A subspace shift technique for solving close-to-critical nonsymmetric algebraic Riccati equations

Bruno Iannazzo[1] and Federico Poloni[2]

November 8, 2010

[1]: Dipartimento di Matematica e Informatica. Via Vanvitelli 1, 06123 Perugia, Italy.
bruno.iannazzo@dmi.unipg.it
[2]: Scuola Normale Superiore. Piazza dei Cavalieri 7, 56126 Pisa, Italy. f.poloni@sns.it

The worst situation in computing the minimal nonnegative solution X_* of a nonsymmetric algebraic Riccati equation $\mathcal{R}(X) = 0$ associated with an M-matrix occurs when the derivative of $\mathcal{R}$ at X_* is near to a singular matrix. When the derivative of $\mathcal{R}$ at X_* is singular, the problem is ill-conditioned and the convergence of the algorithms based on matrix iterations is slow; however, there exist some techniques to remove the singularity and restore well-conditioning and fast convergence. This phenomenon is partially shown also in the close-to-critical case, but the techniques used for the null recurrent case cannot be applied to this setting.

We present a new method to accelerate the convergence and amend the conditioning in close-to-critical cases. The numerical experiments confirm the efficiency of the new method.

Keywords: nonsymmetric algebraic Riccati equation, fluid queue, doubling algorithm, shift technique

1 Introduction

We consider the nonsymmetric algebraic Riccati equation (or NARE)

$$0 = \mathcal{R}(X) := XCX - AX - XD + B, \tag{1}$$

where $X, B \in \mathbb{C}^{m\times n}$, $A \in \mathbb{C}^{m\times m}$, $C \in \mathbb{C}^{n\times m}$, $D \in \mathbb{C}^{n\times n}$.

In certain applications in queueing models [22] and in the numerical solution of transport equations [19], the coefficients of (1) are such that

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

is an M-matrix, either nonsingular or singular irreducible. In this case, we give Equation (1) the acronym M-NARE. We recall that $M \in \mathbb{C}^{n\times n}$ is an M-matrix if it can be written in the form $M = sI_n - N$, where I_n is the identity matrix of size n (denoted also by I if there is no ambiguity), N is a matrix whose elements are nonnegative, for which we use the notation $N \geqslant 0$, and $s \geqslant \rho(N)$, where $\rho(\cdot)$ is the spectral radius of a square matrix. The M-matrix M is singular if $s = \rho(N)$ and nonsingular if $s > \rho(N)$. It can be proved that the eigenvalues of an M-matrix have nonnegative real part [2].

The solutions of the NARE (1) can be put in correspondence with certain n-dimensional invariant subspaces of the matrix

$$\mathcal{H} = \begin{bmatrix} D & -C \\ B & -A \end{bmatrix}. \tag{2}$$

More precisely, a matrix $X \in \mathbb{C}^{m\times n}$ is a solution of (1) if and only if the columns of $\begin{bmatrix} I_n \\ X \end{bmatrix}$ span an invariant subspace of $\mathcal{H}$, in particular it holds that

$$\mathcal{H} \begin{bmatrix} I_n \\ X \end{bmatrix} = \begin{bmatrix} I_n \\ X \end{bmatrix} (D - CX), \tag{3}$$

and the eigenvalues of $D - CX$ are a subset of the eigenvalues of $\mathcal{H}$.

We say that the NARE (1) is *associated with* the matrix $\mathcal{H}$ of (2). Observe that any 2×2 block matrix with square diagonal blocks yields a NARE associated with it.

In the case of an M-NARE, where $\mathcal{H} = \mathcal{J}\mathcal{M}$ for $\mathcal{J} = \begin{bmatrix} I_n & 0 \\ 0 & -I_m \end{bmatrix}$, it can be proved (see [4] and the references therein) that the eigenvalues of $\mathcal{H}$ can be ordered by non increasing real part such that

$$\Re\lambda_1 \geqslant \cdots \geqslant \Re\lambda_{n-1} > \lambda_n \geqslant 0 \geqslant \lambda_{n+1} > \cdots \geqslant \Re\lambda_{m+n},$$

that is, n eigenvalues belong to the closed right half complex plane and the others to the closed left half plane, and the *central eigenvalues*, λ_n and λ_{n+1}, are real and separated from the other eigenvalues. If $\lambda_n = 0 = \lambda_{n+1}$, then there exists only one linearly independent eigenvector for the zero eigenvalue, that is, there is a Jordan block of size 2 relative to the zero eigenvalue in the Jordan canonical form of $\mathcal{H}$.

For these reasons, the matrix $\mathcal{H}$ associated with an M-NARE has a unique n-dimensional invariant subspace corresponding to the n *rightmost* eigenvalues, namely $\lambda_1, \ldots, \lambda_n$, which we call the n-dimensional *semi-unstable invariant subspace* of $\mathcal{H}$ (the term comes from the theory of the symmetric algebraic Riccati equations in dynamical systems [20]).

In the applications, the required solution of the M-NARE is the one for which the columns of $\begin{bmatrix} I_n \\ X \end{bmatrix}$ span the semi-unstable n-dimensional invariant subspace of $\mathcal{H}$, or, equivalently, such that the eigenvalues of $D - CX$ are the n rightmost eigenvalues of $\mathcal{H}$. This solution has been proved to exist and it turns out to be the minimal element-wise nonnegative solution of (1) (see [11]).

When $\mathcal{M}$ is singular irreducible, at least one among λ_n and λ_{n+1} is zero and $\mathcal{M} = \rho(N)I - N$, for some irreducible nonnegative matrix N. By the Perron–Frobenius theorem [2] and the irreducibility assumption, $\ker \mathcal{M}$ and $\ker \mathcal{M}^T$ are one dimensional, spanned by two vectors with positive entries which we call $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ and $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, respectively. We have $\mathcal{M}v = \mathcal{H}v = 0$ and $u^T\mathcal{M} = (u^T\mathcal{J})\mathcal{H} = 0$. We define the *drift* of the Riccati equation as

$$\mu = -u^T\mathcal{J}v = u_2^T v_2 - u_1^T v_1. \tag{4}$$

It can be proved that [4, 11]

- $\mu < 0$ if and only if $\lambda_n = 0 > \lambda_{n+1}$ (positive recurrent case);
- $\mu = 0$ if and only if $\lambda_n = 0 = \lambda_{n+1}$ (null recurrent case);
- $\mu > 0$ if and only if $\lambda_n > 0 = \lambda_{n+1}$ (transient case).

The terms drift, transient, positive and null recurrent come from the fluid queue model where the M-NARE first appeared [22].

Equation (1) is usually solved either by some matrix iteration, e.g., the Cyclic Reduction (CR) [3] or the Structure-preserving Doubling Algorithm (SDA) [8, 14], whose limits yield the required solution or using the ordered Schur form of $\mathcal{H}$ [12].

Both the conditioning of the equation and the convergence speed of the iterations are strictly related to the relative gap between the central eigenvalues of $\mathcal{H}$, i.e., $(\lambda_n - \lambda_{n+1})/\|\mathcal{H}\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm. If $\lambda_n = \lambda_{n+1}$, then the minimal nonnegative solution of equation (1) is ill-conditioned [13] and the convergence of iterations such as CR and SDA, which is quadratic in the generic case, turns to linear [7]. We speak of critical case, since in these cases the required solution X is critical, namely $\mathcal{R}'(X)$ is singular, where $\mathcal{R}'(X)$ is the Fréchet derivative of the operator $\mathcal{R}(X)$.

In such cases, the shift technique of [14, 18] has proved to be useful. It consists in making a special rank-one correction of $\mathcal{H}$, obtaining a new Riccati equation with the same minimal solution. The new equation has better conditioning and the convergence of iterations is quadratic again.

However, ill-conditioning and slow convergence appear also in the close-to-null recurrent case, or, in terms of the central eigenvalues, when $\lambda_n \approx 0, \lambda_{n+1} \approx 0$. This is the worst-case scenario since the numerical solution of the matrix equations is problematic and the use of the shift technique is not recommended since it relies on the computation of an ill-conditioned eigenvector and the sign of the drift μ. When $\mathcal{M}$ is a singular irreducible M-matrix, it is also hard to determine the sign of the drift μ, and thus to classify the queue in the fluid queue models.

These difficulties are the main motivation for this work in which we present a new technique to handle the close-to-null recurrent case. The technique relies on the fact that, for the M-NARE, there exists a unique 2-dimensional invariant subspace of $\mathcal{H}$ associated with the eigenvalues λ_n and λ_{n+1}, which we call the 2-dimensional *central invariant subspace* of $\mathcal{H}$. It is spanned by the two eigenvectors corresponding to λ_n and λ_{n+1}, when $\lambda_n \neq \lambda_{n+1}$ or by the Jordan chain associated to 0 in the critical case.

If the central eigenvalues λ_n and λ_{n+1} of $\mathcal{H}$ are close to each other but well separated from the other eigenvalues of $\mathcal{H}$, then, while the eigenvectors corresponding to λ_n and λ_{n+1} are ill-conditioned, the 2-dimensional central invariant subspace shows good conditioning.

The proposed technique is based on a rank-two modification of $\mathcal{H}$ made by means of the central invariant subspace. We move the central eigenvalues together and we obtain a new NARE, with the same solution as the original or a solution which is a rank-one modification of the solution of the original equation. The new equation has better conditioning and, in certain cases, the convergence of iterations is much faster.

The paper is organized as follows. In Section 2, we review some basic results on NAREs, especially regarding their relation to invariant subspaces. In Section 3, we show how the conditioning of the equation is related to the gap between the central eigenvalues of a NARE. In Sections 4 and 5, we present respectively the Structured Doubling Algorithm and the shift technique, two important tools for the numerical solution of NAREs. In Section 6, we introduce the subspace shift technique, and show how it affects the semi-stable and semi-stable solutions of the Riccati equation. In Section 7, we arrive to an algorithm, and discuss its implementation details and variants. Finally, Section 8 contains our numerical experiments on the subspace shift algorithm.

In the following, $\sigma(M)$ stands for the set of the eigenvalues of $M \in \mathbb{C}^{n\times n}$, and $\|\cdot\|_F$ denotes the Frobenius norm.

2 Invariant subspaces, solvability, and the dual equation

Let $\mathcal{R}(X) = 0$ be the M-NARE associated with $\mathcal{H}$ of (2), and let $\lambda_1, \ldots, \lambda_{m+n}$ be the eigenvalues of $\mathcal{H}$, counted with their algebraic multiplicity and ordered by nonincreasing real part. In the following, we need to deal only with the following cases. Notice that in all of them there is a *splitting* of the eigenvalues with respect to the imaginary axis, i.e., $\Re\lambda_n \geqslant 0 \geqslant \Re\lambda_{n+1}$.

nonsingular case $\Re\lambda_n > 0 > \Re\lambda_{n+1}$;

transient case $\Re\lambda_n > 0 = \Re\lambda_{n+1}$;

positive recurrent case $\Re\lambda_n = 0 > \Re\lambda_{n+1}$;

null recurrent case $\Re\lambda_{n-1} > \Re\lambda_n = 0 = \Re\lambda_{n+1} > \Re\lambda_{n+2}$, and λ_n and λ_{n+1} are the eigenvalues corresponding to a 2×2 Jordan block with eigenvalue 0.

In all the cases above, the invariant subspaces $\mathcal{V}_{\rm su}$ and $\mathcal{V}_{\rm ss}$ associated with the eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ and $\{\lambda_{n+1}, \ldots, \lambda_{n+m}\}$, respectively, are well-defined. We call them *semi-unstable* and *semi-stable* subspace, respectively.

Theorem 1 ([11, 20])**.** *Let*

$$V_{su} = \begin{bmatrix} V_{su}^{(1)} \\ V_{su}^{(2)} \end{bmatrix}, \qquad V_{su}^{(1)} \in \mathbb{R}^{n\times n}, \qquad V_{su}^{(2)} \in \mathbb{R}^{m\times n}$$

be a matrix such that $\mathcal{V}_{su} = \operatorname{span} V_{su}$*. The NARE* (1) *associated with the matrix* $\mathcal{H}$ *of* (2) *admits a solution* X_* *such that* $\sigma(D - CX_*) = \{\lambda_1, \ldots, \lambda_n\}$ *if and only if* $V_{su}^{(1)}$ *is nonsingular, and in this case* $X_* = V_{su}^{(2)} \left(V_{su}^{(1)}\right)^{-1}$*.*

We call X_* the *semi-unstable* solution.
The NARE

$$0 = \mathcal{D}(Y) := YBY - YA - DY + C, \qquad Y \in \mathbb{R}^{n\times m}, \tag{5}$$

is called the *dual equation* to (1). There is a counterpart of Theorem 1 for the dual equation.

Theorem 2 ([11])**.** *Let*

$$V_{ss} = \begin{bmatrix} V_{ss}^{(1)} \\ V_{ss}^{(2)} \end{bmatrix}, \qquad V_{ss}^{(1)} \in \mathbb{R}^{n\times m}, \qquad V_{ss}^{(2)} \in \mathbb{R}^{m\times m}$$

be a matrix such that $\mathcal{V}_{ss} = \operatorname{span} V_{ss}$*. The dual equation admits a solution* Y_* *such that* $\sigma(BY_* - A) = \{\lambda_{n+1}, \ldots, \lambda_{n+m}\}$ *if and only if* $V_{ss}^{(2)}$ *is nonsingular, and in this case* $Y_* = V_{ss}^{(1)} \left(V_{ss}^{(2)}\right)^{-1}$*.*

Notice that

$$\begin{bmatrix} I_n \\ X_* \end{bmatrix}, \qquad \begin{bmatrix} Y_* \\ I_m \end{bmatrix}, \tag{6}$$

span the semi-unstable and semi-stable subspaces, respectively. The two subspaces have zero intersection, unless $\lambda_n = \lambda_{n+1}$, when the unique (up to a scalar multiple) independent eigenvector of $\mathcal{H}$ corresponding to λ_n belongs to both of them.

In the case of M-NAREs, one can prove the existence of X_* and Y_* by a direct argument [15].

3 Gap and conditioning of the Riccati equation

We recall from [10, 23] the classical definition of *separation* between the two square matrices M and N

$$\operatorname{sep}(M, N) := \min_{X\neq 0} \frac{\|MX - XN\|_F}{\|X\|_F},$$

and the bound

$$\operatorname{sep}(M,N) \leqslant \min_{\mu\in\sigma(M),\nu\in\sigma(N)} |\mu-\nu|\,.$$

Let

$$\mathcal{U} = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}, \qquad U_{11}\in\mathbb{R}^{n\times n}, \qquad U_{22}\in\mathbb{R}^{m\times m}$$

be an orthogonal matrix such that

$$\mathcal{U}^T\mathcal{H}\mathcal{U} = \begin{bmatrix} G_{11} & G_{12} \\ 0 & G_{22} \end{bmatrix}, \qquad G_{11}\in\mathbb{C}^{n\times n},$$

where $\mathcal{H}$ is as in (2) and $\sigma(G_{11}) = \{\lambda_1,\lambda_2,\ldots,\lambda_n\}$, then the columns of $\begin{bmatrix} U_{11} \\ U_{21} \end{bmatrix}$ span the semi-unstable space, and thus by Theorem 1 the minimal solution of (1) is $X_* = U_{21}U_{11}^{-1}$.

Let

$$\widetilde{\mathcal{H}} = \mathcal{H} + \Delta\mathcal{H}$$

be a perturbation of $\mathcal{H}$. For $\Delta\mathcal{H}$ sufficiently small $J\widetilde{\mathcal{H}}$ is an M-matrix and the NARE associated with $\widetilde{\mathcal{H}}$ is an M-NARE whose minimal nonnegative solution is denoted by $\widetilde{X}_*$. Then, the following result holds.

Theorem 3 ([16]). *Let $\mathcal{U}^T\Delta\mathcal{H}\mathcal{U}$ be conformably partitioned as*

$$\mathcal{U}^T\Delta\mathcal{H}\mathcal{U} = \begin{bmatrix} \Delta G_{11} & \Delta G_{12} \\ \Delta G_{21} & \Delta G_{22} \end{bmatrix},$$

and

$$\delta = \operatorname{sep}(G_{11},G_{22}) - (\|\Delta G_{11}\|_F + \|\Delta G_{22}\|_F).$$

If $\operatorname{sep}(G_{11},G_{22}) > 0$ and the perturbation $\Delta\mathcal{H}$ is sufficiently small, then

$$\left\|\widetilde{X}^* - X^*\right\|_F \leqslant \frac{2\sqrt{2}\left\|U_{11}^{-1}\right\|_2\|\Delta\mathcal{H}\|_F}{\delta - 2\sqrt{2}\left\|U_{11}^{-1}\right\|_2\|\Delta\mathcal{H}\|_F}\sqrt{1+\|X_*\|_2^2}.$$

This result shows that we may regard to the quantity

$$\frac{\sqrt{1+\|X_*\|_2^2}}{\|X_*\|_F}\frac{2\sqrt{2}\left\|U_{11}^{-1}\right\|_2\|\mathcal{H}\|_F}{\operatorname{sep}(G_{11},G_{22})} \tag{7}$$

as a (Frobenius-norm) condition number for X_*. Since $\sigma(G_{22}) = \{\lambda_{n+1},\lambda_{n+2},\ldots,\lambda_{m+n}\}$, this condition number is necessarily large if the relative gap $(\lambda_n - \lambda_{n+1})/\|\mathcal{H}\|_F$ is small. Therefore, a close-to-critical NARE is always ill-conditioned.

4 The doubling algorithm

In this section we review the Structure-preserving Doubling Algorithm (or SDA) and its application for computing the minimal nonnegative solution X of the M-NARE (1).

4.1 Matrix pencils

We recall some definitions used for dealing with matrix pencils. Given $P, Q \in \mathbb{C}^{n\times n}$, if there exists a full rank matrix $V \in \mathbb{C}^{n\times s}$ and a matrix $\Lambda \in \mathbb{C}^{s\times s}$ such that $QV = PV\Lambda$, we say that the columns of V span an *(s-dimensional) deflating subspace* for the pencil $Pz - Q$. The invariant subspaces of a matrix Q turn out to be deflating subspaces for the pencil $Iz - Q$. If $\varphi(z) = \det(Pz - Q)$ is not identically zero, we say that $Pz - Q$ is a *regular* pencil. The *(generalized) eigenvalues* of the regular pencil $Pz - Q$ are defined as the roots of $\varphi(z)$, complemented with $n - k$ eigenvalues at infinity if $\varphi(z)$ has degree k lower than n.

4.2 Cayley transform

The *Cayley transform* of parameter $0 \neq \gamma \in \mathbb{R}$ is the map

$$\mathcal{C}_\gamma : z \mapsto \frac{z-\gamma}{z+\gamma}.$$

Notice that, for $\gamma > 0$, $\mathcal{C}_\gamma$ maps the open (closed) right half-plane onto the open (closed) unit circle, and the open (closed) left half-plane onto the exterior of the open (closed) unit circle.

One can extend the map $\mathcal{C}_\gamma$ to matrices in a trivial way: if $\gamma \notin \sigma(\mathcal{H})$, then

$$\mathcal{C}_\gamma(\mathcal{H}) := (\mathcal{H} - \gamma I)(\mathcal{H} + \gamma I)^{-1}.$$

If $\gamma \in \sigma(\mathcal{H})$, one can define $\mathcal{C}_\gamma(\mathcal{H})$ only as a pencil $(\mathcal{H} + \gamma I)z - (\mathcal{H} - \gamma I)$.

The matrix $\mathcal{C}_\gamma(\mathcal{H})$ has the same right invariant subspaces as $\mathcal{H}$, while the associated eigenvalues are transformed according to $\mu_i = \mathcal{C}_\gamma(\lambda_i)$, while $\lambda_1, \ldots, \lambda_{n+m}$ are the eigenvalues of $\mathcal{H}$, thus $\sigma(\mathcal{C}_\gamma(\mathcal{H})) = \{\mu_1, \mu_2, \ldots, \mu_{n+m}\}$. In particular, $\mu_1, \ldots, \mu_n$ lie inside the (closed) unit disc, and $\mu_{n+1}, \ldots, \mu_{n+m}$ lie outside. A single or double eigenvalue at 0, if present, is mapped to a single or double one at 1 (the precise statement about the Jordan canonical form of the function of a matrix can be found in [21, Theorem 9.4.7]).

Similarly, the matrix pencil $(\mathcal{H} - \gamma I) - z(\mathcal{H} + \gamma I)$, has generalized eigenvalues $\mu_i = \mathcal{C}_\gamma(\lambda_i)$ for $\lambda_i \neq \gamma$ together with eigenvalues at infinity up to the multiplicity of γ as eigenvalue of $\mathcal{H}$ and the same right deflating subspaces as $\mathcal{H}$. Moreover, if 0 is a double eigenvalue of $\mathcal{H}$ corresponding to a Jordan chain of length 2, then 1 is a double eigenvalue of $(\mathcal{H} - \gamma I) - z(\mathcal{H} + \gamma I)$ corresponding to a Jordan chain of length 2 (for the definition of Jordan chain for a matrix polynomial see [9]).

Therefore, the four cases appearing in the beginning of Section 2 are mapped by the Cayley transform to the following possibilities for $\mathcal{C}_\gamma(\mathcal{H})$ and $(\mathcal{H} - \gamma I) - z(\mathcal{H} + \gamma I)$. Notice that, in general, the μ_i are not ordered by increasing modulus.

nonsingular case $|\mu_i| < 1$ for all $i \leqslant n$ and $|\mu_j| > 1$ for all $j \geqslant n+1$;

transient case $|\mu_i| < 1$ for all $i \leqslant n$, $\mu_{n+1} = 1$ and $|\mu_j| > 1$ for all $j \geqslant n+2$;

positive recurrent case $|\mu_i| < 1$ for all $i \leqslant n-1$, $\mu_n = 1$ and $|\mu_j| > 1$ for all $j \geqslant n+1$;

null recurrent case $|\mu_i| < 1$ for all $i \leqslant n-1$, $\mu_n = \mu_{n+1} = 1$ and $|\mu_j| > 1$ for all $j \geqslant n+2$, and μ_n and μ_{n+1} are the eigenvalues corresponding to a Jordan chain of length 2 for the eigenvalue 1.

We say that matrices $\mathcal{A} \in \mathbb{R}^{(m+n)\times(m+n)}$ (or pencils $\mathcal{E}z - \mathcal{A}$, with $\mathcal{E}, \mathcal{A} \in \mathbb{R}^{(m+n)\times(m+n)}$) have a (n, m) *d-splitting* if their eigenvalues satisfy one of the previous four set of inequalities. The semi-unstable and semi-stable subspaces of $\mathcal{H}$ are mapped to the (well-defined) n-dimensional invariant/deflating subspace associated with $\{\mu_1, \mu_2, \ldots, \mu_n\}$ and the m-dimensional invariant/deflating subspace associated with $\{\mu_{n+1}, \mu_{n+2}, \ldots, \mu_{n+m}\}$, respectively. which we call the *d-semi-stable* and *d-semi-unstable* subspaces.

Notice the slight confusion deriving from the fact that the unstable subspace becomes the d-stable one and vice versa; however, this is necessary if we wish to be consistent with the classical terminology regarding discrete- and continuous-time stability in dynamical systems.

4.3 Outline of SDA

The structured doubling algorithm, in the formulation of [17], is a system of rational matrix iterations defined by

$$\begin{aligned} E_{k+1} &= E_k(I - G_k H_k)^{-1} E_k, \\ F_{k+1} &= F_k(I - H_k G_k)^{-1} F_k, \\ G_{k+1} &= G_k + E_k(I - G_k H_k)^{-1} G_k F_k, \\ F_{k+1} &= H_k + F_k(I - H_k G_k)^{-1} H_k E_k, \end{aligned} \tag{8}$$

with suitable initial values $E_0 \in \mathbb{C}^{n\times n}$, $F_0 \in \mathbb{C}^{m\times m}$, $G_0 \in \mathbb{C}^{n\times m}$, $H_0 \in C^{m\times n}$.

Let $L, M \in \mathbb{C}^{(m+n)\times(m+n)}$ have the structure

$$L = \begin{bmatrix} I_n & -G_0 \\ 0 & F_0 \end{bmatrix}, \quad M = \begin{bmatrix} E_0 & 0 \\ -H_0 & I_m \end{bmatrix}, \tag{9}$$

and be such that the pencil $Lz - M$ has a (n, m) d-splitting in the sense of the previous section.

The SDA can be seen as an algorithm to compute a special basis for two special deflating subspaces of $Lz - M$. Namely, if we define $G_\infty = \lim G_k$ and $H_\infty = \lim H_k$, then

$$\begin{bmatrix} I \\ H_\infty \end{bmatrix}, \qquad \begin{bmatrix} G_\infty \\ I \end{bmatrix} \tag{10}$$

span the right d-semi-stable and d-semi-unstable right deflating subspace of the pencil $Lz - M$ [17]. Observe that it may happen that such kind of bases do not exist, or that some of the matrices to be inverted in (8) are singular, in that case we say that the SDA cannot be applied. In the noncritical case, the two corresponding left deflating subspaces are respectively $\begin{bmatrix} I & -G_\infty \end{bmatrix}$ and $\begin{bmatrix} -H_\infty & I \end{bmatrix}$.

In the case of the M-NARE, the initial values of the SDA can be chosen as

$$\begin{aligned} E_0 =& I - 2\gamma V_\gamma^{-1}, & F_0 =& I - 2\gamma W_\gamma^{-1}, \\ G_0 =& 2\gamma D_\gamma^{-1} C W_\gamma^{-1}, & H_0 =& 2\gamma W_\gamma^{-1} B D_\gamma^{-1}, \\ A_\gamma =& A + \gamma I, & D_\gamma =& D + \gamma I, \\ W_\gamma =& A_\gamma - B D_\gamma^{-1} C, & V_\gamma =& D_\gamma - C A_\gamma^{-1} B, \end{aligned} \tag{11}$$

for a suitable $\gamma > 0$. One can verify that the initial values (11) form a pencil $Lz - M$ whose generalized eigenvalues and right deflating subspaces are precisely the eigenvalues and right invariant subspaces of $(H + \gamma I)z - (H - \gamma I)$; thus, by comparing (10) and (6), one sees that G_∞ and H_∞ are the minimal solutions of (1) and (5) respectively.

The following result ensures the applicability of the SDA in our setting.

Theorem 4 ([14])**.** *In an M-NARE, the SDA can be applied without breakdown (i.e., the quantities $(I - G_k H_k)^{-1}$ and $(I - H_k G_k)^{-1}$ exist at each step), and converges monotonically to the minimal solution (i.e., it holds that $0 \leqslant H_k \leqslant H_{k+1} \leqslant X_*$ and $H_k \to X_*$), provided*

$$\gamma \geqslant \gamma_* = \max\left\{ \max_{1 \leqslant i \leqslant m} a_{ii}, \max_{1 \leqslant i \leqslant n} d_{ii} \right\}. \tag{12}$$

Regarding the convergence rate, we have the following result. Notice that all its hypotheses are satisfied in the case of M-NAREs.

Theorem 5 ([14, 7])**.** *Suppose that we are in one of the four cases in Section 2, the solutions X_* and Y_* exist, the SDA converges and $\gamma \geqslant \gamma_*$ where γ_* is defined in* (12)*. The convergence of the SDA is linear with rate* $1/2$ *in the null recurrent case, and quadratic with rate*

$$\nu = \frac{\max_{i=1,\dots,n} |\mathcal{C}_\gamma(\lambda_i)|}{\min_{j=1,\dots,m} |\mathcal{C}_\gamma(\lambda_{n+j})|} \tag{13}$$

in the other three cases, where λ_i are the eigenvalues of $\mathcal{H}$ different from γ. Moreover, the value of $\gamma \geqslant \gamma_*$ that yields faster convergence is $\gamma = \gamma_*$.

4.4 Central eigenvalues and SDA convergence speed

In order to provide a stricter relation between the two central eigenvalues and the convergence speed, we prove that the minimum and maximum appearing in (13) are attained by the central eigenvalues, if γ is chosen as in (12).

Theorem 6. *Consider the SDA algorithm for an M-NARE, and let γ be chosen according to* (12)*. The minimum and the maximum in* (13) *are attained by $i = n$ and $j = 1$, i.e., the two eigenvalues responsible for the convergence rate of SDA are the central eigenvalues.*

We establish the result with a geometrical reasoning. Let us first assess the following lemma.

Lemma 7. *Let Γ be a closed disc in the complex plane with center $C \in \mathbb{R}$ and radius r. The point in Γ with maximal modulus is one among $C + r$ and $C - r$.*

Proof. (of Lemma 7) Let $C + p \in \mathbb{C}$, $|p| \leqslant r$, be a generic point in the disc. By the triangle inequality, $|C + p| \leqslant |C| + |p| \leqslant |C| + r$, with equality if and only if $|p| = r$ and p has the same argument as C, i.e., either real positive or negative. □

Proof. (of Theorem 4.4) From (12) we have $\gamma I - D \geqslant 0$ and thus $P = \gamma I - D + CX_* \geqslant 0$. Hence we may write $D - CX_* = \gamma I - P$; from the Perron–Frobenius theory of M-matrices, it follows that all the eigenvalues of $D - CX_*$ are contained in the closed disc with center γ and radius $r = \gamma - \lambda_n$; and in particular, the eigenvalue λ_n lies on its boundary. We call this disc Γ, and proceed to prove that $\mathcal{C}_\gamma(\lambda_n)$ has the maximal modulus among all points in in $\mathcal{C}_\gamma(\Gamma)$. The image of Γ under the Cayley transform is a closed disc Γ', which must be contained in the unit disc and symmetric with respect to the real axis. This means that its center (which is not in general $\mathcal{C}_\gamma(\gamma)$) is real. This disc Γ' intersects the real axis in the two points $\mathcal{C}_\gamma(\lambda_n)$ and $\mathcal{C}_\gamma(2\gamma - \lambda_n)$. By the lemma, the point of maximal modulus in Γ' is one among them; direct computation (using $\lambda_n \leqslant \gamma$) shows that it is the former.

A similar reasoning starting from $A - X_* C$ yields that $\min_{j=1,\dots,m} |\mathcal{C}_\gamma(\lambda_{n+j})|$ is achieved by $j = 1$; we need some extra care with the signs, as $\lambda_{n+1} \leqslant 0$, and with the fact that this time the image of the enclosing disc under the Cayley transform is the *outside* of a suitable disc. □

This means that we may replace (13) with

$$\nu = \frac{|\mathcal{C}_\gamma(\lambda_n)|}{|\mathcal{C}_\gamma(\lambda_{n+1})|}.$$

By continuity, this ratio tends to 1 whenever $\lambda_{n+1} - \lambda_n \to 0$, for a fixed γ ; thus, a small gap implies a slow convergence rate for SDA.

5 The shift technique for the M-NARE

The shift technique has been applied in [14] to the M-NARE (1) where $\mathcal{M}$ is a singular irreducible M-matrix, that is, when at least one between λ_n and λ_{n+1} is 0.

Without loss of generality one can assume that $\lambda_n = 0$: the case $\lambda_n > 0 = \lambda_{n+1}$ can be reduced to the case $\lambda_n = 0$ by a simple trick. In fact [14, Lemma 5.1] the matrix X is the minimal nonnegative solution of (1) if and only if $Z = X^T$ is the minimal nonnegative solution of the equation

$$ZC^T Z - ZA^T - D^T Z + B^T = 0; \tag{14}$$

where (1) has positive drift (transient case) if and only if (14) has negative drift (positive recurrent case).

The shift technique is rooted in the following results.

Lemma 8 (Brauer's theorem [6]). *Let (λ, v) be an eigenpair for the matrix T. Let u be a vector with $u^T v = 1$ and s be a scalar. The eigenvalues of the matrix $\widehat{T} := T + svu^T$ are the same as those of T, except λ which is replaced by $\lambda + s$.*

Theorem 9 ([14]). *Let $\mathcal{H}$ be the as in (2) associated with the M-NARE (1) with $\lambda_n = 0$, and let v_n be the eigenvector relative to λ_n; consider the matrix*

$$\widehat{\mathcal{H}} := \mathcal{H} + sv_n u^T,$$

with $u^T v_n = 1$ and $s > 0$. Then, the minimal solution $\widehat{X}_*$ of the NARE associated with $\widehat{\mathcal{H}}$ coincides with the minimal solution X_* of the M-NARE associated with $\mathcal{H}$.

The shift technique consists in computing the eigenvector v_n corresponding to the eigenvalue $\lambda_n = 0$, and using it to construct the modified NARE as in Theorem 9. The matrix $\widehat{\mathcal{H}}$ has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{n-1}, \widehat{\lambda}_n$ where $\widehat{\lambda}_n = s$ (the eigenvalue λ_n has been "shifted" from 0 to s, this justifies the name of the technique). Observe that the gap of $\widehat{\mathcal{H}}$ is larger than that of $\mathcal{H}$; thus, better conditioning and faster convergence of SDA are expected, for a fixed Cayley parameter γ. Numerical experiments [14] show that this technique reduces dramatically the number of steps of the SDA in the critical and close-to-critical case. In the critical case, the convergence from linear becomes quadratic.

However, it can happen that the Riccati equation associated with $\widehat{\mathcal{H}}$ is not an M-NARE; that is, $\widehat{\mathcal{M}} = \mathcal{J}\widehat{\mathcal{H}}$ need not be an M-matrix. Hence, there is no guarantee that the SDA can be performed without a breakdown, even if in practice this method works well and the applicability of SDA is usually assumed [14].

Notice that, since $\lambda_n = 0$, the vector v_n can be computed as $\ker \mathcal{M}$. In principle, the shift technique could be used also for nonsingular M-matrices, i.e., the hypothesis $\lambda_n = 0$ is not actually needed in Theorem 9. Different techniques are needed for the computation of λ_n and v_n in this case; for instance, the power method.

However, in the close-to-critical case the eigenvector v_n is ill-conditioned and therefore it cannot be computed with good accuracy. When in addition $\mathcal{M}$ is singular irreducible, it is not easy to discriminate between $\lambda_n > 0$ and $\lambda_n = 0$, or to identify the critical cases.

To overcome this problem, we present a shift technique which moves together λ_n and λ_{n+1}, without computing explicitly the eigenvectors corresponding to them, but working with the whole 2-dimensional invariant subspace associated with λ_n and λ_{n+1}, which typically shows better conditioning. We call this subspace *central subspace* of $\mathcal{H}$.

6 The subspace shift technique

We describe a modification of the shift technique that is especially tailored for the case in which the gap of the M-NARE is small, i.e., $\mathcal{H}$ has two eigenvalues λ_n and λ_{n+1} very close (or equal) to 0.

The idea of the subspace shift technique is the following. First, we perform a suitable rank-two modification on the matrix $\mathcal{H}$ obtaining a new matrix $\widehat{\mathcal{H}}$ with the same splitting as $\mathcal{H}$ but with a larger gap between the central eigenvalues. This rank-two modification of $\mathcal{H}$ is built using the left and right invariant subspaces corresponding to the central

eigenvalues. Then, using the blocks of $\widehat{\mathcal{H}}$, we derive a new Riccati equation, which we call $\widehat{\mathcal{R}}(X) = 0$. A numerical method (we used the SDA in our examples, but any other algorithm is fine) is then used to compute the solution $\widehat{X}_*$ of $\widehat{\mathcal{R}}(X) = 0$ associated with the invariant subspace of the eigenvalues contained in the right half-plane. Finally, the required solution X_* of the original NARE is obtained from $\widehat{X}_*$ by adding a rank-1 correction.

6.1 Invariant subspaces of $\widehat{\mathcal{H}}$

Let $\mathcal{V}$ (resp. $\mathcal{U}$) be the invariant subspace generated by the right (resp. left) eigenvectors v_n and v_{n+1} (resp. u_n and u_{n+1}) relative to the eigenvalues λ_n and λ_{n+1}. Let $V \in \mathbb{R}^{(m+n)\times 2}$ (resp. $U \in \mathbb{R}^{(m+n)\times 2}$) be an orthonormal basis of $\mathcal{V}$ (resp. $\mathcal{U}$); then there exists $\Lambda \in \mathbb{R}^{2\times 2}$ such that $HV = V\Lambda$, and premultiplying both sides by V^T we get $\Lambda = V^T HV$.

Let now

$$\widehat{\mathcal{H}} := \mathcal{H} + VSU^T, \tag{15}$$

for a matrix $S \in \mathbb{R}^{2\times 2}$.

Theorem 10. *Let $\mathcal{H}$ be as in* (2) *associated with the M-NARE* (1)*, and let $\widehat{\mathcal{H}}$ be defined according to* (15)*. The eigenvalues, eigenvectors, and invariant subspaces of $\widehat{\mathcal{H}}$ (corresponding to noncentral eigenvalues) are the same as those of $\mathcal{H}$, the central subspace of $\mathcal{H}$, that is the invariant subspace $\mathcal{H}V = V\Lambda$ corresponding to the central eigenvalues λ_n, λ_{n+1} is replaced by*

$$\widehat{\mathcal{H}}V = V\widehat{\Lambda}, \qquad\qquad \widehat{\Lambda} = \Lambda + SU^TV.$$

If $\Lambda + SU^TV$ is nondefective, with two eigenpairs $(\widehat{\lambda}_i, x_i)$, $i = n, n+1$, then $(\widehat{\lambda}_i, Vx_i)$ are two eigenpairs of $\widehat{\mathcal{H}}$. If it is defective, with a Jordan chain y_n, y_{n+1}, then Vy_n, Vy_{n+1} is a Jordan pair for $\mathcal{H}$ with the same generalized eigenvalue.

Proof. Suppose $\mathcal{H}W = WM$, and $\lambda_n, \lambda_{n+1} \notin \sigma(M)$. Then, $U^TW = 0$: as λ_n and λ_{n+1} are separated from the other eigenvalues in $\mathcal{H}$ of a M-NARE, U^T and W span respectively the left and right invariant subspace relative to different eigenvalues, and thus are orthogonal. This implies that $\widehat{\mathcal{H}}W = \mathcal{H}W = WM$, that is, W spans an invariant subspace with the same eigenvalues for $\widehat{\mathcal{H}}$ as well.

On the other hand, for the central subspace V we have

$$\widehat{\mathcal{H}}V = (\mathcal{H} + VSU^T)V = V(\Lambda + SU^TV).$$

Multiplying the latter equation by x_i or y_i yields the last statement of the theorem. □

We aim to find a choice of S so that the matrix $\widehat{\Lambda}$ has one positive and one negative eigenvalue, so that the matrix $\widehat{\mathcal{H}}$ has an eigenvalue splitting with respect to the imaginary axis and we fall in one of the four cases of eigenvalue splitting.

We suggest two strategies for enforcing this.

1. Choosing S so that $\det(\widehat{\Lambda}) < 0$. In this case, the two roots of the characteristic polynomial are guaranteed to be real and with different signs. Since U^TV is nonsingular (left and right eigenvector matrices corresponding to separated eigenvalues), we may choose S to obtain any matrix in place of $\widehat{\Lambda}$.

2. In particular, choosing $S = s\Lambda(U^TV)^{-1}$ for a scalar $s > 0$, we obtain $\widehat{\Lambda} = (s+1)\Lambda$. As we see in the following, this is a special situation since the eigenvectors of Λ and $\widehat{\Lambda}$ coincide, and thus the remainder of the solution algorithm is greatly simplified.

6.2 Solution form

Once $\widehat{\mathcal{H}}$ is constructed, one can define a new algebraic Riccati equation $\widehat{\mathcal{R}}(X) = 0$ associated with $\widehat{\mathcal{H}}$; as in the case of the simple shift, $\widehat{\mathcal{R}}(X) = 0$ need not be an M-NARE, and thus convergence is not guaranteed.

Theorem 11. *Let v be the left eigenvector of $\mathcal{H}$ corresponding to an eigenvector λ with $\Re\lambda \geqslant 0$, and let $\widehat{v}$ and $\widehat{u}$ be a left and right eigenvector of $\widehat{\mathcal{H}}$ relative to $\widehat{\lambda}$ with $\Re\widehat{\lambda} \geqslant 0$. Suppose that the Riccati equation associated with $\widehat{H}$ has semi-unstable solution $\widehat{X}_*$. Let*

$$w^T := \widehat{u}_n^T \begin{bmatrix} I \\ \widehat{X}_* \end{bmatrix}, \qquad \Delta v := v - \widehat{v} = \begin{bmatrix} \Delta v_1 \\ \Delta v_2 \end{bmatrix} \text{ with } \Delta v_1 \in \mathbb{R}^n, \qquad r := 1 + w^T \Delta v_1.$$

The Riccati equation associated with $\mathcal{H}$ has a semi-unstable solution X_* if and only if $r \neq 0$, and in this case it is given by

$$X_* = \widehat{X}_* + \frac{1}{r} z w^T, \qquad\qquad z = \begin{bmatrix} -\widehat{X}_* & I \end{bmatrix} \Delta v.$$

Proof. Let $v_1, \dots v_{n-1}, \widehat{v}$ be a set of Jordan chains spanning the semi-unstable space of $\widehat{\mathcal{H}}$. Then,

$$\begin{bmatrix} I \\ \widehat{X}_* \end{bmatrix} = \begin{bmatrix} v_1 & \dots v_{n-1} & \widehat{v} \end{bmatrix} Z$$

for a suitable nonsingular $Z \in \mathbb{R}^{n\times n}$. Since $\widehat{u}^T$ is the left eigenvector corresponding to $\widehat{\lambda}$, we have (up to a normalization factor which has no effect on the final formula) $\widehat{u}^T\widehat{v} = 1$ and $\widehat{u}^T v_i = 0$ for $i = 1, \dots, n-1$. Hence, left-multiplying by $\widehat{u}^T$ yields $e_n^T Z = w^T$. The semi-unstable space of $\mathcal{H}$ is spanned by $\{v_1, \dots, v_{n-1}, v\}$, thus a basis for it is given by

$$\begin{bmatrix} v_1 & \dots v_{n-1} & v \end{bmatrix} Z = \left(\begin{bmatrix} v_1 & \dots v_{n-1} & \widehat{v} \end{bmatrix} + (v - \widehat{v}) e_n^T \right) Z = \begin{bmatrix} I \\ \widehat{X}_* \end{bmatrix} + \Delta v w^T.$$

By Theorem 1, the semi-unstable solution X_* exists if and only if the leading $n \times n$ block, i.e. $I + \Delta v_1 w^T$, is nonsingular. By the Sherman–Morrison formula, said matrix is nonsingular if and only if $r = 0$, and in this case

$$(I + \Delta v_1 w^T)^{-1} = I - \frac{1}{r} \Delta v_1 w^T. \tag{16}$$

By the same theorem, the solution X_* has the form

$$X_* = (\widehat{X}_* + \Delta v_2 w^T)(I + \Delta v_1 w^T)^{-1};$$

now simple algebraic manipulations and (16) lead to the desired expression. □

7 The subspace shift algorithm

7.1 The basic algorithm

Using the formula of Theorem 11, we propose Algorithm 1 for computing the minimal solution of an M-NARE.

Algorithm 1 Subspace-shift algorithm for the solution of an M-NARE

1: Compute orthonormal bases U, V for the invariant subspaces corresponding to λ_n, λ_{n+1}
2: choose any S such that $\widehat{\Lambda} = \Lambda + SU^TV$ has negative determinant
3: compute $\widehat{\mathcal{H}} = \mathcal{H} + VSU^T$
4: solve the NARE $\widehat{\mathcal{R}}(X) = 0$ associated with $\widehat{\mathcal{H}}$.
5: compute the semi-stable eigenvectors x and $\widehat{x}$ for the 2×2 matrices Λ, $\widehat{\Lambda}$, and set $v = Vx$, $\widehat{v} = V\widehat{x}$; similarly, compute the semi-stable left eigenvector y^T of $M = U^T\widehat{\mathcal{H}}U$ and set $u^T = y^TU^T$
6: Recover the solution to the original NARE X_* from the solution $\widehat{X}_*$ using Theorem 11

It is possible to compute the solution to the dual equation with a similar formula, by making use of the unstable eigenvalues v_{u}, $\widehat{v}_{\mathrm{u}}$ and $\widehat{u}_{\mathrm{u}}^T$.

7.2 The case $\Delta v = 0$

As suggested in Section 6, the choice $S = s\Lambda(U^TV)^{-1}$ leads to a simpler development of the algorithm. Since Λ and $\widehat{\Lambda}$ have the same eigenvectors, $v = \widehat{v}$, and thus the rank-1 modifications to $\widehat{X}_*$ in Theorem 11 vanish, i.e., $X_* = \widehat{X}_*$. The new algorithm is exposed here as Algorithm 2. Notice that Algorithm 2, however, does not change the situation

Algorithm 2 Subspace shift algorithm for the solution of an M-NARE, in the case $\Delta v = 0$

1: Compute orthonormal bases U, V for the invariant subspaces corresponding to λ_n, λ_{n+1}
2: Compute $\widehat{\mathcal{H}} = \mathcal{H} + sVV^T\mathcal{H}V(U^TV)^{-1}U^T$
3: Solve the NARE $\widehat{\mathcal{R}}(X) = 0$ associated with $\widehat{\mathcal{H}}$, getting directly the solution of the original M-NARE

when the problem is exactly critical, since the two critical eigenvalues $\lambda_n = \lambda_{n+1} = 0$ do not change when multiplied by a constant. In this case, which is easy to identify given U and V, we may revert to the shift technique.

7.3 Central subspaces or smallest eigenvalues?

The algorithm above can be applied to any M-NARE without additional assumptions. However, two main issues arise.

- It is not clear how to compute the central invariant subspace. A natural candidate in the test problems is the inverse orthogonal iteration [10, Section 7.3.2], a linearly-convergent iteration which computes the invariant subspace relative to the eigenvalues with smallest modulus of a given matrix. However, λ_n and λ_{n+1} need not be the two smallest eigenvalues. In principle, we could have settings such as λ_n and λ_{n-1} very close to each other and to zero, and λ_{n+1} slightly larger in modulus.
- Even if we shift away the two central eigenvalues, the remaining ones (e.g., λ_{n-1} in a setting similar to the case above) could be very close to them, and thus the conditioning and SDA convergence speed are almost unchanged; therefore, the subspace shift can still be applied but is not much useful.

Therefore, it is useful to restrict our interest to the cases in which the method works best. Namely, we set

$$\varepsilon = \max(|\lambda_n|, |\lambda_{n+1}|), \qquad \delta = \min_{j \notin \{n,n+1\}} |\lambda_j|,$$

and focus on the cases in which $\varepsilon \neq 0$ and δ is larger than ε by a significant amount. Notice that, in this case, inverse orthogonal iteration converges to the correct subspace linearly with rate ε/δ, and the distance between the two central eigenvalues of $\widetilde{\mathcal{H}}$ is increased to at least 2δ. Notice that we can easily detect during the algorithm when these assumptions are not satisfied. An algorithm to perform all the relevant checks is reported here as Algorithm 3. In many applications [19, 1], our assumptions hold and

Algorithm 3 Subspace shift algorithm for the solution of an M-NARE, with a general subspace

1: Compute orthonormal bases U, V for the invariant subspaces corresponding to the two eigenvalues of $\mathcal{H}$ with smallest modulus, e.g., with inverse orthogonal iteration
2: If the orthogonal iteration converges too slowly (or does not converge), then there are eigenvalues close to the central ones: report failure
3: If $\det(V^T \mathcal{H} V) > 0$, then the two smallest eigenvalues are not the central ones: report failure
4: Continue as in Algorithm 1 or 2

subspace shift can be applied with computational advantage, as we see in the numerical experiments. Notice that Algorithm 2 apparently works even when $\det(V^T \mathcal{H} V) > 0$: in this case, we are simply shifting away (by multiplying them by $s+1$) two eigenvalues belonging to the same subspace, the d-semi-stable or the d-semi-unstable one. However, when this happens on a close-to-critical problem, then the subspace computed by the method is ill-conditioned (since there are at least three eigenvalues close to zero, and we are shifting away only two of them), and we do not expect the method to yield good results. Moreover, in this case the non-central eigenvalue that we are shifting is not guaranteed to be simple.

7.4 SDA for computing the central invariant subspace

As an alternative to the inverse orthogonal iteration, we may use another run of SDA to compute the central invariant subspace. SDA costs $O(n^3)$ per step, and converges quadratically, whereas the inverse orthogonal iteration costs $O(n^2)$ per step and converges linearly. We call this first application of the algorithm the *inner SDA*. Since in our setting λ_n and λ_{n+1} are the smallest eigenvalues of $\mathcal{H}$, there a disk $\mathcal{D}_r = \{z \in \mathbb{C} : |z| \leqslant r\}$ such that such that λ_n, λ_{n+1} lie in $\mathcal{D}_r$ and the other eigenvalues lie outside $\mathcal{D}_r$. We scale $\mathcal{H}$ by r and re-block the resulting matrix as

$$\mathcal{H}_r = r\mathcal{H} = \begin{bmatrix} D_r & -C_r \\ B_r & -A_r \end{bmatrix},$$

so that $D_r \in \mathbb{C}^{2\times 2}$ and $A_r \in \mathbb{C}^{(m+n-2)\times(m+n-2)}$. Notice that the d-semi-stable space of $\mathcal{H}_r$ is the one associated to λ_n and λ_{n+1}, due to our choice of r. Therefore, when there is no breakdown in SDA, the sequences given by (8), with

$$E_0 = D_r - C_r A_r^{-1} B_r, \quad F_0 = -A_r^{-1}, \quad G_0 = C_r A_r^{-1}, \quad H_0 = A_r^{-1} B_r,$$

are such that $\lim_k H_k = X_r$, $\lim_k G_k = Y_r$, where $\begin{bmatrix} I \\ X_r \end{bmatrix}$ and $\begin{bmatrix} I & -Y_r \end{bmatrix}$ span the right and left central invariant subspaces of $\mathcal{H}$, respectively. After computing them, it suffices to get orthogonal bases of these two subspaces via a QR factorization to get U and V as required in the algorithm.

At first sight, it seems that applying this algorithm requires knowing the value of r in advance. In fact, one sees that the multiplicative constant r factors nicely out of the SDA iteration, and only affects the magnitude of E_k and F_k. Namely, the initial values for a generic r are related to those for $r = 1$, which we denote by a (1) superscript, by

$$E_k = r^{2^k} E_k^{(1)}, \qquad F_k = r^{-2^k} F_k^{(1)}, \qquad G_k = G_k^{(1)}, \qquad H_k = H_k^{(1)}.$$

Therefore, we may compute the SDA iteration for $r = 1$, or generically for any value of r, since the values G_k and H_k do not depend on its choice. To avoid overflow and underflow, it may be useful to renormalize the iterates every few steps, setting

$$\alpha_k \leftarrow \frac{\rho(F_k)^{1/2}}{\rho(E_k)^{1/2}}, \qquad E_k \leftarrow \alpha_k E_k, \qquad F_k \leftarrow \alpha_k^{-1} F_k. \tag{17}$$

7.5 Magnitude of the shift

Another issue which appears in the practical implementation is the selection of s in Algorithm 2. If the chosen value is too small, then the two central eigenvalues do not move significantly and the gap remains small; on the other hand, if the shift is excessively large then $\left\|\widetilde{\mathcal{H}}\right\|_F$ grows, and the conditioning of the shifted Riccati equation degrades, according to (7).

Intuitively, larger values of s improve the conditioning as long as the shifted eigenvalues $(1+s)\lambda_n$ and $(1+s)\lambda_{n+1}$ are central eigenvalues for $\widetilde{\mathcal{H}}$, too. As soon as they become larger (in modulus) than λ_{n-1} and λ_{n+2}, then they are no longer central, and thus increasing them further does not affect directly our conditioning bounds. Therefore, we need to estimate how small they are with respect to the other eigenvalues. We may get an estimate using the convergence speed of the inner inverse power iteration or inner SDA. Notice that, with our assumptions on the eigenvalues, the convergence rate of both algorithms is determined by

$$t = \frac{\varepsilon}{\delta} = \frac{\max(\lambda_n, \lambda_{n+1})}{\delta};$$

hence a rough estimate for t is given by comparing two successive iterates of any of the two inner iterations. The values of λ_n, λ_{n+1} are easily computed, since they are the eigenvalues of the 2×2 matrix $V^T \mathcal{H} V$. We may solve for δ in the equation above, obtaining the magnitude of the smallest eigenvalue besides the central ones. If we choose s such that $(1+s)\min(\lambda_n, \lambda_{n+1}) \geqslant \delta$, then both $(1+s)\lambda_n$ and $(1+s)\lambda_{n+1}$ become larger (in modulus) than δ.

Similarly, for Algorithm 1, we may choose S to make $\widehat{\Lambda}$ any matrix, as pointed out above. In particular, we may ensure that both its eigenvalues have larger modulus than δ.

7.6 Randomization in the inner SDA

When the SDA is applied to a M-NARE, the existence of the two solutions X_* and Y_* is guaranteed by probabilistic arguments [15]. On the other hand, an additional issue that may arise with the inner SDA is that the d-semi-stable and d-semi-unstable invariant subspaces need not have bases in the form (6). To ensure that this happens with high probability, we conjugate $\mathcal{H}_r$ by a random orthogonal matrix before applying the algorithm.

The complete inner SDA used in the implementation is reported as Algorithm 4.

7.7 Solution of the shifted equation

Applying the subspace shift to an M-NARE yields a shifted equation $\widehat{R}(X) = 0$ which may not be an M-NARE. However, the corresponding matrix $\widehat{\mathcal{H}}$ has the same splitting as $\mathcal{H}$, and thus the SDA converges to the required solution (assuming the applicability).

So the SDA converges faster in the shifted case if the parameter γ of the Cayley transform is the same as the one used in the customary SDA, namely γ_* of (12). This remark is necessary since the shifted NARE loose the structure of M-NARE and thus Theorem 4 may not hold.

8 Numerical experiments

We present some numerical examples showing the effectiveness of the algorithms presented in Section 7 in solving close-to-null recurrent M-NARE, when the assumptions of

Algorithm 4 "Inner SDA" for the computation of the central invariant subspaces

1: generate a random $(n+m)\times(n+m)$ orthogonal matrix Q
2: set $\bar{\mathcal{H}} = QHQ^{-1}$
3: partition $\bar{\mathcal{H}} = \begin{bmatrix} \bar{D} & -\bar{C} \\ \bar{B} & -\bar{A} \end{bmatrix}$, with $\bar{D} \in \mathbb{R}^{2\times 2}$, $\bar{A} \in \mathbb{R}^{(m+n-2)\times(m+n-2)}$
4: Compute the starting values $E_0 = \bar{D} - \bar{C}\bar{A}^{-1}\bar{B}$, $F_0 = -\bar{A}^{-1}$, $G = \bar{C}\bar{A}^{-1}$, $H_0 = \bar{A}^{-1}\bar{B}$.
5: **while** G_k and H_k have not converged yet **do**
6: perform a SDA step (8)
7: if breakdown happens in the SDA step, report failure
8: if needed, renormalize E_k and F_k using the procedure (17)
9: **end while**
10: return $Q^{-1}\begin{bmatrix} I \\ H_\infty \end{bmatrix}$ and $\begin{bmatrix} I & -G_\infty \end{bmatrix} Q$

Section 7.3 are fulfilled; that is, when the two central eigenvalues are not both zero and the other eigenvalues are well separated from them. We recall that these assumptions can be identified dynamically by the algorithm.

We report the number of steps required by the customary application of the SDA to the NARE and the number of steps of the inner and outer SDA in the subspace shift algorithm. These steps are the most expensive part of the algorithms, since their asymptotic cost is cubic with respect to the size of the matrices; for instance, for $m = n$, the cost of a step of the SDA is $O(n^3)$ elementary arithmetic operations. The number of steps of the inner SDA is put in parentheses since the same quantities computed by the inner SDA can be also computed in principle by different, less expensive, algorithms.

We estimate the accuracy of the computed solution $\widetilde{X}_*$ by means of the relative error

$$\mathrm{err} = \frac{\left\|\widetilde{X}_* - X_*\right\|_F}{\|X_*\|_F},$$

if the exact solution X_* is available, elsewhere by means of the relative residual

$$\mathrm{res} = \frac{\left\|\mathcal{R}(\widetilde{X}_*)\right\|_F}{\left\|\widetilde{X}_* C \widetilde{X}_* + B\right\|_F + \left\|A\widetilde{X}_* + \widetilde{X}_* D\right\|_F}.$$

In our experiments the Frobenius norm is used.

Test 1. As a first test, we consider the close-to-critical cases of the transport problem treated in [15, 19, 5]. It is an M-NARE with square coefficients of size n and depending on two parameters $0 \leqslant \alpha < 1$ and $0 < c \leqslant 1$ (for the exact definition and the meaning of the parameters see [19]). The problem is critical for $(\alpha, c) = (0, 1)$, and it is close-to-critical if α and c approach simultaneously 0 and 1.

We measure the number of SDA iterations needed to get the best relative residual for several matrix sizes n and choices of the parameters β such that $\alpha = \beta$ and $c = 1 - \beta$.

n	β	gap	SDA its	SDA res	Alg 2 its	Alg 2 res
32	10^{-3}	-0.11	14	$8.8 \cdot 10^{-15}$	(5+)10	$4.0 \cdot 10^{-16}$
32	10^{-6}	$-3.5 \cdot 10^{-3}$	19	$1.0 \cdot 10^{-14}$	(4+)10	$1.1 \cdot 10^{-16}$
32	10^{-12}	$-3.5 \cdot 10^{-6}$	28	$8.1 \cdot 10^{-15}$	(3+)9	$1.1 \cdot 10^{-16}$
128	10^{-3}	-0.11	16	$1.2 \cdot 10^{-13}$	(5+)12	$7.9 \cdot 10^{-15}$
128	10^{-8}	$-3.5 \cdot 10^{-4}$	24	$1.4 \cdot 10^{-13}$	(4+)12	$2.1 \cdot 10^{-16}$

Table 1: Number of iterations for Algorithm 2 vs. SDA on the transport problem

As β approaches zero, the problem becomes close-to-critical; in fact β is strictly related to the relative gap which can be defined as $gap = |\lambda_n - \lambda_{n+1}|/\|\mathcal{H}\|$. The results are reported in Table 8. As one can see the problem is well suited to be solved by our algorithms since the central eigenvalues are well separated from the others.

Test 2. As a second example, we consider an M-NARE associated with a simple weakly transient Markov chain. It is a slight modification of an example of [1].

We define the matrix

$$\mathcal{H} = \begin{bmatrix} 3 & 0 & -1.5 & -1.5 \\ 0 & 3 & -2.9 & -0.1 \\ 2-p & 1 & -3 & p \\ 2-p & 1 & p & -3 \end{bmatrix},$$

such that $\mathcal{J}\mathcal{H}$ is a singular M-matrix for $0 \leqslant p \leqslant 2$. The corresponding M-NARE has the minimal nonnegative solution

$$X_* = \begin{bmatrix} (2-p)/3 & 1/3 \\ (2-p)/3 & 1/3 \end{bmatrix}.$$

Since the eigenvalues of $\mathcal{H}$ are $\{0, 3, p, -p-3\}$, the problem is close-to-critical as p approaches 0.

As before we measure the number of SDA iterations for several values of p together with the relative error. The results are reported in Table 2. As one can see the number of iteration required is dramatically reduced when p tends to 0. However, the accuracy of the solution will not increase. This fact suggest the use of the customary shift technique in singular cases.

References

[1] N. G. Bean, M. M. O'Reilly, and P. G. Taylor. Algorithms for return probabilities for stochastic fluid flows. *Stoch. Models*, 21(1):149–184, 2005.

[2] A. Berman and R. J. Plemmons. *Nonnegative matrices in the mathematical sciences*, volume 9 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. Revised reprint of the 1979 original.

p	SDA its	SDA err	Alg 2 its	Alg 2 err
0.1	9	$4.5 \cdot 10^{-15}$	(5+)4	$6.9 \cdot 10^{-15}$
10^{-2}	12	$1.0 \cdot 10^{-13}$	(4+)4	$3.7 \cdot 10^{-14}$
10^{-4}	18	$3.5 \cdot 10^{-13}$	(3+)4	$3.9 \cdot 10^{-12}$
10^{-8}	23	$3.5 \cdot 10^{-9}$	(3+)1	$1.0 \cdot 10^{-8}$

Table 2: Number of iterations for Algorithm 2 vs. SDA on a fluid queue problem

[3] D. Bini and B. Meini. On the solution of a nonlinear matrix equation arising in queueing problems. *SIAM J. Matrix Anal. Appl.*, 17(4):906–926, 1996.

[4] D. A. Bini, B. Iannazzo, B. Meini, and F. Poloni. Nonsymmetric Algerbraic Riccati Equations Associated with an M-Matrix: Recent Advances and Algorithms. In V. Olshevsky and E. Tyrtyshnikov, editors, *Matrix Methods: Theory, Algorithms and Applications*, pages 176–209. World Scientific, Singapore, 2010.

[5] D. A. Bini, B. Iannazzo, and F. Poloni. A fast Newton's method for a nonsymmetric algebraic Riccati equation. *SIAM J. Matrix Anal. Appl.*, 30(1):276–290, 2008.

[6] A. Brauer. Limits for the characteristic roots of a matrix. IV. Applications to stochastic matrices. *Duke Math. J.*, 19:75–91, 1952.

[7] C.-Y. Chiang, E. K.-W. Chu, C.-H. Guo, T.-M. Huang, W.-W. Lin, and S.-F. Xu. Convergence analysis of the doubling algorithm for several nonlinear matrix equations in the critical case. *SIAM J. Matrix Anal. Appl.*, 31(2):227–247, 2009.

[8] E. K.-W. Chu, H.-Y. Fan, and W.-W. Lin. A structure-preserving doubling algorithm for continuous-time algebraic Riccati equations. *Linear Algebra Appl.*, 396:55–80, 2005.

[9] I. Gohberg, P. Lancaster, and L. Rodman. *Invariant subspaces of matrices with applications*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006.

[10] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[11] C.-H. Guo. Nonsymmetric algebraic Riccati equations and Wiener-Hopf factorization for M-matrices. *SIAM J. Matrix Anal. Appl.*, 23(1):225–242, 2001.

[12] C.-H. Guo. Efficient methods for solving a nonsymmetric algebraic Riccati equation arising in stochastic fluid models. *J. Comput. Appl. Math.*, 192(2):353–373, 2006.

[13] C.-H. Guo and N. J. Higham. Iterative solution of a nonsymmetric algebraic Riccati equation. *SIAM J. Matrix Anal. Appl.*, 29(2):396–412, 2007.

[14] C.-H. Guo, B. Iannazzo, and B. Meini. On the doubling algorithm for a (shifted) nonsymmetric algebraic Riccati equation. *SIAM J. Matrix Anal. Appl.*, 29(4):1083–1100, 2007.

[15] C.-H. Guo and A. J. Laub. On the iterative solution of a class of nonsymmetric algebraic Riccati equations. *SIAM J. Matrix Anal. Appl.*, 22(2):376–391, 2000.

[16] X.-x. Guo and Z.-z. Bai. On the minimal nonnegative solution of nonsymmetric algebraic Riccati equation. *J. Comput. Math.*, 23(3):305–320, 2005.

[17] X.-X. Guo, W.-W. Lin, and S.-F. Xu. A structure-preserving doubling algorithm for nonsymmetric algebraic Riccati equation. *Numer. Math.*, 103(3):393–412, 2006.

[18] C. He, B. Meini, and N. H. Rhee. A shifted cyclic reduction algorithm for quasi-birth-death problems. *SIAM J. Matrix Anal. Appl.*, 23(3):673–691, 2001/02.

[19] J. Juang and W.-W. Lin. Nonsymmetric algebraic Riccati equations and Hamiltonian-like matrices. *SIAM J. Matrix Anal. Appl.*, 20(1):228–243, 1999.

[20] P. Lancaster and L. Rodman. *Algebraic Riccati equations*. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1995.

[21] P. Lancaster and M. Tismenetsky. *The theory of matrices*. Computer Science and Applied Mathematics. Academic Press Inc., Orlando, FL, second edition, 1985.

[22] L. C. G. Rogers. Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. *Ann. Appl. Probab.*, 4(2):390–413, 1994.

[23] G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Computer Science and Scientific Computing. Academic Press Inc., Boston, MA, 1990.