CONVERGENCE ANALYSIS OF THE DOUBLING ALGORITHM
FOR SEVERAL NONLINEAR MATRIX EQUATIONS IN THE
CRITICAL CASE

CHUN-YUEH CHIANG†, ERIC KING-WAH CHU‡, CHUN-HUA GUO§, TSUNG-MING
HUANG¶, WEN-WEI LIN∥, AND SHU-FANG XU∗∗

Abstract. In this paper, we review two types of doubling algorithm and some techniques for
analyzing them. We then use the techniques to study the doubling algorithm for three different
nonlinear matrix equations in the critical case. We show that the convergence of the doubling
algorithm is at least linear with rate $1/2$. As compared to earlier work on this topic, the results we
present here are more general, and the analysis here is much simpler.

Key words. nonlinear matrix equation, minimal nonnegative solution, maximal positive definite
solution, critical case, doubling algorithm, cyclic reduction, convergence rate

AMS subject classifications. 15A24, 15A48, 65F30, 65H10

DOI. 10.1137/080717304

1. Introduction. The doubling algorithm has been studied for various nonlinear
matrix equations in [1, 6, 7, 19, 21, 24, 27, 28, 34]. Its convergence behavior in the
critical case, however, has not been fully investigated. The doubling algorithm is said
to be structure-preserving (and denoted by SDA) because it preserves certain block
structures for matrix pairs (or pencils) related to matrix equations.

In section 2, we review two types of doubling algorithm and some techniques for
analyzing them. The presentation here is more general than in [34] and [24], to al-
low direct application to various matrix equations. In sections 3–5, the techniques
reviewed in section 2 are used to study the convergence behavior of the doubling
algorithm for three different nonlinear matrix equations in the critical case. As com-
pared to previous papers, the results here are obtained with only basic assumptions.
In particular, the results we obtain about a quadratic matrix equation arising from
quasi-birth-death processes are more general than previous results, and the analysis
here is much simpler. A connection between the doubling algorithm and the cyclic
reduction algorithm is also pointed out for that quadratic matrix equation. Some
concluding remarks are made in section 6.

* Received by the editors March 3, 2008; accepted for publication (in revised form) by D. A. Bini
http://www.siam.org/journals/simax/31-2/71730.html
† Department of Mathematics, National Tsinghua University, Hsinchu 300, Taiwan (d907201@oz.
thu.edu.tw).
‡ School of Mathematical Sciences, Monash University, Building 28, VIC 3800, Australia (eric.chu@sci.monash.edu.au).
§ Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada
(chguo@math.uregina.ca). The work of this author was supported in part by a grant from the Natural
Sciences and Engineering Research Council of Canada.
¶ Department of Mathematics, National Taiwan Normal University, Taipei 11677, Taiwan (min@
math.ntnu.edu.tw). The work of this author was partially supported by the National Science Council
and the National Center for Theoretical Sciences in Taiwan.
∥ Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan
(wwlin@math.nctu.edu.tw). The work of this author was partially supported by the National Science
Council and the National Center for Theoretical Sciences in Taiwan.
∗∗LMAM, School of Mathematical Sciences, Peking University, Beijing 100871, China (xsf@pku.
edu.cn).
2. The doubling algorithm. The first three subsections are based on [34], [24], and [27], but the presentation here is more general. The last subsection is directly from [27].

2.1. SDA-1. For a given matrix pair

\[ L_0 = \begin{bmatrix} I & -G_0 \\ 0 & F_0 \end{bmatrix}, \quad M_0 = \begin{bmatrix} E_0 & 0 \\ -H_0 & I \end{bmatrix}, \]

where \(E_0, F_0, G_0, H_0\) are \(n \times n, m \times m, n \times m, m \times n\), respectively, we are going to define

\[ L_k = \begin{bmatrix} I & -G_k \\ 0 & F_k \end{bmatrix}, \quad M_k = \begin{bmatrix} E_k & 0 \\ -H_k & I \end{bmatrix} \]

for all \(k \geq 0\). Assume that \(L_k\) and \(M_k\) have been defined and \(I - G_kH_k\) (and, thus, \(I - H_kG_k\)) is nonsingular for \(k \geq 0\). Then we can define the matrices

\[ \tilde{L}_k = \begin{bmatrix} I & -(G_k + E_k(I - G_kH_k)^{-1}G_kF_k) \\ 0 & F_k(I - H_kG_k)^{-1}F_k \end{bmatrix}, \quad \tilde{M}_k = \begin{bmatrix} E_k(I - G_kH_k)^{-1} & 0 \\ -(H_k + F_k(I - H_kG_k)^{-1}H_kF_k) & I \end{bmatrix}. \]

It is easily verified that \(\tilde{L}_kM_k = \tilde{M}_kL_k\). We then define

\[ L_{k+1} = \tilde{L}_kL_k = \begin{bmatrix} I & -(G_k + E_k(I - G_kH_k)^{-1}G_kF_k) \\ 0 & F_k(I - H_kG_k)^{-1}F_k \end{bmatrix}, \]

\[ M_{k+1} = \tilde{M}_kM_k = \begin{bmatrix} E_k(I - G_kH_k)^{-1}E_k & 0 \\ -(H_k + F_k(I - H_kG_k)^{-1}H_kF_k) & I \end{bmatrix}. \]

Therefore, the sequence \(\{L_k, M_k\}\) can be defined by the following doubling algorithm if no breakdown occurs.

Algorithm 2.1. (SDA-1) Given \(E_0, F_0, G_0, H_0\).

For \(k = 0, 1, \ldots\) compute

\[ E_{k+1} = E_k(I - G_kH_k)^{-1}E_k, \]

\[ F_{k+1} = F_k(I - H_kG_k)^{-1}F_k, \]

\[ G_{k+1} = G_k + E_k(I - G_kH_k)^{-1}G_kF_k, \]

\[ H_{k+1} = H_k + F_k(I - H_kG_k)^{-1}H_kF_k. \]

The algorithm requires about \(\frac{4}{3}m^3 + 6m^2n + 6mn^2 + \frac{14}{3}n^3\) flops each iteration. Note that the flop count is \(\frac{8}{3}n^3\) when \(m = n\).

2.2. SDA-2. For a given matrix pair

\[ L_0 = \begin{bmatrix} -P_0 & I \\ T_0 & 0 \end{bmatrix}, \quad M_0 = \begin{bmatrix} V_0 & 0 \\ Q_0 & -I \end{bmatrix}, \]

where all matrix blocks are \(n \times n\), we are going to define

\[ L_k = \begin{bmatrix} -P_k & I \\ T_k & 0 \end{bmatrix}, \quad M_k = \begin{bmatrix} V_k & 0 \\ Q_k & -I \end{bmatrix} \]

for all \(k \geq 0\). Assume that \(L_k\) and \(M_k\) have been defined and \(Q_k - P_k\) is nonsingular for \(k \geq 0\). Then we can define the matrices

\[ \tilde{L}_k = \begin{bmatrix} I & -V_k(Q_k - P_k)^{-1} \\ 0 & T_k(Q_k - P_k)^{-1} \end{bmatrix}, \quad \tilde{M}_k = \begin{bmatrix} V_k(Q_k - P_k)^{-1} & 0 \\ -T_k(Q_k - P_k)^{-1} & I \end{bmatrix}. \]
It is easily verified that $\tilde{L}_kM_k = \tilde{M}_kL_k$. We then define

$$
L_{k+1} = \tilde{L}_kL_k = \begin{bmatrix}
-P_k + V_k(Q_k - P_k)^{-1}T_k & I \\
T_k(Q_k - P_k)^{-1}T_k & 0
\end{bmatrix},
$$

$$
M_{k+1} = \tilde{M}_kM_k = \begin{bmatrix}
V_k(Q_k - P_k)^{-1}V_k & 0 \\
Q_k - T_k(Q_k - P_k)^{-1}V_k & -I
\end{bmatrix}.
$$

Therefore, the sequence $\{L_k, M_k\}$ can be defined by the following doubling algorithm if no breakdown occurs.

**Algorithm 2.2. (SDA-2)** Given $V_0, T_0, Q_0, P_0$.

For $k = 0, 1, \ldots$, compute

$$
V_{k+1} = V_k(Q_k - P_k)^{-1}V_k,
$$

$$
T_{k+1} = T_k(Q_k - P_k)^{-1}T_k,
$$

$$
Q_{k+1} = Q_k - T_k(Q_k - P_k)^{-1}V_k,
$$

$$
P_{k+1} = P_k + V_k(Q_k - P_k)^{-1}T_k.
$$

This algorithm requires about $\frac{38}{3}n^3$ flops each iteration.

### 2.3. Relation between $L_k$ and $M_k$.

Suppose we have

$$
L_0U = M_0UE,
$$

where the matrix pair $(L_0, M_0)$ is the initialization for either SDA-1 or SDA-2, $E$ is a square matrix, and $U$ is any matrix of suitable dimension.

Premultiplying (2.8) with $L_0$ and using $L_0M_0 = M_0L_0$, we get $L_1U = M_1UE^2$.

In general, we have for each $k \geq 0$

$$
L_kU = M_kUE^{2^k}.
$$

Suppose that there are nonsingular matrices $V$ and $Z$ such that

$$
VL_0Z = J_L, \quad VM_0Z = J_M,
$$

and $J_LJ_M = J_MJ_L$. Then it follows that

$$
M_0ZJ_L = V^{-1}J_MJ_L = V^{-1}J_LJ_M = L_0ZJ_M,
$$

and

$$
M_1ZJ_L^2 = \tilde{M}_0M_0ZJ_L^2 = \tilde{M}_0L_0ZJ_MJ_L = \tilde{L}_0M_0ZJ_LJ_M = \tilde{L}_0L_0ZJ_M^2 = L_1ZJ_M^2.
$$

In general, we have for each $k \geq 0$

$$
M_kZJ_L^{2^k} = L_kZJ_M^{2^k}.
$$

### 2.4. Result on special Jordan blocks.

Let $J_{\omega, p}$ be the $p \times p$ Jordan block with a unimodular eigenvalue $\omega = e^{i\theta}$:

$$
J_{\omega, p} = \begin{bmatrix}
\omega & 1 & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 1 \\
0 & \cdots & \cdots & 0 & \omega
\end{bmatrix}.
$$
When \( p = 2m \), let \( \Gamma_{k,m} \) be determined through the partition

\[
J_{\omega,2m}^{2k} = \begin{bmatrix}
J_{\omega,m}^{2k} & \Gamma_{k,m} \\
0 & J_{\omega,m}^{2k}
\end{bmatrix}.
\]

The following useful Lemma is proved in [27].

**Lemma 2.1.** The matrix \( \Gamma_{k,m} \) is invertible and satisfies

\[
\|\Gamma_{k,m}^{-1} J_{\omega,m}^{2k}\| = O(2^{-k}), \quad \|J_{\omega,m}^{2k} \Gamma_{k,m}^{-1} J_{\omega,m}^{2k}\| = O(2^{-k}) \quad \text{as } k \to \infty.
\]

In the next three sections, we will apply the techniques reviewed in this section to three different nonlinear matrix equations. Although the general approach will be the same, we will need to fully exploit the special properties of each equation. Among other things, the following two issues deserve special attention: (1) Given a nonlinear matrix equation, how should we rewrite it in its equivalent form (2.8)? If possible, we should try to get a form (2.8) that would lead to SDA-2 rather than SDA-1, since SDA-2 is less expensive. (2) How should we choose the matrices \( J_L \) and \( J_M \) in (2.10)? The matrices must satisfy \( J_L J_M = J_M J_L \), and the resulting equation (2.11) and an equation from a similar procedure should be easy to handle together. We will keep these issues in mind when we carry out the convergence analysis for the three equations.

### 3. A special nonlinear matrix equation.

In this section we consider the nonlinear matrix equation (NME)

\[
X + A^T X^{-1} A = Q,
\]

where \( A, Q \in \mathbb{R}^{n \times n} \) with \( Q \) being symmetric positive definite. Various aspects of the NME, like solvability, numerical solution, perturbation, and applications, can be found in [8, 9, 13, 17, 22, 35, 38, 39, 40, 41] and the references therein.

For symmetric matrices \( X \) and \( Y \), we write \( X \geq Y \) (\( X > Y \)) if \( X - Y \) is positive semidefinite (definite). We use this definition of ordering only in this section, and will use the elementwise order in sections 4 and 5. We assume that (3.1) has a symmetric positive definite solution. Then [9] it has a maximal symmetric positive definite solution \( X_+ \) (\( X_+ \geq X \) for any symmetric positive definite solution \( X \) of (3.1)), and \( \rho(X_+^{-1} A) \leq 1 \), where \( \rho(\cdot) \) is the spectral radius.

Let

\[
L_0 = \begin{bmatrix}
0 & I \\
A^T & 0
\end{bmatrix}, \quad M_0 = \begin{bmatrix}
A & 0 \\
Q & -I
\end{bmatrix}.
\]

It is easy to verify that the pencil \( M_0 - \lambda L_0 \) (also denoted by \( (M_0, L_0) \)) is symplectic, i.e.,

\[
M_0 J M_0^T = L_0 J L_0^T \quad \text{for} \quad J = \begin{bmatrix}
0 & I \\
-I & 0
\end{bmatrix}.
\]

Using Algorithm 2.2 with \( V_0 = A, T_0 = A^T, Q_0 = Q, P_0 = 0 \), we have \( T_k = V_k^T, Q_k^T = Q_k, P_k^T = P_k \). So Algorithm 2.2 is simplified to the following, where we have used \( A_k \) for \( V_k \).
Algorithm 3.1. Let \( A_0 = A, Q_0 = Q, P_0 = 0 \).
For \( k = 0, 1, \ldots \), compute
\[
A_{k+1} = A_k(Q_k - P_k)^{-1}A_k, \\
Q_{k+1} = Q_k - A_k^T(Q_k - P_k)^{-1}A_k, \\
P_{k+1} = P_k + A_k(Q_k - P_k)^{-1}A_k^T.
\]

The matrices \( L_k, M_k \) in (2.7) are now given by
\[
L_k = \begin{bmatrix} -P_k & I \\ A_k^T & 0 \end{bmatrix}, \\
M_k = \begin{bmatrix} A_k & 0 \\ Q_k & -I \end{bmatrix}.
\]

It is noted in [34] that the cyclic reduction algorithm in [35] is recovered from Algorithm 3.1 when \( Q_k - P_k \) and \( Q_k \) are replaced by \( Q_k \) and \( X_k \), respectively, where the latter \( Q_k \) and \( X_k \) are the notations used in [35, Algorithm 3.1]. So we know from [35] that \( Q_k - P_k > 0 \) in Algorithm 3.1. Thus, the algorithm is well defined and \( 0 \leq P_k < Q_k \leq Q \). This fact is also proved in [34] without using the results in [35].

It is easy to verify that
\[
M_0 \begin{bmatrix} I \\ X_+ \end{bmatrix} = L_0 \begin{bmatrix} I \\ X_+ \end{bmatrix} X_+^{-1}A.
\]

We are interested in the case with \( \rho(X_+^{-1}A) = 1 \). It follows from [13, Theorem 2.4] that the eigenvalues of \( X_+^{-1}A \) have the following characterization.

Theorem 3.1. For (3.1), the eigenvalues of the matrix \( X_+^{-1}A \) are precisely the eigenvalues of the matrix pencil \( M_0 - \lambda L_0 \) inside or on the unit circle, with half of the (necessarily even) partial multiplicities for each unimodular eigenvalue of the pencil.

In view of the connection between Algorithm 3.1 and the cyclic reduction algorithm in [35], we know from [13] that the sequence \( Q_k \) in Algorithm 3.1 converges to \( X_+ \) at least linearly with rate 1/2, as long as all eigenvalues of \( X_+^{-1}A \) on the unit circle are semisimple. With the tools in section 2, we are going to prove more convergence results for Algorithm 3.1, without any assumption on the unimodular eigenvalues of \( X_+^{-1}A \).

Suppose there are \( r \) Jordan blocks associated with unimodular eigenvalues of \((M_0, L_0)\). Then they have the form
\[
J_{\omega_j, 2m_j} = \begin{bmatrix} J_{\omega_j, m_j} & \Gamma_0, m_j \\ 0 & J_{\omega_j, m_j} \end{bmatrix}, \\
\Gamma_0, m_j \equiv e_{m_j} e_1^T,
\]
where \( \omega_j = e^{i \theta_j} \) for \( j = 1, \ldots, r \).

By the results on Kronecker canonical form for a symplectic pencil (see [11] and [33]), there exist nonsingular matrices \( V \) and \( Z \) such that
\[
VL_0Z = \begin{bmatrix} I_n & 0_n \\ 0_n & J_s^H \oplus I_m \end{bmatrix} \equiv J_L, \\
VM_0Z = \begin{bmatrix} J_s \oplus J_1 & 0_t \oplus \Gamma_0 \\ 0_n & I_t \oplus J_1 \end{bmatrix} \equiv J_M,
\]
where \( J_s \in \mathbb{C}^{t \times t} \) consists of stable Jordan blocks (so \( \rho(J_s) < 1 \)), \( J_1 = J_{\omega_1, m_1} \oplus \cdots \oplus J_{\omega_r, m_r} \), \( \Gamma_0 = \Gamma_0, m_1 \oplus \cdots \oplus \Gamma_0, m_r \), \( m = m_1 + \cdots + m_r \), \( t = n - m \), \( \oplus \) denotes the direct
sum of matrices and \((\cdot)^H\) the conjugate transpose. Moreover, the nonsingular matrix 
\(Z\) can be taken to be of the form \(Z = Z_a Z_b\) with \(Z_a\) symplectic and \(Z_b = I_n \oplus Z_c\).
It follows that \(\text{span}\{Z(:,1:n)\}\) forms the unique weakly stable Lagrangian deflating 
subspace of \((M_0, L_0)\) corresponding to \(J_s \oplus J_1\).

Let \(\Gamma_{k,m}\) be given by (2.13) with \(\omega = \omega_j\) and \(m = m_j\). Since \(J_L J_M = J_M J_L\), we 
have by (2.11)

\[
(3.8) \quad M_k Z \begin{bmatrix} I & 0 \\ 0 & (J_s^H)^{2^k} \oplus I \end{bmatrix} = L_k Z \begin{bmatrix} J_s^{2^k} \oplus J_1^{2^k} & 0 \oplus \Gamma_k \\ 0 & I \oplus J_1^{2^k} \end{bmatrix},
\]
where \(\Gamma_k = \Gamma_{k,m_1} \oplus \cdots \oplus \Gamma_{k,m_n}\).

Similarly, there exist nonsingular matrices \(T\) and \(W\) such that

\[
(3.9) \quad TM_0 W = J_L, \quad TL_0 W = J_M,
\]

and

\[
(3.10) \quad L_k W \begin{bmatrix} I & 0 \\ 0 & (J_s^H)^{2^k} \oplus I \end{bmatrix} = M_k W \begin{bmatrix} J_s^{2^k} \oplus J_1^{2^k} & 0 \oplus \Gamma_k \\ 0 & I \oplus J_1^{2^k} \end{bmatrix}.
\]

By Lemma 2.1 we have

\[
(3.11) \quad \|\Gamma_k^{-1} J_s^{2^k}\| = O\left(2^{-k}\right), \quad \|J_1^{2^k} \Gamma_k^{-1} J_s^{2^k}\| = O\left(2^{-k}\right) \quad \text{as} \quad k \to \infty.
\]

We now prove some convergence results for Algorithm 3.1. Partition \(Z\) and \(W\) as

\[
(3.12) \quad Z = \begin{bmatrix} Z_1 & Z_3 \\ Z_2 & Z_4 \end{bmatrix}, \quad W = \begin{bmatrix} W_1 & W_3 \\ W_2 & W_4 \end{bmatrix},
\]

where \(Z_i, W_i \in \mathbb{C}^{n \times n} \quad (i = 1, \ldots, 4)\).

**THEOREM 3.2.** When \(\rho(X_+^{-1}A) = 1\), the sequences \(\{A_k, Q_k, P_k\}\) generated by 
Algorithm 3.1 satisfy

(a) \(\|A_k\| = O(2^{-k})\); 
(b) \(\|Q_k - X_+\| = O(2^{-k})\) and \(X_+ = Z_2 Z_1^{-1}\); 
(c) \(\|P_k - X_-\| = O(2^{-k})\) for \(X_- = W_2 W_1^{-1}\) if \(W_1\) is invertible; if \(A\) is also 
invertible, then \(X_-\) is a solution of (3.1) and the eigenvalues of \(X_-^{-1}A\) are 
the reciprocals of the eigenvalues of \(X_+^{-1}A\); 
(d) \(Q_k - P_k\) converges to a singular matrix as \(k \to \infty\).

**Proof.** (a) Substituting \(L_k\) and \(M_k\) of (3.3) and \(Z\) of (3.12) into (3.8), we obtain

\[
(3.13) \quad A_k Z_1 = (-P_k Z_1 + Z_2) \left(J_s^{2^k} \oplus J_1^{2^k}\right),
\]

\[
(3.14) \quad A_k Z_3 \left((J_s^H)^{2^k} \oplus I\right) = (-P_k Z_1 + Z_2)(0 \oplus \Gamma_k) + (-P_k Z_3 + Z_4) \left(I \oplus J_1^{2^k}\right),
\]

\[
(3.15) \quad Q_k Z_1 - Z_2 = A_k^T Z_1 \left(J_s^{2^k} \oplus J_1^{2^k}\right),
\]

\[
(3.16) \quad (Q_k Z_3 - Z_4) \left((J_s^H)^{2^k} \oplus I\right) = A_k^T Z_1 (0 \oplus \Gamma_k) + A_k^T Z_3 \left(I \oplus J_1^{2^k}\right).
\]

From (3.6) and (3.7) we have

\[
M_0 \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = L_0 \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} (J_s \oplus J_1).
\]
By Theorem 3.1, $X_+^{-1}A$ is similar to $J_s \oplus J_1$. Then from (3.4) and the uniqueness of weakly stable Lagrangian deflating subspaces of $(M_0, L_0)$ corresponding to $J_s \oplus J_1$, we have

$$
\begin{bmatrix}
Z_1 \\
Z_2
\end{bmatrix} = \begin{bmatrix}
I \\
X_+
\end{bmatrix} R
$$

for a nonsingular matrix $R$. It follows that $Z_1^{-1}$ exists and $X_+ = Z_2 Z_1^{-1}$.

Postmultiplying (3.14) by $(0 \oplus \Gamma_k^{-1} J_1^k) Z_1^{-1}$ and using (3.13), we have

$$
A_k \left[ I - Z_3 \left( 0 \oplus \Gamma_k^{-1} J_1^k \right) Z_1^{-1} \right]
$$

$$
= (-P_k Z_1 + Z_2) \left( J_s^2 \oplus 0 \right) Z_1^{-1} - (-P_k Z_3 + Z_4) \left( 0 \oplus J_1^k \Gamma_k^{-1} J_1^k \right) Z_1^{-1}.
$$

It follows from (3.11) and the boundedness of $\{P_k\}$ that

(3.17) \quad $\|A_k\| = O \left( 2^{-k} \right)$.

(b) Postmultiplying (3.16) by $(0 \oplus \Gamma_k^{-1} J_1^k) Z_1^{-1}$ and using (3.15), we get

$$
Q_k \left[ I - Z_3 \left( 0 \oplus \Gamma_k^{-1} J_1^k \right) Z_1^{-1} \right] - X_+
$$

$$
= \left[ A^T Z_1 \left( J_s^2 \oplus 0 \right) - A^T Z_3 \left( 0 \oplus J_1^k \Gamma_k^{-1} J_1^k \right) - Z_4 \left( 0 \oplus \Gamma_k^{-1} J_1^k \right) \right] Z_1^{-1}.
$$

By (3.11) and (3.17), we have

$$
\|Q_k - X_+\| = O \left( 2^{-k} \right).
$$

(c) Substituting $L_k$ and $M_k$ of (3.3) and $W$ of (3.12) into (3.10), we have

(3.19) \quad $W_2 - P_k W_1 = A_k W_1 \left( J_s^2 \oplus J_1^k \right)$,

(3.20) \quad $W_4 - P_k W_3 = (J_s^2) \Gamma_k^{-1} + I \left( I \oplus J_1^k \right)$.

Let $X_- = W_2 W_1^{-1}$. As before, postmultiplying (3.20) by $(0 \oplus \Gamma_k^{-1} J_1^k) W_1^{-1}$ and using (3.19), we get

$$
X_- - P_k \left[ I - W_3 \left( 0 \oplus \Gamma_k^{-1} J_1^k \right) W_1^{-1} \right]
$$

$$
= \left[ W_4 \left( 0 \oplus \Gamma_k^{-1} J_1^k \right) + A_k W_1 \left( J_s^2 \oplus 0 \right) - A_k W_3 \left( 0 \oplus J_1^k \Gamma_k^{-1} J_1^k \right) \right] W_1^{-1}.
$$

By (3.11) and the result of (a), we have

$$
\|X_- - P_k\| = O \left( 2^{-k} \right).
$$

From (3.9) we get

$$
\begin{bmatrix}
0 & I \\
A^T & 0
\end{bmatrix}
\begin{bmatrix}
I \\
X_-
\end{bmatrix} =
\begin{bmatrix}
A & 0 \\
Q & -I
\end{bmatrix}
\begin{bmatrix}
I \\
X_-
\end{bmatrix} R_-,
$$

where $R_- = W_1 (J_s \oplus J_1) W_1^{-1}$. It follows that

$$
X_- = AR_-,
\quad A^T = (Q - X_-) R_-.
$$
When $A$ is invertible, the matrices $X_+^{-1}A, R_-, X_-$ are all invertible and we obtain

$$X_- + A^T X_-^{-1} A = Q.$$ 

Moreover, the eigenvalues of $X_-^{-1} A$ are the reciprocals of the eigenvalues of $R_-$ (and, thus, $X_+^{-1} A$).

(d) From (3.13) and (3.15), we get

$$-P_k Z_1 \left( J_s^e + J_l^e \right) = A_k Z_1 - Z_2 \left( J_s^e + J_l^e \right),$$

$$Q_k Z_1 \left( J_s^e + J_l^e \right) = Z_2 \left( J_s^e + J_l^e \right) + A_k Z_1 \left( J_s^e + J_l^e \right).$$

This implies that

$$(Q_k - P_k) Z_1 \begin{bmatrix} 0 \\ I_m \end{bmatrix} = A_k Z_1 \begin{bmatrix} 0 \\ J_1^{-2k} \end{bmatrix} + A_k Z_1 \begin{bmatrix} 0 \\ J_1^{2k} \end{bmatrix}.$$ 

Since $0 \leq P_k \leq P_{k+1} \leq Q$, the sequence $P_k$ converges even if $W_1$ is singular. Let $\lim(Q_k - P_k) = R_*$.

Thus, $R_*$ is singular.

The most important conclusion in Theorem 3.2 is that the sequence $Q_k$ from the doubling algorithm converges to $X_+$ at least linearly with rate $1/2$, regardless of the values of $m_j$ ($j = 1, 2, \ldots, r$). This is in sharp contrast with the behavior of Newton’s method. The NME (3.1) is a special case of the discrete algebraic Riccati equation studied in [12]. It is conjectured in [12] that the convergence of Newton’s method is linear with rate $1/\sqrt{2}$, where $q = \max_{1 \leq j \leq r} m_j$. This conjecture is confirmed in numerical experiments on (3.1) with $A$ being a $q \times q$ Jordan block with eigenvalue 1 and $Q = I + A^T A$, for small values of $q$. We know from [13] that $X_+ = I$ in all those examples. Newton’s method is given in [13, Algorithm 3.3].

4. A quadratic matrix equation from quasi-birth-death problems. A discrete-time quasi-birth-death (QBD) process is a Markov chain with state space $\{(i, j) \mid i \geq 0, 1 \leq j \leq n\}$, and with a transition probability matrix of the form

$$P = \begin{bmatrix}
B_0 & B_1 & 0 & 0 & \cdots \\
A_0 & A_1 & A_2 & 0 & \cdots \\
0 & A_0 & A_1 & A_2 & \cdots \\
0 & 0 & A_0 & A_1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix},$$

where $B_0, B_1, A_0, A_1,$ and $A_2$ are $n \times n$ nonnegative matrices such that $P$ is stochastic. In particular, $(A_0 + A_1 + A_2) e = e$, where $e = (1, 1, \ldots, 1)^T$.

We make the standard assumption that the matrix $P$ and the matrix $A = A_0 + A_1 + A_2$ are both irreducible. Thus, $A_0 \neq 0$ and $A_2 \neq 0$. Moreover, there exists a unique positive vector $\alpha$ with $\alpha^T e = 1$ and $\alpha^T A = \alpha^T$. The QBD is positive recurrent if $\alpha^T A_0 e > \alpha^T A_2 e$, transient if $\alpha^T A_0 e < \alpha^T A_2 e$, and null recurrent if $\alpha^T A_0 e = \alpha^T A_2 e$. 

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
The minimal nonnegative solution $G$ of the matrix equation

$$G = A_0 + A_1G + A_2G^2$$

(4.1)

plays an important role in the study of the QBD process (see [32]). We will also need the dual equation

$$F = A_2 + A_1F + A_0F^2,$$

(4.2)

and we let $F$ be its minimal nonnegative solution. It is well known (see [32], for example) that if the QBD is positive recurrent, then $G$ is stochastic and $F$ is substochastic with spectral radius $\rho(F) < 1$; if the QBD is transient, then $F$ is stochastic and $G$ is substochastic with $\rho(G) < 1$; if the QBD is null recurrent, then $G$ and $F$ are both stochastic.

The Latouche–Ramaswami (LR) algorithm [31] and the cyclic reduction (CR) algorithm [5] are both efficient iterative methods for finding the minimal solution $G$. The convergence of these two algorithms is quadratic for positive recurrent and transient QBDs. A convergence analysis has been performed in [15] for the LR algorithm in the null recurrent case under two additional assumptions. The first assumption is that $\lambda = 1$ is a simple eigenvalue of $G$ and $F$ and there are no other eigenvalues of $G$ or $F$ on the unit circle; the second assumption is made under the first assumption and is more technical. The convergence rate for the LR algorithm is the same in view of the relationship between CR and LR, given in [3].

We can also use the doubling algorithm (SDA-1 or SDA-2) to find the minimal solution $G$. We will choose to use SDA-2 since it is less expensive. Moreover, there is a close connection between the CR algorithm and SDA-2. In this section we determine the convergence rate of SDA-2 in the null recurrent case, without the two additional assumptions in [15]. The convergence rate for the CR (or LR) algorithm in the null recurrent case is the same in view of their connections to SDA-2. As compared to [15], the result here is more general and the analysis here is much simpler.

We mention that a doubling algorithm is also derived in [26] for finding the minimal nonnegative solution of a polynomial equation that is more general than (4.1). The algorithm there is different from SDA-2 when applied to (4.1).

The CR algorithm for (4.1), or for $-A_0 + (I - A_1)G - A_2G^2 = 0$, is the following:

**Algorithm 4.1.** Set $T_0 = A_0$, $U_0 = I - A_1$, $V_0 = A_2$, $S_0 = I - A_1$.

For $k = 0, 1, \ldots$, compute

$$T_{k+1} = T_kU_k^{-1}T_k,$$

$$U_{k+1} = U_k - T_kU_k^{-1}V_k - V_kU_k^{-1}T_k,$$

$$V_{k+1} = V_kU_k^{-1}V_k,$$

$$S_{k+1} = S_k - V_kU_k^{-1}T_k.$$

The above CR algorithm is as presented in [3], but with one minor change: if we follow [3] exactly, $T_k$ and $V_k$ here would have to be replaced by $-T_k$ and $-V_k$ for $k \geq 0$.

The following result is known from the discussions in [4] and [32].

**Theorem 4.1.** The sequences $\{T_k\}, \{U_k\}, \{V_k\}, \{S_k\}$ in Algorithm 4.1 are well defined. For each $k \geq 0$, $T_k$ and $V_k$ are nonnegative, and $U_k$ and $S_k$ are nonsingular $M$-matrices. When the QBD is positive recurrent or transient, the sequence $\{S_k\}$ converges quadratically to a nonsingular $M$-matrix $S_\ast$ and $S_\ast^{-1}A_0 = G$. 

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
We note that Algorithm 4.1 may break down if we do not assume the irreducibility of the transition matrix $P$. As an example, we consider

$$A_0 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad A_1 = 0, \quad A_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$  

It is easy to see that $P$ is not irreducible, although $A_0 + A_1 + A_2$ is. For this example, $U_1 = 0$ in Algorithm 4.1, so the algorithm breaks down. The LR algorithm also breaks down for this example.

To use the doubling algorithm to find $G$, we may rewrite (4.1) as

$$\begin{bmatrix} 0 & I \\ A_0 & A_1 - I \end{bmatrix} \begin{bmatrix} I \\ G \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & -A_2 \end{bmatrix} \begin{bmatrix} I \\ G \end{bmatrix} G.$$  

Multiplying the second block row by $-(I - A_1)^{-1}$ and eliminating the $I$ in the $(1, 2)$ block of the leftmost matrix, we get

$$\begin{bmatrix} (I - A_1)^{-1}A_0 & 0 \\ -(I - A_1)^{-1}A_0 & I \end{bmatrix} \begin{bmatrix} I \\ G \end{bmatrix} = \begin{bmatrix} I & -(I - A_1)^{-1}A_2 \\ 0 & (I - A_1)^{-1}A_2 \end{bmatrix} \begin{bmatrix} I \\ G \end{bmatrix} G.$$  

We can then use SDA-1 to find the matrix $G$. However, the less expensive SDA-2 can also be used if we rewrite (4.1) as

$$(4.3) \quad L_0 \begin{bmatrix} I \\ A_2 G \end{bmatrix} = M_0 \begin{bmatrix} I \\ A_2 G \end{bmatrix} G,$$

where

$$L_0 = \begin{bmatrix} 0 & I \\ A_0 & 0 \end{bmatrix}, \quad M_0 = \begin{bmatrix} A_2 & 0 \\ I - A_1 & -I \end{bmatrix}.$$  

It is easily seen that $L_0 - \lambda M_0$ is a linearization of $-A_0 + \lambda (I - A_1) - \lambda^2 A_2$.

If we use SDA-1, the matrix $G$ can be approximated directly by a sequence generated by SDA-1. One may have some concern about the SDA-2 approach: how can one get $G$ if $A_2 G$ is obtained and $A_2$ is singular? This concern will turn out to be unnecessary.

In this section SDA-2 is Algorithm 2.2 with the initialization

$$(4.4) \quad T_0 = A_0, \quad Q_0 = I - A_1, \quad P_0 = 0, \quad V_0 = A_2.$$  

The algorithm generates the sequence $\{L_k, M_k\}$ (see (2.7)) if no breakdown occurs.

It is readily seen that Algorithm 4.1 is recovered from SDA-2 by letting $U_k = Q_k - P_k$ and $S_k = S_0 - P_k$. By Theorem 4.1, $Q_k - P_k = U_k$ are nonsingular $M$-matrices for all $k \geq 0$. So SDA-2 is also well defined.

In view of (2.9) we have for each $k \geq 0$

$$L_k \begin{bmatrix} I \\ A_2 G \end{bmatrix} = M_k \begin{bmatrix} I \\ A_2 G \end{bmatrix} G^{2^k}.$$  

So

$$(4.5) \quad -P_k + A_2 G = V_k G^{2^k}, \quad T_k = Q_k G^{2^k} - A_2 G^{2^k+1}.$$
Similarly we have
\[ \hat{L}_0 \begin{bmatrix} I \\ A_0 F \end{bmatrix} = \hat{M}_0 \begin{bmatrix} I \\ A_0 F \end{bmatrix} F, \]
where
\[ \hat{L}_0 = \begin{bmatrix} 0 & I \\ V_0 & 0 \end{bmatrix}, \quad \hat{M}_0 = \begin{bmatrix} T_0 & 0 \\ Q_0 & -I \end{bmatrix}. \]
It is easily seen that \( \hat{M}_0 - \lambda \hat{L}_0 \) is also a linearization of \(-A_0 + \lambda(I - A_1) - \lambda^2 A_2\).

For each \( k \geq 0 \) we now have
\[ \hat{L}_k \begin{bmatrix} I \\ A_0 F \end{bmatrix} = \hat{M}_k \begin{bmatrix} I \\ A_0 F \end{bmatrix} F^{2^k}, \]
where
\[ \hat{L}_k = \begin{bmatrix} -\hat{P}_k & I \\ V_k & 0 \end{bmatrix}, \quad \hat{M}_k = \begin{bmatrix} T_k & 0 \\ \hat{Q}_k & -I \end{bmatrix} \]
with
\begin{equation}
\hat{P}_k = I - A_1 - Q_k, \quad \hat{Q}_k = I - A_1 - P_k.
\end{equation}
So
\begin{equation}
-\hat{P}_k + A_0 F = T_k F^{2^k}, \quad V_k = \hat{Q}_k F^{2^k} - A_0 F^{2^k+1}.
\end{equation}

We mentioned before that the \( S_k \) in Algorithm 4.1 satisfies \( S_k = S_0 - P_k = I - A_1 - P_k \). So we have \( \hat{Q}_k = S_k \).

When the QBD is positive recurrent or transient, we know by Theorem 4.1 that \( \hat{Q}_k \) converges quadratically to a nonsingular \( M \)-matrix \( Q_* \) and \( \hat{Q}_k^{-1} A_0 = G \). Here we give a quick proof using the doubling algorithm. By the first equation in (4.5) and the second equation in (4.6), we have
\[ \hat{Q}_k - I + A_1 + A_2 G = V_k G^{2^k}. \]
Eliminating \( V_k \) using the second equation in (4.7) gives
\[ \hat{Q}_k \left( I - F^{2^k} G^{2^k} \right) = I - A_1 - A_2 G - A_0 F^{2^k+1} G^{2^k}. \]
It follows that
\[ \lim_{k \to \infty} \sqrt[k]{\| \hat{Q}_k - (I - A_1 - A_2 G) \|} \leq \rho(F) \rho(G) < 1. \]
Since \( \hat{Q}_k = I - A_1 - A_2 G \) is a nonsingular \( M \)-matrix and \( A_0 = \hat{Q}_k G \), we have \( G = \hat{Q}_k^{-1} A_0 \). Similarly, \( Q_k \) converges quadratically to the nonsingular \( M \)-matrix
\[ Q_* = I - A_1 - A_0 F \] and \( F = Q_*^{-1} A_2 \).

Our main purpose of this section, however, is to determine the convergence rate of SDA-2 for the null recurrent case.

We start with a review of an important result about the spectral properties of the quadratic pencil \(-A_0 + \lambda(I - A_1) - \lambda^2 A_2\) and of the matrices \( G \) and \( F \) when the QBD is null recurrent. See Proposition 14 and Theorem 4 of [10] and Theorem 4.10 of [4].
Theorem 4.2. Let the QBD be null recurrent. Then
(a) For some integer \( r \geq 1 \) the quadratic pencil \(-A_0 + \lambda(I - A_1) - \lambda^2 A_2\) has \( n - r \) eigenvalues inside the unit circle, \( n - r \) eigenvalues outside the unit circle (which include eigenvalues at infinity), and \( 2r \) eigenvalues on the unit circle, which are the \( r \)th roots of unity, each with multiplicity two.
(b) The partial multiplicity of each eigenvalue on the unit circle is exactly two.
(c) The eigenvalues of \( G \) are the \( n - r \) eigenvalues of the pencil inside the unit circle plus the \( r \) simple eigenvalues at the \( r \)th roots of unity, the eigenvalues of \( F \) are the reciprocals of the \( n - r \) eigenvalues of the pencil outside the unit circle plus the \( r \) simple eigenvalues at the \( r \)th roots of unity.

Using the Kronecker form for matrix pairs, we have nonsingular matrices \( V \) and \( Z \) such that
\[
VM_0Z = \begin{bmatrix} I_n & 0 \\ 0 & J_2 \oplus I_r \end{bmatrix} = J_M,
\]
\[
VL_0Z = \begin{bmatrix} J_1 \oplus D_r & 0 \oplus I_r \\ 0 & I_{n-r} \oplus D_r \end{bmatrix} = J_L,
\]
where \( J_1 \) and \( J_2 \) are \((n - r) \times (n - r)\) matrices consisting of the Jordan blocks with diagonal elements inside the unit circle, and \( D_r \) is a \( r \times r \) diagonal matrix with the \( r \)th roots of unity on the diagonal.

Similarly, we have nonsingular matrices \( T \) and \( W \) such that
\[
T\hat{L}_0W = \begin{bmatrix} I_n & 0 \\ 0 & J_2 \oplus I_r \end{bmatrix} = \hat{J}_M,
\]
\[
T\hat{M}_0W = \begin{bmatrix} J_1 \oplus D_r & 0 \\ 0 & I_r \oplus I_r \oplus D_r \end{bmatrix} = \hat{J}_L.
\]

We have for each \( k \geq 0 \)
\[
M_k Z J_{L_k}^{2k} = L_k Z J_M^{2k}, \quad \hat{L}_k W \hat{J}_L^{2k} = \hat{M}_k W \hat{J}_M^{2k}.
\]

Let \( Z \) and \( W \) be partitioned as in (3.12). From (4.8) and (4.9) we have
\[
L_0 \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = M_0 \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} (J_1 \oplus D_r).
\]

Comparing this with (4.3) and using Theorem 4.2, we know that \( Z_1 \) is nonsingular and \( Z_2^{Z_1^{-1}} = A_2 G \). Similarly, \( W_3 \) is nonsingular and \( W_3^{W_3^{-1}} = A_0 F \).

Using block matrix multiplication for (4.12), we have
\[
V_k Z_1 \left( J_1^{2k} \oplus D_r^{2k} \right) = -P_k Z_1 + Z_2,
\]
\[
\left( Q_k Z_1 - Z_2 \right) \left( J_1^{2k} \oplus D_r^{2k} \right) = T_k Z_1,
\]
\[
V_k Z_3 \left( 0 \oplus 2^k D_r^{2^{k-1}} \right) = -P_k Z_3 + Z_4 \left( J_2^{2^k} \oplus I \right),
\]
\[
\left( Q_k Z_1 - Z_2 \right) \left( 0 \oplus 2^k D_r^{2^{k-1}} \right) + \left( Q_k Z_3 - Z_4 \right) \left( 0 \oplus 2^k D_r^{2^{k-1}} \right) = T_k Z_3 \left( J_2^{2^k} \oplus I \right),
\]
\[
\left( -P_k W_1 + W_2 \right) \left( J_1^{2^k} \oplus D_r^{2^k} \right) + \left( -P_k W_3 + W_4 \right) \left( 0 \oplus 2^k D_r^{2^{k-1}} \right) = T_k W_1,
\]
\[
V_k W_1 \left( J_1^{2^k} \oplus D_r^{2^k} \right) + V_k W_3 \left( 0 \oplus 2^k D_r^{2^{k-1}} \right) = Q_k W_1 - W_2.
\]
(4.19) \[
(-\tilde{P}_k W_3 + W_4)(I \oplus D_r^{2^k}) = T_k W_3 \left(J_2^{2^k} \oplus I\right),
\]
(4.20) \[
V_k W_3 \left(I \oplus D_r^{2^k}\right) = \left(\hat{Q}_k W_3 - W_4\right) \left(J_2^{2^k} \oplus I\right).
\]
Postmultiplying (4.16) by \(0 \oplus 2^{-k} D_r\) and subtracting the result from (4.14), we get
(4.21) \[
T_k \left(Z_1 - Z_3 (0 \oplus 2^{-k} D_r)\right) = (Q_k Z_1 - Z_2) \left(J_1^{2^k} \oplus 0\right) - (Q_k Z_3 - Z_4) \left(0 \oplus 2^{-k} D_r^{2^k+1}\right).
\]
By (4.19) we have
(4.22) \[
-\tilde{P}_k = -W_4 W_3^{-1} + T_k W_3 \left(J_2^{2^k} \oplus D_r^{-2^k}\right) W_3^{-1}.
\]
Thus, in view of (4.6),
(4.23) \[
Q_k = I - A_1 - W_4 W_3^{-1} + T_k W_3 \left(J_2^{2^k} \oplus D_r^{-2^k}\right) W_3^{-1}.
\]
Inserting (4.23) into (4.21) and letting \(Q_* = I - A_1 - W_4 W_3^{-1}\), we get
\[
T_k \left[Z_1 - Z_3 \left(0 \oplus 2^{-k} D_r\right) - W_3 \left(J_2^{2^k} \oplus D_r^{-2^k}\right) W_3^{-1} \left(Z_1 \left(J_1^{2^k} \oplus 0\right) - Z_3 \left(0 \oplus 2^{-k} D_r^{2^k+1}\right)\right)\right]
= (Q_* Z_1 - Z_2) \left(J_1^{2^k} \oplus 0\right) - (Q_* Z_3 - Z_4) \left(0 \oplus 2^{-k} D_r^{2^k+1}\right),
\]
from which it follows that
\[
\|T_k\| = O \left(2^{-k}\right).
\]
It then follows from (4.23) that
\[
\|Q_k - (I - A_1 - W_4 W_3^{-1})\| = O \left(2^{-k}\right).
\]
Postmultiplying (4.15) by \(0 \oplus 2^{-k} D_r\) and subtracting the result from (4.13), we get
(4.24) \[
-P_k Z_3 + Z_2 - (-P_k Z_3 + Z_4) \left(0 \oplus 2^{-k} D_r\right) = V_k \left(Z_1 \left(J_1^{2^k} \oplus 0\right) - Z_3 \left(0 \oplus 2^{-k} D_r^{2^k+1}\right)\right).
\]
By (4.20),
(4.25) \[
V_k = \left(\hat{Q}_k W_3 - W_4\right) \left(J_2^{2^k} \oplus D_r^{-2^k}\right) W_3^{-1}.
\]
Inserting (4.25) into (4.24) and using \(\hat{Q}_k = I - A_1 - P_k\), we get
\[
-P_k Z_1 + Z_2 - (-P_k Z_3 + Z_4) \left(0 \oplus 2^{-k} D_r\right) = ((I - A_1 - P_k) W_3 - W_4) C_k
\]
for some \(C_k\) with \(\|C_k\| = O(2^{-k})\). Thus,
\[
P_k \left(Z_1 - Z_3 \left(0 \oplus 2^{-k} D_r\right) - W_3 C_k\right) = Z_2 - Z_4 \left(0 \oplus 2^{-k} D_r\right) - ((I - A_1) W_3 - W_4) C_k.
\]
It follows that
\[
\|P_k - Z_2 Z_1^{-1}\| = O \left(2^{-k}\right).
\]
Postmultiplying (4.18) by $0 \oplus 2^{-k}D_r^{1-2^k}$, we get

\[(4.26) \quad V_k W_1 (0 \oplus 2^{-k}D_r) + V_k W_3 (0 \oplus I) = \left( \hat{Q}_k W_1 - W_2 \right) \left( 0 \oplus 2^{-k}D_r^{1-2^k} \right).\]

Postmultiplying (4.20) by $I \oplus 0$, we get

\[(4.27) \quad V_k W_3 (I \oplus 0) = \left( \hat{Q}_k W_3 - W_4 \right) \left( J_2^{2^k} \oplus 0 \right).\]

Adding (4.26) and (4.27) gives

\[V_k \left( W_3 + W_1 (0 \oplus 2^{-k}D_r) \right) = \left( \hat{Q}_k W_1 - W_2 \right) \left( 0 \oplus 2^{-k}D_r^{1-2^k} \right) + \left( \hat{Q}_k W_3 - W_4 \right) \left( J_2^{2^k} \oplus 0 \right).\]

It follows that

\[\|V_k\| = O\left(2^{-k}\right),\]

since $W_3$ is nonsingular and $\{\hat{Q}_k\}$ has been shown to be bounded.

In summary, we have proved the following result.

**Theorem 4.3.** Let the QBD be null recurrent. Then for SDA-2sf we have

\[\|V_k\| = O\left(2^{-k}\right), \quad \|T_k\| = O\left(2^{-k}\right),\]

\[\|Q_k - (I - A_1 - A_0 F)\| = O\left(2^{-k}\right), \quad \|P_k - A_2 G\| = O\left(2^{-k}\right).\]

**Corollary 4.4.** Let $\lim Q_k = Q_*$ and $\lim P_k = P_*$. Then $Q_*$ is nonsingular and $Q_*^{-1}A_2 = F$, $I - A_1 - P_*$ is nonsingular and $(I - A_1 - P_*)^{-1}A_0 = G$. The matrix $Q_* - P_*$ is a singular $M$-matrix.

**Proof.** By Theorem 4.3, $Q_* = I - A_1 - A_0 F$ and $I - A_1 - P_* = I - A_1 - A_2 G$. These two matrices are known to be nonsingular [32]. Since $Q_* F = (I - A_1 - A_0 F) F = A_2$, $Q_*^{-1}A_2 = F$. Since $(I - A_1 - P_*) G = (I - A_1 - A_2 G) G = A_0, (I - A_1 - P_*)^{-1} A_0 = G$. $Q_* - P_*$ is a singular $M$-matrix since

\[(Q_* - P_*) e = (I - A_1 - A_0 F - A_2 G) e = e - (A_1 + A_0 + A_2) e = 0.\]

This completes the proof. \[\Box\]

When the QBD is null recurrent, the interpretation of the CR algorithm as a doubling algorithm has allowed us to show that the minimal solutions $G$ and $F$ can be found by the CR algorithm (or the closely related LR algorithm) simultaneously and with at least linear convergence with rate $1/2$. It is important to note that we no longer need the assumption that the matrices $G$ and $F$ have no eigenvalues on the unit circle other than the simple eigenvalue 1. With that assumption, one would use the shift technique as studied in [25], [16], and [4], and apply the CR algorithm or the LR algorithm to the shifted equation. When $G$ and $F$ have more than one eigenvalue on the unit circle, the shift technique is not helpful and the CR algorithm or the LR algorithm will be applied directly to the equation (4.1).

**5. A nonsymmetric algebraic Riccati equation.** In this section we consider the nonsymmetric algebraic Riccati equation (NARE)

\[(5.1) \quad X CX - XD - AX + B = 0,\]
where \( A, B, C, D \) are real matrices of sizes \( m \times m, m \times n, n \times m, n \times n \), respectively, and the matrix

\[
K = \begin{bmatrix}
D & -C \\
- B & A
\end{bmatrix}
\]

is a nonsingular \( M \)-matrix or an irreducible singular \( M \)-matrix. The NARE arises in the study of Wiener-Hopf factorization of Markov chains [37], and it includes the NARE arising from transport theory [29, 30]. We will also need the dual equation of (5.1)

\[
Y B Y - Y A - D Y + C = 0,
\]

which is in the same form of (5.1).

We will use the elementwise order for matrices: for any matrices \( A = [a_{ij}], B = [b_{ij}] \in \mathbb{R}^{m \times n} \), we write \( A \geq B \) (\( A > B \)) if \( a_{ij} \geq b_{ij} \) (\( a_{ij} > b_{ij} \)) for all \( i, j \).

A basic result about (5.1) and (5.3) is the following [14].

**Theorem 5.1.** If the matrix \( K \) in (5.2) is a nonsingular \( M \)-matrix or an irreducible singular \( M \)-matrix, then the NARE (5.1) and the NARE (5.3) have minimal nonnegative solutions \( X \) and \( Y \), respectively. Moreover, \( D - C X \) and \( A - B Y \) are \( M \)-matrices.

The minimal nonnegative solution of the NARE is the solution of practical interest. There have been a number of methods for finding this solution. The methods and their analyses can be found in [2, 14, 18, 20, 21, 23, 24, 36]. Among the iterative methods, the doubling algorithm proposed in [24] stands out for its overall efficiency. The algorithm is analyzed in [24] for the case when \( K \) is a nonsingular \( M \)-matrix, and is analyzed in [21] for the case when \( K \) is an irreducible singular \( M \)-matrix. When \( K \) is an irreducible singular \( M \)-matrix, we let \( [u_1^T, u_2^T]^T > 0 \) and \( [v_1^T, v_2^T]^T > 0 \) be the right and the left null vectors of \( K \) in (5.2), respectively. If \( u_1^T v_1 \neq u_2^T v_2 \), then the convergence of the doubling algorithm is still quadratic; if \( u_1^T v_1 = u_2^T v_2 \), then the convergence is observed to be linear with rate \( 1/2 \) (see [21]). The latter case will be referred to as the critical case for the NARE. For this critical case, the convergence of Newton’s method has been shown to at least linear with rate \( 1/2 \) [14, 20, 23]. We will reach the same conclusion for the doubling algorithm.

We start with a brief review of the doubling algorithm in [24]. Let

\[
H = \begin{bmatrix}
D & -C \\
B & -A
\end{bmatrix},
\]

and

\[
R = D - C X, \quad S = A - B Y,
\]

where \( X \) and \( Y \) are given in Theorem 5.1. Then the NAREs (5.1) and (5.3) can be rewritten as

\[
H \begin{bmatrix} I_n \\ X \end{bmatrix} = \begin{bmatrix} I_n \\ X \end{bmatrix} R
\]

and

\[
H \begin{bmatrix} Y \\ I_m \end{bmatrix} = \begin{bmatrix} Y \\ I_m \end{bmatrix} (-S).
\]
Applying the Cayley transform to (5.6) with a scalar \( \gamma > 0 \) we have

\[
(H - \gamma I) \begin{bmatrix} I_n \\ X \end{bmatrix} = (H + \gamma I) \begin{bmatrix} I_n \\ X \end{bmatrix} R_\gamma,
\]

where \( R_\gamma = (R + \gamma I_n)^{-1}(R - \gamma I_n) \). Premultiplying the above equation by a proper nonsingular matrix gives

\[
M_0 \begin{bmatrix} I_n \\ X \end{bmatrix} = L_0 \begin{bmatrix} I_n \\ X \end{bmatrix} R_\gamma,
\]

Here \( L_0 \) and \( M_0 \) are given by (2.1) with

\[
E_0 = I_n - 2\gamma V_\gamma^{-1}, \quad F_0 = I_m - 2\gamma W_\gamma^{-1},
\]

\[
G_0 = 2\gamma D_\gamma^{-1} CW_\gamma^{-1}, \quad H_0 = 2\gamma W_\gamma^{-1} BD_\gamma^{-1},
\]

where

\[
A_\gamma = A + \gamma I_m, \quad D_\gamma = D + \gamma I_n,
\]

\[
W_\gamma = A_\gamma - BD_\gamma^{-1} C, \quad V_\gamma = D_\gamma - CA_\gamma^{-1} B.
\]

Similarly,

\[
M_0 \begin{bmatrix} Y \\ I_m \end{bmatrix} S_\gamma = L_0 \begin{bmatrix} Y \\ I_m \end{bmatrix},
\]

where \( S_\gamma = (S + \gamma I_m)^{-1}(S - \gamma I_m) \).

In this section SDA-1 denotes Algorithm 2.1 with \( E_0, F_0, G_0, H_0 \) given by (5.9).

The following result from \([21]\) improves the original results given in \([24]\).

**Theorem 5.2.** Let the matrix \( K \) in (5.2) be a nonsingular \( M \)-matrix or an irreducible singular \( M \)-matrix, and \( X, Y \geq 0 \) be the minimal nonnegative solutions of the NAREs (5.1) and (5.3), respectively. If \( \gamma \) satisfies

\[
\gamma \geq \gamma_0 \equiv \max \left\{ \max_{1 \leq i \leq m} a_{ii}, \max_{1 \leq i \leq n} d_{ii} \right\},
\]

where \( a_{ii} \) and \( d_{ii} \) are the diagonal entries of \( A \) and \( D \), respectively, then the sequence \( \{E_k, F_k, H_k, G_k\} \) in SDA-1 is well defined. Moreover, we have

(a) \( E_0, F_0 < 0 \) and \( E_k, F_k > 0 \) for \( k \geq 1 \);

(b) For \( k \geq 0 \), \( 0 \leq H_k < H_{k+1} < X \), \( 0 \leq G_k < G_{k+1} < Y \);

(c) For \( k \geq 0 \), \( I_m - H_k G_k \) and \( I_n - G_k H_k \) are nonsingular \( M \)-matrices.

From now on we assume that \( K \) in (5.2) is an irreducible singular \( M \)-matrix, and consider the critical case of the NARE (5.1). We always assume that \( \gamma \) satisfies (5.12).

The Kronecker form for the pencil \((M_0, L_0)\) can be determined with the help of the following result [14], where \( \mathbb{C}_- \) and \( \mathbb{C}_+ \) denote the open left and the open right half planes, respectively.

**Theorem 5.3.** For the critical case of the NARE (5.1), the matrix \( H \) has \( n - 1 \) eigenvalues in \( \mathbb{C}_+ \), \( m - 1 \) eigenvalues in \( \mathbb{C}_- \), and two zero eigenvalues with a quadratic divisor. Moreover, \( R \) and \( S \) in (5.5) are irreducible singular \( M \)-matrices (so each of them has a simple eigenvalue 0 and the remaining eigenvalues are in \( \mathbb{C}_+ \)).
In view of Theorem 5.3, the properties of the Cayley transform, and the process leading to (5.8) and (5.11), we know that there are nonsingular matrices $V$ and $Z$ such that

\begin{align}
VL_0Z &= \begin{bmatrix} I_n & 0_{n,m} \\ 0_{m,n} & J_{2,s} \oplus [1] \end{bmatrix} = J_L, \\
VM_0Z &= \begin{bmatrix} J_1 & \Gamma \\ 0_{m,n} & I_{m-1} \oplus [-1] \end{bmatrix} = J_M,
\end{align}

in which

\begin{align}
J_1 &= J_{1,s} \oplus [-1] \sim R_{\gamma}, \\
J_2 &= J_{2,s} \oplus [-1] \sim S_{\gamma}, \\
\Gamma &= 0_{n-1,m-1} \oplus [1] \equiv e_ne_n^T,
\end{align}

where $\rho(J_{1,s}) < 1$, $\rho(J_{2,s}) < 1$, and "$\sim_n$" denotes the similarity transformation. Since $J_LJ_M = J_MJ_L$, for the matrices $L_k$ and $M_k$ given by (2.2) we have by (2.11)

\begin{equation}
M_kZJ_L^{2k} = L_kZJ_M^{2k}.
\end{equation}

On the other hand, there are nonsingular matrices $T$ and $W$ such that

\begin{align}
TL_0W &= \begin{bmatrix} J_2 \\ 0_{n,m} \end{bmatrix} \hat{\Gamma} = \hat{J}_L, \\
TM_0W &= \begin{bmatrix} I_m \\ 0_{n,m} \end{bmatrix} \begin{bmatrix} J_1 \\ 0_{m,n} \end{bmatrix} = \hat{J}_M,
\end{align}

where $\hat{\Gamma} = e_me_m^T$. We now have

\begin{equation}
L_kW\hat{J}_L^{2k} = M_kW\hat{J}_M^{2k}.
\end{equation}

The following result determines the convergence rate of SDA-1 in the critical case.

**Theorem 5.4.** Let $X, Y \geq 0$ be the minimal nonnegative solutions of the NAREs (5.1) and (5.3), respectively, and let $\{E_k, F_k, G_k, H_k\}$ be generated by SDA-1. Then for the critical case

$$
\|E_k\| = O(2^{-k}), \quad \|F_k\| = O(2^{-k}), \quad \|H_k - X\| = O(2^{-k}), \quad \|G_k - Y\| = O(2^{-k}).
$$

**Proof.** Partition the matrices $Z$ and $W$ as

\begin{equation}
Z = \begin{bmatrix} Z_1 & Z_3 \\ Z_2 & Z_4 \end{bmatrix}, \quad W = \begin{bmatrix} W_1 & W_3 \\ W_2 & W_4 \end{bmatrix},
\end{equation}

where $Z_1, Z_3 \in \mathbb{R}^{n \times n}$, and $Z_4, W_2 \in \mathbb{R}^{m \times m}$. Then from (5.13) and (5.14), and from (5.17) and (5.18), we have

\begin{equation}
M_0 \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = L_0 \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} J_1, \quad M_0 \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} J_2 = L_0 \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}.
\end{equation}

Comparing (2.21) with (5.8) and (5.11), and using (5.15), we know that $Z_1$ and $W_2$ are invertible and $X = Z_2Z_1^{-1}$, $Y = W_1W_2^{-1}$.

Note that for $k \geq 1$ we have

$$
\begin{align}
J_L^{2k} &= \begin{bmatrix} I_n & 0 \\ 0 & J_L^{2k} \end{bmatrix}, \quad \hat{J}_L^{2k} = \begin{bmatrix} J_L^{2k} \\ 0 \end{bmatrix}, \\
J_M^{2k} &= \begin{bmatrix} J_M^{2k} & \Gamma_k \\ \Gamma_k & I_m \end{bmatrix}, \quad \hat{J}_M^{2k} = \begin{bmatrix} J_M^{2k} \\ 0 \end{bmatrix}, \\
\hat{J}_L^{2k} &= \begin{bmatrix} J_L^{2k} \\ 0 \end{bmatrix}, \quad \hat{J}_M^{2k} = \begin{bmatrix} J_M^{2k} \\ 0 \end{bmatrix},
\end{align}
$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
where \( \Gamma_k = -2^k \Gamma = -2^k e_n e_n^T \), \( \hat{\Gamma}_k = -2^k \hat{\Gamma} = -2^k e_m e_m^T \). It follows from (5.16) and (5.19) that for \( k \geq 1 \)

\begin{align}
(5.22) & \quad E_k Z_1 = (Z_1 - G_k Z_2) J_1^{2^k}, \\
(5.23) & \quad E_k Z_3 J_2^{2^k} = (Z_1 - G_k Z_2) \Gamma_k + (Z_3 - G_k Z_4), \\
(5.24) & \quad -H_k Z_1 + Z_2 = F_k Z_2 J_1^{2^k}, \\
(5.25) & \quad (-H_k Z_3 + Z_4) J_2^{2^k} = F_k Z_2 \Gamma_k + F_k Z_4, \\
(5.26) & \quad W_1 - G_k W_2 = E_k W_1 J_2^{2^k}, \\
(5.27) & \quad (W_3 - G_k W_4) J_3^{2^k} = E_k W_3 \hat{\Gamma}_k + E_k W_3, \\
(5.28) & \quad F_k W_2 = (W_2 - H_k W_1) J_2^{2^k}, \\
(5.29) & \quad F_k W_4 J_1^{2^k} = (W_2 - H_k W_1) \hat{\Gamma}_k + (W_4 - H_k W_3).
\end{align}

Postmultiplying (5.29) by \( \hat{\Gamma}_k^* = -2^{-k} \Gamma \), the Moore–Penrose pseudo inverse of \( \hat{\Gamma}_k \), subtracting the result from (5.28), and noting that \( \hat{\Gamma}_k \hat{\Gamma}_k^* = 0_{m-1} \oplus [1] \), we get

\begin{align}
(5.30) & \quad F_k \left( W_2 + 2^{-k} W_4 J_3^{2^k} \Gamma \right) = (W_2 - H_k W_1) \left( J_2^{2^k} \oplus [0] \right) + 2^{-k} (W_4 - H_k W_3) \Gamma.
\end{align}

Since \( W_2 \) is invertible and \( \{H_k\} \) is bounded by Theorem 5.2(b), it follows from (5.30) that \( \|F_k\| = O(2^{-k}) \). It then follows from (5.24) that \( \|H_k - X\| = O(2^{-k}) \).

Similarly, postmultiplying (5.23) by \( \Gamma_k^* = -2^{-k} \hat{\Gamma} \), subtracting the result from (5.22), and noting that \( \Gamma_k \Gamma_k^* = 0_{n-1} \oplus [1] \), we get

\begin{align}
(5.31) & \quad E_k \left( Z_1 + 2^{-k} Z_3 J_2^{2^k} \hat{\Gamma} \right) = (Z_1 - G_k Z_2) \left( J_1^{2^k} \oplus [0] \right) + 2^{-k} (Z_3 - G_k Z_4) \hat{\Gamma}.
\end{align}

Since \( Z_1 \) is invertible and \( \{G_k\} \) is bounded by Theorem 5.2(b), it follows from (5.31) that \( \|E_k\| = O(2^{-k}) \). It then follows from (5.26) that \( \|G_k - X\| = O(2^{-k}) \).

We note that \( \lim(I - G_k H_k) = I - Y X \) and \( \lim(I - H_k G_k) = I - X Y \) are both singular \( M \)-matrices (see [21]).

The critical case we have considered is a singular case, and the singularity can be removed by applying a proper shift technique. Indeed, a shift technique has been introduced in [21] and SDA-1 applied to the shifted NARE has quadratic convergence if no breakdown happens. However, whether breakdown is possible remains an open problem in general, although some partial results have been obtained in [21].

Since \( K \) is an irreducible singular \( M \)-matrix, we may assume without loss of generality that \( Ke = 0 \). In this case, one can transform the NARE to a quadratic matrix equation of the type in section 4, but with \( (m+n) \times (m+n) \) matrices in the equation (see [36]). One can then apply CR and LR to the transformed equation (see [2, 18]). A specific shift technique (following [25]) is introduced in [18] to the transformed equation, and quadratic convergence is recovered for the LR algorithm (thus, also for the CR algorithm) if no breakdown happens. It has been shown in [20] that the LR algorithm is indeed well-defined when the shift technique is used. However, when \( m = n \), the computational work required in each iteration is nearly twice that for SDA-1, due to the dimension expansion from \( n \) to \( 2n \). If we use the shift technique in [18] with the CR approach in [2], then no breakdown happens and the complexity is down to \( 34n^3 \) flops each iteration when \( m = n \).

Although it is preferable to use a shift technique for the critical case of the NARE (with an irreducible singular \( M \)-matrix \( K \)), our convergence results in Theorem 5.4...
still provide some insights about the convergence behavior of SDA-1 for nearby NAREs with a nonsingular $M$-matrix $K$ (where the shift technique is no longer applicable). The exact solution of a singular NARE is quite sensitive to the input data in the NARE (see [20]). For the singular NARE and nearby NAREs, it would be reasonable to stop the iteration when $\|H_k - H_{k-1}\| < \epsilon^{1/2}$, where $\epsilon$ is the machine epsilon, and take $H_k$ as an approximation to the exact solution $X$. Further iterations for SDA-1 may not be able to improve the accuracy significantly in view of the perturbation behavior of $X$ and the fact that $I - G_k H_k$ and $I - H_k G_k$ are nearly singular for large $k$. So we are mainly interested in the behavior of SDA-1 for iterations up to the point where $\|H_k - H_{k-1}\| < \epsilon^{1/2}$ (assuming this is achievable). And up to that point, the behavior of SDA-1 for those nearby NAREs would be very much similar to that of SDA-1 for the singular NARE. We use one example to illustrate this point.

**Example 5.1.** Let $T$ be a $16 \times 16$ doubly stochastic matrix given by $T = \frac{1}{1056} \text{magic}(16)$, where \text{magic} is the Matlab function that generates magic squares. Let $K = I - T$, and let the $8 \times 8$ matrices $A, B, C, D$ be determined through (5.2). The matrix $K$ is an irreducible singular $M$-matrix, and we have the critical case for the NARE (5.1). We take $\gamma$ to be the largest diagonal entry of $K$ (which is the last diagonal entry of $K$) and apply SDA-1. We find that $\|H_k - H_{k-1}\| < 10^{-7}$ is satisfied for $k = 24$. The convergence rate of $H_k - X$ is determined through that of $F_k$ (see the proof of Theorem 5.4). We find that the values of $\sqrt[8]{\|F_k\|_\infty}$ are between 0.4924 and 0.5001 for $k = 4 : 24$.

We then increase the (1,1) entry of $K$ by $10^{-12}$. So $K$ is now a nonsingular $M$-matrix. The matrix $D$ is changed accordingly. The change in $K$ does not change the largest diagonal entry of $K$. So we apply SDA-1 to the new NARE with the same $\gamma$. We find that $\|H_k - H_{k-1}\| < 10^{-7}$ is satisfied for $k = 23$, and that the values of $\sqrt[8]{\|F_k\|_\infty}$ are between 0.4924 and 0.5000 for $k = 4 : 21$ (the values are 0.4855 and 0.4570 for $k = 22$ and $k = 23$, respectively). Thus, the (nonterminal and more important) convergence behavior of SDA-1 for this nearby NARE is largely dictated by our theoretical results in Theorem 5.4.

**6. Conclusion.** We have determined the convergence rate of the doubling algorithm in the critical (or singular) case for three different nonlinear matrix equations. It is possible to apply the techniques we reviewed in section 2 to other nonlinear matrix equations. Through this study, we have also gained more insights for the convergence behavior for the doubling algorithm for nearly singular cases.

**REFERENCES**


