Model Order Reduction by Dominant Subspace Projection: Error Bound, Subspace Computation and Circuit Applications*

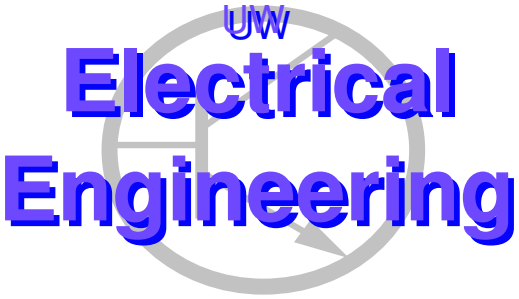
Guoyong Shi and C.-J. Richard Shi

Electrical Engineering Department

University of Washington

Seattle, WA 98195

{gshi,cjshi}@ee.washington.edu



UWEE Technical Report
Number UWEETR-2003-0021
September 2003

Department of Electrical Engineering
University of Washington
Box 352500
Seattle, Washington 98195-2500
PHN: (206) 543-2150
FAX: (206) 543-3842
URL: http://www.ee.washington.edu

Model Order Reduction by Dominant Subspace Projection: Error Bound, Subspace Computation and Circuit Applications*

Guoyong Shi and C.-J. Richard Shi

Electrical Engineering Department
University of Washington
Seattle, WA 98195
{gshi,cjshi}@ee.washington.edu

University of Washington, Dept. of EE, UWEETR-2003-0021

September 2003

Abstract

Balanced truncation is a well-known technique for model order reduction with a known uniform reduction error bound. Its practical use is, however, limited to small-size problems due to its cubic computation complexity. While model order reduction by projection to approximate dominant subspace without balancing has yielded encouraging experimental results, its error bound has not been analyzed. Starting from a frequency-domain solution of the Lyapunov equation, this paper first derives a square-integral reduction error bound for non-balanced dominant subspace projection, which is valid in both the frequency and time domains. Then the computation of approximate dominant subspace via three Krylov subspaces is studied. It is justified analytically that the Krylov subspace by moment matching at low frequency gives rise to a better approximation of dominant subspace than that by moment matching at high frequency. Moreover, upon establishing a new connection between a rational Krylov subspace and waveform matching in the discrete time domain, we point out that it is possible to use a rational Krylov subspace for an even better approximation by choosing an appropriate parameter. The algorithms for approximate dominant subspace computation and their applications to model order reduction are tested by using several circuit examples.

Index Terms – Circuit simulation, dominant subspace, error bound, Krylov subspace, model order reduction, moment matching.

*This work was supported by DARPA NeoCAD Program under Grant No. N66001-01-8920.

1 Introduction

Model order reduction is emerging as an effective technique for the modeling and simulation of large scale integrate circuits (VLSIs) and structures. As the integration level increases and the transistor feature size shrinks, many circuit parasitics can no longer be ignored. Incorporating these parasitics leads to very large scale linear or nonlinear models. Despite of the rapid development of computing resources, direct simulation of such large scale models is often prohibitive.

However, it is a common phenomenon that most large scale models have high degree of redundancy. In practice only an accurate model at certain frequency ranges is of interest. For these reasons, model order reduction can play an important role in reducing the simulation complexity. A variety of model reduction techniques have been developed, mostly in the control literature. Comprehensive reviews can be found in [1, 2] with an emphasis on large scale models. Several popular algorithms are compared in [11].

Two representative reduction techniques widely used in circuit simulation are balanced truncation [7, 19, 25] and moment matching [6, 22, 27]. Balanced truncation yields stable reduced order models with a proven uniform error bound. However, due to its cubic computational complexity, balanced truncation is formidable for large scale model reduction. On the other hand, moment matching has a relatively lower computational complexity and can exploit circuit sparsity. It has been widely used for circuit modeling and analysis [22]. Moment matching for model order reduction has been further popularized by the development of numerically stable computation methods based on Krylov subspaces [6, 20].

Since moment matching only matches moments at some local frequency points, a resulting reduced order model may have fairly large error at some other frequency band. Further a small error bound in the frequency domain might not imply accurate waveform matching in the time domain. For this reason, several recent attempts use modified balanced truncation schemes to improve the global approximation meanwhile to reduce the computation cost [1, 10, 17, 23]. The underlying idea is to use approximately computed dominant controllable and/or observable subspace and then to project the state space of a full order model to the dominant subspace. However, since the exact Gramians are not available, usually balancing is not performed. Some good simulation results have been reported for circuit simulation [17, 23]; however, an error bound by using unbalanced dominant subspace projection remains unknown. The first contribution of this paper is to derive such a bound, which constitutes Section 3.

Dominant subspaces also find applications in nonlinear model order reduction [16]. However, the computation of exact dominant subspace for large scale models, linear or nonlinear, is not a trivial task. For linear time-invariant models, the exact dominant subspace can only be found from the exact Gramian solved from a Lyapunov equation, which has a cubic time complexity. Approximately solving large scale Lyapunov equations has been studied by many researchers in the literature. It has a direct connection to approximate dominant subspace computation [10, 13, 14, 15, 21, 24]. The key idea is to use a low-rank approximation of the Gramian, and apply the efficient computation techniques for Krylov subspace [5]. However, in all of the existing works, the computation of approximate low-rank solution is without exception by forming the Krylov subspace using the matrix pair (A, B) (see eqn. (2)). However, we found in many tests that using the Krylov subspace formed by the pair (A, B) might not be the best choice, given the fact that there are other easily computable Krylov subspaces. In Section 4, after an outline of approximate dominant subspace computation scheme based on Krylov subspace, we

analytically justify that the Krylov subspace formed by the pair $(A^{-1},\ A^{-1}B)$ can in fact have a better approximation. Furthermore, from a new perspective of moment matching in the discrete time domain, we point out that a rational Krylov subspace could possibly yield an even better approximation if an appropriate parameter is chosen. The connection between a rational Krylov subspace and moment matching in the discrete time domain is a new development. We show via this connection that moment matching in the rational Krylov subspace has an implication of waveform matching in the discrete time domain. A comparative study of the approximation effect of different Krylov subspaces in the computation of dominant subspace has hardly been touched in the literature. The preliminary results obtained in Section 4 contribute to the first study along this line.

In Section 5 some practical circuit examples are used to demonstrate that the approximation effect of the three different Krylov subspaces does show the difference in approximating dominant subspace as predicted by the analysis in Section 4. For the evaluation purpose, three measures are used to compare the approximation accuracy. Following the examples of dominant subspace computation, we present several large scale circuit examples for model order reduction by applying the approximate dominant subspace computation technique.

Some terminologies and notations used in this paper are defined below. An asymptotically stable matrix is also called a Hurwitz matrix. Given any matrix $V \in \mathbb{R}^{n\times m}$, the subspace spanned by the columns of V is denoted by $\operatorname{span} V$. The subspace perpendicular to a subspace $\mathcal{S}$ is denoted by $\mathcal{S}^{\perp}$. When we say a matrix V spans a subspace, we mean the columns of matrix V span the subspace, and we call the matrix V the basis matrix. We denote the qth order Krylov subspace generated by two matrices A and B by

$$\mathcal{K}_q(A, B) = \left\{B,\ AB,\ \cdots,\ A^{q-1}B\right\}. \tag{1}$$

We shall also use the convenient matrix manipulation notation used in MATLAB. For example, the notation $M(:, 1{:}q)$ means the matrix formed by taking the first q column from matrix M. The standard basis vectors, i.e. the columns of the identity matrix I, are denoted by e_i. The dimension of e_i should be clear from the context if not specified. A^{T} denotes the transpose of matrix A. $\lambda_i(A)$ denotes the ith eigenvalue of matrix A. If A is symmetry, $\lambda_{\max}(A)$ is the maximal eigenvalue of A. The norm of a vector $\|v\|$ is the conventional Euclidean 2-norm. The norm of a matrix $\|A\|$ is defined by $\|A\| = \lambda^{1/2}(AA^{\mathrm{T}})$. The Frobenius norm of A is defined by $\|A\|_F = [\mathrm{tr}(AA^{\mathrm{T}})]^{1/2}$. The maximum singular value of matrix A is denoted by $\sigma_{\max}(A)$.

2 Preliminary

We consider circuits that can be modeled by linear time-invariant systems

$$\begin{array}{rcl} \dfrac{dx}{dt} & = & Ax + Bu \\ y & = & Cx + Du \end{array} \tag{2}$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^m$ is the input (source) vector, and $y \in R^{\ell}$ is the output (measurement) vector. The transfer function of model (2) is

$$H(s) = C(sI - A)^{-1}B + D. \tag{3}$$

Sometimes it is convenient to use packed notation to represent a linear system and its transfer function

$$H(s) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = C(sI - A)^{-1}B + D. \tag{4}$$

The model order reduction problem is to find a reduced order model

$$\begin{array}{rcl} \dfrac{d\xi}{dt} & = & \hat{A}\xi + \hat{B}u \\ y & = & \hat{C}\xi + Du \end{array} \tag{5}$$

where $\xi \in \mathbb{R}^q$ is the reduced state vector with the reduced order q satisfying $\max\{\ell, m\} \leq q < n$, so that model (5) is a good approximation of the full model (2). The reduced model can also be written as

$$\hat{H}(s) = \left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & D \end{array}\right] = \hat{C}(sI - \hat{A})^{-1}\hat{B} + D. \tag{6}$$

Since D does not play a role in projection-based model reduction, we simply assume $D = 0$ throughout the paper.

A widely accepted model order reduction formulation is by projection. Let W and V be two real matrices in $\mathbb{R}^{n\times q}$ satisfying

$$W^{\mathrm{T}}V = I_q. \tag{7}$$

If we consider the restriction of the state x to span V, we can replace x by $V\xi$ and pre-multiply the first equation in (2) by W^{T}. Then system (2) becomes the reduced order model (5) with

$$\hat{H}(s) = \left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & 0 \end{array}\right] = \left[\begin{array}{c|c} W^{\mathrm{T}}AV & W^{\mathrm{T}}B \\ \hline CV & 0 \end{array}\right]. \tag{8}$$

The quality of a reduced order model obtained by projection can be measured by several criteria. Typical measures are the number of moments matched at some frequency points [6, 20, 22, 27] and the uniform frequency domain error bound used in balanced truncation [7].

Given the LTI system in (2) with the system matrix A Hurwitz, the controllability Gramian is the unique solution of the Lyapunov equation

$$AM + MA^{\mathrm{T}} + BB^{\mathrm{T}} = 0, \tag{9}$$

and the observability Gramian is the unique solution of the dual Lyapunov equation [28]

$$A^{\mathrm{T}}Q + QA + C^{\mathrm{T}}C = 0. \tag{10}$$

If A is Hurwitz, then both Lyapunov equations (9) and (10) have unique solutions and can be expressed respectively by the integrals

$$M = \int_0^\infty e^{At}BB^{\mathrm{T}}e^{A^{\mathrm{T}}t}dt \tag{11}$$

and

$$Q = \int_0^\infty e^{A^{\mathrm{T}}t}C^{\mathrm{T}}Ce^{At}dt. \tag{12}$$

Obviously both Gramians are symmetric and positive semi-definite.

Next we derive an alternate integral expression for the solution of Lyapunov equation (11) in the frequency domain. The frequency domain solution is hardly used in the literature. However, it turns out to be useful for deriving an error bound for unbalanced dominant projection and provides insight for choosing some special Krylov subspace for effective dominant subspace computation.

Lemma 1 *Assume that A is asymptotically stable. The controllability Gramian M can be expressed by an integral in frequency domain, i.e.*

$$M = \frac{1}{2\pi}\int_{-\infty}^{\infty}(j\omega I - A)^{-1}BB^{T}(-j\omega I - A^{T})^{-1}d\omega. \tag{13}$$

Proof : The Lyapunov equation (9) can be rewritten as

$$(j\omega I - A)M + M(-j\omega I - A^{\mathrm{T}}) = BB^{\mathrm{T}},$$

which is equivalent to

$$(j\omega I - A)^{-1}M + M(-j\omega I - A^{\mathrm{T}})^{-1} = (j\omega I - A)^{-1}BB^{\mathrm{T}}(-j\omega I - A^{\mathrm{T}})^{-1}.$$

Taking integral from $\omega = -\infty$ to ∞ yields

$$\int_{-\infty}^{\infty}(j\omega I - A)^{-1}d\omega M + M\int_{-\infty}^{\infty}(-j\omega I - A^{\mathrm{T}})^{-1}d\omega = \int_{-\infty}^{\infty}(j\omega I - A)^{-1}BB^{\mathrm{T}}(-j\omega I - A^{\mathrm{T}})^{-1}d\omega.$$

Since $(sI - A)^{-1}$ vanishes at $s = \infty$, the integration above can be replaced by a contour integral $\oint_{C_R}$ where C_R is the closed path going from $-jR$ to jR on the $j\omega$ axis and following the left semi-circle $Re^{j\theta}$ for $\theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$. For sufficiently large R, this loop encircles all eigenvalues of A but those of $-A$. By Cauchy Theorem we get for R sufficiently large

$$\frac{1}{j2\pi}\oint_{C_R}(sI - A)^{-1}ds = I, \qquad \frac{1}{j2\pi}\oint_{C_R}(sI + A^{\mathrm{T}})^{-1}ds = 0.$$

Then the identity (13) follows immediately. ■

3 Error bound

Suppose that W and V are two projection matrices in $\mathbb{R}^{n\times q}$ satisfying $W^{\mathrm{T}}V = I_q$. The error of model order reduction is defined to be the difference between the full order and a reduced order model in the frequency domain, i.e.,

$$E(s) = H(s) - \hat{H}(s). \tag{14}$$

Several methods have been proposed to estimate the reduction error. A well-known error bound was derived in [7] for balanced truncation. Its practical use is, however, limited to small scale models since balanced truncation is of $\mathcal{O}(n^3)$ computation complexity . Another error bound was derived for moment matching in [3]. However, since moment matching is based on Taylor expansion, an error bound for truncation is not very helpful for evaluating the quality of a reduced order model over a large frequency range.

In this section we derive a new error bound for dominant subspace projection based model order reduction but without balancing. We assume that only the controllability Gramian is available, and its dominant singular subspace is used for reduction. Establishing such an error bound is motivated by practical consideration. In the practice of large scale model reduction, exact Gramians are not available, but the dominant subspaces can be approximately computed by using Krylov subspace method. Without the exact Gramians, the balancing transformation cannot be performed, thus the error bound for balanced truncation cannot be used. However, if a good approximation of the dominant subspace has been computed and used for reduction, knowledge of an approximate error bound of the reduced order model would be very helpful. In the following derivation we assume that an exact dominant controllable subspace is available for reduction projection.

The following lemma gives a characterization of the error defined in (14).

Lemma 2 *For $E(s)$ defined in* (14)*, the following identity holds*

$$E(s) = L(s)\,(I - VW^T)\,F(s) \tag{15}$$

where

$$L(s) = \hat{C}(sI - \hat{A})^{-1}W^T A + C, \qquad F(s) = (sI - A)^{-1}B.$$

By duality, it also holds that

$$E(s) = F_d(s)\,(I - VW^T)\,L_d(s) \tag{16}$$

where

$$F_d(s) = C(sI - A)^{-1}, \qquad L_d(s) = AV(sI - \hat{A})^{-1}\hat{B} + B.$$

Proof : In packed notation we can write

$$E(s) = \left[\begin{array}{cc|c} W^{\mathrm{T}}AV & 0 & W^{\mathrm{T}}B \\ 0 & A & B \\ \hline -CV & C & 0 \end{array}\right].$$

A state transformation $\begin{bmatrix} I & W^{\mathrm{T}} \\ 0 & I \end{bmatrix}$ of this error system leads to

$$E(s) = \left[\begin{array}{cc|c} W^{\mathrm{T}}AV & -W^{\mathrm{T}}A(I - VW^{\mathrm{T}}) & 0 \\ 0 & A & B \\ \hline -CV & C(I - VW^{\mathrm{T}}) & 0 \end{array}\right].$$

This is equivalent to a system in state space

$$\begin{aligned} \dot{\xi} &= \hat{A}\xi - W^{\mathrm{T}}A(I - VW^{\mathrm{T}})x \\ \dot{x} &= Ax + Bu \\ e &= -\hat{C}\xi + C(I - VW^{\mathrm{T}})x \end{aligned}$$

with zero initial conditions. Identity (15) then follows directly from this system.

The dual identity (16) is proven by considering $E^{\mathrm{T}}(s)$ and replacing the triple (A, B, C) by its dual $(A^{\mathrm{T}}, C^{\mathrm{T}}, B^{\mathrm{T}})$. ■

The importance of Lemma 2 lies in the factor $(I - VW^{\mathrm{T}})$. Since $W^{\mathrm{T}}V = I$, this factor is a projector to the subspace $(\operatorname{span} V)^{\perp}$ and its transpose is another projector to the subspace $(\operatorname{span} W)^{\perp}$.

Lemma 2 has two immediate applications. First, it can be used to check the moment matching error if V and W are obtained from moment matching algorithms [6, 27]. For example, if W and V are generated by the q-step Lanczos process as in PVL [6], then the coefficients of s^{-i} in $E(s)$ up to order $2q$ vanish, i.e. $2q$ moments of $H(s)$ and $\hat{H}(s)$ are matched. Second, Lemma 2 can be used to derive an error bound for $E(s)$ if either W or V spans respectively the dominant observable or controllable subspace.

Define the $\mathcal{L}_2$ norm of error as

$$\|E\|_2 := \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{tr}\left[E(j\omega)E^{\mathrm{T}}(-j\omega)\right] d\omega \right\}^{1/2} \tag{17}$$

and the $\mathcal{L}_\infty$ norm of $L(s)$ as [28]

$$\|L\|_\infty := \sup_{\omega \in (-\infty, \infty)} \sigma_{\max}[L(j\omega)]. \tag{18}$$

The next theorem establishes a bound on $\|E\|_2$.

Theorem 1 *Let M be the controllability Gramian and $M = U\Sigma U^T$ be the SVD of M, where U is an orthogonal matrix and $\Sigma = \operatorname{diag}[\sigma_1, \cdots, \sigma_n]$ is a diagonal matrix containing the singular values of M in descending order. Let $1 \le q < n$. If $V = U(:, 1{:}q)$ and $W \in \mathbb{R}^{n\times q}$ satisfying $W^TV = I_q$ are used for reduction projection and assume that the reduced matrix $\hat{A} = W^TAV$ is asymptotically stable. Then we have*

$$\|E\|_2 \le \left(\sum_{i=q+1}^{n} \sigma_i \right)^{1/2} \|L\|_\infty \|I - WV^T\| \tag{19}$$

where $L = L(s)$ is defined in Lemma 2. If $W = V$, then we have

$$\|E\|_2 \le \left(\sum_{i=q+1}^{n} \sigma_i \right)^{1/2} \|L\|_\infty. \tag{20}$$

Proof : Since $\hat{A}$ is asymptotically stable, $\|L\|_\infty$ is finite. Let $V_2 = U(:, q+1{:}n)$ and $\Sigma_2 = \operatorname{diag}[\sigma_{q+1}, \cdots, \sigma_n]$. By the definition of $\|E\|_2$ in (17) and using the identity (15), we get

$$\begin{aligned}
\|E\|_2^2 &\le \|L\|_\infty^2 \operatorname{tr}\left\{ (I - VW^{\mathrm{T}}) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega)F^{\mathrm{T}}(-j\omega)d\omega \right] (I - WV^{\mathrm{T}}) \right\} \\
&= \operatorname{tr}\{(I - VW^{\mathrm{T}})M(I - WV^{\mathrm{T}})\}\|L\|_\infty^2 \\
&= \operatorname{tr}\{(I - VW^{\mathrm{T}})V_2\Sigma_2V_2^{\mathrm{T}}(I - WV^{\mathrm{T}})\}\|L\|_\infty^2 \\
&\le \|L\|_\infty^2 \left(\sum_{i=q+1}^{n} \sigma_i \right) \|I - WV^{\mathrm{T}}\|^2
\end{aligned}$$

Hence, (19) follows. Bound (20) follows from the fact that $\|I - VV^{\mathrm{T}}\| = 1$ for $q < n$. ∎

Remark 1 *By Parseval's theorem, the $\mathcal{L}_2$ norm of error in the frequency domain is equal to that in the time domain. Hence the reduction error bound in Theorem 1 also holds in the time domain. This guarantees a good waveform approximation if the frequency domain approximation is good. Such a fact does not hold for $\mathcal{L}_\infty$ error bound.*

An immediate consequence of Theorem 1 is that if the trailing singular values are zero, then the projection based reduction would not lose any accuracy. Hence the following corollary is straightforward.

Corollary 1 *Under the same condition as in Theorem 1, if $\sigma_i = 0$ for $i = q+1, \cdots, n$, then $\hat{H}(s) = H(s)$, i.e. the reduced order model is equivalent to the full order model.*

Remark 2 *In circuit simulation, passivity is an important property. Since congruence transformation preserves passivity for circuit models with port formulation, one prefers to choose $W = V$ if the columns of V are orthonormal [20]. In Theorem 1 we assumed that $\hat{A} = W^T AV$ is asymptotically stable. However, the stability of $\hat{A} = W^T AV$ is in general not guaranteed, unless W and V are obtained from balanced truncation. In practice sometimes $\hat{A}$ might contain a few unstable poles. In that case some kind of postprocessing is needed, such as dropping the unstable poles by extracting the stable subspace. Since normally $\hat{A}$ is a low-dimensional matrix, the stability check and stable subspace extraction are computationally feasible.*

Remark 3 *Note that Theorem 1 is stated for the controllability Gramian only. A dual result holds for the observability Gramian as well, which follows directly from Lemma 2.*

4 Computation of approximate dominant subspace

In the previous section we established an $\mathcal{L}_2$ error bound in model reduction by unbalanced dominant subspace projection. However, since the exact Gramian is not easily computable for a large model, the dominant subspace must be computed by approximation in practice. Obviously, the better the dominant subspace is approximated, the more accurate the error bound is.

Approximate computation of dominant subspace for large scale models has been studied by many researchers in the literature. Most of the works stem from the idea of low-rank approximate solution of a large scale Lyapunov equation. The low-rank solution of Lyapunov equation was first addressed by Hodel and Poolla [13], where several heuristic algorithms were proposed. Saad [24] specifically analyzed the low-rank approximation by using the Krylov subspace $\mathcal{K}_m(A, B)$, where the Galerkin condition on the residual was considered. Jaimoukha and Kasenally [15] extended Saad's idea on SISO case to MIMO and proposed a GMRES-like solution scheme by using an explicit expression for the residual. In all of these works, the Krylov subspace was formed by the pair (A, B). However, the Galerkin condition does not necessarily imply that this pair is optimal for effective dominant subspace computation. More algorithms along the same line are proposed in [14] and [21].

The key point we would like to make in this section is that the traditionally used Krylov subspace $\mathcal{K}_m(A, B)$ in the literature is in fact not the best choice for effective computation of dominant subspace, especially for large scale models. Such an observation is motivated by the connection between Krylov subspace and moment matching. Since moment matching is in the frequency domain, the integral solution of Gramian in the frequency domain as established

in Lemma 1 turns out be very helpful for gaining insight on choosing an appropriate Krylov subspace for numerical computation.

We start from describing the outline of a general iterative dominant subspace computation scheme. Variants of this scheme have been used in many works. We describe the scheme in a generic way without specifying the matrix and vector used for the Krylov subspace computation. Later on this generic algorithm will be specialized to several different Krylov subspaces to compare the effect of dominant subspace approximation. To better articulate the basic idea, we assume that the model has a single input, i.e. B is a column vector. The multiple input case can be treated similarly.

Dominant Subspace Computation Scheme:

Input Data: A matrix $\Phi \in \mathbb{R}^{n\times n}$, a vector $\eta \in \mathbb{R}^n$, the dimension q of an approximate dominant subspace, and an intermediate integer m with $q \le m < n$.

Step 1. Run the Arnoldi algorithm [5] to generate the basis vectors for the mth order Krylov subspace $\mathcal{K}_m(\Phi, \eta)$. Let $V_m \in \mathbb{R}^{n\times n}$ be the basis matrix, i.e. the matrix formed by the basis vectors.

Step 2. Form the reduced order Lyapunov equation

$$A_m M_m + M_m A_m^{\mathrm{T}} + B_m B_m^{\mathrm{T}} = 0 \tag{21}$$

where

$$A_m = V_m^{\mathrm{T}} A V_m, \qquad B_m = V_m^{\mathrm{T}} B.$$

Step 3. Solve M_m from the Lyapunov equation (21), and find the SVD of M_m, i.e.

$$M_m = U_m \Sigma_m U_m^{\mathrm{T}}$$

where $U_m \in \mathbb{R}^{m\times m}$ is an orthogonal matrix and the singular values of M_m are in descending order in Σ_m.

Step 4. Define $\widetilde{V}_m = V_m U_m$ and extract the leading q column vectors of $\widetilde{V}_m$ to form an $n \times q$ matrix $\widetilde{V}_q$.

Output Data: Matrix $\widetilde{V}_q$ whose columns span an approximate dominant subspace.

Since the algorithm requires solving a Lyapunov equation of dimension m, we recommend that the intermediate dimension m should not be too high (normally no greater than 100) so that solving the Lyapunov equation (21) is kept at low cost.

One could directly return V_m after Step 1 for an approximate subspace. However, V_m generated by the Arnoldi algorithm might not capture the dominance very well. Steps 2 to 4 are solely for a better low-dimensional approximation of dominance subspace with some modest additional computation cost. The reduction from m basis vectors to q dominant basis vectors is called *compaction*.

It is easily verified that

$$\widetilde{A}_q \Sigma_q + \Sigma_q \widetilde{A}_q^{\mathrm{T}} + \widetilde{B}_q \widetilde{B}_q^{\mathrm{T}} = 0 \tag{22}$$

where

$$\widetilde{A}_q = \widetilde{V}_q^{\mathrm{T}} A \widetilde{V}_q, \qquad \widetilde{B}_q = \widetilde{V}_q^{\mathrm{T}} B,$$

and Σ_q contains the leading q singular values of Σ_m. The computation scheme outlined above also gives rise to a low-rank approximation of the Gramian

$$\widetilde{M} = V_m M_m V_m^{\mathrm{T}} = V_m U_m \Sigma_m U_m^{\mathrm{T}} V_m^{\mathrm{T}}. \tag{23}$$

This approximate solution has the property that the Galerkin condition is satisfied [24], i.e.

$$V_m^{\mathrm{T}} R(\widetilde{M}) V_m = 0, \tag{24}$$

where $R(Z)$ is the residual matrix of the Lyapunov equation defined by

$$R(Z) = AZ + ZA^{\mathrm{T}} + BB^{\mathrm{T}}. \tag{25}$$

Note that in the computation scheme described above, we did not specify the choice of Φ and η for the Krylov subspace $\mathcal{K}_m(\Phi, \eta)$. Traditionally Φ and η are chosen to be A and B respectively, mainly because most of the previous works consider the Lyapunov equation in the form of (9) and its solution in the form of (11). However, if A is asymptotically stable (hence invertible), the Lyapunov equation (9) can equivalently be written as

$$A^{-1}M + MA^{-T} + A^{-1}BB^{\mathrm{T}}A^{-T} = 0, \tag{26}$$

which is still a Lyapunov equation. Hence, clearly there is another way of computing the dominant controllable subspace by using the pair $(\Phi, \eta) = \left(A^{-1}, A^{-1}B\right)$. In circuit models, obtaining A or A^{-1} involves almost equal amount of computation, depending on inverting the susceptance matrix or the conductance matrix. An interesting question is, between the two Krylov subspace $\mathcal{K}_m(A, B)$ and $\mathcal{K}_m\left(A^{-1}, A^{-1}B\right)$, which one would in general give rise to a better approximation of the dominant subspace, given a low order m?

We have observed from our numerical test that the Krylov subspace $\mathcal{K}_m\left(A^{-1}, A^{-1}B\right)$ always yields a better approximation to the dominant subspace. This phenomenon can be justified by an analysis in the frequency domain, which is based on moment matching and the integral expression of M in the frequency domain (see (13)).

Let $X(s) = H_x(s)U(s)$ be the Laplace transform of the state x of model (2), where $H_x(s) = (sI - A)^{-1}B$. The Taylor expansion of $H_x(s)$ can take two forms, either in terms of s or $1/s$, i.e.

$$\begin{aligned} H_x(s) &= (sI - A)^{-1}B \\ &= \frac{B}{s} + \frac{AB}{s^2} + \frac{A^2B}{s^3} + \cdots \qquad (27) \\ &= -A^{-1}B - A^{-2}Bs - A^{-3}Bs^2 - \cdots \qquad (28) \end{aligned}$$

The two expansions in (27) and (28) lead to two ways of matching the moments, depending on the leading moments in (27) or (28). Clearly, the related Krylov subspaces are respectively $\mathcal{K}_m(A, B)$ and $\mathcal{K}_m\left(A^{-1}, A^{-1}B\right)$. The projections formed by the basis vectors of the two Krylov subspaces can be used to reduce the matrices A and B while keeping the leading moments matched. More specifically, let V be one of the basis matrices and let A and B be reduced to

$$\hat{A} = V^{\mathrm{T}}AV \qquad \text{and} \qquad \hat{B} = V^{\mathrm{T}}B. \tag{29}$$

One can show that the leading m moments of the two transfer functions, $H_x(s) = (sI - A)^{-1}B$ and $\widetilde{H}_x(s) = V(sI - \hat{A})^{-1}\hat{B}$, match each other, provided that V is an appropriate basis matrix. This result is stated in the next Theorem. To simplify notation, this theorem is stated for the single input case. It can be extended easily to the multiple input case.

Theorem 2 *Assume that the full order model* (2) *is single input.*

(*a*) *If* V *is the basis matrix of the Krylov subspace* $\mathcal{K}_m(A,\, B)$*, then we have*

$$A^i B = V\hat{A}^i\hat{B} = \beta V\hat{A}^i e_1 \tag{30}$$

for $i = 0, \cdots, m-1$*, where* $\beta = \|B\|$*.*

(*b*) *If* V *is the basis matrix of the Krylov subspace* $\mathcal{K}_m\left(A^{-1}, A^{-1}B\right)$ *and* $\hat{A}$ *is invertible, then we have*

$$A^{-i}B = V\hat{A}^{-i}\hat{B} \tag{31}$$

for $i = 1, \cdots, m$*.*

Proof : Item (*a*) follows directly from the Arnoldi algorithm, see [8] for details. The proof of Item (*b*) is not that straightforward, see Appendix A for a complete proof. ■

To be specific, we shall use the term *matching moments at the high frequency* to indicate the moment matching in terms of $1/s^i$, and the term *matching moments at the low frequency* to indicate the moment matching in terms of s^i. Theorem 2 can be used to justify that the Krylov subspace $\mathcal{K}_m(A^{-1}, A^{-1}B)$ better approximates the dominant controllable subspace than $\mathcal{K}_m(A,\, B)$.

Let $V^{(1)}$ and $V^{(2)}$ be the basis matrices of the Krylov subspaces $\mathcal{K}_m(A,\, B)$ and $\mathcal{K}_m\left(A^{-1}, A^{-1}B\right)$, respectively. Let

$$\hat{A}_i = {V^{(i)}}^{\mathrm{T}} A V^{(i)}, \qquad \hat{B}_i = {V^{(i)}}^{\mathrm{T}} B,$$

and $H_x^{(i)}(s) = V^{(i)}\hat{H}_x^{(i)}(s)$, where $\hat{H}_x^{(i)}(s) = (sI - \hat{A}_i)^{-1}\hat{B}_i$ for $i = 1, 2$. Then Theorem 2 implies that the leading m moments of $H_x^{(1)}(s)$ match those of $H_x(s) = (sI - A)^{-1}B$ in terms of $1/s^i$, and the leading m moments of $H_x^{(2)}(s)$ match those of $H_x(s)$ in terms of s^i.

In view of Lemma 1, the controllability Gramian M can be written as

$$M = \frac{1}{2\pi}\int_{-\infty}^{\infty} H_x(j\omega)H_x^*(j\omega)d\omega, \tag{32}$$

where the superscript * indicates conjugate transpose. Let $\hat{M}^{(i)}$ $(i = 1, 2)$ be the solution of the reduced order Lyapunov equation

$$\hat{A}_i\hat{M}^{(i)} + \hat{M}^{(i)}\hat{A}_i^{\mathrm{T}} + \hat{B}_i\hat{B}_i^{\mathrm{T}} = 0.$$

Then the two Krylov subspaces lead to two approximations for the Gramian M, i.e. for $i = 1, 2$,

$$\begin{aligned}\widetilde{M}^{(i)} &= V^{(i)}\hat{M}^{(i)}{V^{(i)}}^{\mathrm{T}} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} V^{(i)}\hat{H}_x^{(i)}(j\omega)\hat{H}_x^{(i)*}(j\omega){V^{(i)}}^{\mathrm{T}}d\omega\end{aligned}$$

Since $(j\omega I - A)^{-1}$ rolls off as $\omega \to \infty$, the contribution of $H_x(j\omega) = (j\omega I - A)^{-1}B$ to the integral of M in (32) mainly comes from the low frequency part, rather than the high frequency part. Thus, in approximation sense, another integral with its integrand matching the moments of $H_x(s)$ at the low frequency should give rise to a better approximation of the Gramian M. Following the preceding analysis, we conclude that the matrix $\widetilde{M}^{(2)}$ in general should be a better

approximation of the Gramian M than $\widetilde{M}^{(1)}$. Consequently, from a subspace point of view, the Krylov subspace $\mathcal{K}_m\left(A^{-1}, A^{-1}B\right)$ better captures the dominant controllable subspace. This conclusion fortunately is consistent with the practice; in practice we are mostly interested in the frequency band from DC up to some gigahertz level, and circuit model order reductions are mostly performed at low frequencies by matching moments of s^i.

Interestingly, in addition to the two options we have discussed so far, there are other options for choosing the pair (Φ, η). Among others, an easily computable Krylov subspace is

$$\mathcal{K}_m\left((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B\right) \tag{33}$$

with $\gamma > 0$. Such a Krylov subspace is called *rational Krylov subspace* in [9]. Note that by choosing $\gamma = 0$ this Krylov subspace reduces to $\mathcal{K}_m\left(A^{-1}, A^{-1}B\right)$.

The rational Krylov subspace has connections to iterative numerical algorithms for solving Lyapunov equations, such as the Smith's algorithm [26] and the ADI algorithm [18]. Grimme [9] used it as a candidate for model order reduction. However, unlike the previous two Krylov subspaces that have direct relations to moment matching, the exact meaning of the rational Krylov subspace is not quite clear in the literature. In the remaining part of this section, we show that this third Krylov subspace is in fact related to moment matching in the discrete-time domain, hence model reduction in this subspace bears the meaning of waveform matching in the discrete-time domain. To the authors' knowledge, this link has not been established in the literature.

One way to discretize the continuous-time model

$$\dot{x} = Ax + Bu \tag{34}$$

is to replace the derivative operator (or s operator) approximately by

$$s \approx \frac{z-1}{hz} = \frac{\gamma(z-1)}{z} \tag{35}$$

where $z = e^{sh}$, h is a small time step, and $\gamma = 1/h$ is the sampling frequency. Then the continuous-time model (34) becomes

$$(\gamma I - A)x_{k+1} = \gamma x_k + Bu_{k+1}, \tag{36}$$

which is the backward Euler integration formula.

Taking z-transform, we get $X(z) = H_x(z)U(z)$, where

$$H_x(z) = \left[(\gamma I - A)z - \gamma I\right]^{-1} BzU(z). \tag{37}$$

If $(\gamma I - A)$ is invertible, $H_x(z)$ can be expanded in terms of z^{-1}

$$H_x(z) = B_\gamma + \gamma A_\gamma B_\gamma z^{-1} + \gamma^2 A_\gamma^2 B_\gamma z^{-2} + \cdots$$

where $A_\gamma = (\gamma I - A)^{-1}$ and $B_\gamma = A_\gamma B$. It is clear that the leading m moments of $H_x(z)$ constitute the Krylov subspace $\mathcal{K}_m\left((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B\right)$.

Note that if the transfer functions of two discrete-time models, $X(z)$ and $\widetilde{X}(z)$, have moments (i.e. coefficients of z^{-i}) matched for $i = 0, \cdots, q-1$, then by the definition of z-transform we have $x_k = \widetilde{x}_k$, for $k = 0, \cdots, q-1$, in the discrete-time domain. This implies that moment

matching in the discrete-time domain can be another approach to model order reduction. While rational Krylov subspace has already been used in practice, the new interpretation in the sense of moment matching in the discrete-time domain would provide a new guideline for choosing γ. We emphasize that in the current interpretation, the parameter γ is the inverse of the time step used for integration.

The next proposition characterizes a property of moment matching in the discrete-time domain. The proof is rather straightforward and can be found in ([9], Chapter 3).

Proposition 1 *Let V be the orthonormal basis matrix of the Krylov subspace in* (33) *(with $\gamma > 0$) obtained from the Arnoldi algorithm. Then the following identities hold:*

$$(\gamma I - A)^{-i} B = V(\gamma I - \hat{A})^{-i} \hat{B} \tag{38}$$

for $i = 1, 2, \cdots, q$, where $\hat{A} = V^T A V$ and $\hat{B} = V^T B$.

If we use V for projection, then the discrete-time reduced order model has the transfer function described in the z-domain:

$$\widetilde{Y}(z) = \hat{C} \left[(\gamma I - \hat{A}) z - \gamma I \right]^{-1} \hat{B} z U(z)$$

where $\hat{C} = CV$. It is easy to see that the leading q moments of $\widetilde{Y}(z)$ match those of $Y(z) = C \left[(\gamma I - A) z - \gamma I \right]^{-1} B z U(z)$, which in the time domain is equivalent to

$$\widetilde{y}_k = y_k, \quad \text{for } k = 0, 1, \cdots, q,$$

given any initial conditions satisfying $x_0 = V \widetilde{x}_0$. That is, given the same input and compatible initial conditions, the discrete-time responses of the full and reduced order models match at least in the first q steps.

Then a natural question is whether a rational Krylov subspace can be used to compute an approximate dominant subspace better than the previous two candidates. At the moment of writing this paper, we do not have a clear justification that this is in general true. However, in our experiments to be presented in the next section, we observed that an appropriate choice of γ did lead to a better approximation of the dominant subspace. The optimal choice of γ is a complicated issue beyond the scope of this paper. An empirical guideline is to choose γ several magnitudes smaller than the magnitude of the dominant mode of the original model. A smaller γ usually has the similar performance as choosing $\gamma = 0$, i.e. moment matching at $s = 0$. An overly large γ leads to bad approximation, because the integration step is too small.

So far we have discussed the computation issues of the dominant controllable subspace. The computation of the dominant *observable* subspace is similar by replacing (A, B) by its dual (A^{T}, C^{T}).

Remark 4 *Another discretization is to approximate the s operator by*

$$s \approx \frac{2(z-1)}{h(z+1)} = \frac{2\gamma(z-1)}{(z+1)} \tag{39}$$

which is equivalent to the trapezoidal rule in numerical integration. It can be verified that the Krylov subspace resulting from this discretization is the same as (33).

5 Numerical experiments

The numerical experiment consists of two parts. In the first part we compare the effectiveness of the three types of Krylov subspaces discussed in the previous section for approximating the dominant subspace. Then in the second part we use approximate dominant subspaces for model order reduction.

Some measures are needed for comparison. To see whether the dominant subspace is correctly approximated, we use three measures. The first measure is to compare the approximate singular values with the exact ones. This requires the computation of the exact Gramian and its singular values, thus is for the demonstration purpose. In practice we can check the convergence of the computed singular values to decide whether or not the approximate dominant subspace is accurate enough. Let σ_i and $\widetilde{\sigma}_i$ be the exact and approximate singular values for $i = 1, \cdots, m$, respectively. Then the total relative error is defined by

$$\sigma_{err} = \sum_{i=1}^{m} \frac{|\sigma_i - \widetilde{\sigma}_i|}{\sigma_i} \tag{40}$$

where we assume $\sigma_i > 0$ for $i = 1, \cdots, m$.

The second measure is to compute the distance between the approximate subspace basis matrix $\widetilde{U}$ and the exact dominant subspace basis matrix U obtained from SVD of M. The distance is defined by

$$\text{dist}(\widetilde{U},\, U) := \left\| \widetilde{U} - U(U^{\mathrm{T}}\widetilde{U}) \right\|_F \tag{41}$$

which measures the mismatch between $\widetilde{U}$ and U. The choice of Frobenius norm is to better differentiate the distance for different basis matrices. Note that this measure also requires the exact Gramian, thus is also for the demonstration purpose. In practice, one can check the convergence of $\widetilde{V}$ by checking the distance between two consecutive basis matrices.

The third measure is to check the residual of the Lyapunov equation, i.e. $R(Z)$ defined in (25), by substituting the approximate Gramian formed as in (23). To reduce the computation complexity, we adopt an idea presented in [21] for residual computation. Let $U \in \mathbb{R}^{n\times q}$ span the dominant subspace and $\widetilde{M} = UM_qU^{\mathrm{T}}$ be the approximate solution, where M_q is the solution of the reduced Lyapunov equation as in (21). Then the residual can be expressed as

$$\begin{aligned} R(\widetilde{M}) &= AUM_qU^{\mathrm{T}} + UM_qU^{\mathrm{T}}A^{\mathrm{T}} + BB^{\mathrm{T}} \\ &= \begin{bmatrix} AU & U & B \end{bmatrix} \begin{bmatrix} 0 & M_q & 0 \\ M_q & 0 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} U^{\mathrm{T}}A^{\mathrm{T}} \\ U^{\mathrm{T}} \\ B^{\mathrm{T}} \end{bmatrix}. \end{aligned}$$

Let

$$QR = \begin{bmatrix} AU & U & B \end{bmatrix}$$

be the QR factorization of the matrix on the right hand side, where R is square and upper-triangular. Then

$$\left\| R(\widetilde{M}) \right\| = \|R\Delta R^{\mathrm{T}}\|$$

where

$$\Delta = \begin{bmatrix} 0 & M_q & 0 \\ M_q & 0 & 0 \\ 0 & 0 & I \end{bmatrix}.$$

Since R is a low-dimensional matrix, this method reduces the computation of residual evaluation significantly. The relative residual is defined to be

$$\varepsilon_{res} = \frac{\left\| R(\widetilde{M}) \right\|}{\|B\|^2}. \tag{42}$$

The first test case is an RLC line with N segments shown in Fig. 1, which represents a discretized model of a transmission line or an interconnect. We shall use this example to test first the dominant subspace computation scheme, then model reduction by using the computed dominant subspace. By modified nodal analysis (MNA) formulation [12], the state vector is $x = (V_1, \cdots, V_N, I_1, \cdots, I_N)^{\mathrm{T}} \in \mathbb{R}^{2N}$. We choose $u = V_s$ to be the input and $y = V_N$ the output.

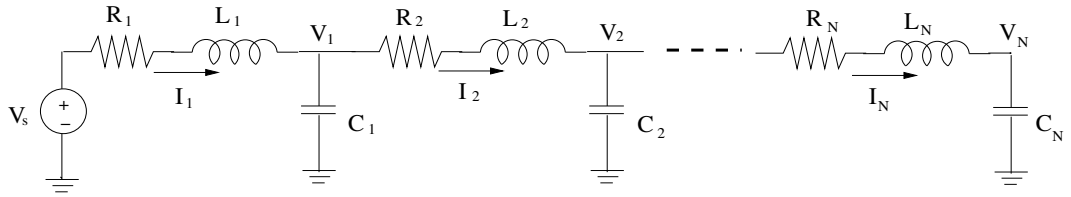


Fig. 1. An RLC line.

The first part of test is to demonstrate that the Krylov subspace $\mathcal{K}_m\left(A^{-1}, A^{-1}B\right)$ performs better than $\mathcal{K}_m(A,\, B)$ in capturing the dominant subspace. Moreover, by choosing an appropriate γ, the the Krylov subspace $\mathcal{K}_m\left((\gamma I - A)^{-1},\, (\gamma I - A)^{-1}B\right)$ performs even better. For easy identification, we represent the first Krylov subspace by $\gamma = 0$, the second one by $\gamma = \infty$, and the third one by the specific γ chosen.

For demonstration purpose, we choose a 200-stage RLC line so that the model order is 400, and the exact Gramian can be solved by Bartels-Stewart algorithm [4]. For simplicity, we assume the RLC values are uniform with $R = 10\,\Omega$, $L_i = 1\,H$, and $C_i = 1\,F$ for all i. Although the state space model is controllable in theory, the computed Gramian only has a rank 31 by MATLAB. That means the controllable space has pronounced dominance and the full order model has quite an amount of redundancy.

TABLE I
MEASURES BY USING THREE KRYLOV SUBSPACES

	$\gamma = \infty$	$\gamma = 0$	$\gamma = 0.002$
σ_{err}	15.69	10.54	8.26
$d(U, \widetilde{U})$	3.29	2.34	2.02
ε_{res}	0.002634	0.009446	0.005470

We compute the approximate dominant controllable subspace by using the three candidate Krylov subspaces, all with order 20. The three measures introduced at the beginning of this section are used for comparison. They are listed in Table I. Fig. 2 shows the approximation of the singular values by using the three optional Krylov subspaces. For better visualization, the 10-based logarithms of the singular values are plotted. We see that the Krylov subspace at $\gamma = 0$

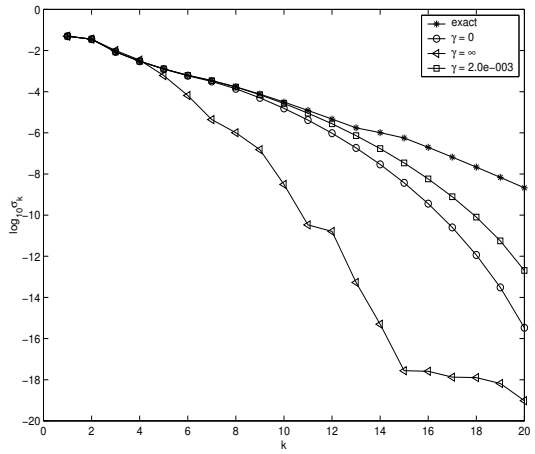


Fig. 2. Approximation of singular values. $\triangleright : \gamma = \infty$; $\circ : \gamma = 0$; $\square : \gamma = 2 \times 10^{-3}$; $\star$: exact.

is better than that at $\gamma = \infty$ by the measures of singular values or subspace distance. For this example, the best approximation is achieved by a rational Krylov subspace at $\gamma = 0.002$. The numerical result matches quite well with our theoretical analysis before. In terms of residual, the Krylov subspace at $\gamma = \infty$ performs the best. This indicates that the residual measure might not be reliable if used as the only measure for dominant subspace computation purpose. Note that one can also use the equivalent Lyapunov equation (26) to compute the residual, which would give a better residual result for the case $\gamma = 0$. But to have a comparison on the same basis, we use the Lyapunov equation (9) only for residual computation.

TABLE II
MEASURES BY USING THREE KRYLOV SUBSPACES

	$\gamma = \infty$	$\gamma = 0$	$\gamma = 3 \times 10^8$
σ_{err}	15.72	11.20	11.00
$d(U, \widetilde{U})$	3.24	2.40	1.99
ε_{res}	0.066	0.321	0.020

Next we choose another set of RLC values for the same circuit, with $R = 20\,\Omega$, $L = 1\,nH$ and $C = 20\,pF$. The test results are summarized in Table II and Fig. 3. For this case, the exact Gramian has a computed rank of 33. We see that the Krylov subspace at $\gamma = 0$ still better captures the dominant singular values than that at $\gamma = \infty$. The Krylov subspace at $\gamma = 3 \times 10^8$ performs better than that at $\gamma = 0$.

The preceding two test cases show that all three Krylov subspaces capture the leading several dominant singular values with very high accuracy, but not very well the trailing singular values. To improve the accuracy, we following the *dominant subspace computation scheme* as presented in Section 4 and choose an intermediate order m relatively large, then use the compaction procedure to get a lower dimensional subspace for a better approximation of dominance. The same circuit example in Fig. 1 is used to test this scheme. We compare different sizes of m and the same q to see the convergence effect.

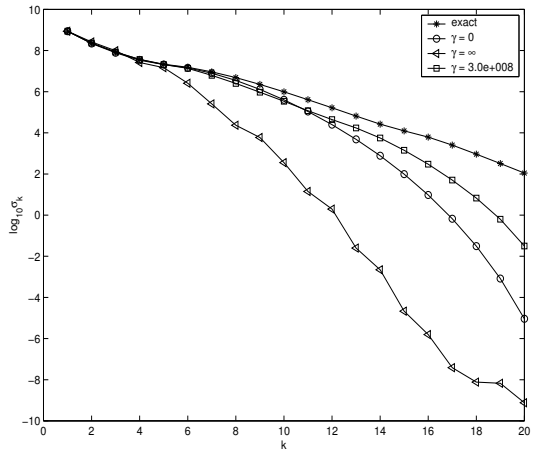


Fig. 3. Approximation of singular values. $\triangleright : \gamma = \infty$; $\circ : \gamma = 0$; $\square : \gamma = 3 \times 10^8$; $\star$: exact.

In this test case we choose a larger model order with $n = 2000$. For such a large scale model, solving the exact Gramian is not practical. Thus we use the first two measures by checking the convergence of the computed singular values and the distance between two consecutive subspaces, and the third measure by computing the residual. The RLC values remain uniform with $R = 30\,\Omega$, $L = 0.1\,nH$, and $C = 5\,pF$. To test the convergence, we choose a sequence of m from 20 up to 100 with increment 20 and all compacted to $q = 20$. Listed in Table III are the the distance measure and the residual measure for different m with $\gamma = 10^8$. The notation $d(U, U_{pre})$ denotes the distance between the computed subspaces at the current and previous steps. The initial subspace is assumed to be the zero subspace. The data in Table III show that the dominant subspace converges quite well. Fig. 4 shows the convergence of the computed singular values.

TABLE III
MEASURES AS A FUNCTION OF m WITH $\gamma = 10^8$.

m	20	40	60	80	100
$d(U, U_{pre})$	4.47	2.12	1.49	1.15	0.10
ε_{res}	0.0354	0.0102	0.0054	4.23×10^{-4}	6.0×10^{-5}

Table IV and Fig. 5 show the test results with $\gamma = 0$. We see that $d(U, U_{pre}) = 1.00$ at $m = 100$, which means the dominant subspace has not converged well yet.

TABLE IV
MEASURES AS A FUNCTION OF m WITH $\gamma = 0$.

m	20	40	60	80	100
$d(U, U_{pre})$	4.47	2.38	1.68	1.30	1.00
ε_{res}	0.2343	0.0764	0.0152	8.28×10^{-3}	5.15×10^{-4}

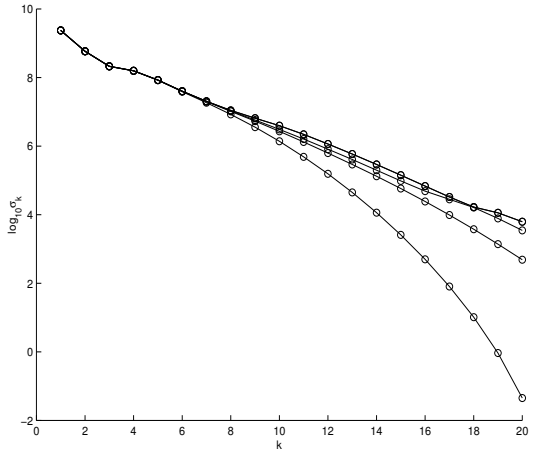


Fig. 4. Convergence of singular values with $\gamma = 10^8$ for $m = 20, 40, \cdots, 100$.

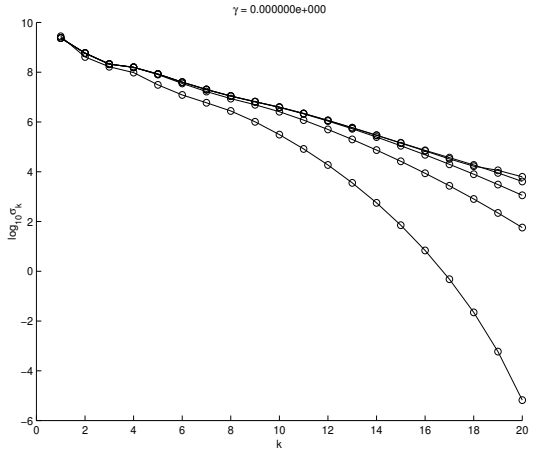


Fig. 5. Convergence of singular values with $\gamma = 0$ for $m = 20, 40, \cdots, 100$.

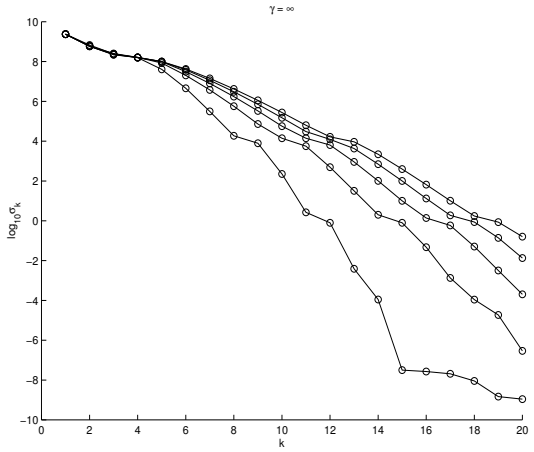


Fig. 6. Convergence of singular values with $\gamma = \infty$ for $m = 20, 40, \cdots, 100$.

To see the performance of the Krylov subspace at $\gamma = \infty$, we test the same example again. The results are shown in Table V and Fig. 6. Clearly the convergence of this option is not as good as the previous two options.

TABLE V
MEASURES AS A FUNCTION OF m WITH $\gamma = \infty$.

m	20	40	60	80	100
$d(U, U_{pre})$	4.47	2.48	1.78	1.46	1.27
ε_{res}	0.0183	0.0064	0.0035	0.0023	0.0016

As a brief summary following the test above, the dominant subspace approximation by the Krylov subspace $\mathcal{K}_m(A, B)$ is not as good as the other two options. This is in consistence with our previous analysis in the frequency domain. For this reason, in the second part of experiment for model order reduction the Krylov subspace $\mathcal{K}_m(A, B)$ will not be used.

The RLC circuit in Fig. 1 is also used for testing model order reduction. Now V_1 is chosen as the output. The uniform RLC values are $R = 20\,\Omega$, $L = 1\,nH$ and $C = 20\,pF$. The full model order is 1000 and the reduced model order is 20. For dominant subspace compaction, we first generate a Krylov subspace with 100th order which is further reduced to 20th by dominance extraction. Shown in Fig. 7 is the reduction result by using the Krylov subspace at $\gamma = 0$. Here only the projection to the dominant controllable subspace V is used. The reduction result becomes better if oblique projection is made, i.e. V is computed from the pair (A, B) and W from the pair $(A^{\mathrm{T}}, C^{\mathrm{T}})$, as shown in Fig. 8. However, since the transformation is not necessarily congruence, the passivity is not guaranteed, which could possibly result in an unstable reduced order model. For comparison, the reduction result by using moment matching up to the 20th order without compaction is also shown in Fig. 9. We see that by matching moments only at the low frequency, the error is very small at the low frequency band, but increases remarkably at the high frequency, while the error by dominant subspace method is rather flat.

The second circuit example is two coupled RLC lines shown in Fig. 10. For this circuit

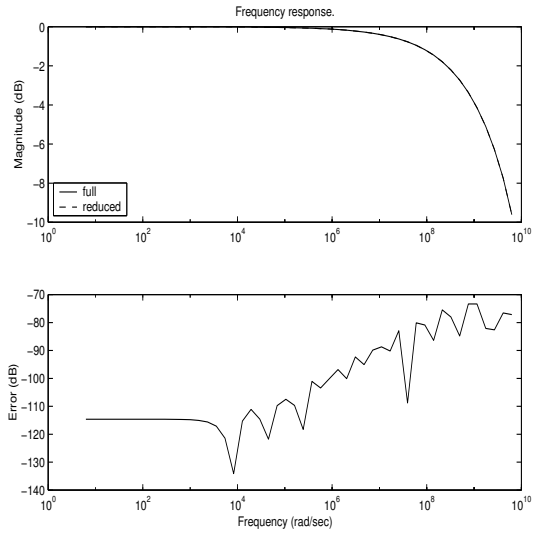


Fig. 7. Reduction of RLC line from 1000th order to 20th order, using Krylov subspace at $\gamma = 0$ (orthogonal projection).

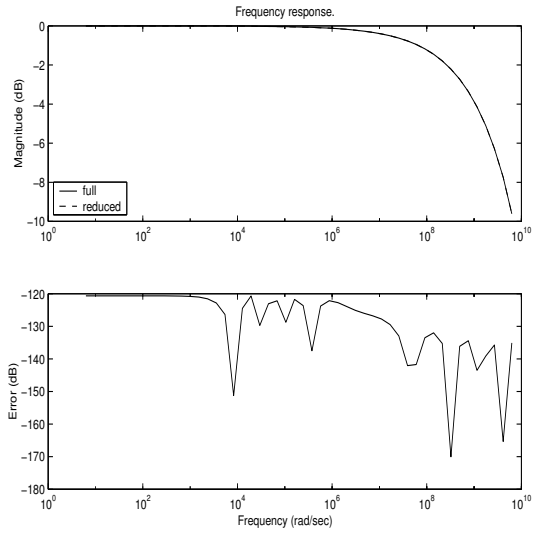


Fig. 8. Reduction of RLC line from 1000th order to 20th order, using Krylov subspace at $\gamma = 0$ (oblique projection).

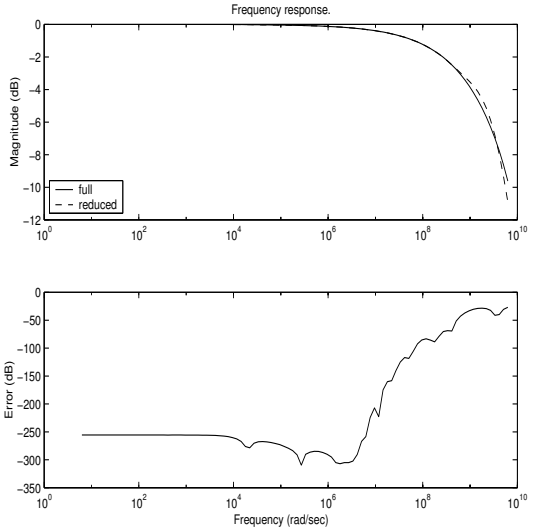


Fig. 9. Reduction of RLC line from 1000th order to 20th order by moment matching (without compaction).

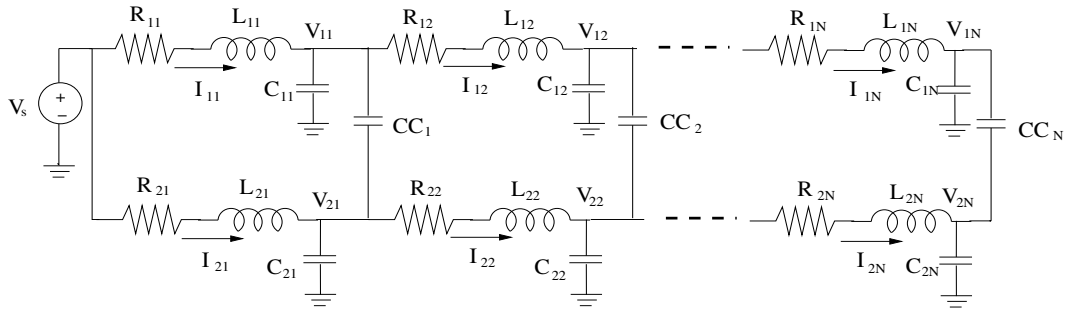


Fig. 10. Two coupled RLC lines.

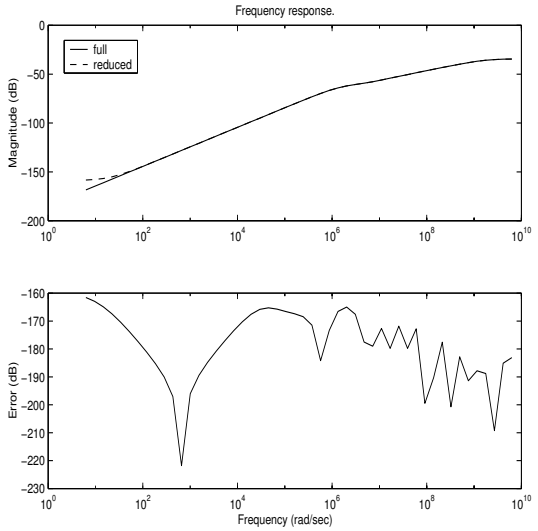


Fig. 11. Reduction of the coupled RLC line from 1200th order to 30th order, using Krylov subspace at $\gamma = 2 \times 10^8$ with oblique projection and compaction.

$x = (V_{11}, \cdots, V_{1n}, V_{21}, \cdots, V_{2N}, I_{11}, \cdots, I_{1N}, I_{21}, \cdots, I_{2N})^{\mathrm{T}} \in \mathbb{R}^{4N}$ is the state vector, $u = V_s$ is the input, and $y = I_{11} + I_{21}$ is chosen as the output. The uniform RLC values are chosen as: $R_{1i} = 20\,\Omega$, $R_{2i} = 15\,\Omega$, $L_{1i} = L_{2i} = 10.0\,nH$, $C_{1i} = C_{2i} = 1.0\,pF$, $CC_i = 20\,pF$ for $i = 1, \cdots, N$, where N is the number of stages. We choose $N = 600$ stages which results in a 1200th order model. The model is first reduced to order 90 by Krylov subspace projections, then further reduced to order 30 by dominance compaction. Here oblique projections to the dominant controllable and observable subspaces are used. The reduction result is shown in Fig. 11 where we choose $\gamma = 2 \times 10^8$. After the reduction, one of the 30 poles is unstable. The unstable pole is removed by projection to the stable subspace. Fig. 11 shows that the reduction error remains rather small even with one unstable pole removed.

6 Conclusion

Large scale models appearing in circuit simulation and other areas bring challenges to conventional model order reduction techniques. Motivated by the limitation of balanced truncation for large scale models, approximate dominant subspaces have been used for model reduction purposes. However, a theoretical error bound for non-balanced dominant subspace projection was unknown. This paper has established an $\mathcal{L}_2$ error bound for non-balancing dominant subspace model order reduction. Moreover, it has investigated the effectiveness of using several different Krylov subspaces possible for the computation of approximate dominant subspace. It has been justified by analysis that the conventionally used Krylov subspace $\mathcal{K}_m(A, B)$ is actually not a good candidate for dominant subspace computation purpose. Rather the other Krylov subspaces obtained from moment matching at low frequency and from discretization are better options for the computation purpose. The theoretical development in this paper has been tested by two large scale circuits to demonstrate the principle. As a further research topic, it would

be interesting to have a deeper understanding on the role played by moment matching in the discrete-time domain for model order reduction.

A Supplementary proof

Proof of Theorem 2, part (b) **:** The Arnoldi algorithm gives rise to the identity

$$A^{-1}V = VH_I + hv_{m+1}e_m^{\mathrm{T}} \tag{43}$$

where H_I is an upper Hessenberg matrix (the subscript I indicates that this Hessenberg matrix corresponds to the inverse of A), v_{m+1} is the $(m+1)$th basis vector in the Arnoldi algorithm and $h \geq 0$ is the normalization factor. By the fact that H_I is upper Hessenberg, it is readily verified that

$$A^{-i}Ve_1 = VH_I^i e_1, \qquad \text{for } i = 1, \cdots, m. \tag{44}$$

It also follows from (43) that

$$I = V^{\mathrm{T}}AVH_I + hV^{\mathrm{T}}Av_{m+1}e_m^{\mathrm{T}}$$

i.e.

$$\hat{A}^{-1} = H_I + h\hat{A}^{-1}V^{\mathrm{T}}Av_{m+1}e_m^{\mathrm{T}}. \tag{45}$$

One can show by induction that

$$\hat{A}^{-i}e_1 = H_I^i e_1, \qquad \text{for } i = 1, \cdots, m-1. \tag{46}$$

Indeed, eqn. (45) implies that $\hat{A}^{-1}e_1 = H_I e_1$. Assume that (46) holds for some i, $1 \leq i \leq m-2$. Then

$$\begin{aligned}
\hat{A}^{-(i+1)}e_1 &= \hat{A}^{-1}H_I^i e_1 \\
&= \left(H_I + h\hat{A}^{-1}V^{\mathrm{T}}Av_{m+1}e_m^{\mathrm{T}}\right)H_I^i e_1 \\
&= H_i^{i+1}e_1
\end{aligned}$$

since H_I being an upper Hessenberg matrix implies that $e_m^{\mathrm{T}}H_I^i e_1 = 0$ for $1 \leq i \leq m-2$.

Let $\beta = \|A^{-1}B\|$. Then $A^{-1}B = \beta Ve_1$, i.e. $B = \beta AVe_1$. Hence, $\hat{B} = \beta\hat{A}e_1$, i.e.

$$\hat{A}^{-1}\hat{B} = \beta e_1. \tag{47}$$

Thereby, $A^{-1}B = V\hat{A}^{-1}\hat{B}$. This shows the identity (31) for $i = 1$. By the identities (46), we obtain for $2 \leq i \leq m$

$$\begin{aligned}
A^{-i}B &= A^{-(i-1)}A^{-1}B \\
&= \beta A^{-(i-1)}Ve_1 = \beta VH_I^{i-1}e_1 \\
&= \beta V\hat{A}^{-(i-1)}e_1 \\
&= V\hat{A}^{-i}\hat{B}
\end{aligned}$$

where we also used (43) and (47). Consequently, the identities (31) hold for $i = 1, \cdots, m$. ■

References

[1] A.C. Antoulas, D.C. Sorensen, and S. Gugercin. A survey of model reduction methods for large-scale systems. In *Structured Matrix in Operator Theory, Numerical Analysis, Control, Signal and Image Processing*, Contemporary Mathematics. AMS Publications, 2001.

[2] Z. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43:9–44, 2002.

[3] Z. Bai, R.D. Slone, W.T. Smith, and Q. Ye. Error bound for reduced system model by padé approximation via the lanczos process. *IEEE Trans. on Computer-Aided Design*, 18(2):133–141, February 1995.

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

[5] D.L. Boley and G.H. Golub. The Lanczos-Arnoldi algorithm and controllability. *Systems & Control Letters*, 4:317–324, 1984.

[6] P. Feldmann and R.W. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. *IEEE Trans. on Computer-Aided Design*, 14(5):639–649, May 1995.

[7] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L_∞ error bounds. *International Journal of Control*, 39(6):1115–1193, 1984.

[8] A. Greenbaum. *Iterative Methods for Solving Linear Systems*. SIAM, Philadelphia, 1997.

[9] E.J. Grimme. *Krylov Projection Methods for Model Reduction*. PhD thesis, ECE Dept., University of Illinoise at Urbana-Champaign, 1997.

[10] T. Gudmundsson and A.J. Laub. Approximate solution of large spase Lyapunov equations. *IEEE Trans. on Automatic Control*, 39(5):1110–1114, 1994.

[11] S. Gugercin and A.C. Antoulas. A comparative study of 7 algorithms for model reduction. In *Proc. 39th IEEE Conf. on Decision and Control*, pages 2367–2372, Sydney, Australia, 2000.

[12] C.-W. Ho, A. E. Ruehli, and P. A. Brennan. The modified nodal approach to network analysis. *IEEE Trans. on Circuits and Systems*, CAS-22(6):504–509, 1975.

[13] A. Hodel and K. Poolla. Heuristic approaches to the solution of very large sparse Lyapunov and algebraic Riccati equations. In *Proc. 27th IEEE Conf. on Decision and Control*, pages 2217–2222, Austin, TX, 1988.

[14] A. Hodel, B. Tenison, and K. Poolla. Numerical solution of the Lyapunov equation by approximate power iteration. *Linear Algebra Appl.*, 236:205–230, 1996.

[15] I.M. Jaimoukha and E.M. Kasenally. Krylov subspace methods for solving large Lyapunov equations. *SIAM J. Numer. Anal.*, 31(1):227–251, 1994.

[16] S. Lall, J.E. Marsden, and S. Glavaski. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *Int. J. on Robust and Nonlinear Control*, 12(5):519–535, 2002.

[17] J.-R. Li, F. Wang, and J. White. An efficient Lyapunov equation-based approach for generating reduced-order models of interconnect. In *Proc. 36th ACM/IEEE Design Automation Conference*, New Orleans, Louisiana, 1999.

[18] A. Lu and E. Wachspress. Solution of Lyapunov equations by ADI iteration. *Computers and Mathematics with Applications*, 21(9):43–58, 1991.

[19] B.C. Moore. Principle component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. on Automatic Control*, AC-26(1):17–32, February 1981.

[20] A. Odabasioglu, M. Celik, and L.T. Pileggi. PRIMA: Passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. on Computer-Aided Design*, 17(8):645–654, August 1998.

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

[22] L.T. Pillage and R.A. Rohrer. Asymptotic waveform evaluation for timing analysis. *IEEE Trans. on Computer-Aided Design*, 9:352–366, April 1990.

[23] P. Rabiei and M. Pedram. Model order reduction of large circuits using balanced truncation. In *IEEE Proc. Asian Pacific Design Automation Conference*, pages 237–240, Feb. 1999.

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

[25] M.G. Safonov and R.Y. Chiang. A Schur method for balanced-truncation model reduction. *IEEE Trans. on Automatic Control*, AC-34(7):729–733, July 1989.

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

[27] C.D. Villemagne and R.E. Skelton. Model reduction using a projection formulation. *Int. J. Control*, 46:2141–2169, 1987.

[28] K. Zhou. *Essentials of Robust Control*. Prentice Hall, Upper Saddle River, New Jersey, 1998.