

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

Guoyong Shi, *Member, IEEE*, and C.-J. Richard Shi, *Senior Member, IEEE*

Abstract—Balanced truncation is a well-known technique for model-order reduction with a known uniform reduction error bound. However, its practical application to large-scale problems is hampered by its cubic computational complexity. While model-order reduction by projection to approximate dominant subspaces without balancing has produced encouraging experimental results, the approximation error bound has not been fully analyzed. In this paper, a square-integral reduction error bound is derived for unbalanced dominant subspace projection by using a frequency-domain solution of the Lyapunov equation. Such an error bound is valid in both the frequency and time domains. Then, a dominant subspace computation scheme together with three Krylov subspace options is introduced. It is analytically justified that the Krylov subspace for moment matching at low frequencies is able to provide a better dominant subspace approximation than the Krylov subspace at high frequencies, while a rational Krylov subspace with a proper real shift parameter is capable of achieving superior approximation than the Krylov subspace at low frequency. A heuristic method of choosing a real shift parameter is also introduced based on its new connection to the discretization of a continuous-time model. The computation algorithm and theoretical analysis are then examined by several numerical examples to demonstrate the effectiveness. Finally, the dominant subspace computation scheme is applied to the model-order reduction of two large-scale interconnect circuit examples.

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

I. INTRODUCTION

MODEL-order reduction is emerging as an effective technique for the modeling and simulation of very large-scale integrated 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 commonly leads to large-scale linear or nonlinear models that are computationally prohibitive even for modern computing resources. Therefore, reducing models before simulation is now becoming a common practice.

Manuscript received September 16, 2003; revised September 16, 2004. This work was supported in part by the U.S. Defense Advanced Research Projects Agency (DARPA) NeoCAD Program under Grant N66001-01-8920 and in part by the National Science Foundation (NSF) CAREER Award under Grant 9985507. This paper was recommended by Associate Editor C.-W. Wu.

The authors are with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: gshi@ee.washington.edu; cjshi@ee.washington.edu).

Digital Object Identifier 10.1109/TCSI.2005.846217

Large-scale full-order models commonly have a high degree of redundancy. Also, in many applications, an accurate model at a limited frequency range is of interest. In these cases, model-order reduction is capable of reducing the model redundancy and providing compact models for efficient simulation. Numerous model-reduction techniques have been developed in the past decades, 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 [3].

Two representative reduction techniques widely used in circuit simulation are balanced truncation [4]–[6] and moment matching [7]–[9]. Balanced truncation yields stable reduced-order models with a proven uniform error bound. However, due to its cubic computational complexity, balanced truncation is not directly applicable to large-scale model reduction. On the other hand, moment matching has a relatively lower computational complexity and can take the advantage of sparsity in circuit models, and has been widely used for integrated circuit modeling and analysis [8]. Moment matching for model-order reduction has been further popularized by the development of numerically stable computation methods based on Krylov subspaces [9], [10]. However, since moment matching only matches moments at some local frequency points, the resulting reduced-order model may have fairly large errors at some other frequency bands. Moreover, a small error bound in the frequency domain does not necessarily imply an accurate waveform matching in the time domain. Maintaining the stability of a reduced-order model is also a critical issue. Many authors have made efforts on extending the Krylov subspace methods in a variety of directions for better results, see [11]–[14] and the references therein.

Knowing the limitations of both techniques, some researchers attempt to use modified balanced truncation schemes to improve the global approximation accuracy while keeping a low computational cost [1], [15]–[17]. The underlying idea is to use approximately computed dominant subspaces and then to project the state space of a full-order model to the dominant subspaces. In this approach efficient computation of the approximate dominant subspace becomes an important task. Unlike the balanced truncation method where the exact Gramians are used for balancing transformation, approximate dominant subspaces are not sufficient for balancing. A natural question in this regard is, if one uses a dominant subspace for model reduction without balancing, what is the error bound? Such an error bound can help

us estimate the model reduction accuracy when *approximate* dominant subspaces are used in practice. Some good experimental results on circuit simulation problems by using the approximate dominant subspaces for model-order reduction publications have been reported [16], [17], but an error bound of using unbalanced dominant subspace projection has not been fully analyzed yet. The first contribution of this paper is the establishment of such an error bound. In Section III, we follow a frequency domain approach and derive a new \mathcal{L}_2 error bound which holds in both the frequency and time domains.

Dominant subspace is not only useful for linear model reduction, but also finds applications in nonlinear model-order reduction [18]. However, direct computation of the exact dominant subspace for large-scale models, linear or nonlinear, is in general not feasible in practice. For linear time-invariant models, the exact dominant subspace can only be obtained from solving the exact Gramian solution of a Lyapunov equation, which is of the cubic time complexity. The high computational complexity has motivated many researchers to study approximate solutions of a large-scale Lyapunov equation [15], [19]–[24]. It has been observed that frequently the Gramian solved from a Lyapunov equation is of low numerical rank [25], which implies that the state space of the full-order model is dominated by a low-dimensional subspace, see [3], [23] for some examples. In fact originally the balanced truncation principle and dominant subspace principle introduced for model-order reduction were motivated by this low-rank phenomenon. Also, because of this phenomenon, the computation of low-dimensional approximate dominant subspace becomes important for practical applications.

The key idea used for a low-rank approximation of a Gramian is to apply an iterative computation technique in the framework of Krylov subspace [26]. Most low-rank approximation techniques proposed in the literature utilized the idea of forming a Krylov subspace with the matrix pair (A, B) [see (2)], [20]–[22]. Only recently have two papers addressed the low-rank approximation using rational Krylov subspace as an extension of the alternate direction implicit (ADI) algorithm [23], [24]. In particular, numerical examples were presented in [24] to demonstrate that different types of Krylov subspaces could give rise to approximate dominant subspaces with different accuracy. But no analysis on the phenomenon was given there. We also found in our experiments similar effects by using different types of Krylov subspaces. In agreement to [24], we found that the Krylov subspace formed by the pair (A, B) always yielded the worst results.

In the second part of this paper, we carry out a careful study on the approximation accuracy of using three different types of Krylov subspaces. In Section IV, we first introduce a general approximate dominant subspace computation scheme based on Krylov subspace and low-order Lyapunov equation solving. Then, we justify analytically that the Krylov subspace formed by the pair $(A^{-1}, A^{-1}B)$ has a better approximation performance than that by the pair (A, B) , and furthermore a rational Krylov subspace with an appropriately chosen real shift parameter can produce superior approximation results than that of the Krylov subspace $(A^{-1}, A^{-1}B)$. We also introduce a heuristic for choosing a real shift parameter by building a connection between a real rational Krylov subspace and the discretization of a

continuous-time model. We show that model reduction in a *real* rational Krylov subspace bears the physical meaning of waveform matching in the discrete-time domain, in contrast to the implication of local approximation in the frequency domain for rational Krylov subspaces with purely imaginary shift parameters.

The computation scheme and theoretical analysis are then examined by numerical examples in Section V. First, we use an interconnect circuit example with different orders and element values to demonstrate that the three types of Krylov subspaces do provide different approximation accuracies in the dominant subspace computation, as predicted by the analysis in Section IV. For the evaluation purpose, three measures are introduced for comparing the approximation accuracy. The effectiveness of applying approximate dominant subspace to large-scale model-order reduction is further demonstrated by using two interconnect circuits.

The terminologies and notations used in this paper are fairly standard. A matrix is called a Hurwitz matrix if it is asymptotically stable. Vectors without transpose are in column convention. Since the column dimension of a matrix is of special interest in this paper, we specifically use a subscript to indicate the column size of a matrix in many places. The subspace spanned by the columns of a matrix $V_m \in \mathbb{R}^{m \times m}$ is denoted by $\text{span } V_m$. The subspace perpendicular to a subspace \mathcal{S} is denoted by \mathcal{S}^\perp . When we say that a matrix V_m spans a subspace, we mean that the columns of matrix V span the subspace, and we call the matrix V_m the basis matrix. The q th-order Krylov subspace generated by two matrices A and B is denoted by

$$\mathcal{K}_q(A, B) := \text{span}[B, AB, \dots, A^{q-1}B]. \quad (1)$$

In some places, we shall use the standard matrix manipulation notation used in MATLAB¹. For example, the notation $M(:, 1 : q)$ means the matrix formed by taking the first q columns from matrix M . The standard basis vectors, i.e., the columns of the identity matrix I , are denoted by e_i whose dimension should be clear from the context if not specified. The transpose of matrix A is denoted by A^T and if A is complex, the conjugate transpose of A is denoted by A^H . The i th eigenvalue of matrix A is denoted by $\lambda_i(A)$ and the maximal eigenvalue of A is denoted by $\lambda_{\max}(A)$ if A is symmetric. The norm of a vector $\|v\|$ is the conventional Euclidean 2-norm, i.e., $\|v\| := (v^T v)^{1/2}$. The norm of a matrix A is defined by $\|A\| := \lambda_{\max}^{1/2}(AA^T)$. The Frobenius norm of A is defined by $\|A\|_F := [\text{tr}(AA^T)]^{1/2}$. The maximum singular value of matrix A is denoted by $\sigma_{\max}(A) := \lambda_{\max}^{1/2}(AA^T)$.

II. PRELIMINARY

We consider circuits and structures that can be modeled by linear time-invariant (LTI) systems

$$\begin{aligned} \frac{dx}{dt} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \quad (2)$$

¹MATLAB is a trademark of The Mathworks, Inc. <http://www.mathworks.com>.

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

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

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

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

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

$$\begin{aligned} \frac{d\xi}{dt} &= \hat{A}\xi + \hat{B}u \\ y &= \hat{C}\xi + Du \end{aligned} \quad (5)$$

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

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

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_q and V_q be two real matrices in $\mathbb{R}^{n \times q}$ satisfying the biorthogonality condition

$$W_q^T V_q = I_q. \quad (7)$$

If we consider the restriction of the state x to span V_q , we can replace x by $V_q \xi$ and premultiply the first equation in (2) by W_q^T . The resulting model (5) is of a reduced order with

$$\hat{H}(s) = \left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & 0 \end{array} \right] = \left[\begin{array}{c|c} W_q^T A V_q & W_q^T B \\ \hline C V_q & 0 \end{array} \right]. \quad (8)$$

The quality of a reduced-order model obtained by projection can be measured by several different criteria. Typical measures are: 1) the number of moments matched at some frequency points [7]–[10] and 2) the uniform frequency domain error bound established for balanced truncation [5].

Given the LTI system in (2) with the system matrix A Hurwitz, two Gramians are important in the context of model-order reduction. The controllability Gramian is the unique solution P of the Lyapunov equation

$$AP + PA^T + BB^T = 0 \quad (9)$$

and the observability Gramian is the unique solution Q of the dual Lyapunov equation [27]

$$A^T Q + Q A + C^T C = 0. \quad (10)$$

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

$$P = \int_0^\infty e^{At} B B^T e^{A^T t} dt \quad (11)$$

and

$$Q = \int_0^\infty e^{A^T t} C^T C e^{At} dt. \quad (12)$$

Clearly, both Gramians are symmetric and positive semi-definite.

An alternate solution of Lyapunov equation (9) is an integral expression in the frequency domain, which is essentially a result of the matrix form Parseval identity. This alternate expression turns out to be useful for deriving an \mathcal{L}_2 error bound for unbalanced dominant projection in the next section. Meanwhile, it provides an analytical justification that a Krylov subspace computed at the low frequencies can provide a better approximate dominant subspace to be discussed in Section IV. Since the authors have not seen this alternate expression in the literature, a formal proof is provided for completeness.

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

$$P = \frac{1}{2\pi} \int_{-\infty}^{\infty} (j\omega I - A)^{-1} B B^T (-j\omega I - A^T)^{-1} d\omega. \quad (13)$$

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

$$(j\omega I - A)P + P(-j\omega I - A^T) = B B^T$$

which is equivalent to

$$\begin{aligned} (j\omega I - A)^{-1} P + P(-j\omega I - A^T)^{-1} \\ = (j\omega I - A)^{-1} B B^T (-j\omega I - A^T)^{-1}. \end{aligned}$$

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

$$\begin{aligned} \int_{-\infty}^{\infty} (j\omega I - A)^{-1} d\omega P + P \int_{-\infty}^{\infty} (-j\omega I - A^T)^{-1} d\omega \\ = \int_{-\infty}^{\infty} (j\omega I - A)^{-1} B B^T (-j\omega I - A^T)^{-1} d\omega. \end{aligned} \quad (14)$$

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 along the $j\omega$ axis and following the left semi-circle $Re^{j\theta}$ for $\theta \in [(\pi/2), (3\pi/2)]$. For sufficiently large R , this loop encircles all eigenvalues of A but not those of $-A$. By Cauchy Theorem we obtain for R sufficiently large that

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

Then, the identity (13) follows immediately by substituting $s = j\omega$ and letting $R \rightarrow \infty$. \blacksquare

III. ERROR BOUND

Suppose that W_q and V_q are two projection matrices in $\mathbb{R}^{n \times q}$ satisfying $W_q^T V_q = I_q$ and the reduced-order model (8) is obtained from these two projection matrices. The error of model-order reduction is defined to be the difference between the full-order model and a reduced-order model in the frequency domain, i.e.,

$$E(s) = H(s) - \hat{H}(s). \quad (15)$$

Several error bounds are available in the literature. A well-known error bound was established in [5] for balanced truncation. Its practical use is, however, limited to small-scale models since the computational complexity of solving Lyapunov equations and singular value decomposition is of $\mathcal{O}(n^3)$. Another error bound was derived for moment matching in [28]. However, since moment matching is based on the Taylor expansion in the local sense, an error bound for the truncated terms does not provide much information on the quality of a reduced-order model in the wideband sense.

Since the theoretically solid error bound for balanced truncation cannot be used if the exact Gramians are not available, it is of interest to consider the error bound if only one exact Gramian is used without performing the balancing transformation. Such an error bound would be useful in practice because, as will be discussed later in this paper, we are able to compute an approximate dominant subspace with high accuracy using specially designed Krylov subspace methods. If an error bound is available for a reduced model obtained from an exactly computed dominant subspace, then this error bound together with the subspace approximation error can be used for an estimate of the model reduction error. Moreover, the better an approximate dominant subspace is computed, the more trustful is the error bound. For large-scale models, the exact error bound is rarely computed because of the computational complexity, nevertheless it still serves as a theoretical measure to justify that an accurately computed approximate dominant subspace is trustful in model-order reduction.

Before establishing the new error bound mentioned above, we first derive a general characterization of the error function defined in (15).

Lemma 2: Suppose that the reduced-order model (8) is obtained from the two projection matrices W_q and V_q in $\mathbb{R}^{n \times q}$ satisfying $W_q^T V_q = I_q$. The following identity holds for the error $E(s)$ defined in (15)

$$E(s) = L(s) (I - V_q W_q^T) F(s) \quad (16)$$

where

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

By duality, an alternate error expression is

$$E(s) = F_d(s) (I - V_q W_q^T) L_d(s) \quad (17)$$

where

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

Proof: In the packed notation, we can write

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

A state transformation

$$\begin{bmatrix} I & W_q^T \\ 0 & I \end{bmatrix}$$

of this error system leads to

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

This is equivalent to a system in the augmented state space

$$\begin{aligned} \dot{\eta} &= \hat{A}\eta - W_q^T A (I - V_q W_q^T) x \\ \dot{x} &= Ax + Bu \\ e &= -\hat{C}\eta + C (I - V_q W_q^T) x \end{aligned}$$

where $\eta = \xi - W_q^T x$, with zero initial conditions. Identity (16) then follows directly from this system.

The dual identity (17) is proven by considering $E^T(s)$ and replacing the triple (A, B, C) by its dual (A^T, C^T, B^T) . ■

The importance of Lemma 2 lies in the factor $(I - V_q W_q^T)$. Since $W_q^T V_q = I_q$, this factor is an oblique projector to the subspace $(\text{span} W_q)^\perp$ along $\text{span} V_q$. The dual factor $(I - W_q V_q^T)$ is another oblique projector to the subspace $(\text{span} V_q)^\perp$ along $\text{span} W_q$.

Lemma 2 has two immediate applications. First, it can be used to check the accuracy of moment matching if V_q and W_q are computed from the moment matching algorithms [7], [9]. For example, if W_q and V_q are generated by the q -step Lanczos process as in PVL [9], 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 is instrumental for deriving an error bound for $E(s)$ if either W_q or V_q spans respectively the dominant observable or controllable subspace, which is the remaining task of this section.

The standard \mathcal{L}_2 norm of the error $E(s)$ is defined as

$$\|E\|_2 := \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{tr}[E(j\omega)E^H(j\omega)] d\omega \right\}^{1/2} \quad (18)$$

where $E^H(j\omega) = E^T(-j\omega)$, and the standard \mathcal{L}_∞ norm of $L(s)$ is defined as [27]

$$\|L\|_\infty := \sup_{\omega \in (-\infty, \infty)} \sigma_{\max}[L(j\omega)]. \quad (19)$$

The next theorem establishes a bound on $\|E\|_2$, where we only consider the controllability Gramian and assume that the exact dominant subspace corresponding to the dominant singular values is available.

Theorem 1: Let P be the controllability Gramian and $P = U\Sigma U^T$ be the SVD (Singular Value Decomposition) of P , where U is an orthonormal matrix and $\Sigma = \text{diag}[\sigma_1, \dots, \sigma_n]$ is a diagonal matrix containing the singular values of P in the descending order. Let $1 \leq q < n$. If $V_q = U(:, 1 : q)$ and $W_q \in \mathbb{R}^{n \times q}$ satisfying $W_q^T V_q = I_q$ are used for reduction

projection and assume that the reduced matrix $\hat{A} = W_q^T A V_q$ is asymptotically stable. Then, we have

$$\|E\|_2 \leq \left(\sum_{i=q+1}^n \sigma_i \right)^{1/2} \|L\|_\infty \|I - W_q V_q^T\| \quad (20)$$

where $L = L(s)$ is defined in Lemma 2. If $W_q = V_q$, then (20) reduces to

$$\|E\|_2 \leq \left(\sum_{i=q+1}^n \sigma_i \right)^{1/2} \|L\|_\infty. \quad (21)$$

Proof: Since \hat{A} is asymptotically stable, the \mathcal{L}_∞ norm $\|L\|_\infty$ is finite. Let $V_q^c = U(:, q+1 : n)$ be the columns of U complementary to V_q and $\Sigma_q^c = \text{diag}[\sigma_{q+1}, \dots, \sigma_n]$. Using the error expression (16) in Lemma 2 and the integral solution P in Lemma 1, we can bound the \mathcal{L}_2 norm $\|E\|_2$ defined in (18) as follows:

$$\begin{aligned} \|E\|_2^2 &\leq \text{tr} \left\{ (I - V_q W_q^T) P (I - W_q V_q^T) \right\} \|L\|_\infty^2 \\ &= \text{tr} \left\{ (I - V_q W_q^T) V_q^c \Sigma_q^c V_q^{cT} (I - W_q V_q^T) \right\} \|L\|_\infty^2 \\ &\leq \|L\|_\infty^2 \left(\sum_{i=q+1}^n \sigma_i \right) \|I - W_q V_q^T\|^2 \end{aligned}$$

where we have used the fact that $(I - V_q W_q^T) V_q = 0$. Hence the error bound (20) follows. Bound (21) is due to the fact that $\|I_q - V_q V_q^T\| = 1$ for $q < n$. ■

Remark 1: By Parseval's theorem [27], the \mathcal{L}_2 norm of error $E(s)$ in the frequency domain is the same as that in the time domain. Therefore, the reduction error bounds in Theorem 1 are valid in both the time and frequency domains. Consequently, a good waveform approximation is guaranteed in the time domain as well if a good frequency domain approximation is established. Note that the \mathcal{L}_∞ error bound of balanced truncation does not have such a property.

An immediate consequence of Theorem 1 is that if the trailing singular values are all zero, then the projection based reduction does not lose any accuracy.

Corollary 1: Under the same conditions as in Theorem 1, if $\sigma_i = 0$ for $i = q+1, \dots, n$, 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 the congruence transformation preserves passivity for circuit models with the port formulation, choosing $W_q = V_q$ is preferred if the columns of V_q are orthonormalized [10]. In Theorem 1 we assumed that $\hat{A} = W_q^T A V_q$ is asymptotically stable. However, the stability of $\hat{A} = W_q^T A V_q$ is in general not guaranteed, unless W_q and V_q are obtained from balanced truncation. In practice sometimes \hat{A} might contain a few unstable poles. If this happens, certain postprocessing is needed, such as dropping the unstable poles by extracting the stable subspace. Since \hat{A} is usually 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.

IV. COMPUTATION OF APPROXIMATE DOMINANT SUBSPACES

In Section III, we established an \mathcal{L}_2 error bound in model reduction by unbalanced dominant subspace projection. However, since the exact Gramians are not easily computable for large-scale models, we have to resort to approximation in practice for dominant subspaces computation. As long as an approximate dominant subspace is computed with sufficient accuracy, the model reduction accuracy can be estimated by the error bound derived above.

Approximate computation of dominant subspaces for large-scale models has been studied by many researchers in the literature. Most of the results stem from the idea of low-rank approximate solution of a high-order Lyapunov equation. The low-rank solution of Lyapunov equation was first addressed by Hodel and Poolla [19], where several heuristic algorithms were proposed. Saad [20] 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 [21] extended Saad's idea on the single-input-single-output case to the multiple-input-multiple-output case and proposed a GMRES-like solution scheme by deriving an explicit expression for the residual. In all of these papers, the Krylov subspace was formed by the pair (A, B) because of their appearance in the Lyapunov equation (9). However, satisfaction of the Galerkin condition does not necessarily imply that the Krylov subspace formed by the pair (A, B) is optimal for effective dominant subspace computation. Other algorithms along the same line are proposed in [22]–[24], where in [23], [24] rational Krylov subspace was used for low-rank approximate solution in the framework of ADI algorithms.

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. Some other Krylov subspace options with a commensurate computation complexity can potentially provide better results. In Section IV-A, we introduce a general dominant subspace computation scheme, which is a different algorithm than the CF-ADI algorithm in [24]. The specific choice of Krylov subspace is not specified in the computation scheme. In Section IV-B, we review the properties of three types of Krylov subspaces that are optional for the computation scheme. These properties are used in Section IV-C for an analytical comparison of their performance in dominant subspace approximation. Finally, in Section IV-D we establish a new connection between a real rational Krylov subspace and the discretization of a continuous-time model and propose a heuristic way of choosing an appropriate real shift parameter for a real rational Krylov subspace.

A. Dominant Subspace Computation Scheme

We start with the outline of a general iterative dominant subspace computation scheme. Algorithms similar to this scheme

have been used in some works [20], [21], but not from the point of view of comparing the performances of different Krylov subspaces. We formulate the scheme in a generic way without specifying the matrices used for the Krylov subspace computation. Later on, this generic algorithm is specialized to several different Krylov subspaces and their performances in dominant subspace approximation are compared analytically. The computation scheme is general enough for multiple-input models.

1) *Dominant Subspace Computation Scheme:*

Input) Two matrices $\Phi \in \mathbb{R}^{n \times n}$ and $\Theta \in \mathbb{R}^{n \times r}$, the dimension q of an approximate dominant subspace, and an intermediate integer m satisfying $r \leq q \leq m < n$.

Step 1) Run the (block) Arnoldi algorithm [26] to generate the basis vectors for the ν th-order Krylov subspace $\mathcal{K}_\nu(\Phi, \Theta)$. Let $V_m \in \mathbb{R}^{n \times m}$ ($m \leq \nu$) be the basis matrix, i.e., $\text{span}V_m = \mathcal{K}_\nu(\Phi, \Theta)$.

Step 2) Form the reduced-order Lyapunov equation

$$A_m P_m + P_m A_m^T + B_m B_m^T = 0 \quad (22)$$

where

$$A_m = V_m^T A V_m, \quad B_m = V_m^T B.$$

Step 3) Solve P_m from the m th-order Lyapunov equation (22) and find the SVD of P_m , i.e.,

$$P_m = U_m \Sigma_m U_m^T$$

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

Step 4) Extract the leading q column vectors of U_m , denoted by $\tilde{U}_q = U_m(:, 1 : q)$, and define $\tilde{V}_q = V_m \tilde{U}_q$.

Output) Matrix \tilde{V}_q whose columns span a q -dimensional 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 large (normally no greater than 100) so that solving the Lyapunov equation (22) is kept at a low cost.

One could choose to return V_m after Step 1) for an approximate subspace as done in [24, Sec. VIII]. However, V_m directly generated by the Arnoldi algorithm without the correction in the other steps might not capture the dominance very well. Steps 2) to 4) in the computation scheme are merely for a better low-dimensional approximation of dominance subspace with some modest additional computation cost. We refer to the reduction procedure from m basis vectors to q dominant basis vectors as *compaction*.

It is easily verified that if \tilde{V}_q is returned from the computation scheme, then $\tilde{A}_q = \tilde{V}_q^T A \tilde{V}_q$ and $\tilde{B}_q = \tilde{V}_q^T B$ satisfy

$$\tilde{A}_q \Sigma_q + \Sigma_q \tilde{A}_q^T + \tilde{B}_q \tilde{B}_q^T = 0 \quad (23)$$

where Σ_q is a diagonal matrix containing the leading q singular values of Σ_m . The computation scheme outlined above also yields a low-rank approximation of the Gramian

$$\tilde{P} = V_m P_m V_m^T = (V_m U_m) \Sigma_m (U_m^T V_m^T). \quad (24)$$

This approximate solution has the property that the Galerkin condition is satisfied [20], i.e.,

$$V_m^T R(\tilde{P}) V_m = 0 \quad (25)$$

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

$$R(X) = AX + XA^T + BB^T. \quad (26)$$

Remark 4: The dominant subspace computation scheme is formulated for the computation of a dominant controllable subspace. It can also be used for the computation of a dominant observable subspace if the dual matrices (A^T, C^T) are used in place of (A, B).

B. *Three Krylov Subspaces for Dominance Approximation*

Note that we did not specify the choice of the matrices Φ and Θ for the Krylov subspace $\mathcal{K}_m(\Phi, \Theta)$ in the computation scheme described above. There are three typical Krylov subspaces optional for the computation scheme. In this section, we briefly review the construction and basic properties of the Krylov subspaces. These properties will be used to examine the approximation performances of the three types of Krylov subspaces.

A traditional option for Φ and Θ is to choose $\Phi = A$ and $\Theta = B$; that is, the Krylov subspace $\mathcal{K}_m(A, B)$ is used for the dominant controllable subspace computation. The choice of this Krylov subspace used in many earlier works is a direct consequence of the Lyapunov equation in the standard form of (9) and its solution in the conventional form of (11).

The second choice of Krylov subspace comes from the following reformulation. If A is asymptotically stable (hence invertible), the Lyapunov equation (9) can equivalently be written as

$$A^{-1}P + PA^{-T} + A^{-1}BB^T A^{-T} = 0 \quad (27)$$

which remains to be a Lyapunov equation. This operation leads to another way of computing the dominant controllable subspace by using the pair $(\Phi, \Theta) = (A^{-1}, A^{-1}B)$. We note that in circuit applications obtaining A or A^{-1} involves almost the equal amount of computation, depending on whether inverting the susceptance matrix or the conductance matrix.

The third option for choosing the pair (Φ, Θ) is by using a rational Krylov subspace [29] with a real shift parameter $\gamma > 0$

$$\mathcal{K}_m((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B). \quad (28)$$

Note that by choosing $\gamma = 0$ this Krylov subspace reduces to the second option $\mathcal{K}_m(A^{-1}, A^{-1}B)$. It should be noted that there are many possible ways of choosing the shift parameter γ , which in general could be complex and distinct in the iteration, for instance, $[(\gamma_1 I - A)^{-1}B, (\gamma_2 I - A)^{-2}B, \dots]$ as typically used in the ADI-type algorithms [23], [24], [30]. Clearly, if a set of distinct shift parameters is used, additional linear solves for $(\gamma_i I - A)^{-1}B$ are required. In this paper we focus ourselves on rational Krylov subspaces with only one real shift parameter γ ,

and compare its performance to the first two options in dominant subspace approximation. An algorithm using distinct shift parameters can be found in [24].

We are mainly interested in one simple question: which one among the three Krylov subspaces could provide the best results for dominant subspace approximation if the dominant subspace computation scheme is used with low orders m and q ? We learned from our numerical experiments that in general the traditionally used Krylov subspace $\mathcal{K}_m(A, B)$ almost always yielded the worst approximation results, the Krylov subspace $\mathcal{K}_m(A^{-1}, A^{-1}B)$ could provide a relatively better approximation result, and the rational Krylov subspace (28) with an appropriately chosen positive γ frequently gave rise to the most superior approximation result. This observation is also consistent with the numerical results reported by Li and White [24]; but no analytical justification was given there. In the next section we shall attempt to provide an analytical justification for the different approximation performances of the three Krylov subspaces. For this purpose, the following three properties associated with moment matching in the three Krylov subspaces will be useful.

The moment matching property of a Krylov subspace is now well-known. Let $X(s) = F(s)U(s)$ be the Laplace transform of the state x of model (2), where $F(s) = (sI - A)^{-1}B$ is the transfer function from input to state. The Taylor expansion of $F(s)$ can take the following three forms:

$$F(s) = \sum_{i=0}^{\infty} A^i B s^{-(i+1)} \quad (29a)$$

$$= - \sum_{i=0}^{\infty} A^{-(i+1)} B s^i \quad (29b)$$

$$= \sum_{i=0}^{\infty} (-1)^i (\gamma I - A)^{-(i+1)} B (s - \gamma)^i \quad (29c)$$

where the first expansion is at $s = \infty$, the second at $s = 0$, and the third at $s = \gamma$. The leading coefficient matrices (vectors) of the terms $s^{-(i+1)}$, s^i , and $(s - \gamma)^i$ in the above three expansions are, respectively, the column vectors forming the three Krylov subspaces we mentioned above.

An important property of (rational) Krylov subspace is that if we reduce the matrices A and B by projection to a low-dimensional Krylov subspace characterized by an orthonormal basis matrix V_q , then the leading q moments of the following two input-to-state transfer functions are matched

$$F(s) = (sI - A)^{-1}B \quad (30a)$$

$$\tilde{F}(s) = V_q^T (sI - \hat{A})^{-1} \hat{B} \quad (30b)$$

where

$$\hat{A} = V_q^T A V_q \quad \text{and} \quad \hat{B} = V_q^T B. \quad (31)$$

are the two reduced matrices.

The moment matching properties of the three Krylov subspaces are listed below for the general multiple-input case. They will be used in the next Section for dominance approximation analysis. By an orthonormal basis matrix V_q of a Krylov

subspace \mathcal{K}_m we mean that V_q has orthonormal columns (i.e., $V_q^T V_q = I_q$) and $\mathcal{K}_m = \text{span} V_q$.

Property 1: If $V_q, q \leq m$, is the orthonormal basis matrix of the Krylov subspace $\mathcal{K}_m(A, B)$, then we have

$$A^i B = V_q \hat{A}^i \hat{B} \quad (32)$$

for $i = 0, \dots, m - 1$, where \hat{A} and \hat{B} are defined in (31).

Property 2: If $V_q, q \leq m$, is the orthonormal basis matrix of the Krylov subspace $\mathcal{K}_m(A^{-1}, A^{-1}B)$ and \hat{A} is invertible, then we have

$$A^{-i} B = V_q \hat{A}^{-i} \hat{B} \quad (33)$$

for $i = 1, \dots, m$, where \hat{A} and \hat{B} are defined in (31).

Property 3: If $V_q, q \leq m$, is the orthonormal basis matrix of the rational Krylov subspace $\mathcal{K}_m((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B)$ with $\gamma > 0$, then we have

$$(\gamma I - A)^{-i} B = V_q (\gamma I - \hat{A})^{-i} \hat{B} \quad (34)$$

for $i = 1, \dots, m$, where \hat{A} and \hat{B} are defined in (31), and γ is chosen such that the matrix inversions exist.

The first property follows directly from the Arnoldi algorithm. Its proof is straightforward and can be found in, for example, [7], [31]. The proof of Property 3 can be found in [29, Sec. 3.1]. The proof of Property 2 is slightly different from the other two, because the reduced matrix $V_q^T A V_q$ rather than $V_q^T A^{-1} V_q$ is formed. For completeness, a proof is provided in Appendix.

To be specific, we shall use the term *moment matching at high frequencies* to indicate the first type of Krylov subspace which is obtained by the Taylor expansion at $s = \infty$, the term *moment matching at low frequencies* to indicate the second type of Krylov subspace which is obtained by the Taylor expansion at $s = 0$, and the term *moment matching at the real γ* for the third type Krylov subspace which is obtained by the Taylor expansion at $s = \gamma$.

C. Comparison of Performances in Dominance Approximation

Our experiment and the numerical examples reported in [24] all show that the three Krylov subspaces examined in the preceding Section have different performances for the approximation of dominant subspaces, especially for large-scale models. The goal of this section is to justify using the tools developed so far that in the computation of approximate dominant subspace for large-scale models, the following two observations are in general true.

- 1) The Krylov subspace $\mathcal{K}_m(A^{-1}, A^{-1}B)$ has a better performance than that of $\mathcal{K}_m(A, B)$.
- 2) The Krylov subspace $\mathcal{K}_m((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B)$ with an appropriately chosen $\gamma > 0$ has a superior performance than that of $\mathcal{K}_m(A^{-1}, A^{-1}B)$.

We first justify the first observation by using Properties 1 and 2 and the the integral expression of P in the frequency domain (see (13)). Let $V^{(\infty)}$ and $V^{(0)}$ be the basis matrices of the Krylov subspaces $\mathcal{K}_m(A, B)$ and $\mathcal{K}_m(A^{-1}, A^{-1}B)$, respectively. Let

$$\hat{A}^{(t)} = V^{(t)T} A V^{(t)} \quad \hat{B}^{(t)} = V^{(t)T} B$$

and $F^{(t)}(s) = V^{(t)}\hat{F}^{(t)}(s)$ where $\hat{F}^{(t)}(s) = (sI - \hat{A}^{(t)})^{-1}\hat{B}^{(t)}$ for $t = 0, \infty$.

Let $\hat{P}^{(t)}$ ($t = 0, \infty$) be the solution of the reduced-order Lyapunov equation

$$\hat{A}^{(t)}\hat{P}^{(t)} + \hat{P}^{(t)}\hat{A}^{(t)T} + \hat{B}^{(t)}\hat{B}^{(t)T} = 0.$$

Then, according to the dominant subspace computation scheme, the two matrices $\tilde{P}^{(t)} := V^{(t)}\hat{P}^{(t)}V^{(t)T}$ for $t = 0, \infty$ are two approximations to the the controllability Gramian P represented as (see Lemma 1)

$$P = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega)F^H(j\omega) d\omega. \quad (35)$$

Then, we have

$$\begin{aligned} \tilde{P}^{(t)} &= V^{(t)}\hat{P}^{(t)}V^{(t)T} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} V^{(t)}\hat{F}^{(t)}(j\omega)\hat{F}^{(t)H}(j\omega)V^{(t)T} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F^{(t)}(j\omega)F^{(t)H}(j\omega) d\omega \\ &\approx \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega)F^H(j\omega) d\omega. \end{aligned}$$

The accuracy of the approximation in the last step can now be evaluated by the moment matching properties of the two associated Krylov subspaces.

Properties 1 and 2 imply that the leading m moments of $F^{(\infty)}(s)$ match those of $F(s) = (sI - A)^{-1}B$ at $s = \infty$, while the leading m moments of $F^{(0)}(s)$ match those of $F(s)$ at $s = 0$. Since $(j\omega I - A)^{-1}$ rolls off as $\omega \rightarrow \infty$, the magnitude of the integral of P in (35) mainly comes from the contribution of $F(j\omega) = (j\omega I - A)^{-1}B$ at the low-frequency part rather than from the high-frequency part. Hence, in the sense of approximation, an integral with its integrand matching the moments of $F(s)$ at the low frequencies and rolling off at the high frequencies must have a better approximation to the Gramian P than the one that matches only the moments of $F(s)$ at the high-frequency which is the rolling off (nonsignificant) part. Observation 1) is thus justified. Fortunately, this result is not against the practice where the circuit operation frequency band is typically below the gigahertz level, never going toward the infinitely high frequencies.

Observation 2) is justified by a different approach that uses another representation of the controllability Gramian P in infinite series. Note that the Lyapunov equation (9) can equivalently be written as

$$(\gamma I - A)P(\gamma I - A)^T = (\gamma I + A)P(\gamma I + A)^T + 2\gamma BB^T.$$

Assuming $(\gamma I - A)$ invertible, we get

$$P = \Phi(\gamma)P\Phi^T(\gamma) + (2\gamma)(\gamma I - A)^{-1}BB^T(\gamma I - A)^T \quad (36)$$

where $\Phi(\gamma) := (\gamma I - A)^{-1}(\gamma I + A)$. This is another type of Lyapunov equation arising from discrete-time systems; its solution can be obtained by the following iteration:

$$P_{k+1} = \Phi(\gamma)P_k\Phi^T(\gamma) + (2\gamma)(\gamma I - A)^{-1}BB^T(\gamma I - A)^T. \quad (37)$$

The convergence of this iteration is guaranteed if A is Hurwitz, since in this case $\Phi(\gamma)$ has spectral radius less than one for any

$\gamma > 0$. This algorithm is known as the Smith algorithm [32]. With zero initial condition $P_0 = 0$, the iteration (37) converges to

$$P = \sum_{i=0}^{\infty} [\Phi^i(\gamma)]M[\Phi^i(\gamma)]^T \quad (38)$$

where

$$M = (2\gamma)(\gamma I - A)^{-1}BB^T(\gamma I - A)^T. \quad (39)$$

Note that the Smith algorithm is rarely used in practical computation due to its slow convergence especially for stiff matrix A , i.e., A has both fast and slow modes, which is typical for many circuit models. Variants of the Smith algorithm lead to the ADI algorithm [30] and the CF-ADI algorithm [24]. Here, we use the series representation of P in (38) and Property 3 to justify Observation 2).

Using the identity

$$\Phi(\gamma) = (\gamma I - A)^{-1}(\gamma I + A) = 2\gamma(\gamma I - A)^{-1} - I. \quad (40)$$

one can easily verify that

$$\mathcal{K}_m(\Phi(\gamma), (\gamma I - A)^{-1}B) = \mathcal{K}_m((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B).$$

Let V_q ($q \leq m$) be the orthonormal basis matrix of the rational Krylov subspace $\mathcal{K}_m((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B)$ and define

$$\hat{\Phi}(\gamma) := (\gamma I - \hat{A})^{-1}(\gamma I + \hat{A})$$

where \hat{A} is as defined in (31). Then, we have the following lemma.

Lemma 3: Given the notations above, the following identities hold:

$$\Phi^i(\gamma)(\gamma I - A)^{-1}B = V_q\hat{\Phi}^i(\gamma)(\gamma I - \hat{A})^{-1}\hat{B} \quad (41)$$

for all $i = 0, \dots, m-1$, where $\gamma > 0$ is chosen such that the two matrix inversions exist and \hat{A} and \hat{B} are defined in (31).

Proof: The identities (41) follow from the moment matching property 3 and the identity (40). ■

It follows directly from Lemma 3 that the following partial-sum identity holds:

$$\sum_{i=0}^{m-1} [\Phi^i(\gamma)]M[\Phi^i(\gamma)]^T = V_q \sum_{i=0}^{m-1} [\hat{\Phi}^i(\gamma)]\hat{M}[\hat{\Phi}^i(\gamma)]^T V_q^T \quad (42)$$

where M is defined in (39) and \hat{M} is defined by

$$\hat{M} = (2\gamma)(\gamma I - \hat{A})^{-1}\hat{B}\hat{B}^T(\gamma I - \hat{A})^T. \quad (43)$$

Now let \hat{P} be the solution of the reduced-order Lyapunov equation

$$\hat{A}\hat{P} + \hat{P}\hat{A}^T + \hat{B}\hat{B}^T = 0.$$

Then, according to the Smith algorithm, the solution \hat{P} can also be expressed as a series if \hat{A} remains Hurwitz, i.e.,

$$\hat{P} = \sum_{i=0}^{\infty} [\hat{\Phi}^i(\gamma)]\hat{M}[\hat{\Phi}^i(\gamma)]^T. \quad (44)$$

Going back to our dominant subspace computation scheme, we can use $V_q\hat{P}V_q^T$ as an approximation to the controllability Gramian P .

The identity (42) is now used to compare the approximation accuracy of $V_q \hat{P} V_q^T \approx P$ by using different γ 's. If an appropriate γ is chosen so that the convergence of (38) is optimal, then the approximation by using $V_q \hat{P} V_q^T$ for P should be superior if V_q is the basis matrix of the rational Krylov subspace $\mathcal{K}_m((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B)$ comparing to other rational Krylov subspaces with other γ . This is because for an optimum γ such that the convergence of (38) is optimal, the partial sum

$$\sum_{i=0}^{m-1} [\Phi^i(\gamma)] M [\Phi^i(\gamma)]^T \quad (45)$$

better dominates the total sum. The partial-sum identity (42) then implies that the subspace dominance is also better captured by the basis matrix V_q if the rational Krylov subspace at an optimum γ is selected in computation. On the other hand, observe that the second type of Krylov subspace $\mathcal{K}_m(A^{-1}, A^{-1}B)$ is the limiting case of a rational Krylov subspace by choosing γ sufficiently small, which is however associated with the *slow* convergence of (38); namely, the dominance captured by the partial-sum (45) for $\gamma \rightarrow 0$ is not as good as that by using an optimum γ . Hence this partial-sum matching approach has justified that Observation 2) is in general true.

D. Choice of γ

In the previous section, we argued that an optimum γ would result in an optimal approximation of the dominant subspace by using a real rational Krylov subspace. However, in general, it is not easy to find an optimum γ for such a purpose. The similar issue of choosing optimum convergence parameters in ADI-type algorithms has been addressed in many places, see [24] and the references therein. The main technique is to solve a rational minimax problem which requires the information of the eigenvalues of the matrix A . This problem is not completely solved when some eigenvalues of A are complex. Also, when A is high dimensional, finding the eigenvalues or a compact region containing all the eigenvalues of A is also expensive in general. To have a guideline for choosing an appropriate parameter γ , we provide a heuristic method here by following a discretization approach.

Rational Krylov subspace has been widely used in model-order reduction. Since moment matching is traditionally formulated as local approximation in the frequency domain, a pure complex $s = j\omega$ or several such points can be used as the shift parameters to improve the approximation accuracy at certain frequency ranges of interest [33]. Also real numbers can be used for shift parameters, see for example [9], [2], which also give good simulation results. However, since moment matching at a real point does not have a direct connection to the local approximation of a frequency response, which is normally along the $j\omega$ axis, the physical meaning of a rational Krylov subspace with *real* shift parameters is not well understood. Grimme made some argument in his thesis for the real shift parameters [29, Sec. 6.2.2], but the interpretation is not as clear as that of pure imaginary shift points. It turns out that interpreting a real shift parameter in terms of discretization bears a better physical meaning for practical applications.

One way to discretize a continuous-time model

$$\dot{x} = Ax + Bu \quad (46)$$

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

$$s \approx \frac{z-1}{hz} = \frac{\gamma(z-1)}{z} \quad (47)$$

where $z = e^{sh}$, h is a small time step, and $\gamma = 1/h$ is the sampling frequency. After the discretization, the continuous-time model (46) becomes

$$(\gamma I - A)x_{k+1} = \gamma x_k + Bu_{k+1} \quad (48)$$

which is the same as the backward Euler integration formula.

Taking z -transform, we obtain $X(z) = F(z)U(z)$, where

$$F(z) = [(\gamma I - A)z - \gamma I]^{-1} BzU(z). \quad (49)$$

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

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

where $A_\gamma = (\gamma I - A)^{-1}$ and $B_\gamma = A_\gamma B$. It is clear that the leading m coefficients in the expansion of $F(z)$ are the matrices that form the rational Krylov subspace $\mathcal{K}_m((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B)$. If a sequence of different γ_i 's are used in the iteration, corresponding to using varying time steps, then a rational Krylov subspace with a set of real shift parameters is formed. Some properties of a rational Krylov subspace with distinct shift parameters are discussed in [24], where the rational Krylov subspaces are obtained by factorizing and reducing the ADI algorithm.

It is interesting to observe that if the transfer functions of two discrete-time models, $Y(z)$ and $\tilde{Y}(z)$, have moments (i.e., coefficients of z^{-i}) matched for $i = 0, \dots, m-1$, then by the definition of z -transform we have $y_k = \tilde{y}_k$, for $k = 0, \dots, m-1$, in the discrete-time domain. This implies that moment matching using a real rational Krylov subspace bears the meaning of waveform matching in the time-domain, rather than the traditional moment matching in the frequency domain. Strictly speaking, a discretized model is only an approximation of the continuous-time model. Nevertheless, by matching the waveform of an approximate dynamic model, the dominant subspace characteristics are still captured to certain degree. This new interpretation of a real shift parameter γ is expected to provide a heuristic way of choosing an appropriate γ for a reasonably good performance in the dominant subspace approximation. We emphasize that in the current interpretation the parameter γ is the inverse of the time step used for integration or discretization.

It is also interesting to note that the iteration formula for P_k in (37) can be viewed as a consequence of discretization of the differential Lyapunov equation

$$\frac{dP(t)}{dt} = AP + PA^T + BB^T$$

whose steady-state solution $P(\infty)$ is the solution of the Lyapunov equation (9). In the same spirit it makes sense to interpret γ as the sampling rate in the discretization.

Although the new interpretation does not give us an optimal choice of γ , it does provide an empirical guideline for choosing an appropriate γ . We recommend to choose a γ several magnitudes smaller than the magnitude of the fastest mode of the full-order model. In our experiment we observed that an overly small γ usually resulted in a performance similar to choosing $\gamma = 0$, i.e., moment matching at $s = 0$, while an exceedingly large γ ended up with very bad approximation because clustered sampling could not capture well the dynamic behavior of the waveform. Such observations are quite consistent to our theoretical analysis.

Remark 5: Another discretization is to approximate the s operator by

$$s \approx \frac{2(z-1)}{h(z+1)} = \frac{2\gamma(z-1)}{(z+1)} \quad (50)$$

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 that from the backward Euler discretization.

V. NUMERICAL EXPERIMENTS

In this section, we report comprehensive numerical results. We first introduce three typical measures in Section V-A for a complete experimental comparison in the subsequent two sections. In Section V-B we demonstrate several numerical results that are consistent with the theoretical predictions made in Section IV-C for the approximation performances of the three types of Krylov subspaces. Then in Section V-C numerical examples further demonstrate that the dominant subspace computation scheme with appropriately chosen Krylov subspaces is very effective for large-scale circuit models. In particular, we demonstrate the convergence effect when a sequence of different dimensions is chosen for intermediate subspaces. Finally, applications of approximately computed dominant subspaces to model-order reduction are reported in Section V-D. All the computations in this section were carried out using MATLAB 6 for the demonstration purpose. When MATLAB becomes inefficient for practical large-scale problems, the free software library SLICOT is recommended for large-scale numerical computation [34].

A. Three Measures for Comparison

We use three measures to compare how accurately the dominant subspace is approximated. The first measure is defined 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 determine whether or not the approximate dominant subspace is sufficiently accurate. Let σ_i and $\tilde{\sigma}_i$ be respectively the exact and approximate singular values for $i = 1, \dots, m$. The total relative error of singular value approximation is defined by

$$\sigma_{\text{err}} = \sum_{i=1}^m \frac{|\sigma_i - \tilde{\sigma}_i|}{\sigma_i} \quad (51)$$

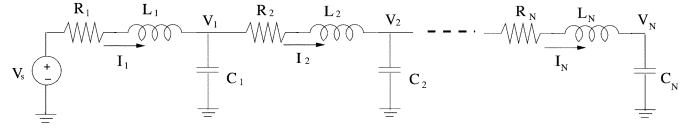


Fig. 1. RLC line.

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

The second measure is defined to compute the distance between the approximate subspace basis matrix $\tilde{U} \in \mathbb{R}^{n \times q}$ and an exact dominant subspace basis matrix $U \in \mathbb{R}^{n \times q}$ obtained from the SVD of P . The distance is defined by

$$\text{dist}(\tilde{U}, U) := \|\tilde{U} - U(U^T \tilde{U})\|_F \quad (52)$$

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

The third measure is defined to be the residual of the Lyapunov equation, i.e., $R(X)$ defined in (26), by substituting the approximate Gramian formed as in (24). To reduce the computational complexity, we adopt an idea presented in [23] for residual computation. Let $U_q \in \mathbb{R}^{n \times q}$ span the dominant subspace and $\tilde{P} = U_q P_q U_q^T$ be the approximate solution, where P_q is the solution of the reduced-order Lyapunov equation as in (22). Then, the residual can be expressed as

$$\begin{aligned} R(\tilde{P}) &= AU_q P_q U_q^T + U_q P_q U_q^T A^T + BB^T \\ &= [AU_q \quad U_q \quad B] \begin{bmatrix} 0 & P_q & 0 \\ P_q & 0 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} U_q^T A^T \\ U_q^T \\ B^T \end{bmatrix}. \end{aligned}$$

Let

$$QR = [AU_q \quad U_q \quad B]$$

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

$$\|R(\tilde{P})\| = \|R\Delta R^T\|$$

where

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

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

$$\varepsilon_{\text{res}} = \frac{\|R(\tilde{M})\|}{\|B\|^2}. \quad (53)$$

B. Comparison of Approximation Without Compaction

The first test case is an RLC line with N segments shown in Fig. 1, which could be a discretized transmission line model or an interconnect model. This example will be used to test the

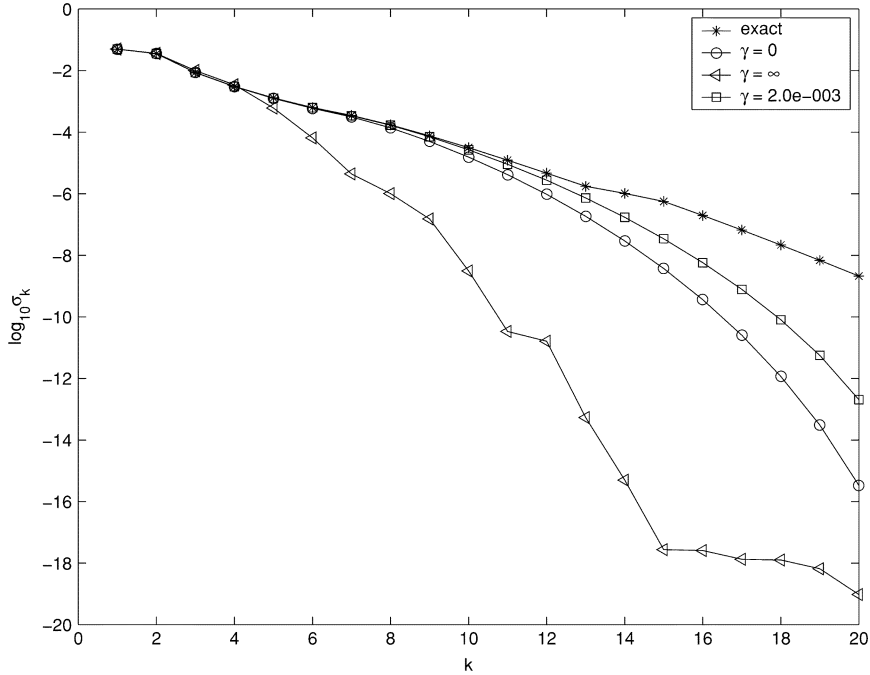


Fig. 2. Approximation of singular values by using $\gamma = \infty$ (\blacktriangleright), $\gamma = 0$ (\circ), and $\gamma = 2 \times 10^{-3}$ (\square).

approximation effects of the three Krylov subspaces, then the effectiveness of the dominant subspace computation scheme, and finally model-order reduction by using computed dominant subspaces. By the modified nodal analysis (MNA) formulation [35], we use $x = (V_1, \dots, V_N, I_1, \dots, I_N)^T \in \mathbb{R}^{2N}$ for the state vector and choose $u = V_s$ and $y = V_N$ to be the input and output, respectively.

This test circuit is first used to demonstrate different approximation performances of the three Krylov subspaces $\mathcal{K}_m((\gamma I - A)^{-1}, (\gamma I - A)^{-1}B)$, $\mathcal{K}_m(A^{-1}, A^{-1}B)$, and $\mathcal{K}_m(A, B)$. For easy identification, we identify the first Krylov subspace by the specific γ chosen, the second one by $\gamma = 0$, and the third one by $\gamma = \infty$. We chose a 200-stage *RLC* line so that the model order is 400, and the exact Gramian is solved by Bartels–Stewart algorithm [36]. For simplicity, we assume that the *RLC* values are uniform with $R = 10 \Omega$, $L_i = 1H$, and $C_i = 1F$ for all i . Although the state–space model is controllable in theory, the Gramian computed from MATLAB has only a rank 31, which means that the controllable space has prominent low-dimensional dominance and the full-order model has quite much redundancy.

We compute the approximate dominant controllable subspace by using the three candidate Krylov subspaces, all with order $q = 20$. To verify the different approximation effects as predicted by the two observations 1) and 2) in Section IV-C, the *compaction* procedure in the dominant subspace computation scheme is not used for this test case. The three measures introduced in the previous section are computed and listed in Table I. Shown in Fig. 2 is 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 $\gamma = 0$ is better than that at $\gamma = \infty$ by the measures of singular values and subspace distance. For this example, the best approximation is achieved by a rational Krylov subspace at $\gamma = 0.002$. The numerical re-

TABLE I
MEASURES BY USING THREE KRYLOV SUBSPACES

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

sult matches very well with our previous theoretical analysis. In terms of residual, the Krylov subspace at $\gamma = \infty$ looks the best, which indicates that the residual measure might not be reliable if used as the only measure for dominant subspace approximation. Note that one can also use the equivalent Lyapunov equation (27) to compute the residual, which in fact resulted in a better residual in our experiment for the case $\gamma = 0$. But to have a comparison on the same basis, we keep on using the Lyapunov equation (9) for the residual comparison.

Next, we choose another set of *RLC* values for the same circuit model, with $R = 20 \Omega$, $L = 1 \text{ nH}$, and $C = 20 \text{ pF}$ to change the model modes but without changing the model order. The test results are summarized in Table II and Fig. 3 where the *compaction* procedure is again not used. 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 $\gamma = \infty$, while the Krylov subspace at $\gamma = 3 \times 10^8$ performs superior among all. We observe that this set of circuit element values has driven the fastest mode to the gigahertz oscillation level, hence a larger γ is chosen here.

The preceding two test cases both demonstrate the consistency with the analytical results established in Section IV.C. Moreover, they show that all three Krylov subspaces can capture the leading several dominant singular values very well, but not for the trailing singular values. To improve the overall approximation accuracy, the *compaction* procedure in the *Dominant Subspace Computation Scheme* can be used.

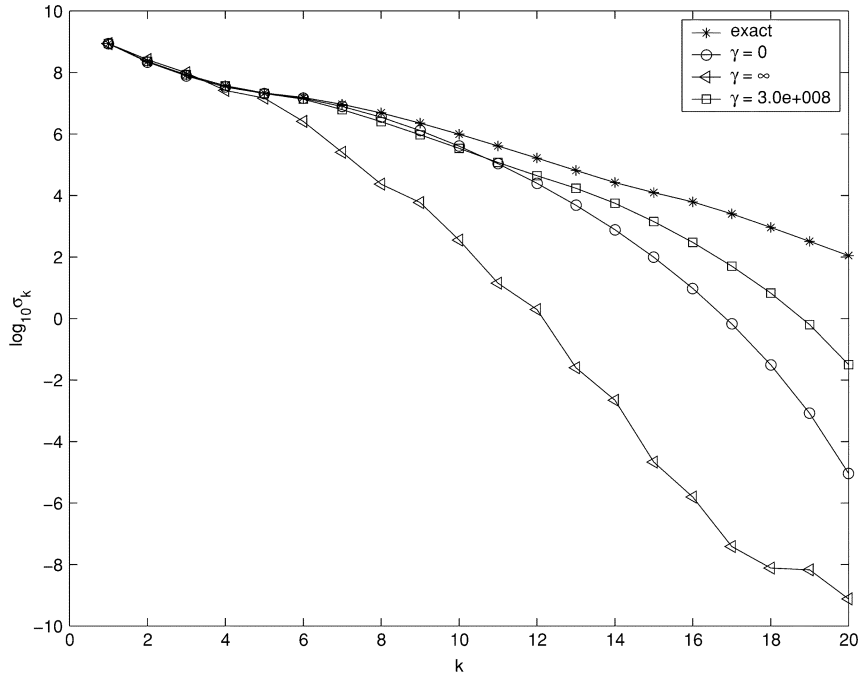


Fig. 3. Approximation of singular values by using $\gamma = \infty$ (\triangleright), $\gamma = 0$ (\circ), and $\gamma = 3 \times 10^8$ (\square).

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
$d(U, \bar{U})$	3.24	2.40	1.99
ϵ_{res}	0.066	0.321	0.020

C. Convergence of the Compaction Procedure

By the compaction procedure we need to choose an intermediate order m slightly larger than q , the dimension of the dominant subspace, then use the steps in the dominant subspace computation scheme to obtain a better dominance approximation. The test cases in this section are used to demonstrate the convergence effect by using a sequence of intermediate orders m .

The same circuit example in Fig. 1 is used again but with a new set of uniform RLC values $R = 30 \Omega$, $L = 0.1$ nH, and $C = 5$ pF. We also choose a larger model order with $n = 2000$. For such a large-scale model, solving the exact Gramian is not feasible. Hence, we compute the distance between two consecutive subspaces and the residual of Lyapunov equation as the measures for comparison. We also plot the computed singular values for visualization of the convergence. To test the convergence, we choose a sequence of m from 20 up to 100 with increment 20, denoted in MATLAB as $m = 20 : 20 : 100$. All intermediate subspaces are compacted to dimension $q = 20$. Listed in Table III are the distance measure and the residual measure for different m but with the same $\gamma = 10^8$ for the shift parameter. The notation $d(U, U_{pre})$ denotes the distance between the subspaces computed at two consecutive m 's, with U for the current m and U_{pre} for the previous m , where both U and U_{pre} are basis matrices in $\mathbb{R}^{n \times q}$. The initial subspace is assumed to be the

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

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
MEASURES AS A FUNCTION OF m FOR $\gamma = 0$

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}

TABLE V
MEASURES AS FUNCTION OF m FOR $\gamma = \infty$

m	20	40	60	80	100
$d(U, U_{pre})$	4.47	2.48	1.78	1.46	1.27
ϵ_{res}	0.0176	0.0064	0.0035	0.0023	0.0016

zero subspace. The data in Table III show that the dominant subspace has well converged. Shown in Fig. 4 is the convergence behavior of the computed singular values.

For comparison, the test results by using $\gamma = 0$ are shown in Table IV and Fig. 5. We see from the table that $d(U, U_{pre}) = 1.00$ at $m = 100$, which means the convergence is not as good as that of $\gamma = 10^8$.

Finally, the same example is tested again to see the performance of the Krylov subspace at $\gamma = \infty$. The results are shown in Table V and Fig. 6. The convergence of this option is clearly worse than the previous two, which again demonstrates the different approximation performances of the three Krylov subspaces as predicted.

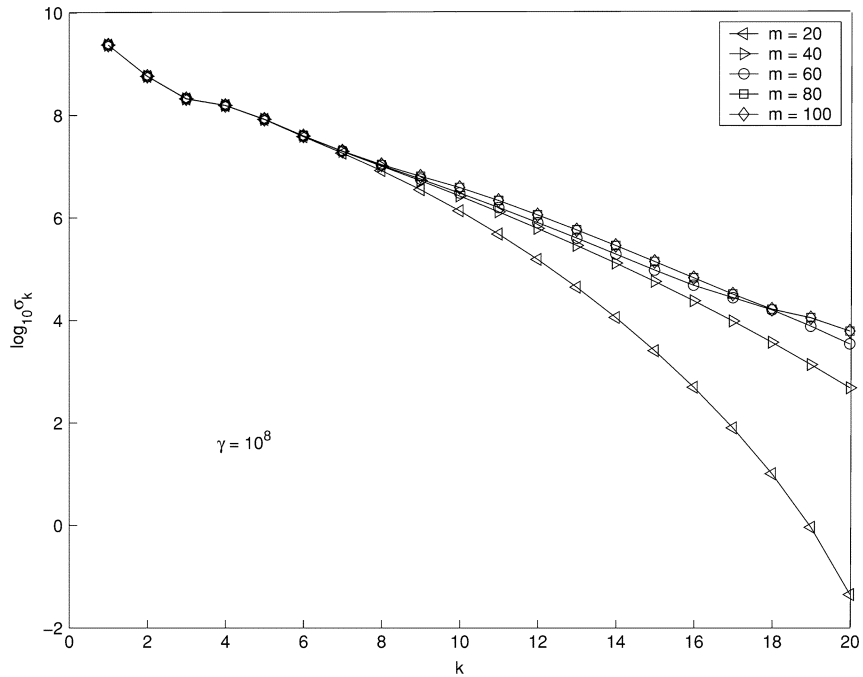


Fig. 4. Convergence of singular values with $\gamma = 10^8$ for $m = 20 : 20 : 100$.

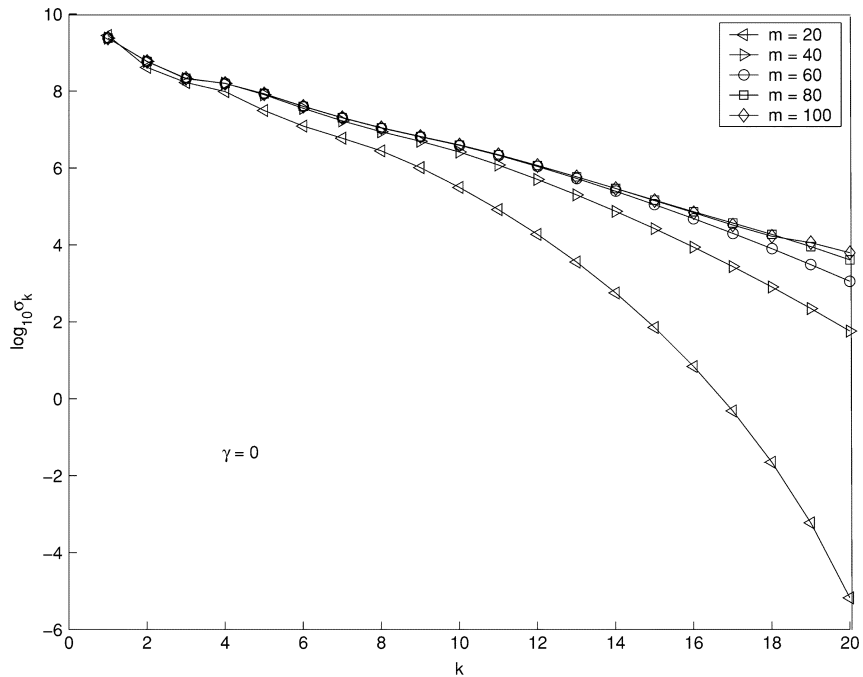


Fig. 5. Convergence of singular values with $\gamma = 0$ for $m = 20 : 20 : 100$.

To summarize, the dominant subspace approximated by the Krylov subspace $\mathcal{K}_m(A, B)$ is always the worst comparing to the other two options. For this reason, in the application to model-order reduction to be presented next, the Krylov subspace $\mathcal{K}_m(A, B)$ will not be used.

D. Application to Model-Order Reduction

The *RLC* circuit in Fig. 1 is now used for testing model-order reduction. This time node V_1 is chosen as the output and the uniform *RLC* values are $R = 20 \Omega$, $L = 1 \text{ nH}$, and $C = 20 \text{ pF}$. The

full model order is 1000 and the reduced model order is 20. For dominant subspace compaction, we first generate a Krylov subspace with an intermediate order 80 which is further reduced to order 20 by compaction. Shown in Fig. 7 is the reduction result by using the Krylov subspace at $\gamma = 0$ and the projection to the dominant *controllable* subspace V_q [i.e., choosing $W_q = V_q$ in (8)]. Shown in Fig. 8 is the reduction result by using the dominant *observable* subspace W_q also computed at $\gamma = 0$ [i.e., choosing $V_q = W_q$ in (8)]. The accuracies of the two reduced models are comparable. The reduction result becomes better if the oblique projection is performed with V_q com-

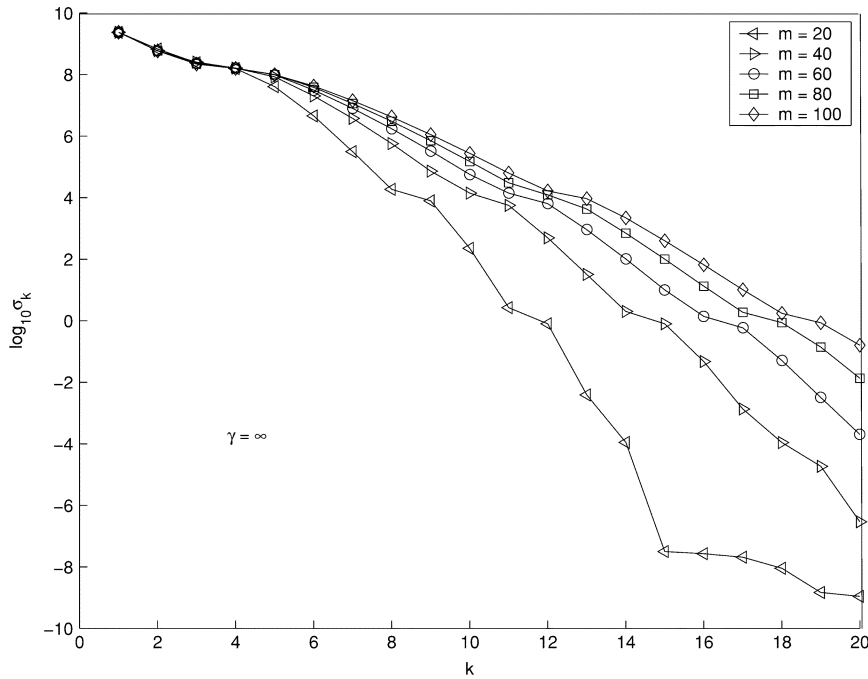


Fig. 6. Convergence of singular values with $\gamma = \infty$ for $m = 20 : 20 : 100$.

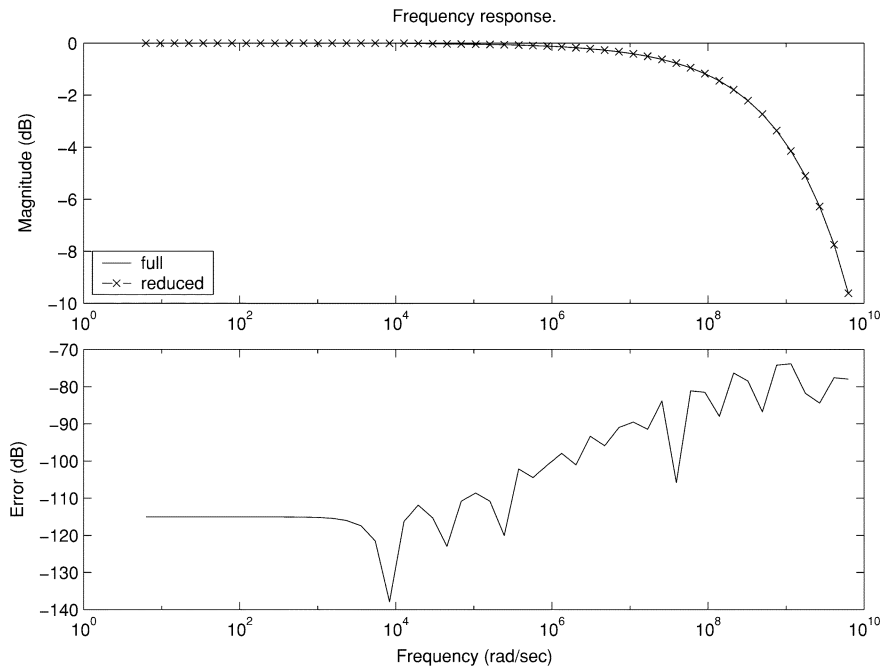


Fig. 7. Reduction of *RLC* line from 1000th order to 20th order using dominant controllable subspace computed at $\gamma = 0$.

puted from the pair (A, B) and W_q from the pair (A^T, C^T) , as shown in Fig. 9, where still $\gamma = 0$ is used. Note that the error in this case is significantly smaller than the previous two. For comparison, the reduction result by using moment matching up to the 20th order without compaction is shown in Fig. 10. We see that by matching moments only at the low frequencies the error, although is very small at the low frequency band, increases remarkably at the high-frequency. Comparatively, the errors from using dominant controllable/observable subspaces are fairly flat.

The second circuit example used for model-order reduction is the two coupled *RLC* lines shown in Fig. 11. For this circuit

$$x = (V_{11}, \dots, V_{1N}, V_{21}, \dots, V_{2N}, I_{11}, \dots, I_{1N}, I_{21}, \dots, I_{2N})^T \in \mathbb{R}^{4N}$$

is the state vector and $u = V_s$ is the input. The output will be specified later. The uniform *RLC* values are chosen as $R_{1i} = 10 \Omega, R_{2i} = 5 \Omega, L_{1i} = L_{2i} = 10.0 \text{ nH}, C_{1i} = C_{2i} = 1.0 \text{ pF}, CC_i = 20 \text{ pF}$ for $i = 1, \dots, N$, where N is the number

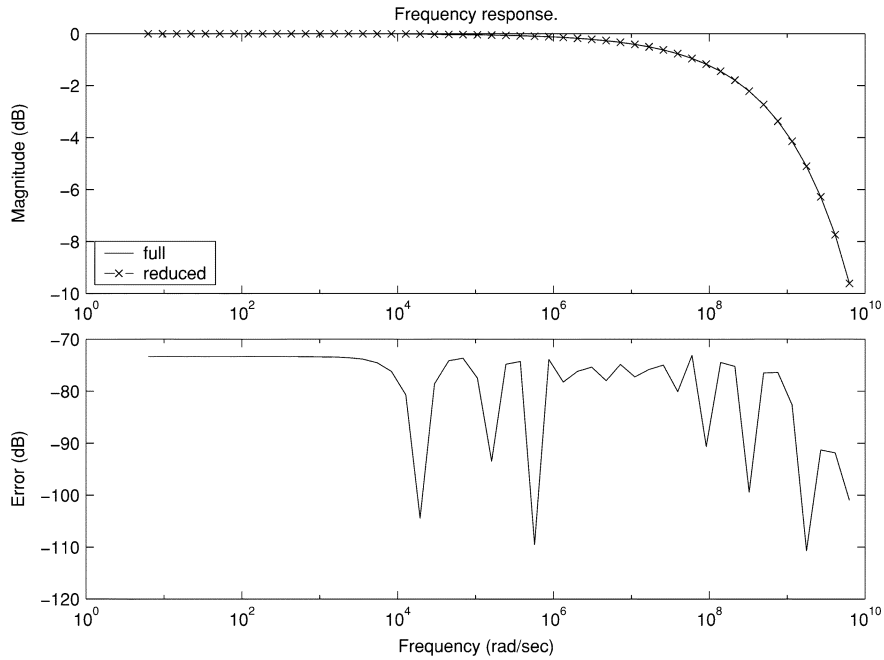


Fig. 8. Reduction of *RLC* line from 1000th order to 20th order, using dominant observable subspace computed at $\gamma = 0$.

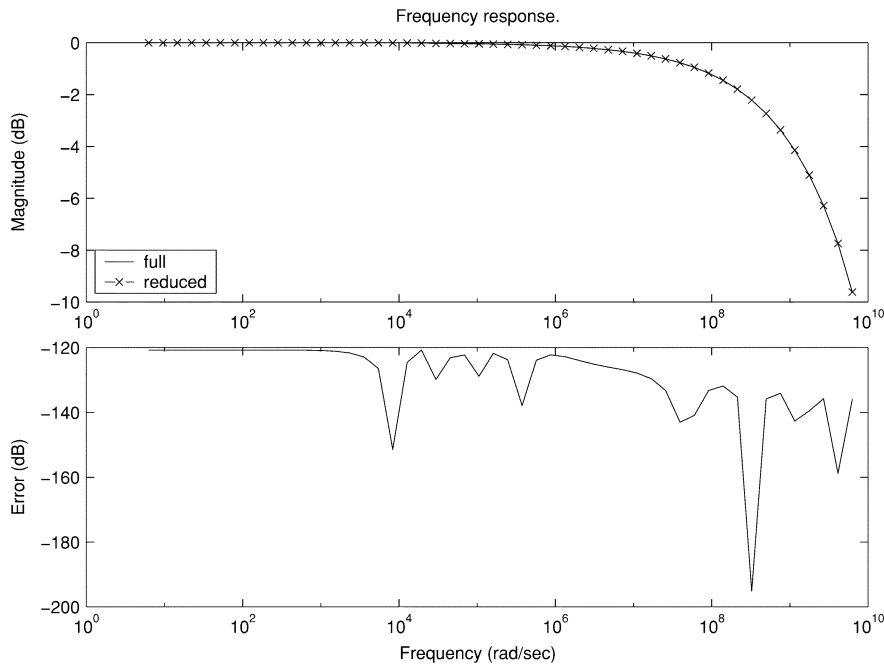


Fig. 9. Reduction of *RLC* line from 1000th order to 20th order using both dominant subspaces computed at $\gamma = 0$ (oblique projection).

of stages. We choose $N = 300$ stages so that the model is of 1200th order. This circuit is used to test the reduction effect by using V_q only, W_q only, and both V_q and W_q . The first two belong to orthogonal projection and the third belongs to oblique projection. It usually happens that if the circuit model is not in the port formulation, the passivity cannot be preserved even using the congruent transformation. This example demonstrates that in case the reduced matrix \hat{A} has unstable poles, removing those few unstable poles by further projecting \hat{A} to the stable subspace would not lose much accuracy for practical applications. Nevertheless, it is important to be aware that this simple technique only works for high-order models.

First we choose $y = V_{1N}$ as the output and compute the dominant controllable and observable subspaces using the dominant subspace computation scheme. We use a shift parameter $\gamma = 2 \times 10^8$ and choose $m = 100$ for the intermediate order and $q = 40$ for the reduced model order. The computed basis matrices for the controllable and observable subspaces are first used separately for model-order reduction and the results are shown in Figs. 12 and 13, respectively. Note that in the case of controllable subspace 8 poles of the reduced matrix are unstable, they are removed by further reducing the model to order 32 after projection to the stable subspace. Also, in the case of observable subspace, six poles of the reduced matrix are unstable, they are

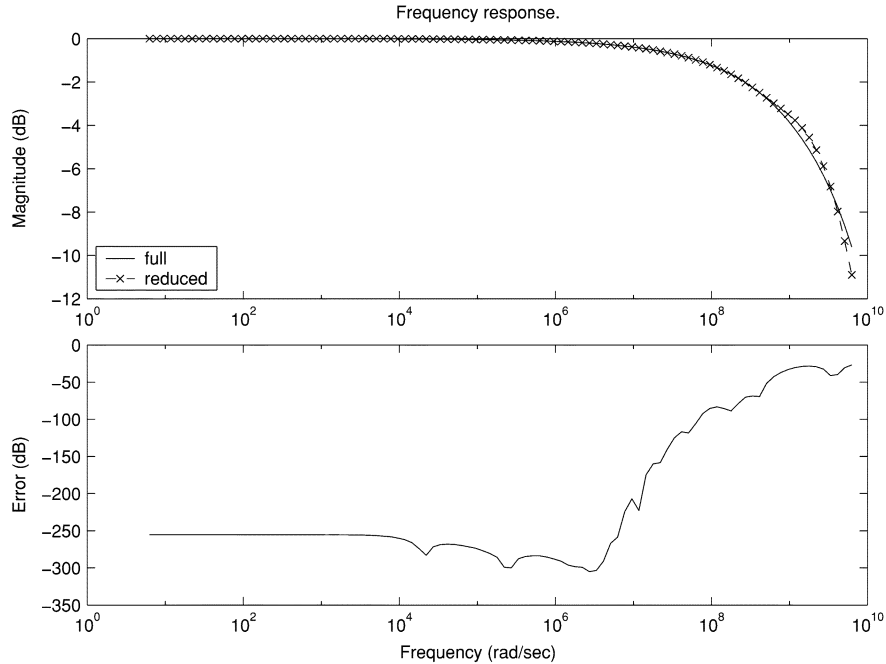


Fig. 10. Reduction of RLC line from 1000th order to 20th order by moment matching at $s = 0$ (without compaction).

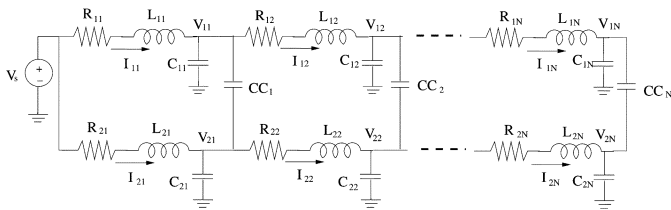


Fig. 11. Two coupled RLC lines.

removed by further reducing the model to order 34. The plots in Figs. 12 and 13 are results with the unstable poles removed, but the accuracy up to a high-frequency point remains very good. The low accuracy at the high-frequency part is probably due to the low reduced order, which in general can be improved by choosing a higher reduced order. Since for this example, the high-frequency part is the significantly rolloff band, the error at that band normally would not cause a serious problem for simulation not reaching that frequency band. Also, we see that both the reduced models by using V_q or W_q have comparable accuracy, implying that in practice there is no reason to favor one from the other for model-order reduction.

Since the outer product of the computed two basis matrices $W_q^T V_q$ is singular for this example, they cannot be used for oblique projection. This is usually caused by the fact that the input/output matrices B and C are inherently orthogonal. In such cases only the orthogonal projections as done above are feasible. However, if we choose a different output, say, $y = I_{11} + I_{21}$, then W_q and V_q have a nonsingular outer product, hence can be used for oblique projection. The reduction result is shown in Fig. 14. In this reduction, two out of the 40 poles are unstable. Hence the model is further reduced to order 38 by a projection to the stable subspace. We can see again from Fig. 14 that the reduction remains quite accurate.

VI. CONCLUSION

Large-scale models appearing in circuit simulation and other areas bring challenges to conventional model-order reduction techniques. To overcome the limitation of balanced truncation for large-scale models, approximate dominant subspaces have been used for practical model-order reduction with many good results. However, a theoretical error bound for unbalanced dominant subspace projection was unknown. This paper has established an \mathcal{L}_2 error bound for model-order reduction using unbalanced dominant subspace. Furthermore, it has investigated the performances of using three types of Krylov subspaces for approximate dominant subspace computation. It is analytically justified that the conventionally used Krylov subspace $\mathcal{K}_m(A, B)$ in fact is not a good candidate for effective dominant subspace computation. Rather, the Krylov subspace that matches moments at low frequencies and a rational Krylov subspace with an appropriately chosen real shift parameter are capable of producing superior approximate dominant subspaces. Numerical experiments have demonstrated that the theoretical analysis well predicts the computation results. Furthermore, numerical results have demonstrated that the approximate dominant subspaces computed from the dominant subspace computation scheme can be used effectively for large-scale model-order reduction.

APPENDIX SUPPLEMENTARY PROOF

Proof of Property 2: We only prove the Property for the single-input case. Its extension to the multiple-input case is straightforward.

The Arnoldi algorithm gives rise to the identity

$$A^{-1}V_m = V_m H + h v_{m+1} e_m^T \quad (54)$$

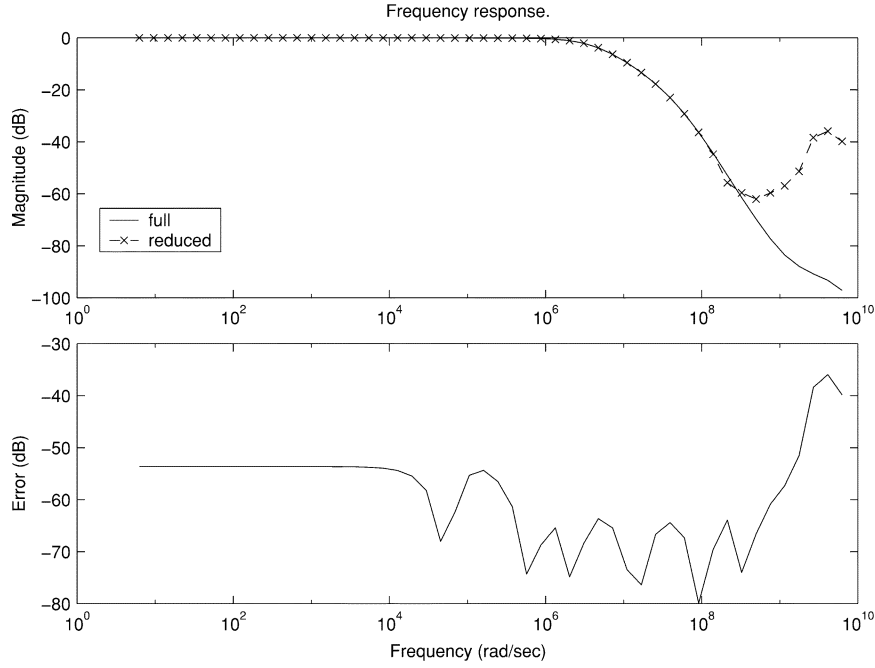


Fig. 12. Reduction of the coupled RLC line from 1200th order to 40th order using dominant controllable subspace computed at $\gamma = 2 \times 10^8$.

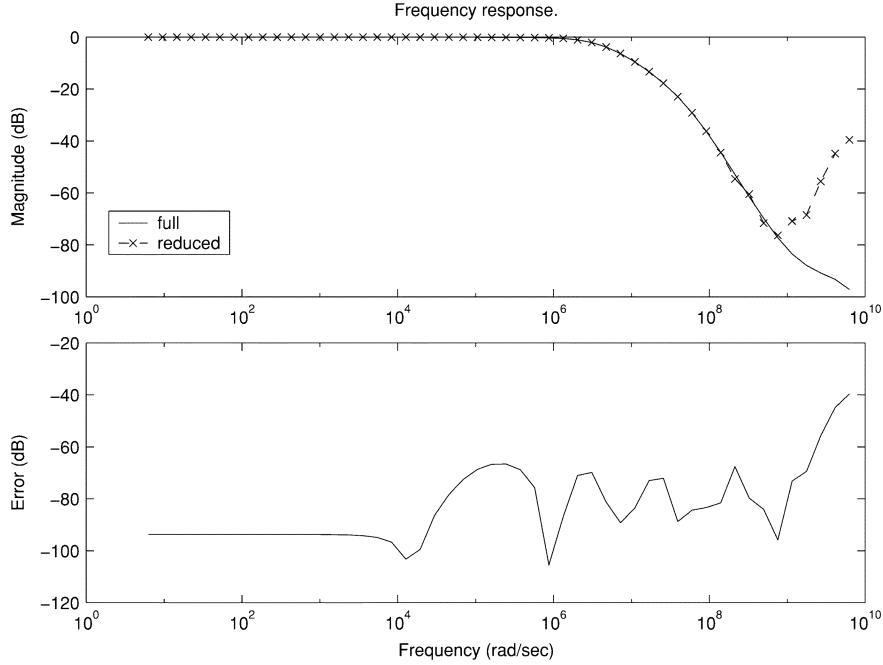


Fig. 13. Reduction of the coupled RLC line from 1200th order to 40th order using dominant observable subspace computed at $\gamma = 2 \times 10^8$.

where H is an upper Hessenberg matrix, 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 is upper Hessenberg, it is readily verified that

$$A^{-i}V_m e_1 = V_m H^i e_1, \quad \text{for } i = 1, \dots, m. \quad (55)$$

It also follows from (54) that

$$I_m = V_m^T A V_m H + h V_m^T A v_{m+1} e_m^T$$

i.e.,

$$\hat{A}^{-1} = H + h \hat{A}^{-1} V_m^T A v_{m+1} e_m^T. \quad (56)$$

One can show by induction that

$$\hat{A}^{-i} e_1 = H^i e_1, \quad \text{for } i = 1, \dots, m - 1. \quad (57)$$

Indeed, (56) implies that $\hat{A}^{-1} e_1 = H e_1$. Assume that (57) holds for some i satisfying $1 \leq i \leq m - 2$. Then

$$\begin{aligned} \hat{A}^{-(i+1)} e_1 &= \hat{A}^{-1} H^i e_1 \\ &= \left(H + h \hat{A}^{-1} V_m^T A v_{m+1} e_m^T \right) H^i e_1 \\ &= H^{i+1} e_1 \end{aligned}$$

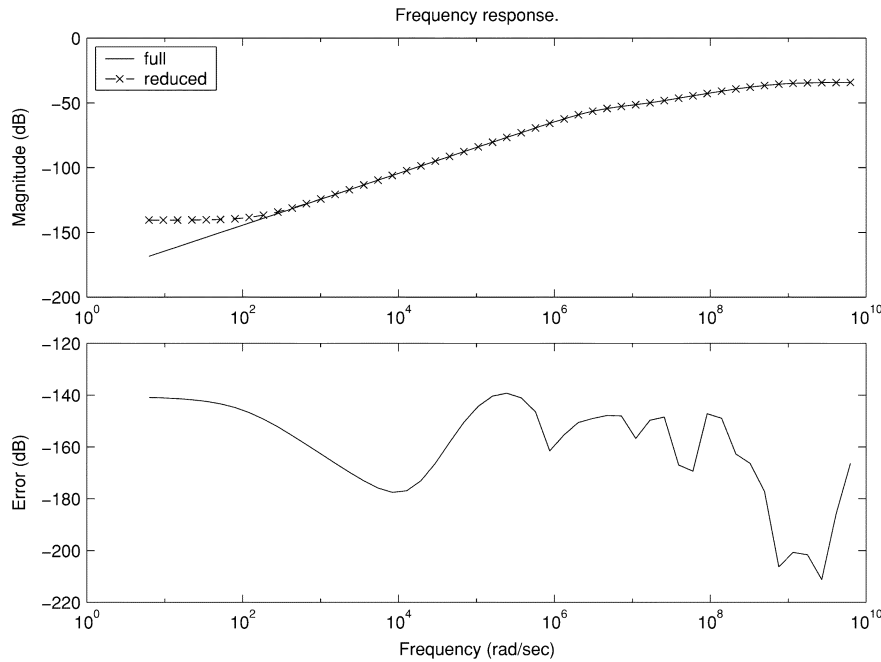


Fig. 14. Reduction of the coupled RLC line from 1200th order to 40th order using both dominant subspaces computed at $\gamma = 2 \times 10^8$ (oblique projection).

because of the fact that H is an upper Hessenberg matrix and $e_m^T H^i e_1 = 0$ for $1 \leq i \leq m-2$.

Let $\beta = \|A^{-1}B\|$. Then, $A^{-1}B = \beta V_m e_1$, i.e., $B = \beta A V_m e_1$, from which we obtain $\hat{B} = \beta \hat{A} e_1$, i.e.,

$$\hat{A}^{-1} \hat{B} = \beta e_1. \quad (58)$$

Thus, we have shown $A^{-1}B = V_m \hat{A}^{-1} \hat{B}$ which is the identity (33) for $i = 1$. Applying the identities (57), we obtain for $2 \leq i \leq m$

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

where we also used (54) and (58). Consequently, the identities (33) hold for $i = 1, \dots, m$. ■

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive comments, in particular, one reviewer for bringing [24] to their attention.

REFERENCES

- [1] A. Antoulas, D. 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*. Providence, RI: AMS Pub., 2001.
- [2] Z. Bai, "Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems," *Appl. Numer. Math.*, vol. 43, pp. 9–44, 2002.
- [3] S. Gugercin and A. Antoulas, "A comparative study of 7 algorithms for model reduction," in *Proc. 39th IEEE Conf. on Decision and Control*, Sydney, NSW, Australia, 2000, pp. 2367–2372.
- [4] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Contr.*, vol. AC-26, no. 1, pp. 17–32, Feb. 1981.
- [5] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L_∞ error bounds," *Int. J. Contr.*, vol. 39, no. 6, pp. 1115–1193, 1984.
- [6] M. Safonov and R. Chiang, "A Schur method for balanced-truncation model reduction," *IEEE Trans. Autom. Contr.*, vol. AC-34, no. 7, pp. 729–733, July 1989.
- [7] C. Vilemagne and R. Skelton, "Model reduction using a projection formulation," *Int. J. Control*, vol. 46, pp. 2141–2169, 1987.
- [8] L. Pillage and R. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. Computer-Aided Design*, vol. 9, no. 4, pp. 352–366, Apr. 1990.
- [9] P. Feldmann and R. Freund, "Efficient linear circuit analysis by Padé approximation via the Lanczos process," *IEEE Trans. Computer-Aided Design*, vol. 14, no. 5, pp. 639–649, May 1995.
- [10] A. Odabasioglu, M. Celik, and L. Pileggi, "PRIMA: Passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. Computer-Aided Design*, vol. 17, no. 8, pp. 645–654, Aug. 1998.
- [11] E. Grimme, D. Sorensen, and P. Dooren, "Model reduction of state space systems via an implicitly restarted Lanczos method," *Numer. Algor.*, vol. 12, pp. 1–31, 1996.
- [12] K. Gallivan and E. Grimme, "A rational Lanczos algorithm for model reduction," *Numer. Algor.*, vol. 12, pp. 33–63, 1996.
- [13] I. Jaimoukha and E. Kasenally, "Implicitly restarted Krylov subspace methods for stable partial realizations," *SIAM J. Matrix Anal. Appl.*, vol. 18, no. 3, pp. 633–652, 1997.
- [14] V. Balakrishnan, Q. Su, and C.-K. Koh, "Efficient balance-and-truncate model reduction for large-scale systems," in *Proc. 2001 IEEE American Control Conf.*, Arlington, VA, 2001, pp. 4746–4751.
- [15] T. Gudmundsson and A. Laub, "Approximate solution of large sparse Lyapunov equations," *IEEE Trans. Autom. Contr.*, vol. 39, no. 5, pp. 1110–1114, Oct. 1994.
- [16] 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 Conf.*, New Orleans, LA, 1999.
- [17] P. Rabiei and M. Pedram, "Model-order reduction of large circuits using balanced truncation," in *Proc. IEEE Asia South Pacific Design Automation Conf. (ASPAC)*, Feb. 1999, pp. 237–240.
- [18] S. Lall, J. Marsden, and S. Glavaski, "A subspace approach to balanced truncation for model reduction of nonlinear control systems," *Int. J. Robust Nonlinear Contr.*, vol. 12, no. 5, pp. 519–535, 2002.
- [19] A. Hodel and K. Poolla, "Heuristic approaches to the solution of very large sparse Lyapunov and algebraic Riccati equations," in *Proc. 27th IEEE Conf. Decision and Control*, Austin, TX, 1988, pp. 2217–2222.

- [20] Y. Saad, "Numerical solution of large Lyapunov equations," in *Signal Processing, Scattering, Operator Theory, and Numerical Methods*, M. Kaashoek, J. Schuppen, and A. Ran, Eds. Boston, MA, 1990, pp. 503–511.
- [21] I. Jaimoukha and E. Kasenally, "Krylov subspace methods for solving large Lyapunov equations," *SIAM J. Numer. Anal.*, vol. 31, no. 1, pp. 227–251, 1994.
- [22] A. Hodel, B. Tenison, and K. Poolla, "Numerical solution of the Lyapunov equation by approximate power iteration," *Linear Algebra Appl.*, vol. 236, pp. 205–230, 1996.
- [23] T. Penzl, "A cyclic low-rank Smith method for large sparse Lyapunov equations," *SIAM J. Sci. Comput.*, vol. 21, no. 4, pp. 1401–1418, 2000.
- [24] J.-R. Li and J. White, "Low rank solution of Lyapunov equations," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 1, pp. 260–280, 2002.
- [25] V. Klema and A. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Trans. Autom. Contr.*, vol. AC-25, no. 2, pp. 164–176, Apr. 1980.
- [26] D. Boley and G. Golub, "The Lanczos-Arnoldi algorithm and controllability," *Syst. Contr. Lett.*, vol. 4, pp. 317–324, 1984.
- [27] K. Zhou, *Essentials of Robust Control*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [28] Z. Bai, R. Stone, W. Smith, and Q. Ye, "Error bound for reduced system model by Padé approximation via the Lanczos process," *IEEE Trans. Computer-Aided Design*, vol. 18, no. 2, pp. 133–141, Feb. 1995.
- [29] E. Grimme, "Krylov Projection methods for model reduction," Ph.D. thesis, ECE Dept., Univ. of Illinois at Urbana-Champaign, 1997.
- [30] A. Lu and E. Wachspress, "Solution of Lyapunov equations by ADI iteration," *Comp. Math. Appl.*, vol. 21, no. 9, pp. 43–58, 1991.
- [31] A. Greenbaum, *Iterative Methods for Solving Linear Systems*. Philadelphia, PA: SIAM, 1997.
- [32] R. Smith, "Matrix equation $XA + BX = C$," *SIAM J. Appl. Math.*, vol. 16, pp. 198–201, 1968.
- [33] E. Chiprout and M. Nakhla, "Analysis of interconnect networks using complex frequency hopping (CFH)," *IEEE Trans. Computer-Aided Design*, vol. 14, no. 2, pp. 186–200, Feb. 1995.
- [34] A. Varga, "Model reduction software in the SLICOT library," in *Proc. IEEE Int. Symp. on Computer Aided Control System Design (CACSD)*, Anchorage, AK, 2000, pp. 183–198.
- [35] C.-W. Ho, A. E. Ruehli, and P. A. Brennan, "The modified nodal approach to network analysis," *IEEE Trans. Circuits Syst.*, vol. CAS-22, no. 6, pp. 504–509, 1975.
- [36] R. Bartels and G. Stewart, "Algorithm 432: Solution of the matrix equation $AX + XB = C$," *Commun. ACM*, vol. 15, pp. 820–826, 1972.



Guoyong Shi (S'99–M'02) received the Bachelor's degree in applied mathematics from Fudan University, Shanghai, China, the M.E. degree in electronics and information science from Kyoto Institute of Technology, Kyoto, Japan, and the Ph.D. degree in electrical engineering from Washington State University, Pullman, in 1987, 1997, and 2002, respectively.

He is currently a Research Associate in the Department of Electrical Engineering, University of Washington, Seattle. He has published over 30 technical articles on a variety of subjects including control theory, job-shop scheduling, neural networks, and circuit simulation. His current research interests include modeling of complex systems and modern techniques for very large-scale integrated circuit simulation.



C.-J. Richard Shi (M'91–SM'99) received the Ph.D. degree in computer science from the University of Waterloo, Waterloo, ON, Canada, in 1994.

He is currently a Professor in Electrical Engineering at the University of Washington, Seattle, which he joined in 1998. From 1994 to 1998, he worked at Analog, Rockwell Semiconductor Systems, and the University of Iowa. His research interests include several aspects of the computer-aided design and test of integrated circuits and systems, with particular emphasis on analog/mixed-signal

and deep-submicron circuit modeling, simulation and design automation.

Dr. Shi is a key contributor to IEEE std 1076.1-1999 (VHDL-AMS) standard for the description and simulation of mixed-signal circuits and systems. He founded IEEE International Workshop on Behavioral Modeling and Simulation (BMAS) in 1997, and has served on the technical program committees of several international conferences. He received the Best Paper Award from the 1999 IEEE/ACM Design Automation Conference, a Best Paper Award from the 1998 IEEE VLSI Test Symposium, a National Science Foundation CAREER Award (2000), and a Doctoral Prize from the Natural Science and Engineering Research Council of Canada (1995). He has been an Associate Editor, as well as a Guest Editor, of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: ANALOG AND DIGITAL SIGNAL PROCESSING. He is currently an Associate Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS.