2010 American Control Conference
Marriott Waterfront, Baltimore, MD, USA
June 30-July 02, 2010

Krylov Subspace Restart Scheme for Solving Large-Scale Sylvester Equations

Mian Ilyas Ahmad, Imad Jaimoukha and Michalis Frangos

Abstract— **The development of efficient interior point methods has greatly enlarged the range of control problems with feasible numerical solution. These methods are nevertheless difficult to solve for large-scale problems. In this paper we suggest the use of a Krylov subspace technique for the efficient low-rank approximate solution to large-scale Sylvester equations. The suggested method is a novel restart scheme which improves further the computation efficiency and storage requirements of the standard Krylov subspace methods for the solution of large-scale Sylvester equations.**

I. Introduction

In this paper, we present a Krylov subspace method for the solution of the Sylvester equation,

$$AP + P\hat{A}^T + B\hat{B}^T = 0, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $\hat{A} \in \mathbb{R}^{\hat{n}\times\hat{n}}$, $B \in \mathbb{R}^n$, $\hat{B} \in \mathbb{R}^{\hat{n}}$ and $P \in \mathbb{R}^{n\times\hat{n}}$. It is assumed that A and $\hat{A}$ are large and sparse. The results presented in this paper are also valid for the solution to large-scale Lyapunov equations which is the case where $\hat{A}=A$ and $\hat{B}=B$.

The Sylvester and Lyapunov equations play an important role in control theory, signal processing, model reduction, filtering and image restoration; see, for example, [5], [4], [17] and the references therein. Existing direct methods for solving the Sylvester equation (e.g. the Bartels Stewart method [3] and the Hammerling Method [14]) are not suitable for sufficiently large values of n or $\hat{n}$ due to their computational complexity. To overcome the computational complexity of the direct methods, iterative algorithms, based on Krylov subspace techniques, have been proposed in the literature, e.g., [15], [13], [21], [16]. Furthermore, the Krylov subspace methods make use of sparsity which is present in many large-scale problems (sparsity for example is a natural result of the discretizing methods of partial differential equations).

Krylov subspace projection methods are based on obtaining low-rank approximate solutions to the large-scale Sylvester equation, such that the approximate solution belongs in a subspace of dimension that is much smaller than the dimension of the space the exact solution belongs to. This technique, as it will be described in a following section, requires the solution of a reduced order Sylvester equation. The reduced order Sylvester equation is obtained by projection using bases of suitable Krylov subspaces obtained by Krylov subspace approximation (reduction) techniques [1]. An extension of the Arnoldi method [2] belonging in the class of Krylov subspace reduction techniques is utilized in this paper. A survey on the Krylov subspace methods for the solution of large-scale Sylvester equation can be found in [6].

The contributions of the paper are listed below;

- Existing Krylov subspace techniques for the solution of the Sylvester equations are based on approximations obtained by the standard Arnoldi method. However the reduced order models obtained by the standard Arnoldi methods provide poor approximations at low frequencies. In this paper we propose the use of the rational Arnoldi algorithm [10], [12]. The rational Arnoldi algorithm constructs interpolating approximations of the original system around multiple interpolation points, which are, in general, more accurate than the approximations obtained by the standard Arnoldi methods. The exploitation of the rational Arnoldi algorithm in the restart scheme was possible due to the derivation of a set of simple Arnoldi-like equations in [8] which describe the rational Arnoldi algorithm.
- In [7], [9] a parameterization of a set of interpolating approximations (and a parameterized set of Arnoldi-like equations in the rational case) was derived in terms of a free parameter. A proper choice of the free parameter allows the preservation of some of the properties of the original system such as stability, passivity, etc. [9]. In this work we utilize the parameterized Arnoldi-like equations derived in [9], for the implementation of the proposed restart scheme; a proper choice of the free parameter on every restart iteration allows the improvement in computation efficiency and storage requirements with respect to the standard Krylov subspace methods for the solution to large-scale Sylvester equations. In addition, we show that, in the case of the Lyapunov equation, the error between the approximated solution obtained by the proposed restart scheme and the actual solution of the Lyapunov equation is decreased monotonically on every restart iteration.

In Section II the projection framework for model reduction is briefly described and additional necessary background information is also given such as the derivation of the parameterized set of approximations in terms of a free parameter and the Krylov subspace technique for the approximate

Mian Ilyas Ahmad and Imad Jaimoukha are with the Department of Electrical and Electronics Engineering, Control and Power Group, Imperial College London, UK. mian.ahmad07@imperial.ac.uk i.jaimouka@imperial.ac.uk

Michalis Frangos is with the Department of Aeronautics and Astronautics, Aerospace Computational Design laboratory, Massachusetts Institute of Technology, USA. frangos@mit.edu

solution to the Sylvester equation. In Section III the proposed restart scheme is analyzed in detail which is based on the low rank residual error expressions derived in Section IV. It is shown in Section V that the approximation through the restart scheme approaches monotonically to the actual solution of the Lyapunov equation. Numerical examples of the proposed method can be found in Section VI and finally, conclusions are drawn in Section VII.

II. Background Theory

In this section, we provide background material for the proposed restart scheme; Krylov subspace projection methods, parameterization of the approximations and Krylov subspace techniques for the solution of the Sylvester equation.

A. Model reduction by projection

Consider a linear system of large state dimension n, with transfer function as,

$$G(s) = (sI - A)^{-1}B. \tag{2}$$

The model reduction problem seeks a system of reduced state dimension $m \ll n$, such that it's transfer function G_m provides an accurate approximation of G. In the literature there exists a variety of methods for model reduction [1]. The projection framework techniques seek approximations of the system which belong to a low m-dimensional subspace $\mathcal{V}_m$. In this paper we use Krylov subspace projection techniques, which are efficient iterative methods suitable for the approximation of very large and sparse models.

Suppose a matrix $V_m \in \mathbb{R}^{n\times m}$ is computed as the basis of the subspace $\mathcal{V}_m$, then the approximation of G is given by

$$G_m(s) = V_m F_m(s), \tag{3}$$

where $F_m(s) = (sI - \tilde{A})^{-1}\tilde{B}$ in which $\tilde{A} \in \mathbb{R}^{m\times m}$ and $\tilde{B} \in \mathbb{R}^m$ are the reduced order matrices. The residual error associated with the approximation in (3) is defined as, $R(F_m(s)) := B - (sI - A)V_m F_m(s)$. Requiring that the residual error is orthogonal to a subspace $\mathcal{W}_m$, such that the Galerkin condition $\mathcal{W}_m \perp R(F_m(s))$ is satisfied, yields the reduced order model of G given by

$$G_m(s) = V_m(sE_m - A_m)^{-1}B_m, \tag{4}$$

where $E_m = W_m^T V_m$, $A_m = W_m^T A V_m$ and $B_m = W_m^T B$ are reduced order matrices and W_m is a basis of $\mathcal{W}_m$.

The quality of the approximation of the reduced model depends on the selection of the bases W_m and V_m. For example, constructing orthogonal bases of suitable Krylov subspaces yields reduced models that retain the dominant moments of the large scale system. In the 'standard' Arnoldi method [2] $\mathcal{W}_m = \mathcal{V}_m$ and the matrix $V_m = W_m$ is constructed as an orthonormal basis of the following Krylov subspace,

$$\mathcal{V}_m = \mathcal{K}_m(A, B) = \mathrm{colsp}([B \;\; AB \;\; \cdots \;\; A^{m-1}B]),$$

where colsp denotes the column span. The resulting reduced model provides a poor approximation at low frequencies and usually the dimension m needs to be very large to improve the approximation. An enhanced version of the Arnoldi method is the rational Arnoldi method [12], which provides accurate approximations at a wide range of frequencies while keeping the order of the approximation m low; for this reason we are motivated in utilizing the rational Arnoldi method.

The rational Arnoldi method constructs orthogonal bases of the union of Krylov subspaces $\mathcal{V}_m$ (and $\mathcal{W}_m$),

$$\mathcal{V}_m \supseteq \bigcup_{k=1}^{K} \mathcal{K}_{m_{s_k}}((s_k I - A)^{-1}, (s_k I - A)^{-1}B) \bigcup \mathcal{K}_{m_\infty}(A, B)$$

where $s_k \in \mathcal{S} \subset \mathbb{C}$ and $\mathcal{S}$ is a set of K distinct interpolating points, with associated multiplicities m_{s_k}. Provided that s_k are not the eigenvalues of A or A_m the reduced model in (4) matches the first m_{s_k} moments of G at s_k. The order of approximation is given by $m = \sum_{k=1}^{K} \alpha_k m_{s_k}$, where $\alpha_k = 1$ if $s_k \in \mathbb{R} \cup \infty$ and $\alpha_k = 2$ otherwise. In the proposed restart scheme we never choose $s_k = \infty$ to construct the basis V_m and the reason for this is explained later in Section IV.

In the case where the basis V_m is orthonormal it has been shown in [8] using Lemma 3.1 that the rational Arnoldi method can be described by a set of equations, known as the Arnoldi-like equations, that are given by,

$$AV_m = V_m A_m + v_{m+1}C_m, \tag{5}$$

$$B = V_m B_m + v_{m+1}D_m, \tag{6}$$

where v_{m+1} is constructed such that $V_{m+1} := [V_m \; v_{m+1}] = \mathrm{colsp}([V_m \; B])$, by orthogonalizing B to V_m. Due to orthonormality of V_{m+1}, $C_m = v_{m+1}^T A V_m$ and $D_m = v_{m+1}^T B$.

In [9], the projection based approximations are parameterized in terms of a free parameter and it is shown that a proper choice of the free parameter can improve the approximation and preserve some of the properties of the original system such as stability, passivity, etc. The proposed restart scheme utilizes such free parameters to perform restarts of the algorithm. Next we describe the approach developed in [9] for the parameterization of the projection based approximations. The orthogonal matrix W_m can be expressed in terms of a free parameter L by using,

$$W_m = V_m + v_{m+1}L^T + V_\perp L_\perp^T.$$

This form of the matrix W_m satisfies the bi-orthogonality condition for any $L \in \mathbb{R}^m$ and $L_\perp \in \mathbb{R}^{m\times m+1}$ where the matrix $V_\perp \in \mathbb{R}^{n\times m+1}$ is orthogonal to $V_{m+1} = [V_m \;\; v_{m+1}]$. Using this parameterized matrix W_m and the Arnoldi-like equations (5) and (6), we get the parameterized Galerkin approximation $G_m^L(s)$ given by,

$$G_m^L(s) = V_m F_m^L(s), \tag{7}$$

where $F_m^L(s) = (sI_m - A_m^L)^{-1}B_m^L$ in which $A_m^L = A_m + LC_m$ and $B_m^L = B_m + LD_m$. It has been shown in [9] that the parameterization of the interpolating functions results in a new set of Arnoldi-like equations given by,

$$AV_m = V_m A_m^L + v_{m+1}^L C_m, \tag{8}$$

$$B = V_m B_m^L + v_{m+1}^L D_m, \tag{9}$$

where $v_{m+1}^L = v_{m+1} - V_m L$. The derivation of the new set of Arnoldi-like equations can be obtained by multiplying the Arnoldi equations (5) and (6) by W_m^T from the left and rearranging. In [9] the authors use the parameterized set of equations to develop accurate reduced order models and preserve model properties.

B. Solution of the Sylvester Equation

In this subsection we describe the use of Krylov subspace techniques for the solution of large scale Sylvester equations. The Sylvester equation (1) has a unique solution P if and only if the eigenvalues of A and $\hat{A}$ satisfy $\lambda(A)+\lambda(\hat{A}) \neq 0$, and is given by

$$P = \frac{1}{2\pi i} \int G(s)\hat{G}^{\tilde{}}(s)ds \tag{10}$$

where $G(s) = (sI - A)^{-1}B$ and $\hat{G}^{\tilde{}}(s) = \hat{G}(-s)^T$ in which $\hat{G}(-s) = (-sI-\hat{A})^{-1}\hat{B}$ (see Theorem 5.2.2 in [17]). In [15], it is proposed that an approximate solution of the Sylvester equation (1) can be computed by using the Galerkin approximation to the functions $G(s)$ and $\hat{G}(s)$, obtained via the standard Arnoldi method. That is,

$$P \approx \mathcal{P} = U_m Y \hat{U}_{\hat{m}}^T, \tag{11}$$

where $U_m \in \mathbb{R}^{n\times m}$ and $\hat{U}_{\hat{m}} \in \mathbb{R}^{\hat{n}\times\hat{m}}$ are the orthogonal basis to the Krylov subspace and the matrix $Y \in \mathbb{R}^{m\times\hat{m}}$. The residual error associated with the above approximation can be written as,

$$\mathcal{R} = AU_m Y \hat{U}_{\hat{m}}^T + U_m Y \hat{U}_{\hat{m}}^T \hat{A}^T + B\hat{B}^T. \tag{12}$$

It is found in [15] that the residual error associated with $\mathcal{P}$ satisfies the Galerkin like condition given by,

$$U_m^T \mathcal{R} \hat{U}_{\hat{m}} = 0, \tag{13}$$

if Y satisfies the following reduced order Sylvester equation,

$$A_m Y + Y \hat{A}_{\hat{m}}^T + B_m \hat{B}_{\hat{m}}^T = 0. \tag{14}$$

The residual error is usually compared with a predefined tolerance to obtain the desired size of the projected reduced Sylvester equation. If the residual error is greater than the tolerance level, the values of m and $\hat{m}$ are increased. The increase in the size of the projected Sylvester equation increases the manipulations with the orthogonal matrices U_m and $\hat{U}_{\hat{m}}$. Also the storage requirements for the orthogonal columns of the matrices U_m and $\hat{U}_{\hat{m}}$ become excessive. These issues are addressed in section III, where the algorithm restarts by truncating the orthogonalization and constructing a set of two new orthogonal matrices.

In [11], the rational Arnoldi method is utilized to obtain an approximate solution of the Sylvester equation, however the computational and storage issues are not considered. Using the rational Arnoldi method an approximate solution of the Sylvester equation of the form (11) can be written as,

$$P \approx \tilde{P} = V_m X \hat{V}_{\hat{m}}^T, \tag{15}$$

where $X \in \mathbb{R}^{m\times\hat{m}}$, $V_m \in \mathbb{R}^{n\times m}$ and $\hat{V}_{\hat{m}} \in \mathbb{R}^{\hat{n}\times\hat{m}}$. The matrices V_m and $\hat{V}_{\hat{m}}$ are the orthogonal matrices computed through the rational Arnoldi method. The corresponding residual error can be expressed in the following form by using the rational Arnoldi equations (5) and (6),

$$\tilde{R} = V_{m+1} S \hat{V}_{\hat{m}+1}^T, \tag{16}$$

where $V_{m+1} = [V_m \quad v_{m+1}]$, $\hat{V}_{\hat{m}+1} = [\hat{V}_{\hat{m}} \quad \hat{v}_{\hat{m}+1}]$ and

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \tag{17}$$

in which $S_{11} = A_m X + X(\hat{A}_{\hat{m}})^T + B_m(\hat{B}_{\hat{m}})^T$, $S_{12} = X\hat{C}_{\hat{m}}^T + B_m \hat{D}_{\hat{m}}^T$, $S_{21} = C_m X + D_m(\hat{B}_{\hat{m}})^T$, $S_{22} = D_m \hat{D}_{\hat{m}}^T$.

It is easy to observe that the Galerkin like condition (13), applied to the residual error (16) for the rational case, is satisfied if a condition similar to (14) is obtained by setting $S_{11} = 0$.

III. Restart Scheme

The purpose of this section is to describe the proposed restart scheme for the solution of the Sylvester equation. Suppose we perform m and $\hat{m}$ iterations of the rational Arnoldi algorithm to obtain an approximation $\tilde{P}$ given by (15). The interpolation points are selected such that $\mathcal{S} = -\lambda_m(A)$ and $\hat{\mathcal{S}} = -\lambda_{\hat{m}}(\hat{A})$, where $\lambda_k(Z)$ represent the k eigenvalues of Z that are close to the imaginary axis. Note that the computation of a subset of eigenvalues of A can be efficiently carried out using Krylov subspace techniques [18], [19]. The residual error associated with the approximate solution $\tilde{P}$ can be written as,

$$\begin{aligned} \tilde{R} &= A\tilde{P} + \tilde{P}\hat{A}^T + B\hat{B}^T, \\ &= AV_m X \hat{V}_{\hat{m}}^T + V_m X \hat{V}_{\hat{m}}^T \hat{A}^T + B\hat{B}^T. \end{aligned}$$

Using the paramaterized Arnoldi-like equations (8) and (9) to replace $AV_m, B, \hat{A}\hat{V}_{\hat{m}}$ and $\hat{B}$ in the above equation, we may write,

$$\tilde{R} = V_{m+1}^L \tilde{S} (\hat{V}_{\hat{m}+1}^{\hat{L}})^T, \tag{18}$$

where $V_{m+1}^L = [V_m \quad v_{m+1}^L]$, $\hat{V}_{\hat{m}+1}^{\hat{L}} = [\hat{V}_{\hat{m}} \quad \hat{v}_{\hat{m}+1}^{\hat{L}}]$ and

$$\tilde{S} = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{bmatrix}, \tag{19}$$

in which $\tilde{S}_{11} = A_m^L X + X(\hat{A}_{\hat{m}}^{\hat{L}})^T + B_m^L(\hat{B}_{\hat{m}}^{\hat{L}})^T$, $\tilde{S}_{12} = X\hat{C}_{\hat{m}}^T + B_m^L \hat{D}_{\hat{m}}^T$, $\tilde{S}_{21} = C_m X + D_m(\hat{B}_{\hat{m}}^{\hat{L}})^T$ and $\tilde{S}_{22} = D_m \hat{D}_{\hat{m}}^T$. Thus we introduced two free parameters L and $\hat{L}$ in the residual error expression. Now to update the approximation $\tilde{P}$, we perform a single restart that results in an updated approximation of the form,

$$\tilde{P}_1 = V_m X \hat{V}_{\hat{m}}^T + V_{m_1} X_1 \hat{V}_{\hat{m}_1}^T, \tag{20}$$

where $V_{m_1} \in \mathbb{R}^{n\times m_1}$ and $\hat{V}_{\hat{m}_1} \in \mathbb{R}^{\hat{n}\times\hat{m}_1}$ are the orthogonal matrices obtained via the rational Arnoldi method with interpolation points as selected previously. Also $X_1 \in \mathbb{R}^{m_1\times\hat{m}_1}$ while $\tilde{P}_1 \in \mathbb{R}^{n\times\hat{n}}$. In the above equation, the number of orthogonal columns in V_{m_1} and $\hat{V}_{\hat{m}_1}$ i.e., m_1 and $\hat{m}_1$ are small values compared to m and $\hat{m}$. Note that once we obtain the product $V_m X \hat{V}_{\hat{m}}^T$, we are not storing the orthogonal matrices V_m and $\hat{V}_{\hat{m}}$ for further computations. That is V_{m_1}

and $\hat{V}_{\hat{m}_1}$ are not orthogonalized with V_m and $\hat{V}_{\hat{m}}$. This saves the computations of reorthogonalization and the storage of the orthogonal columns. The residual error associated with the updated approximate solution can be written as,

$$\begin{aligned}\tilde{R}_1 &= A\tilde{P}_1 + \tilde{P}_1\hat{A}^T + B\hat{B}^T \\ &= AV_mX\hat{V}_{\hat{m}}^T + V_mX\hat{V}_{\hat{m}}^T\hat{A}^T + B\hat{B}^T \\ &+ AV_{m_1}X_1\hat{V}_{\hat{m}_1}^T + V_{m_1}X_1\hat{V}_{\hat{m}_1}^T\hat{A}^T \\ &= AV_{m_1}X_1\hat{V}_{\hat{m}_1}^T + V_{m_1}X_1\hat{V}_{\hat{m}_1}^T\hat{A}^T + \tilde{R}\end{aligned}$$

where $\tilde{R}$ is the residual error corresponding to the initial approximation, $\tilde{P}$. If we choose the free parameters L and $\hat{L}$ such that the residual error $\tilde{R}$ is expressed in the low rank form, preferably rank 1 then we can write,

$$\tilde{R} = B^1(\hat{B}^1)^T \tag{21}$$

where $B^1 \in \mathbb{R}^n$ and $\hat{B}^1 \in \mathbb{R}^{\hat{n}}$. The next section describes the choice of L and $\hat{L}$ for which the residual error is rank 1. Using (21), the residual error associated with the updated approximation becomes,

$$\tilde{R}_1 = AV_{m_1}X_1\hat{V}_{\hat{m}_1}^T + V_{m_1}X^1\hat{V}_{\hat{m}_1}^T\hat{A}^T + B^1(\hat{B}^1)^T. \tag{22}$$

This residual error now has the form of the Sylvester equation and can be solved by using Krylov subspace techniques. Using the Arnoldi-like equations (8) and (9) for A, B^1 and $\hat{A}, \hat{B}^1$, we can express the new residual error in the form of (18). To perform the second update of the approximation, we need to express the new residual error $\tilde{R}_1$ in the low rank form,

$$\tilde{R}_1 = B^2(\hat{B}^2)^T. \tag{23}$$

Similarly we can restart the algorithm K times to get an approximation of the form,

$$\tilde{P}_i = V_mX\hat{V}_{\hat{m}}^T + \sum_{i=1}^{K} V_{m_i}X_i\hat{V}_{\hat{m}_i}^T \tag{24}$$

where $i = 1, 2, ..., K$. The corresponding residual error can then be written as

$$\tilde{R}_i = B^i(\hat{B}^i)^T + AV_{m_i}X_i\hat{V}_{\hat{m}_i}^T + V_{m_i}X_i\hat{V}_{\hat{m}_i}^T\hat{A}^T. \tag{25}$$

The complete restart algorithm for the solution of the Sylvester equation is summarized in algorithm 1. The algorithm performs restarts by expressing the residual error in the low rank form using parameters L and $\hat{L}$. This means that at the last update we can obtain the updated approximation without using the parameters L and $\hat{L}$ and can compute the associated residual error (not necessarily low rank). It is found that the approximation computed through the restart algorithm without using parameters L and $\hat{L}$ at the final restart approaches the approximation to the solution of the Sylvester equation without restart.

Algorithm 1 Restart scheme for the solution of Sylvester equation

Require: A, B, $\mathcal{S} = \{s_1, s_2, ..., s_k\}$ and $\hat{A}$, $\hat{B}$, $\hat{\mathcal{S}} = \{\hat{s}_1, \hat{s}_2, ..., \hat{s}_k\}$

1: Initialize: K, k and p
2: Use A, B, $\mathcal{S}$ and the rational Arnoldi algorithm to obtain A_m, B_m, C_m, D_m, V_m, v_{m+1}
3: Use $\hat{A}$, $\hat{B}$, $\hat{\mathcal{S}}$ and the rational Arnoldi algorithm to obtain $\hat{A}_{\hat{m}}$, $\hat{B}_{\hat{m}}$, $\hat{C}_{\hat{m}}$, $\hat{D}_{\hat{m}}$, $\hat{V}_{\hat{m}}$, $\hat{v}_{\hat{m}+1}$
4: Calculate X by solving the nonsymmetric algebraic Riccati equation (28) and find the parameters L and $\hat{L}$ from (26) and (27)
5: Calculate the approximation $\tilde{P}$ and the corresponding residual error $\tilde{R}$ using (15) and (18) respectively
6: Choose $B^1 = v_{m+1}^L D_m$ and $\hat{B}^1 = \hat{v}_{\hat{m}+1}\hat{D}_{\hat{m}}$
7: **for** $i = 1$ to K **do**
8: Use A, $\hat{A}$, B^i, $\hat{B}^i$, $\mathcal{S}_i = \{s_{k+1}, s_{k+2}, ..., s_{k+p}\}$, $\hat{\mathcal{S}}_i = \{\hat{s}_{k+1}, \hat{s}_{k+2}, ..., \hat{s}_{k+p}\}$ and the rational Arnoldi algorithm to obtain A_{m_i}, B_{m_i}, C_{m_i}, D_{m_i}, $\hat{A}_{\hat{m}_i}$, $\hat{B}_{\hat{m}_i}$, $\hat{C}_{\hat{m}_i}$, $\hat{D}_{\hat{m}_i}$, $\hat{V}_{\hat{m}_i}$, $\hat{v}_{\hat{m}_i+1}$, V_{m_i}, v_{m_i+1}
9: Calculate X_i by solving the updated equation (28) and find the corresponding parameters L_i and $\hat{L}_i$ from (26) and (27)
10: Calculate the updated solution $\tilde{P}_i = \tilde{P}_{i-1} + V_{m_i}X_i\hat{V}_{\hat{m}_i}^T$, where $\tilde{P}_0 = \tilde{P}$
11: Compute the corresponding low rank residual error $\tilde{R}_i = B^{i+1}(\hat{B}^{i+1})^T$.
12: Set k=k+p
13: **end for**

IV. Low Rank Residual Error Analysis

In this section, we describe the choice of the parameters L and $\hat{L}$ to obtain the low rank residual error expression. It follows from (19) that a suitable choice of L, $\hat{L}$ and X is one that makes the $\tilde{S}_{11}$, $\tilde{S}_{12}$ and $\tilde{S}_{21}$ blocks of $\tilde{S}$ zero. Expressions for parameters L and $\hat{L}$ obtained by equating the $\tilde{S}_{12}$ and $\tilde{S}_{21}$ blocks of $\tilde{S}$ to zero are given as

$$L = -(B_m\hat{D}_{\hat{m}}^T + X\hat{C}_{\hat{m}}^T)(D_m\hat{D}_{\hat{m}}^T)^{-1}, \tag{26}$$

$$\hat{L} = -(\hat{B}_{\hat{m}}D_m^T + X^TC_m^T)(\hat{D}_{\hat{m}}D_m^T)^{-1}, \tag{27}$$

provided that D_m and $\hat{D}_{\hat{m}}$ are nonzero, which we will assume. Note that $D_m \neq 0$ and $\hat{D}_{\hat{m}} \neq 0$ provided that ∞ is not chosen to be an interpolating point for (A, B) and $(\hat{A}, \hat{B})$. Using the expressions of L and $\hat{L}$ in the equation $\tilde{S}_{11}$, we get a low order nonsymmetric algebraic Riccati equation given by,

$$A_iX + X\hat{A}_i^T - X\hat{C}_i^TC_iX = 0 \tag{28}$$

where $A_i = A_m - B_mD_m^{-1}C_m$, $\hat{A}_i = \hat{A}_{\hat{m}} - \hat{B}_{\hat{m}}\hat{D}_{\hat{m}}^{-1}\hat{C}_{\hat{m}}$, $C_i = D_m^{-1}C_m$ and $\hat{C}_i = \hat{D}_{\hat{m}}^{-1}\hat{C}_{\hat{m}}$. The low order Riccati equation can be solved for X by using one of the techniques described in Subsection IV-A. The parameters L and $\hat{L}$ can then be computed by using the value of X. For these values of L, $\hat{L}$ and X, the residual error can be written as,

$$\tilde{R} = v_{m+1}^L D_m\hat{D}_{\hat{m}}^T(\hat{v}_{\hat{m}+1}^{\hat{L}})^T. \tag{29}$$

This residual error has rank 1 and can be expressed in the form of (21) by using $B^1 = v^L_{m+1} D_m$ and $\hat{B}^1 = \hat{v}^{\hat{L}}_{\hat{m}+1} \hat{D}_{\hat{m}}$. We can also derive the above low rank residual error expression by equating the Schur complement of S given in (17) to zero and introducing the free parameters L and $\hat{L}$ as expressed in (26) and (27). The parameters L and $\hat{L}$ can be written as $L = -S_{12}S_{22}^{-1}$ and $\hat{L} = -S_{21}^T S_{22}^{-T}$ and thus,

$$\begin{bmatrix} I & L \\ 0 & 1 \end{bmatrix} S \begin{bmatrix} I & 0 \\ \hat{L}^T & 1 \end{bmatrix} = \begin{bmatrix} S_{11} - S_{12}S_{22}^{-1}S_{21} & 0 \\ 0 & S_{22} \end{bmatrix},$$

where $S_{11} - S_{12}S_{22}^{-1}S_{21}$ is the Schur complement of S. Using the above equation and the inverse of the unit triangular matrices, the residual error expression can be written as,

$$\begin{aligned} \tilde{R} &= V_{m+1} \begin{bmatrix} I & -L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} S_{11} - S_{12}S_{22}^{-1}S_{21} & 0 \\ 0 & S_{22} \end{bmatrix} \\ &\qquad \begin{bmatrix} I & 0 \\ -\hat{L}^T & 1 \end{bmatrix} \hat{V}^T_{\hat{m}+1} \\ &= V^L_{m+1} \begin{bmatrix} S_{11} - S_{12}S_{22}^{-1}S_{21} & 0 \\ 0 & S_{22} \end{bmatrix} (\hat{V}^{\hat{L}}_{\hat{m}+1})^T \end{aligned} \quad (30)$$

where $V^L_{m+1} = [V_m \quad v^L_{m+1}]$ and $\hat{V}^{\hat{L}}_{\hat{m}+1} = [\hat{V}_{\hat{m}} \quad \hat{v}^{\hat{L}}_{\hat{m}+1}]$. Now if the Schur complement of S is zero then the residual error is low rank and is given by equation (29).

A. Solution of Nonsymmetric Algebraic Riccati Equation

It is clear that the low rank residual error expression requires the solution of the equation (28). In this subsection, we consider two possible methods for the solution of the Riccati equation.

The first method is a standard approach for solving algebraic Riccati equation (see Theorem 7.1.2 in [17]). The technique involves the use of the following $(m + \hat{m}) \times (m + \hat{m})$ Hamiltonian matrix associated with the algebraic Riccati equation (28),

$$H = \begin{bmatrix} \hat{A}_i^T & -C_i \hat{C}_i^T \\ 0 & -A_i \end{bmatrix}. \quad (31)$$

Suppose a matrix T transforms H into a Jordan form $HT = TJ$, where J is the Jordan matrix with the eigenvalues of H as the diagonal entries and T contains the $m+\hat{m}$ eigenvectors of H. If T is partitioned as $T = [T_1 \ T_2]$ where $T_1 \in \mathbb{R}^{(m+\hat{m})\times m}$ such that it contains the first m eigenvectors of H and T_2 contains the remaining $\hat{m}$ eigenvectors, then we can write

$$HT_1 = T_1 \Lambda \quad (32)$$

where Λ contain the eigenvalues of H corresponding to the eigenvectors in T_1. It is easy to show that each solution to the Quadratic Sylvester equation can be expressed as, $X = T_{21}T_{11}^{-1}$, where $T_{11} \in \mathbb{R}^{m\times m}$ is assumed to be nonsingular and $[T_{11}^T \ T_{21}^T]^T = T_1$. This method works for any choice of Λ and an interesting issue is to investigate an optimal choice of Λ, which will be considered in future work.

The second method transforms the nonsymmetric algebraic Riccati equation (28) in a low order Sylvester equation if the matrix X is square and invertible. It is easy to see that if $Y = X^{-1}$ exists, then the pre-multiplication and the post-multiplication of Y to (28) results in,

$$\hat{A}_i^T Y + Y A_i - \hat{C}_i^T C_i = 0. \quad (33)$$

The above equation can be solved for Y by using direct methods. The required solution is then, $X = Y^{-1}$.

V. Solution of Lyapunov Equation

In this section we describe the restart scheme for the approximate solution of the Lyapunov equation. It is shown that the updated approximations monotonically approaches the actual solution of the Lyapunov equation. The Lyapunov equation can be written as,

$$AQ + QA^T + BB^T = 0, \quad (34)$$

where $Q \in \mathbb{R}^{n\times n}$ is the positive definite solution of the Lyapunov equation if the matrix A is stable. Suppose Q_m is the initial approximate solution of the Lyapunov equation (34) using our technique. Then,

$$AQ_m + Q_m A^T + BB^T - B^1(B^1)^T = 0, \quad (35)$$

where $B^1(B^1)^T$ is the residual error associated with the approximate solution Q_m. Subtracting equation (35) from equation (34), we get,

$$A(Q - Q_m) + (Q - Q_m)A^T + B^1(B^1)^T = 0. \quad (36)$$

As $Q - Q_m$ is the solution of the above Lyapunov equation and A is stable, we have $Q \geq Q_m$. Now after the first restart of the algorithm, the updated approximate solution is of the form,

$$Q^1_{m_1} = Q_m + V^1_{m_1} X^1 (V^1_{m_1})^T. \quad (37)$$

It is clear from the above equation that if $X^1 \geq 0$, the updated approximation satisfies $Q^1_{m_1} \geq Q_m$. As each of the updated approximate solutions has a low rank residual error expression, we can obtain the Lyapunov equations of the form (36) for all the updated approximations. Thus the following relation is obtained among the updated solutions, initial approximate solution and the actual solution.

$$Q \geq Q^K_{m_K} \geq Q^{K-1}_{m_{K-1}} \geq \cdots \geq Q^1_{m_1} \geq Q_m. \quad (38)$$

This shows that the updated approximations monotonically approach the exact solution of the Lyapunov equation.

VI. Numerical Results

In this section, we compare the proposed restart scheme and the standard Krylov method for the approximation of the solution to the Sylvester equation. The approximate solution obtained by the proposed scheme converges to the one obtained by the standard Krylov technique with two important advantages; low computational cost and less storage requirements. These advantages are due to the fact that the proposed scheme avoids the storage of large orthogonal matrices and it also avoids orthogonalization of the bases constructed on each iteration with previously computed bases on each iteration of the restart scheme. A complete comparison of the method in terms of computational cost, storage requirement and convergence rates with other existing methods is out of the scope of the current work and thus omitted.

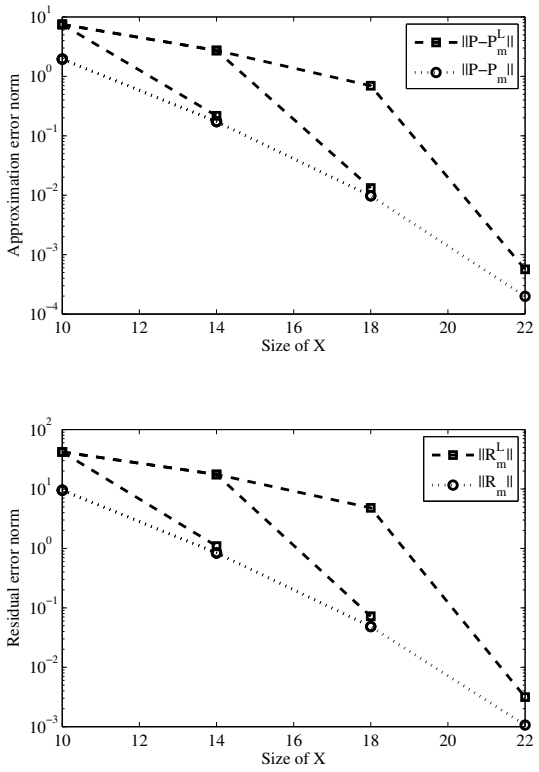

Fig. 1. The 2-norm of the forward error and the residual error are plotted for the restart scheme and the standard Krylov technique to solve the Sylvester equation. P is the actual solution of the Sylvester equation, while P_m^L and P_m denote the approximate solutions with and without restart scheme, respectively. Similarly, R_m^L and R_m express the residual errors with and without restarts. For all sizes of X the proposed scheme (dashed line) has similar values of forward and residual error to that of standard Krylov technique (dotted line) but with reduced computational complexity.

A. Example 1

This example is taken from [20] where the matrices $A \in \mathbb{R}^{n\times n}$ and $\hat{A} \in \mathbb{R}^{\hat{n}\times\hat{n}}$ are in the form

$$A = tridiag(-1 + \frac{10}{n+1}, -2, -1 + \frac{10}{n+1}),$$
$$\hat{A} = tridiag(-1 + \frac{10}{\hat{n}+1}, -2, -1 + \frac{10}{\hat{n}+1}),$$

where $n = 300$ and $\hat{n} = 250$. The vectors B and $\hat{B}$ are randomly selected. Figure 1 shows the norm of the forward error (defined as the error between the actual and the approximate solution) and the residual error for the proposed scheme and the standard Krylov technique. The two strategies are compared at three different projection sizes where the size of the projected Sylvester equation is taken as the same size to that of matrix X. For all cases, the approximation computed through the restart scheme approaches the one obtained through the standard Krylov technique with the advantage that the restart scheme does not involve the orthogonalization cost and the storage of orthogonal columns.

VII. Conclusions

A computationally efficient Krylov based restart scheme has been developed for the approximate solution of the Sylvester and the Lyapunov equations. In case of Lyapunov equations, the updated approximations monotonically converge to the actual solution. The exact expressions for the low rank residual error are computed for each of the updated approximation. Although the approximation and the residual error converge with the number of restarts, they increase the rank of the approximation which, in turn, requires more storage; we aim to address this problem in our future work. Also, the choice of the parameters L and $\hat{L}$ require further investigation to improve the approximation error.

References

[1] A. C. Antoulas. *Approximation of large-scale dynamical systems*. SIAM Publications, 2005.

[2] W.E. Arnoldi. The principle of minimized iterations in the solution of matrix eigenvalue problem. *Quart. Appl. Math*, 9:17–29, 1951.

[3] R.H. Bartels and G.W. Stewart. Solution of the matrix equation $AX + XA = C$: Algorithm 432. *Communications of the ACM*, 15:820–826, 1972.

[4] P. Benner. Factorized solution of Sylvester equations with applications in control. In *Proceedings of International Symposium on Mathematical Theory of Networks and Systems (MTNS2004)*, 2004.

[5] D. Calvetti and L. Reichel. Application of ADI iterative methods to the restoration of noisy images. *SIAM Journal on Matrix Analysis and Applications*, 17:165–186., 1996.

[6] B.N. Datta. Krylov subspace methods for large-scale matrix problems in control. *Future Generation Computer Systems*, 19:12531263, 2003.

[7] M. Frangos. *Adaptive Krylov Subspace Methods for Model Reduction of Large Scale Systems*. PhD thesis, Imperial College London, London,UK, May 2003.

[8] M. Frangos and I.M. Jaimoukha. Adaptive rational interpolation: Arnoldi and Lanczos-like equations. *European Journal of Control*, 14(4):342 – 354, 2008.

[9] M. Frangos and I.M. Jaimoukha. A parameterization of a set of interpolating approximations via the rational Arnoldi algorithm: Passivity, stability and minimum-phase preserving. *submitted to SIAM Journal on Matrix Analysis and Applications (SIMAX)*, 2010.

[10] K. Gallivan, E. Grimme, and P. Van Dooren. A rational Lanczos algorithm for model reduction. *Numerical Algorithms*, 12:33–64, 1996.

[11] K. Gallivan, A. Vandendorpe, and P. Van Dooren. Sylvester equations and projection-based model reduction. *Journal of Computational and Applied Mathematics*, 162:213229, 2004.

[12] E. Grimme. *Krylov projection methods for model reduction*. PhD thesis, University of Illinois at Urbana-Champain, Urbana, Illinois, 1997.

[13] A. El Guennouni, K. Jbilou, and A. J. Riquet. Block Krylov subspace methods for solving large Sylvester equations. *Numerical Algorithms*, 29:75–96, 2002.

[14] S. Hammarling. Numerical solution of the stable, non-negative definite Lyapunov equation. *IMA Journal of Numerical Analysis*, 2:303–323., 1982.

[15] I.M. Jaimoukha and E.M. Kasenally. Krylov subspace methods for solving large Lyapunov equations. *SIAM Journal on Numerical Analysis*, 31:227–251, 1994.

[16] K. Jbilou and A.J. Riquet. Projection methods for large Lyapunov matrix equations. *Linear Algebra and its Applications*, 415:344–358,, 2006.

[17] P. Lancaster and L. Rodman. *Algebraic Riccati Equations*. Clarendon press, Oxford, 1995.

[18] R. B. Lehoucq, D. C. Sorensen, and C. Yang. Arpack users guide. *SIAM, Philadelphia*, 9, 1998.

[19] Y Saad. Iterative methods for sparse linear systems. *SIAM Philadelphia*, 2003.

[20] D.K. Salkuyeh and F. Toutounian. New approaches for solving large Sylvester equations. *Applied Mathematics and Computations*, 173:9–18, 2006.

[21] V. Simoncini. A new iterative method for solving large-scale Lyapunov matrix equations. *SIAM Journal on Scientific Computing*, 29(3):1268–1288, 2008.