
&E_{k+1} = E_k(I - G_kH_k)^{-1}E_k,\ F_{k+1} = F_k(I - H_kG_k)^{-1}F_k,\\
&G_{k+1} = G_k + E_k(I - G_kH_k)^{-1}G_kF_k,\ H_{k+1} = H_k + F_k(I - H_kG_k)^{-1}H_kE_k.
\end{aligned} \tag{28}$$

The iterates are well-defined with $I - H_kG_k$ and $I - G_kH_k$ being nonsingular M-matrices for all k, and $H_k \to X$ and $G_k \to Y$ (respectively, the solutions to (9) and its adjoint) from below quadratically as $k \to \infty$ (see [11, Theorem 5.1]).

When $D^\pm = D$, we have $F^\pm = I - \hat{s}PD$ and $B^\pm = \hat{s}PD$, halving the operation count of doubling.

5. Numerical examples

For comparison, we shall summarize the operation counts per iteration of various iterative methods in Table 5.1. We shall show only the dominant terms, assuming that $n \gg r$. The Sylvester equations

Table 5.1
Operation counts per iteration.

Parameter	P	Method	Flops
$b = 0$	General	Fixed-point iteration for (17)	$4n^3$
	Low-ranked	NBGS (23)	$6r^2n$
$b \neq 0$	General	Fixed-point iteration (26)	$6n^3$
	Low-ranked	Fixed-point iteration (26)	$2n^3$
	General	Newton's method (27)	$41n^3$
	Low-ranked	Newton's method (27)	$34n^3$
		Doubling (28)	$16\frac{2}{3}n^3$

Table 5.2
CPU-times and iteration numbers for Example 1.

	Fixed-point			Newton			Doubling		
n	t_n	r_n	#It	t_n	r_n	#It	t_n	r_n	#It
64	0.125	N/A	42	0.062	N/A	6	0.047	N/A	7
128	0.374	2.99	38	0.421	6.79	5	0.156	3.32	7
256	2.886	7.72	38	2.558	6.08	5	1.435	9.20	7
512	18.80	6.51	40	21.75	8.50	5	10.76	7.50	7
1024	186.9	9.94	43	172.1	7.91	5	97.25	9.04	8

in (27) are assumed to be solved by the Bartels–Stewart algorithm [5]. For $b = 0$ with a low-ranked P, only the fastest method NBGS is considered. The slow fixed-point iteration method is also included for comparison.

We shall consider two randomly generated examples for $n = 64, 128, 256, 512$ and 1024. Example 1 has $\hat{s} = 0.3$, $\hat{b} = 0.4$ and P being full-ranked, and Example 2 has $\hat{s} = 0.3$, $\hat{b} = 0$ and rank $P = 10$. For the examples, the respective assumptions in Theorems 2.3 and 3.2 are satisfied. The numerical computation has been carried out using MATLAB R2008b on a laptop with eps $= 2.2204 \times 10^{-16}$ [16].

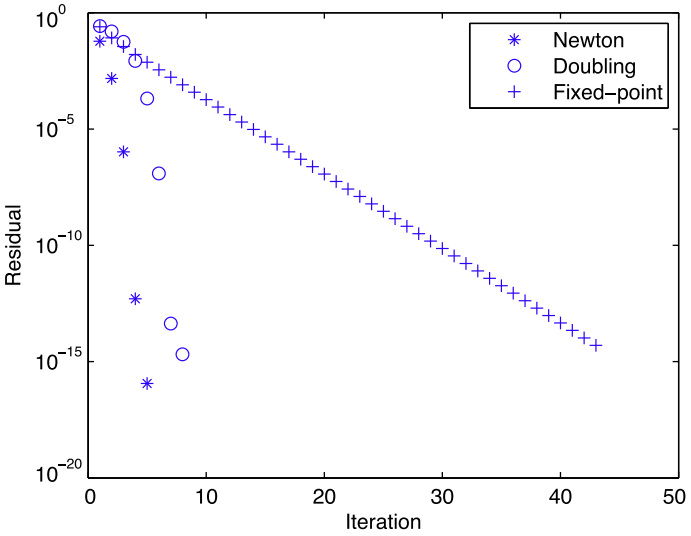

Fig. 3. Residuals for Example 1 ($n = 1024$).

Table 5.3
CPU-times and iteration numbers for Example 2 with NBGS.

n	64	128	256	512	1024
t_n	0.0000	0.0000	0.0312	0.0468	0.0780
r_n	N/A	N/A	N/A	1.50	1.67
#It	9	9	9	9	9

For Example 1, fixed-point iteration, Newton's method and the doubling algorithm have been compared for various values of n. The iterations have been run until convergence with tolerance tol $= 10^{-15}$. The results are summarized in Table 5.2, with t_n denoting the CPU-time, $r_n \equiv t_n/t_{\frac{n}{2}}$ and #It the number of iterations required, for a particular value of n. The iterates are also plotted in Fig. 3 for $n = 1024$.

From Table 5.2 and Fig. 3, it is evident that the doubling algorithm performs better than Newton's method and the fixed-point iteration method is the slowest, as predicted in Table 5.1. The ratios r_n illustrate the $O(n^3)$ complexity of the methods. The graphs in Fig. 3 illustrate the quadratic convergence of the doubling algorithm and Newton's method, with fixed-point iteration obviously converges linearly. Newton's method is faster than doubling in terms of convergence but the latter has an advantage in operation count per iteration by a factor of three, resulting in its better efficiency in terms of CPU-time. Note that the `cputime` command in MATLAB [16] is not an exact reflection of CPU-time consumed and should be used as a rough guide only. Also, we have no control over some parts of the algorithms, such as the inversion of the Sylvester operators by the MATLAB command `lyap` [16] in Newton's method.

For Example 2, only the fastest iteration method NBGS has been tested and the results are summarized in Table 5.3, with tol $= 10^{-15}$. The $O(n)$ complexity of the method is illustrated in the ratios r_n, although cputime in MATLAB fails to register the small amount of CPU-time for smaller values of n.

6. Concluding remarks

For a two-dimensional model in transport theory, we need to solve an integral–differential equation to obtain the Eastern reflection kernel R_E, from which other kernels can be derived. For the steady-state solution, we have derived an NARE from the corresponding integral equation using numerical quadratures. We have proved the existence and uniqueness of the minimal nonnegative solution of the NARE. When $b = 0$ and P is low-ranked, the efficient NBGS method of complexity $O(n)$ solves the NARE efficiently. For the general case when $b \neq 0$, the doubling algorithm is the most efficient, approximately three times more efficient as Newton's method. The numerical results support our theoretical findings.

For future work, we need to consider conditions for existence other than (11), efficient algorithms making better use of the structure of the Riccati equations, the convergence of $X = RD^{-}$ to $\mathcal{R}_E$ and to improve the efficiency of the numerical algorithms for large values of n. Finally, there are other similar models and problems in transport theory [2] worthy of investigation.

Acknowledgements

E. Chu was partially supported by the Centre of Mathematical Modeling and Scientific Computing and the National Centre of Theoretical Sciences, National Chiao Tung University, Republic of China. T. Li was supported by the National Science Foundation, Republic of China.

References

[1] Z.-Z. Bai, Y.-H. Gao, L.-Z. Lu, Fast iterative schemes for nonsymmetric algebraic Riccati equations arising from transport theory, SIAM J. Sci. Comput. 30 (2) (2008) 804–818.
[2] R.E. Bellman, G.M. Wing, An Introduction to Invariant Imbedding, Wiley, New York, 1975.

[3] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, SIAM, Philadelphia, 1994.
[4] C.-Y. Chiang, E.K.-W. Chu, C.-H. Guo, T.-M. Huang, W.-W. Lin, S.-F. Xu, Convergence analysis of the doubling algorithm for several nonlinear matrix equations in the critical case, SIAM J. Matrix Anal. Appl. 31 (2009) 227–247.
[5] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[6] C.-H. Guo, Nonsymmetric algebraic Riccati equations and Wiener–Hopf factorization for M-matrices, SIAM J. Matrix Anal. Appl. 23 (2001) 225–242.
[7] C.-H. Guo, A new class of nonsymmetric algebraic Riccati equations, Linear Algebra Appl. 426 (2007) 636–649.
[8] C.-H. Guo, N.J. Higham, Iterative solution of a nonsymmetric algebraic Riccati equation, SIAM J. Matrix Anal. Appl. 29 (2007) 396–412.
[9] C.-H. Guo, P. Lancaster, Analysis and modification of Newton's method for algebraic Riccati equations, Math. Comp. 67 (1998) 1089–1105.
[10] C.-H. Guo, W.-W. Lin, Convergence rates of some iterative methods for nonsymmetric algebraic Riccati equations arising in transport theory, Linear Algebra Appl. 432 (2010) 283–291.
[11] X.-X. Guo, W.-W. Lin, S.-F. Xu, A structure-preserving doubling algorithm for nonsymmetric algebraic Riccati equations, Numer. Math. 103 (2006) 393–412.
[12] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.
[13] J. Juang, Existence of algebraic matrix Riccati equations arising in transport theory, Linear Algebra Appl. 230 (1995) 89–100.
[14] Y. Lin, L. Bao, Y. Wei, A modified Newton method for solving non-symmetric algebraic Riccati equations arising in transport theory, SIMA J. Numer. Anal. 29 (2008) 215–224.
[15] L.-Z. Lu, Newton iterations for a non-symmetric algebraic Riccati equation, Numer. Linear Algebra Appl. 12 (2005) 191–200.
[16] MathWorks, MATLAB User's Guide, 2002.
[17] V. Mehrmann, H. Xu, Explicit solutions for a Riccati equations from transport theory, SIAM J. Matrix Anal. 30 (2008) 1339–1357.
[18] P. Nelson, D.L. Seth, Integrodifferential equations for the two-Dimensional transition kernels of invariant imbedding, Appl. Math. Comp. 82 (1997) 67–83.
[19] P. Nelson, D.L. Seth, R. Vasudevan, An integrodifferential equation for the two-dimensional reflection kernel, Appl. Math. Comp. 49 (1992) 1–18.