2

Smith-Type Methods for Balanced Truncation of Large Sparse Systems

Serkan Gugercin[1] and Jing-Rebecca Li[2]

[1] Virginia Tech., Dept. of Mathematics, Blacksburg, VA, 24061-0123, USA
gugercin@math.vt.edu
[2] INRIA-Rocquencourt, Projet Ondes, Domaine de Voluceau - Rocquencourt - B.P. 105, 78153 Le Chesnay Cedex, France jingrebecca.li@inria.fr

2.1 Introduction

Many physical phenomena, such as heat transfer through various media, signal propagation through electric circuits, vibration suppression of bridges, the behavior of Micro-Electro-Mechanical Systems (MEMS), and flexible beams are modelled with linear time invariant (LTI) systems

$$\mathbf{\Sigma}: \begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad \Leftrightarrow \quad \mathbf{\Sigma} := \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input and $y(t) \in \mathbb{R}^p$ is the output; moreover $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$ are constant matrices. The number of the states, n, is called the *dimension* or *order* of the system $\mathbf{\Sigma}$. Closely related to this system are two continuous-time *Lyapunov equations*:

$$A\mathcal{P} + \mathcal{P}A^T + BB^T = 0, \quad A^T\mathcal{Q} + \mathcal{Q}A + C^TC = 0. \tag{2.1}$$

The matrices $\mathcal{P} \in \mathbb{R}^{n\times n}$ and $\mathcal{Q} \in \mathbb{R}^{n\times n}$ are called the *reachability* and *observability Gramians*, respectively. Under the assumptions that A is asymptotically stable, i.e. $\lambda_i(A) \in \mathbb{C}_-$ (the open left half-plane), and that $\mathbf{\Sigma}$ is minimal (that is the pairs (A, B) and (C, A) are, respectively, reachable and observable), the Gramians $\mathcal{P}$, $\mathcal{Q}$ are unique and positive definite. In many applications, such as circuit simulation or time dependent PDE control problems, the dimension, n, of $\mathbf{\Sigma}$ is quite large, in the order of tens of thousands or higher, while the number of inputs m and outputs p usually satisfy $m, p \ll n$. In these large-scale settings, it is often desirable to approximate the given system with a much lower dimensional system

$$\mathbf{\Sigma}_r: \begin{cases} \dot{x}_r(t) = A_r x(t) + B_r u(t) \\ y_r(t) = C_r x(t) + D_r u(t) \end{cases} \quad \Leftrightarrow \quad \mathbf{\Sigma}_r := \left[\begin{array}{c|c} A_r & B_r \\ \hline C_r & D_r \end{array}\right]$$

where $A_r \in \mathbb{R}^{r\times r}$, $B_r \in \mathbb{R}^{r\times m}$, $C_r \in \mathbb{R}^{p\times r}$, $D_r \in \mathbb{R}^{p\times m}$, with $r \ll n$. The problem of model reduction is to produce such a low dimensional system $\mathbf{\Sigma}_r$ that has similar response characteristic as the original system $\mathbf{\Sigma}$ to any given input u.

The Lyapunov matrix equations in (2.1) play an important role in model reduction. One of the most effective model reduction approaches, called *balanced truncation* [MOO81, MR76], requires solving (2.1) to obtain $\mathcal{P}$ and $\mathcal{Q}$. A state space transformation based on $\mathcal{P}$ and $\mathcal{Q}$ is then derived to balance the system in the sense that the two Gramians become diagonal and equal. In this new co-ordinate system, states that are difficult to reach are simultaneously difficult to observe. Then, the reduced model is obtained by truncating the states that are both difficult to reach and difficult to observe. When applied to stable systems, balanced truncation preserves stability and provides an *a priori* bound on the approximation error.

For small-to-medium scale problems, balanced truncation can be implemented efficiently using the Bartels-Stewart [BS72] method, as modified by Hammarling [HAM82], to solve the two Lyapunov equations in (2.1). However, the method requires computing a Schur decomposition and results in $\mathcal{O}(n^3)$ arithmetic operations and $\mathcal{O}(n^2)$ storage; therefore, it is not appropriate for large-scale problems.

For large-scale sparse problems, iterative methods are preferred since they retain the sparsity of the problem and are much more suitable for parallelization. The Smith method [SMI68], the alternating direction implicit (**ADI**) iteration method [WAC88a], and the Smith(l) method [PEN00b] are the most popular iterative schemes developed for large sparse Lyapunov equations. Unfortunately, even though the number of arithmetic operations is reduced, all of these methods compute the solution in dense form and hence require $\mathcal{O}(n^2)$ storage.

It is well known that the Gramians $\mathcal{P}$ and $\mathcal{Q}$ often have low numerical rank (i.e. the eigenvalues of $\mathcal{P}$ and $\mathcal{Q}$ decay rapidly). This phenomenon is explained to a large extent in [ASZ02, PEN00a]. One must take advantage of this low-rank structure to obtain approximate solutions in low-rank factored form. In other words, one should construct a matrix $Z \in \mathbb{R}^{n\times r}$ such that $\mathcal{P} \approx ZZ^T$. The matrix Z is called *the approximate low-rank Cholesky factor* of $\mathcal{P}$. If the effective rank r is much smaller than n, i.e. $r \ll n$, then the storage is reduced from $\mathcal{O}(n^2)$ to $\mathcal{O}(nr)$. We note that such low-rank schemes are the only existing methods that can effectively solve very large sparse Lyapunov equations.

Most low-rank methods, such as [HPT96, HR92, JK94, SAA90], are Krylov subspace methods. As stated in [PEN00b], even though these methods reduce the memory requirement, they usually fail to yield approximate solutions of high accuracy. To reach accurate approximate solutions, one usually needs a large number of iterations, and therefore obtain approximations with relatively high numerical ranks; see [PEN00b]. For large-scale sparse Lyapunov equations, a more efficient low-rank scheme based on the ADI iteration was

introduced, independently, by Penzl [PEN00b], and Li and White [LW02]. The method was called the low-rank ADI iteration (**LR-ADI**) in [PEN00b] and the Cholesky factor ADI iteration (**CF-ADI**) in [LW02]. Even though LR-ADI and CF-ADI are theoretically the same, CF-ADI is less expensive and more efficient to implement. Indeed, LR-ADI can be considered as an intermediate step in deriving the CF-ADI algorithm. Another low-rank scheme based on the ADI iteration was also introduced in [PEN00b]. The method is called the cyclic low-rank Smith method (**LR-Smith(l)**) and is a special case of LR-ADI where l number of shifts are re-used in a cyclic manner.

While solving the Lyapunov equation $A\mathcal{P} + \mathcal{P}A^T + BB^T = 0$ where B has m columns, the LR-ADI and the LR-Smith(l) methods add m and $m \times l$ columns respectively to the current solution at each step, where l is the number of shifts. Therefore, for slowly converging iterations and for the case where m is big, e.g. $m = 10$, the number of columns of the approximate low-rank Cholesky factor can exceed manageable memory capacity. To overcome this, Gugercin *et. al.* [GSA03] introduced a Modified LR-Smith(l) method that prevents the number of columns from increasing arbitrarily at each step. In fact, the method only requires the number of columns r which are needed to meet the pre-specified balanced truncation tolerance. Due to the rapid decay of the Hankel singular values, this r is usually quite small relative to n. Consequently the memory requirements are drastically reduced.

This paper surveys Smith-type methods used for solving large-scale sparse Lyapunov equations and consequently for balanced truncation of the underlying large sparse dynamical system. Connections between different Smith-type methods, convergence results, and upper bounds for the approximation errors are presented. Moreover, numerical examples are given to illustrate the performance of these algorithms.

2.2 Balancing and Balanced Truncation

One model reduction scheme that is well grounded in theory is *Balanced Truncation*, first introduced by Mullis and Roberts [MR76] and later in the systems and control literature by Moore [MOO81]. The approximation theory underlying this approach was developed by Glover [GLO84]. Several researchers have recognized the importance of balanced truncation for model reduction because of its theoretical properties. Computational schemes for small-to-medium scale problems already exist. However, the development of computational methods for large-scale settings is still an active area of research; see [GSA03, PEN99, BQQ01, AS02], and the references therein.

2.2.1 The Concept of Balancing

Let $\mathcal{P}$ and $\mathcal{Q}$ be the unique Hermitian positive definite solutions to equations (2.1). The square roots of the eigenvalues of the product $\mathcal{PQ}$ are the

singular values of the Hankel operator associated with $\mathbf{\Sigma}$ and are called the *Hankel singular values*, $\sigma_i(\mathbf{\Sigma})$, of the system $\mathbf{\Sigma}$:

$$\sigma_i(\mathbf{\Sigma}) = \sqrt{\lambda_i(\mathcal{PQ})}.$$

In most cases, the eigenvalues of $\mathcal{P}, \mathcal{Q}$ as well as the Hankel singular values $\sigma_i(\mathbf{\Sigma})$ decay very rapidly. This phenomena is explained to a large extent in [ASZ02].

Define the two functionals J_r and J_o as follows:

$$J_r = \min_{x(-\infty)=0,\ x(0)=x} \|u(t)\|^2, \quad t \leq 0, \tag{2.2}$$

$$J_o = \|y(t)\|^2, \quad x(0) = x_o,\ u(t) = 0,\ t \geq 0. \tag{2.3}$$

The quantity J_r is the minimal energy required to drive the system from the zero state at $t = -\infty$ to the state x at $t = 0$. On the other hand, J_o is the energy obtained by observing the output with the initial state x_o under no input. The following lemma is crucial to the concept of balancing:

Lemma 2.2.1. *Let $\mathcal{P}$ and $\mathcal{Q}$ be the reachability and observability Gramians of the asymptotically stable and minimal system $\mathbf{\Sigma}$ and J_r and J_o be defined as above. Then*

$$J_r = x^T \mathcal{P}^{-1} x$$

and

$$J_o = x_o^T \mathcal{Q}\, x_o.$$

It follows from the above lemma that the states which are difficult to reach, i.e., require a large energy J_r, are spanned by the eigenvectors of $\mathcal{P}$ corresponding to small eigenvalues. Moreover, the states which are difficult to observe, i.e., yield small observation energy J_o, are spanned by the eigenvectors of $\mathcal{Q}$ corresponding to small eigenvalues. Hence Lemma 2.2.1 yields a way to evaluate the degree of reachability and the degree of observability for the states of the given system. One can obtain a reduced model by eliminating the states which are difficult to reach and observe. However, it is possible that the states which are difficult to reach are not difficult to observe and vice-versa. See [ANT05] for more details and examples. Hence the following question arises: Given $\mathbf{\Sigma}$, does there exist a basis where the states which are difficult to reach are simultaneously difficult to observe? It is easy to see from the Lyapunov equations in (2.1) that under a state transformation by a nonsingular matrix T, the Gramians are transformed as

$$\bar{\mathcal{P}} = T\mathcal{P}T^T, \qquad \bar{\mathcal{Q}} = T^{-T}\mathcal{Q}T^{-1}.$$

Hence, the answer to the above question reduces to finding a nonsingular state transformation T such that, in the transformed basis, the Gramians $\bar{\mathcal{P}}$ and $\bar{\mathcal{Q}}$ are equal.

Definition 2.2.2. *The reachable, observable and stable system* $\mathbf{\Sigma}$ *is called balanced if* $\mathcal{P} = \mathcal{Q}$. $\mathbf{\Sigma}$ *is called principal-axis-balanced if*

$$\mathcal{P} = \mathcal{Q} = \Sigma = \operatorname{diag}(\sigma_1 I_{m_1}, \cdots, \sigma_q I_{m_q}), \tag{2.4}$$

where $\sigma_1 > \sigma_2 > \cdots > \sigma_q > 0$, m_i, $i = 1, \cdots, q$, *are the multiplicities of* σ_i, *and* $m_1 + \cdots + m_q = n$.

In the following, by balancing we mean **principal-axis-balancing** unless otherwise stated. It follows from the above definition that balancing amounts to the simultaneous diagonalization of the two positive definite matrices $\mathcal{P}$ and $\mathcal{Q}$.

Let U denote the Cholesky factor of $\mathcal{P}$, i.e., $\mathcal{P} = UU^T$, and let $U^T \mathcal{Q} U = R\Sigma^2 R^T$ be the eigenvalue decomposition of $U^T \mathcal{Q} U$. The following result explains how to compute the balancing transformation T:

Lemma 2.2.3. Principal-Axis-Balancing Transformation:
Given the minimal and asymptotically stable LTI system $\mathbf{\Sigma}$ *with the corresponding Gramians* $\mathcal{P}$ *and* $\mathcal{Q}$, *a principal-axis-balancing transformation* T *is*

$$T = \Sigma^{1/2} R^T U^{-1}. \tag{2.5}$$

The next result gives a generalization of all possible balancing transformations:

Corollary 2.2.4. *Let there be* q *distinct Hankel singular values* σ_i *with multiplicities* m_i. *Every principal-axis-balancing transformation* $\hat{T}$ *has the form* $\hat{T} = VT$ *where* T *is given by (2.5) and* V *is a block diagonal unitary matrix with an arbitrary* $m_i \times m_i$ *unitary matrix as the* i^{th} *block for* $i = 1, \cdots, q$.

2.2.2 Model Reduction by Balanced Truncation

The balanced basis has the property that the states which are difficult to reach are simultaneously difficult to observe. Hence, a reduced model is obtained by truncating the states which have this property, i.e., those which correspond to small Hankel singular values σ_i.

Theorem 2.2.5. *Let the asymptotically stable and minimal system* $\mathbf{\Sigma}$ *have the following balanced realization:*

$$\mathbf{\Sigma} = \left[\begin{array}{c|c} A_b & B_b \\ \hline C_b & D_b \end{array}\right] = \left[\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & D \end{array}\right],$$

with $\mathcal{P} = \mathcal{Q} = \operatorname{diag}(\Sigma_1, \Sigma_2)$ *where*

$$\Sigma_1 = \operatorname{diag}(\sigma_1 I_{m_1}, \cdots, \sigma_k I_{m_k}) \quad \text{and} \quad \Sigma_2 = \operatorname{diag}(\sigma_{k+1} I_{m_{k+1}}, \cdots, \sigma_q I_{m_q}).$$

Then the reduced order model $\mathbf{\Sigma}_r = \left[\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & D \end{array}\right]$ *obtained by balanced truncation is asymptotically stable, minimal and satisfies*

$$\|\mathbf{\Sigma} - \mathbf{\Sigma}_r\|_{\mathcal{H}_\infty} \leq 2\,(\sigma_{k+1} + \cdots + \sigma_q). \tag{2.6}$$

The equality holds if Σ_2 *contains only* σ_q.

The above theorem states that if the neglected Hankel singular values are small, then the systems $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_r$ are guaranteed to be close. Note that (2.6) is an *a priori* error bound. Hence, given an error tolerance, one can decide how many states to truncate without forming the reduced model.

The balancing method explained above is also called Lyapunov balancing since it requires solving two Lyapunov equations. Besides the Lyapunov balancing method, other types of balancing exist such as stochastic balancing [DP84, GRE88a, GRE88b], bounded real balancing , positive real balancing [DP84], LQG balancing [OJ88], and frequency weighted balancing [ENN84, LC92, SAM95, WSL99, ZHO95, VA01, GJ90, GA04]. For a recent survey of balancing related model reduction, see [GA04].

2.2.3 A Numerically Robust Implementation of Balanced Reduction

The above discussion on the balancing transformation and the balanced reduction requires balancing the whole system $\mathbf{\Sigma}$ followed by the truncation. This approach is numerically inefficient and very ill-conditioned to implement. Instead, below we will give another implementation of the balanced reduction which directly obtains a reduced balanced system without balancing the whole system.

Let $\mathcal{P} = UU^T$ and $\mathcal{Q} = LL^T$. This is always possible since both $\mathcal{P}$ and $\mathcal{Q}$ are symmetric positive definite matrices. The matrices U and L are called the *Cholesky factors* of the Gramians $\mathcal{P}$ and $\mathcal{Q}$, respectively. Let $U^T L = ZSY^T$ be a singular value decomposition (SVD). It is easy to show that the singular values of $U^T L$ are indeed the Hankel singular values, hence, we have

$$U^T L = Z\Sigma Y^T$$

where

$$\Sigma = \mathrm{diag}(\sigma_1 I_{m_1}, \sigma_2 I_{m_2}, \ldots, \sigma_q I_{m_q}),$$

q is the number of distinct Hankel singular values, with $\sigma_i > \sigma_{i+1} > 0$, m_i is the multiplicity of σ_i, and $m_1 + m_2 + \cdots + m_q = n$. Let

$$\Sigma_1 = \mathrm{diag}(\sigma_1 I_{m_1}, \sigma_2 I_{m_2}, \ldots, \sigma_k I_{m_k}), \quad k < q, \quad r := m_1 + \cdots + m_k,$$

and define

$$W_1 := LY_1\Sigma_1^{-1/2} \quad \text{and} \quad V_1 := UZ_1\Sigma_1^{-1/2},$$

where Z_1 and Y_1 are composed of the leading r columns of Z and Y, respectively. It is easy to check that $W_1^T V_1 = I_r$ and hence $V_1 W_1^T$ is an oblique projector. We obtain a reduced model of order r by setting

$$A_r = W_1^T A V_1, \quad B_r = W_1^T B, \quad C_r = C V_1.$$

Noting that $\mathcal{P} W_1 = V_1 \Sigma_1$ and $\mathcal{Q} V_1 = W_1 \Sigma_1$ gives

$$\begin{aligned} W_1^T (A\mathcal{P} + \mathcal{P} A^T + B B^T) W_1 &= A_r \Sigma_1 + \Sigma_1 A_r^T + B_r B_r^T \\ V_1^T (A^T \mathcal{Q} + \mathcal{Q} A + C^T C) V_1 &= A_r^T \Sigma_1 + \Sigma_1 A_r + C_r^T C_r. \end{aligned}$$

Thus, the reduced model is balanced and asymptotically stable (due to the Lyapunov inertia theorem) for any $k \leq q$. As mentioned earlier, the formulae above provide a numerically stable scheme for computing the reduced order model based on a numerically stable scheme for computing the Cholesky factors U and L directly in upper triangular and lower triangular form, respectively. It is important to truncate Z, Σ, Y to Z_1, Σ_1, Y_1 prior to forming W_1 or V_1. It is also important to avoid formulae involving inversion of L or U as these matrices are typically ill-conditioned due to the decay of the eigenvalues of the Gramians.

2.3 Iterative ADI Type Methods for Solving Large-Scale Lyapunov Equations

The numerically stable implementation of the balanced truncation method described in Section 2.2.3 requires the solutions to two Lyapunov equations of order n. For small-to-medium scale problems, the solutions can be obtained through the Bartels-Stewart [BS72] method as modified by Hammarling [HAM82]. This method requires the computation of a Schur decomposition, and thus is not appropriate for large-scale problems. The problem of obtaining the full-rank exact solution to a Lyapunov equation is a numerically ill-conditioned problem in the large-scale setting.

As explained previously, $\mathcal{P}$ and $\mathcal{Q}$ often have *numerically* low-rank compared to n. In most cases, the eigenvalues of $\mathcal{P}, \mathcal{Q}$ as well as the Hankel singular values $\sigma_i(\mathbf{\Sigma})$ decay very rapidly, see [ASZ02]. This *low-rank phenomenon* leads to the idea of approximating the Gramians with low-rank approximate Gramians.

In the following, we will focus on the approximate solution of the reachability Lyapunov equation

$$A\mathcal{P} + \mathcal{P} A^T + B B^T = 0, \tag{2.7}$$

where $A \in \mathbb{R}^{n \times n}$ is asymptotically stable and diagonalizable and $B \in \mathbb{R}^{n \times m}$. The discussion applies equally well to the observability Lyapunov equation $A^T \mathcal{Q} + \mathcal{Q} A + C^T C = 0$.

In this section we survey the ADI, Smith, and Smith(l) methods. In these methods the idea is to transform a continuous time Lyapunov equation (2.7) into a discrete time Stein equation using spectral transformations of the type $\omega(\lambda) = \frac{\mu^* - \lambda}{\mu + \lambda}$, where $\mu \in \mathbb{C}_-$ (the open left half-plane). Note that ω is a bilinear transformation mapping the open left half-plane onto the open unit disk with $\omega(\infty) = -1$. The number μ is called the *shift* or the *ADI parameter*.

2.3.1 The ADI Iteration

The alternating direction implicit (ADI) iteration was first introduced by Peaceman and Rachford [PR55] to solve linear systems arising from the discretization of elliptic boundary value problems. In general, the ADI iteration is used to solve linear systems of the form

$$My = b,$$

where M is symmetric positive definite and can be split into the sum of two symmetric positive definite matrices $M = M_1 + M_2$ for which the following iteration is efficient:

$$\begin{aligned} y_0 &= 0, \\ (M_1 + \mu_j I) y_{j-1/2} &= b - (M_2 - \mu_j I) y_{j-1}, \\ (M_2 + \eta_j I) y_j &= b - (M_1 - \eta_j I) y_{j-1/2}, \text{ for } j = 1, 2, \cdots, J. \end{aligned}$$

The ADI shift parameters μ_j and η_j are determined from spectral bounds on M_1 and M_2 to increase the convergence rate. When M_1 and M_2 commute, this is classified as a "model problem".

One should notice that (2.7) is a model ADI problem in which there is a linear system with the sum of two commuting operators acting on the unknown $\mathcal{P}$, which is a matrix in this case. Therefore, the iterates $\mathcal{P}_i^A$ of the ADI iteration are obtained through the iteration steps

$$(A + \mu_i I)\mathcal{P}_{i-1/2}^A = -BB^T - \mathcal{P}_{i-1}^A (A^T - \mu_i I) \tag{2.8}$$

$$(A + \mu_i I)\mathcal{P}_i^A = -BB^T - (\mathcal{P}_{i-1/2}^A)^* (A^T - \mu_i I), \tag{2.9}$$

where $\mathcal{P}_0^A = 0$ and the shift parameters $\{\mu_1, \mu_2, \mu_3, \ldots\}$ are elements of $\mathbb{C}_-$ (here $*$ denotes complex conjugation followed by transposition). These two equations are equivalent to the following single iteration step:

$$\begin{aligned} \mathcal{P}_i^A = (A - \mu_i^* I)(A + \mu_i I)^{-1} \mathcal{P}_{i-1}^A [(A - \mu_i^* I)(A + \mu_i I)^{-1}]^* \\ -2\rho_i (A + \mu_i I)^{-1} BB^T (A + \mu_i I)^{-*}, \end{aligned} \tag{2.10}$$

where $\rho_i = \text{Real}(\mu_i)$. Note that if $\mathcal{P}_{i-1}$ is Hermitian positive semi-definite, then so is $\mathcal{P}_i$.

The spectral radius of the matrix $\left(\prod_{i=1}^{l}(A-\mu_i^* I)(A+\mu_i I)^{-1}\right)$, denoted by ρ_{ADI}, determines the rate of convergence, where l is the number of shifts used. Note that since A is asymptotically stable, $\rho_{ADI} < 1$. Smaller ρ_{ADI} yields faster convergence. The minimization of ρ_{ADI} with respect to shift parameters μ_i is called the *ADI minimax problem*:

$$\{\mu_1, \mu_2, \ldots, \mu_l\} = \arg \min_{\{\mu_1,\ldots,\mu_l\}\in\mathbb{C}_-} \max_{\lambda\in\sigma(A)} \frac{|\,(\lambda-\mu_1^*)\ldots(\lambda-\mu_l^*)\,|}{|\,(\lambda+\mu_1)\ldots(\lambda+\mu_l)\,|}. \quad (2.11)$$

We refer the reader to [EW91, STA91, WAC90, CR96, STA93, WAC88b, PEN00b] for contributions to the solution of the ADI minimax problem. It can be shown that if A is diagonalizable, the l^{th} ADI iterate satisfies the inequality

$$\left\|\mathcal{P}-\mathcal{P}_l^A\right\|_F \le \|W\|_2^2 \left\|W^{-1}\right\|_2^2 \rho_{ADI}^2 \left\|\mathcal{P}\right\|_F, \quad (2.12)$$

where W is the matrix of eigenvectors of A.

The basic computational costs in the ADI iterations are that each individual shift μ_i requires a sparse direct factorization of $(A+\mu_i I)$ and each application of $(A+\mu_i I)^{-1}$ requires triangular solves from that factorization. Moreover, in the case of complex shifts, these operations have to be done in complex arithmetic. To keep the solution $\mathcal{P}$ real, complex conjugate pairs of shifts have to be applied, one followed immediately by the other. However, even with this, one would have to form $(A+\mu_i I)(A+\mu_i^* I) = A^2+2\rho_i A+|\mu_i|^2 I$ in order to keep the factorizations in real arithmetic. This matrix squaring would most likely have an adverse effect on sparsity. In the following, we wish to avoid the additional details required to discuss complex shifts. Therefore, *we will restrict our discussions to real shifts for the remainder of the paper.* If necessary, all of the operations can be made valid for complex shifts.

2.3.2 Smith's Method

For every real scalar $\mu < 0$, the continuous-time Lyapunov equation (2.7) is equivalent to

$$\mathcal{P} = (A-\mu I)(A+\mu I)^{-1}\mathcal{P}(A+\mu I)^{-T}(A-\mu I)^T - 2\mu(A+\mu I)^{-1}BB^T(A^T+\mu I)^{-1}.$$

Then one obtains the Stein equation

$$\mathcal{P} = A_\mu \mathcal{P} A_\mu^T - 2\mu B_\mu B_\mu^T, \quad (2.13)$$

where

$$A_\mu := (A-\mu I)(A+\mu I)^{-1},\ B_\mu := (A+\mu I)^{-1}B. \quad (2.14)$$

Hence using the bilinear transformation $\omega(\lambda) = \frac{\mu-\lambda}{\mu+\lambda}$, the problem has been transformed into discrete time, where the Stein equation (2.13) has the same

solution as the continuous time Lyapunov equation (2.7). Since A is asymptotically stable, $\rho(A_\mu) < 1$ and the sequence $\{\mathcal{P}_i^S\}_{i=0}^{\infty}$ generated by the iteration

$$\mathcal{P}_1^S = -2\mu B_\mu B_\mu^T \text{ and } \mathcal{P}_{j+1}^S = A_\mu \mathcal{P}_j^S A_\mu^T + \mathcal{P}_1^S$$

converges to the solution $\mathcal{P}$. Thus, the Smith iterates can be written as

$$\mathcal{P}_k^S = -2\mu \sum_{j=0}^{k-1} A_\mu^j B_\mu B_\mu^T (A_\mu^j)^T. \tag{2.15}$$

If one uses the same shift through out the ADI iteration, ($\mu_j = \mu$, $j = 1, 2, \dots$), then the ADI iteration reduces to the Smith method. Generally, the convergence of the Smith method is slower than ADI. An accelerated version, the so called squared Smith method, has been proposed in [PEN00b] to improve convergence. However, despite a better convergence, the squared Smith methods destroys the sparsity of the problem which is not desirable in large-scale settings.

2.3.3 Smith(l) Iteration

Penzl [PEN00b] illustrated that the ADI iteration with a single shift converges very slowly, while a moderate increase in the number of shifts l accelerates the convergence nicely. However, he also observed that the speed of convergence is hardly improved by a further increase of l; see Table 2.1 in [PEN00b]. These observations led to the idea of the cyclic Smith(l) iteration, a special case of ADI where l different shifts are used in a cyclic manner, i.e. $\mu_{i+jl} = \mu_i$ for $j = 1, 2, \cdots$.

The Smith(l) iterates are generated by

$$\mathcal{P}_k^{Sl} = \sum_{j=0}^{k-1} A_d^k T (A_d^k)^T, \tag{2.16}$$

where

$$A_d = \prod_{i=1}^{l} (A - \mu_i I)(A + \mu_i I)^{-1} \qquad \text{and} \qquad T = \mathcal{P}_l^A, \tag{2.17}$$

i.e., T is the l^{th} ADI iterate with the shifts $\{\mu_1, \cdots, \mu_l\}$. As in Smith's methods, $\mathcal{P} - A_d \mathcal{P} A_d^T = T$ is equivalent to (2.7), where A_d and T are defined in (2.17).

2.4 Low-rank Iterative ADI-Type Methods

The original versions of the ADI, Smith, and Smith(l) methods outlined above form and store the entire dense solution $\mathcal{P}$ explicitly, resulting in extensive

storage requirement. In many cases the storage requirement is the limiting factor rather than the amount of computation. The observation that $\mathcal{P}$ is numerically low-rank compared to n leads to the low-rank formulations of the ADI iterations, namely, LR-ADI [PEN00b], CF-ADI [LW02], LR-Smith(l) [PEN00b], and Modified LR-Smith(l) [GSA03] where, instead of explicitly forming the solution $\mathcal{P}$, only the low-rank approximate Cholesky factors are computed and stored, reducing the storage requirement to $\mathcal{O}(nr)$ where r is the numerical rank of $\mathcal{P}$.

2.4.1 LR-ADI and CF-ADI Iterations

Recall that the two steps in (2.8) and (2.9) of the ADI iteration can be combined into the single iteration step in (2.10), as rewritten below:

$$\begin{aligned}\mathcal{P}_i^A = (A-\mu_i I)(A+\mu_i I)^{-1}\mathcal{P}_{i-1}^A[(A-\mu_i I)(A+\mu_i I)^{-1}]^T \\ -2\mu_i(A+\mu_i I)^{-1}BB^T(A+\mu_i I)^{-T}.\end{aligned} \tag{2.18}$$

The key idea in the low-rank versions of the ADI method is to rewrite the iterate $\mathcal{P}_i^A$ in (2.18) as an outer product:

$$\mathcal{P}_i^A = Z_i^A(Z_i^A)^T. \tag{2.19}$$

This is always possible since starting with the initial guess $\mathcal{P}_i^A = 0$, the iterates $\mathcal{P}_i^A$ can be shown recursively to be positive definite and symmetric.

Using (2.19) in (2.18) results in

$$\begin{aligned}Z_i^A(Z_i^A)^T = (A-\mu_i I)(A+\mu_i I)^{-1}Z_{i-1}^A[(A-\mu_i I)(A+\mu_i I)^{-1}Z_{i-1}^A]^T \\ -2\mu_i(A+\mu_i I)^{-1}BB^T(A+\mu_i I)^{-T}.\end{aligned} \tag{2.20}$$

Since the left-hand side of (2.20) is an outer product, and the right hand side is the sum of two outer products, Z_i^A can be rewritten as

$$Z_i^A = [\ (A-\mu_i I)(A+\mu_i I)^{-1}Z_{i-1}^A \qquad \sqrt{-2\mu_i}(A+\mu_i I)^{-1}B\ ]. \tag{2.21}$$

Therefore, the ADI algorithm (2.18) can be reformulated in terms of the Cholesky factor Z_i^A as

$$Z_1^A = \sqrt{-2\mu_1}(A+\mu_1 I)^{-1}B, \tag{2.22}$$

$$Z_i^A = [\ (A-\mu_i I)(A+\mu_i I)^{-1}Z_{i-1}^A \qquad \sqrt{-2\mu_i}(A+\mu_i I)^{-1}B\ ]. \tag{2.23}$$

This low-rank formulation of the ADI iteration was independently developed in [PEN00b] and [LW02]. We will call this the LR-ADI iteration as in [PEN00b] since it is the *preliminary* form of the final CF-ADI iteration [LW02]. In the LR-ADI formulation (2.22) and (2.23), at the i^{th} step, the $(i-1)^{\text{st}}$ Cholesky factor Z_{i-1}^A is multiplied from left by $(A-\mu_i I)(A+\mu_i I)^{-1}$.

Therefore, the number of columns to be modified at each step increases by m, the number of columns in B. In [LW02], the steps (2.22) and (2.23) are reformulated to keep the number of columns modified at each step as constant. The resulting algorithm, outlined below, is called the CF-ADI iteration.

The columns of k^{th} LR-ADI iterate Z_i^A can be written out explicitly as

$$Z_k^A = [S_k\sqrt{-2\mu_k}B,\ S_k(T_kS_{k-1})\sqrt{-2\mu_{k-1}}B,\ \cdots,\ S_kT_k\cdots S_2(T_2S_1)\sqrt{-2\mu_1}B]$$

where

$$S_i := (A+\mu_i I)^{-1}, \quad \text{and} \quad T_i := (A-\mu_i I) \quad \text{for} \quad i=1,\ldots,k.$$

Since S_i and T_j commute, i.e.

$$S_iS_j = S_jS_i, \quad T_iT_j = T_jT_i, \quad S_iT_j = T_jS_i, \quad \forall i,j,$$

Z_k^A can be written as

$$Z_k^A = [\ z_k \quad P_{k-1}(z_k), \quad P_{k-2}(P_{k-1}z_k), \quad \cdots\cdots \quad P_1(P_2\cdots P_{k-1}z_k)], \tag{2.24}$$

where

$$z_k := \sqrt{-2\mu_k}\ (A+\mu_k I)^{-1}B, \tag{2.25}$$

$$P_i := \frac{\sqrt{-2\mu_i}}{\sqrt{-2\mu_{i+1}}}\ [I-(\mu_{i+1}+\mu_i)(A+\mu_i I)^{-1}]. \tag{2.26}$$

Since the order of the ADI parameters μ_i is not important, the ordering of μ_i can be reversed resulting in the CF-ADI iteration:

$$Z_1^{CFA} = z_1 = \sqrt{-2\mu_1}\ (A+\mu_1 I)^{-1}B, \tag{2.27}$$

$$z_i = \left(\frac{\sqrt{-2\mu_i}}{\sqrt{-2\mu_{i-1}}}\right)\ \left(I-(\mu_i+\mu_{i-1})(A+\mu_i I)^{-1}\right) z_{i-1}, \tag{2.28}$$

$$Z_i^{CFA} = [Z_{i-1}^{CFA} \quad z_i], \qquad \text{for } i=2,\cdots,k. \tag{2.29}$$

Unlike the LR-ADI iteration (2.22)-(2.23) where at the i^{th} step $(i-1)m$ number of columns need to be modified, the CF-ADI iteration (2.27)-(2.29) requires only that a *constant* number of columns, namely, m, to be modified at each step. Therefore, the implementation of CF-ADI is numerically more efficient compared to LR-ADI.

Define $\mathcal{P}_j^{CFA} := Z_j^{CFA}(Z_j^{CFA})^T$. Clearly, the stopping criterion $\|\mathcal{P}_j^{CFA} - \mathcal{P}_{j-1}^{CFA}\|_2 \leq tol^2$ can be implemented as $\|z_j\|_2 \leq tol$, since

$$\|Z_j^{CFA}(Z_j^{CFA})^T - Z_{j-1}^{CFA}(Z_{j-1}^{CFA})^T\|_2 = \|z_jz_j^T\|_2 = \|z_j\|_2^2.$$

It is not necessarily true that a small z_j implies that all further z_{j+k} will be small, but this has been observed in practice. Relative error can also be used, in which case the stopping criterion is

$$\frac{\|z_j\|_2}{\|Z_{j-1}^{CFA}\|_2} \leq tol.$$

The 2-norm of Z_{j-1}^{CFA}, which is also its largest singular value, can be estimated by performing power iterations to estimate the largest eigenvalue of $Z_{j-1}^{CFA}(Z_{j-1}^{CFA})^T$, taking advantage of the fact that $j \ll n$. This cost is still high, and this estimate should only be used after each segment of several iterations.

The next result shows the relation between the ADI, LR-ADI and CF-ADI iterations. For a proof, see the original source [LW02].

Theorem 2.4.1. *Let $\mathcal{P}_k^A$ be the approximation obtained by k steps of the ADI iteration with shifts $\{\mu_1, \mu_2, \ldots, \mu_k\}$. Moreover, for the same shift selection, let Z_k^A and Z_k^{CFA} be the approximations obtained by the LR-ADI and the CF-ADI iterations as above, respectively. Then,*

$$\mathcal{P}_k^A = Z_k^A(Z_k^A)^T = Z_k^{CFA}(Z_k^{CFA})^T.$$

2.4.2 LR-Smith(l) Iteration

The ADI, LR-ADI, and CF-ADI iterations are of interest if a sequence $\{\mu_i\}_{i=1}^k$ of different shifts is available. When the number of shift parameters is limited, the cyclic low-rank Smith method (LR-Smith(l)) is a more efficient alternative. As in the LR-ADI formulation of the ADI iteration, the key idea is to write the i^{th} Smith(l) iterate as

$$\mathcal{P}_i^{Sl} = Z_i^{Sl}(Z_i^{Sl})^T. \tag{2.30}$$

Given the l cyclic-shifts $\{\mu_1, \mu_2, \ldots, \mu_l\}$, the LR-Smith($l$) method consists of two steps. First the iterate Z_1^{Sl} is obtained by an l step low-rank ADI iteration; i.e. $P_l^A = Z_l^A(Z_l^A)^T$ is the low-rank l step ADI iterate. Then, the LR-Smith(l) method is initialized by

$$Z_1^{Sl} = B_d = Z_l^A, \tag{2.31}$$

followed by the actual LR-Smith(l) iteration:

$$\begin{aligned} Z^{(i+1)} &= A_d Z^{(i)} \\ Z_{i+1}^{Sl} &= [\ Z_i^{Sl} \quad Z^{(i+1)}\ ], \end{aligned} \tag{2.32}$$

where A_d is defined in (2.17). It then follows that

$$Z_k^{Sl} = [\ B_d \ \ A_d B_d \ \ A_d^2 B_d \ \ \cdots \ \ A_d^{k-1} B_d\ ]. \tag{2.33}$$

One should notice that while k step LR-ADI and CF-ADI iterations require k matrix factorizations, a k step LR-Smith(l) iteration computes only l matrix

factorizations. Moreover, the equality (2.33) reveals that similar to the CF-ADI iteration, the number of columns to be modified at the i^{th} step of the LR-Smith(l) iteration is constant, equal to the number of columns of B_d, namely $l \times m$. If the shifts $\{\mu_1, \cdots, \mu_l\}$ are used in a cyclic manner, the cyclic LR-Smith(l) iteration gives the same approximation as the LR-ADI iteration.

Remark 2.4.2. A system theoretic interpretation of using l cyclic shifts (the Smith(l) iteration) is that the continuous time system

$$\mathbf{\Sigma} = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

which has order n, m inputs, and p outputs is embedded into a discrete time system

$$\mathbf{\Sigma}_d = \left[\begin{array}{c|c} A_d & B_d \\ \hline C_d & D_d \end{array}\right]$$

which has order n, lm inputs, and lp outputs; they have the same reachability and observability Gramians $\mathcal{P}$ and $\mathcal{Q}$. Therefore, at the cost of increasing the number of inputs and outputs, one reduces the spectral radius $\rho(A_d)$ and hence increases the convergence.

Remark 2.4.3. Assume that we know all the eigenvalues of A and the system

$$\mathbf{\Sigma} = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

is single input single output, i.e. $B, C^T \in \mathbb{R}^n$. Then choosing $\mu_i = \lambda_i(A)$ for $i = 1, \cdots, n$ results in

$$A_d = 0 \qquad \text{and} \qquad \mathcal{P} = \mathcal{P}_1^{Sl} = \mathcal{P}_l^A.$$

In other words, the exact solution $\mathcal{P}$ of (2.7) is obtained at the first step. The resulting discrete time system has n inputs, and n outputs.

Convergence Results for the Cyclic LR-Smith(l) Iteration

In this section some convergence results for the Cyclic LR-Smith(l) iteration are presented. For more details, we refer the reader to the original source [GSA03].

Let Z_k^{Sl} be the k^{th} LR-Smith(l) iterate as defined in (2.33) corresponding to the Lyapunov equation

$$A\mathcal{P} + \mathcal{P}A^T + BB^T = 0.$$

Similar to Z_k^{Sl}, let Y_k^{Sl} be the k^{th} LR-Smith(l) iterate corresponding to the observability Lyapunov equation

$$A^T\mathcal{Q} + \mathcal{Q}A + C^TC = 0$$

for the same cyclic shift selection as used in computing Z_k^{Sl}.

Denote by $\mathcal{P}_k^{Sl}$ and $\mathcal{Q}_k^{Sl}$ the k step LR-Smith(l) iterates for $\mathcal{P}$ and $\mathcal{Q}$ respectively, i.e., $\mathcal{P}_k^{Sl} = Z_k^{Sl}(Z_k^{Sl})^T$ and $\mathcal{Q}_k^{Sl} = Y_k^{Sl}(Y_k^{Sl})^T$. Similar to (2.12), the following result holds:

Proposition 2.4.4. *Let $E_{kp} := \mathcal{P} - \mathcal{P}_k^{Sl}$ and $E_{kq} = \mathcal{Q} - \mathcal{Q}_k^{Sl}$ and $A = W(\Lambda)W^{-1}$ be the eigenvalue decomposition of A. The k step LR-Smith(l) iterates satisfy*

$$\begin{aligned} 0 &\leq \operatorname{trace}(E_{kp}) = \operatorname{trace}(\mathcal{P} - \mathcal{P}_k^{Sl}) \leq K\, m\, l\, (\rho(A_d))^{2k} \operatorname{trace}(\mathcal{P}) \quad (2.34)\\ 0 &\leq \operatorname{trace}(E_{kq}) = \operatorname{trace}(\mathcal{Q} - \mathcal{Q}_k^{Sl}) \leq K\, p\, l\, (\rho(A_d))^{2k} \operatorname{trace}(\mathcal{Q}), \quad (2.35)\end{aligned}$$

where

$$K = \kappa(W)^2, \tag{2.36}$$

and $\kappa(W)$ denotes the 2-norm condition number of W.

Since the low-rank Cholesky factors Z_k^{Sl} and Y_k^{Sl} will be used for balanced truncation of the underlying dynamical system, it is important to see how well the exact Hankel singular values are approximated. Let σ_i and $\hat{\sigma}_i$ denote the Hankel singular values resulting from the full-rank exact Gramians and the low-rank approximate Gramians, respectively, i.e.,

$$\sigma_i^2 = \lambda_i(\mathcal{P}\mathcal{Q}) \text{ and } \hat{\sigma}_i^2 = \lambda_i(\mathcal{P}_k^{Sl}\mathcal{Q}_k^{Sl}). \tag{2.37}$$

The following lemma holds:

Lemma 2.4.5. *Let σ_i and $\hat{\sigma}_i$ be given by (2.37). Define $\hat{n} = kl\min(m,p)$. Then,*

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n}\sigma_i^2 - \sum_{i=1}^{\hat{n}}\hat{\sigma}_i^2\\ &\leq K\, l\, (\rho(A_d))^{2k}\Big(K\ \min(m,p)(\rho(A_d))^{2k}\operatorname{trace}(\mathcal{P})\operatorname{trace}(\mathcal{Q})\\ &\quad + m\operatorname{trace}(\mathcal{P})\sum_{i=0}^{k-1}\|C_dA_d^i\|_2^2 + p\operatorname{trace}(\mathcal{Q})\sum_{i=0}^{k-1}\|A_d^iB_d\|_2^2\Big) \end{aligned} \tag{2.38}$$

where K is as defined in (2.36).

As mentioned in [GSA03], these error bounds critically depend on $\rho(A_d)$ and K. Hence when $\rho(A_d)$ is almost 1 and/or A is highly non-normal, the bounds may be pessimistic. On the other hand, when $\rho(A_d)$ is small, for example less than 0.9, the convergence of the iteration is extremely fast and also the error bounds are tight.

2.4.3 The Modified LR-Smith(l) Iteration

It follows from the implementations of the LR-ADI, the CF-ADI, and the LR-Smith(l) iterations that at each step the number of columns of the current iterates is increased by m for the LR-ADI and CD-ADI methods, and by $m \times l$ for the LR-Smith(l) method. Hence, when m is large, i.e. for MIMO systems, or when the convergence is slow, i.e., $\rho(A_d)$ is close to 1, the number of columns of Z_k^A, Z_k^{CFA}, and Z_k^{Sl} might exceed available memory. In light of these observations, Gugercin *et. al.* [GSA03] introduced a modified LR-Smith(l) iteration where the number of columns in the low-rank Cholesky factor does not increase unnecessarily at each step. The idea is to compute the singular value decomposition of the iterate at each step and, given a tolerance τ, to replace the iterate with its best low-rank approximation as outlined below.

Let Z_k^{Sl} be the k^{th} LR-Smith(l) iterate as defined in (2.33) corresponding to the Lyapunov equation $A\mathcal{P}+\mathcal{P}A^T+BB^T=0$. Let the short singular value decomposition (S-SVD) of Z_k^{Sl} be

$$Z_k^{Sl} = V\Phi F^T,$$

where $V \in \mathbb{R}^{n\times(mlk)}$, $\Phi \in \mathbb{R}^{(mlk)\times(mlk)}$, and $F \in \mathbb{R}^{(mlk)\times(mlk)}$. Then the S-SVD of $\mathcal{P}_k^{Sl} = Z_k^{Sl}(Z_k^{Sl})^T$ is given by $\mathcal{P}_k^{Sl} = V\Phi^2V^T$. Therefore, it is enough to store only V and Φ, and

$$\tilde{Z}_k := V\Phi$$

is also a low-rank Cholesky factor for $\mathcal{P}_k^{Sl}$.

For a pre-specified tolerance value $\tau > 0$, assume that until the k^{th} step of the algorithm all the iterates $\mathcal{P}_i^{Sl}$ satisfy

$$\frac{\sigma_{min}(\mathcal{P}_i^{Sl})}{\sigma_{max}(\mathcal{P}_i^{Sl})} > \tau^2 \text{ or equivalently } \frac{\sigma_{min}(Z_i^{Sl})}{\sigma_{max}(Z_i^{Sl})} = \frac{\sigma_{min}(\tilde{Z}_i)}{\sigma_{max}(\tilde{Z}_i)} > \tau$$

for $i = 1, 2, \cdots, k$, where σ_{min} and σ_{max} denote the minimum and maximum singular values, respectively. It readily follows from the implementation of the LR-Smith(l) method that at the $(k+1)^{st}$ step, the approximants Z_{k+1}^{Sl} and $\mathcal{P}_{k+1}^{Sl}$ are given by

$$Z_{k+1}^{Sl} = [Z_k^{Sl} \quad A_d^k B_d] \qquad \text{and} \qquad \mathcal{P}_{k+1}^{Sl} = \mathcal{P}_k^{Sl} + A_d^k B_d B_d^T (A_d^k)^T.$$

Decompose $A_d^k B_d$ into the two spaces $Im(V)$ and $(Im(V))^{\perp}$; i.e., write

$$A_d^k B_d = V\Gamma + \hat{V}\Theta, \tag{2.39}$$

where $\Gamma \in \mathbb{R}^{(mlk)\times(ml)}$, $\Theta \in \mathbb{R}^{(ml)\times(ml)}$, $V^T\hat{V} = 0$ and $\hat{V}^T\hat{V} = I_{ml}$. Define the matrix

$$\hat{Z}_{k+1} = [\ V \quad \hat{V}\ ] \underbrace{\begin{bmatrix} \Phi & \Gamma \\ 0 & \Theta \end{bmatrix}}_{\hat{S}}. \tag{2.40}$$

Let $\hat{S}$ have the following SVD: $\hat{S} = T\hat{\Phi}Y^T$. Then it follows that $\tilde{Z}_{k+1}$ is given by

$$\tilde{Z}_{k+1} = \tilde{V}\hat{\Phi}, \qquad \tilde{V} = [\ V \quad \hat{V}\ ]T, \tag{2.41}$$

where $\tilde{V} \in \mathbb{R}^{n\times((k+1)ml)}$ and $\hat{\Phi} \in \mathbb{R}^{((k+1)ml)\times((k+1)ml)}$. Note that computation of $\tilde{Z}_{k+1}$ requires the knowledge of $\tilde{Z}_k$, which is already available, and the SVD of $\hat{S}$, which is easy to compute. Next, partition $\hat{\Phi}$ and $\tilde{V}$ conformally:

$$\tilde{Z}_{k+1} = [\ \tilde{V}_1 \quad \tilde{V}_2\ ] \begin{bmatrix} \hat{\Phi}_1 & \\ & \hat{\Phi}_2 \end{bmatrix} \text{ so that } \frac{\hat{\Phi}_2(1,1)}{\hat{\Phi}_1(1,1)} < \tau. \tag{2.42}$$

Then, the $(k+1)^{st}$ low-rank Cholesky factor is approximated by

$$\tilde{Z}_{k+1} \approx \tilde{V}_1\hat{\Phi}_1. \tag{2.43}$$

$\tilde{Z}_{k+1}$ in (2.43) is the $(k+1)^{st}$ modified LR-Smith(l) iterate. In computing $\tilde{Z}_{k+1}$, the singular values which are less than the given tolerance τ are truncated. Hence, in going from the k^{th} to the $(k+1)^{st}$ step, the number of columns of $\tilde{Z}_{k+1}$ generally does not increase. An increase will only occur if more than r singular values of $\tilde{Z}_{k+1}$ are above the tolerance $\tau\sigma_1$. In the worst case, at most ml additional columns will be added at any step which is the same as the unmodified LR-Smith(l) iteration discussed in Section 2.4.1.

Using $\tilde{Z}_{k+1}$ in (2.43), the $(k+1)^{st}$ step modified low-rank Smith Gramian is given by

$$\tilde{\mathcal{P}}_{k+1} := \tilde{Z}_{k+1}(\tilde{Z}_{k+1})^T = \tilde{V}_1\hat{\Phi}_1\hat{\Phi}_1^T\tilde{V}_1^T.$$

Convergence Properties of the Modified LR-Smith(l) Iteration

Let $\tilde{\mathcal{P}}_k$ and $\tilde{\mathcal{Q}}_k$ be the k step modified LR-Smith(l) solutions to the two Lyapunov equations $A\mathcal{P} + \mathcal{P}A^T + BB^T = 0$, $\quad A^T\mathcal{Q} + \mathcal{Q}A + C^TC = 0$, respectively, where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{p\times n}$. Moreover let $\mathcal{I}_\mathcal{P}$ denote the set of indices i for which some columns have been eliminated from the i^{th} step approximant during the modified Smith iteration:

$$\mathcal{I}_\mathcal{P} = \{i : \text{such that in (2.42) } \hat{\Phi}_2 \neq 0 \text{ for } \tilde{Z}_i,\ i = 1, 2, \cdots, k\}.$$

Then for each $i \in \mathcal{I}_\mathcal{P}$, let $n_i^\mathcal{P}$ denote the number of the neglected singular values. Similarly define $\mathcal{I}_\mathcal{Q}$ and $n_i^\mathcal{Q}$. The following convergence result holds [GSA03].

Theorem 2.4.6. *Let $\mathcal{P}_k^{Sl}$ be the k^{th} step LR-Smith(l) iterate. $\Delta_{kp} := \mathcal{P}_k^{Sl} - \tilde{\mathcal{P}}_k$, the error between the LR-Smith(l) and Modified LR-Smith(l) iterates, satisfies*

$$\|\Delta_{kp}\| = \|\mathcal{P}_k^{Sl} - \tilde{\mathcal{P}}_k\| \leq \tau^2 \sum_{i \in \mathcal{I}_{\mathcal{P}}} (\sigma_{max}(\tilde{Z}_i))^2, \tag{2.44}$$

where τ is the tolerance value of the modified LR-Smith(l) algorithm. Moreover, define $\tilde{E}_{kp} = \mathcal{P} - \tilde{\mathcal{P}}_k$, the error between the exact solution and the k^{th} Modified LR-Smith(l) iterates. Then,

$$\begin{aligned} 0 &\leq \operatorname{trace}(\tilde{E}_{kp}) \\ &\leq K\, m\, l\, (\rho(A_d))^{2k} \operatorname{trace}(\mathcal{P}) + \tau^2 \sum_{i \in \mathcal{I}_{\mathcal{P}}} n_i^{\mathcal{P}}\ (\sigma_{max}(\tilde{Z}_i))^2, \end{aligned} \tag{2.45}$$

where K is given by (2.36).

Note that the error $\|\Delta_{kp}\|$ is in the order of $\mathcal{O}(\tau^2)$. This means with a lower number of columns in the approximate Cholesky factor, the Modified Smith method will yield almost the same accuracy as the exact Smith method.

The next result concerns the convergence of the computed Hankel singular values in a way analogous to Lemma 2.4.5.

Lemma 2.4.7. *Let σ_i and $\tilde{\sigma}_i$ denote Hankel singular values resulting from the full-rank exact Gramians $\mathcal{P}$ and $\mathcal{Q}$ and from the modified LR-Smith(l) approximants $\tilde{\mathcal{P}}_k$ and $\tilde{\mathcal{Q}}_k$ respectively: ${\sigma_i}^2 = \lambda_i(\mathcal{P}\mathcal{Q})$ and $\tilde{\sigma}_i^2 = \lambda_i(\tilde{\mathcal{P}}_k\tilde{\mathcal{Q}}_k)$. Define $\hat{n} = kl\min(m,p)$. Then,*

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} \sigma_i^2 - \sum_{i=1}^{\hat{n}} \tilde{\sigma}_i^2 \\ &\leq K\, l\, (\rho(A_d))^{2k} \Big(K\ \min(m,p)(\rho(A_d))^{2k} \operatorname{trace}(\mathcal{P}) \operatorname{trace}(\mathcal{Q}) \\ &\quad + m \operatorname{trace}(\mathcal{P}) \sum_{i=0}^{k-1} \|C_d A_d^j\|_2^2 + p \operatorname{trace}(\mathcal{Q}) \sum_{i=0}^{k-1} \|A_d^j B_d\|_2^2 \Big) \\ &\quad + \tau_{\mathcal{P}}^2 \|\mathcal{Q}_k^{Sl}\|_2 \sum_{i \in \mathcal{I}_{\mathcal{P}}} n_i^{\mathcal{P}}\ (\sigma_{max}(\tilde{Z}_i))^2 \\ &\quad + \tau_{\mathcal{Q}}^2 \|\mathcal{P}_k^{Sl}\|_2 \sum_{i \in \mathcal{I}_{\mathcal{Q}}} n_i^{\mathcal{Q}}\ (\sigma_{max}(\tilde{Y}_i))^2) \end{aligned} \tag{2.46}$$

where $\tau_{\mathcal{P}}$ and $\tau_{\mathcal{Q}}$ are the given tolerance values; and K is as defined in (2.36).

Once again the bounds in Lemma 2.4.5 and Lemma 2.4.7 differ only by the summation of terms of $\mathcal{O}(\tau_{\mathcal{P}}^2)$ and $\mathcal{O}(\tau_{\mathcal{Q}}^2)$.

2.4.4 ADI Parameter Selection

As the selection of good parameters is vitally important to the successful application of the ADI and derived algorithms, in this section we discuss two possible approaches. Both seek to solve the minimax problem (2.11), in other words, minimizing the right hand side of the error bound in (2.12).

Because it is not practical to assume the knowledge of the complete spectrum of the matrix A, i.e., not practical to solve (2.11) over $\lambda \in \sigma(A)$, the first approach [WAC95] solves a different problem. It begins by bounding the spectrum of A inside a domain $\mathcal{R} \subset \mathbb{C}_-$, in other words,

$$\lambda_1(A), \cdots, \lambda_n(A) \in \mathcal{R} \subset \mathbb{C}_-,$$

and then solves the following rational minimax problem:

$$\min_{\mu_1,\mu_2,\cdots,\mu_l} \max_{x\in\mathcal{R}} \left| \prod_{j=1}^{l} \frac{(\mu_j - x)}{(\mu_j + x)} \right|, \tag{2.47}$$

where the maximization is done over $x \in \mathcal{R}$ (rather than $\lambda \in \sigma(A)$). In this general formulation, $\mathcal{R}$ can be any region in the open left half plane.

If the eigenvalues of A are strictly real, then one takes the domain $\mathcal{R}$ to be a line segment, with the end points being the extremal eigenvalues of A. In this case the solution to (2.47) is known (see [WAC95]). Power and inverse iterations can be used to estimate the extremal eigenvalues of A at a low cost.

If A has complex eigenvalues, finding a good domain $\mathcal{R}$ which provides an efficient covering of the spectrum of A can be involved, since the convex hull of the spectrum of an arbitrary stable matrix can take on widely varying shapes. Typically one estimates extremal values of the spectrum of A, along the real and the imaginary axes, and then assumes that the spectrum is bounded inside some region which can be simply defined by the extremal values one has obtained.

However, even after a good $\mathcal{R}$ has been obtained, there remains the serious difficulty of solving (2.47). The solution to (2.47) is not known when $\mathcal{R}$ is an arbitrary region in the open left half plane. However, the problem of finding optimal and near-optimal parameters for a few given shapes was investigated in several papers [IT95, EW91, STA91, STA93, WAC62, WAC95] and we give some of the useful results below.

In particular, we summarize a parameter selection procedure from [WAC95] which defines the spectral bounds a, b, and α for the matrix A as

$$a = \min_i(Re\{\gamma_i\}), \quad b = \max_i(Re\{\gamma_i\}), \quad \alpha = \tan^{-1} \max_i \left| \frac{Im\{\gamma_i\}}{Re\{\gamma_i\}} \right|, \tag{2.48}$$

where $\gamma_1, \cdots, \gamma_n$ are the eigenvalues of $-A$. It is assumed that the spectrum of $-A$ lies entirely inside a region which was called in that reference the "elliptic function domain" determined by the numbers a, b, α. The specific definition

of the "elliptic function domain" can be found in [WAC95]. If this assumption does not hold, one should try to apply a more general parameter selection algorithm. If it does hold, then let

$$\cos^2\beta = \frac{2}{1+\frac{1}{2}(\frac{a}{b}+\frac{b}{a})},$$
$$m = \frac{2\cos^2\alpha}{cos^2\beta} - 1.$$

If $m < 1$, the parameters are complex, and are given in [EW91, WAC95]. If $m \geq 1$, the parameters are real, and we define

$$k' = \frac{1}{m+\sqrt{m^2-1}}, \quad k = \sqrt{1-k'^2}.$$

Note $k' = \frac{a}{b}$ if all the eigenvalues of A are real. Define the elliptic integrals K and v as,

$$F[\psi, k] = \int_0^{\psi} \frac{dx}{\sqrt{1-k^2\sin^2 x}},$$
$$K = K(k) = F\left[\frac{\pi}{2}, k\right], \qquad v = F\left[\sin^{-1}\sqrt{\frac{a}{bk'}}, k'\right].$$

The number of ADI iterations required to achieve $\rho^2_{ADI} \leq \epsilon_1$ is given by $l = \left\lceil \frac{K}{2v\pi} \log \frac{4}{\epsilon_1} \right\rceil$, and the ADI parameters are given by

$$\mu_j = -\sqrt{\frac{ab}{k'}} dn\left[\frac{(2j-1)K}{2l}, k\right], \quad j = 1, 2, \cdots, l, \tag{2.49}$$

where $dn(u,k)$ is the elliptic function. It was noted in [LW91] that for many practical problems ADI converges in a few iterations with these parameters.

A second approach to the problem of determining ADI parameters is a heuristic one and was given in [PEN00b]. It chooses potential parameters from a list $\mathcal{S} = \{\rho_1, \rho_2, \cdots, \rho_k\}$ which is taken to be the union of the Ritz values of A and the reciprocals of the Ritz values of A^{-1}, obtained by two Arnoldi processes, with A and A^{-1}. From this list $\mathcal{S}$, one chooses the list of l ADI parameters, $\mathcal{L}$, in the following way. First, we define the quantity

$$s_{\mathcal{M}}(x) := \frac{|(x-\mu_1)\times\cdots\times(x-\mu_m)|}{|(x+\mu_1)\times\cdots\times(x+\mu_m)|},$$

where $\mathcal{M} = \{\mu_1, \cdots, \mu_m\}$. The algorithm proceeds as follows:

1. Find i such that $\max\limits_{x\in\mathcal{S}} s_{\rho_i}(x) = \min\limits_{\rho_i\in\mathcal{S}} \max\limits_{x\in\mathcal{S}} s_{\rho_i}(x)$ and let

$$\mathcal{L} := \begin{cases} \{\rho_i\} & \text{if } \rho_i \text{ real,} \\ \{\rho_i, \bar{\rho}_i\} & \text{otherwise.} \end{cases}$$

2. While $\text{card}(\mathcal{L}) < l$, find i such that $s_{\mathcal{L}}(\rho_i) = \max_{x \in \mathcal{L}} s_{\mathcal{L}}(x)$ and let

$$\mathcal{L} := \begin{cases} \mathcal{L} \cup \{\rho_i\} & \text{if } \rho_i \text{ real,} \\ \mathcal{L} \cup \{\rho_i, \bar{\rho}_i\} & \text{otherwise.} \end{cases}$$

The procedure is easy to implement and good results have been obtained [PEN00b].

2.5 Smith's Method and Eigenvalue Decay Bounds for Gramians

As discussed earlier, in most cases, the eigenvalues of the reachability and observability Gramians $\mathcal{P}, \mathcal{Q}$, as well as the Hankel singular values, i.e., $\sqrt{\lambda_i(\mathcal{PQ})}$, decay very rapidly. In this section, we briefly review the results of [ASZ02, ZHO02] and reveal the connection to convergence of Smith-type iterations. We will again consider the Lyapunov equation

$$A\mathcal{P} + \mathcal{P}A^T + BB^T = 0, \tag{2.50}$$

where $B \in \mathbb{R}^{n \times m}$ with $m \ll n$, $A \in \mathbb{R}^{n \times n}$ is asymptotically stable, and the pair (A, B) is reachable.

2.5.1 Eigenvalue Decay Bounds for the Solution $\mathcal{P}$

Given the Lyapunov equation (2.50), let an l step ADI iteration be computed using the shifts μ_i, with $\mu_i < 0$ where $i = 1, \ldots, l$ and $lm < n$. Then it simply follows from (2.10) that

$$\text{rank}(\mathcal{P}_{i-1}^A) \leq \text{rank}(\mathcal{P}_i^A) \leq \text{rank}(\mathcal{P}_{i-1}^A) + m.$$

Hence, at the l^{th} step, one has

$$\text{rank}(\mathcal{P}_l^A) \leq lm.$$

Then by Schmidt-Mirsky theorem and considering $\mathcal{P}_l^A$ as a low-rank approximation to $\mathcal{P}$, one simply obtains

$$\frac{\lambda_{lm+1}(\mathcal{P})}{\lambda_1(\mathcal{P})} \leq \|A_d\|_2^2,$$

where A_d is given by (2.17). The following result holds:

Theorem 2.5.1. *Given the above set-up, let A be diagonalizable. Then, eigenvalues of the solution $\mathcal{P}$ to the Lyapunov equation (2.50) satisfy*

$$\frac{\lambda_{lm+1}(\mathcal{P})}{\lambda_1(\mathcal{P})} \leq K(\rho(A_d))^2, \tag{2.51}$$

where $lm < n$, K is given by (2.36), $\rho_{ADI} = \rho(A_d)$ as before and the shifts μ_i are chosen by solving the ADI minimax problem (2.11).

See the original source [ASZ02] and [ZHO02] for details and a proof.

2.5.2 Connection Between Convergence of the Smith Iteration and Theorem 2.5.1

Smith (or ADI) type iterations try to approximate the exact Gramian $\mathcal{P}$ with a low-rank version in which the convergence of the iteration is given by either (2.12) or Proposition (2.4.4). Hence, if $\rho(A_d)$ is close to 1 and/or K is big, we expect slow convergence. The slow convergence leads to more steps in the Smith iteration, and, consequently, the rank of the approximant is higher. Since $\mathcal{P}$ is positive definite, in turn, this means that eigenvalues of $\mathcal{P}$ do not decay rapidly. Therefore $\rho(A_d) \approx 1$ and/or K is big mean that $\lambda_i(\mathcal{P})$ might decay slowly. This final remark is consistent with the above decay bound (2.51). These relations are expected since (2.51) is derived via the ADI iteration.

As stated in [ZHO02] and [ASZ02], (2.51) yields the following remarks:

1. If $\lambda_i(A)$ are clustered in the complex plane, choosing the shifts μ_i as the clustered points yields a small $\rho(A_d)$, and consequently fast decay of $\lambda_i(\mathcal{P})$. Hence, the convergence of an ADI-type iteration is fast.
2. If $\lambda_i(A)$ have mostly dominant real parts, then the decay rate is again fast. Hence, as above, the convergence of an ADI-type iteration is fast.
3. If $\lambda_i(A)$ have mostly dominant imaginary parts, while the real parts are relatively small, the decay rate $\lambda_i(\mathcal{P})$ is slow. Then an ADI iteration converges slowly.

These observations agree with the numerical simulations. In Example 2.7.2, the Smith(l) method is applied to a CD player example, a system of order 120, where the eigenvalues of A are scattered in the complex plane with dominant complex parts. Even with a high number of shifts, $\rho(A_d)$ cannot be reduced less than 0.98, and the Smith methods converge very slowly. Indeed, an exact computation of $\mathcal{P}$ reveals that $\mathcal{P}$ does not have rapidly decaying eigenvalues. Also, it was shown in [ASG01] that the Hankel singular values of this system decay slowly as well, and the CD player was among the hardest models to approximate. These results are consistent with item 3. above.

Item 2. is encountered in Example 2.7.2, where the Smith method is applied to a model of order 1006. 1000 of the eigenvalues are real and only the remaining 6 are complex. By choosing the shifts as the complex eigenvalues, $\rho(A_d)$ is reduced to a small value and convergence is extremely fast. Indeed, using the modified Smith method, the exact Gramians are approximated very well with low-rank Gramians having rank of only 19. We note that the shifts are even not the optimal ones.

2.6 Approximate Balanced Truncation and its Stability

Recall the implementation of balanced truncation presented in Section 2.2.3. An exact balanced truncation requires the knowledge of Cholesky factors U

and L of the Gramians $\mathcal{P}$ and $\mathcal{Q}$, i.e. $\mathcal{P} = UU^T$ and $\mathcal{Q} = LL^T$ where $\mathcal{P}$ and $\mathcal{Q}$ are the solutions to the two Lyapunov equations

$$A\mathcal{P} + \mathcal{P}A^T + BB^T = 0 \quad \text{and} \quad A^T\mathcal{Q} + \mathcal{Q}A + C^TC = 0.$$

As mentioned earlier, in large-scale settings, obtaining U and L is a formidable task. In this section, we will discuss approximate balanced truncation of large-sparse dynamical systems, where the approximate low-rank Cholesky factors are used in place of the exact Gramians in computing the reduced-order model. Hence, we will replace the full-rank Cholesky factors U and L with the low-rank ones, namely $\tilde{U}$ and $\tilde{L}$ which are obtained through a k step Smith-type iteration. For details see [GSA03]. For simplicity, let us assume that the original model is SISO. Proceeding similarly to Section 2.2.3, let $\tilde{U}^T\tilde{L} = \tilde{Z}\tilde{\Sigma}\tilde{Y}^T$ be the singular value decomposition (SVD) with $\tilde{\Sigma} = \mathrm{diag}(\tilde{\sigma}_1, \cdots, \tilde{\sigma}_k)$ where $\tilde{\sigma}_i$ are the approximate Hankel singular values with $\tilde{\sigma}_1 > \tilde{\sigma}_2 > \cdots > \tilde{\sigma}_k$. Here we have assumed, for the brevity of the discussion, that the Hankel singular values are distinct. Now define

$$\tilde{W}_1 := \tilde{L}\tilde{Y}_1\tilde{\Sigma}_1^{-1/2} \quad \text{and} \quad \tilde{V}_1 := \tilde{U}\tilde{Z}_1\tilde{\Sigma}_1^{-1/2},$$

where $\tilde{Z}_1$ and $\tilde{Y}_1$ are composed of the leading r columns of $\tilde{Z}$ and $\tilde{Y}$ respectively, and $\tilde{\Sigma}_1 = \mathrm{diag}(\tilde{\sigma}_1, \cdots, \tilde{\sigma}_r)$. We note that the equality $\tilde{W}_1^T\tilde{V}_1 = I_r$ still holds and hence that $\tilde{V}_1\tilde{W}_1^T$ is an oblique projection. The approximately balanced reduced model $\tilde{\mathbf{\Sigma}}_r$ of order r is obtained as

$$\tilde{A}_r = \tilde{W}_1^T A\tilde{V}_r, \quad \tilde{B}_r = \tilde{W}_1^T B, \quad C_r = C\tilde{V}_1, \quad \text{and} \quad \tilde{D}_r = D.$$

To examine the stability of this reduced model, we first define the error term in $\mathcal{P}$. Define Δ as

$$\Delta := \tilde{U}\tilde{U}^T - UU^T = \tilde{\mathcal{P}} - \mathcal{P}.$$

Then one can show that

$$\tilde{A}_r\tilde{\Sigma}_1 + \tilde{\Sigma}_1\tilde{A}_r^T + \tilde{B}\tilde{B}_r{}^T = \tilde{W}_1^T(A\Delta + \Delta A^T)\tilde{W}_1 \tag{2.52}$$

We know that $\tilde{\Sigma}_1 > 0$. Hence to apply Lyapunov's inertia theorem, we need

$$\tilde{B}\tilde{B}_r{}^T - \tilde{W}_1^T(A\Delta + \Delta A^T)\tilde{W}_1 = \tilde{W}_1^T(BB^T - A\Delta - \Delta A^T)\tilde{W}_1 \geq 0. \tag{2.53}$$

Unfortunately, this is not always satisfied, and therefore *one cannot guarantee the stability of the reduced system.* However, we would like to note many researchers have observed that this does not seem to be a difficulty in practice; in most cases approximate balanced truncation via a Smith-type iteration yields a stable reduced system and instability is not an issue; see, for example, [GSA03], [GA01], [PEN99], [LW01], [LW99] and the references there in.

Let $\mathbf{\Sigma}_r = \left[\begin{array}{c|c} A_r & B_r \\ \hline C_r & D \end{array}\right]$ and $\tilde{\mathbf{\Sigma}}_r = \left[\begin{array}{c|c} \tilde{A}_r & \tilde{B}_r \\ \hline \tilde{C}_r & D \end{array}\right]$ be the r^{th} order reduced systems obtained by exact and approximate balancing, respectively. Now we examine

closeness of $\mathbf{\Sigma}_r$ to $\tilde{\mathbf{\Sigma}}_r$. Define $\Delta_V := V_1 - \tilde{V}_1$ and $\Delta_W := W_1 - \tilde{W}_1$, and let $\|\Delta_V\| \leq \tau$ and $\|\Delta_W\| \leq \tau$ where τ is a small number; in other words, we assume that $\tilde{V}_1$ and $\tilde{W}_1$ are close to V_1 and W_1, respectively. Under certain assumptions (see [GSA03]), one can show that

$$\|\mathbf{\Sigma}_r - \tilde{\mathbf{\Sigma}}_r\|_\infty \leq \tau \left(\|C_r\| \, \|B_r\| \, \|A_r\| \, \left(\|W_1\| + \|V_1\| \right) + \|\mathbf{\Sigma}_1\|_\infty \|B_r\| + \|\mathbf{\Sigma}_2\|_\infty \|C_r\| \right) + \mathcal{O}(\tau^2) \quad (2.54)$$

where $\mathbf{\Sigma}_1 := \left[\begin{array}{c|c} A_r & I \\ \hline C_r & \end{array}\right]$ and $\mathbf{\Sigma}_2 := \left[\begin{array}{c|c} A_r & B_r \\ \hline I & \end{array}\right]$. Hence for small τ, i.e., when $\tilde{V}_1$ and $\tilde{W}_1$ are, respectively, close to V_1 and W_1, we expect $\mathbf{\Sigma}_r$ to be close to $\tilde{\mathbf{\Sigma}}_r$. Indeed as the examples in Section 2.7 show, $\tilde{\mathbf{\Sigma}}_r$ behaves much better than the above upper bound predicts and $\tilde{\mathbf{\Sigma}}_r$, the approximately balanced system using low-rank Gramians, is almost the same as the exactly balanced system. These observations reveal the effectiveness of the Smith-type methods for balanced truncation of large-sparse dynamical systems.

2.7 Numerical Examples

In this section we give numerical results on the CF–ADI method as well as the LR-Smith(l) and Modified LR-Smith(l) methods.

2.7.1 CF–ADI and the Spiral Inductor

We begin with the CF–ADI approximation to the Lyapunov equation

$$A\mathcal{X} + \mathcal{X}A^T + BB^T = 0.$$

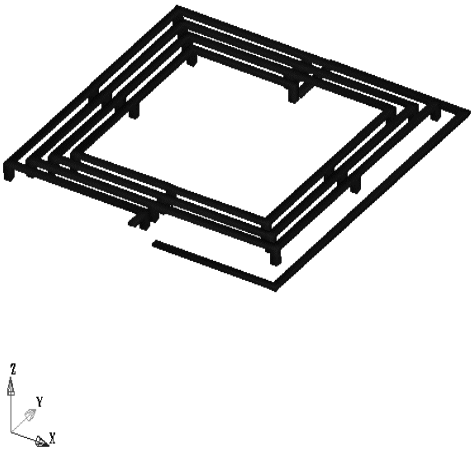
The example in Figure 2.1 comes from the inductance extraction of an on-chip planar square spiral inductor suspended over a copper plane [KWW98], shown in Figure 1(a). (See Chapter 23 for a detailed description of the spiral inductor.) The original order 500 system has been symmetrized according to [SKEW96]. The matrix A is a symmetric 500×500 matrix, and the input coefficient matrix $B \in \mathbb{R}^n$ has one column.

Because A is symmetric, the eigenvalues of A are real and good CF–ADI parameters are easy to find. The procedure given in Section 2.4.4 was followed. CF–ADI was run to convergence in this example, which took 20 iterations.

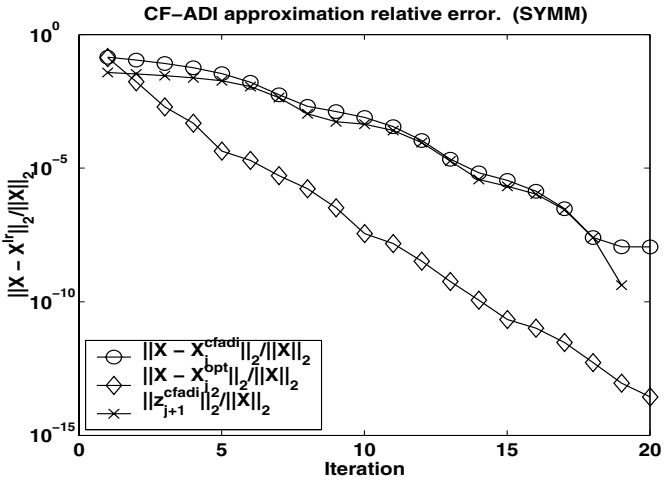
Figure 1(b) shows the relative 2-norm error of the CF–ADI approximation, i.e.

$$\frac{\|\mathcal{X} - \mathcal{X}_j^{cfadi}\|_2}{\|\mathcal{X}\|_2},$$

where $\mathcal{X}$ is the exact solution to $A\mathcal{X} + \mathcal{X}A^T + BB^T = 0$ and $\mathcal{X}_j^{cfadi}$ is the jth CF–ADI approximation, for $j = 1, \cdots, 20$. To illustrate the quality of



(a) Spiral inductor



(b) CF–ADI approximation

Fig. 2.1. Spiral inductor, a symmetric system.

the low-rank approximation, we compare it with the optimal 2-norm rank-j approximation to X[GVL96], denoted X_j^{opt}, obtained from the singular value decomposition of exact solution X. At $j = 20$, the relative error of the CF–ADI approximation has reached 10^{-8}, which is about the same size as the error of the optimal rank 11 approximation. The error estimate $\|z_{j+1}^{CFA}\|_2^2$ approximates the actual error $\|\mathcal{X} - \mathcal{X}_j^{cfadi}\|$ closely for all j.

2.7.2 LR-Smith(l) and Modified LR-Smith(l) Methods

In this section we apply LR-Smith(l) and Modified LR-Smith(l) methods to two dynamical systems. In each example, both the LR-Smith(l) iterates $\mathcal{P}_k^{Sl}$, and $\mathcal{Q}_k^{Sl}$; and the modified LR-Smith(l) iterates $\tilde{\mathcal{P}}_k$, and $\tilde{\mathcal{Q}}_k$ are computed. Also balanced reduction is applied using the full rank Gramians $\mathcal{P}$, $\mathcal{Q}$ and the approximate Gramians $\mathcal{P}_k^{Sl}$, $\mathcal{Q}_k^{Sl}$; and $\tilde{\mathcal{P}}_k$, $\tilde{\mathcal{Q}}_k$. The resulting reduced order systems are compared.

CD Player Model

This example is described in Chapter 24, Section 4, this volume. The full order model (FOM) describes the dynamics of a portable CD player, is of order 120, and single-input single-output. The eigenvalues of A are scattered in the complex plane with relatively large imaginary parts. This makes it harder to obtain a low $\rho(A_d)$. A single shift results in $\rho(A_d) = 0.99985$. Indeed, even with a high number of multiple shifts, $l = 40$, $\rho(A_d)$ could not be reduced to less than 0.98. Hence only a single shift is considered. This observation agrees with the discussion in Section 2.5 that when the eigenvalues of A are scattered in the complex plane, ADI-type iterations converge slowly. LR-Smith(l) and the modified LR-Smith(l) iterations are run for $k = 70$ iterations. For the Modified Smith(l) iteration, the tolerance values are chosen to be

$$\tau_{\mathcal{P}} = 1 \times 10^{-6} \qquad \text{and} \qquad \tau_{\mathcal{Q}} = 8 \times 10^{-6}.$$

The low-rank LR-Smith(l) yields Cholesky factors Z_k^{Sl} and Y_k^{Sl} with 70 columns. On the other hand, the modified LR-Smith(l) yields low-rank Cholesky factors $\tilde{Z}_k$ and $\tilde{Y}_k$ with only 25 columns. To check the closeness of modified Smith iterates to the exact Smith iterates, we compute the following relative error norms:

$$\frac{\|\mathcal{P}_k^{Sl} - \tilde{\mathcal{P}}_k\|}{\|\mathcal{P}_k^{Sl}\|} = 4.13 \times 10^{-10}, \text{ and} \qquad \frac{\|\mathcal{Q}_k^{Sl} - \tilde{\mathcal{Q}}_k\|}{\|\mathcal{Q}_k^{Sl}\|} = 2.33 \times 10^{-10}.$$

Although the number of columns of the Cholesky factor have been reduced from 70 to 25, the Modified Smith method yields almost the same accuracy. We also look at the error between the exact and approximate Gramians:

$$\frac{\|\mathcal{P} - \mathcal{P}_k^{Sl}\|}{\|\mathcal{P}\|} = \frac{\|\mathcal{P} - \tilde{\mathcal{P}}_k\|}{\|\mathcal{P}\|} = 3.95 \times 10^{-3},$$

$$\frac{\|\mathcal{Q} - \mathcal{Q}_k^{Sl}\|}{\|\mathcal{Q}\|} = \frac{\|\mathcal{Q} - \tilde{\mathcal{Q}}_k\|}{\|\mathcal{Q}\|} = 8.24 \times 10^{-1}.$$

Next, we reduce the order of the FOM to $r = 12$ by balanced truncation using both the approximate and the exact solutions. $\mathbf{\Sigma}_k$, $\mathbf{\Sigma}_k^{Sl}$ and $\tilde{\mathbf{\Sigma}}_k$ denote

the 12^{th} order reduced systems obtained through balanced reduction using the exact Cholesky factors Z and Y; the LR-Smith(l) iterates Z_k^{Sl} and Y_k^{Sl}; and the modified LR-Smith(l) iterates $\tilde{Z}_k$ and $\tilde{Y}_k$ respectively. Also $\mathbf{\Sigma}$ denotes the FOM.

Figure 2.2 depicts the amplitude Bode plots of the FOM $\mathbf{\Sigma}$ and the reduced balanced systems $\mathbf{\Sigma}_k$, $\mathbf{\Sigma}_k^{Sl}$ and $\tilde{\mathbf{\Sigma}}_k$. As can be seen, although relative error between the exact and the approximate Gramians are not very small, $\mathbf{\Sigma}_k^{Sl}$ and $\tilde{\mathbf{\Sigma}}_k$ show a very similar behavior to $\mathbf{\Sigma}_k$. This observation reveals that even if the relative error in the approximate Gramians are big, if the dominant eigenspace of $\mathcal{PQ}$, and hence the largest HSV are matched well, approximate balanced truncation performs very closely to the exact balanced truncation. Similar observations can be found in [GA01, GUG03]. The amplitude Bode plots of the error systems $\mathbf{\Sigma} - \mathbf{\Sigma}_k$, $\mathbf{\Sigma} - \mathbf{\Sigma}_k^{Sl}$ and $\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}}_k$ are illustrated in Figure 2.3. It is also important to note that since the errors between $\tilde{\mathcal{P}}_k$ and $\mathcal{P}_k^{Sl}$, and $\tilde{\mathcal{Q}}_k$ and $\mathcal{Q}_k^{Sl}$ are small, $\mathbf{\Sigma}_k^{Sl}$ and $\tilde{\mathbf{\Sigma}}_k$ are almost equal as expected. The relative $\mathcal{H}_\infty$ norms of the error systems are tabulated in Table 2.1.

Table 2.1. Numerical Results for CD Player Model

$\|\mathbf{\Sigma} - \mathbf{\Sigma}_k\|_{\mathcal{H}_\infty}$	$\|\mathbf{\Sigma} - \mathbf{\Sigma}_k^{Sl}\|_{\mathcal{H}_\infty}$	$\left\|\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}}_k\right\|_{\mathcal{H}_\infty}$	$\left\|\mathbf{\Sigma}_k^{Sl} - \tilde{\mathbf{\Sigma}}_k\right\|_{\mathcal{H}_\infty}$
9.88×10^{-4}	9.71×10^{-4}	9.69×10^{-4}	5.11×10^{-6}

$\|\mathbf{\Sigma}_k - \mathbf{\Sigma}_k^{Sl}\|_{\mathcal{H}_\infty}$	$\left\|\mathbf{\Sigma}_k - \tilde{\mathbf{\Sigma}}_k\right\|_{\mathcal{H}_\infty}$
1.47×10^{-4}	1.47×10^{-4}

A Random System

This model is from [PEN99] and the example from [GSA03, GUG03]. The FOM is a dynamical system of order 1006. The state-space matrices of the full-order model $\mathbf{\Sigma} = \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$ are given by

$$A = \mathrm{diag}(A_1, A_2, A_3, A_4), \quad B^T = C = [\ \underbrace{10\ \cdots\ 10}_{6}\ \underbrace{1\ \cdots 1}_{1000}]$$

where

$$A_1 = \begin{bmatrix} -1 & 100 \\ -100 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -1 & 200 \\ -200 & -1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} -1 & 400 \\ -400 & -1 \end{bmatrix},$$
$$\text{and} \quad A_4 = \mathrm{diag}(-1, \cdots, -1000). \qquad .$$

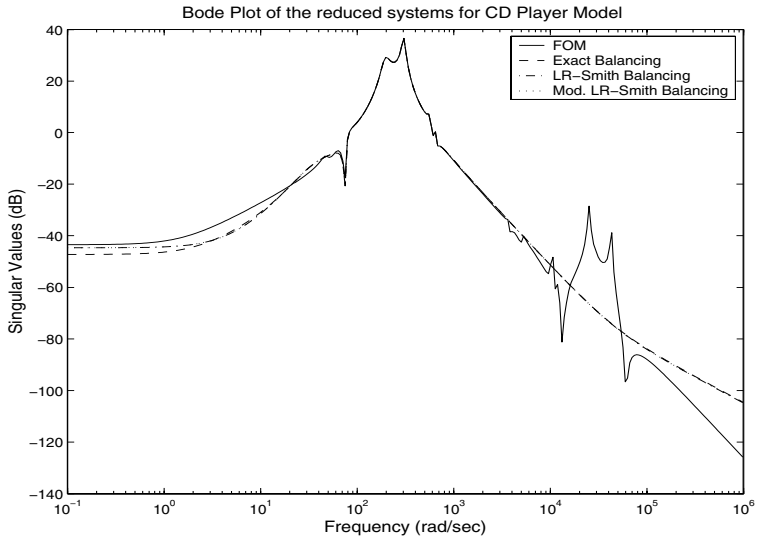


Fig. 2.2. The amplitude Bode plots of the FOM $\mathbf{\Sigma}$ and the reduced systems $\mathbf{\Sigma}_k$ (Exact Balancing), $\mathbf{\Sigma}_k^{Sl}$ (LR-Smith Balancing) and $\tilde{\mathbf{\Sigma}}_k$ (Mod. LR-Smith Balancing) for the CD Player Model

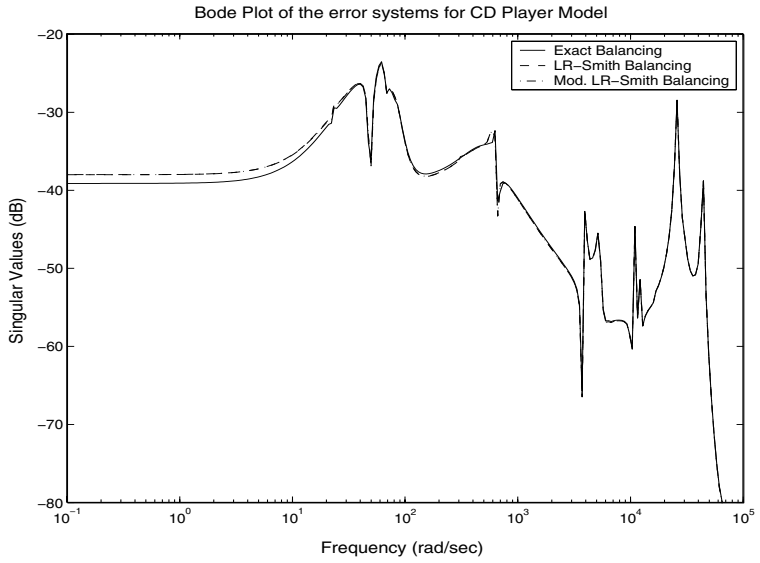


Fig. 2.3. The amplitude Bode plots of error systems $\mathbf{\Sigma} - \mathbf{\Sigma}_k$ (Exact Balancing), $\mathbf{\Sigma} - \mathbf{\Sigma}_k^{Sl}$ (LR-Smith Balancing) and $\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}}_k$ (Mod. LR-Smith Balancing) for the CD Player Model

The spectrum of A is

$$\sigma(A) = \{-1, -2, \cdots, -1000, -1 \pm 100\jmath, -1 \pm 200\jmath, -1 \pm 400\jmath\}.$$

LR-Smith(l) and modified LR-Smith(l) methods are applied using $l = 10$ cyclic shifts. Six of the shifts are chosen so that the 6 complex eigenvalues of A are eliminated. This shift selection reduces the ADI spectral radius $\rho(A_d)$ to 0.7623, and results in a fast convergence. Once more, the numerical results support the discussion in Section 2.5. Since the eigenvalues are mostly real, with an appropriate choice of shifts, the spectral radius can be easily reduced to a small number yielding a fast convergence. Both LR-Smith(l) and the modified LR-Smith(l) iterations are run for $k = 30$ iterations with the tolerance values

$$\tau_{\mathcal{P}} = \tau_{\mathcal{Q}} = 3 \times 10^{-5}$$

for the latter. The resulting LR-Smith(l) and modified LR-Smith(l) Cholesky factors has 300 and 19 columns, respectively. Even though the number of columns in the modified method is much less than the exact LR-Smith(l) method, almost there is no lost of accuracy in the computed Gramian as the following numbers show:

$$\frac{\|\mathcal{P}_k^{Sl} - \tilde{\mathcal{P}}_k\|}{\|\mathcal{P}_k^{Sl}\|} = 1.90 \times 10^{-8}, \text{ and } \qquad \frac{\|\mathcal{Q}_k^{Sl} - \tilde{\mathcal{Q}}_k\|}{\|\mathcal{Q}_k^{Sl}\|} = 3.22 \times 10^{-8}.$$

The errors between the exact and computed Gramians are as follows:

$$\frac{\|\mathcal{P} - \mathcal{P}_k^{Sl}\|}{\|\mathcal{P}\|} = 4.98 \times 10^{-10}, \quad \frac{\|\mathcal{P} - \tilde{\mathcal{P}}_k\|}{\|\mathcal{P}\|} = 1.88 \times 10^{-8}$$

$$\frac{\|\mathcal{Q} - \mathcal{Q}_k^{Sl}\|}{\|\mathcal{Q}\|} = 4.98 \times 10^{-10}, \quad \frac{\|\mathcal{Q} - \tilde{\mathcal{Q}}_k\|}{\|\mathcal{Q}\|} = 3.21 \times 10^{-8}.$$

Unlike the CD Player model, since $\rho(A_d)$ is small, the iterations converge fast, and both $\mathcal{P}_k^{Sl}$ and $\tilde{\mathcal{P}}_k$ ($\mathcal{Q}_k^{Sl}$ and $\tilde{\mathcal{Q}}_k$) are very close to the exact Gramian $\mathcal{P}$ (to $\mathcal{Q}$).

We reduce the order of the FOM to $r = 11$ using both exact and approximate balanced truncation. As in the CD Player example, $\mathbf{\Sigma}_k$, $\mathbf{\Sigma}_k^{Sl}$ and $\tilde{\mathbf{\Sigma}}_k$ denote the reduced systems obtained through balanced reduction using the exact Cholesky factors Z and Y; the LR-Smith(l) iterates Z_k^{Sl} and Y_k^{Sl}; and the modified LR-Smith(l) iterates $\tilde{Z}_k$ and $\tilde{Y}_k$ respectively. Figure 2.4 depicts the amplitude Bode plots of the FOM $\mathbf{\Sigma}$ and the reduced systems $\mathbf{\Sigma}_k$, $\mathbf{\Sigma}_k^{Sl}$ and $\tilde{\mathbf{\Sigma}}_k$. As Figure 2.4 illustrates, all the reduced models match the FOM quite well. More importantly, the approximate balanced truncation using the low-rank Gramians yields almost the same result as the exact balanced truncation. These results once more prove the effectiveness of the Smith-type methods. The amplitude Bode plots of the error systems $\mathbf{\Sigma} - \mathbf{\Sigma}_k$, $\mathbf{\Sigma} - \mathbf{\Sigma}_k^{Sl}$ and $\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}}_k$

are illustrated in Figure 2.5 and all the relative $\mathcal{H}_\infty$ norms of the error systems are tabulated in Table 2.2. As in the previous example, $\mathbf{\Sigma}_k^{Sl}$ and $\tilde{\mathbf{\Sigma}}_k$ are almost identical. The relative $\mathcal{H}_\infty$ norm of the error $\mathbf{\Sigma}_k^{Sl} - \tilde{\mathbf{\Sigma}}_k$ is $\mathcal{O}(10^{-9})$. We note that $\mathbf{\Sigma}_k^{Sl}$ has been obtained using a Cholesky factor with 300 columns; on the other hand $\tilde{\mathbf{\Sigma}}_k$ has been obtained using a Cholesky factor with only 19 columns, which proves the effectiveness of the modified Smith's method.

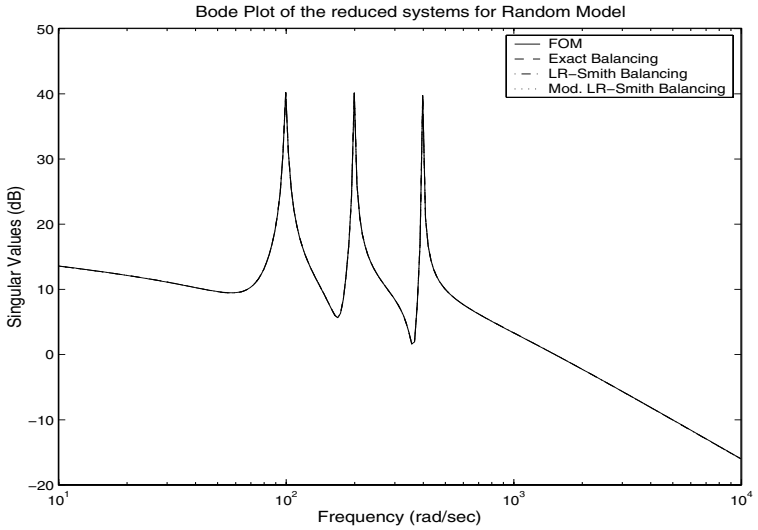


Fig. 2.4. The amplitude Bode plots of the FOM $\mathbf{\Sigma}$ and the reduced systems $\mathbf{\Sigma}_k$ (Exact Balancing), $\mathbf{\Sigma}_k^{Sl}$ (LR-Smith Balancing) and $\tilde{\mathbf{\Sigma}}_k$ (Mod. LR-Smith Balancing) for the Random Model

Table 2.2. Numerical Results for the Random Model

$\|\mathbf{\Sigma} - \mathbf{\Sigma}_k\|_{\mathcal{H}_\infty}$	$\left\|\mathbf{\Sigma} - \mathbf{\Sigma}_k^{Sl}\right\|_{\mathcal{H}_\infty}$	$\left\|\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}}_k\right\|_{\mathcal{H}_\infty}$	$\left\|\mathbf{\Sigma}_k^{Sl} - \tilde{\mathbf{\Sigma}}_k\right\|_{\mathcal{H}_\infty}$
1.47×10^{-4}	1.47×10^{-4}	1.47×10^{-4}	2.40×10^{-9}

$\|\mathbf{\Sigma}_k - \mathbf{\Sigma}_k^{Sl}\|_{\mathcal{H}_\infty}$	$\|\mathbf{\Sigma}_k - \tilde{\mathbf{\Sigma}}_k\|_{\mathcal{H}_\infty}$
7.25×10^{-11}	7.25×10^{-11}

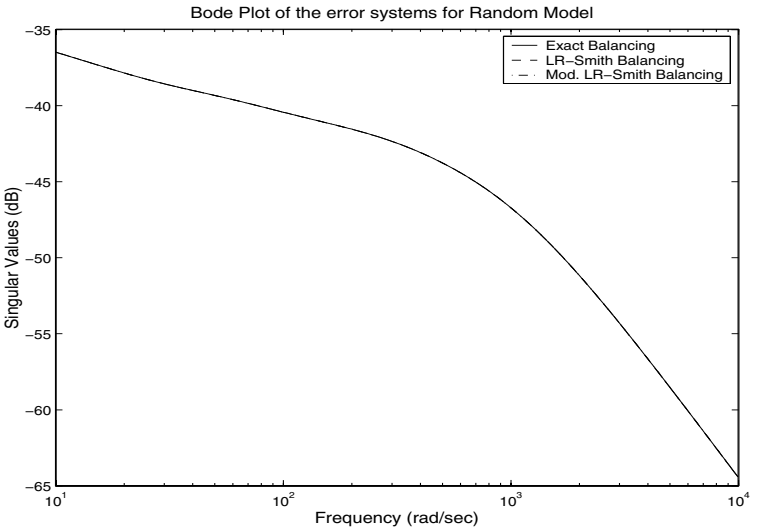


Fig. 2.5. The amplitude Bode plots of error systems $\mathbf{\Sigma} - \mathbf{\Sigma}_k$ (Exact Balancing), $\mathbf{\Sigma} - \mathbf{\Sigma}_k^{Sl}$ (LR-Smith Balancing) and $\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}}_k$ (Mod. LR-Smith Balancing) for the Random Model

2.8 Conclusions

We have reviewed several low-rank methods to solve Lyapunov equations which are based on Smith-type methods, with the goal of facilitating the efficient model reduction of large-scale linear systems. The low-rank methods covered included the Low-Rank ADI method, the Cholesky Factor ADI method, the Low-Rank Smith(l) method, and the modified Low-Rank Smith (l) method. The low-rank factored versions of the ADI method reduced the work required from $\mathcal{O}(n^3)$ to $\mathcal{O}(n)$ for sparse matrices and the required storage from $\mathcal{O}(n^2)$ to $O(nr)$ where r is the numerical rank of the solution. Because these low-rank methods produce the Cholesky factor of the solution to the Lyapunov equation, they are especially well-suited to be used in conjunction with approximate balanced truncation to reduce large-scale linear systems.

References

[ASG01] A. C. Antoulas, D. C. Sorensen, and S. Gugercin. A survey of model reduction methods for large-scale systems. *Contemporary Mathematics, AMS Publications*, **280**, 193–219 (2001).

[ASZ02] A.C. Antoulas, D.C. Sorensen, and Y.K. Zhou. On the decay rate of Hankel singular values and related issues. Systems and Control Letters, **46:5**, 323–342 (2002).

[AS02] A.C. Antoulas and D.C. Sorensen. The Sylvester equation and approximate balanced reduction. Linear Algebra and Its Applications, **351–352**, 671–700 (2002).

[ANT05] A.C. Antoulas. Lectures on the approximation of linear dynamical systems. Advances in Design and Control, SIAM, Philadelphia (2005).

[BS72] R. H. Bartels and G. W. Stewart. Solution of the matrix equation $AX + XA = C$: Algorithm 432. Comm. ACM, **15**, 820–826 (1972).

[BQ99] P. Benner and E. S. Quintana-Ortí. Solving stable generalized Lyapunov equation with the matrix sign function. *Numerical Algorithms*, **20**, 75–100 (1999).

[BQQ01] P. Benner, E. S. Quintana-Ortí, and G. Quintana-Ortí. Efficient Numerical Algorithms for Balanced Stochastic Truncation. International Journal of Applied Mathematics and Computer Science, **11**:5, 1123–1150 (2001).

[CR96] D. Calvetti and L, Reichel. Application of ADI iterative methods to the restoration of noisy images. SIAM J. Matrix Anal. Appl., **17**, 165–186 (1996).

[DP84] U.B. Desai and D. Pal. A transformation approach to stochastic model reduction. IEEE Trans. Automat. Contr., vol **AC-29**, 1097–1100 (1984).

[EW91] N. Ellner and E. Wachspress. Alternating direction implicit iteration for systems with complex spectra. SIAM J. Numer. Anal., **28**, 859–870 (1991).

[ENN84] D. Enns. Model reduction with balanced realizations: An error bound and a frequency weighted generalization. In *Proc.* 23^{rd} *IEEE Conf. Decision and Control* (1984).

[GRE88a] M. Green. A relative error bound for balanced stochastic truncation. IEEE Trans. Automat. Contr., **AC-33**:10, 961–965 (1988).

[GRE88b] M. Green. Balanced stochastic realizations. Journal of Linear Algebra and its Applications, **98**, 211–247 (1988).

[GJ90] W. Gawronski and J.-N. Juang. Model reduction in limited time and frequency intervals. Int. J. Systems Sci., **21**:2, 349–376 (1990).

[GLO84] K. Glover. All Optimal Hankel-norm Approximations of Linear Mutilvariable Systems and their L^{∞}-error Bounds. Int. J. Control, **39**, 1115–1193 (1984).

[GVL96] G. Golub. and C. Van Loan. Matrix computations, 3rd Ed., Johns Hopkins University Press, Baltimore, MD (1996).

[GSA03] S. Gugercin, D.C. Sorensen, and A.C. Antoulas. A modified low-rank Smith method for large-scale Lyapunov equations. Numerical Algorithms, **32**:1, 27–55 (2003).

[GA01] S. Gugercin and A. C. Antoulas. Approximation of the International Space Station 1R and 12A models. In *Proc.* 40^{th} *CDC* (2001).

[GUG03] S. Gugercin. Projection methods for model reduction of large-scale dynamical systems. Ph.D. Dissertation, ECE Dept., Rice University, Houston, TX, USA, May 2003.

[GA04] S. Gugercin and A.C. Antoulas. A survey of model reduction by balaned truncation and some new results. Int. J. Control, **77**:8, 748–766 (2004).

[IT95] M.-P. Istace and J.-P. Thiran. On the third and fourth Zolotarev problems in the complex plane. SIAM J. Numer. Anal., **32**:1, 249–259 (1995).

[HAM82] S. Hammarling. Numerical solution of the stable, non-negative definite Lyapunov equation. IMA J. Numer. Anal., **2**, 303–323 (1982).

[HPT96] A. S. Hodel, K.P. Poola, and B. Tenison. Numerical solution of the Lyapunov equation by approximate power iteration. Linear Algebra Appl., **236**, 205–230 (1996).

[HR92] D. Y. Hu and L. Reichel. Krylov subspace methods for the Sylvester equation. Linear Algebra Appl., **172**, 283–313, (1992).

[JK94] I. M. Jaimoukha and E. M. Kasenally. Krylov subspace methods for solving large Lyapunov equations. SIAM J. Numerical Anal., **31**, 227–251 (1994).

[JK97] I.M. Jaimoukha, E.M. Kasenally. Implicitly restarted Krylov subspace methods for stable partial realizations. SIAM J. Matrix Anal. Appl., **18**, 633–652 (1997).

[KWW98] M. Kamon and F. Wang and J. White. Recent improvements for fast inductance extracton and simulation [packaging]. *Proceedings of the IEEE 7th Topical Meeting on Electrical Performance of Electronic Packaging*, 281–284 (1998).

[LW02] J.-R. Li and J. White. Low rank solution of Lyapunov equations. SIAM J. Matrix Anal. Appl., **24**:1, 260–280 (2002).

[LW99] J.-R. Li and J. White. Efficient model reduction of interconnect via approximate system Gramians. In *Proc. IEEE/ACM Intl. Conf. CAD*, 380–383, San Jose, CA (1999).

[LW01] J.-R. Li and J. White. Reduction of large-circuit models via approximate system Gramians. Int. J. Appl. Math. Comp. Sci., **11**, 1151–1171 (2001).

[LC92] C.-A Lin and T.-Y Chiu. Model reduction via frequency weighted balanced realization. Control Theory and Advanced Technol., **8**, 341–351 (1992).

[LW91] A. Lu and E. Wachspress. Solution of Lyapunov equations by alternating direction implicit iteration. Comput. Math. Appl., **21**:9, 43–58 (1991).

[MOO81] B. C. Moore. Principal Component Analysis in Linear System: Controllability, Observability and Model Reduction. *IEEE Transactions on Automatic Control*, **AC-26**, 17–32 (1981).

[MR76] C. T. Mullis and R. A. Roberts. Synthesis of minimum roundoff noise fixed point digital filters. IEEE Trans. on Circuits and Systems, **CAS-23**, 551–562, (1976).

[OJ88] P.C. Opdenacker and E.A. Jonckheere. A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds. IEEE Trans. Circuits and Systems, (1988).

[PR55] D. W. Peaceman and H. H. Rachford. The numerical solutions of parabolic and elliptic differential equations. J. SIAM, **3**, 28–41 (1955).

[PEN00a] T. Penzl. Eigenvalue Decay Bounds for Solutions of Lyapunov Equations: The Symmetric Case. Systems and Control Letters, **40**: 139–144 (2000).

[PEN00b] T. Penzl. A cyclic low-rank Smith method for large sparse Lyapunov equations. SIAM J. Sci. Comput., **21**;4, 1401–1418 (2000).

[PEN99] T. Penzl. Algorithms for model reduction of large dynamical systems. Technical Report SFB393/99-40, Sonderforschungsbereich 393

Numerische Simulation auf massiv parallelen Rechnern, TU Chemnitz (1999). Avaliable from `http://www.tu-chemnitz.de/sfb393/sfb99pr.html`.

[ROB80] J. D. Roberts. Linear model reduction and solution of the algebraic Riccati equation by use of the sign function. *International Journal of Control,* **32**, 677–687 (1980).

[SAA90] Y. Saad. Numerical solution of large Lyapunov equations. In *Signal Processing, Scattering, Operator Theory and Numerical Methods*, M. Kaashoek, J.V. Schuppen, and A. Ran, eds., Birkhäuser,Boston, MA, 503–511 (1990).

[SKEW96] M. Silveira, M. Kamon, I. Elfadel and J. White. A coordinate-transformed Arnoldi algorithm for generating guaranteed stable reduced-order models of RLC circuits. In *Proc. IEEE/ACM Intl. Conf. CAD, San Jose, CA*, 288–294 (1996).

[SMI68] R. A. Smith. Matrix Equation, $XA + BX = C$. SIAM J. Appl. Math, **16**, 198–201 (1968).

[SAM95] V. Sreeram, B.D.O Anderson and A.G. Madievski. Frequency weighted balanced reduction technique: A generalization and an error bound. In *Proc.* 34^{th} *IEEE Conf. Decision and Control* (1995).

[WSL99] G. Wang, V. Sreeram and W.Q. Liu. A new frequency weighted balanced truncation method and an error bound. IEEE Trans. Automat. Contr., **44**;9, 1734–1737 (1999).

[STA91] G. Starke. Optimal alternating direction implicit parameters for non-symmetric systems of linear equations. SIAM J. Numer. Anal., **28**:5, 1431–1445 (1991).

[STA93] G. Starke. Fejer-Walsh points for rational functions and their use in the ADI iterative method. J. Comput. Appl. Math., **46**, 129–141, (1993).

[VA01] A. Varga and B.D.O Anderson. Accuracy enhancing methods for the frequency-weighted balancing related model reduction. In *Proc.* 40^{th} *IEEE Conf. Decision and Control* (2001).

[WAC62] E. Wachspress. Optimum alternating-direction-implicit iteration parameters for a model problem. J. Soc. Indust. Appl. Math., **10**, 339–350 (1962).

[WAC88a] E. Wachspress. Iterative solution of the Lyapunov matrix equation. Appl. Math. Lett., **1**, 87–90 (1988).

[WAC88b] E. Wachspress. The ADI minimax problem for complete spectra. Appl. Math. Lett., **1**, 311–314 (1988).

[WAC90] E. Wachspress. The ADI minimax problem for complex spectra. In *Iterative Methods for Large Linear Systems,* D. Kincaid and L. Hayes, edss, Academic Press, San Diego, 251–271 (1990).

[WAC95] E. Wachspress. The ADI model problem. *Self published,* Windsor, CA (1995).

[ZHO02] Y, Zhou. Numerical methods for large scale matrix equations with applications in LTI system model reduction. Ph. D. Thesis, CAAM Department, Rice University, Houston, TX, USA, May (2002).

[ZHO95] K. Zhou. Frequency-weighted $\mathcal{L}_\infty$ norm and optimal Hankel norm model reduction. IEEE Trans. Automat. Contr., **40**:10, 1687–1699 (1995).