The Projected Generalized Sylvester Equations: Numerical Solution and Applications

YIQIN LIN
Hunan University of Science and Engineering
Institute of Computational Mathematics
Yangzhitang Road 130, 425199 Yongzhou
CHINA
yqlin@aliyun.com

LIANG BAO
East China University of Science and Technology
Department of Mathematics
Meilong Road 130, 200237 Shanghai
CHINA
nlbao@aliyun.com

Abstract: In this paper we consider the numerical solution of large-scale projected generalized continuous-time and discrete-time Sylvester equations with low-rank right-hand sides. First, we present the results on the sufficient conditions for the existence, uniqueness, and analytic formula of the solutions of these equations. Second, we review the low-rank alternating direction implicit method and the low-rank cyclic Smith method for solving the projected generalized continuous-time Sylvester equation, and propose a numerical method for the projected generalized discrete-time Sylvester equation. Third, we show that the solutions of these two equations are useful for computing the $\mathbb{HL}_2$ inner product of two descriptor systems. Finally, we present some numerical experiments.

Key–Words: Projected generalized Sylvester equation, Matrix pencil, Alternating direction implicit method, Smith method, Descriptor system, $\mathbb{HL}_2$ inner product

1 Introduction

In this paper we consider the projected generalized continuous-time Sylvester equation (PGCTSE)

$$AX\widetilde{E} + EX\widetilde{A} + P_l F\widetilde{P}_r = 0, \quad X = P_r X\widetilde{P}_l \tag{1}$$

and the projected generalized discrete-time Sylvester equation (PGDTSE)

$$AY\widetilde{A} - EY\widetilde{E} = (I - P_l)F(I - \widetilde{P}_r), \quad P_r Y\widetilde{P}_l = 0 \tag{2}$$

where $E, A \in \mathbb{R}^{n\times n}$, $\widetilde{E}, \widetilde{A} \in \mathbb{R}^{m\times m}$, $F \in \mathbb{R}^{n\times m}$, and $X, Y \in \mathbb{R}^{n\times m}$ are the sought-after solutions. Here, P_r and $\widetilde{P}_r$ are the spectral projectors onto the right deflating subspaces corresponding to the finite eigenvalues of the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$, respectively, while P_l and $\widetilde{P}_l$ are the spectral projectors onto the left deflating subspaces corresponding to the finite eigenvalues of $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$, respectively.

We assume that the matrices E and $\widetilde{E}$ are singular, but the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ are regular, i.e., $\det(\lambda E - A)$ and $\det(\lambda\widetilde{E} - \widetilde{A})$ are not identically zero. Under this assumption, the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ have the Weierstrass canonical forms [11]: there exist nonsingular $n \times n$ matrices W, T and $m \times m$ matrices $\widetilde{W}, \widetilde{T}$ such that

$$E = W\begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}T, \qquad A = W\begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix}T, \tag{3}$$

$$\widetilde{E} = \widetilde{W}\begin{bmatrix} I & 0 \\ 0 & \widetilde{N} \end{bmatrix}\widetilde{T}, \qquad \widetilde{A} = \widetilde{W}\begin{bmatrix} \widetilde{J} & 0 \\ 0 & I \end{bmatrix}\widetilde{T}, \tag{4}$$

where $J \in \mathbb{R}^{n_f\times n_f}$, $\widetilde{J} \in \mathbb{R}^{m_f\times m_f}$, $N \in \mathbb{R}^{(n-n_f)\times(n-n_f)}$ and $\widetilde{N} \in \mathbb{R}^{(m-m_f)\times(m-m_f)}$ are block diagonal matrices with each diagonal block being a Jordan block. The eigenvalues of J and $\widetilde{J}$ are the finite eigenvalues of the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$, respectively, while N and $\widetilde{N}$ correspond to the eigenvalue at infinity. The indices ν and $\tilde{\nu}$ of nilpotency of N and $\widetilde{N}$ are called the indices of the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$, respectively. Using (3) and (4), P_l, P_r, $\widetilde{P}_l$ and $\widetilde{P}_r$ can be expressed as

$$P_l = W\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}W^{-1}, \quad P_r = T^{-1}\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}T, \tag{5}$$

$$\widetilde{P}_l = \widetilde{W}\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\widetilde{W}^{-1}, \quad \widetilde{P}_r = \widetilde{T}^{-1}\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\widetilde{T}. \tag{6}$$

A number of numerical solution methods have been proposed for the standard /generalized Lyapunov and Sylvester equations. Two classical direct methods are the Bartels-Stewart method [3] and the Hammarling method [14]. These methods need to compute the real Schur forms/generalized real Schur forms of the

underlying matrices/matrix pencils by means of the QR/QZ algorithm [13]. Besides direct methods, we mention, among several iterative methods, the Smith method [24], the alternating direction implicit iteration (ADI) method [19, 32], the Smith(l) method [21], the low-rank Smith method [12, 21, 23], the Cholesky factor ADI method [4, 6, 8, 17], and the (generalized) matrix sign function method [5, 9, 10, 7, 16, 30].

As $m = n$, $\widetilde{A} = A^T$, and $\widetilde{E} = E^T$, (1) is referred to as the projected generalized continuous-time Lyapunov equation and (2) as the projected generalized discrete-time Lyapunov equation, which arise in stability analysis and control design problems for descriptor systems including the characterization of controllability and observability properties, balanced truncation model order reduction, determining the minimal and balanced realizations as well as computing $\mathbb{H}_2$ and Hankel norms; see [1, 15, 20, 22, 27, 31] and the references therein. If the pencil $\lambda E - A$ is c-stable, i.e., all its finite eigenvalues have negative real part, then the projected generalized Lyapunov equations have unique solutions for each F, and if, additionally, F is symmetric and semi-positive definite, then the solutions are symmetric and semi-positive definite, see, e.g., [26] for details. Lyapunov equations are useful in solving Riccati equations, see [2, 33].

Recently, several numerical methods have been proposed in the literature for solving the projected generalized Lyapunov equations. In [25], two direct methods, the generalized Bartels-Stewart method and the generalized Hammarling method, were proposed for the projected generalized Lyapunov equations. The generalized Hammarling method is designed to obtain the Cholesky factors of the solutions. These two methods are based on the generalized real Schur form of the pencil $\lambda E - A$, and require $O(n^3)$ flops and $O(n^2)$ memory. Iterative methods to solve the projected generalized Lyapunov equations have also been proposed. Stykel [31] extended the ADI method and the Smith method to the projected equations. Moreover, their low-rank versions were also presented, which could be used to compute low-rank approximations to the solutions. These methods are especially suitable for large sparse equations with low-rank F. Another iterative method for the projected generalized Lyapunov equations is the modified generalized matrix sign function method [30]. Unlike the classical generalized matrix sign function method, the variant converges quadratically independent of the index of the underlying matrix pencil, see [30] for more details. In [18], the modified generalized matrix sign function method is extended to the PGCTSE (1). This method requires the inverses of two matrices at each iteration and has a computational complexity $O(\max\{n^3, m^3\})$, and thus is not suitable for large sparse problems. The Krylov subspace method for the projected generalized Lyapunov equations is proposed in [34].

In this paper, we firstly present the results on the sufficient conditions for the existence, uniqueness, and analytic formula of the solutions of the PGCTSE and the PGDTSE. Then, we review the low-rank ADI method and the low-rank cyclic Smith method for solving the PGCTSE with low-rank right-hand sides. We propose a numerical method for the PGDTSE. Finally, we show that the solutions of these two equations are useful for computing the $\mathbb{HL}_2$ inner product of two descriptor systems.

Throughout this paper, we adopt the following notation. The square identity and zero matrices are denoted by I and 0, respectively. The spaces of $m \times n$ real matrices are denoted by $\mathbb{R}^{m\times n}$. The 2-norm and the Frobenius matrix norm are denoted by $\|\cdot\|_2$ and $\|\cdot\|_F$, respectively. The superscript " $\cdot^T$ " denotes the transposition of a vector or a matrix. Moreover, $\rho(A)$ is the spectral radius of square matrix A and $\kappa(A) = \|A\|_2\|A^{-1}\|_2$ is the spectral condition number of A. The open left is denoted by $\mathbb{C}^-$.

The remainder of the paper is organized as follows. In Section 2, we give the sufficient conditions for the existence, uniqueness, and analytic formula of the solutions of the PGCTSE and the PGDTSE. In Section 3, we propose the low-rank ADI method and the low-rank cyclic Smith method to the PGCTSE with low-rank right-hand sides. In Section 4, we present a numerical method for the PGDTSE. In Section 5, we present an application. Section 6 is devoted to numerical experiments. Some concluding remarks are given in the last section.

2 Solution of PGCTSE and PGDTSE

In this section, we consider the solutions of the PGCTSE (1) and the PGDTSE (2).

The following theorem gives sufficient conditions for the existence, uniqueness, and analytic formula of the solution of the PGCTSE (1); see, e.g., [18].

Theorem 1 *Let $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ be regular pencils with finite eigenvalues $\{\zeta_1, \zeta_2, \cdots, \zeta_{n_f}\}$ and $\{\eta_1, \eta_2, \cdots, \eta_{m_f}\}$ counted according to their multiplicities, respectively. Then, the projected generalized continuous-time Sylvester equation (1) has a unique solution for every F if $\zeta_i + \eta_j \neq 0$ for any $i = 1, 2, \cdots, n_f$ and $j = 1, 2, \cdots, m_f$.*

Moreover, if $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ are c-stable, i.e., all their finite eigenvalues have negative real part,

then X can be expressed as

$$X = \frac{1}{2\pi}\int_{-\infty}^{\infty}(-i\omega E - A)^{-1}P_l F\widetilde{P}_r(i\omega\widetilde{E} - \widetilde{A})^{-1}d\omega.$$

We now consider the solution of the PGDTSE (2). Similarly, we have the following result.

Theorem 2 *Let $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ be regular pencils with A and $\widetilde{A}$ nonsingular. Then, the PGDTSE (2) has a unique solution Y for every F. Moreover, the solution Y can be expressed as*

$$\begin{aligned} Y &= \sum_{k=0}^{\min\{\nu,\tilde{\nu}\}-1}(A^{-1}E)^k A^{-1}(I-P_l)F \\ &\quad\cdot(I-\widetilde{P}_r)\widetilde{A}^{-1}(\widetilde{E}\widetilde{A}^{-1})^k \qquad (7)\\ &= \frac{1}{2\pi}\int_0^{2\pi}(e^{-i\omega}E - A)^{-1}(I-P_l)F(I-\widetilde{P}_r) \\ &\quad\cdot(e^{i\omega}\widetilde{E} - \widetilde{A})^{-1}d\omega. \qquad (8)\end{aligned}$$

Proof: Define $\widehat{Y} = TY\widetilde{W}$ and $\widehat{F} = W^{-1}F\widetilde{T}^{-1}$, and partition $\widehat{Y}$ and $\widehat{F}$ appropriately as

$$\widehat{Y} = \begin{bmatrix} \widehat{Y}_{11} & \widehat{Y}_{12} \\ \widehat{Y}_{21} & \widehat{Y}_{22}\end{bmatrix}, \quad \widehat{F} = \begin{bmatrix} \widehat{F}_{11} & \widehat{F}_{12} \\ \widehat{F}_{21} & \widehat{F}_{22}\end{bmatrix}.$$

Using (5) and (6), we can rewrite the second equation of (2) into

$$\begin{aligned} 0 &= P_r Y\widetilde{P}_l = T^{-1}\begin{bmatrix} I & 0\\ 0 & 0\end{bmatrix}TY\widetilde{W}\begin{bmatrix} I & 0\\ 0 & 0\end{bmatrix}\widetilde{W}^{-1} \\ &= T^{-1}\begin{bmatrix} I & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix} \widehat{Y}_{11} & \widehat{Y}_{12}\\ \widehat{Y}_{21} & \widehat{Y}_{22}\end{bmatrix}\begin{bmatrix} I & 0\\ 0 & 0\end{bmatrix}\widetilde{W}^{-1}.\end{aligned}$$

Thus, this shows that $\widehat{Y}_{11} = 0$.

By using the expressions of $E, A, E, \widetilde{A}$ in (3) and (4), we obtain

$$\begin{aligned} AY\widetilde{A} &= W\begin{bmatrix} J & 0\\ 0 & I\end{bmatrix}TY\widetilde{W}\begin{bmatrix} \widetilde{J} & 0\\ 0 & I\end{bmatrix}\widetilde{T} \\ &= W\begin{bmatrix} J & 0\\ 0 & I\end{bmatrix}\begin{bmatrix} 0 & \widehat{Y}_{12}\\ \widehat{Y}_{21} & \widehat{Y}_{22}\end{bmatrix}\begin{bmatrix} \widetilde{J} & 0\\ 0 & I\end{bmatrix}\widetilde{T} \\ &= W\begin{bmatrix} 0 & J\widehat{Y}_{12}\\ \widehat{Y}_{21}\widetilde{J} & \widehat{Y}_{22}\end{bmatrix}\widetilde{T}, \qquad (9)\end{aligned}$$

$$\begin{aligned} EY\widetilde{E} &= W\begin{bmatrix} I & 0\\ 0 & N\end{bmatrix}TY\widetilde{W}\begin{bmatrix} I & 0\\ 0 & \widetilde{N}\end{bmatrix}\widetilde{T} \\ &= W\begin{bmatrix} I & 0\\ 0 & N\end{bmatrix}\begin{bmatrix} 0 & \widehat{Y}_{12}\\ \widehat{Y}_{21} & \widehat{Y}_{22}\end{bmatrix}\begin{bmatrix} I & 0\\ 0 & \widetilde{N}\end{bmatrix}\widetilde{T} \\ &= W\begin{bmatrix} 0 & \widehat{Y}_{12}\widetilde{N}\\ N\widehat{Y}_{21} & N\widehat{Y}_{22}\widetilde{N}\end{bmatrix}\widetilde{T}, \qquad (10)\end{aligned}$$

$$\begin{aligned} &(I-P_l)F(I-\widetilde{P}_r) \\ &= W\begin{bmatrix} 0 & 0\\ 0 & I\end{bmatrix}W^{-1}F\widetilde{T}^{-1}\begin{bmatrix} 0 & 0\\ 0 & I\end{bmatrix}\widetilde{T} \\ &= W\begin{bmatrix} 0 & 0\\ 0 & I\end{bmatrix}\begin{bmatrix} \widehat{F}_{11} & \widehat{F}_{12}\\ \widehat{F}_{21} & \widehat{F}_{22}\end{bmatrix}\begin{bmatrix} 0 & 0\\ 0 & I\end{bmatrix}\widetilde{T} \\ &= W\begin{bmatrix} 0 & 0\\ 0 & \widehat{F}_{22}\end{bmatrix}\widetilde{T}. \qquad (11)\end{aligned}$$

By (9), (10), and (11), the PGDTSE (2) can be equivalently written as

$$\begin{cases} N\widehat{Y}_{21} - \widehat{Y}_{21}\widetilde{J} = 0, \\ J\widehat{Y}_{12} - \widehat{Y}_{12}\widetilde{N} = 0, \\ \widehat{Y}_{22} - N\widehat{Y}_{22}\widetilde{N} = \widehat{F}_{22}. \end{cases} \qquad (12)$$

Since A and $\widetilde{A}$ are nonsingular, J and $\widetilde{J}$ have no zero eigenvalue. Thus, the first two equations in (12), which are standard continuous-time Sylvester equations, have unique trivial solutions, respectively; see [11].

The third equation in (12) is a standard discrete-time Sylvester equation, and has a unique solution, which can be formulated as

$$\begin{aligned} \widehat{Y}_{22} &= \sum_{k=0}^{\min\{\nu,\tilde{\nu}\}-1}N^k\widehat{F}_{22}\widetilde{N}^k \\ &= \frac{1}{2\pi}\int_0^{2\pi}(e^{-i\omega}N - I)^{-1}\widehat{F}_{22}(e^{i\omega}\widetilde{N} - I)^{-1}d\omega.\end{aligned}$$

By the expression of $\widehat{Y}_{22}$ and $\widehat{Y} = TY\widetilde{W}$, it is easy to verify that (7) and (8) hold. □

3 Iterative method for PGCTSE

The ADI method and the Smith method are among the most popular iterative methods for solving large sparse standard Lyapunov or Sylvester matrix equations; see, e.g., [8, 17, 19, 21, 24, 32]. Recently, these iterative methods are extended to projected generalized Lyapunov equations by Stykel [31]. In this section, we review the ADI method and the Smith method for iteratively solving the PGCTSE (1), see [35].

3.1 The ADI method

In the remainder of this paper, we always assume that the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ are c-stable, i.e., all their finite eigenvalues have negative real part. In this case, the matrices A and $\widetilde{A}$ are nonsingular.

Multiplying the first equation in (1) on the left by A^{-1} and on the right by $\widetilde{A}^{-1}$, we get the following projected standard Sylvester equation

$$\begin{aligned}&(A^{-1}E)X + X\widetilde{E}\widetilde{A}^{-1} + A^{-1}P_lF\widetilde{P}_r\widetilde{A}^{-1} = 0,\\ &X = P_rX\widetilde{P}_l.\end{aligned} \tag{13}$$

The iterates X_i of the ADI iteration for (13) are usually generated by the alternating solution of two linear systems with multiple right-hand sides

$$\begin{aligned}&(A^{-1}E + \beta_iI)X_{i-\frac{1}{2}} =\\ &-X_{i-1}(\widetilde{E}\widetilde{A}^{-1} - \beta_iI) - A^{-1}P_lF\widetilde{P}_r\widetilde{A}^{-1},\\ &X_i(\widetilde{E}\widetilde{A}^{-1} + \alpha_iI) =\\ &-(A^{-1}E - \alpha_iI)X_{i-\frac{1}{2}} - A^{-1}P_lF\widetilde{P}_r\widetilde{A}^{-1},\end{aligned}$$

where $X_0 = 0$ and the shift parameters $\{\alpha_i\}_{i=1}^k$ and $\{\beta_i\}_{i=1}^k$ are elements of $\mathbb{C}^-$. These two equations are equivalent to the following single iteration step:

$$\begin{aligned}X_i &= (A^{-1}E + \beta_iI)^{-1}(A^{-1}E - \alpha_iI)X_{i-1}\\ &\cdot(\widetilde{E}\widetilde{A}^{-1} - \beta_iI)(\widetilde{E}\widetilde{A}^{-1} + \alpha_iI)^{-1}\\ &-(\alpha_i + \beta_i)(A^{-1}E + \beta_iI)^{-1}A^{-1}\\ &\cdot P_lF\widetilde{P}_r\widetilde{A}^{-1}(\widetilde{E}\widetilde{A}^{-1} + \alpha_iI)^{-1}.\end{aligned} \tag{14}$$

We can rewrite the iteration (14) as

$$\begin{aligned}X_i &= (E + \beta_iA)^{-1}(E - \alpha_iA)X_{i-1}(\widetilde{E} - \beta_i\widetilde{A})(\widetilde{E} + \alpha_i\widetilde{A})^{-1}\\ &\quad -(\alpha_i + \beta_i)(E + \beta_iA)^{-1}P_lF\widetilde{P}_r(\widetilde{E} + \alpha_i\widetilde{A})^{-1}.\end{aligned} \tag{15}$$

Let X denote the exact solution of (1). Then it is easy to verify that the error matrix $X - X_i$ obeys the recursion

$$\begin{aligned}X - X_i &= (E + \beta_iA)^{-1}(E - \alpha_iA)(X - X_{i-1})\\ &\quad\cdot(\widetilde{E} - \beta_i\widetilde{A})(\widetilde{E} + \alpha_i\widetilde{A})^{-1}\\ &= \cdots\\ &= \mathcal{A}_iX\widetilde{\mathcal{A}}_i,\end{aligned} \tag{16}$$

where

$$\mathcal{A}_i = P_r(E+\beta_iA)^{-1}(E-\alpha_iA)\cdots(E+\beta_1A)^{-1}(E-\alpha_1A), \tag{17}$$

and

$$\widetilde{\mathcal{A}}_i = (\widetilde{E}-\beta_1\widetilde{A})(\widetilde{E}+\alpha_1\widetilde{A})^{-1}\cdots(\widetilde{E}-\beta_i\widetilde{A})(\widetilde{E}+\alpha_i\widetilde{A})^{-1}\widetilde{P}_l. \tag{18}$$

By using (3), (4), (5), and (6), we obtain

$$\mathcal{A}_i = T^{-1}\begin{bmatrix} J_i & 0\\ 0 & 0\end{bmatrix}T, \quad \widetilde{\mathcal{A}}_i = \widetilde{W}\begin{bmatrix} \widetilde{J}_i & 0\\ 0 & 0\end{bmatrix}\widetilde{W}^{-1}, \tag{19}$$

where

$$\begin{aligned}J_i &= (I + \beta_iJ)^{-1}(I - \alpha_iJ)\cdots(I + \beta_1J)^{-1}(I - \alpha_1J),\\ \widetilde{J}_i &= (I - \beta_1\widetilde{J})(I + \alpha_1\widetilde{J})^{-1}\cdots(I - \beta_i\widetilde{J})(I + \alpha_i\widetilde{J})^{-1}.\end{aligned}$$

This implies that if $\{\alpha_i\}_{i=1}^k$ contains all finite eigenvalues (multiple eigenvalues counted by their algebraic multiplicities) of $A^{-1}E$ or if $\{\beta_i\}_{i=1}^k$ contains all finite eigenvalues of $\widetilde{A}^{-1}\widetilde{E}$, then $X - X_i \equiv 0$. This is due to the Cayley-Hamilton theorem [11], which states that $p(A) \equiv 0$ for A's characteristic polynomial $p(\lambda) = \det(\lambda I - A)$.

From equation (16), we can see that the parameters $\{\alpha_i\}_{i=1}^k$ and $\{\beta_i\}_{i=1}^k$ should be chosen to achieve $\min\limits_{\alpha_j,\beta_j} \|\mathcal{A}_i\|_2 \cdot \|\widetilde{\mathcal{A}}_i\|_2$.

About the ith approximate solution X_i of the ADI method, by using the equations (16)-(19), we have the following estimate.

Theorem 3 *Assume J and $\widetilde{J}$, in the Weierstrass form (3) and (4) of the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$, are diagonal. Then*

$$\|X - X_i\|_2 \leqslant \kappa(T)\kappa(\widetilde{W})\rho(\mathcal{A}_i)\rho(\widetilde{\mathcal{A}}_i)\|X\|_2.$$

Note that the iterate X_i is explicitly computed by the ADI iteration (15), so the storage requirement is $O(mn)$. One should notice that in many cases the storage requirement is the limiting factor rather than the amount of computation. We note that low-rank schemes are the only existing methods that can effectively solve large-scale Lyapunov/Sylvester equations.

Assume that the low-rank right-hand side F has the factored form $F = BC$ with $B \in \mathbb{R}^{n\times r}$ and $C \in \mathbb{R}^{r\times m}$. Instead of explicitly forming X_i, the low-rank method compute and store approximate solutions in low-rank factored form. If the numerical rank l of X is much smaller than $\min\{m, n\}$, i.e. $l \ll \min\{m, n\}$, then the storage is reduced from $O(mn)$ to $O(ml)$ or $O(nl)$.

The key idea in the low-rank version of the ADI iteration is to rewrite the iteration X_i in (15) as an outer product:

$$X_i = U_iH_iV_i.$$

This is always possible since starting with the initial guess $X_0 = 0_{n\times m}$. The low-rank ADI (LR-ADI) method is based on (15). Replacing X_{i-1} with $U_{i-1}H_{i-1}V_{i-1}$, (15) can be reformulated in terms of the low-rank factors as

$$X_i = U_iH_iV_i,$$

where

$$U_i = \left[\begin{array}{cc} (E+\beta_i A)^{-1}P_l B & (E+\beta_i A)^{-1}(E-\alpha_i A)U_{i-1} \end{array} \right],$$
$$H_i = \left[\begin{array}{cc} -(\alpha_i+\beta_i)I & \\ & H_{i-1} \end{array} \right],$$
$$V_i = \left[\begin{array}{c} C\widetilde{P}_r(\widetilde{E}+\alpha_i\widetilde{A})^{-1} \\ V_{i-1}(\widetilde{E}-\beta_i\widetilde{A})(\widetilde{E}+\alpha_i\widetilde{A})^{-1} \end{array} \right].$$

From the fact that Y_0, H_0 and Z_0 are all zero matrices, it can be seen that U_i is $n \times kr$, H_i is $kr \times kr$ and V_i is $kr \times m$. Thus the rank of X_i is no more than kr. Since the order of the ADI parameters $\{\alpha_i\}_{i=1}^k$ and $\{\beta_i\}_{i=1}^k$ is not important, the ordering of $\{\alpha_i\}_{i=1}^k$ and $\{\beta_i\}_{i=1}^k$ can be reversed. As shown in [8], we have the following iterative scheme

$$U_i = \left[\begin{array}{cccc} U^{(1)} & U^{(2)} & \cdots & U^{(i)} \end{array} \right],$$
$$V_i = \left[\begin{array}{cccc} V^{(1)} & V^{(2)} & \cdots & V^{(i)} \end{array} \right],$$

where

$$\begin{aligned} U^{(1)} &= (E+\beta_1 A)^{-1}P_l B, \\ U^{(j+1)} &= (E+\beta_{j+1}A)^{-1}(E-\alpha_j A)U^{(j)}, \\ V^{(1)} &= C\widetilde{P}_r(\widetilde{E}+\alpha_1\widetilde{A})^{-1}, \\ V^{(j+1)} &= V^{(j)}(\widetilde{E}-\beta_j\widetilde{A})(\widetilde{E}+\alpha_{j+1}\widetilde{A})^{-1}. \end{aligned}$$

So we have

$$\begin{aligned} X_i &= U_i H_i V_i, \\ H_i &= \mathrm{diag}\left(-(\alpha_1+\beta_1)I_r, \cdots, -(\alpha_i+\beta_i)I_r\right). \end{aligned}$$

The LR-ADI method for solving the PGCTSE (1) with $F = BC$ is described in Algorithm 1.

Algorithm 1. The LR-ADI method

Input: $E, A \in \mathbb{R}^{n\times n}$, $\widetilde{E}, \widetilde{A} \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times r}$ and $C \in \mathbb{R}^{r\times m}$ with $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ being c-stable; the ADI shifts $\{\alpha_1, \alpha_2, \cdots, \alpha_k\}$ and $\{\beta_1, \beta_2, \cdots, \beta_k\}$ computed by Algorithm 2.

Output: U, H and V such that $X = UHV$ is an approximate solution of the PGCTSE (1) with $F = BC$.

1. Compute $U_1 = (E+\beta_1 A)^{-1}P_l B$, $V_1 = C\widetilde{P}_r(\widetilde{E} + \alpha_1\widetilde{B})^{-1}$, and $H_1 = -(\alpha_1+\beta_1)I_r$;
2. Set $U = U_1$, $V = V_1$, and $H = H_1$;
3. For $i = 1, 2, \cdots$
   $U_{i+1} = (E+\beta_{i+1}A)^{-1}(E-\alpha_i A)U_i$;
   $V_{i+1} = V_i(\widetilde{E}-\beta_i\widetilde{A})(\widetilde{E}+\alpha_{i+1}\widetilde{A})^{-1}$;
   $H_{i+1} = -(\alpha_{i+1}+\beta_{i+1})I_r$;
   $U = [U, U_{i+1}]$;
   $V = \left[\begin{array}{c} V \\ V_{i+1} \end{array} \right]$;
   $H = \mathrm{diag}(H, H_{i+1})$;
   End For

We make a few comments on Algorithm 1:

(i) Since the computation of the Weierstrass canonical form is sensitive under small perturbations, we should make use of the generalized real Schur factorization to compute the spectral projectors, see, for example, [25]. For large-scale problems, the computation of the spectral projectors by the generalized real Schur factorization may be very expensive. However, in some applications the spectral projectors can be expressed in explicit form by using the special block structure of the matrices E, A, $\widetilde{E}$ and $\widetilde{A}$; see numerical examples in this paper or the reference [31].

(ii) If the number of shift parameters k is smaller than the number of iterations required to obtain a prescribed tolerance, then we reuse these parameters in a cyclic manner.

In the remainder of this subsection, we will consider the choice of shift parameters, which is vitally important to the successful application of the LR-ADI iteration. The rate of convergence is dominated by spectral radii of matrices $\mathcal{A}_i$ and $\widetilde{\mathcal{A}}_i$ given by (19). The minimization of these spectral radii with respect to the parameters $\{\alpha_j\}_{j=1}^i$ and $\{\beta_j\}_{j=1}^i$ leads to the generalized ADI minimax problem

$$\min_{\substack{\{\alpha_1,\ldots,\alpha_i\}\in\mathbb{C}^- \\ \{\beta_1,\ldots,\beta_i\}\in\mathbb{C}^-}} \max_{\substack{x\in\mathbb{E}_f \\ y\in\mathbb{F}_f}} \prod_{j=1}^{i} \left| \frac{(1-\alpha_j x)(1-\beta_j y)}{(1+\beta_j x)(1+\alpha_j y)} \right|,$$

where $\mathbb{E}_f$ and $\mathbb{F}_f$ denote two sets of finite eigenvalues of the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$, respectively. In practice since the eigenvalues of the pencils $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ are unknown and expensive to compute, $\mathbb{E}_f$ and $\mathbb{F}_f$ are often replaced by domains that contain the eigenvalues of $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$, respectively. Note that A and $\widetilde{A}$ are nonsingular, this problem is equivalent to find $\{\alpha_j\}_{j=1}^i$ and $\{\beta_j\}_{j=1}^i$ such that

$$\min_{\substack{\{\alpha_1,\ldots,\alpha_i\}\subset\mathbb{C}^- \\ \{\beta_1,\ldots,\beta_i\}\subset\mathbb{C}^-}} \max_{\substack{x\in\mathbb{E}\setminus\{0\} \\ y\in\mathbb{F}\setminus\{0\}}} \prod_{j=1}^{i} \left| \frac{(x-\alpha_j)(y-\beta_j)}{(x+\beta_j)(y+\alpha_j)} \right|, \quad (20)$$

where $\mathbb{E}$ and $\mathbb{F}$ denote spectrums of the matrices $A^{-1}E$ and $\widetilde{A}^{-1}\widetilde{E}$, respectively.

To compute the suboptimal ADI shift parameters for the standard Lyapunov equation with $\widetilde{A} = A^T$ and $E = \widetilde{E} = I$, a heuristic algorithm has been proposed in [21]. It chooses suboptimal ADI parameters from a set R, which is taken to be the union of the Ritz values of A and the reciprocals of the Ritz values of A^{-1}, obtained by two Arnoldi processes, with A and A^{-1}. As shown in [21], the Ritz values obtained by the Arnoldi process with A tend to be located near the "outer" eigenvalues, i.e., the eigenvalues near the convex hull of the spectrum. In particular, the eigenvalues of large magnitude are usually approximated well. In contrast, they are generally poor approximations to the eigenvalues near the origin. Therefore, one computes the reciprocals of the Ritz values obtained by the Arnoldi process with A^{-1} to approximate the eigenvalues near the origin.

In [8], this idea has been extended to Sylvester equations. A heuristic procedure which is easy to implement has been proposed in [8]. It can also be naturally extended to the generalized problem (20). Here, we need to compute the largest and smallest (in modulus) non-zero approximate eigenvalues of $A^{-1}E$ and $\widetilde{A}^{-1}\widetilde{E}$, respectively. Note that E and $\widetilde{E}$ are assumed to be singular, inverses of $A^{-1}E$ and $\widetilde{A}^{-1}\widetilde{E}$ do not exist. In [31], Stykel proposed a strategy to overcome this difficult. Define

$$\begin{aligned} P &= P_r(EP_r - AQ_r)^{-1} \\ &= (P_lE - Q_lA)^{-1}P_l = T^{-1}\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}W^{-1} \end{aligned}$$

and

$$\begin{aligned} \widetilde{P} &= \widetilde{P}_r(\widetilde{E}\widetilde{P}_r - \widetilde{A}\widetilde{Q}_r)^{-1} \\ &= (\widetilde{P}_l\widetilde{E} - \widetilde{Q}_l\widetilde{A})^{-1}\widetilde{P}_l = \widetilde{T}^{-1}\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\widetilde{W}^{-1}, \end{aligned}$$

where T, $\widetilde{T}$, W and $\widetilde{W}$ are the transformation matrices as in (3) and (4). A simple calculation gives that

$$PA = T^{-1}\begin{bmatrix} J & 0 \\ 0 & 0 \end{bmatrix}T, \quad \widetilde{P}\widetilde{A} = \widetilde{T}^{-1}\begin{bmatrix} \widetilde{J} & 0 \\ 0 & 0 \end{bmatrix}\widetilde{T}.$$

Then it is clear that the reciprocals of the smallest non-zero eigenvalues of $A^{-1}E$ and $\widetilde{A}^{-1}\widetilde{E}$ are the largest eigenvalues of PA and $\widetilde{P}\widetilde{A}$, respectively. Thus, we can run two Arnoldi processes with the matrices PA and $\widetilde{P}\widetilde{A}$ to compute the smallest non-zero eigenvalues of $A^{-1}E$ and $\widetilde{A}^{-1}\widetilde{E}$, respectively. In some applications, similar to the projectors P_r, P_l, $\widetilde{P}_r$ and $\widetilde{P}_l$, the matrices P and $\widetilde{P}$ can be also obtained in explicit form by making use of the special block structure of the matrices E, A, $\widetilde{E}$ and $\widetilde{A}$.

The algorithm for choosing $\{\alpha_j\}_{j=1}^k$ and $\{\beta_j\}_{j=1}^k$ is summarized in Algorithm 2. For more details about this algorithm, the interesting reader is referred to [8].

Algorithm 2. Choose ADI parameters

Input: $E, A \in \mathbb{R}^{n\times n}$, $\widetilde{E}, \widetilde{A} \in \mathbb{R}^{m\times m}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^m$ with $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ being c-stable.

Output: ADI parameters $\{\alpha_1, \alpha_2, \ldots, \alpha_i\}$ and $\{\beta_1, \beta_2, \ldots, \beta_i\}$.

1. Run the Arnoldi process with $A^{-1}E$ on b to obtain the set $\mathbb{E}_A^+$ of Ritz values.
2. Run the Arnoldi process with PA on b to get the set $\mathbb{E}_A^-$ of Ritz values.
3. Set $\mathbb{E} = \mathbb{E}_A^+ \cup (1/\mathbb{E}_A^-)$.
4. Run the Arnoldi process with $\widetilde{A}^{-1}\widetilde{E}$ on c to obtain the set $\mathbb{F}_B^+$ of Ritz values.
5. Run the Arnoldi process with $\widetilde{P}\widetilde{A}$ on c to obtain the set $\mathbb{F}_B^-$ of Ritz values.
6. Set $\mathbb{F} = (\mathbb{F}_B^+) \cup (1/\mathbb{F}_B^-)$.
7. Set $\{\alpha_1, \beta_1\} = \arg \min\limits_{\substack{\alpha\in\mathbb{E}\\ \beta\in\mathbb{F}}} \max\limits_{\substack{x\in\mathbb{E}\\ y\in\mathbb{F}}} \left|\frac{(x-\alpha)(y-\beta)}{(x+\beta)(y+\alpha)}\right|$,
8. For $i = 2, 3, \cdots, k$

   Set $\{\alpha_i, \beta_i\} = \arg \min\limits_{\substack{\alpha\in\mathbb{E}'\\ \beta\in\mathbb{F}'}} \max\limits_{\substack{x\in\mathbb{E}\\ y\in\mathbb{F}}} \left|\frac{(x-\alpha)(y-\beta)}{(x+\beta)(y+\alpha)}\right| \prod_{j=1}^{i-1}\left|\frac{(x-\alpha_j)(y-\beta_j)}{(x+\beta_j)(y+\alpha_j)}\right|$,

   where $\mathbb{E}'$ is $\mathbb{E}$ with $\alpha_1, \ldots, \alpha_{i-1}$ deleted, and similarly for $\mathbb{F}'$,

   End For

3.2 The Smith method

The Smith method is derived from the projected discrete-time Sylvester equation

$$\begin{cases} \mathcal{A}X\widetilde{\mathcal{A}} - X - (\alpha+\beta)P_r\mathcal{B}C\widetilde{P}_l = 0, \\ X = P_rX\widetilde{P}_l, \end{cases} \tag{21}$$

where

$$\begin{aligned} &\mathcal{A} = (E+\beta A)^{-1}(E-\alpha A), \\ &\widetilde{\mathcal{A}} = (\widetilde{E}-\beta\widetilde{A})(\widetilde{E}+\alpha\widetilde{A})^{-1}, \\ &\mathcal{B} = (E+\beta A)^{-1}B, \quad C = C(\widetilde{E}+\alpha\widetilde{A})^{-1}, \end{aligned}$$

The equation (21) is equivalent to the PGCTSE (1) with $F = BC$ for any two parameters $\alpha, \beta \in \mathbb{C}^-$.

From (21), we obtain the Smith iteration

$$\begin{cases} X_0 = P_r \mathcal{B} C \widetilde{P}_l, \\ X_i = \mathcal{A} X_{i-1} \widetilde{\mathcal{A}} - (\alpha + \beta) P_r \mathcal{B} C \widetilde{P}_l. \end{cases}$$

If all of the shifts in the ADI iteration (15) are constant, i.e., $\alpha_i = \alpha$, $\beta_i = \beta$ $(i = 1, 2, \cdots)$, then the ADI iteration reduces to the Smith method. In [21], Penzl illustrated that the ADI iteration with a single shift (Smith method) converges very slowly, while a moderate increase in the number of shifts s accelerates the convergence nicely. However, it is also observed that the speed of convergence is hardly improved by a further increase of s; see Table 2.1 in [21]. These observations lead to the idea of the cyclic Smith(s) iteration, a special case of ADI where s different shifts are used in a cyclic manner.

The low-rank scheme based on the Smith(s) iteration was also introduced in [21]. The method is called the cyclic low-rank Smith method (LR-Smith(s)) and is a special case of LR-ADI, where s shifts are reused in a cyclic manner. This idea can be generalized for the equation (1). The generalized LR-Smith method consists of two steps. First the matrices Y_s, H_s and Z_s are obtained by an s step generalized LR-ADI iteration. Then one solves the discrete-time Sylvester equation

$$\mathcal{A}_s X \mathcal{B}_s - X + U_s H_s V_s = 0,$$

where $\mathcal{A}_s$ and $\mathcal{B}_s$ is as in (17) and (18) with $i = s$, respectively.

The generalized LR-Smith method for the PGCTSE (1) with $F = BC$ is described in Algorithm 3.

Algorithm 3. The LR-Smith(s) method

Input: $E, A \in \mathbb{R}^{n\times n}$, $\widetilde{E}, \widetilde{A} \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times r}$ and $C \in \mathbb{R}^{r\times m}$ with $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ being c-stable.

Output: U, H and V such that $Y = UHV$ is an approximate solution of the PGCTSE (1) with $F = BC$.

1. Compute U_s, H_s and V_s by using Algorithm 1 with $k = s$;
2. Set $U = \bar{U} = U_s$, $H = \bar{H} = H_s$ and $V = \bar{V} = V_s$;
3. For $i = 1, 2, \cdots$
   For $j = 1, 2, \cdots, s$
   $\bar{U} = (E + \beta_j A)^{-1}(E - \alpha_j A)\bar{U}$;
   $\bar{V} = \bar{V}(\widetilde{E} - \beta_j \widetilde{A})(\widetilde{E} + \alpha_j \widetilde{A})^{-1}$;
   End For
   $U = [U, P_r \bar{U}]$;
   $V = \begin{bmatrix} V \\ \bar{V}\widetilde{P}_l \end{bmatrix}$;
   $H = \mathrm{diag}(H, \bar{H})$;
   End For

4 Numerical method for PGDTSE

In this section we present a numerical method for solving the PGDTSE (2) with $F = BC$.

From (7) and $F = BC$, it follows that the solution Y can be expressed as

$$Y = \sum_{i=0}^{\min\{\nu,\tilde{\nu}\}-1} (A^{-1}E)^i A^{-1}(I - P_l)BC(I - \widetilde{P}_r)\widetilde{A}^{-1}(\widetilde{E}\widetilde{A}^{-1})^i.$$

Thus, Y can be reformulated in terms of the low-rank factors as

$$Y = UV,$$

where

$$\begin{aligned} U &= [A^{-1}(I - P_l)B, A^{-1}EA^{-1}(I - P_l)B, \\ &\quad \cdots, (A^{-1}E)^{\min\{\nu,\tilde{\nu}\}-1}A^{-1}(I - P_l)B], \\ V &= \begin{bmatrix} C(I - \widetilde{P}_r)\widetilde{A}^{-1} \\ C(I - \widetilde{P}_r)\widetilde{A}^{-1}\widetilde{E}\widetilde{A}^{-1} \\ \vdots \\ C(I - \widetilde{P}_r)\widetilde{A}^{-1}(\widetilde{E}\widetilde{A}^{-1})^{\min\{\nu,\tilde{\nu}\}-1} \end{bmatrix}. \end{aligned}$$

Therefore, we obtain a method for solving the PGDTSE as described in Algorithm 4.

Algorithm 4. Numerical method for PGDTSE

Input: $E, A \in \mathbb{R}^{n\times n}$, $\widetilde{E}, \widetilde{A} \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times r}$ and $C \in \mathbb{R}^{r\times m}$ with $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ being c-stable.

Output: U and V such that $Y = UV$ is a solution of the PGDTSE (2) with $F = BC$.

1. Compute $U_1 = A^{-1}(I - P_l)B$ and $V_1 = C(I - \widetilde{P}_r)\widetilde{A}^{-1}$;
2. Set $U = U_1$ and $V = V_1$;
3. For $i = 1, \cdots, \min\{\nu, \tilde{\nu}\} - 1$
   $U_{i+1} = A^{-1}EU_i$;
   $V_{i+1} = V_i\widetilde{E}\widetilde{A}^{-1}$;
   $U = [U, U_{i+1}]$;
   $V = \begin{bmatrix} V \\ V_{i+1} \end{bmatrix}$;
   End For

5 Applications

In this section, we show that the PGCTSE (1) and the PGDTSE (2) play an important role in computing the $\mathbb{HL}_2$ inner production of two c-stable descriptor systems.

Let $H(s) = C(sE - A)^{-1}B$ and $\widetilde{H}(s) = \widetilde{C}(s\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}$ be two c-stable systems. As shown in [27], $H(s)$ and $\widetilde{H}(s)$ can be decomposed into

$$H(s) = H_{sp}(s) + P(s), \quad \widetilde{H}(s) = \widetilde{H}_{sp}(s) + \widetilde{P}(s),$$

where

$$\begin{aligned} H_{sp}(s) &= CP_r(sE - A)^{-1}P_lB, \\ P(s) &= C(I - P_r)(sE - A)^{-1}(I - P_l)B, \\ \widetilde{H}_{sp}(s) &= \widetilde{C}\widetilde{P}_r(s\widetilde{E} - \widetilde{A})^{-1}\widetilde{P}_l\widetilde{B}, \\ \widetilde{P}(s) &= \widetilde{C}(I - \widetilde{P}_r)(s\widetilde{E} - \widetilde{A})^{-1}(I - \widetilde{P}_l)\widetilde{B}. \end{aligned}$$

Here, $H_{sp}(s), \widetilde{H}_{sp}(s)$ are called the strictly proper parts of $H(s), \widetilde{H}(s)$, while $P(s), \widetilde{P}(s)$ are called the polynomial parts of $H(s), \widetilde{H}(s)$, respectively.

Define the $\mathbb{HL}_2$ inner product of $H(s)$ and $\widetilde{H}(s)$ by

$$\begin{aligned} &\langle H(s), \widetilde{H}(s)\rangle_{\mathbb{HL}_2} = \langle H_{sp}(s), \widetilde{H}_{sp}(s)\rangle_{\mathbb{H}_2} + \langle P(s), \widetilde{P}(s)\rangle_{\mathbb{L}_2} \\ = \ & \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{trace}\left(\overline{H_{sp}(i\omega)}\widetilde{H}_{sp}^T(i\omega)\right)d\omega \\ & + \frac{1}{2\pi}\int_0^{2\pi} \mathrm{trace}\left(\overline{P(e^{i\omega})}\widetilde{P}^T(e^{i\omega})\right)d\omega, \end{aligned}$$

where the notation trace$(\cdot)$ denotes the trace of a matrix. Then, the $\mathbb{HL}_2$ norm of $H(s)$ induced by the $\mathbb{HL}_2$ inner product is

$$\begin{aligned} \|H(s)\|_{\mathbb{HL}_2} &= \sqrt{\|H_{sp}(s)\|_{\mathbb{H}_2}^2 + \|P(s)\|_{\mathbb{L}_2}^2} \\ &= \sqrt{\langle H_{sp}(s), H_{sp}(s)\rangle_{\mathbb{H}_2} + \langle P(s), P(s)\rangle_{\mathbb{L}_2}}. \end{aligned}$$

The following theorem establishes a connection between the $\mathbb{HL}_2$ inner product of two c-stable systems and the solutions of two projected generalized Sylvester equations.

Theorem 4 *Assume that $H(s) = C(sE - A)^{-1}B$ and $\widetilde{H}(s) = \widetilde{C}(s\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}$ are two c-stable systems. Let X be the solution of the PGCTSE*

$$AX\widetilde{E}^T + EX\widetilde{A}^T + P_lB\widetilde{B}^T\widetilde{P}_l^T = 0, \ X = P_rX\widetilde{P}_r^T \quad (22)$$

and Y be the solution of the PGDTSE

$$AY\widetilde{A}^T - EY\widetilde{E}^T = (I - P_l)B\widetilde{B}^T(I - \widetilde{P}_l)^T, \ P_rY\widetilde{P}_r^T = 0. \quad (23)$$

Then, we have

(1) $\langle H_{sp}(s), \widetilde{H}_{sp}(s)\rangle_{\mathbb{H}_2} = \mathrm{trace}(CX\widetilde{C}^T)$;

(2) $\langle P(s), \widetilde{P}(s)\rangle_{\mathbb{L}_2} = \mathrm{trace}(CY\widetilde{C}^T)$;

(3) $\langle H(s), \widetilde{H}(s)\rangle_{\mathbb{HL}_2} = \mathrm{trace}(C(X + Y)\widetilde{C}^T)$.

Proof: Following the similar arguments as in [18] and Section 2, we can show that the solutions of (22) and (23) can be expressed as

$$X = \frac{1}{2\pi}\int_{-\infty}^{\infty}(-i\omega E - A)^{-1}P_lB\widetilde{B}^T\widetilde{P}_l^T(i\omega\widetilde{E}^T - \widetilde{A}^T)^{-1}d\omega$$

and

$$\begin{aligned} Y &= \frac{1}{2\pi}\int_0^{2\pi}(e^{-i\omega}E - A)^{-1}(I - P_l)B\widetilde{B}^T \\ &\quad \cdot(I - \widetilde{P}_l)^T(e^{i\omega}\widetilde{E} - \widetilde{A})^{-T}d\omega. \end{aligned}$$

Thus, we have

$$\begin{aligned} &\langle H_{sp}(s), \widetilde{H}_{sp}(s)\rangle_{\mathbb{H}_2} \\ = \ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{trace}\left(\overline{H_{sp}(i\omega)}\widetilde{H}_{sp}^T(i\omega)\right)d\omega \\ = \ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{trace}\Big(C(-i\omega E - A)^{-1}P_lB \\ & \cdot \widetilde{B}^T\widetilde{P}_l^T(i\omega\widetilde{E}^T - \widetilde{A}^T)^{-1}\widetilde{C}^T\Big)d\omega \\ = \ & \mathrm{trace}(CX\widetilde{C}^T), \end{aligned}$$

$$\begin{aligned} &\langle P(s), \widetilde{P}(s)\rangle_{\mathbb{L}_2} \\ = \ & \frac{1}{2\pi}\int_0^{2\pi}\mathrm{trace}\left(P(e^{-i\omega})\widetilde{P}^T(e^{i\omega})\right)d\omega \\ = \ & \frac{1}{2\pi}\int_0^{2\pi}\mathrm{trace}\Big(C(e^{-i\omega}E - A)^{-1}(I - P_l) \\ & \cdot B\widetilde{B}^T(I - \widetilde{P}_l)^T(e^{i\omega}\widetilde{E} - \widetilde{A})^{-T}\widetilde{C}^T\Big)d\omega \\ = \ & \mathrm{trace}(CY\widetilde{C}^T). \end{aligned}$$

□

Similarly to Theorem 4, we easily obtain the following result.

Theorem 5 *Let $H(s) = C(sE - A)^{-1}B$ and $\widetilde{H}(s) = \widetilde{C}(s\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}$ be two c-stable systems. Let X be the solution of the PGCTSE*

$$\widetilde{E}^TXA + \widetilde{A}^TXE + \widetilde{P}_r^T\widetilde{C}^TCP_r = 0, \ X = \widetilde{P}_l^TXP_l \quad (24)$$

and Y be the solution of the PGDTSE

$$\widetilde{A}^TYA - \widetilde{E}^TYE = (I - \widetilde{P}_r)^T\widetilde{C}^TC(I - P_r), \ \widetilde{P}_l^TYP_l = 0. \quad (25)$$

Then, we have

(1) $\langle H_{sp}(s), \widetilde{H}_{sp}(s)\rangle_{\mathbb{H}_2} = \text{trace}(\widetilde{B}^T X B)$;

(2) $\langle P(s), \widetilde{P}(s)\rangle_{\mathbb{L}_2} = \text{trace}(\widetilde{B}^T Y B)$;

(3) $\langle H(s), \widetilde{H}(s)\rangle_{\mathbb{HL}_2} = \text{trace}(\widetilde{B}^T (X+Y) B)$.

As a consequence of Theorem 4 and Theorem 5, we obtain the following result, which is also given in [29].

Corollary 6 *Assume that $H(s) = C(sE - A)^{-1}B$ is a c-stable system. Let $\mathcal{G}_{pc}, \mathcal{G}_{po}$ be the proper controllability and observability Gramians, and $\mathcal{G}_{ic}, \mathcal{G}_{io}$ the improper controllability and observability Gramians, i.e., $\mathcal{G}_{pc}, \mathcal{G}_{po}$ are the unique symmetric, positive-semidefinite solutions of the PGCTLEs*

$$\begin{aligned} &A\mathcal{G}_{pc}E^T + E\mathcal{G}_{pc}A^T + P_l B B^T P_l^T = 0, \\ &\mathcal{G}_{pc} = P_r \mathcal{G}_{pc} P_r^T, \end{aligned} \tag{26}$$

$$\begin{aligned} &A^T\mathcal{G}_{po}E + E^T\mathcal{G}_{po}A + P_r^T C^T C P_r = 0, \\ &\mathcal{G}_{po} = P_l^T \mathcal{G}_{po} P_l, \end{aligned} \tag{27}$$

and $\mathcal{G}_{ic}, \mathcal{G}_{io}$ are the unique symmetric, positive-semidefinite solutions of the PGDTLEs

$$\begin{aligned} &A\mathcal{G}_{ic}A^T - E\mathcal{G}_{ic}E^T = (I - P_l)BB^T(I - P_l)^T, \\ &P_r \mathcal{G}_{ic} P_r^T = 0, \end{aligned} \tag{28}$$

$$\begin{aligned} &A^T\mathcal{G}_{io}A - E^T\mathcal{G}_{io}E = (I - P_r)^T C^T C (I - P_r), \\ &P_l^T \mathcal{G}_{io} P_l = 0. \end{aligned} \tag{29}$$

Then, we have

(1) $\|H_{sp}(s)\|_{\mathbb{H}_2} = \sqrt{\text{trace}(C\mathcal{G}_{pc}C^T)} = \sqrt{\text{trace}(B^T\mathcal{G}_{po}B)}$;

(2) $\|P(s)\|_{\mathbb{L}_2} = \sqrt{\text{trace}(C\mathcal{G}_{ic}C^T)} = \sqrt{\text{trace}(B^T\mathcal{G}_{io}B)}$;

(3) $\|H(s)\|_{\mathbb{HL}_2} = \sqrt{\text{trace}(C(\mathcal{G}_{pc} + \mathcal{G}_{ic})C^T)} = \sqrt{\text{trace}(B^T(\mathcal{G}_{po} + \mathcal{G}_{io})B)}$.

We now consider the $\mathbb{HL}_2$ inner product of two c-stable single input single output (SISO) systems.

Theorem 7 *Assume that $H(s) = C(sE - A)^{-1}B$ and $\widetilde{H}(s) = \widetilde{C}(s\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}$ are two c-stable SISO systems. Let X be the solution of the PGCTSE*

$$AX\widetilde{E} + EX\widetilde{A} + P_l B\widetilde{C}\widetilde{P}_r = 0, \quad X = P_r X \widetilde{P}_l \tag{30}$$

and Y be the solution of the PGDTSE

$$AY\widetilde{A} - EY\widetilde{E} = (I - P_l)B\widetilde{C}(I - \widetilde{P}_r), \quad P_r Y \widetilde{P}_l = 0. \tag{31}$$

Then, we have

(1) $\langle H_{sp}(s), \widetilde{H}_{sp}(s)\rangle_{\mathbb{H}_2} = \text{trace}(CX\widetilde{B})$;

(2) $\langle P(s), \widetilde{P}(s)\rangle_{\mathbb{L}_2} = \text{trace}(CY\widetilde{B})$;

(3) $\langle H(s), \widetilde{H}(s)\rangle_{\mathbb{HL}_2} = \text{trace}(C(X+Y)\widetilde{B})$.

Proof: Since $H(s)$ and $\widetilde{H}(s)$ are SISO systems, it follows that

$$\begin{aligned} & G_{sp}(-i\omega)\widetilde{G}_{sp}^T(i\omega) \\ = \; & G_{sp}(-i\omega)\widetilde{G}_{sp}(i\omega) \\ = \; & C(-i\omega E - A)^{-1} P_l B \widetilde{C}(i\omega\widetilde{E} - \widetilde{A})^{-1}\widetilde{P}_l\widetilde{B} \\ = \; & C(-i\omega E - A)^{-1} P_l B \widetilde{C}\widetilde{P}_r(i\omega\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}, \end{aligned}$$

$$\begin{aligned} & P(e^{-i\omega})\widetilde{P}^T(e^{i\omega}) \\ = \; & P(e^{-i\omega})\widetilde{P}(e^{i\omega}) \\ = \; & C(e^{-i\omega}E - A)^{-1}(I - P_l)B\widetilde{C}(e^{i\omega}\widetilde{E} - \widetilde{A})^{-1}(I - \widetilde{P}_l)\widetilde{B} \\ = \; & C(e^{-i\omega}E - A)^{-1}(I - P_l)B\widetilde{C}(I - \widetilde{P}_r)(e^{i\omega}\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}. \end{aligned}$$

Then, we can prove the theorem by following the same arguments as the proof of Theorem 4. □

The following result, which results directly from (7), shows that the $\mathbb{HL}_2$ norm of a c-stable SISO system can be formulated by the solutions of a PGCTSE and a PGDTSE.

Corollary 8 *Let $H(s) = C(sE - A)^{-1}B$ be a c-stable SISO system. Let X be the solution of the PGCTSE*

$$AXE + EXA + P_l B C P_r = 0, \quad X = P_r X P_l$$

and Y be the solution of the PGDTSE

$$AYA - EYE = (I - P_l)BC(I - P_r), \quad P_r Y P_l = 0.$$

Then, we have

(1) $\|H_{sp}(s)\|_{\mathbb{H}_2} = \sqrt{CXB}$;

(2) $\|P(s)\|_{\mathbb{L}_2} = \sqrt{CYB}$;

(3) $\|H(s)\|_{\mathbb{HL}_2} = \sqrt{C(X+Y)B}$.

6 Numerical experiments

In this section, we present some numerical examples to illustrate the performance of the LR-ADI method and the LR-Smith(s) method for the PGCTSE (1). In the following examples, we compare the numerical behavior of these two methods with respect to the number of iterations (ITERs), the CPU time (in seconds), and the relative residual (RES). Here the relative residual is defined by

$$\text{RES} \equiv \frac{\|AX_i\widetilde{E} + EX_i\widetilde{A} + P_l F \widetilde{P}_r\|_F}{\|P_l F \widetilde{P}_r\|_F},$$

Table 1: MVP, LSS, and LU at every iteration step

	LR-ADI	LR-Smith(s)
MVP for $E-\alpha A$	r	rs^2
MVP for $(\widetilde{E}-\beta\widetilde{A})^T$	r	rs^2
LSS for $E+\beta A$	r	rs^2
LSS for $(\widetilde{E}+\alpha\widetilde{A})^T$	r	rs^2
LU for $E+\beta A$	1	s
LU for $(\widetilde{E}+\alpha\widetilde{A})^T$	1	s

where X_i denotes the approximate solution obtained by the LR-ADI method or the LR-Smith method.

Let MVP denote the number of matrix vector products, LSS denote the number of linear system solvers, and LU the number of the LU decomposition of matrices at every iteration for these two methods. Table 1 has a rough count of the expenses of LR-ADI and LR-Smith(s) at every iteration. Only the major expenses are considered. From Table 1, we can see that LR-Smith(s) needs more MVP, LSS, and LU than LR-ADI at every iteration, so LR-ADI is more efficient than LR-Smith(s) for the computation cost at every iteration.

All numerical experiments are performed on an Intel Pentium Dual E2160 with CPU 1.80GHz and RAM 2GB under the Window XP operating system and the usual double precision, where the floating point relative accuracy is 2.22×10^{-16}.

6.1 Example 1

For the first experiment, we consider the 2D instationary Stokes equation that describes the flow of an incompressible fluid in a domain. We use the spatial discretization of this equation by the finite difference method on a uniform staggered grid to generate two descriptor systems

$$\begin{aligned} E\dot{x}(t) &= Ax(t)+Bu(t),\\ y(t) &= Cx(t), \end{aligned} \tag{32}$$

and

$$\begin{aligned} \widetilde{E}\dot{x}(t) &= \widetilde{A}x(t)+\widetilde{B}u(t),\\ y(t) &= \widetilde{C}x(t). \end{aligned}$$

The matrices in (32) are given by

$$E=\begin{bmatrix} I & 0\\ 0 & 0\end{bmatrix}\in\mathbb{R}^{n\times n},\quad A=\begin{bmatrix} A_{11} & A_{12}\\ A_{21} & 0\end{bmatrix}\in\mathbb{R}^{n\times n}.$$

These matrices are sparse and have a special block structure. Using this structure, the projectors P_l and P_r onto the left and right deflating subspaces of the pencil $\lambda E-A$ can be computed as

$$P_l=\begin{bmatrix}\Pi & -\Pi A_{11}A_{12}(A_{21}A_{12})^{-1}\\ 0 & 0\end{bmatrix},$$

$$P_r=\begin{bmatrix}\Pi & 0\\ -(A_{21}A_{12})^{-1}A_{21}A_{11}\Pi & 0\end{bmatrix},$$

where $\Pi = I-A_{12}(A_{21}A_{12})^{-1}A_{21}$ is the orthogonal projector onto the kernel of A_{21} along the image of A_{12}, see [28]. The matrices A_{12} and A_{21} have full rank and the pencil $\lambda E-A$ is of index 2. The pencil $\lambda\widetilde{E}-\widetilde{A}$ has the same structure as $\lambda E-A$. In our experiment the state space dimensions of the problems are $n=2132$ and $m=1280$, respectively. The matrix F in (1) is $F=B\widetilde{C}$ with $B\in\mathbb{R}^n$ and $\widetilde{C}^T\in\mathbb{R}^m$.

Table 2: LR-ADI, $k=15$ vs. LR-Smith(5)

	LR-ADI	LR-Smith(5)
ITERs	40	11
MVP for $E-\alpha A$	40	275
MVP for $(\widetilde{E}-\beta\widetilde{A})^T$	40	275
LSS for $E+\beta A$	40	275
LSS for $(\widetilde{E}+\alpha\widetilde{A})^T$	40	275
LU for $E+\beta A$	40	55
LU for $(\widetilde{E}+\alpha\widetilde{A})^T$	40	55
CPU	25.48	37.21
RES	8.1879e-13	5.9044e-14

The numerical results are presented in Table 2. We point out that the main computational cost is attributed to the LU decomposition of matrices. So, the CPU time for the LR-Smith(5) method is not much more than that required by the LR-ADI method with $k=15$. The convergence curves for these two methods are exhibited Figure 1.

We also computed the solution of the PGDTSE (2) with $F=B\widetilde{C}$. Since the indices of the pencils in this example is 2, two iteration step of Algorithm 4 is required. It costs only 2.41 seconds and obtains the relative residual 2.2104e-15.

6.2 Example 2

For the second experiment, we consider a holonomically constrained damped mass-spring system with g masses as in [30]. The vibration of this system is described by the descriptor system (32) with the matrices

$$E=\begin{bmatrix} I & 0 & 0\\ 0 & M & 0\\ 0 & 0 & 0\end{bmatrix},\quad A=\begin{bmatrix} 0 & I & 0\\ K & D & -N^T\\ N & 0 & 0\end{bmatrix},$$

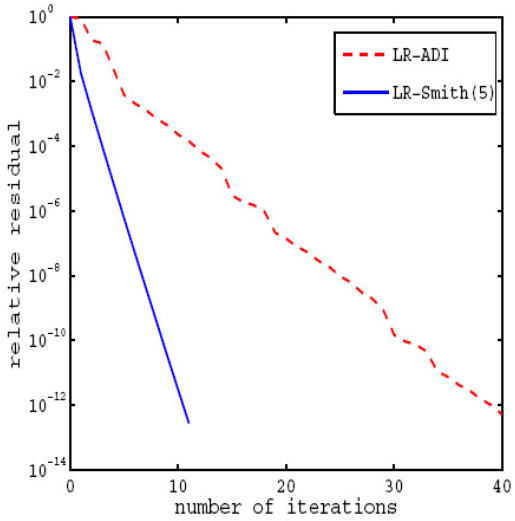

Figure 1: Convergence curves of LR-ADI and LR-Smith(5) for Example 1.

where $M = \mathrm{diag}(m_1, m_2, \cdots, m_g)$ is the symmetric positive definite mass matrix, $K \in \mathbb{R}^{g\times g}$ is the tridiagonal stiffness matrix, $D \in \mathbb{R}^{g\times g}$ is the tridiagonal damping matrix, and N is the matrix of constraints. The spectral projectors P_l and P_r can be expressed as

$$P_l = \begin{bmatrix} \Pi & 0 & -\Pi M^{-1}DQ \\ -\Pi^T D(I-\Pi) & \Pi^T & -\Pi^T(K+D\Pi M^{-1}D)Q \\ 0 & 0 & 0 \end{bmatrix},$$

$$P_r = \begin{bmatrix} \Pi & 0 & 0 \\ -\Pi M^{-1}D(I-\Pi) & \Pi & 0 \\ Q^T(K\Pi - D\Pi M^{-1}D(I-\Pi)) & Q^T D\Pi & 0 \end{bmatrix}.$$

Here $Q = M^{-1}N^T(NM^{-1}N^T)^{-1}$ and $\Pi = I - M^{-1}N^T(NM^{-1}N^T)^{-1}N = I - QN$ is a projector onto the kernel of N along the image of $M^{-1}N^T$. If N has full row rank, the the pencil $\lambda E - A$ is of index 3. We use the same method to generate the pencil $\lambda\widetilde{E} - \widetilde{A}$. In this experiment the state space dimensions of the problems are $n = 1261$ and $m = 1161$, respectively.

The computational results were reported in Figure 2 and Table 3. We note that the iteration steps of the LR-Smith(5) method is 17 while the steps of LR-ADI is 39, but the CPU time of LR-Smith(5) method is much more than that of LR-ADI method. This is because the computation cost at every iteration of two methods is very different.

7 Conclusions

In this paper, we have reviewed the low-rank alternating direction implicit method and the low-rank cyclic Smith method for solving the PGCTSE with low-rank right-hand sides. We have proposed a numerical method for the PGDTSE. Moreover, we show that these two equations are useful for computing the $\mathbb{HL}_2$ inner product of two c-stable descriptor systems. Numerical experiments presented in this paper show the effectiveness of these methods.

Table 3: LR-ADI, $k = 15$ vs. LR-Smith(5)

	LR-ADI	LR-Smith(5)
ITERs	39	17
MVP for $E - \alpha A$	39	425
MVP for $(\widetilde{E} - \beta\widetilde{A})^T$	39	425
LSS for $E + \beta A$	39	425
LSS for $(\widetilde{E} + \alpha\widetilde{A})^T$	39	425
LU for $E + \beta A$	39	85
LU for $(\widetilde{E} + \alpha\widetilde{A})^T$	39	85
CPU	41.67	112.86
RES	7.3098e-11	1.4239e-10

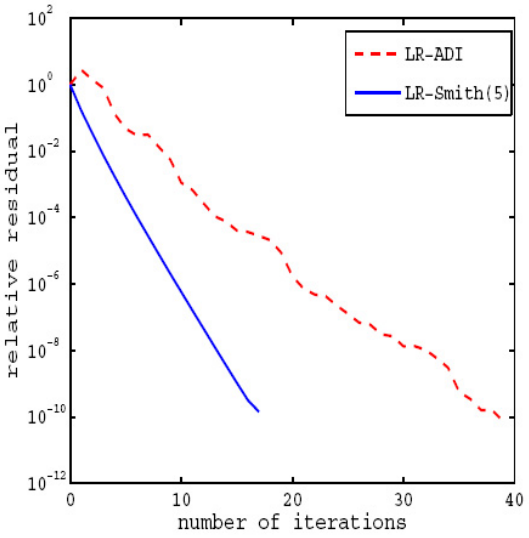

Figure 2: Convergence curves of LR-ADI and LR-Smith(5) for Example 2.

Acknowledgements: The authors would like to thank Prof. Genqi Xu and three anonymous referees for their helpful suggestions, which greatly improved the paper. Yiqin Lin is supported by the Academic Leader Training Plan of Hunan Province in University and the Talent Support Plan of Hunan University of Science and Engineering. Liang Bao is supported by the National Natural Science Foundation of China (grants No. 10926150 and No. 11101149).

References:

[1] B. Anderson and J. Moore, *Optimal Control–Linear Quadratic Methods*, Prentice–Hall, Englewood Cliffs, NJ 1990.

[2] S. Balaji, Solution of nonlinear Riccati differential equation using Chebyshev wavelets, *WSEAS Transactions on Mathematics* 13, 2014, pp. 441–451.

[3] R. Bartels and G. Stewart, Solution of the equation $AX+XB$ =, *Comm. ACM* 15, 1972, pp. 820–826.

[4] U. Baur, Low rank solution of data-sparse Sylvester equations, *Numer. Lin. Alg. Appl.* 15, 2008, pp. 837–851.

[5] P. Benner, *Contributions to the Numerical Solution of Algebraic Riccati Equations and Related Eigenvalue Problems*, Logos Verlag, Berlin, 1997.

[6] P. Benner, Factorized solution of Sylvester equations with applications in control, *Proc. Intl. Symp. Math. Theory Networks and Syst*, MTNS, 2004.

[7] P. Benner and J.-R. Li, and T. Penzl, Numerical solution of large Lyapunov equations, Riccati equations, and linear-quadratic control problems, *Numer. Lin. Alg. Appl.* 15, 2008, pp. 755–777.

[8] P. Benner, R. Li and N. Truhar, On the ADI method for Sylvester equations, *J. Comput. Appl. Math.* 233, 2009, pp. 1035–1045.

[9] P. Benner and E. Quintana-Ortí, Solving stable generalized Lyapunov equations with the matrix sign function, *Numer. Algorithms* 20, 1999, pp. 75–100.

[10] P. Benner, E. S. Quintana-Ortí and G. Quintana-Ortí, Solving stable Sylvester equations via rational iterative schemes, *J. Sci. Comput.* 28, 2006, pp. 51–83.

[11] F. Gantmacher, *Theory of Matrices*, Chelsea, New York 1959.

[12] S. Gugercin, D. C. Sorensen and A. C. Antoulas, A modified low-rank Smith method for large-scale Lyapunov equations, *Numer. Algorithms* 32, 2003, pp. 27–55.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, MD 1996.

[14] S. J. Hammarling, Numerical solution of the stable non-negative definite Lyapunov equation, *IMA J. Numer. Anal.* 2, 1982, pp. 303–323.

[15] U. Helmke and J. Moore, *Optimization and Dynamical Systems*, Springer, London 1994.

[16] C. Kenney and A. Laub, The matrix sign function, *IEEE Trans. Automat. Control* 40, 1995, pp. 1330-1348.

[17] J. Li and J. White, Low rank solution of Lyapunov equations, *SIAM J. Matrix Anal. Appl.* 24, 2002, pp. 260–280.

[18] Y. Lin, L. Bao, and Y. Wei, Matrix sign function methods for solving projected generalized continuous-time Sylvester equations, *IEEE Trans. Automat. Control* 55, 2010, pp. 2629–2634.

[19] A. Lu and E. Wachspress, Solution of Lyapunov equations by alternating direction implicit iteration, *Comput. Math. Appl.* 21, 1991, pp. 43–58.

[20] V. Mehrmann, *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solution*, Springer, Heidelberg 1991.

[21] T. Penzl, A cyclic low-rank Smith method for large sparse Lyapunov equations, *SIAM J. Sci. Comput.* 21, 2000, pp. 1401–1418.

[22] P. Petkov, N. Christov and M. Konstantinov, *Computational Methods for Linear Control Systems*, Prentice-Hall, Hertfordshire, UK 1991.

[23] J. Sabino, *Solution of Large-Scale Lyapunov Equations via the Block Modified Smith Method*, Rice University, PhD Thesis, Houston, Texas 2006.

[24] R. Smith, Matrix equation $XA + BX = C$, *SIAM J. Appl. Math.* 16, 1968, pp. 198–201.

[25] T. Stykel, Numerical solution and perturbation theory for generalized Lyapunov equations, *Linear Algebra Appl.* 349, 2002, pp. 155–185.

[26] T. Stykel, Stability and inertia theorems for generalized Lyapunov equations, *Linear Algebra Appl.* 355, 2002, pp. 297–314.

[27] T. Stykel, Gramian-based model reduction for descriptor systems, *Math. Control Signals Systems* 16, 2004, pp. 297–319.

[28] T. Stykel, Balanced truncation model reduction for semidiscretized Stokes equation, *Linear Algebra Appl.* 415, 2006, pp. 262–289.

[29] T. Stykel, On some norms for descriptor systems, *IEEE Trans. Automat. Control* 51, 2006, pp. 842–847.

[30] T. Stykel, A modified matrix sign function method for projected Lyapunov equations, *Systems Control Lett.* 56, 2007, pp. 695–701.

[31] T. Stykel, Low-rank iterative methods for projected generalized Lyapunov equations, *Elect. Trans. Numer. Anal.* 30, 2008, pp. 187–202.

[32] E. Wachspress, Iterative solution of the Lyapunov matrix equation, *Appl. Math. Lett.* 107, 1988, pp. 87–90.

[33] B. Zheng, The Riccati sub-ode method for fractional differential-difference equations, *WSEAS Transactions on Mathematics* 13, 2014, pp. 192–200.

[34] Y. Zhou, L. Bao, and Y. Lin, Krylov subspace type methods for solving projected generalized continuous-time Lyapunov equations, *WSEAS Transactions on Mathematics* 11, 2012, pp. 1114–1126.

[35] Y. Zhou, L. Bao, and Y. Lin, An alternating direction implicit method for solving projected generalized continuous-time Sylvester equations, *Math. Problems Eng.* 13, 2013, pp. 1–8.