A Parameter Free ADI-like Method for the Numerical Solution of Large Scale Lyapunov Equations *

RYAN NONG AND DANNY C. SORESENSEN[†]

ryannong@caam.rice.edu
sorensen@rice.edu

May 15, 2009

Abstract

An algorithm is presented for constructing an approximate numerical solution to a large scale Lyapunov equation in low rank factored form. The algorithm is based upon a synthesis of an approximate power method and an alternating direction implicit (ADI) method. The former is parameter free and tends to be efficient in practice, but there is little theoretical understanding of its convergence properties. The ADI method has a well understood convergence theory, but the method relies upon selection of shift parameters and a poor shift selection can lead to very slow convergence in practice. The algorithm presented here uses an approximate power method iteration to obtain a basis update. It then constructs a re-weighting of this basis update to provide a factorization update that satisfies ADI-like convergence properties.

1 Introduction

This paper proposes a new algorithm for obtaining approximate solutions $\mathcal{P}$ to large scale Lyapunov equations of the form

$$\mathbf{A}\mathcal{P} + \mathcal{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0}, \quad (1.1)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$ with $\mathbf{A}$ stable (all eigenvalues in the open left half-plane) and $p \leq n$. It is well known that the solution $\mathcal{P}$ is symmetric and positive semidefinite and that $\mathcal{P}$ is definite if and only if $\text{rank}[\lambda\mathbf{I} - \mathbf{A}, \mathbf{B}] = n$ for all $\lambda \in \mathbb{C}$. However, even when $\mathcal{P}$ is positive definite, it is often the case that the eigenvalues of $\mathcal{P}$ decay rapidly, so that $\mathcal{P}$ may be well approximated by a low rank matrix. This situation arises in numerous applications involving control of dynamical systems. In that setting, there is considerable interest in obtaining approximate solutions $\mathcal{P} \approx \mathbf{L}\mathbf{L}^T$ in *low rank factored form*, meaning that $\mathbf{L} \in \mathbb{R}^{n \times k}$ is a rank k matrix factor with $k \ll n$. If such an approximation has been obtained, one can efficiently construct an approximate balanced reduction of a large scale linear time invariant system [11, 9, 6].

*This work was supported in part by NSF Grants ACI-0325081 and CCF-0634902 and AFOSR grant FA9550-09-1-0225.

[†]Department of Computational and Applied Mathematics, MS 134, Rice University, Houston, Texas 77251-1892.

A number of techniques have been proposed and investigated. Two methods of particular interest are approximate power methods [7] and ADI/Smith methods [15, 11, 9]. Approximate power methods can be quite effective in practice but very little is known about convergence. On the other hand, there is an elegant convergence theory for ADI methods that is complete and well understood. Unfortunately, the performance of ADI methods is heavily dependent upon shift parameters and a poor shift selection can easily lead to very slow convergence. There is a theory using elliptic functions for optimal shift selection in the case that $\mathbf{A}$ is symmetric [3, 13]. However, the theory is incomplete for the non-symmetric case. Some heuristics can be found in [11, 13].

Here, we derive an algorithm for obtaining an approximate solution in low rank factored form that retains ADI guaranteed convergence properties without the need to select parameters. Parameter selection is done automatically through the approximate power method.

Another parameter free approach to solving Lyapunov equations is offered by Simoncini [14]. This approach is based upon approximating $\mathbf{e}^{\mathbf{A}t}\mathbf{B}$ from a specially constructed rational Krylov space

$$\mathcal{K} = \text{Span}\{\mathbf{B}, \mathbf{A}^j\mathbf{B}, \mathbf{A}^{-j}\mathbf{B} \mid j = 1 : k\}.$$

The approximate solution takes the form

$$\mathcal{P} = \mathbf{V}\mathbf{P}\mathbf{V}^T \text{ with } \mathbf{H}\mathbf{P} + \mathbf{P}\mathbf{H}^T + \hat{\mathbf{B}}\hat{\mathbf{B}}^T = \mathbf{0},$$

where $\mathbf{H} = \mathbf{V}^T\mathbf{A}\mathbf{V}$, $\hat{\mathbf{B}} = \mathbf{V}^T\mathbf{B}$ and $\mathbf{V}$ is an orthogonal matrix with $\text{Range}(\mathbf{V}) = \mathcal{K}$. This relies upon how well

$$\mathbf{V} \left[\int_0^\infty \mathbf{e}^{\mathbf{H}t} \hat{\mathbf{B}}\hat{\mathbf{B}}^T \mathbf{e}^{\mathbf{H}^T t} dt \right] \mathbf{V}^T$$

approximates

$$\mathcal{P} = \int_0^\infty \mathbf{e}^{\mathbf{A}t} \mathbf{B}\mathbf{B}^T \mathbf{e}^{\mathbf{A}^T t} \mathbf{B} dt.$$

Several refinements are included to provide an efficient implementation of this basic idea.

Our approach here is quite different and more closely related to ADI. However, it does make use of solutions to the small dense Lyapunov equations obtained by projecting the full equation onto a Krylov subspace.

2 The Alternating Direction Implicit Method

This section gives a brief derivation of the ADI/Smith method for Lyapunov equations. We shall use $\mathbf{M}^T$ to denote the transpose of a matrix $\mathbf{M}$ and $\mathbf{M}^*$ to denote complex-conjugate transpose. Although the notation $\mathbf{M}^*$ could be used throughout, we wish to reserve it for those cases where $\mathbf{M}$ could be complex. This is done to emphasize where real arithmetic can be used.

From the original equation

$$\mathbf{A}\mathcal{P} + \mathcal{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0},$$

applying a complex shift μ with $\rho \equiv \text{Real}(\mu) < 0$ from the left gives

$$\mathcal{P} = -(\mathbf{A} + \mu\mathbf{I})^{-1} [\mathcal{P}(\mathbf{A} - \bar{\mu}\mathbf{I})^* + \mathbf{B}\mathbf{B}^*]. \quad (2.1)$$

Applying the shift μ from the right (alternate direction) gives

$$\mathcal{P} = -[(\mathbf{A} - \bar{\mu}\mathbf{I})\mathcal{P} + \mathbf{B}\mathbf{B}^*](\mathbf{A} + \mu\mathbf{I})^{-*}. \quad (2.2)$$

Substitute the formula for $\mathcal{P}$ in (2.2) into the right hand side of (2.1) to get a Stein equation of the form

$$\mathbf{A}\mathcal{P} + \mathcal{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0 \iff \mathcal{P} = \mathbf{A}_\mu\mathcal{P}\mathbf{A}_\mu^* + \mathbf{B}_\mu\mathbf{B}_\mu^*,$$

where

$$\mathbf{A}_\mu = (\mathbf{A} - \bar{\mu}\mathbf{I})(\mathbf{A} + \mu\mathbf{I})^{-1}, \quad \mathbf{B}_\mu = \sqrt{2|\rho|}(\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{B}.$$

A convenient aspect of this formulation is that the analytic solution can be written as an infinite matrix series

$$\mathcal{P} = \sum_{j=0}^{\infty} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^* (\mathbf{A}_\mu^j)^* = \mathbf{L}\mathbf{L}^*,$$

where $\mathbf{L} = [\mathbf{B}_\mu, \mathbf{A}_\mu\mathbf{B}_\mu, \mathbf{A}_\mu^2\mathbf{B}_\mu, \dots]$, which expresses the solution in factored form. Since $\mathbf{A}$ is asymptotically stable, it follows that the spectral radius of $\mathbf{A}_\mu$ is less than 1. This leads to a very straightforward analysis of convergence. Putting

$$\mathcal{P}_m = \sum_{j=0}^{m-1} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^* (\mathbf{A}_\mu^j)^*$$

gives

$$\mathcal{P}_{m+1} = \mathbf{A}_\mu\mathcal{P}_m\mathbf{A}_\mu^* + \mathbf{B}_\mu\mathbf{B}_\mu^*,$$

and it is easily shown that

$$\mathcal{E}_{m+1} = \mathbf{A}_\mu\mathcal{E}_m\mathbf{A}_\mu^* = \mathbf{A}_\mu^{m+1}\mathcal{P}(\mathbf{A}_\mu^{m+1})^* \rightarrow 0,$$

where $\mathcal{E}_m = \mathcal{P} - \mathcal{P}_m$ for $m = 1, 2, 3, \dots$.

The Low-Rank Smith method developed by Penzl [11] repeatedly updates the factored form $\mathcal{P}_m = \mathbf{L}_m\mathbf{L}_m^*$ via

$$\begin{aligned} \mathbf{L}_{m+1} &= [\mathbf{B}_\mu, \mathbf{A}_\mu\mathbf{L}_m] \\ &= [\mathbf{L}_m, \mathbf{A}_\mu^{m+1}\mathbf{B}_\mu]. \end{aligned}$$

The second formulation of the update may be implemented by

$$\begin{aligned} \mathbf{Z} &= \mathbf{B}_\mu; \quad \mathbf{L} = []; \\ \text{while ('not converged')} \\ \quad \mathbf{L} &\leftarrow [\mathbf{L}, \mathbf{Z}]; \\ \quad \mathbf{Z} &\leftarrow \mathbf{A}_\mu\mathbf{Z}; \\ \text{end} \end{aligned}$$

Since $\mathbf{L}$ typically has many more columns than $\mathbf{B}$, this is far less expensive than the first formulation but is potentially unstable when $\mathbf{A}$ is highly non-normal. Both formulations may require prohibitively large amounts of storage.

An asymptotic convergence rate is obtained from the spectral radius $\rho(\mathbf{A}_\mu)$, and one may attempt to select an optimal value of μ to minimize this spectral radius. Usually a single shift μ is not sufficient to obtain a spectral radius significantly less than one. Therefore multiple shifts are recommended. However, each individual shift requires a matrix factorization, which must be stored or re-computed each time a particular shift is applied. In [6], a Modified Smith Method was developed to overcome the storage difficulties and also to enable applications of multiple shifts with only one matrix factorization per shift.

Let $\mathcal{P}_m = \sum_{j=0}^{m-1} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^* (\mathbf{A}_\mu^j)^*$ be the approximate solution obtained after m applications of the shift μ . Then

$$\begin{aligned} \mathbf{A}\mathcal{P}_m + \mathcal{P}_m\mathbf{A}^* + \mathbf{B}\mathbf{B}^* &= \mathbf{A}(\mathcal{P}_m - \mathcal{P}) + (\mathcal{P}_m - \mathcal{P})\mathbf{A}^* \\ &= -\mathbf{A} \left(\sum_{j=m}^{\infty} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^* (\mathbf{A}_\mu^j)^* \right) - \left(\sum_{j=m}^{\infty} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^* (\mathbf{A}_\mu^j)^* \right) \mathbf{A}^* \\ &= -\mathbf{A}_\mu^m (\mathbf{A}\mathcal{P} + \mathcal{P}\mathbf{A}^*) (\mathbf{A}_\mu^m)^* \\ &= \mathbf{A}_\mu^m \mathbf{B} (\mathbf{A}_\mu^m \mathbf{B})^*. \end{aligned}$$

It follows that the difference $\mathcal{P} - \mathcal{P}_m$ satisfies the modified Lyapunov equation

$$\mathbf{A}(\mathcal{P} - \mathcal{P}_m) + (\mathcal{P} - \mathcal{P}_m)\mathbf{A}^* + \mathbf{B}_m \mathbf{B}_m^* = \mathbf{0},$$

where $\mathbf{B}_m \equiv \mathbf{A}_\mu^m \mathbf{B}$. Hence, after m steps, one may replace $\mathbf{B}$ by $\mathbf{B}_m$ and proceed to solve

$$\mathbf{A}\mathbf{Z} + \mathbf{Z}\mathbf{A}^* + \mathbf{B}_m \mathbf{B}_m^* = \mathbf{0}$$

to obtain

$$\mathcal{P} = \mathcal{P}_m + \mathbf{Z}.$$

The Modified Smith Method propagates a low rank SVD approximation to $\mathbf{L}_m$ and aggregates the shift application so that all instances of a particular shift are applied before the next shift is brought in and applied. At each shift application, the SVD of $\mathbf{L}$ is updated and truncated in a manner that maintains sufficient accuracy of the SVD approximation to $\mathbf{L}$. This modified Smith method is much faster and requires far less storage than the original method. The steps of an iteration (with the initialization $\mathbf{Z} = \mathbf{B}_\mu$) are

$$\begin{aligned} \mathbf{Z} &\leftarrow \mathbf{A}_\mu \mathbf{Z}; \\ [\mathbf{V}, \mathbf{S}, \mathbf{Q}] &= \text{svd}([\mathbf{Z}, \mathbf{L}_m]); \text{ (via SVD updating)} \\ \mathbf{L}_{m+1} &\leftarrow \mathbf{V}_k \mathbf{S}_k; \text{ (} \sigma_{k+1} < \text{tol} \cdot \sigma_1 \text{)} \end{aligned}$$

The key to performance, however, is still the selection of shifts to minimize the spectral radius of the product $\mathbf{A}_{\mu_1} \mathbf{A}_{\mu_2} \cdots \mathbf{A}_{\mu_k}$. The complete details of this implementation and its performance are available in [6].

3 The Approximate Power Method

A very different approach called the Approximate Power Method was suggested by Hodel et al. [7]. The idea is to utilize a subspace iteration technique to approximately compute the dominant invariant subspace of $\mathcal{P}$. The

difficulty with such an approach is that the matrix $\mathcal{P}$ is not available and thus approximate matrix-matrix products of the form $\mathbf{Z} = \mathcal{P}\mathbf{U}$ must be provided indirectly. The idea is to use the form of the Lyapunov equation and note that if $\mathcal{P}$ is a solution and $\mathbf{U}$ is any $n \times k$ matrix then

$$\mathbf{A}\mathcal{P}\mathbf{U} + \mathcal{P}\mathbf{A}^T\mathbf{U} + \mathbf{B}\mathbf{B}^T\mathbf{U} = \mathbf{0}.$$

Simply adding and subtracting the term $\mathcal{P}\mathbf{U}\mathbf{U}^T\mathbf{A}^T\mathbf{U}$ gives

$$\mathbf{A}\mathcal{P}\mathbf{U} + \mathcal{P}\mathbf{U}\mathbf{U}^T\mathbf{A}^T\mathbf{U} + \mathbf{B}\mathbf{B}^T\mathbf{U} + \mathcal{P}(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{A}^T\mathbf{U} = \mathbf{0}.$$

If we think of $\mathcal{P}(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{A}^T\mathbf{U}$ as a small error term, then

$$\mathbf{A}\mathcal{P}\mathbf{U} + \mathcal{P}\mathbf{U}\mathbf{H}^T + \mathbf{B}\mathbf{B}^T\mathbf{U} \approx \mathbf{0}, \quad \text{where } \mathbf{H} = \mathbf{U}^T\mathbf{A}\mathbf{U}.$$

Thus, solving

$$\mathbf{A}\mathbf{Z} + \mathbf{Z}\mathbf{H}^T + \mathbf{B}\mathbf{B}^T\mathbf{U} = \mathbf{0} \tag{3.1}$$

gives an approximation

$$\mathbf{Z} \approx \mathcal{P}\mathbf{U}$$

to the desired matrix-matrix product.

The simplest form of the Approximate Power Method iterates this calculation with

$$\mathbf{H}_j = \mathbf{U}_j^T\mathbf{A}\mathbf{U}_j, \quad [\mathbf{U}_{j+1}, \mathbf{S}_{j+1}, \mathbf{W}_{j+1}] = \text{svd}(\mathbf{Z}_j) \quad \text{for } j = 1, 2, \dots$$

where $\mathbf{Z}_j$ is obtained by solving (3.1). Variants of this pose the results of equation (3.1) as an update equation (see [19]).

This iteration can often provide very good results and convergence is rapid in many cases. However, there is no assurance that the discarded error term is small, and when it is not, the convergence may be slow or non-existent. A variant of the method is proposed in [17], where the term $\mathcal{P}(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{A}^T\mathbf{U}$ is not discarded but instead approximated with $\mathcal{P}$ being replaced by a low-rank approximation. Computational evidence is given to indicate that this modification improves convergence in practice, but no mathematical proof is provided to guarantee this improvement.

4 A Parameter Free ADI Method

The ultimate goal of both the APM and the ADI methods is to construct a basis $\mathbf{U}$ for the dominant invariant subspace of $\mathcal{P}$ corresponding to its largest eigenvalues. As shown in Figure 1, our synthesis of the APM and ADI methods consists of four basic steps at each iteration. At step j of an outer iteration, an orthogonal basis $\mathbf{U}_j$ for an approximate invariant subspace of $\mathcal{P}$ is assumed to be available. This provides an approximate factorization $\mathcal{P}_j = \mathbf{L}_j\mathbf{L}_j^T$ with $\mathbf{L}_j = \mathbf{U}\mathbf{S}$ and $\mathcal{P}_j\mathbf{U} = \mathbf{U}\mathbf{S}^2$. Thus, $\mathbf{U}$ is an ortho-normal basis for the range of $\mathcal{P}_j$. A selected subspace of $\text{Range}(\mathbf{U})$ is spanned by the columns of an ortho-normal matrix $\mathbf{U}_j$ typically having fewer columns than $\mathbf{U}$. We think of $\mathbf{U}_j$ as an approximate basis for the dominant invariant subspace of $\mathcal{P}_j$. However, other

Step 1: Solve the reduced order Lyapunov equation for $\hat{\mathcal{P}}_j$

$$\text{Solve } \mathbf{H}_j \hat{\mathcal{P}}_j + \hat{\mathcal{P}}_j \mathbf{H}_j^T + \hat{\mathbf{B}}_j \hat{\mathbf{B}}_j^T = \mathbf{0}.$$

Step 2: (APM step) Solve a projected Sylvester equation for $\mathbf{Z}_j$

$$\mathbf{A} \mathbf{Z}_j + \mathbf{Z}_j \mathbf{H}_j^T + \mathbf{B}_j \hat{\mathbf{B}}_j^T = \mathbf{0}, \text{ with } \mathbf{H}_j = \mathbf{U}_j^T \mathbf{A} \mathbf{U}_j, \hat{\mathbf{B}}_j = \mathbf{U}_j^T \mathbf{B}_j.$$

Step 3: Modify $\mathbf{B}_j$

$$\text{Update } \mathbf{B}_{j+1} \leftarrow (\mathbf{I} - \mathbf{Z}_j \hat{\mathcal{P}}_j^{-1} \mathbf{U}_j^T) \mathbf{B}_j.$$

Step 4: (ADI step) Update factorization and basis $\mathbf{U}$

$$\text{Re-scale } \mathbf{Z}_j \leftarrow \mathbf{Z}_j \hat{\mathcal{P}}_j^{-1/2}.$$

$$\text{Update (and truncate) } [\mathbf{U}, \mathbf{S}] \leftarrow \text{svd}[\mathbf{L}_j, \mathbf{Z}_j].$$

$$\mathbf{L}_{j+1} = \mathbf{U} \mathbf{S}.$$

$$\mathbf{U}_{j+1} \leftarrow \mathbf{U}(:, 1 : k_j), \text{ basis for dominant subspace of dimension } k_j.$$

Figure 1: A Parameter Free Lyapunov Solver

choices are possible and these will be discussed in Section 5. The number of columns of $\mathbf{U}_j$ is unspecified and can, for example, be a small number and vary at each outer step j . At step j , the four inner iteration steps shown in Figure 1 are performed. (Initially, $\mathbf{L}_1 = \mathbf{0}$ and $\mathbf{B}_1 = \mathbf{B}$.)

The motivation for these steps resides in the following results. The re-scaling at Step 4 provides an update to the factorization that will always maintain $\mathcal{P} \succeq \mathcal{P}_j$. The modification to $\mathbf{B}_j$ at Step 3 provides an updated Lyapunov equation for the difference $\mathcal{P} - \mathcal{P}_{j+1}$. Each subsequent iteration of these four basic steps provides an incremental update to the approximate factorization such that $\mathcal{P} \succeq \mathcal{P}_{j+1} \succeq \mathcal{P}_j$. Hence, the iteration will automatically be convergent (although not necessarily to $\mathcal{P}$). The following lemmas will establish these properties. In this discussion, it is assumed that $\hat{\mathcal{P}}$ is always positive definite. In fact, it is possible to modify this algorithm to assure that $\text{cond}(\hat{\mathcal{P}})$ is uniformly bounded. This amounts to selecting a subspace corresponding to the leading eigenvalues of $\hat{\mathcal{P}}$ and projecting onto this subspace. The following results are also valid for this modification, which will be introduced in detail when implementation issues are discussed in Section 5.

Lemma 4.1 *Let $\mathbf{Z}_j, \hat{\mathcal{P}}_j, \mathbf{B}_j$ and $\mathbf{H}_j$ be defined as in Steps 1-3 in Figure 1, and let $\mathbf{G}_j = \hat{\mathcal{P}}_j^{-1}$. Then*

$$\mathbf{A} \mathbf{Z}_j \mathbf{G}_j \mathbf{Z}_j^T + \mathbf{Z}_j \mathbf{G}_j \mathbf{Z}_j^T \mathbf{A}^T + \mathbf{B}_j \mathbf{B}_j^T = \mathbf{B}_{j+1} \mathbf{B}_{j+1}^T,$$

where $\mathbf{B}_{j+1} \equiv (\mathbf{I} - \mathbf{Z}_j \mathbf{G}_j \mathbf{U}_j^T) \mathbf{B}_j$.

Proof: First note that

$$\mathbf{A} \mathbf{Z}_j \mathbf{G}_j \mathbf{Z}_j^T = -[\mathbf{Z}_j \mathbf{H}_j^T + \mathbf{B}_j \mathbf{B}_j^T \mathbf{U}_j] \mathbf{G}_j \mathbf{Z}_j^T.$$

Thus

$$\begin{aligned} \mathbf{A}\mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T + \mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T\mathbf{A}^T &= -\mathbf{Z}_j[\mathbf{H}_j^T\mathbf{G}_j + \mathbf{G}_j\mathbf{H}_j]\mathbf{Z}_j^T - \mathbf{B}_j\mathbf{B}_j^T\mathbf{U}_j\mathbf{G}_j\mathbf{Z}_j^T - \mathbf{Z}_j\mathbf{G}_j\mathbf{U}_j^T\mathbf{B}_j\mathbf{B}_j^T \\ &= \mathbf{Z}_j[\mathbf{G}_j\mathbf{U}_j^T\mathbf{B}_j\mathbf{B}_j^T\mathbf{U}_j\mathbf{G}_j]\mathbf{Z}_j^T - \mathbf{B}_j\mathbf{B}_j^T\mathbf{U}_j\mathbf{G}_j\mathbf{Z}_j^T - \mathbf{Z}_j\mathbf{G}_j\mathbf{U}_j^T\mathbf{B}_j\mathbf{B}_j^T, \end{aligned}$$

and hence

$$\begin{aligned} \mathbf{A}\mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T &+ \mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T\mathbf{A}^T + \mathbf{B}_j\mathbf{B}_j^T \\ &= \mathbf{Z}_j\mathbf{G}_j\mathbf{U}_j^T\mathbf{B}_j\mathbf{B}_j^T\mathbf{U}_j\mathbf{G}_j\mathbf{Z}_j^T - \mathbf{B}_j\mathbf{B}_j^T\mathbf{U}_j\mathbf{G}_j\mathbf{Z}_j^T - \mathbf{Z}_j\mathbf{G}_j\mathbf{U}_j^T\mathbf{B}_j\mathbf{B}_j^T + \mathbf{B}_j\mathbf{B}_j^T \\ &= (\mathbf{I} - \mathbf{Z}_j\mathbf{G}_j\mathbf{U}_j^T)\mathbf{B}_j\mathbf{B}_j^T(\mathbf{I} - \mathbf{U}_j\mathbf{G}_j^{-1}\mathbf{Z}_j^T) \\ &= \mathbf{B}_{j+1}\mathbf{B}_{j+1}^T. \end{aligned}$$

□

This result leads immediately to an equation for the difference $\mathcal{P} - \mathcal{P}_j$.

Lemma 4.2 *Let the hypothesis of Lemma 4.1 hold. Then*

$$\mathbf{A}\mathcal{P}_j + \mathcal{P}_j\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{B}_j\mathbf{B}_j^T, \quad (4.1)$$

and

$$\mathbf{A}(\mathcal{P} - \mathcal{P}_j) + (\mathcal{P} - \mathcal{P}_j)\mathbf{A}^T + \mathbf{B}_j\mathbf{B}_j^T = \mathbf{0}. \quad (4.2)$$

As a consequence, $\mathcal{P} \succeq \mathcal{P}_{j+1} \succeq \mathcal{P}_j \succeq \mathbf{0}$ for $j = 1, 2, \dots$

Proof: The proof will be an induction. Since $\mathbf{B}_1 = \mathbf{B}$ and $\mathcal{P}_1 = \mathbf{0}$, Equations (4.1) and (4.2) are trivially satisfied when $j = 1$. Assume Equations (4.1) and (4.2) are true for some $j \geq 1$. It follows from Lemma 4.1 that

$$\begin{aligned} \mathbf{A}\mathcal{P}_j + \mathcal{P}_j\mathbf{A}^T + \mathbf{B}\mathbf{B}^T &= \mathbf{B}_j\mathbf{B}_j^T \\ &= \mathbf{B}_{j+1}\mathbf{B}_{j+1}^T - [\mathbf{A}\mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T + \mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T\mathbf{A}^T]. \end{aligned}$$

Hence,

$$\mathbf{A}(\mathcal{P}_j + \mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T) + (\mathcal{P}_j + \mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T)\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{B}_{j+1}\mathbf{B}_{j+1}^T.$$

Since $\mathcal{P}_{j+1} = \mathcal{P}_j + \mathbf{Z}_j\mathbf{G}_j\mathbf{Z}_j^T$, this provides the inductive step

$$\mathbf{A}\mathcal{P}_{j+1} + \mathcal{P}_{j+1}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{B}_{j+1}\mathbf{B}_{j+1}^T, \quad (4.3)$$

to establish Equation (4.1). Equation (4.2) readily follows from Equation (4.1) since $\mathbf{B}\mathbf{B}^T = -\mathbf{A}\mathcal{P} - \mathcal{P}\mathbf{A}^T$. Equation (4.2) implies $\mathcal{P} \succeq \mathcal{P}_j$ for all j and clearly $\mathcal{P}_{j+1} \succeq \mathcal{P}_j \succeq \mathbf{0}$ since $\mathbf{G}_j$ is positive definite for all j . □

It has been shown that the iteration will provide a sequence of symmetric positive semidefinite matrices $\mathcal{P}_j = \mathbf{L}_j\mathbf{L}_j^T$ which satisfy $\mathcal{P} \succeq \mathcal{P}_{j+1} \succeq \mathcal{P}_j$ for all $j = 1, 2, \dots$. The following lemma will establish that the sequence $\mathcal{P}_j$ is convergent. This lemma is proved in far greater generality for bounded symmetric linear operators in Hilbert space in Riesz and Sz.-Nagy [12](p. 263), where the result is attributed to J.P. Vigier in his Ph.D. thesis [18].

Lemma 4.3 Suppose $\mathcal{P}_\ell$ is a sequence of symmetric matrices such that $\mathcal{P} \succeq \mathcal{P}_{\ell+1} \succeq \mathcal{P}_\ell$, where $\mathcal{P}$ is a fixed symmetric matrix. Then $\lim_{\ell \rightarrow \infty} \mathcal{P}_\ell = \mathcal{P}_o \preceq \mathcal{P}$ exists.

Proof: Let $\rho_{ij}^{(\ell)}$ denote the ij -th element of $\mathcal{P}_\ell$ and let ρ_{ij} denote the ij -th element of $\mathcal{P}$. The definition of the partial ordering $\succeq$ implies that $\mathbf{e}_j^T \mathcal{P} \mathbf{e}_j \geq \mathbf{e}_j^T \mathcal{P}_{\ell+1} \mathbf{e}_j \geq \mathbf{e}_j^T \mathcal{P}_\ell \mathbf{e}_j$ and hence that $\rho_{jj} \geq \rho_{jj}^{(\ell+1)} \geq \rho_{jj}^{(\ell)}$. Thus $\{\rho_{jj}^{(\ell)}\}$ is an increasing sequence bounded above and hence convergent to a limit $\rho_{jj}^{(o)} \leq \rho_{jj}$ for each $j = 1, 2, \dots, n$. Moreover, the fact that $\mu_{ii} + \mu_{jj} \geq 2|\mu_{ij}|$ for any symmetric positive semi-definite matrix $\mathbf{M} = (\mu_{ij})$ will imply that

$$(\rho_{ii}^{(\ell+1)} - \rho_{ii}^{(\ell)}) + (\rho_{jj}^{(\ell+1)} - \rho_{jj}^{(\ell)}) \geq 2|\rho_{ij}^{(\ell+1)} - \rho_{ij}^{(\ell)}|,$$

for all ℓ and for all (i, j) . Observe that

$$\sum_{\ell=1}^{\infty} (\rho_{ii}^{(\ell+1)} - \rho_{ii}^{(\ell)}) + (\rho_{jj}^{(\ell+1)} - \rho_{jj}^{(\ell)}) \leq (\rho_{ii} - \rho_{ii}^{(1)}) + (\rho_{jj} - \rho_{jj}^{(1)})$$

since the series is telescoping. Therefore, the series

$$\sum_{\ell=1}^{\infty} |\rho_{ij}^{(\ell+1)} - \rho_{ij}^{(\ell)}| \leq \frac{1}{2} [(\rho_{ii} - \rho_{ii}^{(1)}) + (\rho_{jj} - \rho_{jj}^{(1)})]$$

is convergent. Now, since

$$|\rho_{ij}^{(m_2)} - \rho_{ij}^{(m_1)}| \leq \sum_{\ell=m_1}^{m_2-1} |\rho_{ij}^{(\ell+1)} - \rho_{ij}^{(\ell)}|$$

for any positive integers $m_1 < m_2$, it follows that the sequence $\{\rho_{ij}^{(\ell)}\}$ is Cauchy and hence convergent for each (i, j) . This concludes the proof. $\square$

Another consequence of Lemma 4.1 is that the iteration is norm decreasing for $\mathbf{B}_j$ in the sense that $\|\mathbf{B}_{j+1}\|_F < \|\mathbf{B}_j\|_F$ for all j .

Lemma 4.4 Let the hypothesis of Lemma 4.1 hold, and assume $\mathbf{B}_{j+1}$ be defined as in Lemma 4.2 and assume $\mathbf{v}^T(\mathbf{A} + \mathbf{A}^T)\mathbf{v} \leq \gamma_{max} < 0$ for all vectors $\mathbf{v}$ of norm one. Then $\|\mathbf{B}_{j+1}\|_F \leq \|\mathbf{B}_j\|_F + \gamma_{max} \text{trace}(\mathbf{Z}_j \mathbf{G}_j \mathbf{Z}_j^T) < \|\mathbf{B}_j\|_F$ for all j .

Proof: Let $\mathbf{Z}_j \mathbf{G}_j \mathbf{Z}_j^T = \mathbf{V}_j \hat{\mathbf{S}}_j \mathbf{V}_j^T$ where $\hat{\mathbf{S}}_j$ is diagonal with positive diagonal elements $\hat{\sigma}_i$ and $\mathbf{V}_j^T \mathbf{V}_j = \mathbf{I}$. From Lemma 4.1,

$$\begin{aligned} \|\mathbf{B}_{j+1}\|_F^2 &= \text{trace}(\mathbf{B}_{j+1} \mathbf{B}_{j+1}^T) \\ &= \text{trace}[\mathbf{A} \mathbf{V}_j \hat{\mathbf{S}}_j \mathbf{V}_j^T + \mathbf{V}_j \hat{\mathbf{S}}_j \mathbf{V}_j^T \mathbf{A}^T] + \text{trace}(\mathbf{B}_j \mathbf{B}_j^T). \end{aligned}$$

However,

$$\begin{aligned}
\text{trace}[\mathbf{A}\mathbf{V}_j\hat{\mathbf{S}}_j\mathbf{V}_j^T + \mathbf{V}_j\hat{\mathbf{S}}_j\mathbf{V}_j^T\mathbf{A}^T] &= \text{trace}[\mathbf{V}_j^T\mathbf{A}\mathbf{V}_j\hat{\mathbf{S}}_j + \hat{\mathbf{S}}_j\mathbf{V}_j^T\mathbf{A}^T\mathbf{V}_j] \\
&= \sum_{i=1}^{k_j} \mathbf{v}_i^T(\mathbf{A} + \mathbf{A}^T)\mathbf{v}_i\hat{\sigma}_i \\
&\leq \gamma_{\max}\text{trace}[\hat{\mathbf{S}}_j] \\
&< 0.
\end{aligned}$$

□

Given the lemmas above, the following results show that the convergence is guaranteed when $\mathbf{B} = \mathbf{b}$, a vector.

Lemma 4.5 Assume $\mathbf{A}\mathbf{X} = \mathbf{X}\Lambda$ is diagonalizable and let the hypothesis of Lemma 4.1 hold. Suppose also that $\mathbf{B} = \mathbf{b}$ is a vector and that $r = 1$ throughout the iteration, i.e., that $\mathbf{U}_j = \mathbf{u}_j$ with $\mathcal{P}_j\mathbf{u}_j = \mathbf{u}_j\sigma_j$, where σ_j is the dominant eigenvalue of $\mathcal{P}_j$. Let $\mathbf{b}_j$ be the right hand side at step j . Then

$$\|\mathbf{b}_{j+1}\| \leq \rho_j \|\mathbf{b}_j\|,$$

where $\rho_j = \rho((\gamma_j\mathbf{I} - \mathbf{A})(\gamma_j\mathbf{I} + \mathbf{A})^{-1})$ is the spectral radius, with $\gamma_j = \frac{1}{2}\mathbf{u}_j^T(\mathbf{A} + \mathbf{A}^T)\mathbf{u}_j < 0$. Moreover, there is a constant $\rho < 1$ such that $\rho_j \leq \rho$ for all j sufficiently large.

Proof: From the definition of the quantities in the iteration, we have

$$\mathbf{A}\mathbf{z}_j + \mathbf{z}_j\gamma_j + \mathbf{b}_j\beta_j = 0 \quad \text{and} \quad \gamma_j\hat{\eta}_j + \hat{\eta}_j\gamma_j + \beta_j^2 = 0,$$

where $\beta_j = \mathbf{u}_j^T\mathbf{b}_j$ and $\eta_j = \hat{\mathcal{P}}_j$ is used to emphasize that $\hat{\mathcal{P}}_j$ is a scalar in this case. Thus

$$\mathbf{z}_j = -(\gamma_j\mathbf{I} + \mathbf{A})^{-1}\mathbf{b}_j\beta_j \quad \text{and} \quad \hat{\eta}_j = -\frac{\beta_j^2}{2\gamma_j},$$

so that

$$\begin{aligned}
\mathbf{b}_{j+1} &= \mathbf{b}_j - \mathbf{z}_j(\hat{\eta}_j)^{-1}\mathbf{u}_j^T\mathbf{b}_j \\
&= \mathbf{b}_j - (\gamma_j\mathbf{I} + \mathbf{A})^{-1}\mathbf{b}_j\beta_j\left(\frac{2\gamma_j}{\beta_j^2}\right)\beta_j \\
&= \mathbf{b}_j - 2\gamma_j(\gamma_j\mathbf{I} + \mathbf{A})^{-1}\mathbf{b}_j \\
&= (\gamma_j\mathbf{I} + \mathbf{A})^{-1}[(\gamma_j\mathbf{I} + \mathbf{A}) - 2\gamma_j\mathbf{I}]\mathbf{b}_j \\
&= -(\gamma_j\mathbf{I} + \mathbf{A})^{-1}(\gamma_j\mathbf{I} - \mathbf{A})\mathbf{b}_j.
\end{aligned}$$

Since $\mathbf{A}$ is stable and $\gamma_j < 0$, the eigenvalues of $\mathbf{A}$ are mapped to the interior of the unit disc under the Cayley transformation $\mathbf{C}_j = (\gamma_j\mathbf{I} - \mathbf{A})(\gamma_j\mathbf{I} + \mathbf{A})^{-1}$. Hence, the spectral radius satisfies $\rho_j = \rho((\gamma_j\mathbf{I} - \mathbf{A})(\gamma_j\mathbf{I} + \mathbf{A})^{-1}) < 1$ for all j . Since $\gamma_j \rightarrow \gamma < 0$ is convergent and since $\rho((\gamma\mathbf{I} - \mathbf{A})(\gamma\mathbf{I} + \mathbf{A})^{-1}) < 1$, there is a number $\rho < 1$ such

that $\rho_j \leq \rho$ for all j sufficiently large. Moreover, a standard argument will show that $\|\mathbf{b}_j\| \leq \rho^{(j-m)}\|\mathbf{b}\|$ for all $j > m$, where m is a positive integer such that $\text{cond}(\mathbf{X}) \prod_{i=1}^m \rho_i < 1$. $\square$

Note that a consequence of this result is that $\mathbf{b}_j \rightarrow \mathbf{0}$ and hence that $\mathcal{P}_j \rightarrow \mathcal{P}$ (i.e., convergence to the true $\mathcal{P}$ at a linear rate).

A more general result will provide a direct relationship to the ADI method. In fact, it can be shown that this parameter free method provides exactly the same update as ADI would construct with shifts given by the eigenvalues of the projected matrix $\mathbf{H}$. The advantage of the parameter free approach is that these complex shifts are automatically selected and they are applied indirectly through the solution of the Sylvester equation, which may be done in real arithmetic.

It is necessary to establish a formula for the update provided by ADI with a set of shifts μ_j , $j = 1, 2, \dots, k$. First, suppose that μ is a shift with $\rho = \text{Real}(\mu) < 0$. Then

$$\mathcal{P} = \mathbf{A}_\mu \mathcal{P} \mathbf{A}_\mu^* + \mathbf{B}_\mu \mathbf{B}_\mu^* \text{ and thus } \mathcal{P} = 2|\rho| \sum_{j=0}^{\infty} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^* (\mathbf{A}_\mu^j)^*,$$

where

$$\mathbf{A}_\mu = (\mathbf{A} - \bar{\mu}\mathbf{I})(\mathbf{A} + \mu\mathbf{I})^{-1}, \quad \mathbf{B}_\mu = (\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{B}.$$

Let $\mathbf{E} = 2|\rho|\mathbf{B}_\mu \mathbf{B}_\mu^*$. Then

$$\begin{aligned} \mathbf{A}(\mathcal{P} - \mathbf{E}) + (\mathcal{P} - \mathbf{E})\mathbf{A}^T &= 2|\rho| \left[\mathbf{A} \left(\sum_{j=1}^{\infty} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^T (\mathbf{A}_\mu^j)^* \right) + \left(\sum_{j=1}^{\infty} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^T (\mathbf{A}_\mu^j)^* \right) \mathbf{A}^T \right] \\ &= 2|\rho| \mathbf{A}_\mu \left[\mathbf{A} \left(\sum_{j=0}^{\infty} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^T (\mathbf{A}_\mu^j)^* \right) + \left(\sum_{j=0}^{\infty} \mathbf{A}_\mu^j \mathbf{B}_\mu \mathbf{B}_\mu^T (\mathbf{A}_\mu^j)^* \right) \mathbf{A}^T \right] \mathbf{A}_\mu^* \\ &= \mathbf{A}_\mu [\mathbf{A}\mathcal{P} + \mathcal{P}\mathbf{A}^T] \mathbf{A}_\mu^* \\ &= -\mathbf{A}_\mu [\mathbf{B}\mathbf{B}^T] \mathbf{A}_\mu^*. \end{aligned}$$

Thus,

$$\mathbf{A}(\mathcal{P} - \mathbf{E}) + (\mathcal{P} - \mathbf{E})\mathbf{A}^T + \mathbf{A}_\mu \mathbf{B}\mathbf{B}^T \mathbf{A}_\mu^* = \mathbf{0}.$$

Now, define

$$\mathbf{B}_{\mu_i} \equiv \left(\prod_{\ell=1}^i \mathbf{A}_{\mu_\ell} \right) \mathbf{B}, \quad \mathbf{B}^{(i)} \equiv \sqrt{2|\rho_i|} (\mathbf{A} + \mu_i \mathbf{I})^{-1} \mathbf{B}_{\mu_{i-1}}, \quad \text{and} \quad \mathbf{E}_i \equiv \mathbf{B}^{(i)} \mathbf{B}^{(i)*}, \quad (4.4)$$

with $\mathbf{B}_{\mu_0} \equiv \mathbf{B}$. Then

$$\mathbf{A}(\mathcal{P} - \mathcal{E}_i) + (\mathcal{P} - \mathcal{E}_i)\mathbf{A}^T + \mathbf{B}_{\mu_i} \mathbf{B}_{\mu_i}^* = \mathbf{0}, \quad \text{for } i = 1, 2, \dots, k, \quad (4.5)$$

where $\mathcal{E}_i \equiv \mathbf{E}_1 + \mathbf{E}_2 + \dots + \mathbf{E}_i$.

The discussion leading to Equations (4.4),(4.5) are valid for arbitrary $\mathbf{B}$. However, when $\mathbf{B}$ is a vector it is possible to establish a direct relationship with the ADI method.

Lemma 4.6 Suppose that $\mathbf{B} = \mathbf{b}$ is a vector. Let $\mathbf{H}_j = \mathbf{U}_j^T \mathbf{A} \mathbf{U}_j \in \mathbb{R}^{k \times k}$ and $\hat{\mathbf{B}}_j = \mathbf{U}_j^T \mathbf{B}_j$ where $\mathbf{B}_j$ is the right hand side at Step j . Then the update $\mathcal{P}_{j+1} = \mathcal{P}_j + \mathbf{Z}_j \hat{\mathcal{P}}^{-1} \mathbf{Z}_j^T$ is precisely the same as the update that would be obtained by applying k steps of ADI with shifts $\{\mu_1, \mu_2, \dots, \mu_k\} = \sigma(\mathbf{H}_j)$ (the spectrum of $\mathbf{H}_j$).

Proof: From Equations (4.4),(4.5), the update obtained by applying shifts $\{\mu_1, \mu_2, \dots, \mu_k\} = \sigma(\mathbf{H}_j)$ is of the form $\mathbf{L}_j \mathbf{L}_j^T$, where

$$\mathbf{L}_j = [\mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \dots, \mathbf{B}^{(k)}], \text{ with } \mathbf{B}^{(i)} \equiv \sqrt{2\rho_i}(\mathbf{A} + \mu_j \mathbf{I})^{-1} \left(\prod_{\ell=1}^{i-1} \mathbf{A}_{\mu_\ell} \right) \mathbf{B}_j,$$

and with $\rho_i = |\text{Real}(\mu_i)|$.

Now, regardless of how the solution to $\mathbf{A}\mathbf{Z} + \mathbf{Z}\mathbf{H}_j^T + \mathbf{B}_j \hat{\mathbf{B}}_j^T = \mathbf{0}$ is obtained, it is mathematically equivalent to the solution obtained by the ADI method. This solution is given by $\mathbf{Z}_j = \mathbf{L}_j \hat{\mathbf{L}}_j^T$ where

$$\hat{\mathbf{B}}^{(i)} \equiv \sqrt{2\rho_i}(\hat{\mathbf{H}}_j + \mu_j \mathbf{I})^{-1} \left(\prod_{\ell=1}^{i-1} \hat{\mathbf{H}}_{\mu_\ell} \right) \hat{\mathbf{B}}_j \text{ and } \hat{\mathbf{L}}_j = [\hat{\mathbf{B}}^{(1)}, \hat{\mathbf{B}}^{(2)}, \dots, \hat{\mathbf{B}}^{(k)}].$$

Moreover, $\hat{\mathcal{P}}_j = \hat{\mathbf{L}}_j \hat{\mathbf{L}}_j^T$ is the solution to

$$\mathbf{H}_j \hat{\mathcal{P}} + \hat{\mathcal{P}} \mathbf{H}_j^T + \hat{\mathbf{B}}_j \hat{\mathbf{B}}_j^T = \mathbf{0}.$$

Thus

$$\mathbf{Z}_j \hat{\mathcal{P}}^{-1} \mathbf{Z}_j^T = \mathbf{L}_j \mathbf{L}_j^T,$$

which is the same as would have been obtained by the ADI method with shifts $\{\mu_1, \mu_2, \dots, \mu_k\}$. $\square$

This result, together with the fact that $\mathcal{P}_j$ is convergent, can be used to establish a rate of convergence involving the spectral radius $\rho(\prod_{i=1}^k \mathbf{A}_{\mu_i})$, where the $\{\mu_i\}$ are the limits of the eigenvalues of $\mathbf{H}_j$. Unfortunately, the connection of the parameter free iteration with ADI is not so straightforward in the case that $\mathbf{B}$ has $m > 1$ columns, and a rate of convergence has not yet been established for this case.

5 Implementation Details

Computing both the update $\mathbf{Z}\mathbf{G}\mathbf{Z}^T$ and the modification $\mathbf{B} \leftarrow (\mathbf{I} - \mathbf{Z}\mathbf{G}\mathbf{U}^T)\mathbf{B}$ involves some delicate numerical issues requiring care to avoid numerical instabilities. The problem arises when $\hat{\mathcal{P}}$ becomes ill-conditioned.

Recall that $\hat{\mathcal{P}}$ is obtained from the equation $\mathbf{H}\hat{\mathcal{P}} + \hat{\mathcal{P}}\mathbf{H}^T + \hat{\mathbf{B}}\hat{\mathbf{B}}^T = \mathbf{0}$. The convergence of $\mathcal{P}_j$ may be used to show that $\mathbf{Z}\mathbf{G}\mathbf{U}^T\mathbf{B}$ must be well behaved in exact arithmetic even though $\mathbf{G} = \hat{\mathcal{P}}^{-1}$ becomes arbitrarily large in norm as the iteration progresses. However, in finite precision this term can be problematic and must be handled with care.

One way to deal with this term is to control the conditioning of $\hat{\mathcal{P}}$. The eigensystem $\hat{\mathcal{P}} = \mathbf{Q}\hat{\mathbf{S}}\mathbf{Q}^T$, with $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \dots \geq \hat{\sigma}_m$ being the diagonal elements of $\hat{\mathbf{S}}$, may be used. The strategy is to maintain a uniformly bounded condition number for $\hat{\mathcal{P}}$ via truncation. Namely, if $\hat{\sigma}_k \geq \tau \hat{\sigma}_1 \geq \hat{\sigma}_{k+1}$ then the search subspace is reduced to order k using the leading k columns of $\mathbf{Q}$. This is done prior to solving for $\mathbf{Z}$ at Step 2.

Step 1: Solve the reduced order Lyapunov equation for $\hat{\mathcal{P}}_j$

i) Solve $\mathbf{H}_j \hat{\mathcal{P}}_j + \hat{\mathcal{P}}_j \mathbf{H}_j^T + \hat{\mathbf{B}}_j \hat{\mathbf{B}}_j^T = \mathbf{0}$.

ii) Reduce size of subspace so that $\text{cond}(\hat{\mathcal{P}}_j) < \frac{1}{\tau}$.

Step 2: (APM step) Solve a projected Sylvester equation for $\mathbf{Z}_j$

$$\mathbf{A} \mathbf{Z}_j + \mathbf{Z}_j \mathbf{H}_j^T + \mathbf{B}_j \hat{\mathbf{B}}_j^T = \mathbf{0}, \text{ with } \mathbf{H}_j = \mathbf{U}_j^T \mathbf{A} \mathbf{U}_j, \hat{\mathbf{B}}_j = \mathbf{U}_j^T \mathbf{B}_j.$$

Step 3: Modify $\mathbf{B}_j$

$$\text{Update } \mathbf{B}_{j+1} \leftarrow (\mathbf{I} - \mathbf{Z}_j \hat{\mathcal{P}}_j^{-1} \mathbf{U}_j^T) \mathbf{B}_j.$$

Step 4: (ADI step) Update factorization and basis $\mathbf{U}$

$$\text{Re-scale } \mathbf{Z}_j \leftarrow \mathbf{Z}_j \hat{\mathcal{P}}_j^{-1/2}.$$

$$\text{Update (and truncate) } [\mathbf{U}, \mathbf{S}] \leftarrow \text{svd}[\mathbf{L}_j, \mathbf{Z}_j].$$

$$\mathbf{L}_{j+1} = \mathbf{U} \mathbf{S}.$$

$$\mathbf{U}_{j+1} \leftarrow \mathbf{U}(:, 1 : k_j), \text{ basis for dominant subspace of dimension } k_j.$$

Figure 2: A Parameter Free Lyapunov Solver

5.1 Direct Solution of the Projected Sylvester Equation

Computing the update $\mathbf{Z}$ requires solution of the projected Sylvester equation

$$\mathbf{A} \mathbf{Z} + \mathbf{Z} \mathbf{H}^T + \mathbf{B} \hat{\mathbf{B}}^T = \mathbf{0}, \text{ where } \mathbf{H} = \mathbf{U}^T \mathbf{A} \mathbf{U} \text{ and } \hat{\mathbf{B}} = \mathbf{U}^T \mathbf{B}. \quad (5.1)$$

A minor variant of the Bartels–Stewart algorithm may be used. If $\mathbf{H}^T = \mathbf{Q} \mathbf{R} \mathbf{Q}^*$ is a Schur decomposition of $\mathbf{H}^T$, then Equation (5.1) transforms to

$$\mathbf{A} \tilde{\mathbf{Z}} + \tilde{\mathbf{Z}} \mathbf{R} + \mathbf{B} \tilde{\mathbf{B}}^T = \mathbf{0}, \text{ where } \tilde{\mathbf{Z}} = \mathbf{Z} \mathbf{Q} \text{ and } \tilde{\mathbf{B}} = \mathbf{Q}^* \hat{\mathbf{B}}.$$

Denote the elements of $\mathbf{R}$ as $\rho_{i,j}$ and the columns of $\tilde{\mathbf{Z}}$ as $\mathbf{z}_j$. Then this may be solved as follows:

for $j = 1 : k$,

$$\text{Solve } (\mathbf{A} + \rho_{jj} \mathbf{I}) \mathbf{z}_j = -\mathbf{B} \tilde{\mathbf{B}}^T \mathbf{e}_j - \sum_{i=1}^{j-1} \mathbf{z}_i \rho_{i,j};$$

end

$$\text{Put } \mathbf{Z} = \tilde{\mathbf{Z}} \mathbf{Q}^*.$$

This scheme was essentially developed in [5], where just one of the two matrices in the Sylvester equation is reduced all the way to Schur form. The idea was also suggested in [17] for the case of large sparse $\mathbf{A}$ and small dense $\mathbf{H}$.

This approach has the advantage that the assumption that $\mathbf{A} + \mathbf{A}^T$ is negative definite is no longer needed. The algorithm is valid regardless of the stability of $\mathbf{H}$ so long as $\sigma(\mathbf{A}) \cap \sigma(-\mathbf{H}) = \emptyset$. On the other hand, each step requires the sparse direct factorization of the matrix $\mathbf{A} - \rho_{jj}\mathbf{I}$, and this must be done in complex arithmetic whenever an eigenvalue ρ_{jj} of $\mathbf{H}$ is complex. In any event, with this approach to solving the Sylvester equation, the entire algorithm requires no choice of parameters.

5.2 Iterative Solution of the Sylvester Equation in Real Arithmetic

It is possible to completely avoid multiple sparse matrix factorizations in complex arithmetic through an invariant subspace approach. An alternative is to use the fact that $\mathbf{Z} = \mathbf{X}\mathbf{Y}^{-1}$ where

$$\begin{bmatrix} \mathbf{A} & \mathbf{M} \\ \mathbf{0} & -\mathbf{H}^T \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \hat{\mathbf{R}},$$

and where $\mathbf{M} = \mathbf{B}\hat{\mathbf{B}}^T$ with $\sigma(\hat{\mathbf{R}}) = \sigma(-\mathbf{H}^T)$. One can show that $\mathbf{Y}$ is nonsingular if and only if $\sigma(\hat{\mathbf{R}}) = \sigma(-\mathbf{H}^T)$. In this approach, we assume that $\mathbf{H}$ is stable. However, this may not always be the case if $\mathbf{A} + \mathbf{A}^T$ is not negative definite. Thus, in the situation where $\mathbf{H}$ is not stable, we must deflate the unwanted subspace out of $\mathbf{H}$ and this can be done as follows: Compute a Schur decomposition of $\mathbf{H}$ as in

$$\mathbf{H} \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix},$$

where $\mathbf{R}_{11}$ is strictly stable and $\mathbf{R}_{22}$ anti-stable. Now update $\mathbf{H}$ and $\mathbf{U}$ as follows:

$$\begin{aligned} \mathbf{H} &\leftarrow \mathbf{R}_{11} \\ \mathbf{U} &\leftarrow \mathbf{U}\mathbf{Q}_1. \end{aligned}$$

To motivate these updates, recall that $\mathbf{H} = \mathbf{U}^T \mathbf{A} \mathbf{U}$. From the Schur decomposition,

$$\mathbf{H}\mathbf{Q}_1 = \mathbf{Q}_1\mathbf{R}_{11}.$$

Therefore,

$$\mathbf{R}_{11} = \mathbf{Q}_1^T \mathbf{H} \mathbf{Q}_1 = (\mathbf{U}\mathbf{Q}_1)^T \mathbf{A} (\mathbf{U}\mathbf{Q}_1).$$

After the deflation, $\mathbf{H}$ is guaranteed stable although the size k of the search subspace may have been reduced.

Now, define

$$\tilde{\mathbf{A}} \equiv \begin{bmatrix} \mathbf{A} & \mathbf{M} \\ \mathbf{0} & -\mathbf{H}^T \end{bmatrix},$$

and consider the Cayley transformation

$$\mathbf{C}_\mu \equiv (\mu\mathbf{I} - \tilde{\mathbf{A}})^{-1}(\mu\mathbf{I} + \tilde{\mathbf{A}}), \quad (5.2)$$

where μ is a real positive number. Under this transformation, an eigenvalue λ of $\tilde{\mathbf{A}}$ is mapped to an eigenvalue $\omega = \frac{\mu+\lambda}{\mu-\lambda}$ of $\mathbf{C}_\mu$. These eigenvalues have the same eigenvector. The eigenvalues of $\tilde{\mathbf{A}}$ consist of $\sigma(\tilde{\mathbf{A}}) = \sigma(\mathbf{A}) \cup \sigma(-\mathbf{H}^T)$. Since both $\mathbf{A}$ and $\mathbf{H}$ are stable, the eigenvalues of $\mathbf{A}$ are mapped strictly to the interior of the unit

disc, while the eigenvalues of $-\mathbf{H}^T$ are mapped strictly to the exterior of the unit disc. An iterative method (such as the implicitly restarted Arnoldi method [16] available in ARPACK [8]) may be used to rapidly compute the k eigenvalues of largest magnitude for $\mathbf{C}_\mu$. These will be the transformed eigenvalues of $-\mathbf{H}^T$ which are the k eigenvalues that are exterior to the unit disk. We emphasize that the point of this calculation is to construct a basis corresponding to the invariant subspace corresponding to the *known eigenvalues* of $\tilde{\mathbf{A}}$ (i.e., the eigenvalues of $-\mathbf{H}^T$).

Repeated application of the Cayley transformed matrix to vectors requires only a single sparse direct factorization which can be done in real arithmetic. The parameter μ may be chosen to enhance convergence of the iterative eigenvalue method. If a sparse direct factorization is too expensive, an iterative approach can be used to apply the Cayley transformation to approximate a matrix-vector product with $\mathbf{C}_\mu$.

As indicated by the analysis of polynomial restart Krylov methods given in [1], convergence to the k eigenvalues outside the unit disk will be optimally enhanced if these transformed eigenvalues are distanced as far as possible from the unit circle. That is, an optimal convergence rate will be achieved if we are able to choose a real shift μ in the Cayley transformation that is constructed to maximize the minimum distance of these k eigenvalues to the unit circle. The following discussion will explain how this is done.

Lemma 5.1 *Let $z = a + ib \in \mathbb{C}$ with $a \neq 0$ be a fixed complex number and let μ vary in $(0, \infty)$. Consider the following Cayley transformation*

$$w(\mu) = \frac{\mu + z}{\mu - z}. \quad (5.3)$$

1. *If $b \neq 0$, then (5.3) maps $(0, \infty)$ onto a circular arc $\mathcal{C}$ centered at $i\frac{a}{b}$ with radius $r = \sqrt{1 + (\frac{a}{b})^2}$. In addition, $w \rightarrow -1$ as $\mu \rightarrow 0^+$ and $w \rightarrow 1$ as $\mu \rightarrow \infty$.*
2. *If $b = 0$, i.e., $z = a \in \mathbb{R}$, there are the following two cases:*
 - (a) *If $a > 0$, then (5.3) maps $(0, a)$ onto $(-\infty, -1)$ with $\lim_{\mu \rightarrow 0^+} w = -1$ and $\lim_{\mu \rightarrow a^-} w = -\infty$, and (a, ∞) onto $(1, \infty)$ with $\lim_{\mu \rightarrow a^+} w = \infty$ and $\lim_{\mu \rightarrow \infty} w = 1$.*
 - (b) *If $a < 0$, then (5.3) maps $(0, \infty)$ onto $(-1, 1)$ with $\lim_{\mu \rightarrow 0^+} w = -1$ and $\lim_{\mu \rightarrow \infty} w = 1$.*

Proof:

1. Let $b \neq 0$. It is well known that w maps the real line to a circle. If this circle is centered at $w_0 \in \mathbb{C}$ with radius r , then $|w(\mu) - w_0|^2 = r^2$ for any real value μ . It is easily seen that $w(0) = -1$ and $w(\infty) = 1$. If $w_0 = c + id$, it follows that

$$r^2 = |w(0) - w_0|^2 = (1 + c)^2 + d^2 \quad \text{and} \quad r^2 = |w(\infty) - w_0|^2 = (1 - c)^2 + d^2.$$

This implies $0 = (1 + c)^2 - (1 - c)^2 = 4c$ so that $c = 0$ and $r^2 = 1 + d^2$. Thus, the image of w is a circle centered on the imaginary axis at $w_0 = id$ with radius $r = \sqrt{1 + d^2}$. It remains to show that $d = \frac{a}{b}$.

Let $z = a + ib$. Then for μ real,

$$w(\mu) = \frac{\mu + z}{\mu - z} = \frac{\mu^2 - a^2 - b^2}{(\mu - a)^2 + b^2} + i \frac{2b\mu}{(\mu - a)^2 + b^2} \quad (5.4)$$

The imaginary axis bisects the circle and intersects it at the purely imaginary points $w(\mu_0)$ and $w(-\mu_0)$ with $\mu_0 = \sqrt{a^2 + b^2} = |z|$. The center id must satisfy $id = \frac{1}{2}(w(\mu_0) + w(-\mu_0))$ and then (5.4) implies

$$2d = \frac{2b\mu_0}{(\mu_0 - a)^2 + b^2} - \frac{2b\mu_0}{(\mu_0 + a)^2 + b^2} = (2b\mu_0) \left[\frac{1}{2\mu_0(\mu_0 - a)} - \frac{1}{2\mu_0(\mu_0 + a)} \right] \quad (5.5)$$

$$= b \left[\frac{1}{\mu_0 - a} - \frac{1}{\mu_0 + a} \right] = \frac{2ab}{\mu_0^2 - a^2} = 2\frac{a}{b}, \quad (5.6)$$

where $b^2 = (\mu_0 + a)(\mu_0 - a) = \mu_0^2 - a^2$ is used in the denominators.

This argument shows the Cayley transformation in (5.3) maps $(0, \infty)$ onto a circular arc $\mathcal{C}$ centered at $w_0 = i\frac{a}{b}$ with radius $r = \sqrt{1 + (\frac{a}{b})^2}$. It is straightforward to verify that $w \rightarrow -1$ as $\mu \rightarrow 0^+$ and that $w \rightarrow 1$ as $\mu \rightarrow \infty$.

2. For $b = 0$, the Cayley transformation in (5.3) is a differentiable real-valued function in μ on the specified intervals. Thus, it is straightforward to verify the claims.

□

The monotonicity of $|w(\mu)|$ is established in the following lemma.

Lemma 5.2 *Let $z = a + ib \in \mathbb{C}$ with $a \neq 0$ be a fixed complex number and let μ vary in $(0, \infty)$. Consider the Cayley transformation in (5.3).*

1. For $b \neq 0$, there are two cases:

- (a) *If $a > 0$, then $|w|$ attains its maximum value of $\sqrt{1 + (\frac{a}{b})^2} + |\frac{a}{b}|$ at $\mu_0 = |z| = \sqrt{a^2 + b^2}$. In addition, the function $|w|$ monotonically increases from 1 to its maximum value in $(0, \mu_0)$ and monotonically decreases from its maximum value down to 1 in (μ_0, ∞) .*
- (b) *If $a < 0$, then $|w|$ attains its minimum value of $\sqrt{1 + (\frac{a}{b})^2} - |\frac{a}{b}|$ at $\mu_0 = |z| = \sqrt{a^2 + b^2}$. In addition, the function $|w|$ monotonically decreases from 1 down to its minimum value in $(0, \mu_0)$ and monotonically increases from its minimum value to 1 in (μ_0, ∞) .*

2. For $b = 0$, there are two cases:

- (a) *If $a > 0$, then $\lim_{\mu \rightarrow a} |w| = \infty$. In addition, the function $|w|$ monotonically increases from 1 to ∞ in $(0, a)$ and monotonically decreases from ∞ down to 1 in (a, ∞) .*
- (b) *If $a < 0$, then $|w|$ attains its minimum value of 0 at $\mu_0 = |z| = |a| = -a$. In addition, the function $|w|$ monotonically decreases from 1 down to its minimum value in $(0, -a)$ and monotonically increases from its minimum value to 1 in $(-a, \infty)$.*

Proof: When $b \neq 0$, the proof of Lemma 5.1 showed the image $w(\mu)$ of the real line $-\infty < \mu < \infty$ is a circle bisected by the imaginary axis with $\mu = \mu_0 = |z| = \sqrt{a^2 + b^2}$, giving the point of intersection for the image of $\mu \in (0, \infty)$. This point gives either a maximum or a minimum of $|w(\mu)|$ depending on the sign of a , as established in the following discussion.

Since $|w| > 0$, studying the monotonicity of $|w|$ is equivalent to studying that of $|w|^2$. Consider

$$|w|^2 = \bar{w}w = \frac{(\mu^2 - a^2 - b^2)^2 + (2\mu b)^2}{((\mu - a)^2 + b^2)^2}.$$

It is straightforward to verify the following:

For $a > 0$,

$$\frac{d}{d\mu}|w|^2 > 0 \text{ for } \mu \in (0, \mu_0) \quad \text{and} \quad \frac{d}{d\mu}|w|^2 < 0 \text{ for } \mu \in (\mu_0, \infty). \quad (5.7)$$

For $a < 0$,

$$\frac{d}{d\mu}|w|^2 < 0 \text{ for } \mu \in (0, \mu_0) \quad \text{and} \quad \frac{d}{d\mu}|w|^2 > 0 \text{ for } \mu \in (\mu_0, \infty). \quad (5.8)$$

Claims 1(a) and 1(b) are established with these results.

When $b = 0$, Conditions (5.7) and (5.8) are still valid and establish monotonicity. The limiting values of $|w(\mu)|$ as $\mu \rightarrow \mu_0$ can be verified directly from the formula for $w(\mu)$ to obtain Claims 2(a) and 2(b). $\square$

Our objective is to enhance convergence of the iterative eigenvalue solver while computing an invariant subspace of $\mathbf{C}_\mu$ corresponding to the k eigenvalues of largest magnitude. This is accomplished by selecting an optimal choice for the parameter μ in the Cayley transformation (5.2). With this selection of μ , the desired eigenvalues are separated as much as possible from the closed unit disk which contains all of the unwanted eigenvalues of $\mathbf{C}_\mu$.

Lemmas 5.1 and 5.2 lead to the following result, which shows that the optimal μ lies in an interval determined by eigenvalues of largest and smallest magnitude.

Lemma 5.3 *If z_s and z_l are eigenvalues of $\mathbf{H}$ with smallest and largest magnitude respectively, then the optimal value of μ lies in the interval $[|z_s|, |z_l|]$ and solves*

$$\max_{\mu \in [|z_s|, |z_l|]} \min_{z \in \sigma(-\mathbf{H}^T)} \left| \frac{\mu + z}{\mu - z} \right|. \quad (5.9)$$

Proof: Define

$$w(\mu; z) \equiv \frac{\mu + z}{\mu - z}.$$

Since $\mathbf{H}$ is stable, all eigenvalues of $-\mathbf{H}^T$ have positive real part and Lemma 5.2 implies $|w(\mu; z)|$ achieves its maximum at $\mu = |z| \in [|z_s|, |z_l|]$ for each $z \in \sigma(-\mathbf{H}^T)$. Moreover, $|w(\mu; z)|$ is strictly increasing on $(0, |z|)$ and strictly decreasing on $(|z|, \infty)$.

Suppose $\mu < |z_s|$. Let $\hat{z} \in \sigma(-\mathbf{H}^T)$ satisfy $|w(|z_s|; \hat{z})| \leq |w(|z_s|; z)|$ for all $z \in \sigma(-\mathbf{H}^T)$. Then

$$\max_{\eta \in [|z_s|, |z_l|]} \min_{z \in \sigma(-\mathbf{H}^T)} |w(\eta; z)| \geq |w(|z_s|; \hat{z})| > |w(\mu; \hat{z})|,$$

since $|w(\mu; \hat{z})|$ is strictly increasing on $(0, |z_s|)$. A similar argument applies to the right endpoint $|z_l|$. $\square$

In [10] we provide an efficient algorithm to solve (5.9) in $\mathcal{O}(|\sigma(\mathbf{H})|)$ steps, where $|\sigma(\mathbf{H})|$ is the cardinality of $\sigma(\mathbf{H})$. The development of this algorithm is too lengthy and tedious to include here. However, we will illustrate how the algorithm essentially works with a simple example. This example consists of a matrix $\mathbf{H}$ with a spectrum of only three complex pairs: $-2.5 \pm i$, $-4.5 \pm 4i$ and $-8 \pm 4i$. Since $|w(\mu, z)| = |w(\mu, \bar{z})|$ for any real μ , we can replace $\sigma(-\mathbf{H}^T)$ by $\mathcal{S} = \{z_1, z_2, z_3\} = \{2.5 + i, 4.5 + 4i, 8 + 4i\}$. In Figure 3(a), the modulus of w in (5.4) is plotted for each of these three eigenvalues z_j as μ varies. Most of the features that determine the steps to find the optimal μ occur in the region inside the rectangular which can be seen in Figure 3(b). Let w_j denote the curve $w(z, \mu_j)$, $j = 1, 2, 3$. The algorithm first computes the intersection of $|w_1|$ and $|w_3|$, which is at $\mu_1 = 5.0812$ and denoted by a circle at position 1 in Figure 3(b). Next, the algorithm looks straight down and sees a smaller value on $|w_2|$ at position 2. Since the signs of the derivatives of $|w_2|$ and $|w_3|$ at μ_1 are the same, the algorithm replaces $|w_3|$ by $|w_2|$. It then starts over by computing the intersection of $|w_1|$ and $|w_2|$, which is at $\mu_2 = 5.3852$ and denoted by a circle at position 3 in Figure 3(b). As it looks down and sees no smaller value than its current value, the algorithm returns $\mu_2 = 5.3852$ as the optimal shift.

Figure 3: A demonstration on how to solve (5.9). Most of the activities of computing the optimal shift occur inside the rectangular in (a), which can be seen in (b).

For more details and convergence analysis of the algorithm, see [10]. We want to mention that, in general at each step of the algorithm, there are only a few function evaluations and comparisons. In addition, in practice, it takes no more than a handful of steps for the algorithm to return the optimal value of μ . Hence, the total computational cost at the termination of the algorithm is insignificant. Given that an optimal value for the parameter μ is computed, the eigenvalue problem can be solved effectively for an invariant subspace as indicated. Then the solution of the projected Sylvester equation is obtained.

Another approach to solve the projected Sylvester equation is to use the ADI method. It can be shown that selecting an optimal value for a single real shift in the ADI method for the projected Sylvester equation is equivalent to selecting an optimal value for the parameter μ for the iterative eigenvalue solver as presented above. Essentially the same analysis can be carried out for computing an optimal value for a single real shift in the ADI method. Moreover, this analysis can be generalized to choose an optimal set of multiple real shifts for the ADI method for

the Sylvester equation [10].

5.3 Stopping Rules

The ultimate goal of the algorithm is to construct a basis for the dominant invariant subspace of the solution of the Lyapunov equation. To achieve this goal, the algorithm uses an approximate power step at each iteration to obtain a basis update. Thus, the algorithm will halt when the updates become insignificant relative to a prescribed tolerance tol . From the analysis, at each iteration, the term $\mathbf{Z}_j \mathbf{Z}_j^T$ (with an updated $\mathbf{Z}_j$) provides the absolute error between the current and preceding approximations of $\mathcal{P}$. A practical consequence is that the algorithm may be stopped if in Step 4 in Figure 2,

$$\|\mathbf{Z}_j\|_2 \leq \sqrt{\text{tol}} \|\mathbf{L}_j, \mathbf{Z}_j\|_2. \quad (5.10)$$

This is implemented as follows: At iteration j in Figure 2, with $\hat{\mathcal{P}}_j$ obtained in its Cholesky factors $\hat{\mathcal{P}}_j = \mathbf{R}_j \mathbf{R}_j^T$ from Step 1 and $\mathbf{Z}_j$ from Step 2, the singular values of the rescaled $\mathbf{Z}_j$ in Step 4 are available via the generalized singular value decomposition of $[\mathbf{Z}_j, \mathbf{R}_j]$. Let these singular values reside in Σ . With $[\mathbf{U}, \mathbf{S}] \leftarrow \text{svd}[\mathbf{L}_j, \mathbf{Z}_j]$, then (5.10) becomes

$$\|\Sigma\|_2 \leq \sqrt{\text{tol}} \|\mathbf{S}\|_2. \quad (5.11)$$

There is no cost associated with making this stopping test, as all the required quantities are readily available.

In addition, Lemma 4.2 shows that $\mathbf{B}_j \mathbf{B}_j^T$ is the Lyapunov equation residual at iteration j . Thus, to test for convergence, the residual norm is also monitored and is very cheap to compute at each iteration. This residual norm will be monotone decreasing due to Lemma 4.4. The residual norm stopping rule is

$$\|\mathbf{B}_j\|_2 \leq \sqrt{\text{tol}} \|\mathbf{B}\|_2. \quad (5.12)$$

6 Numerical Demonstrations

We apply the algorithm in Figure 2 to the following three systems: a system whose state matrix is obtained from a streamlined upwind Petrov-Galerkin (SUPG) discretization of the 2-dimensional steady-state advection-diffusion equation with Dirichlet boundary conditions studied in [4], a randomly generated power grid studied in [20] and an atmospheric storm track model documented in [2] as part of a collection of benchmark examples for model reduction of linear time invariant dynamical systems. We use the eigenvalue problem approach to compute the solution of the projected Sylvester equations for the basis updates. Note that all of these three systems are non-normal. Within the three, only the SUPG system satisfies the assumption of $\mathbf{A} + \mathbf{A}^T$ being negative definite. The numerical ranges of the other two systems do cross over the right half-plane. However, during the computation, we observe that the projected matrices $\mathbf{H}_j$ are always stable. In case an indefinite matrix $\mathbf{H}_j$ is encountered during the computation, a deflation step as presented earlier should be sufficient treatment as long as the state matrix $\mathbf{A}$ is mildly non-normal.

In Figure 4, an SUPG system of size 400 is considered. The dimension k of the search subspace is 20 throughout. After 8 iterations, the parameter free ADI-like algorithm returns an approximate solution with a relative residual norm $\|\mathbf{B}_j\|/\|\mathbf{B}\|$ of $\mathcal{O}(10^{-8})$.

Figure 4: The spectrum and convergence history of the SUPG system using the PFADI method

Figure 5: The spectrum and convergence history of the power grid system using the PFADI method

In Figure 5, a randomly generated power grid of size 276 is considered. The dimension k of the search subspace is 20 through out. After 23 iterations, the parameter free ADI-like algorithm returns an approximate solution with a relative residual norm $\|\mathbf{B}_j\|/\|\mathbf{B}\|$ of $\mathcal{O}(10^{-6})$.

The atmospheric storm track model of size 598 is considered in Figure 6. We also use the same size of 20 for the search subspace dimension. After 11 iterations, the parameter free ADI-like algorithm returns an approximate solution with a relative residual norm $\|\mathbf{B}_j\|/\|\mathbf{B}\|$ of $\mathcal{O}(10^{-5})$.

7 Concluding Remarks

We have proposed a parameter-free algorithm to approximately compute the solution of a large-scale Lyapunov equation in low-rank factored form. The algorithm is based upon a synthesis of an approximate power method (APM) and an alternating direction implicit (ADI) method. Specifically, it uses an APM iteration to obtain a

Figure 6: The spectrum and convergence history of the storm track system using the PFADI method

basis update and then constructs a re-weighting of this basis update to provide a factorization update that satisfies ADI-like convergence properties. We have shown that when $\mathbf{B}$ is a vector, this parameter free method provides exactly the same update as the ADI method would construct with shifts given by the eigenvalues of a projected matrix. Thus, the advantage of the method is that the ADI convergence properties are guaranteed without the need to select shifts. Shift selection is done automatically via solving a projected Sylvester equation and can be done in real arithmetic effectively and optimally as demonstrated. Regarding the stopping rules, all the needed components are readily available in the algorithm. Thus, no major extra computation is required.

Note that in order to control the growth in size of $\mathbf{L}$, the SVD of $\mathbf{L}$ is updated and truncated toward certain accuracy in Step 4. However, in some cases, this may cause stagnation to occur in the algorithm if the accuracy required for the algorithm to halt is finer than that required by the SVD truncation. We recommend these two accuracy requirements be compatible.

An immediate step from this parameter free algorithm is to develop a parameter-free approximate balancing algorithm for model reduction of large-scale LTI dynamical systems where the two controllability and observability Lyapunov equations are solved simultaneously. The desired projection matrices for the balanced truncation model reduction then can be constructed using the two dominant invariant subspaces of the two Gramians. This algorithm will be presented in our future work.

8 Acknowledgment

The authors wish to thank Professor Mark Embree for several helpful discussions and comments.

References

- [1] C. A. Beattie, M. Embree, D.C. Sorensen, Convergence of Polynomial Restart Krylov Methods for Eigenvalue Computation, *SIAM Review*, **47**, (3), pp. 492-515, (2005).

- [2] Y. Chahlaoui and P. van Dooren, *A collection of benchmark examples for model reduction of linear time invariant dynamical systems*, SLICOT Working Note 2002-2, (2002).
- [3] N.S. Ellner and E.L. Wachspress, *Alternating direction implicit iteration for systems with complex spectra*, SIAM J. Num. Anal. **28**(3): 859-870, (1991)
- [4] B. Fischer, A. Ramage, D. Silvester and A. Wathen, *Towards parameter-free streamline upwinding for advection-diffusion problems*, Comp. Methods Appl. Mech. Eng., **179**: 185-202 (1999).
- [5] G.H. Golub, S. Nash, and C. Van Loan, *A Hessenberg-Schur method for the problem $\mathbf{AX} + \mathbf{XB} = \mathbf{C}$* , IEEE Transactions on Automatic Control, **24**: 909-913, (1979).
- [6] S. Gugercin, D. C. Sorensen, and A. C. Antoulas, *A modified low-rank Smith method for large-scale Lyapunov equations*, Numerical Algorithms, **32**: Issue 1, pp. 27-55, Jan. 2003.
- [7] A.S. Hodel, B. Tenison and K.R. Poolla, *Numerical solution of the Lyapunov equation by approximate power iteration*, Linear Algebra and Its Application, **236**: 205-230, (1996)
- [8] R. Lehoucq, D.C. Sorensen, and C. Yang, *ARPACK Users Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi methods*, SIAM Publications, Philadelphia, (1998). (software available at <http://www.caam.rice.edu/software/ARPACK>)
- [9] J.-R. Li and J. White, *Low rank solution of Lyapunov equations*, SIAM J. Matrix Anal. Appl. **24**: 260-280, (2002)
- [10] R. Nong and D.C. Sorensen, *New Results in Numerically Solving Large-Scale Algebraic Sylvester Equations*, in preparation
- [11] T. Penzl, *A cyclic low-rank Smith method for large sparse Lyapunov equations*, SIAM J. Sci. Comput. **21**: 1401-1418, (2000)
- [12] F. Riesz and B. Sz.-Nagy, *Functional Analysis, 2nd. ed.* (translated by L.F. Boron), Ungar, New York (1955)
- [13] J. Sabino, *Solution of large-scale Lyapunov equations via the block modified Smith method*, Ph.D. Thesis, Rice University, June 2006
- [14] V. Simoncini, *A new iterative method for solving large-scale Lyapunov matrix equations*, SIAM J. Sci. Comput. **29**: 1268-1288, (2007)
- [15] R.A. Smith, *Matrix equation $\mathbf{XA} + \mathbf{BA} = \mathbf{C}$* , SIAM J. Appl. Math., **16**: (1968)
- [16] D.C. Sorensen, *Implicit application of polynomial filters in a k -step Arnoldi method*, SIAM J. Matrix Anal. Applic., **13**: 357-385 (1992).
- [17] D. Vasilyev and J. White, *A more reliable reduction algorithm for behavior model extraction*, Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, 813-820, (2005)
- [18] J.P. Vigier, *Etude sur les suites infinies d'operateurs hermitiens*, These (Geneve, 1946).

- [19] Y.K. Zhou and D.C. Sorensen, Approximate Implicit Subspace Iteration with Alternating Directions for LTI system Model Reduction, *Numerical Linear Algebra with Applications*, Special Issue on Large Scale Matrix Equations **15**: 873-886 (2008)
- [20] Q. Zhou, K. Sun, K. Mohanram and D.C. Sorensen, *Large power grid analysis using domain decomposition*, Design, Automation and Test in Europe, (2006).