A Modified Schur Method for Robust Pole Assignment in State Feedback Control *

Zhen-Chen Guo† Yun-feng Cai† Jian Qian‡ Shu-fang Xu†

December 20, 2013

Abstract

Recently, a SCHUR method was proposed in [8] to solve the robust pole assignment problem in state feedback control. It takes the departure from normality of the closed-loop system matrix A_c as the measurement of robustness, and intends to minimize it via the real Schur form of A_c. The SCHUR method works well for real poles, but when non-real poles are involved, it does not produce the real Schur form of A_c and can be problematic. In this paper, we propose a modified Schur method, which improve the efficiency of the SCHUR method when real poles are assigned, more importantly, when non-real poles are assigned, not only does this method produce the real Schur form of A_c, but also leads to a relatively small departure from normality of A_c. Numerical examples show that our modified Schur method produce better or at least comparable results than existing methods, with less computational costs.

Key words. pole assignment, state feedback control, robustness, departure from normality, real Schur form

AMS subject classification. 15A18, 65F18, 93B55.

*This research was supported in part by NSFC under grant 61075119.

†LMAM & School of Mathematical Sciences, Peking University, Beijing, 100871, China

‡School of Sciences, Beijing University of Posts and Telecommunications, Beijing, 100876, China

Email addresses: guozhch06@gmail.com (Z.C. Guo), yfcai@math.pku.edu.cn (Y.F. Cai), jqian104@gmail.com (J. Qian), xsf@pku.edu.cn (S.F. Xu)

1 Introduction

Let the matrix pair (A, B) denote the time invariant linear system with dynamic state equation

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{1.1}$$

where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ are the open-loop system matrix and the input matrix, respectively. The dynamic behavior of (1.1) is governed by the poles(eigenvalues) of A. In order to change the dynamic behavior of the open-loop system (1.1) in some desirable way (to achieve stability or to speed up response), one need to modify the poles of (1.1). This may be achieved by state-feedback control

$$u(t) = Fx(t), \tag{1.2}$$

where the feedback matrix $F \in \mathbb{R}^{m\times n}$ is to be chosen such that the the closed-loop system

$$\dot{x}(t) = (A + BF)x(t) \equiv A_c x(t) \tag{1.3}$$

has desirable poles.

Mathematically, the *state-feedback pole assignment problem* can be stated as:

State-Feedback Pole Assignment Problem (SFPA) Given $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and a set of n complex numbers $\mathfrak{L} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, closed under complex conjugation, find an $F \in \mathbb{R}^{m\times n}$ such that $\lambda(A + BF) = \mathfrak{L}$, where $\lambda(A + BF)$ is the eigenvalue set of $A + BF$.

A necessary and sufficient condition for the solvability of the **SFPA** for any set $\mathfrak{L}$ of n self-conjugate complex numbers is that (A, B) is controllable, or equivalently, the controllability matrix $\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}$ is of full row rank [25–27]. Many algorithms have been proposed to solve the **SFPA**, such as the invariant subspace method [18], the QR-like method [15,16], etc.. We refer readers to [3,4,7,10,12,17,20,24] for some other approaches.

When $m > 1$, the solution to the **SFPA** is generally not unique. We may then utilize the freedoms of F to achieve some other desirable properties of the closed-loop system. In applications, one desirable character for system design is that the eigenvalues of the closed-loop system matrix A_c are insensitive to perturbations, which leads to the following *state-feedback robust pole assignment problem*:

State-Feedback Robust Pole Assignment Problem (SFRPA) Find a solution $F \in \mathbb{R}^{m\times n}$ to the **SFPA**, such that the closed-loop system is robust, that is, the eigenvalues of A_c are as insensitive to perturbations on A_c as possible.

The key to solve the **SFRPA** is to choose an appropriate measure of robustness formulated in quantitative form. Some measures can be found in [5,8,9,13,25], such as the `condition number`

`measurement` $\kappa(X) = \|X\|_F\|X^{-1}\|_F$, where X is the eigenvector matrix of A_c, the `departure from normality` $\Delta_F(A_c) = \sqrt{\|A_c\|_F^2 - \sum_{j=1}^n |\lambda_j|^2}$ and so on. Ramar and Gourishankar [19] make an early contribution to the **SFRPA** and since then many optimization methods have been proposed based on different measurements [5,6,8,9,13,14,23]. The most classic methods should be those proposed by Kautsky, Nichols and Van Dooren in [13], where $\kappa(X)$ is used as the measure of robustness of the closed-loop system. However, method 0 in [13] may fail to converge, method 1 may suffer from slow convergence, and method 2/3 may not perform well on ill-conditioned problems. Based on method 0 in [13], Tits and Yang [23] propose a method for solving the **SFRPA** by trying to maximize the absolute value of determinant of the eigenvector matrix X. The optimization processes are iterative, and hence generally expensive. Recently, Chu [8] puts forward a Schur method for the **SFRPA** by tending to minimize the departure from normality of the closed-loop matrix A_c via the Schur decomposition of A_c. It computes the matrices X and T column by column, where $A_c = XTX^{-1}$, X, T are real and T is quasi-upper triangular, such that the strictly block upper triangular elements of matrix T are minimized in each step. If $\lambda_1, \ldots, \lambda_n$ are all real, the SCHUR method in [8] will generate an orthogonal matrix X, that is, $A_c = XTX^{-1}$ is the Schur decomposition of A_c. This implies that the departures from normality of A_c and T are the same. Hence the strategy aiming to minimize the departure from normality of T is also pliable to A_c. However, if there are non-real poles to be assigned, it cannot generate an orthogonal X, then the departure from normality of A_c is generally not identical to that of T. Hence although it tends to minimize the departure from normality of T, that of A_c may still be large.

In this paper, based on [8] we will propose a modified Schur method, where poles are assigned via real Schur decomposition of $A_c = XTX^\top$, with X being real orthogonal and T being real quasi-upper triangular. In each step(assigning a real pole or a pair of conjugate poles), one optimization problem arises, so as to minimize the departure from normality of T. When assigning a real pole, we improve the efficiency of the SCHUR method in [8] by computing the SVD of a matrix, instead of computing the GSVD of a matrix pencil. When assigning a pair of conjugate poles, by exploring the properties of optimization problem, we provide an efficient way to obtain its sub-optimal solution. Numerical examples show that our method outperforms the **SCHUR** method when non-real poles are involved. We also compare our method with the MATLAB function **place** (an implementation of Method 1 in [13]) and **robpole** (an implementation of the method in [23]). Numerical results show that our method can produce results in similar accuracy and robustness, while with much lower computational costs.

The paper is organized as follows. In Section 2, we give some preliminaries which will be used in subsequent sections. Our method is developed in Section 3, including both the real case and the conjugate complex case. Numerical results are presented in Section 4. Some concluding remarks are finally drawn in Section 5.

2 Preliminaries

In this section, we will review the parametric solutions to the **SFPA**, and the departure from normality.

2.1 Solutions to the SFPA

The parametric solutions to the **SFPA** can be expressed in several ways. In this paper, as in [8], we will formulate it by using the real Schur decomposition of $A_c = A + BF$. Assume that the real Schur decomposition of $A + BF$ is

$$A + BF = XTX^\top, \tag{2.1}$$

where $X \in \mathbb{R}^{n\times n}$ is orthogonal, $T \in \mathbb{R}^{n\times n}$ is quasi-upper triangular with only 1-by-1 and 2-by-2 diagonal blocks.

Without loss of generality, we may assume that B is of full column rank. Let

$$B = Q\begin{bmatrix} R \\ 0 \end{bmatrix} = [Q_1 \quad Q_2]\begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1 R \tag{2.2}$$

be the QR decomposition of B, where $Q \in \mathbb{R}^{n\times n}$ is orthogonal, $Q_1 \in \mathbb{R}^{n\times m}$, and $R \in \mathbb{R}^{m\times m}$ is nonsingular and upper triangular.

It follows from (2.1) that

$$AX + BFX - XT = 0. \tag{2.3}$$

Pre-multiplying (2.3) by $\mathrm{diag}(R^{-1}, I_{n-m})\,[Q_1 \quad Q_2]^\top$ on both sides gives

$$\begin{cases} R^{-1}Q_1^\top AX + FX - R^{-1}Q_1^\top XT = 0, \\ Q_2^\top(AX - XT) = 0. \end{cases} \tag{2.4}$$

Consequently, if we get an orthogonal matrix X and a quasi-upper triangular matrix T from the second equation of (2.4), then a solution F to the **SFPA** can be obtained from the first equation of (2.4) as

$$F = R^{-1}Q_1^\top(XTX^\top - A). \tag{2.5}$$

2.2 Departure from normality

In this paper, we adopt the departure from normality of $A_c = A+BF$ as the measure of robustness of the closed-loop system as in [8], which is defined as ([11, 22])

$$\Delta_F(A_c) = \sqrt{\|A_c\|_F^2 - \sum_{j=1}^{n} |\lambda_j|^2},$$

where $\lambda_1, \ldots, \lambda_n$ are the poles to be assigned, and hence eigenvalues of A_c. Now let D be the block diagonal part of T with only 1-by-1 and 2-by-2 blocks on its diagonal. Each 1-by-1 block of D admits a real eigenvalue d_j of T, while each 2-by-2 block of D admits a pair of conjugate eigenvalues $d_j = \alpha_j + i\beta_j, d_{j+1} = \bar{d}_j$ and is of the form $D_j = \begin{bmatrix} \alpha_j & \delta_j\beta_j \\ -\frac{\beta_j}{\delta_j} & \alpha_j \end{bmatrix} \in \mathbb{R}^{2\times 2}$ with $\delta_j\beta_j \neq 0$, where δ_j is a real number. Let $N = T - D = [\breve{v}_1 \quad \breve{v}_2 \quad \ldots \quad \breve{v}_n]$ be the strictly quasi-upper triangular part of T with $\breve{v}_k = \begin{bmatrix} v_k^\top & 0 \end{bmatrix}^\top, v_k \in \mathbb{R}^{k-1}$or $\mathbb{R}^{k-2}$. Direct calculations give rise to

$$\Delta_F^2(A_c) = \Delta_F^2(T) = \|N\|_F^2 + \sum_j (\delta_j - \frac{1}{\delta_j})^2\beta_j^2, \tag{2.6}$$

where the summation is over each 2-by-2 block of D.

When all poles $\lambda_1, \ldots, \lambda_n$ are real, the second part of $\Delta_F^2(A_c)$ in (2.6) will vanish. However, when some poles are non-real, not only the strictly block upper triangular part N contributes to the departure from normality, but also the block diagonal part D. When some $|\delta_j|$ is large or close to zero, the second term can be pretty large, which means that the second term is not negligible.

3 Solving SFRPA via the real Schur form

In this section, we will solve the **SFRPA** by finding an orthogonal matrix $X = [x_1 \quad x_2 \quad \ldots \quad x_n]$ and a quasi-upper triangular matrix $T = D + N$ satisfying the second equation of (2.4), such that $\Delta_F^2(A_c)$ in (2.6) is minimized. Obtaining a global optimization solution to the problem $\min \Delta_F^2(A_c)$ is rather difficult. In this paper, we will propose an efficient method to get a sub-optimal solution, which balances the contributions of N and D to the departure from normality. As in [8], we computes the matrices X and T column by column.

Assume that we have already obtained $X_j = [x_1 \quad x_2 \quad \ldots \quad x_j] \in \mathbb{R}^{n\times j}$ and $T_j \in \mathbb{R}^{j\times j}$ satisfying

$$Q_2^\top(AX_j - X_jT_j) = 0, \qquad X_j^\top X_j = I_j, \tag{3.1}$$

with T_j being quasi-upper triangular and $\lambda(T_j) = \{\lambda_k\}_{k=1}^{j}$. We then are to assign the pole λ_{j+1} (if λ_{j+1} is real) or poles $\lambda_{j+1}, \bar{\lambda}_{j+1}$ (if λ_{j+1} is non-real) to get x_{j+1}, $\breve{v}_{j+1}$ or x_{j+1}, x_{j+2}, $\breve{v}_{j+1}, \breve{v}_{j+2}$, such that the departure from normality of A_c is minimized in some sense. This procedure is repeated until all columns of X and T are obtained, and then a solution F to the **SFRPA** can be computed from (2.5). In the following subsections we will distinguish into two different cases when λ_{j+1} is real or non-real.

Before this, we should show how to get the first one(two) column(s) of X and T. If λ_1 is real, the first column of T is then $\lambda_1 e_1$, or $T_1 = \lambda_1$, and the first column x_1 of X must satisfy

$$Q_2^\top (A - \lambda_1 I_n) x_1 = 0, \tag{3.2}$$

and $\|x_1\|_2 = 1$. Let the columns of $S \in \mathbb{R}^{n\times r}$ be an orthonormal basis of $\mathcal{N}(Q_2^\top (A - \lambda_1 I_n))$, then x_1 can be chosen to be any unit vector in $\mathcal{R}(S)$. We take

$$x_1 = (S\,[1 \quad \dots \quad 1]^\top)/\|S\,[1 \quad \dots \quad 1]^\top\|_2 \tag{3.3}$$

in our algorithm as in [8], and then initially set $X_1 = x_1, T_1 = \lambda_1$.

If $\lambda_1 = \alpha_1 + i\beta_1$ is non-real, to get the real Schur form, we must assign $\bar{\lambda}_1 = \alpha_1 - i\beta_1$ together with λ_1. Thus, $T_2 = \begin{bmatrix} \alpha_1 & \delta_1\beta_1 \\ -\beta_1/\delta_1 & \alpha_1 \end{bmatrix}$ with $0 \neq \delta_1 \in \mathbb{R}$. Then the first two columns $x_1, x_2 \in \mathbb{R}^n$ of X are needed to be chosen to satisfy

$$Q_2^\top (A\,[x_1 \quad x_2] - [x_1 \quad x_2]\,T_2) = 0, \quad x_1^\top x_2 = 0, \quad \|x_1\|_2 = \|x_2\|_2 = 1, \tag{3.4}$$

so that $(\delta_1 - \frac{1}{\delta_1})^2\beta_1^2$ is minimized. Obviously, it achieves its minimum when $\delta_1 = 1$. Let the columns of $S \in \mathbb{C}^{n\times r}$ be an orthonormal basis of $\mathcal{N}(Q_2^\top (A - \lambda_1 I_n))$, and $S_1 = \mathrm{Re}(S)$, $S_2 = \mathrm{Im}(S)$. Direct calculations show that such x_1, x_2 satisfying (3.4) with $\delta_1 = 1$ can be obtained by

$$x_1 = [S_1 \quad -S_2]\,[\gamma_1 \quad \dots \quad \gamma_r \quad \zeta_1 \quad \dots \quad \zeta_r]^\top, \qquad x_2 = [S_2 \quad S_1]\,[\gamma_1 \quad \dots \quad \gamma_r \quad \zeta_1 \quad \dots \quad \zeta_r]^\top,$$

with $x_1^\top x_2 = 0$ and $\|x_1\|_2 = \|x_2\|_2 = 1$. Clearly,

$$\begin{aligned}
&x_1^\top x_2 + x_2^\top x_1 \\
&= [\gamma_1 \quad \dots \quad \gamma_r \quad \zeta_1 \quad \dots \quad \zeta_r] \begin{bmatrix} S_1^\top S_2 + S_2^\top S_1 & S_1^\top S_1 - S_2^\top S_2 \\ S_1^\top S_1 - S_2^\top S_2 & -(S_1^\top S_2 + S_2^\top S_1) \end{bmatrix} [\gamma_1 \quad \dots \quad \gamma_r \quad \zeta_1 \quad \dots \quad \zeta_r]^\top, \\
&x_1^\top x_1 - x_2^\top x_2 \\
&= [\gamma_1 \quad \dots \quad \gamma_r \quad \zeta_1 \quad \dots \quad \zeta_r] \begin{bmatrix} S_1^\top S_1 - S_2^\top S_2 & -(S_1^\top S_2 + S_2^\top S_1) \\ -(S_1^\top S_2 + S_2^\top S_1) & S_2^\top S_2 - S_1^\top S_1 \end{bmatrix} [\gamma_1 \quad \dots \quad \gamma_r \quad \zeta_1 \quad \dots \quad \zeta_r]^\top.
\end{aligned} \tag{3.5}$$

Note that the two matrices in the above two equations are symmetric Hamiltonian systems owning special properties. So we exhibit some simple results about symmetric Hamiltonian system which will be used here and when assigning the complex poles. Both results can be verified directly, and we omit the proof.

Lemma 3.1. *Let $A, B \in \mathbb{R}^{n\times n}$ satisfying $A^\top = A, B^\top = B$. If λ is an eigenvalue of $\begin{bmatrix} A & B \\ B & -A \end{bmatrix}$ and $\begin{bmatrix} x^\top & y^\top \end{bmatrix}^\top$ is the corresponding eigenvector, then*

$$\begin{bmatrix} A & B \\ B & -A \end{bmatrix} \begin{bmatrix} x & -y \\ y & x \end{bmatrix} = \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} \lambda & \\ & -\lambda \end{bmatrix},$$

and

$$\begin{bmatrix} B & -A \\ -A & -B \end{bmatrix} \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \lambda & \\ & -\lambda \end{bmatrix}.$$

Lemma 3.2. *(Property of Two Hamiltonian Systems) Let $A, B \in \mathbb{R}^{n\times n}$ be symmetric, and let $\begin{bmatrix} A & B \\ B & -A \end{bmatrix} = U \operatorname{diag}(\Theta, -\Theta) U^\top$ be the spectral decomposition, where $\Theta = \operatorname{diag}(\theta_1, \theta_2, \ldots, \theta_n)$ with $\theta_j \geq 0, j = 1 : n$. If the j-th column u_j and the $(n+j)$-th column u_{n+j} of U satisfy $u_{n+j} = \begin{bmatrix} & -I_n \\ I_n & \end{bmatrix} u_j$, then $\begin{bmatrix} B & -A \\ -A & -B \end{bmatrix} = U \begin{bmatrix} 0 & -\Theta \\ -\Theta & 0 \end{bmatrix} U^\top$.*

Applying Lemma 3.2 to the two symmetric Hamiltonian systems appeared in (3.5), that is

$$\begin{aligned} \begin{bmatrix} S_1^\top S_2 + S_2^\top S_1 & S_1^\top S_1 - S_2^\top S_2 \\ S_1^\top S_1 - S_2^\top S_2 & -(S_1^\top S_2 + S_2^\top S_1) \end{bmatrix} &= U \operatorname{diag}(\Theta, -\Theta) U^\top, \\ \begin{bmatrix} S_1^\top S_1 - S_2^\top S_2 & -(S_1^\top S_2 + S_2^\top S_1) \\ -(S_1^\top S_2 + S_2^\top S_1) & S_2^\top S_2 - S_1^\top S_1 \end{bmatrix} &= U \begin{bmatrix} 0 & -\Theta \\ -\Theta & 0 \end{bmatrix} U^\top, \end{aligned}$$

then if we let

$$\begin{bmatrix} \gamma_1 & \ldots & \gamma_r & \zeta_1 & \ldots & \zeta_r \end{bmatrix}^\top = U \begin{bmatrix} \mu_1 & \ldots & \mu_r & \nu_1 & \ldots & \nu_r \end{bmatrix}^\top,$$

$x_1^\top x_2 + x_2^\top x_1 = \sum_{j=1}^r \theta_j(\mu_j^2 - \nu_j^2)$ and $x_1^\top x_1 - x_2^\top x_2 = -2\sum_{j=1}^r \theta_j \mu_j \nu_j$ follow. Without loss of generality, we may assume that $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_r \geq 0$, then by taking

$$\mu_3 = \nu_3 = \ldots = \mu_r = \nu_r = 0, \quad \mu_1 = -\nu_1 = \sqrt{\frac{\theta_2}{\theta_1}\mu_2^2}, \quad \mu_2 = \nu_2, \tag{3.6}$$

with μ_2 satisfying

$$\|x_1\|_2 = \|x_2\|_2 = \| \begin{bmatrix} S_1 & -S_2 \end{bmatrix} U \begin{bmatrix} \mu_1 & \nu_1 & \mu_2 & \nu_2 & 0 & \ldots & 0 \end{bmatrix}^\top \|_2 = 1, \tag{3.7}$$

it is easy to verify that (3.4) holds. Hence, we can still choose initial vectors x_1 and x_2, so that $(\delta_1 - \frac{1}{\delta_1})^2 \beta_1^2 = 0$. We then initially set

$$X_2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix}, \qquad T_2 = \begin{bmatrix} \alpha_1 & \beta_1 \\ -\beta_1 & \alpha_1 \end{bmatrix}. \tag{3.8}$$

Now assume that (3.1) has been satisfied with $j \geq 1$, we will then assign the next pole λ_{j+1}. The case when λ_{j+1} is real is discussed in the following subsection, and the the case when λ_{j+1} is non-real is discussed in the next subsection.

3.1 Assigning a real pole

Let λ_{j+1} be a real pole, then the $(j+1)$-th diagonal element of T must be λ_{j+1}. Comparing the $(j+1)$-th column of $Q_2^\top AX - Q_2^\top XT = 0$ gives

$$Q_2^\top A x_{j+1} - Q_2^\top X_j v_{j+1} - \lambda_{j+1} Q_2^\top x_{j+1} = 0. \tag{3.9}$$

Recalling the definition of the departure from normality of A_c in (2.6) and noting that we are now computing the $(j+1)$-th columns of X and T, it is then natural to consider the following optimization problem:

$$\min_{\|x_{j+1}\|_2=1} \|v_{j+1}\|_2^2 \tag{3.10}$$

$$\text{s.t.} \begin{cases} Q_2^\top A x_{j+1} - \lambda_{j+1} Q_2^\top x_{j+1} - Q_2^\top X_j v_{j+1} = 0, \\ X_j^\top x_{j+1} = 0, \end{cases} \tag{3.11}$$

which can be rewritten as

$$\min_{\|x_{j+1}\|_2=1} \|v_{j+1}\|_2^2 \tag{3.12}$$

$$\text{s.t.} \begin{bmatrix} Q_2^\top (A - \lambda_{j+1} I_n) & -Q_2^\top X_j \\ X_j^\top & 0 \end{bmatrix} \begin{bmatrix} x_{j+1} \\ v_{j+1} \end{bmatrix} = 0. \tag{3.13}$$

Denote

$$M_{j+1} = \begin{bmatrix} Q_2^\top (A - \lambda_{j+1} I_n) & -Q_2^\top X_j \\ X_j^\top & 0 \end{bmatrix}, \tag{3.14}$$

and let $r = \dim \mathcal{N}(M_{j+1})$. From the controllability of (A, B), we know that $Q_2^\top (A - \lambda_{j+1} I_n)$ is of full row rank. So $n - m \leq \operatorname{rank}(M_{j+1}) \leq n - m + j$ and $\mathcal{N}(M_{j+1}) \neq \emptyset$ ([8]). Suppose that the columns of $S = \begin{bmatrix} S_1^\top & S_2^\top \end{bmatrix}^\top$ with $S_1 \in \mathbb{R}^{n \times r}, S_2 \in \mathbb{R}^{j \times r}$ form an orthonormal basis of $\mathcal{N}(M_{j+1})$, then (3.13) shows that

$$x_{j+1} = S_1 y, \quad v_{j+1} = S_2 y, \qquad \forall y \in \mathbb{R}^r. \tag{3.15}$$

Consequently, the optimization problem (3.12) subject to (3.13) equals to the following problem:

$$\min_{y^\top S_1^\top S_1 y = 1} y^\top S_2^\top S_2 y. \tag{3.16}$$

Note that the discussions above can also be found in [8], and the constrained optimization problem (3.16) is solved by using the GSVD of the matrix pencil (S_1, S_2). We will propose a simpler approach here. Actually, since $S^\top S = I_r$, we have

$$S_2^\top S_2 = I_r - S_1^\top S_1.$$

Thus the problem (3.16) is equivalent to

$$\min_{y^\top S_1^\top S_1 y = 1} y^\top y, \tag{3.17}$$

where the minimum value is acquired when y is an eigenvector of $S_1^\top S_1$ corresponding to its largest eigenvalue and satisfies $y^\top S_1^\top S_1 y = 1$. Once such y is obtained, x_{j+1} and v_{j+1} can be computed by (3.15). We may then update X_j and T_j as

$$X_{j+1} = [X_j \quad x_{j+1}] \in \mathbb{R}^{n\times(j+1)}, \qquad T_{j+1} = \begin{bmatrix} T_j & v_{j+1} \\ 0 & \lambda_{j+1} \end{bmatrix} \in \mathbb{R}^{(j+1)\times(j+1)}, \tag{3.18}$$

and continue with the next pole λ_{j+2}.

3.2 Assigning a pair of conjugate poles

In this subsection, we will consider the case that λ_{j+1} is non-real. To obtain a real matrix F from the real Schur form of $A + BF$, we would assign λ_{j+1} and $\lambda_{j+2} = \bar{\lambda}_{j+1}$ simultaneously to get the $(j+1)$-th and $(j+2)$-th columns of X and T.

3.2.1 Initial optimization problem

Assume that $\lambda_{j+1} = \alpha_{j+1} + i\beta_{j+1}$ $(\beta_{j+1} \neq 0)$ and let $D_\delta = \begin{bmatrix} \alpha_{j+1} & \delta\beta_{j+1} \\ -\beta_{j+1}/\delta & \alpha_{j+1} \end{bmatrix}$ be the diagonal block in T whose eigenvalues are λ_{j+1} and $\bar{\lambda}_{j+1}$. By comparing the $(j+1)$-th and $(j+2)$-th columns of $Q_2^\top AX - Q_2^\top XT = 0$, we have

$$Q_2^\top A\,[x_{j+1} \quad x_{j+2}] - Q_2^\top X_j\,[v_{j+1} \quad v_{j+2}] - Q_2^\top\,[x_{j+1} \quad x_{j+2}]\,D_\delta = 0. \tag{3.19}$$

Recalling the form of $\Delta_F^2(A_c)$ in (2.6), it is then natural to consider the following optimization problem:

$$\min_{\delta, v_{j+1}, v_{j+2}} \|v_{j+1}\|_2^2 + \|v_{j+2}\|_2^2 + \beta_{j+1}^2(\delta - \frac{1}{\delta})^2 \tag{3.20a}$$

$$\text{s.t.} \quad Q_2^\top(A\,[x_{j+1} \quad x_{j+2}] - X_j\,[v_{j+1} \quad v_{j+2}] - [x_{j+1} \quad x_{j+2}]\,D_\delta) = 0, \tag{3.20b}$$

$$X_j^\top\,[x_{j+1} \quad x_{j+2}] = 0, \tag{3.20c}$$

$$[x_{j+1} \quad x_{j+2}]^\top\,[x_{j+1} \quad x_{j+2}] = I_2. \tag{3.20d}$$

The constraints (3.20b) and (3.20d) are nonlinear. In [8], the author solves this optimization problem by taking $\delta = 1$ and neglecting the orthogonal requirement on x_{j+1} and x_{j+2}. These simplify the problem significantly. However, it cannot lead to the real Schur form of the closed-loop system matrix A_c, since x_{j+1} is generally not orthogonal to x_{j+2}. Moreover, the minimum value of the simplified optimization problem in [8] may be much greater than that of the original problem (3.20).

We may rewrite the optimization problem (3.20) into another equivalent form. If we write $\delta = \frac{\delta_2}{\delta_1}$ with $\delta_1, \delta_2 > 0$, and set $D_0 = \begin{bmatrix} \alpha_{j+1} & \beta_{j+1} \\ -\beta_{j+1} & \alpha_{j+1} \end{bmatrix}$, then $D_\delta = \begin{bmatrix} 1/\delta_1 & \\ & 1/\delta_2 \end{bmatrix} D_0 \begin{bmatrix} \delta_1 & \\ & \delta_2 \end{bmatrix}$. Redefine $x_{j+1} \triangleq \frac{x_{j+1}}{\delta_1}, x_{j+2} \triangleq \frac{x_{j+2}}{\delta_2}, v_{j+1} \triangleq \frac{v_{j+1}}{\delta_1}, v_{j+2} \triangleq \frac{v_{j+2}}{\delta_2}$, then the optimization problem (3.20) is equivalent to

$$\min_{\delta_1, \delta_2, v_{j+1}, v_{j+2}} \|\delta_1 v_{j+1}\|_2^2 + \|\delta_2 v_{j+2}\|_2^2 + \beta_{j+1}^2 (\frac{\delta_1}{\delta_2} - \frac{\delta_2}{\delta_1})^2 \tag{3.21a}$$

$$\text{s.t.} \quad Q_2^\top (A \begin{bmatrix} x_{j+1} & x_{j+2} \end{bmatrix} - X_j \begin{bmatrix} v_{j+1} & v_{j+2} \end{bmatrix} - \begin{bmatrix} x_{j+1} & x_{j+2} \end{bmatrix} D_0) = 0, \tag{3.21b}$$

$$X_j^\top \begin{bmatrix} x_{j+1} & x_{j+2} \end{bmatrix} = 0, \tag{3.21c}$$

$$\begin{bmatrix} x_{j+1} & x_{j+2} \end{bmatrix}^\top \begin{bmatrix} x_{j+1} & x_{j+2} \end{bmatrix} = \begin{bmatrix} 1/\delta_1^2 & \\ & 1/\delta_2^2 \end{bmatrix}. \tag{3.21d}$$

Here the constraint (3.21b) becomes linear. Once a solution to the optimization problem (3.21) is obtained, we need to redefine

$$v_{j+1} \triangleq \frac{v_{j+1}}{\|x_{j+1}\|_2}, \quad v_{j+2} \triangleq \frac{v_{j+2}}{\|x_{j+2}\|_2}, \quad x_{j+1} \triangleq \frac{x_{j+1}}{\|x_{j+1}\|_2}, \quad x_{j+2} \triangleq \frac{x_{j+2}}{\|x_{j+2}\|_2}$$

as the corresponding columns of T and X.

The constraints (3.21b) and (3.21c) are linear. All vectors $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$ satisfying these two constraints can be found via the null space of the matrix

$$M_{j+1} = \begin{bmatrix} Q_2^\top (A - (\alpha_{j+1} + i\beta_{j+1}) I_n) & -Q_2^\top X_j \\ X_j^\top & 0 \end{bmatrix}. \tag{3.22}$$

Specifically, for any $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$ satisfying (3.21b) and (3.21c), direct calculations show that $M_{j+1} \begin{bmatrix} x_{j+1} + i x_{j+2} \\ v_{j+1} + i v_{j+2} \end{bmatrix} = 0$. Conversely, for any vector $\begin{bmatrix} z^\top & w^\top \end{bmatrix}^\top \in \mathcal{N}(M_{j+1})$, the vectors $x_{j+1} = \mathrm{Re}(z), x_{j+2} = \mathrm{Im}(z), v_{j+1} = \mathrm{Re}(w), v_{j+2} = \mathrm{Im}(w)$ satisfy (3.21b) and (3.21c). The constraint (3.21d) shows that $x_{j+1}^\top x_{j+2} = 0$. For any vector $\begin{bmatrix} z^\top & w^\top \end{bmatrix}^\top \in \mathcal{N}(M_{j+1})$ with $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ being linearly independent, we may then orthogonalize $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ by the Jacobi transformation as follows to get x_{j+1} and x_{j+2} satisfying $x_{j+1}^\top x_{j+2} = 0$. Let $\varrho_1 = \|\mathrm{Re}(z)\|_2^2$, $\varrho_2 = \|\mathrm{Im}(z)\|_2^2$, $\gamma = \mathrm{Re}(z)^\top \mathrm{Im}(z)$ and $\tau = \frac{\varrho_2 - \varrho_1}{2\gamma}$, and define t as

$$t = \begin{cases} 1/(\tau + \sqrt{1+\tau^2}), & \text{if} \quad \tau \geq 0, \\ -1/(-\tau + \sqrt{1+\tau^2}), & \text{if} \quad \tau < 0. \end{cases}$$

Let $c = 1/\sqrt{1+t^2}$, $s = tc$. Then x_{j+1} and x_{j+2} obtained by

$$[x_{j+1} \quad x_{j+2}] = [\mathrm{Re}(z) \quad \mathrm{Im}(z)] \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \tag{3.23}$$

satisfy $x_{j+1}^\top x_{j+2} = 0$. Moreover, if we let

$$[v_{j+1} \quad v_{j+2}] = [\mathrm{Re}(w) \quad \mathrm{Im}(w)] \begin{bmatrix} c & s \\ -s & c \end{bmatrix}, \tag{3.24}$$

then $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$ satisfy (3.21b) and (3.21c). Hence, we can get $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$ satisfying the constrains (3.21b)-(3.21d) in this way. Furthermore,

$$1/\delta_1^2 = \|x_{j+1}\|_2^2 = \|x\|_2^2 - \omega, \quad 1/\delta_2^2 = \|x_{j+2}\|_2^2 = \|y\|_2^2 + \omega, \tag{3.25}$$

where $x = \mathrm{Re}(z)$, $y = \mathrm{Im}(z)$, $\omega = \frac{2(x^\top y)^2}{\|y\|_2^2 - \|x\|_2^2 + \sqrt{4(x^\top y)^2 + (\|y\|_2^2 - \|x\|_2^2)^2}}$ if $\|x\|_2 < \|y\|_2$; and $w = \frac{2(x^\top y)^2}{\|y\|_2^2 - \|x\|_2^2 - \sqrt{4(x^\top y)^2 + (\|y\|_2^2 - \|x\|_2^2)^2}}$ if $\|x\|_2 \geq \|y\|_2$.

3.2.2 The suboptimal strategy

It is hard to get an optimal solution to (3.21) since it is a nonlinear optimization problem with quadratic constraints. Even such an optimal solution can be found, the cost will be expensive. So instead of finding an optimal solution, we prefer to get a suboptimal solution with less computational cost.

Let the columns of $S = \begin{bmatrix} S_1^\top & S_2^\top \end{bmatrix}^\top \in \mathbb{C}^{(n+j)\times r}$ with $S_1 \in \mathbb{C}^{n\times r}$ and $S_2 \in \mathbb{C}^{j\times r}$ form an orthonormal basis of $\mathcal{N}(M_{j+1})$, and let $S_1 = U\Sigma V^*$ be the SVD of S_1. Since $S_1^* S_1 + S_2^* S_2 = I_r$, it follows that $S_2^* S_2 = V(I_r - \Sigma^*\Sigma)V^*$. For any vector $\begin{bmatrix} z^\top & w^\top \end{bmatrix}^\top \in \mathcal{N}(M_{j+1})$ with $z \in \mathbb{C}^n$ and $w \in \mathbb{C}^j$, there exists $b \in \mathbb{C}^r$ such that $z = S_1 b = U(\Sigma V^* b)$ and $w = S_2 b$. Hence

$$\|z\|_2 \leq \sigma_1 \|b\|_2 \qquad \text{and} \qquad \|w\|_2^2 \geq (1 - \sigma_1^2)\|b\|_2^2,$$

where σ_1 is the largest singular value of S_1. Now suppose that the real part and the imagine part of z are linearly independent satisfying $\|\mathrm{Re}(z)\|_2 \leq \|\mathrm{Im}(z)\|_2$, and x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2} are obtained from the the Jacobi orthogonal process (3.23), (3.24). Define $C = \frac{\|z\|_2}{\|x_{j+1}\|_2}$, then $C \geq \sqrt{2}$ and the objective function in (3.21a) becomes

$$\begin{aligned} &\|\delta_1 v_{j+1}\|_2^2 + \|\delta_2 v_{j+2}\|_2^2 + \beta_{j+1}^2 (\frac{\delta_1}{\delta_2} - \frac{\delta_2}{\delta_1})^2 \\ =&\frac{C^2}{C^2-1}\frac{\|w\|_2^2}{\|z\|_2^2} + \frac{C^4 - 2C^2}{C^2-1}\frac{\|v_{j+1}\|_2^2}{\|z\|_2^2} + \beta_{j+1}^2 (C^2 - 3 + \frac{1}{C^2-1}). \end{aligned} \tag{3.26}$$

Obviously,

$$\frac{C^2}{C^2-1}\frac{\|w\|_2^2}{\|z\|_2^2} \leq \frac{C^2}{C^2-1}\frac{\|w\|_2^2}{\|z\|_2^2} + \frac{C^4 - 2C^2}{C^2-1}\frac{\|v_{j+1}\|_2^2}{\|z\|_2^2} \leq C^2 \frac{\|w\|_2^2}{\|z\|_2^2}. \tag{3.27}$$

So the objective function in (3.21a) depends on $\frac{\|w\|_2^2}{\|z\|_2^2}$ and C with $\min \frac{\|w\|_2^2}{\|z\|_2^2} = \frac{1-\sigma_1^2}{\sigma_1^2}$. In our sub-optimal strategy, we will first take b from $\text{span}\{Ve_1\}$, where e_i is the i-th column of the identity matrix. With this choice, $\frac{\|w\|_2^2}{\|z\|_2^2}$ achieves its minimum value. And the following theorem shows the relevant results.

Theorem 3.1. *With the notations above, let u_1 be the first column of U and assume that $Re(u_1)$ and $Im(u_1)$ are linearly independent. Let x_{j+1} and x_{j+2} be the vectors obtained from $Re(u_1)$ and $Im(u_1)$ via the Jacobi orthogonal process*

$$[x_{j+1} \quad x_{j+2}] = [Re(u_1) \quad Im(u_1)] \begin{bmatrix} c & s \\ -s & c \end{bmatrix},$$

and let

$$[v_{j+1} \quad v_{j+2}] = [Re(w) \quad Im(w)] \begin{bmatrix} c & s \\ -s & c \end{bmatrix},$$

where $w = S_2 V e_1/\sigma_1$. Then $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$ satisfy the constrains (3.21b)-(3.21d)*, and the value of the corresponding objective function in* (3.21a) *will be no larger than*

$$\frac{1}{\min\{\|x_{j+1}\|_2^2, \|x_{j+2}\|_2^2\}}(\frac{1-\sigma_1^2}{\sigma_1^2} + \beta_{j+1}^2).$$

Proof. The first part of the theorem is obvious. To prove the second part, note that here $b = \frac{Ve_1}{\sigma_1}$, $\|z\|_2 = \|u_1\|_2 = 1, \|w\|_2^2 = \frac{1-\sigma_1^2}{\sigma_1^2}$. If $\|\text{Re}(u_1)\|_2 \leq \|\text{Im}(u_1)\|_2$, it then follows directly from (3.26), (3.27) and $C^2 - 3 + \frac{1}{C^2-1} \leq C^2$ with $C = \frac{1}{\|x_{j+1}\|_2}$. The case when $\|\text{Re}(u_1)\|_2 \geq \|\text{Im}(u_1)\|_2$ can be proved similarly. □

Theorem 3.1 shows that if $\text{Re}(u_1)$ and $\text{Im}(u_1)$ are linearly independent, and $\min\{\|x_{j+1}\|_2, \|x_{j+2}\|_2\}$ is not small, the above procedure will generate $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$ satisfying the constrains (3.21b)-(3.21d), and the value of the corresponding objective function in (3.21a) is not too large. We then take these $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$ as the suboptimal solution. However, if $\text{Re}(u_1)$ and $\text{Im}(u_1)$ are linearly dependent, we cannot get orthogonal x_{j+1} and x_{j+2} via the Jacobi orthogonal process. Even if $\text{Re}(u_1)$ and $\text{Im}(u_1)$ are linearly independent, the resulted $\min\{\|x_{j+1}\|_2, \|x_{j+2}\|_2\}$ might be fairly small, which means that the value of the objective function might be large. In this case, we would choose b from $\text{span}\{Ve_1, Ve_2\}$.

Define

$$\begin{aligned} \tilde{x}_1 + i\tilde{y}_1 = z_1 = u_1 = \frac{S_1 V e_1}{\sigma_1}, \qquad & w_1 = \frac{S_2 V e_1}{\sigma_1}, \\ \tilde{x}_2 + i\tilde{y}_2 = z_2 = u_2 = \frac{S_1 V e_2}{\sigma_2}, \qquad & w_2 = \frac{S_2 V e_2}{\sigma_2}, \end{aligned} \tag{3.28}$$

where σ_1, σ_2 are the first two greatest singular values of S_1. Let $b = \begin{bmatrix} \frac{Ve_1}{\sigma_1} & \frac{Ve_2}{\sigma_2} \end{bmatrix} \begin{bmatrix} \gamma_1 + i\zeta_1 \\ \gamma_2 + i\zeta_2 \end{bmatrix}$ with $\gamma_1^2 + \gamma_2^2 + \zeta_1^2 + \zeta_2^2 = 1$, then

$$x + iy = z = S_1 b = [z_1 \quad z_2] \begin{bmatrix} \gamma_1 + i\zeta_1 \\ \gamma_2 + i\zeta_2 \end{bmatrix}, \quad w = S_2 b = [w_1 \quad w_2] \begin{bmatrix} \gamma_1 + i\zeta_1 \\ \gamma_2 + i\zeta_2 \end{bmatrix}. \tag{3.29}$$

Denoting $\tilde{X} = [\tilde{x}_1 \quad \tilde{x}_2]$, $\tilde{Y} = [\tilde{y}_1 \quad \tilde{y}_2]$, it can be easily verified that

$$x = \begin{bmatrix} \tilde{X} & -\tilde{Y} \end{bmatrix} [\gamma_1 \quad \gamma_2 \quad \zeta_1 \quad \zeta_2]^\top, \qquad y = \begin{bmatrix} \tilde{Y} & \tilde{X} \end{bmatrix} [\gamma_1 \quad \gamma_2 \quad \zeta_1 \quad \zeta_2]^\top, \tag{3.30}$$

and

$$x^\top y + y^\top x = [\gamma_1 \quad \gamma_2 \quad \zeta_1 \quad \zeta_2] \begin{bmatrix} \tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X} & \tilde{X}^\top \tilde{X} - \tilde{Y}^\top \tilde{Y} \\ \tilde{X}^\top \tilde{X} - \tilde{Y}^\top \tilde{Y} & -(\tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X}) \end{bmatrix} [\gamma_1 \quad \gamma_2 \quad \zeta_1 \quad \zeta_2]^\top, \tag{3.31}$$

$$x^\top x - y^\top y = [\gamma_1 \quad \gamma_2 \quad \zeta_1 \quad \zeta_2] \begin{bmatrix} \tilde{X}^\top \tilde{X} - \tilde{Y}^\top \tilde{Y} & -(\tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X}) \\ -(\tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X}) & \tilde{Y}^\top \tilde{Y} - \tilde{X}^\top \tilde{X} \end{bmatrix} [\gamma_1 \quad \gamma_2 \quad \zeta_1 \quad \zeta_2]^\top. \tag{3.32}$$

Obviously, the two matrices in (3.31) and (3.32) are symmetric Hamiltonian systems and they satisfy the property in Lemma 3.2. Hence we can get the following lemma.

Lemma 3.3. *Let ϕ_m, ϕ_M be the two smallest singular values of $\begin{bmatrix} \tilde{Y} & \tilde{X} \end{bmatrix}$ and $\begin{bmatrix} p_1 \\ q_1 \end{bmatrix}, \begin{bmatrix} p_2 \\ q_2 \end{bmatrix}$ be the corresponding right singular vectors respectively. Define*

$$\Omega = \begin{bmatrix} p_1 & p_2 & -q_1 & -q_2 \\ q_1 & q_2 & p_1 & p_2 \end{bmatrix}, \tag{3.33}$$

$\Phi = \mathrm{diag}(\phi_1, \phi_2, -\phi_1, -\phi_2)$ with $\phi_1 = 1 - 2\phi_m^2$, $\phi_2 = 1 - 2\phi_M^2$, then

$$\begin{bmatrix} \tilde{X}^\top \tilde{X} - \tilde{Y}^\top \tilde{Y} & -(\tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X}) \\ -(\tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X}) & \tilde{Y}^\top \tilde{Y} - \tilde{X}^\top \tilde{X} \end{bmatrix} = \Omega \Phi \Omega^\top, \tag{3.34}$$

and

$$\begin{bmatrix} \tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X} & \tilde{X}^\top \tilde{X} - \tilde{Y}^\top \tilde{Y} \\ \tilde{X}^\top \tilde{X} - \tilde{Y}^\top \tilde{Y} & -(\tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X}) \end{bmatrix} = \Omega \left(\begin{array}{cc|cc} & & \phi_1 & \\ & & & \phi_2 \\ \hline \phi_1 & & & \\ & \phi_2 & & \end{array} \right) \Omega^\top.$$

Proof. Since $(\tilde{X}^\top - i\tilde{Y}^\top)(\tilde{X} + i\tilde{Y}) = [z_1 \quad z_2]^* [z_1 \quad z_2] = I_2$, so $\tilde{X}^\top \tilde{X} + \tilde{Y}^\top \tilde{Y} = I_2$ and $\tilde{X}^\top \tilde{Y} = \tilde{Y}^\top \tilde{X}$. Thus

$$\begin{bmatrix} \tilde{X}^\top \tilde{X} - \tilde{Y}^\top \tilde{Y} & -(\tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X}) \\ -(\tilde{X}^\top \tilde{Y} + \tilde{Y}^\top \tilde{X}) & \tilde{Y}^\top \tilde{Y} - \tilde{X}^\top \tilde{X} \end{bmatrix} = \begin{bmatrix} I_2 - 2\tilde{Y}^\top \tilde{Y} & -2\tilde{Y}^\top \tilde{X} \\ -2\tilde{X}^\top \tilde{Y} & I_2 - 2\tilde{X}^\top \tilde{X} \end{bmatrix} = I_4 - 2 \begin{bmatrix} \tilde{Y}^\top \\ \tilde{X}^\top \end{bmatrix} \begin{bmatrix} \tilde{Y} & \tilde{X} \end{bmatrix}.$$

Then the results follow from the above equations and Lemma 3.2. □

Now by defining

$$[\mu_1 \quad \mu_2 \quad \nu_1 \quad \nu_2]^\top = \Omega^\top [\gamma_1 \quad \gamma_2 \quad \zeta_1 \quad \zeta_2]^\top, \tag{3.35}$$

we have

$$x^\top y + y^\top x = 2\phi_1\mu_1\nu_1 + 2\phi_2\mu_2\nu_2, \qquad x^\top x - y^\top y = \phi_1(\mu_1^2 - \nu_1^2) + \phi_2(\mu_2^2 - \nu_2^2). \tag{3.36}$$

Theorem 3.2. *With the notations above, there exist $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathbb{R}$ such that $x^\top y = 0$ and $\|x\|_2 = \|y\|_2 = \frac{\sqrt{2}}{2}$. For these $\mu_1, \mu_2, \nu_1, \nu_2$, let $\gamma_1, \gamma_2, \zeta_1, \zeta_2$ be computed from* (3.35)*, where Ω is as in* (3.33)*. Then $x_{j+1} = x, x_{j+2} = y, v_{j+1} = Re(w)$ and $v_{j+2} = Im(w)$, where w is computed by* (3.29)*, satisfy the constrains* (3.21b)-(3.21d)*, and the value of the corresponding objective function in* (3.21a) *will be no larger than $\frac{2(1-\sigma_2^2)}{\sigma_2^2}$.*

Proof. It is easy to check that all solutions of the following system of equations

$$\begin{cases} \phi_1\mu_1\nu_1 + \phi_2\mu_2\nu_2 & = 0, \\ \phi_1(\mu_1^2 - \nu_1^2) + \phi_2(\mu_2^2 - \nu_2^2) & = 0, \\ \mu_1^2 + \mu_2^2 + \nu_1^2 + \nu_2^2 & = 1. \end{cases} \tag{3.37}$$

are

$$\begin{cases} \mu_2 = \pm\sqrt{\dfrac{\phi_1}{\phi_1+\phi_2} - \nu_2^2} \\ \mu_1 = -\sqrt{\dfrac{\phi_2}{\phi_1}}\nu_2 \\ \nu_1 = \pm\sqrt{\dfrac{\phi_2}{\phi_1+\phi_2} - \dfrac{\phi_2}{\phi_1}\nu_2^2} \end{cases} \quad \text{and} \quad \begin{cases} \mu_2 = \pm\sqrt{\dfrac{\phi_1}{\phi_1+\phi_2} - \nu_2^2} \\ \mu_1 = \sqrt{\dfrac{\phi_2}{\phi_1}}\nu_2 \\ \nu_1 = \mp\sqrt{\dfrac{\phi_2}{\phi_1+\phi_2} - \dfrac{\phi_2}{\phi_1}\nu_2^2} \end{cases} \tag{3.38}$$

with $\nu_2^2 \le \frac{\phi_1}{\phi_1+\phi_2}$. Note (3.36) and $\|x\|_2^2 + \|y\|_2^2 = 1$, so with the values in (3.38), it holds that $x^\top y = 0$ and $\|x\|_2 = \|y\|_2 = \frac{\sqrt{2}}{2}$. Since $\begin{bmatrix} z^\top & w^\top \end{bmatrix}^\top \in \mathcal{N}(M_{j+1})$, so $\left[\begin{smallmatrix} x_{j+1} & x_{j+2} \\ v_{j+1} & v_{j+2} \end{smallmatrix}\right] = \left[\begin{smallmatrix} x & y \\ \mathrm{Re}(w) & \mathrm{Im}(w) \end{smallmatrix}\right]$ satisfy the constrains (3.21b)-(3.21d) with $\delta_1 = \delta_2 = \frac{\sqrt{2}}{2}$. Hence

$$\begin{aligned} &\|\delta_1 v_{j+1}\|_2^2 + \|\delta_2 v_{j+2}\|_2^2 + \beta_{j+1}^2(\frac{\delta_1}{\delta_2} - \frac{\delta_2}{\delta_1})^2 \\ =&2\|w\|_2^2 = 2(\gamma_1^2 + \zeta_1^2)\frac{1-\sigma_1^2}{\sigma_1^2} + 2(\gamma_2^2 + \zeta_2^2)\frac{1-\sigma_2^2}{\sigma_2^2} \le \frac{2(1-\sigma_2^2)}{\sigma_2^2}, \end{aligned}$$

which completes the proof of the theorem. □

From the proof of Theorem 3.2 we can see that with these choice of $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$, the value of the corresponding objective function is just $2\|w\|_2^2$. Define $\xi_1 = p_1^\top \Xi p_1, \xi_2 = p_2^\top \Xi p_2, \eta_1 = q_1^\top \Xi q_1, \eta_2 = q_2^\top \Xi q_2, \zeta_{12} = q_1^\top \Xi p_2, \zeta_{21} = q_2^\top \Xi p_1$, with $\Xi = \mathrm{diag}\{(1-\sigma_1^2)/\sigma_1^2, (1-\sigma_2^2)/\sigma_2^2\}$, it then

follows

$$\|w\|_2^2 = \begin{cases} \frac{\phi_2}{\phi_1+\phi_2}(\xi_1+\eta_1) + \frac{\phi_1}{\phi_1+\phi_2}(\xi_2+\eta_2) + 2\sqrt{\frac{\phi_2}{\phi_1}}\frac{\phi_1}{\phi_1+\phi_2}(\zeta_{21}-\zeta_{12}) & \text{if } (\mu_1\nu_2) \le 0, \\ \frac{\phi_2}{\phi_1+\phi_2}(\xi_1+\eta_1) + \frac{\phi_1}{\phi_1+\phi_2}(\xi_2+\eta_2) + 2\sqrt{\frac{\phi_2}{\phi_1}}\frac{\phi_1}{\phi_1+\phi_2}(\zeta_{12}-\zeta_{21}) & \text{if } (\mu_1\nu_2) > 0. \end{cases} \tag{3.39}$$

So in order to get a smaller $\|w\|_2$, we can take $\mu_1, \mu_2, \nu_1, \nu_2$ satisfying $\mu_1\nu_2 \le 0$ if $\zeta_{21} \le \zeta_{12}$, and $\mu_1\nu_2 > 0$ if $\zeta_{21} > \zeta_{12}$.

Till now we have proposed two strategies for computing $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$. The first strategy computes $x_{j+1}, x_{j+2}, v_{j+1}$ and v_{j+2} by using the Jacobi orthogonal process (3.23) and (3.24) with $z = u_1$ and $w = \frac{S_2 V e_1}{\sigma_1}$. The second strategy first computes $\mu_1, \mu_2, \nu_1, \nu_2$ by (3.38) satisfying $\mu_1\nu_2 \le 0$ if $\zeta_{21} \le \zeta_{12}$, and $\mu_1\nu_2 > 0$ if $\zeta_{21} > \zeta_{12}$, and then compute $\gamma_1, \gamma_2, \zeta_1, \zeta_2$ from (3.35), where Ω is as in (3.33), and finally set $x_{j+1} = x, x_{j+2} = y, v_{j+1} = \mathrm{Re}(w)$ and $v_{j+2} = \mathrm{Im}(w)$, where x, y, w are computed by (3.29). We cannot tell which strategy is better. So we suggest to apply both strategies, compare the corresponding values of the objective function and adopt the one which gives better results. Specifically, if the value of the objective function corresponding to the first strategy is smaller, we would update X_j and T_j as

$$X_{j+2} = [X_j \quad \delta_1 x_{j+1} \quad \delta_2 x_{j+2}] \in \mathbb{R}^{n\times(j+2)}, \qquad T_{j+2} = \begin{bmatrix} T_j & \delta_1 v_{j+1} & \delta_2 v_{j+2} \\ 0 & \alpha_{j+1} & \delta\beta_{j+1} \\ 0 & -\frac{1}{\delta}\beta_{j+1} & \alpha_{j+1} \end{bmatrix} \in \mathbb{R}^{(j+2)\times(j+2)}, \tag{3.40}$$

where $\delta_1 = \frac{1}{\|x_{j+1}\|_2}, \delta_2 = \frac{1}{\|x_{j+2}\|_2}, \delta = \frac{\delta_2}{\delta_1}$. Otherwise, we update X_j and T_j as

$$X_{j+2} = \begin{bmatrix} X_j & \sqrt{2}x & \sqrt{2}y \end{bmatrix} \in \mathbb{R}^{n\times(j+2)}, \quad T_{j+2} = \begin{bmatrix} T_j & \sqrt{2}\mathrm{Re}(w) & \sqrt{2}\mathrm{Im}(w) \\ 0 & \alpha_{j+1} & \beta_{j+1} \\ 0 & -\beta_{j+1} & \alpha_{j+1} \end{bmatrix} \in \mathbb{R}^{(j+2)\times(j+2)}, \tag{3.41}$$

with x, y and w defined as in (3.29). This completes the assignment of the complex conjugate poles $\lambda_{j+1}, \lambda_{j+2}$, and we can then continue with the next pole λ_{j+3}.

These two strategies essentially choose z from $\mathcal{R}(u_1)$ and $\mathcal{R}([u_1 \quad u_2])$, respectively. If the results by these two strategies are not satisfactory, theoretically, we can choose z from a higher dimensional space, i.e. $z \in \mathrm{span}\{u_1, u_2, \ldots, u_k\}, k \ge 3$, with u_l being the l-th column of U. However the resulted optimization problem is much more complicated. More importantly, numerical examples show that these two strategies with $k = 1, 2$ can produce fairly satisfying results for most problems.

3.3 Algorithm

In this part, we will give the framework of our algorithm.

Algorithm 1 Framework of our **schur-rob** algorithm.

Input:

A, B and $\mathfrak{L} = \{\lambda_1, \ldots, \lambda_n\}$ (complex conjugate poles appear in pairs).

Output:

The feedback matrix F.

If λ_1 is real, compute x_1 by (3.2) and (3.3) and set $X_1 = x_1, T_1 = \lambda_1, j = 1$. If λ_1 is non-real, compute x_1, x_2 by (3.6), (3.7), and set X_2, T_2 as in (3.8), $j = 2$.
while $j < n$ **do**
  if λ_{j+1} is real **then**
    Find $S = \begin{bmatrix} S_1^\top & S_2^\top \end{bmatrix}^\top$, whose columns are an orthonormal basis of $\mathcal{N}(M_{j+1})$ in (3.14);
    Compute y from (3.17);
    Compute x_{j+1} and v_{j+1} by (3.15), update X_j and T_j as (3.18) and set $j = j + 1$.
  else
    Find $S = \begin{bmatrix} S_1^\top & S_2^\top \end{bmatrix}^\top$, whose columns are an orthonormal basis of $\mathcal{N}(M_{j+1})$ in (3.22);
    Compute the SVD of S_1 as $S_1 = U\Sigma V^*$;
    if $\mathrm{Re}(u_1)$ and $\mathrm{Im}(u_1)$ are linearly independent **then**
      Compute $x_{j+1}, x_{j+2}, v_{j+1}, v_{j+2}$ by (3.23) and (3.24) with $z = u_1, w = \frac{S_2 V e_1}{\sigma_1}$;
      Set $\delta_1 = \frac{1}{\|x_{j+1}\|_2}, \delta_2 = \frac{1}{\|x_{j+2}\|_2}$ and $\delta = \frac{\delta_2}{\delta_1}$;
      Compute $dep_1 = \|\delta_1 v_{j+1}\|_2^2 + \|\delta_2 v_{j+2}\|_2^2 + \beta_{j+1}^2(\delta - \frac{1}{\delta})^2$;
    else
      Set $dep_1 = \infty$;
    end if
    Let $\tilde{X} = [\tilde{x}_1 \quad \tilde{x}_2]$, $\tilde{Y} = [\tilde{y}_1 \quad \tilde{y}_2]$, with $\tilde{x}_1, \tilde{y}_1, \tilde{x}_2, \tilde{y}_2$ defined as in (3.28), and compute the spectral decomposition (3.34);
    Computes $\mu_1, \mu_2, \nu_1, \nu_2$ by (3.38) satisfying $\mu_1\nu_2 \le 0$ if $\zeta_{21} \le \zeta_{12}$, and $\mu_1\nu_2 > 0$ if $\zeta_{21} > \zeta_{12}$, and then compute $\gamma_1, \gamma_2, \zeta_1, \zeta_2$ from (3.35), where Ω is as in (3.33);
    Set $x_{j+1} = x, x_{j+2} = y, v_{j+1} = \mathrm{Re}(w)$ and $v_{j+2} = \mathrm{Im}(w)$, where x, y, w are computed by (3.29) and compute $dep_2 = 2[(\gamma_1^2 + \zeta_1^2)\frac{1-\sigma_1^2}{\sigma_1^2} + (\gamma_2^2 + \zeta_2^2)\frac{1-\sigma_2^2}{\sigma_2^2}]$;
    if $dep_1 < dep_2$, update X_j and T_j as in (3.40) and set $j = j + 1$; otherwise, update them as in (3.41) and set $j = j + 2$.
  end if
end while
Set $X = X_n, T = T_n$, and compute F by (2.5).

4 Numerical Examples

In this section, we will give some numerical examples to illustrate the performance of our **schur-rob** algorithm, and compare it with the **SCHUR** algorithm [8], the MATLAB functions **place** [13] and **robpole** [23]. Each algorithm computes a feedback matrix F such that the eigenvalues of $A+BF$ are those given in $\mathfrak{L}$, and $A+BF$ is robust. When applying **robpole** to all test examples, we set the maximum number of sweep to be the default value 5. All calculations are carried out on an Intel(R)Core(TM)i3, dual core, 2.27 GHz machine, with 2.00 GB RAM. MATLAB R2012a is used with machine epsilon $\epsilon \approx 2.2 \times 10^{-16}$.

Example 4.1. *Let*

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0.5 & 0.5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathfrak{L} = \{0.5, 0.5, 0.5 + 0.01i, 0.5 - 0.01i\}.$$

Applying the ***SCHUR*** *algorithm to this example gives*

$$X_1 = \begin{bmatrix} -8.165e-1 & 5.774e-1 & 0 & 0 \\ 0 & 0 & -7.071e-1 & -1.211e-19 \\ 0 & 0 & 7.071e-1 & -6.132e-20 \\ 5.774e-1 & 8.165e-1 & 0 & 0 \end{bmatrix},$$

$$T_1 = \begin{bmatrix} 5.000e-1 & 0 & 5.551e-17 & 2.708e-35 \\ & 5.000e-1 & 2.776e-17 & -6.019e-38 \\ & & 5.000e-1 & 1.000e-2 \\ & & -1.000e-2 & 5.000e-1 \end{bmatrix}.$$

Obviously, X_1 is nearly singular since the 2-norm of the forth column is $1.3574e{-}19$, almost zero. Consequently, X_1 is far from orthonormal . The departure from normality of $A + BF_1$ would be as large as $7.752e+16$, with the computed feedback being

$$F_1 = \begin{bmatrix} -1.000e+0 & 0 & 0 & 0 \\ 0 & (3.876e+16)-1 & 3.876e+16 & 0 \\ 0 & -3.876e+16 & (-3.876e+16)-1 & 0 \end{bmatrix}.$$

However, our algorithm will obtain an orthogonal X_2 and a quasi-upper triangular T_2

$$X_2 = \begin{bmatrix} -8.165e-1 & 5.774e-1 & 0 & 0 \\ 0 & 0 & -7.071e-1 & -7.071e-1 \\ 0 & 0 & 7.071e-1 & -7.071e-1 \\ 5.774e-1 & 8.165e-1 & 0 & 0 \end{bmatrix},$$

$$T_2 = \begin{bmatrix} 5.000e-1 & 0 & 0 & -4.082e-1 \\ & 5.000e-1 & 0 & -5.774e-1 \\ & & 5.000e-1 & 1.000e-2 \\ & & -1.000e-2 & 5.000e-1 \end{bmatrix}.$$

The computed feedback matrix is

$$F_2 = \begin{bmatrix} -5.000e-1 & 0 & 0 & 0 \\ 0 & -5.000e-1 & 1.000e-2 & 0 \\ 0 & 1.000e-2 & -5.000e-1 & 0 \end{bmatrix},$$

*and the departure from normality of the closed-loop system matrix $A+BF_2$ is $7.071e-1$, which is much smaller than that obtained by the **SCHUR** algorithm. So our algorithm can not only obtain the real Schur form of the closed-loop system matrix A_c, but also leads to a much smaller departure from normality of A_c. This example illustrates that with the existence of non-real poles, our algorithm will generally produce better results than the **SCHUR** algorithm.*

We then compare our **schur-rob** algorithm with the MATLAB functions **place** and **robpole**, by applying them on some benchmark sets. The precision and the robustness of all results computed by these algorithms will be displayed. Here the precision refers to the accuracy of the eigenvalues of $A_c = A+BF$, compared with the prescribed poles in $\mathfrak{L}$. Precisely, we list the relative errors $\min_{1\le j\le n}(-\log(|\frac{\lambda_j-\hat{\lambda}_j}{\lambda_j}|))$ (denoted as "$prec_m$") and $\max_{1\le j\le n}(-\log(|\frac{\lambda_j-\hat{\lambda}_j}{\lambda_j}|))$ (denoted as "$prec_M$") with $\hat{\lambda}_j (j = 1,\ldots,n)$ being the eigenvalue of the computed closed-loop system matrix A_c. The robustness is, however, more complicated, since different measures of robustness are used in these algorithms. Specifically, let the spectral decomposition and the real Schur decomposition of $A+BF$ respectively be

$$A+BF = X\Lambda X^{-1}, \qquad A+BF = UTU^{\top},$$

where Λ is a diagonal matrix whose diagonal elements are those in $\mathfrak{L}$, U is orthogonal, and T is the real Schur form. The MATLAB function **place** tends to minimize $\|X^{-1}\|_F$ and **robpole** aims to maximum $|\det(X)|$. Both measures are closely related to the condition number $\kappa_F(X) = \|X\|_F\|X^{-1}\|_F$. While our **schur-rob** algorithm tries to minimize the departure from normality of $A_c = A+BF$. Hence, in the following tables, we list both measures of robustness, i.e. the departure from normality of A_c (denoted as "dep.") and the condition number of X (denoted as "$\kappa_F(X)$"), for all three algorithms. We also list the Frobenius norm of the feedback matrix F (denoted as "$\|F\|_F$"), which is also regarded as a measure of robustness in some literature.

The benchmark sets we test include eleven illustrated examples from [5], ten multi-input CARE examples and nine multi-input DARE examples in benchmark collections [1,2]. All examples are numbered in the order as they appear in the references.

Example 4.2. *The first benchmark set includes eleven small examples from [5]. Applying the three algorithms on these examples, all algorithms produce comparable precisions of the assigned poles, which are greater than 10, and we omit the results here. Table 4.1 lists the three measures of robustness dep., $\kappa_F(X)$, $\|F\|_F$ by the three algorithms for five examples. The results are generally comparable. The remaining six examples are not listed in the table, as the results of the three algorithms applying on these examples are quite similar.*

num.	*dep.*			$\kappa_F(X)$			$\|F\|_F$		
	place	**robpole**	**schur-rob**	**place**	**robpole**	**schur-rob**	**place**	**robpole**	**schur-rob**
2	3.0e+1	3.0e+1	3.8e+1	5.3e+1	5.3e+1	1.3e+2	4.1e+2	2.3e+2	1.6e+2
3	3.7e+1	3.9e+1	7.2e+1	5.3e+1	5.6e+1	1.2e+2	5.9e+1	4.9e+1	5.0e+1
5	7.4e-1	7.4e-1	7.2e-1	1.5e+2	1.5e+2	1.9e+3	4.9e+0	5.4e+0	4.0e+0
8	1.3e+1	5.0e+0	7.5e+0	3.7e+1	6.2e+0	1.2e+1	1.6e+1	2.9e+1	2.7e+1
9	1.2e+1	1.2e+1	1.8e+1	2.4e+1	2.4e+1	5.8e+1	8.5e+2	8.2e+2	1.5e+3
10	2.5e-3	3.6e-1	2.4e-1	4.0e+0	4.1e+0	4.1e+0	1.9e+0	5.3e+0	1.9e+0

Table 4.1: Robustness of the closed-loop system matrix for the examples from [5]

	$prec_m$									
	1	2	3	4	5	6	7	8	9	10
place	14	14	11	11	11	9	14	11	13	11
robpole	14	14	12	13	12	11	14	14	13	10
schur-rob	14	14	12	8	9	6	14	14	12	9
	$prec_M$									
	1	2	3	4	5	6	7	8	9	10
place	15	15	15	16	16	14	15	15	15	14
robpole	15	15	15	16	16	15	15	15	16	15
schur-rob	15	16	15	14	16	13	16	15	15	14

Table 4.2: Accuracy for CARE examples

	$prec_m$								
	1	2	3	4	5	6	7	8	9
place	-	15	14	14	7	11	5	-	13
robpole	-	15	14	14	7	11	1	-	13
schur-rob	15	15	15	15	8	10	4	-	12
	$prec_M$								
	1	2	3	4	5	6	7	8	9
place	-	15	15	15	15	15	14	15	16
robpole	-	15	15	15	15	14	14	15	15
schur-rob	16	15	15	15	15	14	11	15	15

Table 4.3: Accuracy for DARE examples

Now we apply the three algorithms on ten CARE and nine DARE examples from the SLICOT CARE/DARE benchmark collections ([1, 2]). Tables 4.2- 4.5 present the numerical results, respectively. The "-"s in the first column in Tables 4.3 and the first row in Tables 4.5 corresponding to ***place*** *and* ***robpole*** *mean that both algorithms fail to output a solution, because the algebraic multiplicity of some pole is greater than m. Note that the "$prec_m$" in the eighth column in Table 4.3 are also " -"s, which indicates that there exists at least one eigenvalue of $A+BF$, which owns no relative accuracy corresponding to the assigned poles. From Table 4.2, we know that the maximum relative accuracy $prec_m$ of the poles in example* 4 *and* 5 *corresponding to* ***schur-rob*** *is the smallest. And the reason is that there are semi-simple eigenvalues in these two examples. So how to dispose the issue that semi-simple eigenvalues can achieve higher relative accuracy deserves further exploitation and we will treat it in a separate paper. For the sixth column in Table 4.2, "$prec_m$" from our algorithm is also smaller than those obtained from the other two algorithms for the existence of poles which are relatively bad separated from the imaginary axis. And this is the weakness of our algorithm. From Tables 4.4 and 4.5, we know that all algorithms produce comparable robustness of the closed-loop system matrix.*

We now test the three algorithms on some random examples generated by the MATLAB

num.	*dep.*			$\kappa_F(X)$			$\|F\|_F$		
	place	**robpole**	**schur-rob**	**place**	**robpole**	**schur-rob**	**place**	**robpole**	**schur-rob**
1	5.2e+0	5.2e+0	7.6e+0	7.4e+0	7.3e+0	1.1e+1	4.2e+0	4.4e+0	7.5e+0
2	3.0e-1	2.9e-1	3.0e-1	8.0e+0	8.0e+0	8.2e+0	2.3e+1	1.6e+1	3.2e+1
3	7.3e+2	5.7e+2	1.4e+2	4.3e+1	4.2e+1	9.2e+2	1.8e+5	1.4e+5	3.4e+4
4	1.5e+6	7.5e+5	1.1e+5	1.7e+15	2.2e+7	9.0e+7	2.2e+3	2.2e+2	1.2e+2
5	2.9e+6	2.9e+6	7.3e+6	8.5e+4	8.9e+4	2.0e+6	2.8e+1	2.8e+1	2.9e+1
6	2.3e+7	2.3e+7	2.3e+7	4.8e+6	3.2e+6	3.2e+8	3.4e+6	2.6e+6	1.2e+7
7	7.6e+0	8.1e+0	7.5e+0	1.6e+1	1.6e+1	3.3e+1	7.6e+0	8.4e+0	7.8e+0
8	2.2e+1	2.0e+1	2.1e+1	9.8e+1	9.0e+1	5.7e+2	2.3e+1	2.0e+1	2.1e+1
9	6.1e+0	6.0e+0	8.4e+0	1.5e+2	1.4e+2	6.5e+3	6.3e+0	2.7e+0	2.4e+1
10	4.9e+9	3.8e+9	2.2e+10	2.3e+6	2.3e+6	4.3e+6	3.0e+13	7.3e+13	9.5e+13

Table 4.4: Robustness of the closed-loop system matrix for ten CARE examples

num.	*dep.*			$\kappa_F(X)$			$\|F\|_F$		
	place	**robpole**	**schur-rob**	**place**	**robpole**	**schur-rob**	**place**	**robpole**	**schur-rob**
1	-	-	1.0e-1	-	-	7.1e+15	-	-	1.0e+0
2	2.2e-1	2.2e-1	2.5e-1	5.2e+0	5.2e+0	5.5e+0	2.1e+0	2.5e+0	2.3e+0
3	3.9e-1	3.9e-1	1.3e+0	4.9e+0	5.0e+0	5.6+0	1.5e+3	1.5e+3	9.8e+2
4	4.3e-1	3.6e-1	3.4e-1	5.4e+0	5.3e+0	7.2e+0	1.3e+1	2.0e+1	1.3e+1
5	1.7e+0	1.7e+0	1.7e+0	1.8e+1	1.8e+1	1.8e+1	6.7e-1	1.7e+1	6.7e-1
6	1.4e+0	1.3e+0	2.0e+0	1.3e+1	1.2e+1	3.8e+1	6.9e+4	6.0e+4	3.0e+5
7	2.3e+1	1.8e+1	1.9e+1	2.3e+8	2.9e+8	1.7e+9	1.5e+2	3.8e+1	1.9e+1
8	4.3e+7	3.9e+12	9.8e+0	9.2e+292	1.3e+308	5.6e+292	4.3e+7	3.9e+12	6.5e+0
9	8.9e+0	8.0e+0	9.9e+0	3.4e+2	3.0e+2	2.2e+4	1.5e+2	1.3e+2	1.5e+2

Table 4.5: Robustness of the closed-loop system matrix for nine DARE examples

function **randn**.

Example 4.3. *This test set includes* 33 *examples where* n *varies from 3 to 25 increased by 2, and* m *is set to be* $2, \lfloor \frac{n}{2} \rfloor, n-1$ *for each* n*. The examples are generated as following. We first randomly generate the matrices* A, B *and* F *by the MATLAB function* ***randn****, and then get* $\mathfrak{L}$ *using the MATLAB function* ***eig****, that is,* $\mathfrak{L} = eig(A + BF)$*. We then apply the three algorithms* ***place****,* ***robpole*** ***schur-rob*** *on the* A, B *and* $\mathfrak{L}$ *as input.*

Fig. 4.1 to Fig. 4.5 respectively display the departure from normality of the computed A_c*, the condition number of the eigenvector matrix* X*, Frobenius norm of the feedback* F*, relative accuracy of the poles and CPU time of the three algorithms applied on these randomly generated examples. In these figures, the* x*-axis represents the number of the* 33 *different sizes* (n, m)*, and the values along the* y*-axis are the mean values over 50 trials for a certain* (n, m)*.*

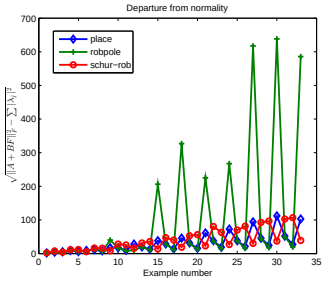

Fig. 4.1: *dep.* over 50 trials

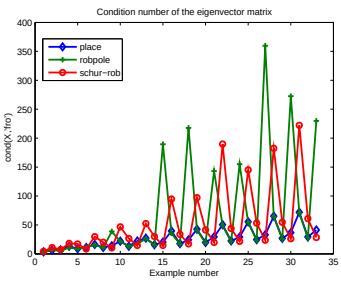

Fig. 4.2: $\kappa_F(X)$ over 50 trials

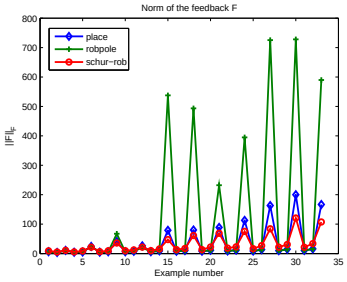

Fig. 4.3: $\|F\|_F$ over 50 trials

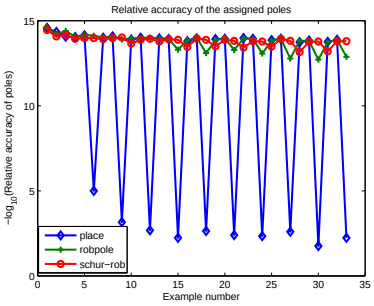

Fig. 4.4: Accuracy over 50 trials

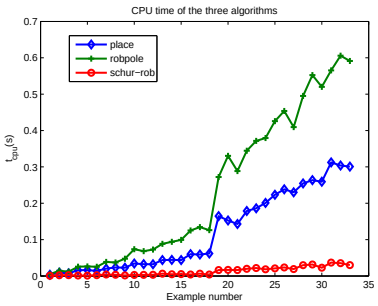

Fig. 4.5: CPU time over 50 trials

All these figures show that our ***schur-rob*** *algorithm can produce comparable or even better results as the other two algorithms, but with much less CPU time.*

5 Conclusion

Pole assignment problem for multi-input control is generally under-determined. And utilizing this freedom to make the closed-loop system matrix to be insensitive to perturbations as far as possible evokes the state-feedback robust pole assignment problem (**SFRPA**) arising. In this paper, a new direct method based on [8] is proposed to solve the **SFRPA**, which obtains the real Schur form of the closed-loop system matrix and tends to minimize its departure from normality via solving standard eigen-problems. Many numerical examples show that our algorithm can produce comparable or even better results than existing methods, but with much less computational costs.

Acknowledgements

The authors would like to express our gratitude to Professor Tits and Dr.Sima for providing some codes needed in this paper. And we also want to thank Dr.Yang for his selfless help.

References

[1] J. Abels and P. Benner, CAREX - A collection of benchmark examples for continuous-time algebraic Riccati equations (Version 2.0), Katholieke Universiteit Leuven, ESAT/SISTA, Leuven, Belgium, SLICOT Working Note 1999-14, Nov. 1999. [Online]. Available: `http://www.slicot.de/REPORTS/SLWN1999-14.ps.gz`.

[2] J. Abels and P. Benner, DAREX - A collection of benchmark examples for discrete-time algebraic Riccati equations (Version 2.0), Katholieke Universiteit Leuven, ESAT/SISTA, Leuven, Belgium, SLICOT Working Note 1999-16, Dec. 1999. [Online]. Available: `http://www.slicot.de/REPORTS/SLWN1999-16.ps.gz`.

[3] A.N. Andry, E.Y. Shapiro and J.C. Chung, Eigenstructure assignment for linear systems, *IEEE Transactions on Aerospace and Electronic Systems*, 19(1983), 711–729.

[4] S.P. Bhattacharyya and E. De Souza, Pole assignment via Sylvester's equation, *Systems & Control Letters*, 1(1982), 261–263.

[5] R. Byers and S.G. Nash, Approaches to robust pole assignment, *International Journal of Control*, 49(1989), 97–117.

[6] R.K. Cavin and S.P. Bhattacharyya, Robust and well-conditioned eigenstructure assignment via sylvester's equation, *Optimal Control Applications and Methods*, 4(1983), 205–212.

[7] E.K.W. Chu, A pole-assignment algorithm for linear state feedback, *System & Control Letters*, 7(1986), 289–299.

[8] E.K.W. Chu, Pole assignment via the schur form, *Systems & Control Letters*, 56(2007), 303–314.

[9] A. Dickman, On the robustness of multivariable linear feedback systems in state-space representation, *IEEE Transactions on Automatic Control*, 32(1987), 407–410.

[10] M. Fahmy and J. O'Reilly, On eigenstructure assignment in linear multivariable systems, *IEEE Transactions on Automatic Control*, 27(1982), 690–693.

[11] P. Henrici, Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices, *Numerische Mathematik* , 4(1962), 24–40.

[12] S.K. Katti, Pole placement in multi-input systems via elementary transformations, *International Journal of Control*, 37(1983), 315–347.

[13] J. Kautsky, N.K. Nichols and P. Van Dooren, Robust pole assignment in linear state feedback, *International Journal of Control*, 41(1985), 1129–1155.

[14] J. Lam and W.Y. Van, A gradient flow approach to the robust pole-placement problem, *International Journal of Robust and Nonlinear Control*, 5(1995), 175–185.

[15] G.S. Miminis and C.C. Paige, A direct algorithm for pole assignment of time-invariant multi-input linear systems using state feedback, *Automatica*, 24(1988), 343–356.

[16] G.S. Miminis and C.C. Paige, A QR-like approach for the eigenvalue assignment problem, in *Proceedings of the 2nd Hellenic Conference on Mathematics and Informatics*, Athens, Greece, Sept., 1994.

[17] R.V. Patel and P. Misra, Numerical algorithms for eigenvalue assignment by state feedback, *Proceedings of the IEEE* , 72(1984), 1755–1764.

[18] P.Hr. Petkov, N.D. Christov and M.M. Konstantinov, A computational algorithm for pole assignment of linear multiinput systems, *IEEE Transactions on Automatic Control*, 31(1986), 1044–1047.

[19] K. Ramar and V. Gourishankar, Utilization of the design freedom of pole assignment feedback controllers of unrestricted rank, *International Journal of Control*, 24(1976), 423–430.

[20] D.G. Retallack and A.G.J. MacFarlane, Pole-shifting techniques for multivariable feedback systems, *Proceedings of the Institution of Electrical Engineers*, 117(1970), 1037–1038.

[21] V. Sima, A.L. Tits and Y. Yang, Computational experience with robust pole assignment algorithms,in *Proceedings of the 2006 IEEE Conference on Computer Aided Control Systems Design*, Munich, Germany, Oct. 4-6, 2006.

[22] G.W. Stewart and J.G. Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.

[23] A.L. Tits and Y. Yang, Globally convergent algorithms for robust pole assignment by state feedback, *IEEE Transactions on Automatic Control* 41(1996), 1432–1452.

[24] A. Varga, A Schur method for pole assignment, *IEEE Transactions on Automatic Control*, 26(1981), 517–519.

[25] S.F. Xu, *An Introduction to Inverse Algebraic Eigenvalue Problems*, Peking University Press, Beijing, and Vieweg, Braunschweig, 1998.

[26] W.M. Wonham, On pole assignment in multi-input controllable linear systems, *IEEE Transactions on Automatic Control*, 12(1967), 660–665.

[27] W.M. Wonham, *Linear Multivariable Control: A Geometric Approach*, 2nd ed., Springer-Verlag, New York, 1979.