Convergence Analysis of the Structure-Preserving Doubling Algorithms for Nonlinear Matrix Equations

Chun-Yueh Chiang

Department of Mathematics
Chiao Tung University, Hsinchu

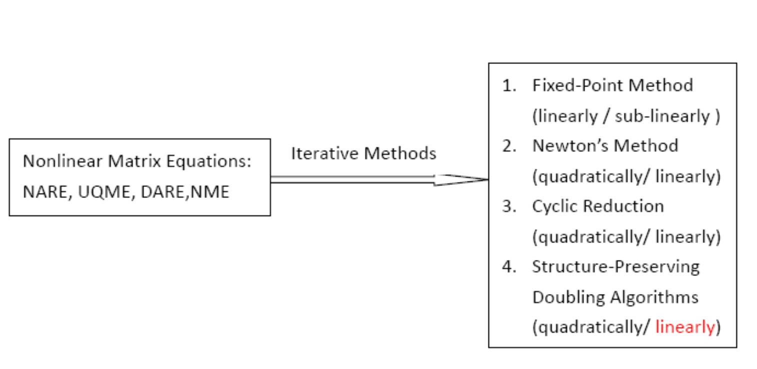
Oct., 13, 2008

Outline

Outline

1. Introduction
   - Matrix background
   - SDA
2. NARE
   - Introduction and Preliminaries
   - Convergence analysis of SDA algorithm
   - Numerical Examples
3. Another Equations
   - UQME
   - DARE
   - NME
4. Concluding Remarks

Our Work


Nonlinear Matrix Equations:
NARE, UQME, DARE,NME
Iterative Methods
1. Fixed-Point Method
(linearly / sub-linearly)
2. Newton's Method
(quadratically/ linearly)
3. Cyclic Reduction
(quadratically/ linearly)
4. Structure-Preserving
Doubling Algorithms
(quadratically/ linearly)

Procedure

Matrix Eqs.

SLF-1,2

$$\mathcal{M}\begin{bmatrix} I \\ X \end{bmatrix} = \mathcal{L}\begin{bmatrix} I \\ X \end{bmatrix} R$$

Cayley trans.

$$\mathcal{H}\begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix} R$$

Ham.-LF

Doubling algs.

$$\mathcal{M}_k\begin{bmatrix} I \\ X \end{bmatrix} = \mathcal{L}_k\begin{bmatrix} I \\ X \end{bmatrix} R^{2^k}$$

SDA-1,2

$\rho(R) < 1$, convergence quadratically.

$\rho(R) = 1$, no information.

Sketch the Proof

Let $R \sim J_1 \oplus J_s$, where $\rho(J_1) = 1$, $\rho(J_s) < 1$. If the eigenvalues in J_1 are not semi-simple. For convenient, let $J_1 = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$ with $|\lambda| = 1$. It is easily seen that $J_1^{2^k} = \begin{bmatrix} \lambda^{2^k} & 2^k \lambda^{2^{k-1}} \\ 0 & \lambda^{2^k} \end{bmatrix}$. Moreover, we have

$$\mathcal{M}_k U = \mathcal{L}_k U \begin{bmatrix} J_s^{2^k} & & \\ & \lambda^{2^k} & 2^k \lambda^{2^{k-1}} \\ & & \lambda^{2^k} \end{bmatrix}.$$

If $O(\mathcal{M}_k) = O(\mathcal{L}_k) = 1$, we should get some information by comparing both sides.

Our tools

In this thesis, we only using elementary matrix theory:

- Weierstrass canonical form
  NARE, UQME, DARE.
- Symplectic triangular Kronecker canonical form
  NME.

We need choose matrices $\mathcal{J}_{\mathcal{M}}$ and $\mathcal{J}_{\mathcal{L}}$ such that

- $\mathcal{M} - \lambda\mathcal{L} \sim \mathcal{J}_{\mathcal{M}} - \lambda\mathcal{J}_{\mathcal{L}}$,
- $\mathcal{J}_{\mathcal{M}}\mathcal{J}_{\mathcal{L}} = \mathcal{J}_{\mathcal{L}}\mathcal{J}_{\mathcal{M}}$.

Cayley Transformation

Let $\gamma > 0$, we define the function $C_\gamma : \mathbb{C}/\{-\gamma\} \to \mathbb{C}$ if $\forall \lambda \in \mathbb{C}$ then $C_\gamma(\lambda) = \frac{\lambda-\gamma}{\lambda+\gamma}$. It is easily seen that

$$\lambda \in \mathbb{C}_+ \Longleftrightarrow |C_\gamma(\lambda)| < 1.$$

If C_γ is defined by $C_\gamma(A) = (A + \gamma I)^{-1}(A - \gamma I)$, where $A \in \mathbb{C}^{n\times n}$ with $-\gamma \notin \sigma(A)$, we also have

$$\sigma(A) \subset \mathbb{C}_+ \Longleftrightarrow \rho(C_\gamma(A)) < 1.$$

Beginning

Finding one or more roots of a matrix equation $F(x) = 0$ is one of the more commonly occurring problems of applied mathematics. In most cases explicit solutions are not available, the numerical methods for finding the roots are called iterative methods. Two classical methods

1. Fixed-Point Iteration

$$X_{k+1} = G(X_k), \quad X_0 \text{ is given},$$

   where $G(X) \equiv X - F(X)$.

2. Newton's Iteration

$$X_{k+1} = X_k - (F'_{X_k})^{-1}(F(X_k)), \quad X_0 \text{ is given},$$

   where F'_Z denoted the Fréchet derivative of F at Z.

Doubling Algorithm

Basic idea:

1. Single iteration

$$X_1 \to X_2 \to X_3 \cdots \to X_k \to \cdots .$$

2. Double iteration

$$X_1 \to X_2 \to X_4 \cdots \to X_{2^k} \to \cdots .$$

Example 1-1

We now consider the simple iteration:

$$f_k = af_{k-1} + b,$$

where $a, b \in \mathbb{R}$ with $|a| < 1$, f_1 is given.

It is easily seen that

$$
\begin{aligned}
f_2 &= af_1 + b \\
f_4 &= af_3 + b = a^2 f_2 + b(1+a) \\
f_8 &= (a^2)^2 f_4 + b(1+a)(1+a^2) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
f_{2^k} &= a^{2^{k-1}} f_{2^{k-1}} + b(1+a)\cdots(1+a^{2^{k-2}}).
\end{aligned}
$$

We get

$$
\begin{cases}
g_k & \equiv f_{2^k} = a^{2^{k-1}} g_{k-1} + b_{k-1}, \quad g_1 = f_1 \\
b_k & = (1+a^{2^{k-2}}) b_{k-1}, \quad b_1 = b.
\end{cases}
$$

Let $f_* \equiv \dfrac{b}{1-a}$, we have

$f_k - f_* = a(f_{k-1} - f_*) = \ldots = a^{k-1}(f_1 - f_*)$.

Since

$b_k = (1 + a^{2^{k-2}}) b_{k-1} = \cdots = (1 + a^{2^{k-2}}) \cdots (1 + a) b = b \dfrac{1 - a^{2^{k-1}}}{1 - a}$

The new iteration is satisfying

$$g_k - f_* = a^{2^{k-1}} g_{k-1} + b \frac{1 - a^{2^{k-1}}}{1 - a} - \frac{b}{1 - a} = a^{2^{k-1}} (g_{k-1} - f_*).$$

Thus, we have

$$\limsup_{k \to \infty} \sqrt[k]{\| f_k - f_* \|} = |a|,$$

$$\limsup_{k \to \infty} \sqrt[2^k]{\| g_k - f_* \|} = |a|.$$

In this talk, we review two types of structure-preserving doubling algorithm (and denoted by SDA). Moreover, we use the techniques to study the SDA in the following four different nonlinear matrix equations in the critical case.

(1) Nonsymmetric algebraic Riccati equation (NARE)

$$XCX - XD - AX + B = 0,$$

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

(2) Unilateral quadratic matrix equation (UQME)

$$A_0 + A_1X + A_2X^2 = X,$$

where $A_0, A_1, A_2 \in \mathbb{R}^{n\times n}$.

(3) Discrete-time algebraic Riccati equation (DARE)

$$-X+A^\top XA+Q-(C+B^\top XA)^\top(R+B^\top XB)^{-1}(C+B^\top XA)=0,$$

where $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times m}$, $C\in\mathbb{R}^{m\times n}$, $Q=Q^\top\in\mathbb{R}^{n\times n}$ and $R=R^\top\in\mathbb{R}^{m\times m}$.

(4) Nonlinear matrix equation (NME)

$$X+A^\top X^{-1}A=Q,$$

where $A,Q\in\mathbb{R}^{n\times n}$.

The canonical form of a matrix pencil

Given a matrix pencil $\mathcal{M} - \lambda\mathcal{L} \in \mathbb{C}^{n\times n}$, we say that $\mathcal{M} - \lambda\mathcal{L}$ is regular if there exist a scalar $\lambda_0 \in \mathbb{C}$ such that $\det(\mathcal{M} - \lambda_0\mathcal{L}) \neq 0$. The following theorem generalizes the Jordan canonical form of a single matrix to a regular pencils.

Theorem 1-2: Weierstrass canonical form [Gantmach77]

Let $\mathcal{M} - \lambda\mathcal{L}$ be regular. Then there are nonsingular U and V such that

$$U\mathcal{M}V = \begin{bmatrix} J & \\ & I \end{bmatrix}, \quad U\mathcal{L}V = \begin{bmatrix} I & \\ & N \end{bmatrix},$$

where J and N are in Jordan canonical form and N is nilpotent.

Let

$$J_p(\lambda) \equiv \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix} \in \mathbb{C}^{p\times p}. \qquad (1.1)$$

Assume that the finite eigenvalues of $\mathcal{M} - \lambda\mathcal{L}$ are $J_1 \oplus J_2$, $J_2 = J_{\lambda_1,2m_1} \oplus \cdots \oplus J_{\lambda_k,2m_k}$. we write

$$J_{\lambda_i,2m_i} = \begin{bmatrix} J_{\lambda,m} & \Gamma_{m_i} \\ 0 & J_{\lambda,m} \end{bmatrix},$$

where $\Gamma_{m_i} = e_{m_i} e_{m_i}^\top$. After some permutations we have

$$P^\top J_2 P = \left[\begin{array}{ccc|ccc} J_{\lambda_1,m_1} & & & \Gamma_{m_1} & & \\ & \ddots & & & \ddots & \\ & & J_{\lambda_k,m_k} & & & \Gamma_{m_k} \\ \hline & & & J_{\lambda_1,m_1} & & \\ & & & & \ddots & \\ & & & & & J_{\lambda_k,m_k} \end{array}\right]$$

for a permutation matrix P. By Theorem 1-2, we have

$$\mathcal{M} - \lambda\mathcal{L} \sim \left[\begin{array}{ccc|c} J_1 & & & \\ & \widetilde{J}_2 & \Gamma & \\ & & \widetilde{J}_2 & \\ \hline & & & I \end{array}\right] - \lambda \left[\begin{array}{c|c} I & \\ \hline & N \end{array}\right]$$

$$\sim \left[\begin{array}{cc|cc} J_1 & & 0 & \\ & \widetilde{J}_2 & & \Gamma \\ \hline & & I & \\ & & & \widetilde{J}_2 \end{array}\right] - \lambda \left[\begin{array}{c|cc} I & & \\ \hline & N & \\ & & I \end{array}\right].$$

Structure-preserving doubling algorithm

Let the matrix pencil $\mathcal{M} - \lambda\mathcal{L} \in \mathbb{R}^{(n+m)\times(n+m)}$ and define the left null space

$$\mathcal{N}(\mathcal{M},\mathcal{L}) \equiv \left\{ \begin{bmatrix} \mathcal{M}_* & \mathcal{L}_* \end{bmatrix}; \mathcal{M}_*, \mathcal{L}_* \in \mathbb{R}^{(n+m)\times(n+m)}, \operatorname{rank} \begin{bmatrix} \mathcal{M}_* & \mathcal{L}_* \end{bmatrix} = n+m, \right.$$
$$\left. \begin{bmatrix} \mathcal{M}_* & \mathcal{L}_* \end{bmatrix} \begin{bmatrix} \mathcal{L} \\ -\mathcal{M} \end{bmatrix} = 0 \right\}.$$

Since $\operatorname{rank} \begin{bmatrix} \mathcal{L} \\ -\mathcal{M} \end{bmatrix} \leq n+m$, thus $\operatorname{nullity} \begin{bmatrix} \mathcal{L} \\ -\mathcal{M} \end{bmatrix}^\top \geq n+m$ and it follows that $\mathcal{N}(\mathcal{M},\mathcal{L}) \neq \phi$. For any given $\begin{bmatrix} \mathcal{M}_* & \mathcal{L}_* \end{bmatrix} \in \mathcal{N}(\mathcal{M},\mathcal{L})$, define

$$\widetilde{\mathcal{M}} = \mathcal{M}_*\mathcal{M}, \quad \widetilde{\mathcal{L}} = \mathcal{L}_*\mathcal{L}.$$

The transformation

$$\mathcal{M} - \lambda\mathcal{L} \to \widetilde{\mathcal{M}} - \lambda\widetilde{\mathcal{L}}$$

is called a doubling transformation.

An important feature of this kind of transformation is that it is structure-preserving, eigenspace-preserving, and eigenvalue-squaring [Lin06]. If

$$\mathcal{M}x = \lambda \mathcal{L}x,$$

then

$$\begin{aligned} \widetilde{\mathcal{M}}x &= \mathcal{M}_*\mathcal{M}x = \lambda \mathcal{M}_*\mathcal{L}x \\ &= \lambda \mathcal{L}_*\mathcal{M}x = \lambda^2 \mathcal{L}_*\mathcal{L}x = \lambda^2 \widetilde{\mathcal{L}}x. \end{aligned}$$

We quote the basic properties in the following theorem.

Theorem 1-3: Properties [Lin06]

Assume that the matrix pencil $\widetilde{\mathcal{M}} - \lambda\widetilde{\mathcal{L}}$ is the result of a doubling transformation of the pencil $\mathcal{M} - \lambda\mathcal{L}$. The matrix pencil $\mathcal{M} - \lambda\mathcal{L}$ has the Weierstrass canonical form

$$U\mathcal{M}V = \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix}, \quad U\mathcal{L}V = \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}$$

where U, V are nonsingular, J is a Jordan block, and N is a nilpotent matrix, then there exists a nonsingular $\widetilde{U}$ such that

$$\widetilde{U}\widetilde{\mathcal{M}}V = \begin{bmatrix} J^2 & 0 \\ 0 & I \end{bmatrix}, \quad \widetilde{U}\widetilde{\mathcal{L}}V = \begin{bmatrix} I & 0 \\ 0 & N^2 \end{bmatrix}.$$

Symplectic-like pencil form

- A pencil $\mathcal{M}_1 - \lambda\mathcal{L}_1$ is said to be in first standard symplectic-like pencil form (SLF-1) if it has the form

$$\mathcal{M}_1 = \begin{bmatrix} E & 0 \\ -H & I \end{bmatrix}, \quad \mathcal{L}_1 = \begin{bmatrix} I & -G \\ 0 & F \end{bmatrix}, \qquad (1.2)$$

where $E \in \mathbb{R}^{n\times n}$, $F \in \mathbb{R}^{m\times m}$, $H \in \mathbb{R}^{m\times n}$ and $G \in \mathbb{R}^{n\times m}$. Suppose that $I - GH$ is nonsingular, we can take the left null matrix $\begin{bmatrix} \mathcal{M}_{1,*} & \mathcal{L}_{1,*} \end{bmatrix}$, where

$$\mathcal{M}_{1,*} = \begin{bmatrix} E(I - GH)^{-1} & 0 \\ -F(I - HG)^{-1}H & I \end{bmatrix}, \quad \mathcal{L}_{1,*} = \begin{bmatrix} I & -E(I - GH)^{-1}G \\ 0 & F(I - HG)^{-1} \end{bmatrix}.$$

The doubling transformation is given by

$$\widetilde{\mathcal{M}_1} \equiv \mathcal{M}_{1,*}\mathcal{M}_1 = \begin{bmatrix} E(I-GH)^{-1}E & 0 \\ -(H+F(I-HG)^{-1}HE) & I \end{bmatrix},$$

$$\widetilde{\mathcal{L}_1} \equiv \mathcal{L}_{1,*}\mathcal{L}_1 = \begin{bmatrix} I & -(G+E(I-GH)^{-1}GF) \\ 0 & F(I-HG)^{-1}F \end{bmatrix}.$$

Therefore, we define the sequence $\{\mathcal{M}_{1,k}, \mathcal{L}_{1,k}\}$ by the following structured doubling algorithm-1 (SDA-1) if no breakdown occurs.

Algorithm of SDA-1

Given

$$E_0, F_0, G_0, H_0,$$

for $k = 0, 1, \ldots,$

Set

$$\begin{aligned}
E_{k+1} &= E_k(I_n - G_k H_k)^{-1} E_k, && (1.3a)\\
F_{k+1} &= F_k(I_m - H_k G_k)^{-1} F_k, && (1.3b)\\
G_{k+1} &= G_k + E_k(I_n - G_k H_k)^{-1} G_k F_k, && (1.3c)\\
H_{k+1} &= H_k + F_k(I_m - H_k G_k)^{-1} H_k E_k. && (1.3d)
\end{aligned}$$

End of algorithm

For the SDA-1, we have the following count for one iteration:

Calculation in SDA-1	Flops
$G_k H_k$	$2n^2m$
$H_k G_k$	$2m^2n$
LU decomposition of $I_n - G_k H_k$	$\frac{2}{3}n^3$
LU decomposition of $I_m - H_k G_k$	$\frac{2}{3}m^3$
$temp1 = E_k(I_n - G_k H_k)^{-1}$	$2n^3$
$temp2 = F_k(I_m - H_k G_k)^{-1}$	$2m^3$
$E_{k+1} = temp1 * E_k$	$2n^3$
$F_{k+1} = temp2 * F_k$	$2m^3$
$G_{k+1} = G_k + temp1 * G_k * F_k$	$4n^2m$
$H_{k+1} = H_k + temp2 * H_k * E_k$	$4m^2n$
The total count $=$	$\frac{14}{3}(m^3 + n^3) + 6(m^2n + n^2m)$

We have ignored any $O(n^2)$ operation counts and the memory counts, note that the flop count is $\frac{64}{3}n^3$ when $m = n$.

- A pencil $\mathcal{M}_2 - \lambda\mathcal{L}_2$ is said to be in second standard symplectic-like pencil form (SLF-2) if it has the form

$$\mathcal{M}_2 = \begin{bmatrix} V & 0 \\ Q & -I \end{bmatrix}, \quad \mathcal{L}_2 = \begin{bmatrix} -P & I \\ T & 0 \end{bmatrix}, \tag{1.4}$$

where $V \in \mathbb{R}^{n\times n}$, $P \in \mathbb{R}^{n\times n}$, $Q \in \mathbb{R}^{m\times n}$ and $T \in \mathbb{R}^{m\times n}$. Suppose that $Q - P$ is nonsingular, we can take the left null matrix $\begin{bmatrix} \mathcal{M}_{2,*} & \mathcal{L}_{2,*} \end{bmatrix}$, where

$$\mathcal{M}_{2,*} = \begin{bmatrix} V(Q-P)^{-1} & 0 \\ -T(Q-P)^{-1} & I \end{bmatrix}, \quad \mathcal{L}_{2,*} = \begin{bmatrix} I & -V(Q-P)^{-1} \\ 0 & T(Q-P)^{-1} \end{bmatrix}.$$

The doubling transformation is given by

$$\widetilde{\mathcal{M}}_2 \equiv \mathcal{M}_{2,*}\mathcal{M}_2 = \begin{bmatrix} V(Q-P)^{-1}V & 0 \\ Q - T(Q-P)^{-1}V & -I \end{bmatrix},$$

$$\widetilde{\mathcal{L}}_2 \equiv \mathcal{L}_{2,*}\mathcal{L}_2 = \begin{bmatrix} -(P + V(Q-P)^{-1}T) & I \\ T(Q-P)^{-1}T & 0 \end{bmatrix}.$$

Therefore, we define the sequence $\{\mathcal{M}_{2,k}, \mathcal{L}_{2,k}\}$ by the following structured doubling algorithm-2 (SDA-2) if no breakdown occurs.

Algorithm of SDA-2

Given

$$V_0, T_0, Q_0, P_0,$$

for $k = 0, 1, \ldots,$

Set

$$\begin{aligned}
V_{k+1} &= V_k(Q_k - P_k)^{-1}V_k, && (1.5a)\\
T_{k+1} &= T_k(Q_k - P_k)^{-1}T_k, && (1.5b)\\
Q_{k+1} &= Q_k - T_k(Q_k - P_k)^{-1}V_k, && (1.5c)\\
P_{k+1} &= P_k + V_k(Q_k - P_k)^{-1}T_k. && (1.5d)
\end{aligned}$$

End of algorithm

For the SDA-2, we have the following count for one iteration:

Calculation in SDA-2	Flops
LU decomposition of $Q_k - P_k$	$\frac{2}{3}n^3$
$temp1 = V_k(Q_k - P_k)^{-1}$	$2n^3$
$temp2 = T_k(Q_k - P_k)^{-1}$	$2n^3$
$V_{k+1} = temp1 * V_k$	$2n^3$
$T_{k+1} = temp2 * T_k$	$2n^3$
$G_{k+1} = Q_k - T_k(Q_k - P_k)^{-1}V_k$	$2n^3$
$H_{k+1} = P_k + V_k(Q_k - P_k)^{-1}T_k$	$2n^3$
The total count =	$\frac{38}{3}n^3$

We have ignored any $O(n^2)$ operation counts and the memory counts.

Outline

1. Introduction
   - Matrix background
   - SDA
2. NARE
   - Introduction and Preliminaries
   - Convergence analysis of SDA algorithm
   - Numerical Examples
3. Another Equations
   - UQME
   - DARE
   - NME
4. Concluding Remarks

NARE

We consider the nonsymmetric algebraic Riccati equation (NARE) in $X \in \mathbb{R}^{m\times n}$:

$$XCX - XD - AX + B = 0 \tag{2.1}$$

and its dual equation in $Y \in \mathbb{R}^{n\times m}$:

$$YBY - YA - DY + C = 0, \tag{2.2}$$

where $A \in \mathbb{R}^{m\times m}$, $D \in \mathbb{R}^{n\times n}$, B and $C^\top \in \mathbb{R}^{m\times n}$, arised from transport theory and the Wiener-Hopf factorization.

Some relevant definitions are given as follows.

Definition 2-1: Z and M matrix

(i) 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.

(ii) A matrix $A \in \mathbb{R}^{n \times n}$ is said to be a Z-matrix if all its off-diagonal elements are non-positive. A Z-matrix A is called a nonsingular M-matrix if $A = sI - B$ with $B \geq 0$ and $s > \rho(B)$, where $\rho(B)$ is the spectral radius of B; if $s = \rho(B)$, then A is called a singular M-matrix.

Some results about the M-matrix.

Lemma 2-2: Properties

Let A be an $n \times n$ Z-matrix,
$\mathbb{C}_+ \equiv \{z; \mathrm{Re}(z) > 0\}, \mathbb{C}_0 \equiv \{z; \mathrm{Re}(z) = 0\}$ and $\sigma(A)$ denote the spectrum of A,. The following statements are equivalent:

(a) A is a nonsingular M-matrix;

(b) $A^{-1} \geq 0$;

(c) $Av > 0$ holds for some $v > 0$;

(d) $\sigma(A) \subseteq \mathbb{C}_+$.

And the following statements are equivalent:

(a) A is a singular M-matrix;

(b) $\sigma(A) \subseteq \mathbb{C}_+ \cup \mathbb{C}_0$, $\sigma(A) \cap \mathbb{C}_0 \neq \phi$.

Consider the standard assumption

$$\textbf{(H)} \qquad \mathcal{K} = \begin{bmatrix} D & -C \\ -B & A \end{bmatrix} \text{ is an irr. sing. M-matrix.} \qquad (2.3)$$

We have

Theorem 2-3: Existence of solution [Guo01]

If the matrix $\mathcal{K}$ is an irr. M-matrix, then the NARE and its dual equation have minimal nonneg. solutions X and Y, resp., such that $D - CX$ and $A - BY$ are irr. M-matrices. Moreover, if $\mathcal{K}$ is a nonsingular M-matrix, then $D - CX$ and $A - BY$ are irreducible nonsingular M-matrices.

Define the Hamiltonian-like matrix

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

and let

$$R \equiv D - CX, \quad S \equiv A - BY,$$

Then the NARE and its dual eq. can be rewritten by

$$\mathcal{H} \begin{bmatrix} I_n \\ X \end{bmatrix} = \begin{bmatrix} I_n \\ X \end{bmatrix} R \tag{2.4}$$

and

$$\mathcal{H} \begin{bmatrix} Y \\ I_m \end{bmatrix} = \begin{bmatrix} Y \\ I_m \end{bmatrix} (-S),$$

respectively.

Spectral analysis

We now consider more strictly condition that $\mathcal{K}$ is a nonsingular M-matrix in (**H**), by Lemma 2-2, there exist a positive vector $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n+m} \end{bmatrix}$ such that $\mathcal{K}v > 0$. Let $V \equiv \operatorname{diag}(v_1, \ldots, v_{n+m}) \in \mathbb{R}^{(n+m)\times(n+m)}$ and $W \equiv V^{-1}\mathcal{H}V$. The Gershgorin's disk of W is

$$|\lambda - d_{ii}| \le \frac{1}{v_i}\Big(\sum_{1\le j\le n, j\ne i} v_j(-d_{ij}) + \sum_{1\le j\le m} v_{j+n}c_{ij}\Big), \quad 1 \le i \le n,$$

$$|\lambda + a_{ii}| \le \frac{1}{v_{i+n}}\Big(\sum_{1\le j\le n} v_j b_{ij} + \sum_{1\le j\le m, j\ne i} v_{j+n}(-a_{ij})\Big), \quad 1 \le i \le m.$$

The condition $\mathcal{K}v > 0$ implies that

$$d_{ii} - \frac{1}{v_i}\Big(\sum_{1\le j\le n, j\ne i} v_j(-d_{ij}) + \sum_{1\le j\le m} v_{j+n}c_{ij}\Big) > 0, \quad 1 \le i \le n,$$

$$a_{ii} - \frac{1}{v_{i+n}}\Big(\sum_{1\le j\le n} v_j b_{ij} + \sum_{1\le j\le m, j\ne i} v_{j+n}(-a_{ij})\Big) > 0, \quad 1 \le i \le m.$$

That is, there are n Gershgorin's disks in $\mathbb{C}_+$, m Gershgorin's disks in $\mathbb{C}_-$ and we conclude that there are exactly n eigenvalues in $\mathbb{C}_+$, the remain m eigenvalues in $\mathbb{C}_-$.

On the other hand, since X and Y are minimal nonnegative solutions of (2.1) and (2.2), respectively. We have

$$\begin{bmatrix} I & 0 \\ -X & I \end{bmatrix} \mathcal{H} \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} = \begin{bmatrix} D - CX & -C \\ 0 & -(A - XC) \end{bmatrix},$$

$$\begin{bmatrix} I & -Y \\ 0 & I \end{bmatrix} \mathcal{H} \begin{bmatrix} I & Y \\ 0 & I \end{bmatrix} = \begin{bmatrix} D - BY & 0 \\ B & -(A - BY) \end{bmatrix},$$

since $\sigma(D - CX) \subset \mathbb{C}_{>}$ and $\sigma(-(A - BY)) \subset \mathbb{C}_{<}$, we get $\sigma(A - XC) = \sigma(A - BY) \subset \mathbb{C}_{+}$, $\sigma(D - BY) = \sigma(D - CX) \subset \mathbb{C}_{+}$ and $\sigma(\mathcal{H}) = \sigma(D - CX) \cup \sigma(-(A - BY))$.

- Now under the assumption (**H**): $\mathcal{K}$ is irreducible singular M-matrix.
    1. By P-F theorem, nullity$(D - CX) \leq 1$, nullity$(A - BY) \leq 1$.
    2. By the continuity argument we know that there are at least $n - 1$ eigenvalues of $\mathcal{H}$ in $\mathbb{C}_+$, at least $m - 1$ eigenvalues of $\mathcal{H}$ in $\mathbb{C}_-$.

- Now under the assumption (**H**): $\mathcal{K}$ is irreducible singular M-matrix.
  1. By P-F theorem, nullity$(D - CX) \leq 1$, nullity$(A - BY) \leq 1$.
  2. By the continuity argument we know that there are at least $n-1$ eigenvalues of $\mathcal{H}$ in $\mathbb{C}_+$, at least $m-1$ eigenvalues of $\mathcal{H}$ in $\mathbb{C}_-$.
- It is interested when zero is a double eigenvalues of $\mathcal{H}$, what is the Jordan blocks associated with zero eigenvalue?
  - $\mathcal{H} = \begin{bmatrix} I_n & 0 \\ 0 & -I_m \end{bmatrix} \mathcal{K} \Rightarrow \mathrm{Ker}(\mathcal{H}) = \mathrm{Ker}(\mathcal{K})$.
  - Apply P-F Theorem, we have nullity$(\mathcal{K}) = 1$, nullity$(D - CX)$ and nullity$(A - BY)$ are less than 1.

Therefore, if zero is a double eigenvalue of $\mathcal{H}$ then the elementary divisors of $\mathcal{H}$ corresponding to zero have degrees 2 (or $\mathcal{H} \sim J_1 \oplus J_2$, where $J_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and J_2 consists of Jordan blocks associated with nonzero eigenvalues).

More precisely, let $[v_1^\top, v_2^\top]^\top > 0$ and $[u_1^\top, u_2^\top]^\top > 0$ be the right and the left null vectors of $\mathcal{K}$ in (2.3), respectively. The following essential result of [Guo01] determines the signs of real parts of all eigenvalues of $\mathcal{H}$ in (2.4). Recall that $R = D - CX$, $S = A - BY$ and $\sigma(\mathcal{H}) = \sigma(R) \cup \sigma(S)$.

Theorem 2-4: Spectral properties [Guo01]

Assume that **(H)** holds. Then

(1) Critical case: If $u_1^\top v_1 = u_2^\top v_2$, then $0 \in \sigma(R) \cap \sigma(S)$. Moreover, nullity$(R)$ =nullity$(S) = 1$.

(2) Non-critical case: If $u_1^\top v_1 > u_2^\top v_2$, then $0 \in \sigma(R)$ and $0 \notin \sigma(S)$. Moreover, nullity$(R) = 1$.

(3) Non-critical case: If $u_1^\top v_1 < u_2^\top v_2$, then $0 \in \sigma(S)$ and $0 \notin \sigma(R)$. Moreover, nullity$(S) = 1$

Since $\mathcal{K}$ in (2.3) is a irr. sing. M-matrix, it follows from Theorem 2-3 that R and S are M-matrices. By Lemma 2-2, it implies $\sigma(R) \subset \mathbb{C}_+ \cup \mathbb{C}_0$, $\sigma(S) \subset \mathbb{C}_+ \cup \mathbb{C}_0$. Using a Cayley transf. with some $\gamma > 0$, we can transform (2.4) into the form

$$(\mathcal{H} - \gamma I) \begin{bmatrix} I_n \\ X \end{bmatrix} = (\mathcal{H} + \gamma I) \begin{bmatrix} I_n \\ X \end{bmatrix} R_\gamma, \tag{2.5}$$

where

$$R_\gamma = (R + \gamma I_n)^{-1}(R - \gamma I_n).$$

Since $\sigma(R) \subset \mathbb{C}_+$, we have that $\rho(R_\gamma) \leq 1$ for any $\gamma > 0$.

Since

$$A_\gamma = A + \gamma I_m \quad \text{and} \quad D_\gamma = D + \gamma I_n$$

are nonsing. M-matrices for any $\gamma > 0$. Let

$$W_\gamma = A_\gamma - BD_\gamma^{-1}C, \qquad V_\gamma = D_\gamma - CA_\gamma^{-1}B,$$

be the Schur complements of $\mathcal{K} + \gamma I$. It is well-known that W_γ and V_γ are also nonsing. M-matrices.

Applying the Sherman-Morrison-Woodbury formula , we have

$$\begin{aligned} V_\gamma^{-1} =& (D_\gamma - CA_\gamma^{-1}B)^{-1} \\ =& D_\gamma^{-1}(I_n - CA_\gamma^{-1}BD_\gamma^{-1})^{-1} \\ =& D_\gamma^{-1}\left[I_n + C(I_m - A_\gamma^{-1}BD_\gamma^{-1}C)^{-1}A_\gamma^{-1}BD_\gamma^{-1}\right] \\ =& D_\gamma^{-1} + D_\gamma^{-1}CW_\gamma^{-1}BD_\gamma^{-1}. \end{aligned}$$

Now let

$$L_1 = \begin{bmatrix} D_\gamma^{-1} & 0 \\ -BD_\gamma^{-1} & I_m \end{bmatrix}, \quad L_2 = \begin{bmatrix} I_n & 0 \\ 0 & -W_\gamma^{-1} \end{bmatrix}, \quad L_3 = \begin{bmatrix} I_n & D_\gamma^{-1}C \\ 0 & I_m \end{bmatrix}.$$

Then direct multiplication gives rise to

$$\mathcal{M} = L_3 L_2 L_1 (\mathcal{H} - \gamma I) = \begin{bmatrix} E_\gamma & 0 \\ -H_\gamma & I_m \end{bmatrix},$$

$$\mathcal{L} = L_3 L_2 L_1 (\mathcal{H} + \gamma I) = \begin{bmatrix} I_n & -G_\gamma \\ 0 & F_\gamma \end{bmatrix},$$

where

$$E_\gamma = I_n - 2\gamma V_\gamma^{-1}, \qquad G_\gamma = 2\gamma D_\gamma^{-1} C W_\gamma^{-1},$$

$$F_\gamma = I_m - 2\gamma W_\gamma^{-1}, \qquad H_\gamma = 2\gamma W_\gamma^{-1} B D_\gamma^{-1}.$$

Clearly, after these transf., (2.1) is transformed into

$$\mathcal{M}\begin{bmatrix} I_n \\ X \end{bmatrix} = \mathcal{L}\begin{bmatrix} I_n \\ X \end{bmatrix} R_\gamma, \tag{2.6}$$

Similarly, if $Y \geq 0$ is the minimal nonneg. sol. of the dual eq. , then

$$\mathcal{M}\begin{bmatrix} Y \\ I_m \end{bmatrix} S_\gamma = \mathcal{L}\begin{bmatrix} Y \\ I_m \end{bmatrix}, \tag{2.7}$$

where $S_\gamma = (S + \gamma I_m)^{-1}(S - \gamma I_m)$, with $S = A - BY$ being a M-matrix and $\rho(S_\gamma) \leq 1$.

Theorem 2-5: Select γ [X.-X Guo06]

Assume that $\mathcal{K}$ in (**H**) is an irr. M-matrix. Let $E_\gamma, F_\gamma, H_\gamma, G_\gamma, R_\gamma, S_\gamma$ be as above. If γ satisfies

$$\gamma \geq \gamma_0 \equiv \max\{\max_{1\leq i\leq m} a_{ii}, \max_{1\leq i\leq n} d_{ii}\}, \tag{2.8}$$

then $-E_\gamma, -F_\gamma, -R_\gamma, -S_\gamma \geq 0$ with $-E_\gamma e, -F_\gamma e, -R_\gamma e, -S_\gamma e > 0$. Moreover, $I_m - H_\gamma G_\gamma$ and $I_n - G_\gamma H_\gamma$ are nonsing. M-matrices.

Algorithm of SDA-1 for NARE

Set

$$E_0 = E_\gamma, F_0 = F_\gamma, G_0 = G_\gamma, H_0 = H_\gamma,$$

for $k = 0, 1, \ldots,$

Set

$$\begin{aligned} E_{k+1} &= E_k(I_n - G_k H_k)^{-1} E_k, && (2.9a) \\ F_{k+1} &= F_k(I_m - H_k G_k)^{-1} F_k, && (2.9b) \\ G_{k+1} &= G_k + E_k(I_n - G_k H_k)^{-1} G_k F_k, && (2.9c) \\ H_{k+1} &= H_k + F_k(I_m - H_k G_k)^{-1} H_k E_k. && (2.9d) \end{aligned}$$

End of algorithm

To ensure that this iteration is well defined, $I_n - G_k H_k$ and $I_m - H_k G_k$ must be nonsing. for all k.

Theorem 2-6: Convergence of SDA-1 [X.-X Guo06]

Assume that $\mathcal{K}$ in (**H**) is an irr. M-matrix. Let $X, Y \geq 0$ be the minimal nonneg. sols. of NARE and its dual eq., respectively. Let

$$R_\gamma = (R + \gamma I_n)^{-1}(R - \gamma I_n), \quad S_\gamma = (S + \gamma I_m)^{-1}(S - \gamma I_m),$$

where $R = D - CX$, $S = A - BY$. If the parameter γ satisfies

$$\gamma \geq \max\left\{ \max_{1 \leq i \leq m} a_{ii}, \max_{1 \leq i \leq n} d_{ii} \right\},$$

then the sequences $\{E_k,\ F_k,\ G_k,\ H_k\}$ generated by SDA are well defined, and

(a) $E_k = (I_n - G_k X)R_\gamma^{2^k} \geq 0$, with $E_k e > 0$;

(b) $F_k = (I_m - H_k Y)S_\gamma^{2^k} \geq 0$, with $F_k e > 0$;

(c) $I_m - H_k G_k$, $I_n - G_k H_k$ are nonsing. M-matrices;

(d) $0 \leq H_k \leq H_{k+1} \leq X$,

$$0 \leq X - H_k = (I_m - H_k Y)S_\gamma^{2^k} X R_\gamma^{2^k} \leq S_\gamma^{2^k} X R_\gamma^{2^k};$$

(e) $0 \leq G_k \leq G_{k+1} \leq Y$,

$$0 \leq Y - G_k = (I_n - G_k X)R_\gamma^{2^k} Y S_\gamma^{2^k} \leq R_\gamma^{2^k} Y S_\gamma^{2^k}.$$

Convergence analysis

By Weierstrass canonical form of $\mathcal{M} - \lambda\mathcal{L}$, there are nonsingular $\mathcal{Q}$ and $\mathcal{Z}$ such that

$$\mathcal{Q}\mathcal{M}\mathcal{Z} = \begin{bmatrix} J_1 & \Gamma \\ 0_{m,n} & I_m \end{bmatrix} \equiv \mathcal{J}_{\mathcal{M}}, \tag{2.10}$$

$$\mathcal{Q}\mathcal{L}\mathcal{Z} = \begin{bmatrix} I_n & 0_{n,m} \\ 0_{m,n} & J_2 \end{bmatrix} \equiv \mathcal{J}_{\mathcal{L}}, \tag{2.11}$$

- Case (1) ($0 \in \sigma(R) \cap \sigma(S)$):

$$J_1 = J_{1,s} \oplus [-1], \quad \Gamma = 0_{n-1,m-1} \oplus [1] \equiv e_n e_m^\top, \quad J_2 = J_{2,s} \oplus [-1];$$

- Case (2) ($0 \in \sigma(R)$, S is nonsing.):

$$J_1 = J_{1,s} \oplus [-1], \quad \Gamma = 0_{n,m}, \quad J_2 = J_{2,s};$$

- Case (3) ($0 \in \sigma(S)$, R is nonsing.):

$$J_1 = J_{1,s}, \quad \Gamma = 0_{n,m}, \quad J_2 = J_{2,s} \oplus [-1],$$

where

$$\rho(J_{1,s}) < 1, \quad \rho(J_{2,s}) < 1, \quad J_1 \overset{s}{\sim} R_\gamma \text{ and } J_2 \overset{s}{\sim} S_\gamma.$$

It is easy to check that $\mathcal{J}_\mathcal{M}\mathcal{J}_\mathcal{L} = \mathcal{J}_\mathcal{L}\mathcal{J}_\mathcal{M}$ in the case (1)–(3).

Theorem 2-7: Noncritical case

Assume that (**H**) holds and satisfies the case (2) or case (3) of Theorem 2-4. Let $X, Y \geq 0$ be minimal nonneg. sols. of NARE and its dual equation. Then $\{H_k, G_k\}_{k=1}^{\infty}$ generated by SDA algorithm satisfies

$$\|X - H_k\|_1 \leq \|X\|_1 \|S_\gamma^{2^k}\|_1 \|R_\gamma^{2^k}\|_1 \to 0, \qquad \text{quadratically},$$
$$\|Y - G_k\|_1 \leq \|Y\|_1 \|S_\gamma^{2^k}\|_1 \|R_\gamma^{2^k}\|_1 \to 0, \qquad \text{quadratically}.$$

It remains to show that H_k and G_k converge linearly to X and Y, resp., for the case (1). From (2.10)-(2.11) one can derive

$$\mathcal{M}\mathcal{Z}\mathcal{J}_{\mathcal{L}} = \mathcal{Q}^{-1}\mathcal{J}_{\mathcal{L}}\mathcal{J}_{\mathcal{M}} = \mathcal{L}\mathcal{Z}\mathcal{J}_{\mathcal{M}},$$

because $\mathcal{J}_{\mathcal{M}}$ and $\mathcal{J}_{\mathcal{L}}$ commute with each other. Let $\{(\mathcal{M}_k, \mathcal{L}_k)\}_{k=1}^{\infty}$ be the sequence of symplectic-like pairs in SSF-1 with

$$\mathcal{M}_k = \begin{bmatrix} E_k & 0_{n,m} \\ -H_k & I_m \end{bmatrix}, \quad \mathcal{L}_k = \begin{bmatrix} I_n & -G_k \\ 0_{m,n} & F_k \end{bmatrix}, \tag{2.12}$$

generated by SDA algorithm with $\mathcal{M}_0 = \mathcal{M}$ and $\mathcal{L}_0 = \mathcal{L}$.

Define

$$\mathcal{M}_k^* = \begin{bmatrix} E_k(I - G_k H_k)^{-1} & 0 \\ -F_k(I - H_k G_k)^{-1} H_k & I_m \end{bmatrix},$$

$$\mathcal{L}_k^* = \begin{bmatrix} I_n & -E_k(I - G_k H_k)^{-1} G_k \\ 0 & F_k(I - H_k G_k)^{-1} \end{bmatrix}.$$

Then we have $\mathcal{M}_k^* \mathcal{L}_k = \mathcal{L}_k^* \mathcal{M}_k$, and

$$\mathcal{M}_{k+1} = \mathcal{M}_k^* \mathcal{M}_k, \qquad \mathcal{L}_{k+1} = \mathcal{L}_k^* \mathcal{L}_k.$$

It follows that

$$\begin{aligned} \mathcal{M}_1 \mathcal{Z} \mathcal{J}_{\mathcal{L}}^2 &= \mathcal{M}_0^* \mathcal{M}_0 \mathcal{Z} \mathcal{J}_{\mathcal{L}}^2 = \mathcal{M}_0^* \mathcal{L}_0 \mathcal{Z} \mathcal{J}_{\mathcal{M}} \mathcal{J}_{\mathcal{L}} \\ &= \mathcal{L}_0^* \mathcal{M}_0 \mathcal{Z} \mathcal{J}_{\mathcal{L}} \mathcal{J}_{\mathcal{M}} = \mathcal{L}_0^* \mathcal{L}_0 \mathcal{Z} \mathcal{J}_{\mathcal{M}}^2 = \mathcal{L}_1 \mathcal{Z} \mathcal{J}_{\mathcal{M}}^2. \end{aligned}$$

By inductive process we have

$$\mathcal{M}_k \mathcal{Z} \mathcal{J}_{\mathcal{L}}^{2^k} = \mathcal{L}_k \mathcal{Z} \mathcal{J}_{\mathcal{M}}^{2^k}. \tag{2.13}$$

If we interchange the role of $\mathcal{M}$ and $\mathcal{L}$ in (2.10)-(2.11) and consider the symplectic pair $(\mathcal{L}, \mathcal{M})$, there are nonsing. P and Y such that

$$\begin{aligned} \mathcal{P}\mathcal{L}\mathcal{Y} &= \begin{bmatrix} J_2 & \hat{\Gamma} \\ 0_{n,m} & I_n \end{bmatrix} \equiv \hat{\mathcal{J}}_{\mathcal{L}}, \\ \mathcal{P}\mathcal{M}\mathcal{Y} &= \begin{bmatrix} I_m & 0_{m,n} \\ 0_{n,m} & J_1 \end{bmatrix} \equiv \hat{\mathcal{J}}_{\mathcal{M}}, \end{aligned}$$

where $\widehat{\Gamma} = e_m e_n^\top$. Similar arguments as above produce

$$\mathcal{L}\mathcal{Y}\hat{\mathcal{J}}_{\mathcal{M}} = \mathcal{M}\mathcal{Y}\hat{\mathcal{J}}_{\mathcal{L}}.$$

consequently, it holds

$$\mathcal{L}_k\mathcal{Y}\hat{\mathcal{J}}_{\mathcal{M}}^{2^k} = \mathcal{M}_k\mathcal{Y}\hat{\mathcal{J}}_{\mathcal{L}}^{2^k}. \tag{2.14}$$

Theorem 2-8: Main Theorem

Assume that (**H**) holds and satisfies the case (1). Let $\{E_k, F_k, G_k, H_k\}_{k=1}^{\infty}$ be the sequence generated by SDA and $\{\mathcal{M}_k, \mathcal{L}_k\}_{k=1}^{\infty}$ be defined in (2.12). Suppose $\mathcal{Z}$ and $\mathcal{Y}$ satisfy (2.13) and (2.14), respectively. Denote

$$\mathcal{Z} = \begin{bmatrix} Z_1 & Z_3 \\ Z_2 & Z_4 \end{bmatrix}, \quad \mathcal{Y} = \begin{bmatrix} Y_1 & Y_3 \\ Y_2 & Y_4 \end{bmatrix},$$

where $Z_1, Y_3 \in \mathbb{R}^{n\times n}$ and $Z_4, Y_2 \in \mathbb{R}^{m\times m}$. Then Z_1 and Y_2 are invertible, and

$$\begin{aligned} \|X - H_k\|_1 &= O(\|J_{2,s}^{2^k}\|_1) + O(2^{-k}) \to 0, \qquad \text{as } k \to \infty, \\ \|Y - G_k\|_1 &= O(\|J_{1,s}^{2^k}\|_1) + O(2^{-k}) \to 0, \qquad \text{as } k \to \infty. \end{aligned}$$

where $X = Z_2 Z_1^{-1}$ and $Y = Y_1 Y_2^{-1}$.

proof

It is easily seen that $[I, X^\top]^\top$ and $[Z_1^\top, Z_2^\top]^\top$ as well as $[Y^\top, I^\top]^\top$ and $[Y_1^\top, Y_2^\top]^\top$ span the unique stable and the unstable subspaces of $(\mathcal{M}, \mathcal{L})$ to $\sigma(J_1)$ and $\sigma(J_2)$, resp.. Then we have that $X = Z_2 Z_1^{-1}$ and $Y = Y_1 Y_2^{-1}$. Form (2.10)-(2.11) and the case (1) follows that

$$\mathcal{J}_{\mathcal{L}}^{2^k} = \begin{bmatrix} I_n & 0 \\ 0 & J_2^{2^k} \end{bmatrix}, \quad \mathcal{J}_{\mathcal{M}}^{2^k} = \begin{bmatrix} J_1^{2^k} & \Gamma_k \\ 0 & I_m \end{bmatrix},$$

where $\Gamma_k = -2^k \Gamma = -2^k e_n e_m^\top$.

Substituting $(\mathcal{M}_k, \mathcal{L}_k)$ and $\mathcal{Z}$ into (2.13), and comparing both sides, we obtain

$$
\begin{aligned}
E_k Z_1 &= (Z_1 - G_k Z_2) J_1^{2^k}, && (2.15a)\\
E_k Z_3 J_2^{2^k} &= (Z_1 - G_k Z_2)\Gamma_k + (Z_3 - G_k Z_4), && (2.15b)\\
-H_k Z_1 + Z_2 &= F_k Z_2 J_1^{2^k}, && (2.15c)\\
(-H_k Z_3 + Z_4) J_2^{2^k} &= F_k Z_2 \Gamma_k + F_k Z_4. && (2.15d)
\end{aligned}
$$

Similarly, we have

$$F_k Y_2 = (Y_2 - H_k Y_1) J_2^{2^k}, \tag{2.16a}$$
$$F_k Y_4 J_1^{2^k} = (Y_2 - H_k Y_1)\hat{\Gamma}_k + (Y_4 - H_k Y_3), \tag{2.16b}$$
$$Y_1 - G_k Y_2 = E_k Y_1 J_2^{2^k}, \tag{2.16c}$$
$$(Y_3 - G_k Y_4) J_1^{2^k} = E_k Y_1 \hat{\Gamma}_k + E_k Y_3, \tag{2.16d}$$

where $\hat{\Gamma}_k = -2^k \hat{\Gamma} = -2^k e_n e_m^\top$, and its pseudo-inverse $\hat{\Gamma}_k^\dagger = -2^{-k} e_n e_m^\top$.

Postmultiplying (2.16b) by $\hat{\Gamma}_k^{\dagger} Y_2^{-1}$. We get

$$(Y_2 - H_k Y_1)\hat{\Gamma}_k \hat{\Gamma}_k^{\dagger} Y_2^{-1} = F_k Y_4 J_1^{2^k} \hat{\Gamma}_k^{\dagger} Y_2^{-1} - (Y_4 - H_k Y_3)\hat{\Gamma}_k^{\dagger} Y_2^{-1}. \tag{2.17}$$

Substituting (2.17) into (2.16a) we have

$$\begin{aligned} F_k(I_m - Y_4 J_1^{2^k} \hat{\Gamma}_k^{\dagger} Y_2^{-1}) &= F_k(I_m - Y_4(0_{n-1,m-1} \oplus [-2^{-k}]) Y_2^{-1}) \\ &= (Y_2 - H_k Y_1)(J_{2,s}^{2^k} \oplus [0]) Y_2^{-1} \\ &- (Y_4 - H_k Y_3)\hat{\Gamma}_k^{\dagger} Y_2^{-1}. \end{aligned} \tag{2.18}$$

Since $\|H_k\|_1 \leq \|X\|_1$, by Theorem 2-6(d), it follows from (2.18) that

$$\|F_k\|_1 \leq O(\|J_{2,s}^{2^k}\|_1) + O(2^{-k}) \to 0, \qquad k \to \infty. \tag{2.19}$$

By (2.19) and the boundness of $\|J_1^{2^k}\|_1$, the matrix in (2.15c) can be estimated by

$$\|X - H_k\|_1 \leq O(\|J_{2,s}^{2^k}\|_1) + O(2^{-k}) \to 0,$$

linearly at least with rate $\frac{1}{2}$, as $k \to \infty$, where $X = Z_2 Z_1^{-1}$.

Similarly, postmultiplying (2.15b) by $\Gamma_k^\dagger$ and substituting it into (2.15a), we get

$$
\begin{aligned}
E_k(I_n - Z_3 J_2^{2^k}\Gamma_k^\dagger Z_1^{-1}) &= E_k(I_n - Z_3(0_{m-1,n-1}\oplus[-2^{-k}])Z_1^{-1}) \\
&= (Z_1 - G_k Z_2)(J_{1,s}^{2^k}\oplus[0])Z_1^{-1} \\
&- (Z_3 - G_k Z_4)\hat{\Gamma}_k^\dagger Z_1^{-1}. \qquad (2.20)
\end{aligned}
$$

Since $\|G_k\|_1 \leq \|Y\|_1$, by Theorem 4.1(e), from (2.20) follows that

$$\|E_k\|_1 \leq O(\|J_{1,s}^{2^k}\|_1) + O(2^{-k}) \to 0, \qquad k \to \infty. \tag{2.21}$$

By (2.21) and the boundness of $\|J_1^{2^k}\|_1$, the matrix in (2.16c) can be estimated by

$$\|Y - G_k\|_1 \leq O(\|J_{1,s}^{2^k}\|_1) + O(2^{-k}) \to 0,$$

linearly with rate $\frac{1}{2}$ as $k \to \infty$, where $Y = Y_1 Y_2^{-1}$. ∎

Numerical Examples

In this subsection,

- We compare the ITs, CPU, NRes of NM with SDA-1, where the "normalized" residuals (NRes) is defined by

$$\mathrm{NRes} = \frac{\|\widetilde{X}C\widetilde{X} - \widetilde{X}D - A\widetilde{X} + B\|_\infty}{\|\widetilde{X}\|_\infty \left(\|\widetilde{X}\|_\infty \|C\|_\infty + \|D\|_\infty + \|\widetilde{A}\|_\infty\right) + \|B\|_\infty},$$

  where $\widetilde{X} \equiv X_{SDA}$.

- In test examples, the IT counts for the SDA algorithm are increased by one, accounting for the additional work on computing initial matrices E_0, F_0, G_0 and H_0. Furthermore, we take $\gamma = [\gamma_0] + 1$, where γ_0 is defined in (2.8).

All implementations were run in MATLAB (version 7.0) on a PC Pentium IV (3.4GHZ) with the machine precision 2.2×10^{-16}.

Newton's method (NM). Given an initial $X_0 = 0$. For $k = 0, 1, 2 \ldots$ until X_k converges, compute X_{k+1} from X_k by solving the Sylvester equation

$$(A - X_k C)X_{k+1} + X_{k+1}(D - CX_k) = B - X_k CX_k.$$

Note that we use the Bartels-Stewart method [Bartels72] to solve Sylvester equations, where the computational cost at each Newton step is approximately $60n^3$ flops when $m = n$ (SDA-1$=\frac{64n^3}{3}$). In the case (1), Guo has been shown that the convergence rate of Newton's iteration is linearly with rate $\frac{1}{2}$.

Example 2-9 [Juang98]

Let

$$A = \mathrm{diag}(\delta_1, \ldots, \delta_n) - eq^\top, \ B = ee^\top, \ C = qq^\top,$$
$$D = \mathrm{diag}(d_1, \ldots, d_n) - qe^\top,$$

in which

$$\delta_i = \frac{1}{c\omega_i(1+\alpha)}, \ d_i = \frac{1}{c\omega_i(1-\alpha)}, \ q = (q_1, \ldots, q_n)^\top$$

with $q_i = \frac{c_i}{2\omega_i}$, where $0 < c \leq 1$, $0 \leq \alpha < 1$ and

$$0 < \omega_n < \cdots < \omega_1 < 1, \ \sum_{i=1}^{n} c_i = 1, \quad c_i > 0.$$

- It has shown in [Guo01] that for $c = 1$ and $\alpha = 0$ the matrix $\mathcal{H}$ has a double zero eigenvalue with quadratic divisor which satisfies the case (1) of Theorem 2-4. Thus, from Theorem 2-8 the SDA algorithm converges linearly, respectively, to minimal nonnegative solutions of (2.1) and (2.2) with a rate at least $\frac{1}{2}$.
- We take $n = 50, 100, 200, 300, 400, 500$, $c = 1$ and $\alpha = 0$. The IT counts, CPU times and NRes for SDA and NM are listed in Table 1.

Table: Table 1: Numerical results for Example 3.

Methods		NM	SDA	Methods		NM	SDA
$n = 50$	IT	26	27	n=100	IT	24	26
	CPU	0.36	0.06		CPU	2.8	0.64
	NRes	3.7E-16	7.2E-16		NRes	1.1E-15	1.1-15
$n = 200$	IT	23	26	n=300	IT	25	28
	CPU	33	4		CPU	150	14
	NRes	3.8E-15	1.5E-15		NRes	1.5E-15	1.3E-15
$n = 400$	IT	26	28	n=500	IT	25	31
	CPU	420	41		CPU	880	67
	NRes	3.1E-15	1.6E-15		NRes	4.7E-15	3.2E-15

Example 2-10

Let $\mathcal{R} \in \mathbb{R}^{2n\times 2n}$ be a doubly stochastic matrix (i.e., $\mathcal{R} \geq 0$, $\mathcal{R}e = \mathcal{R}^\top e = e$) generated by the Matlab code

$$\mathcal{R} = \frac{1}{n(4n^2+1)} \mathrm{magic}(2n).$$

Let $\mathcal{K} = a(I_{2n} - \mathcal{R})$, where a is a randomly chosen positive number. Then $\mathcal{K}$ is a singular M-matrix. Let

$$\mathcal{K} = \begin{bmatrix} D & -C \\ -B & A \end{bmatrix} \quad \text{and} \quad \mathcal{H} = \begin{bmatrix} D & -C \\ B & -A \end{bmatrix} \tag{2.22}$$

with $A, B, C, D \in \mathbb{R}^{n\times n}$.

From the fact that $e^\top \mathcal{K} = 0$, and $\mathcal{K}e = 0$, the condition of Case (1) in Theorem 2-4 holds. Therefore, $\mathcal{H}$ has a double zero eigenvalue with quadratic divisor. Theorem 2-8 shows that SDA converges linearly to minimal nonnegative solutions X and Y, respectively. The numerical results are listed in Table 2.

Table: Table 2: Numerical results for Example 6.

Methods		NM	SDA	Methods		NM	SDA
$n = 50$	IT	24	31	n=100	IT	24	33
	CPU	0.74	0.36		CPU	6.3	1.9
	NRes	9.3E-15	8.7E-15		NRes	4.1E-14	2.6-14
$n = 200$	IT	25	33	n=300	IT	25	34
	CPU	58	15		CPU	280	50
	NRes	1.2E-14	6.4E-14		NRes	7.2E-14	1.3E-14
$n = 400$	IT	27	34	n=500	IT	28	34
	CPU	560	110		CPU	1800	190
	NRes	2.1E-13	1.8E-13		NRes	8.9E-13	1.6E-13

Outline

UQME

We now consider the unilateral quadratic matrix equation

$$A_0 + A_1 X + A_2 X^2 = X \tag{3.1}$$

and its dual equation

$$A_2 + A_1 Y + A_0 Y^2 = Y \tag{3.2}$$

arising from discrete-time quasi-birth-death processes (QBDs).

- Standard assumptions

  $$(A_0 + A_1 + A_2)e = e, \quad A_0, A_2 \geq 0,\ A_1 \geq 0,$$

  and A_0, A_2 are two nonzero matrices. we also need:

  $$A = A_0 + A_1 + A_2 \text{ is irreducible.}$$

UQME

We now consider the unilateral quadratic matrix equation

$$A_0 + A_1X + A_2X^2 = X \tag{3.1}$$

and its dual equation

$$A_2 + A_1Y + A_0Y^2 = Y \tag{3.2}$$

arising from discrete-time quasi-birth-death processes (QBDs).

- Standard assumptions

  $$(A_0 + A_1 + A_2)e = e, \quad A_0, A_2 \geq 0, \; A_1 \geq 0,$$

  and A_0, A_2 are two nonzero matrices. we also need:

  $$A = A_0 + A_1 + A_2 \text{ is irreducible.}$$

- Under the standard assumptions, it is well known that (3.1) has at least one solution in the set $\{X \geq 0; Xe \leq e\}$ (i.e., the set of substochastic matrices). The desired solution X is the minimal nonnegative solution.

Definition

There exist an unique Perron vector $\alpha > 0$ with $\alpha^\top A = \alpha$ and $\alpha^\top A = \alpha^\top$. Let $\mu \equiv \alpha^\top (A_0 - A_2) e$, the QBD is

1. Positive recurrent if and only if $\mu > 0$.
2. Transient if and only if $\mu < 0$.
3. Null recurrent if and only if $\mu = 0$.

Let

$$\mathcal{M} = \begin{bmatrix} 0 & I \\ A_0 & A_1 - I \end{bmatrix}, \qquad \mathcal{L} = \begin{bmatrix} I & 0 \\ 0 & -A_2 \end{bmatrix}, \qquad (3.3)$$

and the characteristic polynomial of $\mathcal{M} - \lambda \mathcal{L}$ is

$$\Gamma(\lambda) \equiv \det(\mathcal{M} - \lambda \mathcal{L}).$$

We have the following fundamental results.

Spectral analysis

Theorem 3-1: Spectral properties [Bini05]

Let the QBD is null recurrent, and the open unit circle $O \equiv \{z; |z| < 1\}$. Then, for some integer $r \geq 1$ we have

1. $\sigma(\mathcal{M}, \mathcal{L}) = \sigma_s \cup \sigma_c \cup \sigma_u$, where $\sigma_s = \{\lambda_i^s\}_{i=1}^{n-r} \subseteqq O$, $\sigma_c = \{\lambda_i^c\}_{i=1}^{2r} \subseteqq \mathrm{bd}(O)$ and $\sigma_u = \{\lambda_i^u\}_{i=1}^{n-r}$ with $|\lambda_i^u| > 1$ for $i = 1, \ldots, n-r$.
2. $\{\lambda_i^c\}_{i=1}^{2r}$ are exactly the rth roots of unity, each with multiplicity two. That is, $\{\lambda_i^c\}_{i=1}^{2r} = \{\lambda_1^c, \lambda_1^c, \cdots, \lambda_r^c, \lambda_r^c\}$. The partial multiplicity of each eigenvalues on the unit circle is exactly two.
3. The spectral set $\sigma(X) = \{\lambda_i^s\}_{i=1}^{n-r} \cup \{\lambda_i^c\}_{i=1}^{r}$, and $\sigma(Y) = \{\frac{1}{\lambda_i^u}\}_{i=1}^{n-r} \cup \{\lambda_i^c\}_{i=1}^{r}$. (Here $\frac{1}{\infty} = 0$)

Latouche Ramaswami reduction

Since X is satisfying the equation (3.1) and $I - A_1$ is nonsingular, we write

$$X = \widetilde{A}_0 + \widetilde{A}_2 X^2 \tag{3.4}$$

where $\widetilde{A}_0 = (I - A_1)^{-1} A_0$, $\widetilde{A}_2 = (I - A_1)^{-1} A_2$. Post-multiplying (3.4) by X and by X^2, we get

$$X^2 = \widetilde{A}_0 X + \widetilde{A}_1 X^3, \tag{3.5}$$

$$X^3 = \widetilde{A}_0 X^2 + \widetilde{A}_1 X^4. \tag{3.6}$$

Pre-multiply (3.4) by $\widetilde{A}_0$ and pre-multiply (3.6) by $\widetilde{A}_2$, sum the equations obtained in this way with (3.5). We get

$$X^2 = \widetilde{A}_0^2 + (\widetilde{A}_0 \widetilde{A}_2 + \widetilde{A}_2 \widetilde{A}_0) X^2 + \widetilde{A}_2^2 X^4.$$

LR process

If the matrix $I - \widetilde{A}_0\widetilde{A}_2 - \widetilde{A}_2\widetilde{A}_0$ is nonsingular, the later equation allows one to express X^2 as a function of X^4 as

$$X^2 = \widetilde{A}_0^{(1)} + \widetilde{A}_2^{(1)} X^4$$

where

$$\widetilde{A}_0^{(1)} = (I - \widetilde{A}_0\widetilde{A}_2 - \widetilde{A}_2\widetilde{A}_0)^{-1}\widetilde{A}_0^2,\ \widetilde{A}_2^{(1)} = (I - \widetilde{A}_0\widetilde{A}_2 - \widetilde{A}_2\widetilde{A}_0)^{-1}\widetilde{A}_2^2.$$

Assuming that all matrices which must be inverted are nonsingular, this process can be inductively repeated by generating successive expressions of X as functions of $X^2, X^4, X^8 \ldots, X^{2^k}, \ldots$. We obtained LR algorithm as follows:

LR algorithm

Algorithm of LR for UQNE

Set

$$\begin{aligned} \widetilde{A}_0^{(0)} &= (I - A_1)^{-1} A_0; \\ \widetilde{A}_2^{(0)} &= (I - A_1)^{-1} A_2; \\ B_0^{(0)} &= \widetilde{A}_0^{(0)}; \\ B_2^{(0)} &= \widetilde{A}_2^{(0)}. \end{aligned}$$

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

Set

$$C_k = \widetilde{A}_0^{(k)} \widetilde{A}_2^{(k)} + \widetilde{A}_2^{(k)} \widetilde{A}_0^{(k)}; \tag{3.7a}$$

$$\widetilde{A}_0^{(k+1)} = (I - C_k)^{-1} (\widetilde{A}_0^{(k)})^2; \tag{3.7b}$$

$$\widetilde{A}_2^{(k+1)} = (I - C_k)^{-1} (\widetilde{A}_2^{(k)})^2; \tag{3.7c}$$

$$B_0^{(k+1)} = B_0^{(k)} + B_2^{(k)} \widetilde{A}_0^{(k+1)}; \tag{3.7d}$$

$$B_2^{(k+1)} = B_2^{(k)} \widetilde{A}_2^{(k+1)}. \tag{3.7e}$$

Theorem 3-2: Convergence Theorem of LR [Bini05,Lat93,Lat99]

Let $\mu \neq 0$ and

$$\xi \equiv \min\{|z|;\ \Gamma(z) = 0, |z| > 1\} = \min\{|z|;\ z \in \sigma(\mathcal{M}, \mathcal{L})/\mathrm{Cl}(O)\} > 1,$$
$$\eta \equiv \max\{|z|;\ \Gamma(z) = 0, |z| < 1\} = \max\{|z|;\ z \in \sigma(\mathcal{M}, \mathcal{L}) \cap O\} < 1,$$

then we have

I. If $\mu < 0$, i.e., the QBD is transient. Then we have

$$\limsup_{k\to\infty} \sqrt[2^k]{\|\widetilde{A}_0^{(k)} - e\beta^\top\|} \leq \xi^{-1},$$
$$\limsup_{k\to\infty} \sqrt[2^k]{\|\widetilde{A}_2^{(k)}\|} \leq \xi^{-1},$$
$$\limsup_{k\to\infty} \sqrt[2^k]{\|I - C_k\|} \leq \xi^{-1},$$
$$\limsup_{k\to\infty} \sqrt[2^k]{\|X - B_0^{(k)}\|} \leq \xi^{-1},$$

where β is a nonnegative left eigenvector of X corresponding eigenvalue one which sum of components is equal 1.

II. If $\mu > 0$, i.e., the QBD is positive recurrent. Then we have

$$\limsup_{k\to\infty} \sqrt[2^k]{\|\widetilde{A}_0^{(k)}\|} \leq \eta,$$

$$\limsup_{k\to\infty} \sqrt[2^k]{\|\widetilde{A}_2^{(k)} - e\gamma^\top\|} \leq \eta,$$

$$\limsup_{k\to\infty} \sqrt[2^k]{\|I - C_k\|} \leq \eta,$$

$$\limsup_{k\to\infty} \sqrt[2^k]{\|X - B_0^{(k)}\|} \leq \eta,$$

where γ is a nonnegative left eigenvector of Y corresponding eigenvalue one which sum of components is equal 1.

Cyclic reduction

Rewrite the matrix equation (3.1) as
$-A_0 + (I - A_1)X - A_2X^2 = 0$, we have the infinite system

$$\begin{bmatrix} I - A_1 & -A_0 & & 0 \\ -A_2 & I - A_1 & -A_2 & 0 \\ & -A_0 & I - A_1 & \ddots \\ 0 & & \vdots & \vdots \end{bmatrix} \begin{bmatrix} X \\ X^2 \\ X^3 \\ \vdots \end{bmatrix} = \begin{bmatrix} A_0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \qquad (3.8)$$

Let $S_0 = I - A_0$, $U_0 = I - A_1$, $V_0 = A_2$ and $T_0 = A_0$, the general procedure is adequately motivated by considering the finite system

$$\begin{array}{rcccccccccl} & S_0X & - & V_0X^2 & & & & & & = & A_0 \\ - & T_0X & + & U_0X^2 & - & V_0X^3 & & & & = & 0 \\ & & - & T_0X^2 & + & U_0X^3 & - & V_0X^4 & & = & 0 \\ & & & & - & T_0X^3 & + & U_0X^4 & - \; V_0X^5 & = & 0 \end{array} \tag{3.9}$$

For (3.9), we pre-multiply the 2^{nd} equation by $-T_0V_0^{-1}$ and add it to the 1^{st} equation. Pre-multiply the 2^{nd}, 3^{rd} and 4^{th} equations by $T_0U_0^{-1}$, I and $V_0U_0^{-1}$, respectively, and add the resulting equations to obtain

$$\begin{array}{rcrcrcl} & S_1 X & - & V_1 X^3 & & & = & A_0 \\ - & T_1 X & + & U_1 X^3 & - & V_1 X^5 & = & 0 \end{array}$$

where

$$\begin{aligned} T_1 &= T_0 U_0^{-1} T_0; \\ U_1 &= U_0 - T_0 U_0^{-1} V_0 - V_0 U_0^{-1} T_0; \\ V_1 &= V_0 U_0^{-1} V_0; \\ S_1 &= S_0 - V_0 U_0^{-1} T_0. \end{aligned}$$

By recursively applying above step (If U_k is nonsingular), we generate the sequence of infinite block tridiagnal systems

$$\begin{bmatrix} S_k & -T_k & & 0 \\ -V_k & U_k & -T_k & \\ & -V_k & U_k & -T_k \\ 0 & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} X \\ X^{2^k+1} \\ X^{2^{k+1}+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} A_0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad k \geq 0. \tag{3.10}$$

then CR algorithm is given as

CR algorithm

Algorithm of CR for UQME

Set

$$\begin{aligned} T_0 &= A_0; \\ U_0 &= I - A_1; \\ V_0 &= A_2; \\ S_0 &= I - A_1. \end{aligned}$$

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

Set

$$\begin{aligned} T_{k+1} &= T_k U_k^{-1} T_k; && (3.11a) \\ U_{k+1} &= U_k - T_k U_k^{-1} V_k - V_k U_k^{-1} T_k; && (3.11b) \\ V_{k+1} &= V_k U_k^{-1} V_k; && (3.11c) \\ S_{k+1} &= S_k - V_k U_k^{-1} T_k. && (3.11d) \end{aligned}$$

End of algorithm

Theorem 3-3: Convergence Theorem of CR [Bini05]

Let ξ and η be defined in Theorem 3-2, then

I. If $\mu < 0$, i.e., the QBD is transient. Then we have

$$\limsup_{k\to\infty} \sqrt[2^k]{\|X - S_k^{-1}A_0\|} \leq \xi^{-1}.$$

$$\limsup_{k\to\infty} \sqrt[2^k]{\|Y - A_2 S_k^{-1}\|} \leq \xi^{-1}.$$

II. If $\mu > 0$, i.e., the QBD is positive recurrent. Then we have

$$\limsup_{k\to\infty} \sqrt[2^k]{\|X - S_k^{-1}A_0\|} \leq \eta.$$

$$\limsup_{k\to\infty} \sqrt[2^k]{\|Y - A_2 S_k^{-1}\|} \leq \eta.$$

$\mu = 0$

In case $\mu = 0$, the LR algorithm still converges to the minimal nonnegative solution X. However, the convergence of the LR algorithm is linear with rate $1/2$ in the case $\mu = 0$, And the proof need additional assumptions. Here some definitions must be given. A positive vectors x is called probability vector of X if the sum of components equal 1 , and $x^\top X = x^\top$. It is well known that probability vector is unique if it exist. We state the convergence result in this case whose proof can be found in [Guo99].

$\mu = 0$

Theorem 3-4: Critical case [Bini05,Guo99]

In the case $\mu = 0$. Let x and y are the unique probability vector of X and Y, respectively. For any limit points (L_0, L_2) of $\{\widetilde{A}_0^{(k)}, \widetilde{A}_2^{(k)}\}$, we have $L_0 = ax^\top$ and $L_2 = (e-a)y^\top$ for some nonnegative vector $a \le e$. We make two extra conditions:

1. $\sigma(\mathcal{M}, \mathcal{L}) \cap \mathrm{bd}(O) = \{1\}$. (3.12)
2. Each limit point $ax^\top$ of $\{\widetilde{A}_0^{(k)}\}$ is such that $0 < y^\top a < 1$. (3.13)

Then, under the assumptions (3.12) and (3.13), we have

$$\limsup_{k\to\infty} \sqrt[k]{\|B_0^{(k)} - X\|_\infty} = \frac{1}{2}.$$

SDA-1

We rewrite (3.1) and (3.2), respectively, into

$$\begin{bmatrix} 0 & I \\ A_0 & A_1 - I \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & -A_2 \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} R_q, \tag{3.14a}$$

$$\begin{bmatrix} 0 & I \\ A_0 & A_1 - I \end{bmatrix} \begin{bmatrix} Y \\ I \end{bmatrix} S_q = \begin{bmatrix} I & 0 \\ 0 & -A_2 \end{bmatrix} \begin{bmatrix} Y \\ I \end{bmatrix}, \tag{3.14b}$$

where $X = R_q \geq 0$ and $Y = S_q \geq 0$. We now assume that

$$\begin{aligned} (I - A_1)^{-1} A_0 e &= A_0 e + A_1 A_0 e + \cdots > 0, \\ (I - A_1)^{-1} A_2 e &= A_2 e + A_1 A_2 e + \cdots > 0. \end{aligned} \tag{3.15}$$

SDA-1 algorithm

Algorithm of SDA-1 for UQME

Set

$$\begin{aligned} E_0 &= (I - A_1)^{-1} A_0, \\ F_0 &= (I - A_1)^{-1} A_2, \\ G_0 &= (I - A_1)^{-1} A_0, \\ H_0 &= (I - A_1)^{-1} A_2. \end{aligned}$$

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

Set

$$E_{k+1} = E_k (I - G_k H_k)^{-1} E_k, \tag{3.16a}$$

$$F_{k+1} = F_k (I - H_k G_k)^{-1} F_k, \tag{3.16b}$$

$$G_{k+1} = G_k + E_k (I - G_k H_k)^{-1} G_k F_k, \tag{3.16c}$$

$$H_{k+1} = H_k + F_k (I - H_k G_k)^{-1} H_k E_k. \tag{3.16d}$$

Theorem 3-5: Main results 1: Convergence of SDA-1

Assume that (3.15) and (3.12) holds. Let $X, Y \geq 0$ be minimal nonnegative solutions of (3.1) and (2.14), respectively. Then the sequence $\{E_k, F_k, G_k, H_k\}$ generated by SDA algorithm is well-defined with $E_0 = E$, $F_0 = F$, $G_0 = F$ and $H_0 = E$. Furthermore, for the case $\mu = 0$ it holds

$$\|X - H_k\|_1 \leq O(\|Y_s\|_1^{2^k}) + O(2^{-k}) \to 0, \tag{3.17a}$$

$$\|Y - G_k\|_1 \leq O(\|X_s\|_1^{2^k}) + O(2^{-k}) \to 0, \tag{3.17b}$$

at least linearly with rate $\frac{1}{2}$ as $k \to \infty$, where $\sigma(X) = \sigma(X_s) \cup \{1\}$ with $\rho(X_s) < 1$ and $\sigma(Y) = \sigma(Y_s) \cup \{1\}$ with $\rho(Y_s) < 1$; for the case $\mu > 0$ (positive recurrent) and the case $\mu < 0$ (transient) it holds

$$\|X - H_k\|_1 \leq \|X\|_1 \|Y^{2^k}\|_1 \|X^{2^k}\|_1 \to 0, \tag{3.18a}$$

$$\|Y - G_k\|_1 \leq \|X\|_1 \|Y^{2^k}\|_1 \|X^{2^k}\|_1 \to 0, \tag{3.18b}$$

as $k \to \infty$, where $\rho(X) \leq 1$, $\rho(Y) < 1$ for the case $\mu > 0$, and $\rho(X) < 1$, $\rho(Y) \leq 1$ for the case $\mu < 0$.

SDA-2 algorithm

To use SDA-2 to find X, we may rewrite (3.1) as

$$\mathcal{L}_0 \begin{bmatrix} I \\ A_2 X \end{bmatrix} = \mathcal{M}_0 \begin{bmatrix} I \\ A_2 X \end{bmatrix} X,$$

where $\mathcal{L}_0 \equiv \begin{bmatrix} 0 & I \\ A_0 & 0 \end{bmatrix}$, $\mathcal{M}_0 \equiv \begin{bmatrix} A_2 & 0 \\ I - A_1 & -I \end{bmatrix}$. As mentioned before, the pencil $\mathcal{L}_0 - \lambda \mathcal{M}_0$ is a linearization of $-A_0 + \lambda(I - A_1) - \lambda^2 A_2$. If we use SDA-1, the matrix X 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 X if $A_2 X$ is obtained and A_2 is singular? This concern will turn out to be unnecessary. We given the SDA-2 for solving the QBD processed.

SDA-2 algorithm

Algorithm of SDA-2 for UQME

Set

$$\begin{aligned} V_0 &= A_2, \\ T_0 &= A_0, \\ Q_0 &= I - A_1, \\ P_0 &= 0. \end{aligned}$$

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

Set

$$V_{k+1} = V_k(Q_k - P_k)^{-1}V_k, \tag{3.19a}$$

$$T_{k+1} = T_k(Q_k - P_k)^{-1}T_k, \tag{3.19b}$$

$$Q_{k+1} = Q_k - T_k(Q_k - P_k)^{-1}V_k, \tag{3.19c}$$

$$P_{k+1} = P_k + V_k(Q_k - P_k)^{-1}T_k. \tag{3.19d}$$

Theorem 3-6: Main results 2: Convergence of SDA-2

Let the QBD be null-recurrent. Then for SDA-2 we have

$$\limsup_{k\to\infty} \sqrt[k]{\|V_k\|} \leq \frac{1}{2},$$
$$\limsup_{k\to\infty} \sqrt[k]{\|T_k\|} \leq \frac{1}{2},$$
$$\limsup_{k\to\infty} \sqrt[k]{\|Q_k - (I - A_1 - A_0 Y)\|} \leq \frac{1}{2},$$
$$\limsup_{k\to\infty} \sqrt[k]{\|P_k - A_2 X\|} \leq \frac{1}{2}.$$

Corollary 3-7: Using SDA-2 to finding X and Y

Let the limits

$$\lim_{k\to\infty} Q_k = Q_*, \quad \lim_{k\to\infty} P_k = P_*.$$

Then, Q_* and $I - A_1 - P_*$ are nonsingular matrices. The minimal nonnegative solutions of (3.1) and (2.14) are

$$X = (I - A_1 - P_*)^{-1} A_0, \quad Y = Q_*^{-1} A_2,$$

respectively. Moreover, the matrix $Q_* - P_*$ is a singular M-matrix.

DARE

The subsection concerns with SDA-1 for finding the symmetric almost stabilizing solution X_s of a discrete-time algebraic Riccati equation (DARE) of the form

$$\mathcal{R}(X) \equiv -X+A^\top XA+Q-(C+B^\top XA)^\top (R+B^\top XB)^{-1}(C+B^\top XA) = 0, \tag{3.20}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{m\times n}$, $Q = Q^\top \in \mathbb{R}^{n\times n}$ and $R = R^\top \in \mathbb{R}^{m\times m}$, respectively. A symmetric solution $X \in \mathbb{R}^{n\times n}$ of (3.20) is called stabilizing (resp., almost stabilizing) if $R + B^\top XB$ is invertible and all the eigenvalues of the closed-loop matrix $A_F \equiv A + BF$ are in the open (resp., closed) unit disk, where

$$F = -(R + B^\top XB)^{-1}(B^\top XA + C). \tag{3.21}$$

Definition

Let extended symplectic pencil (ESP) $\mathcal{M} - \lambda\mathcal{L}$ associated with the DARE (3.20) is defined by

$$\mathcal{M} = \begin{bmatrix} A & 0 & B \\ -Q & I & -C^\top \\ C & 0 & R \end{bmatrix}, \quad \mathcal{L} = \begin{bmatrix} I & 0 & 0 \\ 0 & A^\top & 0 \\ 0 & -B^\top & 0 \end{bmatrix}. \qquad (3.22)$$

If a symmetric matrix $X \in \mathbb{R}^{n\times n}$ satisfies the DARE (3.20) and all eigenvalues of the closed-loop matrix A_F are in the closed unit disk, then we have

Standard assumptions

$$\mathcal{M}\begin{bmatrix} I \\ X \\ F \end{bmatrix} = \mathcal{L}\begin{bmatrix} I \\ X \\ F \end{bmatrix}\Phi, \tag{3.23}$$

where the matrix F is as in (3.21) and $\Phi = A_F = A + BF$.
We make two mild assumptions:

(P) : If X is an almost stabilizing solution to DARE (3.20) and all unimodular eigenvalues of A_F are semi-simple.

(A) : All elementary divisors of unimodular eigenvalues of $\mathcal{M} - \lambda\mathcal{L}$ are of degree two.

Spectral analysis

Lemma 3-8: Spectral Properties [Guo98]

Let λ be a complex number with $|\lambda| = 1$ and X be a solution of (3.20) with $R + B^\top XB > 0$. If

$$\mathrm{rank}[\lambda I - A, B] = n, \tag{3.24}$$

then the elementary divisors of $A + BF$ corresponding to λ have degrees $k_1, k_2, \ldots, k_s (1 \leq k_1 \leq \cdots \leq k_s \leq n)$ if and only if the elementary divisors of $\mathcal{M} - \lambda\mathcal{L}$ corresponding to λ have degrees $2k_1, \ldots, 2k_s$.

Spectral analysis

Theorem 3-9: Spectral Properties [Ionescu92]

Suppose that the ESP (3.22) is regular, then we have:

1. $\deg\det(\mathcal{M} - \lambda\mathcal{L}) \leq 2n$.
2. If $\lambda \neq 0$ is a generalized eigenvalue of $\mathcal{M} - \lambda\mathcal{L}$, then $1/\lambda$ is also a generalized eigenvalue of the same multiplicity.
3. If $\lambda = 0$ is a generalized eigenvalue of $\mathcal{M} - \lambda\mathcal{L}$ with multiplicity r, then $\lambda = \infty$ is a generalized eigenvalue of multiplicity $m + r$.

A trick

We can select an appropriate matrix $Y = Y^\top \in \mathbb{R}^{n\times n}$ such that $R + B^\top YB$ is invertible. After some elementary block row operators are applied on both sides of (3.23), we obtain

$$\begin{bmatrix} (I - G_0 Y)A - B\widehat{R}^{-1}C & 0 & 0 \\ -Q + C^\top \widehat{R}^{-1}(C + B^\top YA) & I & 0 \\ C + B^\top YA & 0 & \widehat{R} \end{bmatrix} \begin{bmatrix} I \\ X \\ F \end{bmatrix}$$
$$= \begin{bmatrix} I - G_0 Y & G_0 & 0 \\ C^\top \widehat{R}^{-1} B^\top Y & A^\top - C^\top \widehat{R}^{-1} B^\top & 0 \\ B^\top Y & -B^\top & 0 \end{bmatrix} \begin{bmatrix} I \\ X \\ F \end{bmatrix} \Phi, \qquad (3.25)$$

A trick

where $\widehat{R} = R + B^\top YB$ and $G_0 = B\widehat{R}^{-1}B^\top$. Next, post-multiplying the second columns of the matrix pair in (3.25) by Y, and then adding them to the first columns, it follows that

$$\begin{bmatrix} (I - G_0Y)A - B\widehat{R}^{-1}C & 0 & 0 \\ \widetilde{Q} & I & 0 \\ C + B^\top YA & 0 & \widehat{R} \end{bmatrix} \begin{bmatrix} I \\ X - Y \\ F \end{bmatrix} = \begin{bmatrix} I & G_0 & 0 \\ A^\top Y & A^\top - C^\top \widehat{R}^{-1}B^\top & 0 \\ 0 & -B^\top & 0 \end{bmatrix} \begin{bmatrix} I \\ X - Y \\ F \end{bmatrix} \Phi \tag{3.26}$$

with $\widetilde{Q} = -Q + C^\top \widehat{R}^{-1}(C + B^\top YA) + Y$.

A trick

Then the above matrix pair in (3.26) is pre-multiplied by the following block elementary matrix

$$\mathcal{E} = \begin{bmatrix} I & 0 & 0 \\ -A^\top Y & I & 0 \\ 0 & 0 & I \end{bmatrix},$$

we thus have

$$\begin{bmatrix} A_0 & 0 & 0 \\ -H_0 & I & 0 \\ C + B^\top YA & 0 & \widehat{R} \end{bmatrix} \begin{bmatrix} I \\ X - Y \\ F \end{bmatrix} = \begin{bmatrix} I & G_0 & 0 \\ 0 & A_0^\top & 0 \\ 0 & -B^\top & 0 \end{bmatrix} \begin{bmatrix} I \\ X - Y \\ F \end{bmatrix} \Phi \tag{3.27}$$

A trick

with

$$A_0 = (I - G_0 Y)A - B\widehat{R}^{-1}C, \tag{3.28a}$$

$$G_0 = B\widehat{R}^{-1}B^\top, \tag{3.28b}$$

$$H_0 = Q - Y - C^\top \widehat{R}^{-1}(C + B^\top YA) + A^\top Y(I - G_0 Y)A - A^\top YB\widehat{R}^{-1}C. \tag{3.28c}$$

Consider the matrix pair $(\mathcal{M}_0, \mathcal{L}_0)$ in standard symplectic form (SSF), where

$$\mathcal{M}_0 = \begin{bmatrix} A_0 & 0 \\ -H_0 & I \end{bmatrix}, \quad \mathcal{L}_0 = \begin{bmatrix} I & G_0 \\ 0 & A_0^\top \end{bmatrix} \tag{3.29}$$

which satisfies $\mathcal{M}_0 \mathcal{J} \mathcal{M}_0^\top = \mathcal{L}_0 \mathcal{J} \mathcal{L}_0^\top$, where G_0 and H_0 are symmetric matrices.

A trick

By Lemma 3-8, Theorem 3-9, (**A**) and (3.27), it is obvious that the spectrum of $(\mathcal{M}_0, \mathcal{L}_0)$ is the same of $(\mathcal{M}, \mathcal{L})$ except m infinite eigenvalues. The generalized eigenvalues of $(\mathcal{M}_0, \mathcal{L}_0)$ can be arranged as

$$\underbrace{0, \ldots, 0}_{r}, \lambda_{r+1}, \ldots, \lambda_{\ell}, \underbrace{\omega_1, \omega_1, \ldots \omega_{n-\ell}, \omega_{n-\ell}}_{\text{unimodular eigenvalues}}, \overline{\lambda_{\ell}^{-1}}, \ldots, \overline{\lambda_{r+1}^{-1}}, \underbrace{\infty, \ldots, \infty}_{r},$$

where the eigenvalues λ_i are inside the unit circle except the origin, $i = r+1, \ldots, \ell$.

A trick

From (3.27)–(3.28c), we immediately obtain

$$\mathcal{M}_0 \begin{bmatrix} I \\ X - Y \end{bmatrix} = \mathcal{L}_0 \begin{bmatrix} I \\ X - Y \end{bmatrix} \Phi. \tag{3.30}$$

The DARE associated with the symplectic matrix pair $(\mathcal{M}_0, \mathcal{L}_0)$ in SSF is

$$\widehat{X} = A_0^\top \widehat{X}(I + G_0\widehat{X})^{-1} A_0 + H_0, \tag{3.31}$$

on which the efficient SDA algorithm [Chu04,Lin06] can be applied. Note that if $\widehat{X}$ is the symmetric solution to the above DARE (3.31), then $X = \widehat{X} + Y$ is the symmetric solution to the DARE (3.20).

SDA-1

Algorithm of SDA-1 for DARE

Select a symmetric matrix Y such that $\widehat{R} \equiv R + B^\top YB$ is invertible.
Set

$$\begin{aligned} A_0 &= (I - GY)A - B\widehat{R}^{-1}C, \\ G_0 &= B\widehat{R}^{-1}B^\top, \\ H_0 &= Q - Y - C^\top\widehat{R}^{-1}B^\top YA - A^\top YB\widehat{R}^{-1}C - C^\top\widehat{R}^{-1}C + A^\top Y(I - GY)A \end{aligned}$$

for $k = 0, 1, \ldots,$ compute
Set

$$\begin{aligned} A_{k+1} &= A_k(I + G_kH_k)^{-1}A_k^\top, \\ G_{k+1} &= G_k + A_kG_k(I + H_kG_k)^{-1}A_k^\top, \\ H_{k+1} &= H_k + A_k^\top(I + H_kG_k)^{-1}H_kA_k; \end{aligned}$$

End of algorithm

NM for DARE

Algorithm of Newton's Method for DARE

`Choose`

A matrix L_0 such that $A_0 \equiv A - BL_0$ is d-stable;

`Solve`

$$X_0 := dlyap(A_0^\top, Q + L_0^\top RL_0 - C^\top L_0 - L_0^\top C); \tag{3.33}$$

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

`Set`

$$\begin{aligned} L_{k+1} &= (R + B^\top X_k B)^{-1}(C + B^\top X_k A) && (3.34a) \\ A_{k+1} &= A - BL_{k+1}; && (3.34b) \\ X_{k+1} &= \mathtt{dlyap}(A_k^\top, Q + L_{k+1}^\top RL_{k+1} - C^\top L_{k+1} - L_{k+1}^\top C) && (3.34c) \end{aligned}$$

End of algorithm

Convergence of NM for DARE

Let (A, B) be d-stabilizable pair and assume that there is a symmetric solution $\tilde{X}$ of the inequality $\mathcal{R}(X) \geqslant 0$ for which $R + B^\top \tilde{X} B > 0$. For any $L_0 \in \mathbb{C}^{m\times n}$ such that $A_0 = A - BL_0$ is d-stable, starting with the symmetric matrix X_0 determined by (3.32), the recursion (3.34b) determines a sequence of symmetric matrices $\{X_k\}$ for which $A - B(R + B^\top X_k B)^{-1}(C + B^\top X_k A)$ is d-stable for $k = 0, 1, \ldots,$ and $X_0 \geq X_1 \geq \ldots$, and $\lim_{k\to\infty} X_k = X_+$.

Theorem 3-10: Main Theorem

Suppose that the $(\mathcal{M}, \mathcal{L})$ in (3.22) satisfies the assumption **(A)** and that the DARE (3.20) has an almost stabilizing solution X_s with property **(P)**. Let $Z_{2b} = Z_2(:, 1 : \mu)$ and $Z_{4a} = Z_4(:, 1 : \ell)$. If the matrix $[Z_{4a}\ Z_{2b}] \in \mathbb{R}^{n\times n}$ is invertible, then Z_1 is invertible, $\widehat{X_s} = Z_2 Z_1^{-1}$ is an almost stabilizing solution of DARE (3.31), and the sequences $\{A_k, G_k, H_k\}$ generated by Algorithm of SDA-1 for DARE is satisfying

(1) $\limsup\limits_{k\to\infty} \sqrt[k]{\|A_k\|} \leq \frac{1}{2}$.

(2) $\limsup\limits_{k\to\infty} \sqrt[k]{\|H_k - \widehat{X_s}\|} \leq \frac{1}{2}$, i.e., $H_k \to \widehat{X_s}$ linearly with rate less than or equal to $\frac{1}{2}$. Moreover, $X_s = \widehat{X_s} + Y$.

Corollary 3-11: Stabilizing and Maximal solution

Assume that (A, B) is d-stabilizable and that the same conditions as in Theorem 15 hold. If the DARE (3.20) has a maximal solution X_+, then it must coincide with the almost stabilizing solution X_s computed by SDA.

NME

- In this subsection, we are interest in the study of the nonlinear matrix equation (NME)

$$X + A^\top X^{-1} A = Q, \tag{3.35}$$

and its dual equation

$$Y + AY^{-1}A^\top = Q, \tag{3.36}$$

where $A, Q \in \mathbb{R}^{n\times n}$ with Q being symmetric positive definite.

NME

- In this subsection, we are interest in the study of the nonlinear matrix equation (NME)
$$X + A^\top X^{-1} A = Q, \tag{3.35}$$
and its dual equation
$$Y + AY^{-1}A^\top = Q, \tag{3.36}$$
where $A, Q \in \mathbb{R}^{n \times n}$ with Q being symmetric positive definite.
- NMEs occur frequently in many applications, that include control theory, ladder networks, dynamic programming, stochastic filtering and statics [Anderson90,Zhan96]. Notable examples include algebraic Riccati equations [Chu05,Chu04,Hwang05,Hwang07,Lin06], quadratic matrix equations [Guo01,Guo04,GuoLan99]
$$AX^2 + BX + C = 0,$$
where A, B, C are given coefficient matrices.

Existence of solution

Theorem 3-12: Necessary and Sufficient condition [EngRanRij93]

The NME (3.35) has a symmetric positive definite solution if and only if $\psi(\lambda) \equiv \lambda A + Q + \lambda^{-1} A^\top$ is regular ($\det \psi(\lambda) \neq 0$ for some $\lambda \in \mathbb{C}$), and $\psi(\lambda) \geq 0$, for all $|\lambda| = 1$. In that case $\psi(\lambda)$ factors as

$$\psi(\lambda) = (C_0^* + \lambda^{-1} C_1^*)(C_0 + \lambda C_1)$$

with $\det(C_0) \neq 0$, then $X = C_0^* C_0$ is a solution of (3.35). Every positive definite solution is obtained in this way.

Maximal and Minimal solution

Theorem 3-13: Existence [EngRanRij93]

If (3.35) has a symmetric positive definite solution, then it has a maximal and minimal symmetric positive definite solution X_+ and X_-, respectively. Moreover, for the maximal solution X_+, we have $\rho(X_+^{-1}A) \leq 1$; for any other symmetric positive definite solution X, we have $\rho(X^{-1}A) > 1$. Here $\rho(\cdot)$ denotes the spectral radius.

Spectral analysis

Consider the NME (3.35) and define

$$\mathcal{M} \equiv \begin{bmatrix} A & 0 \\ Q & -I \end{bmatrix}, \quad \mathcal{L} \equiv \begin{bmatrix} 0 & I \\ A^\top & 0 \end{bmatrix}. \tag{3.37}$$

It is well known that the pencil $\mathcal{M} - \lambda\mathcal{L}$ is symplectic, i.e., it satisfies

$$\mathcal{M}\mathcal{J}\mathcal{M}^\top = \mathcal{L}\mathcal{J}\mathcal{L}^\top, \quad \text{with} \quad \mathcal{J} \equiv \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

and $\lambda \in \sigma(\mathcal{M}, \mathcal{L})$ if and only if $1/\lambda \in \sigma(\mathcal{M}, \mathcal{L})$.
Let $S = X^{-1}A$. Then (3.35) can be rewritten as

$$\mathcal{M} \begin{bmatrix} I \\ X \end{bmatrix} = \mathcal{L} \begin{bmatrix} I \\ X \end{bmatrix} S. \tag{3.38}$$

Spectral analysis

It is widely known that $\lambda \neq 0$ is an eigenvalue of $X_+^{-1}A$ if and only if $(\lambda, \frac{1}{\lambda})$ are eigenvalues of $\mathcal{M} - \lambda\mathcal{L}$. A unimodular λ is an eigenvalue of $X_+^{-1}A$ with algebraic multiplicity k if and only if it is an eigenvalue of $\mathcal{M} - \lambda\mathcal{L}$ with algebraic multiplicity $2k$. The follows Theorem given that the eigenvalues of the matrix $R = X_+^{-1}A$ have the following characterization.

Spectral analysis

Theorem 3-14: Spectral Properties [Guo01]

For (3.35), the eigenvalues of the matrix $X_+^{-1}A$ are precisely the eigenvalues of the matrix pencil $\mathcal{M} - \lambda\mathcal{L}$ inside or on the unit circle, with half of the partial multiplicities for each unimodular eigenvalue.

NM for NME

Algorithm of Newton's Method for NME

`Set`

$$X_0 = Q;$$

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

`Set`

$$L_k = X_{k-1}^{-1} A; \tag{3.39a}$$

$$X_k = \texttt{dlyap}(L_k^\top, Q - 2L_k^\top A); \tag{3.39b}$$

End of algorithm

Convergence of NM

Theorem 3-14: Convergence of NM [Guo01]

If (3.35) has a positive definite solution, then Newton's Algorithm for NME determines a nondecreasing sequence of symmetric matrices $\{X_k\}$ for which $\rho(L_k) < 1$ and $\lim_{k\to\infty} X_k = X_+$. The convergence is quadratic if $\rho(X_+^{-1}A) < 1$. If $\rho(X_+^{-1}A) = 1$ and all eigenvalues of $X_+^{-1}A$ on the unit circle are semi-simple, then the convergence is either quadratic or linear with rate $1/2$.

When $\rho(X_+^{-1}A)$ has non-semisimple unimodular eigenvalues, Newton's method is still convergent but the rate of convergence is only conjectured to be $\frac{1}{\sqrt[p]{2}}$, where p is the size of the largest Jordan blocks associated with unimodular eigenvalues. The conjecture was made in [Guo98] for DARE.

CR for NME

Algorithm of CR algorithm for NME

Set

$$\begin{aligned} A_0 &= A; \\ Q_0 &= 0; \\ X_0 &= 0; \\ Y_0 &= 0, \end{aligned}$$

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

Set

$$A_{k+1} = A_k Q_k^{-1} A_k; \tag{3.40a}$$

$$Q_{k+1} = Q_k - A_k Q_k^{-1} A_k^\top - A_k^\top Q_k^{-1} A_k; \tag{3.40b}$$

$$X_{k+1} = X_k - A_k^\top Q_k^{-1} A_k; \tag{3.40c}$$

$$Y_{k+1} = Y_k - A_k Q_k^{-1} A_k^\top. \tag{3.40d}$$

Convergence of CR

Theorem 3-15: Convergence of CR [Meini02]

The sequence of matrices $\{Q_k\}$, $\{X_k\}$ and $\{Y_k\}$ are positive definite, nonincreasing. Moreover, let $\sigma = \rho(X_+^{-1}A)$ we have

$$\limsup_{k\to\infty} \sqrt[2^k]{\|X_k - X_+\|} \le \sigma^2,$$
$$\limsup_{k\to\infty} \sqrt[2^k]{\|Y_k - Y_+\|} \le \sigma^2,$$
$$\limsup_{k\to\infty} \sqrt[2^k]{\|A_k\|} \le \sigma^2.$$

The matrices Q_k and Q_k^{-1} are bounded in norm.

Theorem 3-16: Critical case [Guo01]

If $\sigma = 1$ and all eigenvalues of $X_+^{-1}A$ on the unit circle are semisimple, then the sequence $\{X_k\}$ produced by CR algorithm converges to X_+ and the convergence is at least with rate $1/2$.

In Theorem 3-16, convergence is guaranteed under a addition assumption, i.e., all elementary divisors of unimodular eigenvalues of $\mathcal{M} - \lambda\mathcal{L}$ are of degree two. In our algorithm (SDA-2), we can show more convergence results without any assumption on the unimodular eigenvalues of $X_+^{-1}A$.

SDA-2 for NME

Algorithm of SDA-2 for NME

Set

$$\begin{aligned} A_0 &= A; \\ Q_0 &= Q; \\ P_0 &= 0. \end{aligned}$$

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

Set

$$A_{k+1} = A_k(Q_k - P_k)^{-1}A_k; \quad (3.41a)$$

$$Q_{k+1} = Q_k - A_k^\top(Q_k - P_k)^{-1}A_k; \quad (3.41b)$$

$$P_{k+1} = P_k + A_k(Q_k - P_k)^{-1}A_k^\top. \quad (3.41c)$$

End of algorithm

SDA-2 for NME

Theorem 3-17: Convergence of SDA-2 [Lin06]

Assume the NME (3.35) has a symmetric positive definite solution X. Then the matrix sequence $\{A_k, Q_k, P_k\}$ generated by SDA-2 Algorithm for NME is well-defined and satisfies

(i) $A_k = (X - P_k)R^{2^k}$;

(ii) $0 \leq P_k \leq P_{k+1} < X$ and
$Q_k - P_k = (X - P_k) + A_k^\top (X - P_k)^{-1} A_k > 0$;

(iii) $X \leq Q_{k+1} \leq Q_k \leq Q$ and
$Q_k - X = (R^\top)^{2^k}(X - P_k)R^{2^k} \leq (R^\top)^{2^k} X R^{2^k}$.

Critical case

Theorem 3-18: $\rho(X_+^{-1}A) = 1$

Assume that the NME (3.35) has a symmetric positive definite solution. Let X_+ and X_- are its maximal and minimal solution, respectively. If $\rho(X_+^{-1}A) = 1$, then the sequence $\{A_k, Q_k, P_k\}$ generated by SDA-2 Algorithm for NME satisfies

(i) $\|A_k\| = O\left(\rho(J_s)^{2^k}\right) + O(2^{-k}) \to 0$ as $k \to \infty$;

(ii) $\|Q_k - X_+\| = O\left(\rho(J_s)^{2^k}\right) + O(2^{-k}) \to 0$ as $k \to \infty$ where $X_+ = Z_2 Z_1^{-1}$ (and therefore Z_1 is nonsing.) solves (3.35);

Critical case

(iii) $\|P_k - \widetilde{X}_-\| = O\left(\rho(J_s)^{2^k}\right) + O(2^{-k}) \to 0$ as $k \to \infty$ where $\widetilde{X}_- = Y_2 Y_1^{-1}$ if Y_1 is invertible; moreover, if in addition, A is invertible, then $\widetilde{X}_-$ solves (3.35) and equals X_-;
(iv) $\{Q_k - P_k\} \to$ singular matrix as $k \to \infty$.

Outline

Contributions

This thesis is concerned with the important topic of iterative methods for some nonlinear matrix equations. In particular it deals with structure-preserving doubling algorithm, we have contributed to

1. We show that the convergence of the SDA is at least linear with rate $\frac{1}{2}$ in the critical case.
2. As compared to previous papers, the results here are obtained with only basic assumptions, only using elementary matrix theory. The results we present here are more general, and the analysis is much simpler.

Contributions

- NARE
  - Spectral analysis
- DARE
  - Applying SDA-1 to the more general DARE

$$-X + A^\top XA + Q - (C + B^\top XA)^\top (R + BXB)^{-1} (C + B^\top XA) = 0,$$

with singular R.

Different than early results

In [Huang07], use SDA-1 for the analyzing of the weakly stabilizing Hermitian sol. of

- CARE
$$-XGX + A^*X + XA + H = 0,$$
- DARE
$$X = A^*X(I + GX)^{-1}A + H.$$

Different: Rewrite DARE as

$$\begin{bmatrix} A & 0 \\ -H & I \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I & G \\ 0 & A^* \end{bmatrix} \Phi,$$

where $\Phi \equiv (I + GX)^{-1}A$.

- Assumption property **(P)**, $\rho(\Phi) \leq 1$ and each unimodular eigs. has a half of the partial multiplicity of $\mathcal{M} - \lambda\mathcal{L}$ corresp. to the same eigs.
- SDA-1 has no breakdown. That is, $\{A_k, G_k, H_k\}$ is well defined, or $(I + G_k H_k)$ is nonsingular.

Future works

Some future works include

1. In the critical case, the convergence of SDA in ours experiments has been observed to be exactly linear with rate $\frac{1}{2}$. In our proof, this thesis only show that the convergence of the SDA is at least linear with rate $\frac{1}{2}$ in the critical case. We believe that the convergence rate is exactly $\frac{1}{2}$ in the critical case, but we have no proof for this.
2. NARE
   - Arising from Transport theory:

$$XCX - XD - AX + B = 0,$$

where $A, B, C, D \in \mathbb{R}^{n \times n}$ are given by

$$A = \Delta - eq^\top,\ B = ee^\top,\ C = qq^\top,\ D = E - qe^\top,$$

Future works:NARE

where

$$
\begin{aligned}
e &= (1, 1, \ldots, 1)^\top, \\
q &= (q_1, q_2, \ldots, q_n)^\top, \quad \text{with } q_i = \frac{c_i}{2\omega_i}, \\
\Delta &= \operatorname{diag}(\delta_1, \delta_2, \ldots, \delta_n), \quad \text{with } \delta_i = \frac{1}{c\omega_i(1+\alpha)}, \\
E &= \operatorname{diag}(e_1, e_2, \ldots, e_n), \quad \text{with } \delta_i = \frac{1}{c\omega_i(1-\alpha)}.
\end{aligned}
$$

Parameters:

$$0 < c \leq 1;\ 0 \leq \alpha < 1;\ 0 < \omega_n < \cdots < \omega_1 < 1;\ c_i > 0 \text{ with } \sum_{i=1}^{n} c_i = 1.$$

Future works:NARE

Recently research: Write the NARE as
$T \circ X \equiv \Delta X + XE = (Xq + e)(q^\top X + e) \equiv uv^\top.$

- Construct an explicit formula via the inversion of a Cauchy matrix: [Juang98],[Mehrmann08].
- Vector iteration: [Lu/SIMAX/05], [Bai08]

$$u = u \circ (Pv) + e, \quad v = v \circ (\widetilde{P}u) + e.$$

- Newton's iteration: [Lu/NLAA/05]

$$f(\begin{bmatrix} u \\ v \end{bmatrix}) \equiv \begin{bmatrix} u - u \circ (Pv) - e \\ v - v \circ (\widetilde{P}u) - e \end{bmatrix}.$$

- SDA (1 or 2)? How to apply?

Future works:NARE

Another condition of the existence of minimal noneg. sol.. Let the assumption (**H1**)

$$B, C \geq 0;\ A, D \text{ are Z-mat.};\ \exists X \geq 0, \text{ s.t. } \mathcal{R}(X) \leq 0.$$

We can show that there exist mini. noneg. sol. under assumption (**H1**). But (**H1**) $\not\Leftrightarrow$ (**H1**).

Future works:NME

MME

- Existence of sol. of NME

$$X + A^\top X^{-1} A = Q,$$

where $Q = Q^\top$ (not necessarily positive definite, this problem arising from $\top$-palindromic eigenvalue problem).

- NME arising from the time delay systems [Faßbender07]

$$X + K X^{-1} M = C,$$

where $\prod K \prod = M$ and $\prod C \prod = C$ (Centrosymmetric),

$$\prod = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}.$$

Future works:NME

- Existence, methods?
- SLF-2

$$\begin{bmatrix} M & 0 \\ C & -I \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} 0 & I \\ K & 0 \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} (X^{-1}M).$$

  This DA has no preserving the centrosymmetric of M, C, K; $M_n \to 0$ not implies $K_n \to 0$.

- NM Method, given X_0, Solve the Sylvester eq.

$$X_{k+1} + (KX_k^{-1})X_{k+1}(X_k^{-1}M) = C.$$

  How is taking X_0?

Future works

3. We are interested in finding the positive semi-definition solution X of the stochastic algebraic Riccati equation (SARE)

$$A^\top X + XA^\top + C^\top XC - (XB + C^\top XD)(R + D^\top XD)^{-1}$$
$$(B^\top X + D^\top XC) + Q = 0, \tag{4.1}$$

with the constricted condition

$$R + D^\top XD > 0,$$

- where A, B, C, D, R and Q are $n \times n$ matrices, the matrices Q and R are symmetric. The equation (4.1) has investigated an important research field in control theory and has found interesting applications.

- where A, B, C, D, R and Q are $n \times n$ matrices, the matrices Q and R are symmetric. The equation (4.1) has investigated an important research field in control theory and has found interesting applications.
- To our best knowledge, the solvability of the SARE (4.1), even for the standard case (i.e., $Q \geq 0$ and $R > 0$ while $D \neq 0$), remains an unexplored and open problem. Recently,
  - In [Rami00], the authors proposed a numerical algorithm to compute the maximal solution to the SARE (4.1), based on a semi-definite programming.
  - The Newton procedure may be applied to SARE (4.1) under some assumptions [Ivanov07].

- where A, B, C, D, R and Q are $n \times n$ matrices, the matrices Q and R are symmetric. The equation (4.1) has investigated an important research field in control theory and has found interesting applications.
- To our best knowledge, the solvability of the SARE (4.1), even for the standard case (i.e., $Q \geq 0$ and $R > 0$ while $D \neq 0$), remains an unexplored and open problem. Recently,
  - In [Rami00], the authors proposed a numerical algorithm to compute the maximal solution to the SARE (4.1), based on a semi-definite programming.
  - The Newton procedure may be applied to SARE (4.1) under some assumptions [Ivanov07].
- In our future work, we expect to develop a structure-preserving method, which can be analytical and computational approaches the SARE (4.1). Moreover, the critical case $R + D^\top XD \geq 0$ can also be solved (the symbol "-1" should be replaced with "$\dagger$", i.e., the Moore Penrose inverse of $R + D^\top XD$).

Thank you for your attention!