UNIVERSITY OF CALIFORNIA
RIVERSIDE

Advanced Non-Krylov Subspace Model Order Reduction Techniques for Interconnect Circuits

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Boyuan Yan

December 2009

Dissertation Committee:

Dr. Sheldon X.-D. Tan, Chairperson
Dr. Yingbo Hua
Dr. Jianlin Liu

The Dissertation of Boyuan Yan is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Advanced Non-Krylov Subspace Model Order Reduction Techniques for Interconnect Circuits

by

Boyuan Yan

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, December 2009
Dr. Sheldon X.-D. Tan, Chairperson

Model order reduction (MOR) is an efficient technique to reduce the complexity of dynamical systems while producing a good approximation of the input and output behavior. Classical MOR approaches such as Krylov subspace and truncated balanced realization methods have been well developed in the areas of system, control, and applied math for general systems in state-space equations. In recent years, MOR techniques using Krylov subspace algorithm have been studied intensively in the field of electronic design automation (EDA) for interconnect analysis.

Interconnects in integrated circuits (IC) can be extracted as RLC circuits, which are described by a class of state-space equations with special structure properties such as symmetry, positive semi-definiteness and sparsity. As a result, to reduce the complexity of interconnect circuits, we can take advantage of the special structures to simplify the classical MOR methods. On the other hand, there are also some special requirements for interconnect reduction: scalability to large problems, passivity and structure preserving, and application to circuits with massive ports. In this thesis, we present several non-Krylov

subspace MOR techniques for interconnect analysis.

First, we present new methods based on classical balanced truncation for interconnect analysis: we generalize the simultaneous diagonalization algorithm for first-order balanced truncation to overcome the high computing costs; we also propose a passive second-order balanced truncation technique (and its fast version) to preserve both passivity and structure information inherent to circuit formulation.

Second, we propose two new methods to perform passive reduction: we present new algorithm based on the Caratheodory extension, which has a similar computational cost as the Krylov subspace based methods but ensures the passivity of reduced model without any restriction on the internal structure of state-space equation; we also propose the concept of conditional passivity and a method to generate frequency band-limited passive reduced models.

Finally, we work on long-standing problem of reducing interconnect circuits with massive ports. We propose a decentralized MOR scheme, where a multi-input multi-output (MIMO) system is decoupled into a number of subsystems in terms of outputs. The decoupling process and terminal reduction are based on the relative gain array (RGA), which measures the degree of interaction of each input-output pair. The reduction scheme can lead to passive reduction and is suitable for resistive coupling dominatant networks like power grids and substrate networks.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Circuit simulation tasks in the field of electronic design automation (EDA), such as the accurate prediction of interconnect effects at the board and chip level, or analog circuit analysis with full accounting of parasitic elements, may require the solution of large linear networks. These networks can become extremely large, especially when circuits are automatically extracted from layout, or contain models of distributed elements, such as transmission lines, ground planes, antennas, and other three-dimensional structures. The use of SPICE-like simulators would be inefficient or even prohibitive for such large problems.

Model order reduction (MOR) is an efficient technique to reduce the complexity of dynamical systems while producing a good approximation of the input and output behavior. Classical MOR approaches such as Krylov subspace and truncated balanced realization methods have been well developed in the areas of system, control, and applied math for general systems in state-space equations. In recent years, MOR techniques using Krylov

subspace algorithm have been studied intensively in the field of EDA for interconnect analysis.

Interconnects in integrated circuits (IC) can be extracted as RLC circuits, which are described by a class of state-space equations with special structure properties such as symmetry, positive semi-definiteness and sparsity. As a result, to reduce the complexity of interconnect circuits, we can take advantage of the special structures to simplify the classical MOR methods. On the other hand, there are also some special requirements for interconnect reduction: scalability to large problems, passivity and structure preserving, and application to circuits with massive ports.

From formulation point of view, MOR techniques in the field of EDA can be classified into first-order based methods (using modified nodal analysis, MNA) and second-order based methods (using nodal analysis, NA). Existing methods mainly project the original system onto a subspace. In terms of projection subspace, these approaches are divided into two broad categories, namely moment matching based methods (Krylov subspace methods) and balanced truncation based methods. In the former case, the system is projected onto a subspace to match dominant moments while in the latter case the system is projected onto a subspace both easily controllable and easily observable.

1.2 Krylov subspace methods and existing problems

Moment matching based approaches have been a great success in the past due to its efficiency and scalability [17, 50, 22, 40, 19, 53]. Due to the introduction of Krylov subspace [17, 50, 22], implicit moment matching can be performed in a projection framework with very good numerical stability. As a result, moment matching idea [45] can be applied

to very large-scale problems. Another reason for the success is that, when applied to a special class of systems (like interconnect circuits in MNA formulation), those methods can be modified to preserve passivity [40, 19, 53]. So the advantages of Krylov-subspace methods are as follows:

(1) Take advantage of sparsity: fast and applicable to very large-scale problems.

(2) Take advantage of structure: preserve passivity and reciprocity for circuits in certain formulations

While suitable for reduction of large-scale circuits, Krylov subspace methods have the following drawbacks:

(1) The reduced model is not as compact as desired

(2) Lack of passivity guarantees for general structure system

(3) Degrades as the number of inputs increases

The three problems have not been solved well, which motivated the study in this thesis: We propose two fast balanced truncation methods via double gramian approximation to mitigate problem (1). For problem (2), we present a positive-real rational interpolation based model reduction via Carathéodory extension. We also propose the concept of conditional passivity and a method to generate frequency band-limited passive reduced models. All existing MOR methods suffer from problem (3) more or less, which is inherent to the centralized formulation. In this study, we propose a decentralized framework to decompose a centralized system with massive ports.

1.3 Non-Krylov subspace solutions

1.3.1 Fast balanced truncation via double gramians approximation

While suitable for reduction of large-scale circuits, Krylov subspace methods may generate models not as compact as desired. Therefore, another approach, truncated balanced realization (TBR), or balanced truncation (BT), which has been well developed in the control community [38, 20, 30, 1], has been studied intensively recently [32, 33, 31, 42, 43, 49, 44, 54]. The classical balanced truncation approaches [38, 30] produce nearly optimal models with controllable a *priori* global error bound [20]. More recently, algorithms [42] based on positive-real balanced truncation were presented to compute guaranteed passive reduced models of controllable accuracy, which is highly appreciated by posing no constrains on the internal structure of the state-space. So the advantages of balanced truncation methods are as follows:

(1) Generate compact models with wideband accuracy

(2) Positive-real versions can preserve passivity for system in general formulation

However, those classical methods are too expensive to directly apply to large-scale problems due to the cubic cost to solve two Lyapunov equations or Riccati equations and the eigenvalue problem. The main drawbacks of the balanced truncation are as follows:

(1) Can not take advantage of sparsity and very expensive $O(n^3)$

(2) Can not take advantage of structure and preserve structure for circuits in certain formulations

(3) Degrades as the number of inputs increases (less sensitive than Krylov subspace method)

There has been significant effort devoted to mitigate this difficulty recently, which has

led to two classes of approximate balanced truncation methods. The first class is based on the approximate balancing by iterative low rank solution of Lyapunov equation [32, 33, 31, 58, 62, 54]. The second class is based on the low rank gramian approximation [43, 49, 44].

Both classes of methods seek to find the dominant subspace of one approximate gramian. For example, in PMTBR [43], the system is reduced by projecting onto the approximate dominant controllable subspace without explicitly computing the gramian, which can be obtained much more cheaply by using singular value decomposition (SVD). Although no rigorous global error bounds exist as the classical method, those methods often produce a better approximation over a wide frequency range than Krylov subspace methods at the similar cost. In addition, one gramian based methods can take advantage of structure to preserve passivity like Krylov subspace methods.

However, considering only one gramian can lead to large errors as both controllability and observability gramians and their corresponding subspaces can be quite different for general interconnect circuits as shown in [54]. Both classes of methods have difficulty in considering both gramians as this requires the eigendecomposition of the product of two gramians, which is a $n \times n$ matrix, where n is the size of the problem. Hence the resulting method will have the similar computational cost, which is $O(n^3)$, of solving Lyapunov equations in the standard balanced truncation method.

In this study, we propose two fast balanced truncation methods to consider both gramians at the cost similar to Krylov subspace methods. The first solution is proposed for general dynamic system, where the system is balanced in terms of two approximate first-order gramians as achieved in the classical balanced truncation methods. This method is very accurate but can not preserve passivity and structure. The second solution is proposed for circuits in second-order formulation, where the system is balanced in terms of two ap-

proximate second-order gramians. While less accurate, this method can preserve passivity, structure, reciprocity inherent to RLC circuits.

Solution 1: double first-order gramians approximation

The first solution can balance a first-order system in general structure in terms of two approximate gramians as achieved in the classical balanced truncation method. The proposed algorithm is based on a generalized SVD-based balancing scheme such that the dominant subspace of the approximate gramian product can be obtained in a very efficient way without explicitly forming the gramians. The novelty of the new method is that we can keep the similar computing costs of the single gramian method. Experimental results on a number of published benchmarks show that the proposed method is much more accurate than the single gramian method with similar computing costs.

Solution 2: double second-order gramians approximation

While positive-real balanced truncation methods [42] can preserve passivity without posing any constrains on the internal structure of the state-space, this generality is less appreciated for RLC reduction, where special internal structure is available to preserve passivity less expensively. Also less appreciated is that structure information inherent to RLCK circuits such as symmetry, positive semi-definiteness and sparsity, cannot be preserved.

As we know, a linear circuit can be equivalently formulated in the form of a first-order system or a second-order system. In fact, it is better to formulate an RLC circuit as a second-order system. One reason is that all matrices in NA are not only positive semi-definite but also symmetric, which makes it easy to preserve structure information inherent to RLC circuits like reciprocity [48]. Another reason is that while the inductance matrix in

MNA is usually very large and dense, the susceptance matrix in NA is diagonally dominant and can be sparsified by a simple truncation method without disrupting the positive definiteness [53]. However, in the past several years, while second-order moment-matching based approaches have been successfully developed from ENOR [48] to SAPOR [53], second-order TBR-type methods still remain an open problem. In this study,

In this study, we propose a novel model order reduction approach, SBPOR (Second-order Balanced truncation for Passive Order Reduction), which is the first second-order balanced truncation method proposed for passive reduction of RLCK circuits. By exploiting the special structure information in the circuit formulation, second-order Gramians are defined based on a symmetric first-order realization in descriptor from. As a result, SBPOR can perform the traditional balancing with passivity-preserving congruency transformation at the cost of solving one generalized Lyapunov equation. Owing to the second-order formulation, SBPOR also preserves the structure information inherent to RLCK circuits. We further propose, SOGA (Second-Order Gramian Approximation version of SBPOR), to mitigate high computational cost of solving Lyapunov equation. Experimental results demonstrate that SBPOR and SOGA are globally more accurate than the Krylov subspace based approaches.

1.3.2 Carathéodory extension and waveform shaping

Krylov subspace methods can no preserve passivity for general systems. In order to preserve passivity for system with arbitrary internal structure, the positive realness of the transfer function of reduced system should be enforced. Toward this goal, algorithms [42] based on positive-real balanced truncation were presented to compute guaranteed passive reduced models of controllable accuracy. However, those methods are too expensive to directly ap-

ply to large-scale problems due to the cubic cost.

In this study, we present a novel rational interpolation based reduction framework for reducing the dynamic systems described in any internal structure. The new method is based on the Carathéodory extension, which ensures the interpolating function is passive without any restriction on the circuit structure. It has similar moment matching properties and similar computational cost as the Krylov subspace based reduction methods. Experimental results demonstrate that the proposed method can be orders of magnitude faster than the positive-real balanced truncation approach for reducing general structure systems with comparable and even better results.

In addition, we propose a new approach to enforcing the passivity of a reduced system of general passive linear time invariant circuits. Instead of making the reduced models passive for infinite frequencies, the new method works on the signal waveform driving reduced models. It slightly shapes the waveforms of the signal such that the resulting signal spectra are band limited to the frequency range in which the reduced system is passive. As a result, the reduced models only need to be band-limited passive (also called conditionally passive), which can be achieved much easier than traditional passivity for a reduced system.

1.3.3 Decentralized framework

Unfortunately, the efficiency of model order reduction degrades as the number of ports increases. The reason for the degradation is fundamental and does not depend on any particular reduction algorithm [18]. For Krylov-subspace based algorithms, the cost associated with model computation is directly proportional to the number of inputs, i.e. to the number of columns in the transfer function matrix. For example, in the PRIMA algorithm [40], if only two (block) moments are to be matched at each port, and the network has 1000 ports,

the resulting reduced model will have 2000 states. Similarly, in the TBR algorithm, for systems with many inputs, many states may be needed because of the high dimension of the controllable subspace.

There has been significant effort devoted to mitigating this difficulty recently, which has led to two classes of approaches. The first class takes advantage of the information of input signal. An extended Krylov subspace (EKS) method was proposed [55], which constructs a transformation matrix based on the dynamics of the circuit as well as the source excitations. More recently, an approximated truncated balanced realizations (TBR) procedure was proposed [49, 44] to obtain compact reduced models by exploiting the correlation information of input signals. However, since the modeling process depends on the input signal, once the pattern of input signal has been changed, the model needs to be rebuilt. More important, they become less useful when input information is unavailable *a priori*. The second type of approaches [16, 18, 35, 34] are based on the SVD of matrix-valued transfer function. The pioneering work is SVDMOR method [16, 18], which exploits the fact that the matrix transfer function may be numerically low rank. However, since a full matrix-valued transfer function is rarely low rank, it is still hard to obtain a compact model.

The reason for the degradation is that existing approaches are based on a *centralized* framework, where each input-output pair is implicitly assumed to be equally interacted and the matrix-valued transfer function has to be assumed to be fully populated. In this paper, a *decentralized* model order reduction scheme is proposed, where a multi-input multi-output (MIMO) system is decoupled into a number of subsystems and each subsystem corresponds to one output and several dominant inputs. The decoupling process is based on the relative gain array (RGA), which measures the degree of interaction of each input-output pair. Our experimental results on a number of interconnect circuits show that most of the input-

output interactions are usually insignificant, which can lead to extremely compact models even for systems with massive ports. The reduction scheme is very amenable for parallel computing and localized simulation as each decoupled subsystem can be reduced and simulated independently.

1.4 Contribution of this thesis

The next Chapter provides an in-depth overview of the model reduction methods for linear dynamic system. The rest of the dissertation contains the following major contributions:

Chapter 3 and 4 are dedicated to mitigating the existing problem 1 of Krylov subspace methods. In the Chapter 3, we generalize the simultaneous diagonalization algorithm for first-order balanced truncation to overcome the high computing costs. In the Chapter 4, we propose a second-order balanced truncation technique (and its approximate version) to preserve both passivity and structure information inherent to circuit formulation.

Chapter 5 and 6 are dedicated to mitigating the existing problem 2 of Krylov subspace methods. In the Chapter 5, we present a MOR technique based on the Carathéodory extension, which has similar computational cost as the Krylov subspace based methods but ensures the passivity of reduced model without any restriction on the internal structure of state-space equation. In the Chapter 6, we propose the concept of conditional passivity and a method to generate band-limited passive reduced models.

A solution of existing problem 3 is presented in Chapter 7. To make the existing MOR methods applicable to interconnects with massive ports, a decentralized MOR scheme is proposed, where a multi-input multi-output (MIMO) system is decoupled into a number of subsystems in terms of outputs. The decoupling process is based on the relative gain array

(RGA), which measures the degree of interaction of each input-output pair. The reduction scheme is very amenable for localized simulation and parallel computing as each decoupled subsystem can be reduced and simulated independently.

Conclusion ends the dissertation.

Chapter 2

Model order reduction techniques for dynamical systems

2.1 Model order reduction in a nutshell

2.1.1 Dynamic system models

The behavior of linear time-invariant (LTI) systems in many engineering problems can be described by state-space equations in descriptor form (E, A, B, C, D)

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2.1}$$

where $E, A \in R^{n\times n}$, $B \in R^{n\times p}$, $C \in R^{p\times n}$, $D \in R^{p\times p}$, $x(t) \in R^n$, $u(t), y(t) \in R^p$. When E equals identity matrix I, the state-space equations become the standard form

(A, B, C, D)

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2.2}$$

In fact, many LTI systems can also be described by a set of second-order differential equations (M, D, K, B, C)

$$\begin{aligned} &M\ddot{q}(t) + D\dot{q}(t) + Kq(t) = Bu(t) \\ &y(t) = Cq(t) \end{aligned} \tag{2.3}$$

where $M, D, K \in R^{n\times n}$, $B \in R^{n\times p}$, $C \in R^{p\times n}$,$u(t), y(t) \in R^p$, $q(t) \in R^n$. The behavior of the systems can be completely characterized by the state-space equations. However, in some cases, we are only interested in the input-output behavior. In such cases, transfer functions are needed. The transfer function associated with the first-order system (2.1) in the Laplace domain is given by

$$H(s) = C(sE - A)^{-1}B \tag{2.4}$$

which becomes

$$H(s) = C(sI - A)^{-1}B \tag{2.5}$$

for a standard system (2.2). For the second-order model (2.3), the transfer function is given by

$$H(s) = C(Ms^2 + Ds + K)^{-1}B \tag{2.6}$$

2.1.2 Passivity

Passivity is an important property of many physical systems. Passive systems can not produce energy internally. When modeling passive systems, non-passive reduced models may generate unbounded responses in transient simulation. For linear dynamic system, passivity requires the transfer functions to be positive-real when the input and output signals are port voltages and currents. For scattering-parameter (s-parameter) systems, passivity requires bounded-real for s-parameter matrices.

Necessary and sufficient condition

The system is passive if and only if its transfer function $H(s)$ is positive real [6], which means

$$\begin{aligned} &\text{(1) } H(s) \text{ is analytic for } Re(s) > 0 \\ &\text{(2) } \overline{H(s)} = H(\overline{s}) \text{ for } s \in C \\ &\text{(3) } H(s) + H(s)^H \geq 0 \text{ for } Re(s) > 0 \end{aligned} \tag{2.7}$$

where $\overline{H}$ denotes complex conjugate, H^H denotes Hermitian (complex conjugate and transpose), and ≥ 0 denotes positive semi-definiteness in a matrix context.

Sufficient condition

Given a dynamic system model, if the system matrices are positive semi-definite and the input matrix and output matrix equal, then the state-space model is in a *passive form* [27,

40, 48]:

for first-order model (2.1), A, E are positive semi-definite and $B = C^T$

for second-order model (2.3), M, D, K are positive semi-definite and $B = C^T$ (2.8)

In such a passive form, the transfer function will be positive-real, which means the system is passive. This sufficient condition is important because RLCK circuits can be formulated into such a passive form. Since the passive form can be inherited by the reduced model via an orthogonal projection, passivity can be easily preserved [27, 40, 48, 19, 53, 44, 60].

2.1.3 Model order reduction

The complexity of the system can be characterized by the size, n, of the model. In electrical engineering, civil engineering, or aeronautics, the size, n, is often very, even prohibitively, large that many analysis and design problems can not be solved within a reasonable computing time. Model order reduction is the technique to solve this problem by constructing a reduced model $H_r(s)$ of size $r \ll n$

$$\begin{aligned} E_r\dot{x}(t) &= A_r x(t) + B_r u(t) \\ y(t) &= C_r x(t) + Du(t) \end{aligned} \tag{2.9}$$

where $E_r, A_r \in R^{r\times r}$, $B_r \in R^{r\times p}$, $C_r \in R^{p\times r}$, $D \in R^{p\times p}$, $x(t) \in R^r$, $u(t), y(t) \in R^p$, for the first-order model (2.1), or

$$\begin{aligned} &M_r\dot{q}(t) + D_r q(t) + K_r \int q(t) = B_r u(t) \\ &y(t) = C_r q(t) \end{aligned} \tag{2.10}$$

where $M_r, D_r, K_r \in R^{r\times r}$,$B_r \in R^{r\times p}$, $C_r \in R^{p\times r}$, $u(t), y(t) \in R^p$, $q(t) \in R^r$, for the second-order model (2.3).

Such a low order system will have approximately the same output y as the original system to the input u of interest. The transfer functions are often used as a metric for approximation. If $\|H(s)-H_r(s)\| < \epsilon$ in some appropriate norm, for some given allowable error ϵ and frequency range of interest s, the reduced model is accepted as accurate.

In addition, it is important to preserve system properties like stability and passivity in model order reduction. Given a passive system, we hope the reduced system is also passive. Otherwise, the reduced system may cause nonphysical behavior when it is simulated with other subsystems even if it is stable.

Currently, most model order reduction methods are projection based. Given two projection matrices $W, V \in R^{n\times r}$, for the first-order model (2.9), we have

$$E_r = W^T EV \quad A_r = W^T AV \quad B_r = W^T B \quad C_r = CV \tag{2.11}$$

and for the second-order model (2.10), we have

$$M_r = W^T MV \quad D_r = W^T DV \quad K_r = W^T KV \quad B_r = W^T B \quad C_r = CV \tag{2.12}$$

where W is the left projector and V is the right projector.

Typically, W and V span useful subspaces. Different choices of W and V will result in different model reduction approaches (Krylov subspace based methods, balanced truncation methods, etc). If $W \neq V$, the projection is an oblique (Petrov-Galerkin) projection. If $W = V$, the projection is an orthogonal (Galerkin) projection. Usually, oblique projection is better in terms of accuracy as both subspaces are used (e.g. PVL [17] and TBR [38]).

However, orthogonal projection is widely used in practice because it can be used to preserve important properties like passivity for systems with special state-space formulation (e.g.PRIMA [40] and PMTBR [44]).

2.2 Krylov subspace methods (moment-matching)

2.2.1 Moments

Given a state-space model in descriptor form (E, A, B, C, D) in (2.1), the transfer function moments $H^{(0)}(s_0), H^{(1)}(s_0), \ldots$, at the frequency point s_0 are defined as terms in the Taylor series of the transfer function $H(s)$ near the point s_0

$$H(s) = H^{(0)}(s_0) + H^{(1)}(s_0)(s - s_0) + H^{(2)}(s_0)(s - s_0)^2 + \ldots \tag{2.13}$$

The moments are directly related to the matrices of derivatives of the transfer function:

$$H^{(k)}(s_0) = \frac{1}{k!}\frac{d^k}{ds^k}H(s)|_{s=s_0} \tag{2.14}$$

and for the state-space model (E, A, B, C, D) in (2.1), we can take the derivative of transfer function

$$H^{(k)}(s_0) = C((A - s_0E)^{-1}E)^k(A - s_0E)^{-1}B \tag{2.15}$$

2.2.2 Two-sided Krylov subspace method

The Krylov subspace $K_m(\mathcal{A}, \mathcal{R})$ generated by a matrix $\mathcal{A}$ and matrix $\mathcal{R}$, of order m, is the space spanned by the set of vectors ($\mathcal{R}, \mathcal{AR}, \mathcal{A}^2\mathcal{R}, \ldots, \mathcal{A}^{m-1}\mathcal{R}$). Given

$$K_m((s_0E - A)^{-1}E, (s_0E - A)^{-1}B) \subseteq colsp(V) \tag{2.16}$$

and

$$K_m((s_0E - A)^{-T}E^T, (s_0E - A)^{-T}C^T) \subseteq colsp(W) \tag{2.17}$$

, if the reduced system $H_r(s)$ is obtained by an oblique projection $(W^TEV, W^TAV, W^TB, CV, D)$, then we have

$$H^{(k)}(s_0) = H_r^{(k)}(s_0), k = 0, \ldots, 2m \tag{2.18}$$

where $H_r(s_0)$ denotes mth moments of the transfer function of the reduced system. This method is called two-sided Krylov subspace method and W and V can be constructed by numerical algorithms like Lanczos [17] or two-sided Arnoldli.

2.2.3 One-sided Krylov subspace method

Another implementation is to use only one Krylov subspace V and the reduced system $H_r(s)$ is obtained by an orthogonal projection $(W^TEV, W^TAV, W^TB, CV, D)$. Then we have

$$H^{(k)}(s_0) = H_r^{(k)}(s_0), k = 0, \ldots, m \tag{2.19}$$

This method is called one-sided Krylov subspace method and V can be constructed by Arnoldli algorithm. Compared with two sided case, one sided Krylov subspace method is

less accurate. However, if the state-space model is in a passive form [27, 40], one sided Krylov subspace can be relied to preserve passivity.

2.3 Balanced truncation methods

In this section, we review the classical balanced truncation methods developed in the control community for general dynamic systems.

2.3.1 Lyapunov balancing

Lyapunov balancing was introduced to the system and control society by [38]. Given a stable minimal linear time invariant (LTI) system in standard state-space form (A, B, C, D) in (2.2), the controllability gramian X and observability gramian Y are as follows

$$X = \int_0^\infty e^{A\tau} B B^T e^{A^T \tau} d\tau \quad Y = \int_0^\infty e^{A^T \tau} C^T C e^{A\tau} d\tau \tag{2.20}$$

It is easy to verify that they are the unique symmetric positive definite solutions to the Lyapunov equations

$$\begin{aligned} AX + XA^T + BB^T &= 0 \\ A^T Y + YA + C^T C &= 0 \end{aligned} \tag{2.21}$$

The controllability and observability gramians X and Y are related to the energy demanded to control and observe the system.

Controllability

Given any state x_0 at $t = 0$, if the system is controllable, there is a signal $u(t)$ with the smallest energy (measured by L_2 norm)

$$\|u(t)\|_2 = \sqrt{\int_{-\infty}^{0} u^T(t)u(t)dt} \tag{2.22}$$

which could drive the system from zero initial condition at $t = -\infty$ to x_0. Assuming $x(-\infty) = 0$, the zero-state response is

$$x(t) = \int_{-\infty}^{t} e^{A(t-\tau)}u(\tau)d\tau \tag{2.23}$$

The controllability gramian X is connected to the solution of the minimum L_2 norm problem

$$\begin{aligned} &min_{u\in L_2[-\infty,0]}\|u(t)\|_2^2 \\ &\text{subject to } x(0) = \textstyle\int_{-\infty}^{0} e^{-A\tau}u(\tau)d\tau = x_0 \end{aligned} \tag{2.24}$$

The solution to this problem is

$$u(t) = B^T e^{-A^T t}\left(\int_{-\infty}^{0} e^{-A\tau}BB^T e^{-A^T\tau}d\tau\right)^{-1} x_0 = B^T e^{-A^T t}X^{-1}x_0 \tag{2.25}$$

So the minimal energy needed to reach x_0 is

$$\|u(t)\|_2^2 = x_0^T X^{-1}x_0 \tag{2.26}$$

Now the optimization problem becomes a quadratic form, which means the size of the eigenvalues of X describes how much input energy is needed to control the associated state

eigenvector. In other words, if x_0 is picked as one of the eigenvectors of X, the energy needed in the input will be exactly the inverse of the corresponding eigenvalue. As a result, the largest eigenvalue will correspond to the state easiest to control.

Observability

Observability shares the similar definition of controllability. Given any state x_0 at $t = 0$, we want to observe how much energy (measured by L_2 norm) there will be from the output signal if the system is released from x_0 with zero input for $t \geq 0$. The observability gramian Y is related to the solution of the maximum L_2 norm problem

$$\begin{aligned} &max_{y \in L_2[0,+\infty]} \|y(t)\|_2^2 \\ &\text{subject to } x(0) = x_0 \end{aligned} \tag{2.27}$$

The zero-input response is

$$y(t) = Cx(t) = Cx(0)e^{At} \tag{2.28}$$

The L_2 norm of the output signal when the system is released from x_0 is

$$\|y(t)\|_2^2 = x_0^T \left(\int_0^\infty e^{A^T t} C^T C e^{At} dt \right) x_0 = x_0^T Y x_0 \tag{2.29}$$

which means the size of the eigenvalues of Y describes how much output energy is produced when the associated state eigenvector is in free evolution. In other words, if x_0 is picked as one of the eigenvectors of Y, the energy observed in the output will be exactly the corresponding eigenvalue. As a result, the largest eigenvalue will correspond to the state easiest to observe.

Balanced truncation

Given a dynamic system, the state-space representation is not unique. Any nonsingular linear transformation $x = T\tilde{x}$ can be applied to the system (A, B, C, D) to obtain a new state-space representation $(\tilde{A}, \tilde{B}, \tilde{C}, D)$

$$\begin{aligned} \dot{\tilde{x}}(t) &= \tilde{A}\tilde{x}(t) + \tilde{B}u(t) \\ y(t) &= \tilde{C}\tilde{x}(t) + Du(t) \end{aligned} \tag{2.30}$$

where

$$\tilde{A} = T^{-1}AT \quad \tilde{B} = T^{-1}B \quad \tilde{C} = CT \tag{2.31}$$

Such a transformation is known as a similarity transformation, which does not change the input-output behavior of the system. It is easy to see both representations (A, B, C, D) and $(\tilde{A}, \tilde{B}, \tilde{C}, D)$ have the same transfer function $H(s)$.

A *balanced* realization is a special state-space representation, where the controllability and observability gramians are diagonal and equal. The balancing transformation can be computed by calculating the eigenmodes of the gramian product XY

$$XY = T\Lambda T^{-1} \tag{2.32}$$

It can be seen that the eigenvetors of XY are the basis vectors that describe the balancing transformation as follows. From (2.20) and (2.31), we obtain the following expressions for the gramians of the transformed system

$$\tilde{X} = T^{-1}XT^{-T}; \quad \tilde{Y} = T^{T}YT \tag{2.33}$$

For a balanced system, we require $\tilde{X} = \tilde{Y} = \Sigma$, where Σ is a diagonal matrix. From (2.33), we can write

$$T^{-1}X = \Sigma T^{T}; \quad YT = T^{-T}\Sigma \tag{2.34}$$

or

$$T^{-1}XYT = \Sigma^{2} \tag{2.35}$$

which means the transformation T, which balances the system, contains the eigenvectors of the gramian product XY as its columns.

From the gramian expression (2.32) and (2.35), it can be seen that the the eigenvalues λ_i contained in the diagonal matrix Λ are positive real numbers, and $\sigma_i = \sqrt{\lambda_i}$ are known as the Hankel singular values of the system. The eigenvectors of XY correspond to states through which the input is transmitted to the output. The magnitudes of the Hankel singular values describe the relative importance of these states and are independent of the particular realization of the system. States corresponding to the small Hankel singular values are difficult to control and difficult to observe. Such states are less involved in the energy transfer from inputs to outputs.

Therefore, a general idea of balanced truncation is to transform the system into a balanced form $(\tilde{A}, \tilde{B}, \tilde{C}, D)$, where the states which are difficult to control are also difficult to observe, and to discard the parts of the dynamics that correspond to those week states. We may partition Σ into

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \tag{2.36}$$

and conformally partition the transformed matrices as

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}; \quad \tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}; \quad \tilde{C} = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix} \tag{2.37}$$

The reduced model of order r (A_r, B_r, C_r, D) is obtained by taking the $r \times r$, $r \times p$, $q \times r$ leading blocks of $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, respectively

$$A_r = \tilde{A}_{11} \quad B_r = \tilde{B}_1 \quad C_r = \tilde{C}_1 \tag{2.38}$$

This truncation leads to a balanced reduced-order system (A_r, B_r, C_r, D).

The most desirable property of TBR method is that the reduced model has bounded error throughout all the frequency domain. Specifically, the error in the transfer function of the order r approximation is bounded by [20]

$$\|H(s) - H_r(s)\| \leq 2 \sum_{i=r+1}^{n} \sigma_k \tag{2.39}$$

To summarize, the standard balanced truncation algorithm flow chart [30] is shown in Fig. 2.1. An approach with improved numerical properties may be found in [47].

Balancing in descriptor form

Given a state-space model in descriptor form (E, A, B, C, D) in (2.1), if the matrix E is singular, the system may not be proper. In this case, there are infinite eigenvalues and the transfer function can be represented as a sum of proper transfer function and a matrix of

ALGORITHM: STANDARD BALANCED TRUNCATION METHOD
Input: $H : (A, B, C, D)$
Output: $H_r : (A_r, B_r, C_r, D)$

1. Compute $X > 0$ and $Y > 0$
2. Cholesky factorization $X = L_x L_x^T$ and $Y = L_y L_y^T$
3. Compute SVD $U\Sigma V = L_y^T L_x$
4. Compute $T = L_x V \Sigma^{-1/2}$ and $T^{-1} = \Sigma^{-1/2} U^T L_y^T$
5. Compute the balanced realizations $\tilde{A} = T^{-1}AT$,$\tilde{B} = T^{-1}B$,$\tilde{C} = CT$.
6. Truncate to form the reduced system (A_r, B_r, C_r, D)

Figure 2.1: Balanced truncation algorithm.

polynomials

$$G(s) = G_p(s) + \sum_{i>0} G_i s^i \tag{2.40}$$

where $G_p(s)$ is a matrix of proper rational functions of s. The proper and polynomial parts of the transfer function can be separated by the projection of the system onto deflating subspaces of the pair (E, A) corresponding to finite and infinite eigenvalues, respectively [52]. The polynomial terms should be exactly preserved by the reduced system and the proper rational term $G_p(s)$, where E is nonsingular, can be reduced by classical balanced truncation.

Given the system in descriptor form with nonsingular E, controllability and observability gramians [54] can be computed by solving generalized Lyapunov equations

$$\begin{aligned} EXA^T + AXE^T + BB^T &= 0 \\ E^TYA + A^TYE + C^TC &= 0 \end{aligned} \tag{2.41}$$

Similarly, we want to find a matrix T to perform a similarity transformation to diagonalize the product XY. After the similarity transformation, the system is balanced. We may conformally partition the transformed matrices as

$$\begin{aligned} T^{-1}ET &= \begin{bmatrix} \tilde{E}_{11} & \tilde{E}_{12} \\ \tilde{E}_{21} & \tilde{E}_{22} \end{bmatrix} \quad T^{-1}AT = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix} \\ T^{-1}B &= \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix} \qquad\qquad CT = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix} \end{aligned} \tag{2.42}$$

and $(\tilde{E}_{11}, \tilde{A}_{11}, \tilde{B}_1, \tilde{C}_1, D)$ is the reduced order system. In fact, this reduction is mathematically equivalent to performing balanced truncation on the system $(E^{-1}A, E^{-1}B, C, D)$. However, the computation steps are numerically better conditioned via generalized Lyapunov equations.

Balancing from a projection point of view

Note that, an alternative interpretation of balanced truncation is to project the system onto a subspace both easily controllable and easily observable, which is just the dominant eigenspace of the matrix XY corresponding to the r largest eigenvalues. If we partition T and T^{-1} as

$$T^{-1} = \begin{bmatrix} W^T \\ \tilde{W}^T \end{bmatrix}; T = \begin{bmatrix} V & \tilde{V} \end{bmatrix} \tag{2.43}$$

and substitute (2.43) into (2.42), then we have

$$\tilde{E}_{11} = W^T EV; \quad \tilde{A}_{11} = W^T AV; \quad \tilde{B}_1 = W^T B; \quad \tilde{C}_1 = CV \tag{2.44}$$

which unifies the balancing and truncating operations into one projection step. Since the left projector W and right projector V are not equal generally $W \neq V$, the projection is an oblique projection.

Balancing a symmetric system

Given a state-space model in descriptor form (E, A, B, C, D) in (2.1), the state-space model is *symmetric* if

$$E = A^T \quad E = E^T \quad B = C^T \tag{2.45}$$

In this symmetric case, both Lyapunov equations (2.41) are the same and both gramians are equivalent $Y = X$. Since the gramian X is symmetric, it is orthogonally diagonalizable, i.e., there exists $T^{-1} = T^T$ such that $T^T X T = \Sigma$. Then, we have

$$T^T X X T = T^T X T T^T X T = \Sigma^2 \tag{2.46}$$

which means, in this symmetric case, the eigenspace of gramian product XX is exactly the eigenspace of each gramian X. In this case, we only need to project onto the dominant eigenspace of one gramian. Since either gramian is symmetric, the left projector W and right projector V are equal, $W = V$ and the projection (2.44) becomes an orthogonal projection

$$\tilde{E}_{11} = V^T E V; \quad \tilde{A}_{11} = V^T A V; \quad \tilde{B}_1 = V^T B; \quad \tilde{C}_1 = C V \tag{2.47}$$

2.3.2 Riccati balancing

Lyapunov balancing preserves the stability of the system, but passivity might not be preserved. To keep the passivity properties of a system, Riccati balancing [13] is needed. If a system (A, B, C, D) is positive real (passive), it will satisfy the positive real (PR) equations [21]

$$\begin{aligned} AP + PA^T &= -B_l B_l^T \\ PC^T - B &= -B_l D_l^T \\ -D - D^T &= -D_l D_l^T \end{aligned} \tag{2.48}$$

where $P = P^T > 0$. A dual pair of positive real equations are as follows

$$\begin{aligned} A^T Q + QA &= -C_r^T C_r \\ QB - C^T &= -C_r^T D_r \\ -D - D^T &= -D_r^T D_r \end{aligned} \tag{2.49}$$

where $Q = Q^T > 0$. The above equations can be rewritten as a dual pair of Riccati equations, and then solved for P and Q

$$\begin{aligned} AP + PA^T + (PC^T - B)(D + D^T)^{-1}(CP - B^T) = 0 \\ A^T Q + QA + (QB - C^T)(D + D^T)^{-1}(B^T Q - C) = 0 \end{aligned} \tag{2.50}$$

Riccati balancing can now be achieved by substituting (P, Q) with (X, Y) in the balanced truncation algorithm. Since the reduced system also satisfies the (PR) equations, passivity is preserved. Riccati balancing has been applied to interconnect reduction in the positive-real TBR (PR-TBR) method [42].

Similar to Lyapunov balancing, Riccati balancing also has physical interpretations in

terms of energy. Let $s(u(t), y(t))$ be the supply function, which describes the rate at which power is supplied to the system and typically is defined such that $s(u(t), y(t)) > 0$ implies a positive amount of energy input, while $s(u(t), y(t)) < 0$ means energy is extracted from the system back to the environment. When the system inputs and outputs are currents or voltages, i.e., when the system transfer function represents impedance or admittance matrices, we may use the supply function $s(u(t), y(t)) = u(t)^T y(t)$.

The input energy gramian P is associated with the following optimization problem

$$inf \left(\int_{-\infty}^{0} s(u(t), y(t)) dt \right) = x_0^T P^{-1} x_0 \tag{2.51}$$

which minimize the amount of energy that must be injected into the system, in order to control the system to state x_0 at time 0. In this setting, the sizes of the eigenvalues of R describe how much energy is needed to control the associated state eigenvector. Small eigenvalues of P implies that a large amount of energy is needed to reach the associated mode.

Similarly, the output energy gramian Q is associated with the following optimization problem

$$sup \left(- \int_{0}^{\infty} s(u(t), y(t)) dt \right) = x_0^T Q x_0 \tag{2.52}$$

which maximize the amount of energy which can be extracted from the system in free evolution from x_0 at time 0. Also, the sizes of the eigenvalues of Q describe how much energy can be extracted from the system in free evolution. Small eigenvalues of Q implies that a small amount of energy can be extracted from the associated model.

For the positive-real case, the error bound is given by [10]

$$\|H(s) - H_r(s)\| \leq \lambda_{max}(D + D^T) \sum_{i=r+1}^{n} \frac{2\sigma_k}{(1-\sigma_k)^2}(1 + \sum_{j=1}^{k} \frac{2\sigma_j}{1-\sigma_j})^2 \tag{2.53}$$

2.3.3 Second-order balancing

Consider a second-order LTI stable system (M, D, K, B, C) in (2.3) with M assumed to be nonsingular, the general idea of reducing the second-order system is to transform the second-order system into the equivalent first-order system, from which the balancing matrices are obtained. The second-order gramians [37] were defined based on the first-order realization in a standard state-space form $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ (2.2) with *2n*-dimensional state $x^T = [q^T \dot{q}^T]$, where

$$\mathcal{A} = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix}; \quad \mathcal{B} = \begin{bmatrix} 0 \\ M^{-1}B \end{bmatrix}; \quad \mathcal{C} = \begin{bmatrix} P & Q \end{bmatrix} \tag{2.54}$$

The first-order realization has the same input-output behavior as the second-order system. Although a first-order MOR approach, like classic balanced truncation [38], can be applied to reduce (2.54), the reduced model is no longer a second-order. To perform the reduction directly on the second-order equations (2.3), one needs to define gramians for second-order systems. Similar to the first order gramian definition (2.24), the second order controllability gramian definition is based on the following optimization problem [37]

$$\begin{aligned} &min_{\dot{q}(0)\in R^n, u\in L_2[-\infty,0]} \left(\int_{-\infty}^{0} u^T(t)u(t)dt \right) \\ &\text{subject to } M\ddot{q}(t) + D\dot{q}(t) + Kq(t) = Bu(t);\ q(0) = q_0 \end{aligned} \tag{2.55}$$

which minimizes the necessary energy to reach the given q_0 over all past inputs $u \in L_2[-\infty, 0]$ and initial $\dot{q}(0) \in R^n$. First, we minimize the energy over all past inputs $u \in L_2[-\infty, 0]$, the solution of which has been available based on the optimization problem related to the first-order gramian (2.24)

$$min_{\dot{q}(0)\in R^n}(min_{u\in L_2[-\infty,0]}\left(\int_{-\infty}^{0} u^T(t)u(t)dt\right)) = min_{\dot{q}(0)\in R^n}(x_0^T\mathcal{X}^{-1}x_0) \tag{2.56}$$

If we compatibly partition the controllability gramian of the first-order realization (2.54) $\mathcal{X}$ and its inverse $\mathcal{X}^{-1}$ as

$$\mathcal{X} = \begin{bmatrix} R_1 & R_2 \\ R_2^T & R_3 \end{bmatrix}; \quad \mathcal{X}^{-1} = \begin{bmatrix} \tilde{R}_1 & \tilde{R}_2 \\ \tilde{R}_2^T & \tilde{R}_3 \end{bmatrix} \tag{2.57}$$

then we minimize the energy over initial $\dot{q}(0) \in R^n$

$$min_{\dot{q}(0)\in R^n}(x_0^T\mathcal{X}^{-1}x_0) = min_{\dot{q}_0\in R^n}(\begin{bmatrix} q_0^T & \dot{q}_0^T \end{bmatrix}\begin{bmatrix} \tilde{R}_1 & \tilde{R}_2 \\ \tilde{R}_2^T & \tilde{R}_3 \end{bmatrix}\begin{bmatrix} q_0 \\ \dot{q}_0 \end{bmatrix}) \tag{2.58}$$

By annihilating the gradient, we can obtain the minimum energy $q_0^T(\tilde{R}_1 - \tilde{R}_2\tilde{R}_3^{-1}\tilde{R}_2^T)q_0$. Since $\tilde{R}_1 - \tilde{R}_2\tilde{R}_3^{-1}\tilde{R}_2^T$ is the Schur complement of $\tilde{R}_3$, we have $\tilde{R}_1 - \tilde{R}_2\tilde{R}_3^{-1}\tilde{R}_2^T = R_1^{-1}$ and

$$min_{\dot{q}(0)\in R^n}(x_0^T\mathcal{X}^{-1}x_0) = q_0^T R_1^{-1} q_0 \tag{2.59}$$

So the optimum for the problem (2.55) is $q_0^T R_1^{-1} q_0$ and thus the controllability gramian of the second-order system is $X_2 = R_1$. Similarly, the second-order observability gramian

definition is based on the following optimization problem

$$\begin{aligned} &max_{\dot{q}(0)\in R^n, y\in L_2[0,\infty]} \left(\int_0^\infty y^T(t)y(t)dt\right) \\ &\text{subject to } M\ddot{q}(t) + D\dot{q}(t) + Kq(t) = Bu(t),\ q(0) = q_0 \end{aligned} \tag{2.60}$$

If we compatibly partition the observability gramian of the first-order realization (2.54) as

$$\mathcal{Y} = \begin{bmatrix} N_1 & N_2 \\ N_2^T & N_3 \end{bmatrix} \tag{2.61}$$

then the observability gramian of the second-order system is $Y_2 = N_1$. The eigenvalues of the gramian product X_2Y_2 are invariant under a similarity transformation. Let W and V be the dominant left and right eigenvectors of the gramian product X_2Y_2. A reduced second-order model can be obtained as follows $(M_r, D_r, K_r, B_r, C_r)$ in which

$$M_r = W^T MV; \quad D_r = W^T DV; \quad K_r = W^T KV; \quad B_r = W^T B; \quad C_r = CV \tag{2.62}$$

However, in order to preserve the symmetry and stability of the original system, an orthogonal projection is performed in [37] as follows

$$M_r = V^T MV; \quad D_r = V^T DV; \quad K_r = V^T KV; \quad B_r = V^T B; \quad C_r = CV \tag{2.63}$$

where the equations are left multiplied by V instead of W. Unfortunately, since $W \neq V$ for a non-symmetric system (2.54), the resulting gramian product X_2Y_2 will not be balanced and accuracy is sacrificed. In fact, this issue has been resolved [60], which is to be presented

in the following section, where second-order systems are in a symmetric form

$$M = M^T; \quad D = D^T; \quad K = K^T; \quad B = C^T \tag{2.64}$$

2.4 RLC circuit equations

RLC circuits are a special class of dynamic systems. Corresponding circuit formulations are dynamic system models with special internal structures. In this section, we present first and second-order circuit formulations.

Second-order circuit formulation (C, G, Γ, B)

The nodal analysis(MNA) circuit equations are shown as follows

$$\begin{aligned} C\dot{v}(t) + Gv(t) + \Gamma \int v(t) &= Bi(t) \\ y(t) &= B^T v(t) \end{aligned} \tag{2.65}$$

where $i(t)$, $y(t) \in R^p$ are input currents and output voltages; $v(t) \in R^n$ are nodal voltages; $G, C, \Gamma \in R^{n \times n}$ are matrices of conductance, capacitance and susceptance respectively; $B \in R^{n \times p}$ is the input matrix and its transpose $B^T \in R^{p \times n}$ is the output matrix. An important property in second order formulation is that the system matrices are both symmetric and positive semi-definite

$$C = C^T \geq 0 \quad G = G^T \geq 0 \quad \Gamma = \Gamma^T \geq 0 \tag{2.66}$$

which means the formulation fulfills both the sufficient conditions of passivity (2.8) and the symmetric conditions (2.64) for second-order systems.

First-order passive circuit formulation $(\mathcal{C}, \mathcal{G}, \mathcal{B})$

The modified nodal analysis(MNA) circuit equations are shown as follows

$$\begin{aligned} \mathcal{C}\dot{x}(t) &= -\mathcal{G}x(t) + \mathcal{B}i(t) \\ y(t) &= \mathcal{B}^T x(t) \end{aligned} \tag{2.67}$$

where

$$\mathcal{C} = \begin{bmatrix} C & 0 \\ 0 & L \end{bmatrix} \quad \mathcal{G} = \begin{bmatrix} G & E \\ -E^T & 0 \end{bmatrix} \quad \mathcal{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \tag{2.68}$$

The matrices have the following properties:

$$\mathcal{C} = \mathcal{C}^T \geq 0 \quad \mathcal{G} + \mathcal{G}^T \geq 0 \tag{2.69}$$

Hence the formulation is in a passive form described by the sufficient conditions of passivity (2.8). However, such a formulation is not in a symmetric form (2.45) because $\mathcal{G}$ is not symmetric.

First-order symmetric circuit formulation $(\mathcal{C}_s, \mathcal{G}_s, \mathcal{B})$

It is easy to see the formulation(2.68) can be rewritten into a symmetric formulation $(\mathcal{C}_s, \mathcal{G}_s, \mathcal{B})$

$$\begin{aligned} \mathcal{C}_s\dot{x}(t) &= -\mathcal{G}_s x(t) + \mathcal{B}i(t) \\ y(t) &= \mathcal{B}^T x(t) \end{aligned} \tag{2.70}$$

where

$$\mathcal{C}_s = \begin{bmatrix} C & 0 \\ 0 & -L \end{bmatrix} \quad \mathcal{G}_s = \begin{bmatrix} G & E \\ E^T & 0 \end{bmatrix} \quad \mathcal{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \tag{2.71}$$

Since both $\mathcal{G}_s$ and $\mathcal{C}_s$ are symmetric, this formulation falls into the class of systems in descriptor form (2.1) with additional symmetric conditions (2.45). However, since $\mathcal{G}_s$ and $\mathcal{C}_s$ are no longer positive semi-definite, the sufficient conditions of passivity (2.8) are violated.

It is easy to verify the formulations (2.67), (2.70) and (2.65), have the same transfer function

$$H(s) = B^T(Cs + G + \Gamma/s)^{-1}B \tag{2.72}$$

Hence they are equivalent in terms of input-output behavior and either (2.67) or (2.70) can be viewed as a first-order realization of (2.65).

There is always a tradeoff in the first-order circuit formulation, either symmetric (implying accuracy) or positive semi-definite (implying passivity). Both can be obtained simultaneously only when the circuits are RC/RL circuits, where the formulations (2.65), (2.67),and (2.70) equal

$$\begin{aligned} C\dot{v}(t) &= -Gv(t) + Bi(t) \\ y(t) &= B^T v(t) \end{aligned} \tag{2.73}$$

with $C = C^T \geq 0$ and $G = G^T \geq 0$.

Chapter 3

Fast first-order balanced truncation via double gramians approximation

The classical balanced truncation approaches [38, 30] produce nearly optimal models with controllable a *priori* global error bound [20]. However, those classical methods are too expensive to directly apply to large-scale problems due to the cubic cost to solve two Lyapunov equations. There has been significant effort devoted to mitigate this difficulty recently, which has led to two classes of approximate balanced truncation methods. The first class is based on the approximate balancing by iterative low rank solution of Lyapunov equation [32, 33, 31, 62, 54]. The second class is based on the low rank gramian approximation [43, 49, 44].

The low rank gramian approximation methods were proposed in [58, 43]. In [58], both controllability and observability gramians are computed in a sampling based method and eigendecomposition is performed on the product of the two gramians to compute the projection matrix, which is still very expensive. In [43], which is called the PMTBR method, only controllability gramian is computed in a similar sampling way and the system is re-

duced by projecting onto the approximate dominant controllable subspace only, which can be obtained much more cheaply by using singular value decomposition (SVD). Although no rigorous global error bounds exist as the classical method, those methods often produce a better approximation over a wide frequency range than Krylov subspace methods at the similar cost.

As shown in [54], considering only one gramian can lead to large errors as both controllability and observability gramians and their corresponding subspaces can be quite different for general interconnect circuits. Considering both gramians requires eigendecomposition of the product of two gramians, which is a $n \times n$ matrix, where n is the size of the problem. Hence the resulting method will have the similar computational cost, which is $O(n^3)$, of solving Lyapunov equations in the standard TBR method.

In this section, we propose a fast balanced truncation method where the system is balanced in terms of two approximate gramians as achieved in the classical balanced truncation method. The novelty of the new method is that we can keep the similar computing costs of the single gramian method. The proposed algorithm is based on a generalized SVD-based balancing scheme, which is the extension of the classical balanced truncation method [30], where the balancing transformation is determined through the SVD of the product of Cholesky factors of gramians without explicitly forming the gramian product. In the proposed method, instead of Cholesky factors, different factors are applied such that the dominant invariant subspace of the approximate gramian product can be obtained in a very efficient way without explicitly forming the gramians. Experimental results on a number of published benchmarks show that the proposed method is much more accurate than the single gramian approximation method at the similar computing cost.

3.1 Gramian approximation method

To mitigate high computing costs of classic BT methods for solving large problems in VLSI design, gramian approximation methods have been proposed, where the approximate dominant subspace of a gramian can be obtained in a very efficient way.

3.1.1 Gramian expression in frequency domain

Given the state-space model (A, B, C, D), in frequency domain, the controllability gramian X can be computed from the expression

$$X = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (j\omega I - A)^{-1} B B^T (j\omega I - A)^{-H} dw \tag{3.1}$$

and the observability gramian Y can be computed from the expression

$$Y = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (j\omega I - A^T)^{-1} C^T C (j\omega I - A^T)^{-H} dw \tag{3.2}$$

where superscript H denotes Hermitian transpose.

3.1.2 Single gramian approximation

Willcox proposed the sampling-based method to compute the approximate gramians to avoid the Lyapunov equations[58]. But the method can not avoid the expensive eigende-composition of the gramian product. To mitigate this problem, Phillips proposed single gramian approximation method (PMTBR) [44], where the approximate dominant subspace of controllability gramian (3.1) can be obtained in a very efficient way.

Specifically, let ω_k be kth sampling point. If we define

$$z_{c_k} = (j\omega_k I - A)^{-1} B \tag{3.3}$$

then X can be approximated as

$$\hat{X} = \sum w_k z_{c_k} z_{c_k}^H = Z_c W^2 Z_c^H \tag{3.4}$$

where $Z_c = [z_{c_1}, z_{c_2}, \ldots, z_{c_N}]$ and W a diagonal matrix with diagonal entries $W_{kk} = \sqrt{w_k}$. The weight w_k may come from a specific numerical quadrature method. In fact, we can set $w_k = 1$. Since $\hat{X}$ is symmetric, it is orthogonally diagonalizable. If we perform the SVD on

$$Z_c = U \Sigma V^T \tag{3.5}$$

then we have

$$\hat{X} = Z_c {Z_c}^H = U \Sigma^2 U^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & \Sigma_2^2 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \tag{3.6}$$

where $U^T U = I$. If the quadrature rule is accurate, $\hat{X}$ will converge to X and the dominant eigenspace of $\hat{X}$ converges to the dominant eigenspace of X by perturbation analysis of invariant subspaces. As a result, the dominant eigenvectors U_1 can be used as the projection matrix and the reduced model (A_r, B_r, C_r, D) can be obtained as

$$A_r = U_1^T A U_1 \quad B_r = U_1^T B \quad C_r = C U_1 \tag{3.7}$$

Note that the method does not need to form the $n \times n$ gramian $Z_c Z_c^H$ explicitly and neither does it need to perform the eigenvalue decomposition at the cost of $O(n^3)$. Instead, it only performs the SVD on a $n \times Np$ thin matrix Z_c instead ($Np \ll n$).

Given N sampling points and p inputs, the cost of SVD on matrix $Z_{n\times Np}$ is $O(n(Np)^2)$. In addition, it takes N matrix factorizations and Np matrix solves. The total cost is $O(n(Np)^2 + Nn^\beta + Npn^\alpha)$ (typically, $1.1 \leq \beta \leq 1.5$ and $1 \leq \alpha \leq 1.2$ for circuits) [44], which is dominated by $O(Nn^\beta) < O(n^2)$.

Note that, the same procedure can be performed based on observability gramian (3.2) as well if we define $z_{o_k} = (j\omega_k I - A^T)^{-1}C^T$ and perform an SVD on $Z_o = [z_{o_1}, \ldots, z_{o_N}]$.

3.1.3 Existing problems

PMTBR only uses controllability gramian and it works well for symmetric systems where both controllability and observability gramians are the same. But it may not work well for general unsymmetrical systems like RLC interconnect systems as shown in [54].

To consider two gramians, the most straightforward way is to compute $\hat{X} = Z_c Z_c^H$ and $\hat{Y} = Z_o Z_o^H$ respectively and perform an eigendecomposition on the product $\hat{X}\hat{Y}$ [58]. However, the eigendecomposition has to be performed on a $n \times n$ full matrix $\hat{X}\hat{Y}$, which still has the computational order of $O(n^3)$.

Another possible way is to use cross-gramian X_{CG}, which contains both controllability and observability information in a single matrix. In the frequency domain, X_{CG} is expressed as

$$X_{CG} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} (j\omega I - A)^{-1}BC(j\omega I - A)^{-1}dw \tag{3.8}$$

which can be approximated as $\hat{X}_{CG} = \sum z_{c_k} z_{o_k}^H = Z_c Z_o^H$. In this case, we do not need to compute $\hat{X}$, $\hat{Y}$, and $\hat{X}\hat{Y}$. However, to determine the dominant subspace of X_{CG}, we still need to perform an eigendecomposition on a $n \times n$ full matrix $Z_c Z_o^H$.

3.2 New double gramian approximation method

In this section, we present the new double gramians approximation method.

3.2.1 Classical SVD-based balancing algorithm

In classical balanced truncation [38], the most straightforward way to determine the balancing transformation T is to perform an eigendecomposition of the gramian product XY and T is the invariant subspace.

However, in practice, an SVD-based method was proposed in [30], where the balancing transformation T is determined through computing the singular value decomposition (SVD) of a certain product of matrices without explicitly forming the gramian product. The algorithm is shown in Fig. 3.1.

In this algorithm, given the controllability gramian $X > 0$ and observability gramian $Y > 0$, the Cholesky factors are computed first. Let L_c and L_o denote the lower triangular Cholesky factors of the gramians X and Y

$$X = L_c L_c^T \quad Y = L_o L_o^T \tag{3.9}$$

Then the singular value decomposition of the product of the Cholesky factors is computed as

$$L_o^T L_c = U \Sigma V^T \tag{3.10}$$

where $U^T U = I$ and $V^T V = I$. The balancing transformation T and T^{-1} are given as

$$T = L_c V \Sigma^{-1/2} \quad T^{-1} = \Sigma^{-1/2} U^T L_o^T \tag{3.11}$$

Therefore, instead of explicitly forming the gramian product XY and performing an eigen-

CLASSICAL SVD-BASED BALANCING METHOD:
Input: $H : (A, B, C, D)$
Output: $H_r : (A_r, B_r, C_r, D)$

1. Compute $X > 0$ and $Y > 0$
2. Cholesky factorization $X = L_c L_c^T$ and $Y = L_o L_o^T$
3. Compute SVD of $U\Sigma V = L_o^T L_c$
4. Compute $T = L_c V \Sigma^{-1/2}$ and $T^{-1} = \Sigma^{-1/2} U^T L_o^T$
5. Compute the balanced realizations $\tilde{A} = T^{-1}AT$,$\tilde{B} = T^{-1}B$,$\tilde{C} = CT$
6. Truncate to form the reduced system (A_r, B_r, C_r, D)

Figure 3.1: Classical SVD-based balancing method.

decomposition, the invariant subspace T can be determined in a less expensive and more efficient way in the classical balanced truncation algorithm.

3.2.2 Generalized SVD-based balancing algorithm

In this paper, we propose a generalized SVD-based balancing scheme. In the classical SVD-based balancing algorithm, the Cholesky factors are used. However, we show that Cholesky factors are not the only choice to compute the balancing transformation. In fact, there is no restriction on the structure of matrix factors at all. We first show the following result:

Theorem 1 *Assume the gramians X and Y can be factorized as*

$$X = Z_c Z_c^T \quad Y = Z_o Z_o^T \tag{3.12}$$

where Z_c and Z_o are matrix factors with arbitrary structure. Then the singular value de-

composition of the product of the factors is computed as

$$Z_o^T Z_c = U\Sigma V^T \tag{3.13}$$

where $U^T U = I$ *and* $V^T V = I$.

In this case, the balancing transformation T *and* T^{-1} *are given as*

$$T = Z_c V \Sigma^{-1/2} \quad T^{-1} = \Sigma^{-1/2} U^T Z_o^T \tag{3.14}$$

Given (3.12) and (3.13), it can be shown that T^{-1} is an inverse matrix of T in (3.14)

$$\begin{aligned} T^{-1}T &= (\Sigma^{-1/2}U^T Z_o^T)(Z_c V \Sigma^{-1/2}) \\ &= \Sigma^{-1/2}U^T(Z_o^T Z_c)V\Sigma^{-1/2} \\ &= \Sigma^{-1/2}U^T(U\Sigma V^T)V\Sigma^{-1/2} \\ &= \Sigma^{-1/2}(U^T U)\Sigma(V^T V)\Sigma^{-1/2} \\ &= \Sigma^{-1/2}\Sigma\Sigma^{-1/2} \\ &= I \end{aligned} \tag{3.15}$$

and T is exactly the invariant subspace of gramian product XY (2.35)

$$\begin{aligned} T^{-1}XYT &= (\Sigma^{-1/2}U^T Z_o^T)(XY)(Z_c V \Sigma^{-1/2}) \\ &= (\Sigma^{-1/2}U^T Z_o^T)(Z_c Z_c^T)(Z_o Z_o^T)(Z_c V \Sigma^{-1/2}) \\ &= \Sigma^{-1/2}U^T(Z_o^T Z_c)(Z_c^T Z_o)(Z_o^T Z_c)V\Sigma^{-1/2} \\ &= \Sigma^{-1/2}U^T(U\Sigma V^T)(V\Sigma U^T)(U\Sigma V^T)V\Sigma^{-1/2} \\ &= \Sigma^{-1/2}\Sigma\Sigma\Sigma\Sigma^{-1/2} \\ &= \Sigma^2 \end{aligned} \tag{3.16}$$

Therefore, the factors of gramians are no longer limited to Cholesky factors.

3.2.3 New double gramians approximation method

Now, we apply the proposed generalized SVD-based balancing algorithm for double gramians approximation.

Let ω_k be kth sampling point. If we define

$$\begin{aligned} z_{c_k} &= (j\omega_k I - A)^{-1} B \\ z_{o_k} &= (j\omega_k I - A^T)^{-1} C^T \end{aligned} \tag{3.17}$$

then X and Y can be approximated as

$$\begin{aligned} \hat{X} &= \sum z_{c_k} z_{c_k}^H = Z_c Z_c^H \\ \hat{Y} &= \sum z_{o_k} z_{o_k}^H = Z_o Z_o^H \end{aligned} \tag{3.18}$$

where $Z_c = [z_{c_1}, z_{c_2}, \ldots, z_{c_N}]$ and $Z_o = [z_{o_1}, z_{o_2}, \ldots, z_{o_N}]$.

According to the generalized SVD-based balancing algorithm, Z_c and Z_o can be used as the factors to compute the balancing transformation T. We perform a singular value decomposition on

$$Z_o^H Z_c = U \Sigma V^H \tag{3.19}$$

Notice that, given N sampling points and p inputs and q outputs, Z_c and Z_o are $n \times Np$ and $n \times Nq$ matrices, respectively, where $Np \ll n$ and $Nq \ll n$. Assume $m = max(Np, Nq)$, the dimension of matrix $Z_o^T Z_c$ is smaller than $m \times m$ and the cost of SVD

is $O(m^3)$, which is much smaller than $O(n^3)$ when $m \ll n$. Then we have

$$Z_o{}^H Z_c = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & \Sigma_2^2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{3.20}$$

where U_1 and V_1 are dominant subspace corresponding to the first r largest singular values Σ_1. Then the right projection matrix and left projection matrix are given by

$$T_r = Z_c V_1 \Sigma_1^{-1/2} \quad T_l = \Sigma_1^{-1/2} U_1^T Z_o^T \tag{3.21}$$

T_r and T_l are $n \times r$ and $r \times n$ matrices respectively, corresponding to the dominant invariant subspace of the approximate gramian product $\hat{X}\hat{Y}$. The reduced model (A_r, B_r, C_r, D) can be obtained as

$$A_r = T_l A T_r \quad B_r = T_l B \quad C_r = C T_r \tag{3.22}$$

The resulting double gramians approximation algorithm, called DGA, is given in Fig. 3.2.

Given N sampling points and p inputs and q outputs, assume $m = max(Np, Nq)$, the cost of SVD on matrix $Z_o^T Z_c$ is $O(m^3)$ as analyzed before. In addition, it takes $2N$ matrix factorizations and $pN + qN$ matrix solves to obtain Z_c and Z_o, and $O(pqn)$ to obtain $Z_o^T Z_c$.

The total cost is $O(m^3 + 2Nn^\beta + (p+q)Nn^\alpha + pqn)$ (typically, $1.1 \leq \beta \leq 1.5$ and $1 \leq \alpha \leq 1.2$ for circuits). As $m \ll n$, the cost is still dominated by $O(2Nn^\beta)$, which is about twice the cost of single gramian approximation method PMTBR [44] but still in the same growth order. In fact, the cost of DGA is less than twice the cost of PMTBR as the cost of SVD process in DGA is $O(m^3)$, which is much less than $O(n(Np)^2)$ in PMTBR, where $Np \leq m$ and $m \ll n$.

Practically, we notice that PMTBR uses incremental QR on Z_c to find the project ma-

trix. But it will have the same computational costs of SVD on Z_c. Incremental QR can be used for the proposed method on $Z_o^T Z_c$ also.

We want to emphasize that although the proposed method is slower than the PMTBR as DGA computes two approximate gramians, *it does not mean that PMTBR will achieve the same accuracy of the proposed method if both have the exactly same computing costs (for instance PMTBR samples twice of the DGA method)*. The proposed method addresses the fundamental problem of *using only single gramians* in the PMTBR-like method.

NEW DOUBLE GRAMIANS APPROXIMATION METHOD (DGA):
Input: $H : (A, B, C, D)$
Output: $H_r : (A_r, B_r, C_r, D)$

1. Select N sampling points $z_k (k = 1, 2, \ldots, N)$
2. Compute $Z_c = [z_{c_1}, z_{c_2}, \ldots, z_{c_N}]$ where $z_{c_k} = (j\omega_k I - A)^{-1} B$
3. Compute $Z_o = [z_{o_1}, z_{o_2}, \ldots, z_{o_N}]$ where $z_{o_k} = (j\omega_k I - A^T)^{-1} C^T$
4. Perform SVD on matrix ${Z_o}^H Z_c = U \Sigma V^T$
5. Compute right projection matrix T_r and left projection matrix T_l as $T_r = Z_c V_1 \Sigma_1^{-1/2}$ and $T_l^T = \Sigma_1^{-1/2} U_1^T Z_o^T$
6. Project onto the dominant invariant subspace of the approximate gramian product $A_r = T_l A T_r, B_r = T_l B, C_r = C T_r$

Figure 3.2: New double gramians approximation method (DGA).

3.2.4 Practical Implementation

Descriptor systems

A special class of dynamic systems is the RLC interconnect circuit described by state-space equations in descriptor form

$$\begin{aligned} C\dot{x}(t) &= -Gx(t) + Bu(t) \\ y(t) &= Lx(t) \end{aligned} \tag{3.23}$$

In this case, the controllability gramian and observability gramian can be obtained from generalized Lyapunov equations [54, 44] and the frequency domain expressions are [44]

$$\begin{aligned} X &= \tfrac{1}{2\pi}\int_{-\infty}^{+\infty}(j\omega C+G)^{-1}BB^T(j\omega C+G)^{-H}dw \\ Y &= \tfrac{1}{2\pi}\int_{-\infty}^{+\infty}(j\omega C+G^T)^{-1}L^TL(j\omega C+G^T)^{-H}dw \end{aligned} \tag{3.24}$$

As a result, z_{c_k}'s and z_{o_k}'s in Fig. 3.2 will be replaced by

$$z_{c_k} = (j\omega_k C+G)^{-1}B \quad z_{o_k} = (j\omega_k C+G^T)^{-1}L^T \tag{3.25}$$

Just as PMTBR, the complications present in applying standard balanced truncation to problems with singular descriptor matrix C vanish in the proposed method.

Passivity

Similar to classical balanced truncation [30], the proposed double gramians approximation method does not preserve the passivity for general dynamical systems. But post-passivity-enforcement process can be carried out to ensure the passivity [12, 54], which is out of scope of this paper.

3.3 Experimental results

In this section, we present experimental results on four benchmark examples used in published papers, which are NOT symmetric. The proposed double gramians approximation method, called DGA, is compared with existing single gramian approximation method PMTBR and Krylov subspace method PRIMA at the same reduced order. Note that, the sampling points (total number and their locations) for both DGA and PMTBR are exactly

Table 3.1: Reduction CPU time comparison of PMTBR and DGA (seconds).

n	152	1520	150002
DGA	0.014597	0.06298	8.810724
PMTBR	0.008164	0.03228	5.070467

the same.

3.3.1 Two RLC lines

The first and second examples are RLC lines used in [54] (in Fig.2 and Fig.4). The two RLC lines are of the same order 1502 but with different topologies. In both examples, input signal $u(t)$ is the voltage at the first node and output is the current flowing through the voltage source. The state vector consists of node voltages, inductor currents, and currents through the voltage source. The MNA formulation for the two lines results in two systems (G, C, B, L) with unsymmetric matrices G. In both examples, the parameters are $R = 0.1, L = 2, C = 15$ and the reduced orders are set to be 10.

The results for the first and second RLC lines are shown in Fig. 3.3(above) and Fig. 3.3(below), respectively. Clearly, we see that DGA is much more accurate that PMTBR and PRIMA in both examples. The reason why PRIMA's results are quite off is that it approximates only dominant controllable states [54]. We then compare the CPU time of DGA and PMTBR. The reduction CPU times are shown in Table 3.1, where the n is the order of the RLC line and the reduced order is 10. From Table 3.1, we can see, the reduction time of DGA is less than twice the reduction time of PMTBR. This is because DGA is much less expensive than PMTBR in the SVD process although DGA has to take twice matrix factorizations and solving.

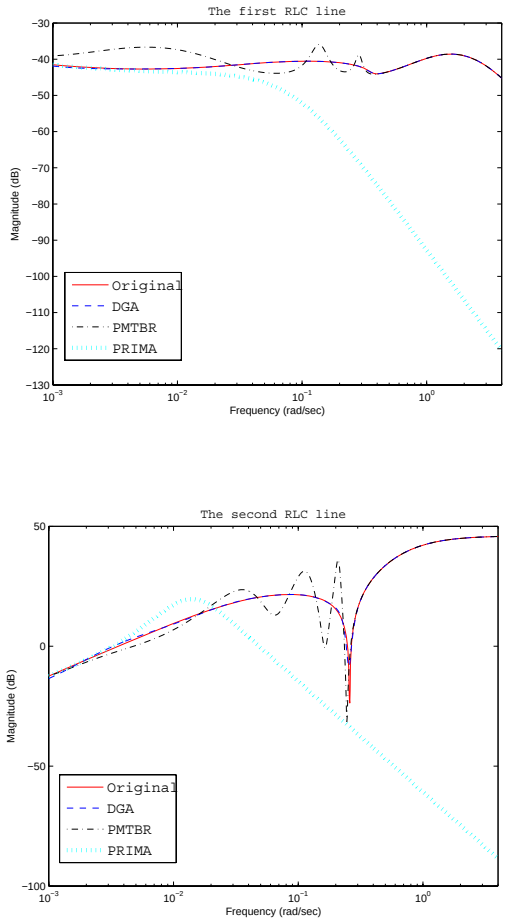

Figure 3.3: Comparison results on the two RLC line examples.

3.3.2 A transmission line model

The third example is a transmission line model from [32, 9], which is not symmetric and has an order of 256. There are 2 inputs and 2 outputs in this model, which results in a 2×2 transfer matrix $H(s)$. The reduced orders are 30.

The results for the diagonal terms $h_{11}(s)$ and $h_{22}(s)$ of $H(s)$ are shown in Fig. 3.4(above) and Fig. 3.4(below), respectively. For this unsymmetric example, the proposed DGA method produces the best wideband approximation again. PRIMA is very accurate in low frequency range but not accurate beyond 10^{11} rad/sec. DGA is much more accurate than

PMTBR also in this case.

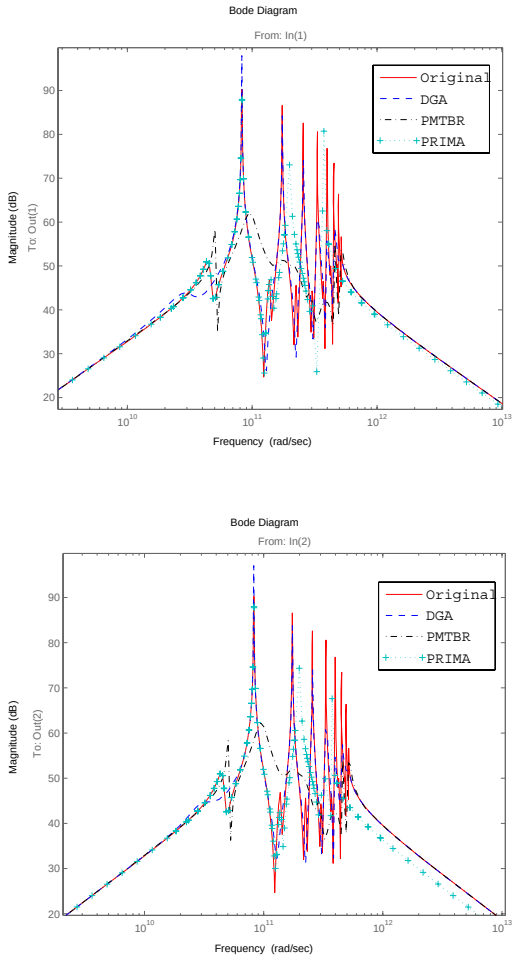

Figure 3.4: Comparison result on the transmission line example.

3.3.3 An international space station example

In addition to interconnect modeling, model reduction is being used to generate compact models of various dynamic systems [54]. The forth example is a structural model of component 1r (Russian service module) of the International Space Station from [9]. This example is also not symmetric and has an order of 270. There are 3 inputs and 3 outputs in this model, which results in a 3×3 transfer matrix $H(s)$. The reduced orders are 25.

The results for the diagonal term $h_{22}(s)$ and the off diagonal term $h_{32}(s)$ are shown in Fig. 3.5(above) and Fig. 3.5(below), respectively. The same conclusion can be drawn here. PRIMA is only accurate in low frequency range, which can match up to about 10 rad/sec for $h_{22}(s)$ and about 1 rad/sec for $h_{32}(s)$. However, DGA still has the excellent wideband accuracy and the performance is much better than PMTBR.

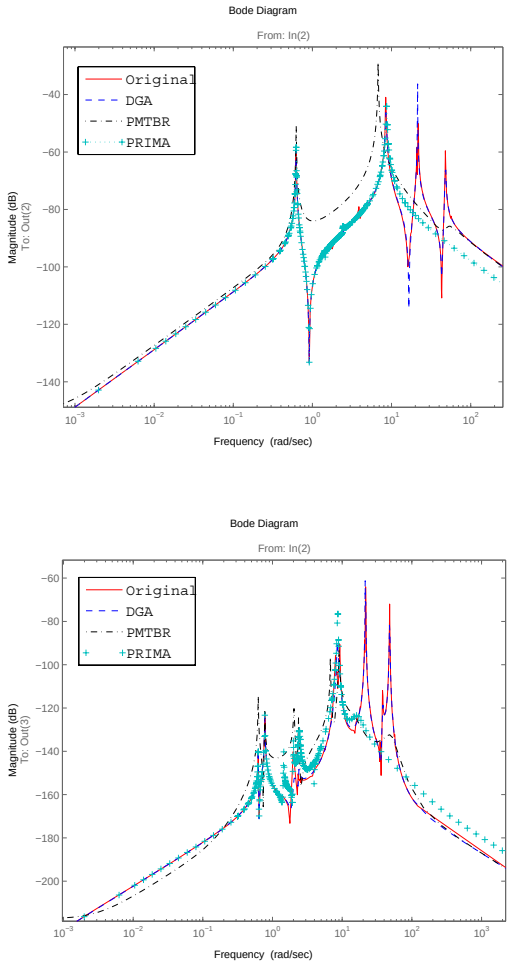

Figure 3.5: Comparison result on the ISS example.

Chapter 4

Second-order balanced truncation for passive order reduction of RLC circuits and its fast version

Model order reduction (MOR) is an efficient technique to reduce the circuit complexity while producing a good approximation of the input-output behavior when an RLCK circuit is formulated in the second-order form, inductance (or partial inductance) will be represented in its inverse form, which is called susceptance. Susceptance coupling are shown to be more localized than inductance coupling and its matrix is diagonally dominant like capacitance matrix [14]. Hence, susceptance matrix can be sparsified much easily without loss of stability, which, however, is difficult in general for the inductance matrix [24]. The new susceptance element (called ”K” element) can be stamped back into the circuit matrix using the SPICE-compatible equivalent circuits [25]. Model order reduction techniques for second-order systems, which are more suitable for reducing RLCK circuits, have been developed in the past [48, 53].

However, existing second-order MOR techniques are mainly based on Krylov-subspace methods, which in general have difficulties to generate reduced models with global accuracy. Therefore, another approach, truncated balanced realization (TBR), or balanced truncation (BT), which was originally developed in the control community [38], has been studied intensively for interconnect reduction recently [42, 44, 49, 56, 57, 59]. The idea of TBR method is to first transform an original system into a new coordinate such that each state in this coordinate is equally controllable and observable before the consequent truncation of the weak states. To perform the passive reduction, positive-real TBR (PR-TBR) was applied in [42], which solves more expensive quadratic matrix equations. PR-TBR has no constrains on the internal structure of the state-space equations. But it also does not preserve any structure information inherent to RLCK circuits such as symmetry, positive semi-definiteness and sparsity, during the reduction process. Another issue is that existing balanced truncation techniques for interconnect reduction are first-order based and cannot handle RLCK circuits formulated as second-order systems.

In the control literature [37], Meyer and Srinivasan introduced a second-order balanced truncation method where second-order gramians are defined based on Moore's first-order balanced truncation method. However, in order to preserve the stability of original system, congruency transformation instead of similarity transformation is performed. As a result, the transformed system is not really balanced, which sacrifices the accuracy.

In this section, we propose a new balanced truncation method, SBPOR (Second-order Balanced truncation for Passive Order Reduction), for passive reduction of RLCK circuits. By exploiting the symmetric positive definiteness of the system matrices in the second-order circuit formulation, the new approach resolves the issue existing in [37] by defining second-order gramians based on a symmetric first-order realization. As a result, balancing

and reduction can be achieved via only congruency transformation without any accuracy degradation. In contrast to the first-order balanced truncation approaches, SBPOR can also preserve the structure information inherent to RLCK circuits and only needs to solve one linear matrix equation instead of two quadratic matrix equations. Furthermore, to mitigate the high computational cost of solving Lyapunov equation, a Second-Order Gramian Approximation version, SOGA, is proposed to generalize the existing first-order gramian approximation technique PMTBR [44] to second-order systems.

4.1 The SBPOR Algorithm

In this section, we introduce the new second-order balanced truncation method SBPOR and its gramian approximation version.

4.1.1 Symmetric realization in descriptor form

Consider a second-order LTI stable system

$$\begin{aligned} M\ddot{q}(t) + D\dot{q}(t) + Kq(t) &= Bu(t) \\ y(t) &= Pq(t) + Q\dot{q}(t) \end{aligned} \tag{4.1}$$

where $u(t) \in R^p$, $y(t) \in R^q$, $q(t) \in R^n$, $B \in R^{n\times p}$, $P, Q \in R^{q\times n}$, $M, D, K \in R^{n\times n}$ with M assumed to be nonsingular.

The general idea of reducing the second-order system is to transform the second-order system first into the equivalent first-order system, from which the balancing matrices are obtained. To this end, the second-order gramians in [37] were defined based on the first-order realization in a standard state-space form (2.1) with *2n*-dimensional state $x^T =$

$[q^T \dot{q}^T]$, where

$$\mathcal{A} = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} 0 \\ M^{-1}B \end{bmatrix}$$
$$\mathcal{C} = \begin{bmatrix} P & Q \end{bmatrix} \tag{4.2}$$

As mentioned before, RLCK circuits can be formulated in a second-order form (4.1) with special structure $M = C$, $D = G$, $K = \Gamma$, $P = 0$, $Q = B^T$

$$\begin{aligned} C\ddot{q}(t) + G\dot{q}(t) + \Gamma q(t) &= Bu(t) \\ y(t) &= B_2^T \dot{q}(t) \end{aligned} \tag{4.3}$$

where $u(t), y(t) \in R^p$ are input currents and output voltages; $\dot{q(t)} \in R^n$ are nodal voltages; $G, C, \Gamma \in R^{n\times n}$ are matrices of conductance, capacitance and susceptance respectively and $C = C^T > 0, G = G^T \geq 0, \Gamma = \Gamma^T \geq 0$; $B \in R^{n\times p}$ is the input matrix and its transpose $B^T \in R^{p\times n}$ is the output matrix. Note that C is assumed to be invertible [37, 29].

The key idea in this paper is that instead of using the first-order realization (4.2), we choose another first-order realization in descriptor form [29] with *2n*-dimensional state $x^T = [q^T, \dot{q}^T]$

$$\begin{aligned} \mathcal{E}\dot{x}(t) &= \mathcal{A}x(t) + \mathcal{B}u(t) \\ y(t) &= \mathcal{B}^T x(t) \end{aligned} \tag{4.4}$$

where

$$\mathcal{E} = \begin{bmatrix} -\Gamma & 0 \\ 0 & C \end{bmatrix}, \mathcal{A} = \begin{bmatrix} 0 & -\Gamma \\ -\Gamma & -G \end{bmatrix}, \mathcal{B} = \begin{bmatrix} 0 \\ B \end{bmatrix} \tag{4.5}$$

Note that, since C, G, Γ are all symmetric, it follows $\mathcal{A} = \mathcal{A}^{\mathcal{T}}, \mathcal{E} = \mathcal{E}^{\mathcal{T}}$, which means such a first-order realization is symmetric.

Controllability and observability gramians in descriptor form can be computed from a pair of generalized Lyapunov equations [49]. However, in this symmetric case, both gramians are equal and only one equation is to be solved

$$\mathcal{E}\mathcal{X}\mathcal{A}^T + \mathcal{A}\mathcal{X}\mathcal{E}^T + \mathcal{B}\mathcal{B}^T = 0 \tag{4.6}$$

If we compatibly partition the gramians as

$$\mathcal{X} = \mathcal{Y} = \begin{bmatrix} R & S \\ S^T & F \end{bmatrix} \tag{4.7}$$

then the second-order gramians are also equal

$$X_2 = Y_2 = R \tag{4.8}$$

Since gramian is symmetric, R is orthogonally diagonalizable, i.e., there exists $T^{-1} = T^T$ such that

$$T^T R T = \Sigma \tag{4.9}$$

As a result, the second-order gramian product RR can be orthogonally diagonalized as

$$T^T R R T = (T^T R T)(T^T R T) = (\Sigma)^2 \tag{4.10}$$

Note that the eigenspace of the gramian product is exactly the eigenspace of each gramian.

If we partition the matrices in (4.9) as

$$\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} R \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tag{4.11}$$

where Σ_1 contains the first r largest eigenvalues of gramian R and V_1 are corresponding eigenvectors, a reduced model can be obtained as follows

$$\begin{aligned} C_r\ddot{q}(t) + G_r\dot{q}(t) + \Gamma_r q(t) &= B_r u(t) \\ y &= B_{2r}^T \dot{q}(t) \end{aligned} \tag{4.12}$$

where $C_r = V_1^T C V_1, G_r = V_1^T G V_1, \Gamma_r = V_1^T \Gamma V_1, B_r = V_1^T B$. This kind of transformation is known as congruency transformation, which preserves symmetry and definiteness of matrices such that $C_r = C_r^T \geq 0, G_r = G_r^T \geq 0, \Gamma_r = \Gamma_r^T \geq 0$, implying the reduced-order system has guaranteed stability, passivity, and reciprocity [53]. The basic algorithm flow for SBPOR is given in Fig. 4.1.

4.1.2 Second-order gramian approximation

We also propose a second-order gramian approximation technique to mitigate high computational cost. Practically, we find that Γ can easily become singular, which will make both $\mathcal{A}$ and $\mathcal{E}$ in (4.5) singular. To mitigate this problem, we propose a little different symmetric realization. If we define $x^T = [E_l^T q^T, \dot{q}^T]$, we have the following new realization:

$$\begin{aligned} \mathcal{C}\dot{x}(t) &= -\mathcal{G}x(t) + \mathcal{B}u(t) \\ y(t) &= \mathcal{B}^T x(t) \end{aligned} \tag{4.13}$$

ALGORITHM 1: SBPOR
Input: C, G, Γ, B
Output: C_r, G_r, Γ_r, B_r

1. Form the symmetric first-order realization in descriptor form (4.4)
2. Solve $\mathcal{E}\mathcal{X}^T\mathcal{A}^T + \mathcal{A}\mathcal{X}\mathcal{E}^T + \mathcal{B}\mathcal{B}^T = 0$ for $\mathcal{X}$
3. Partition $\mathcal{X}$ as:
   $\mathcal{X} = \begin{bmatrix} R & S \\ S^T & F \end{bmatrix}$
4. Compute SVD of the second-order gramian:
   $R = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$
5. Form the reduced model as
   $C_r = V_1^T C V_1, G_r = V_1^T G V_1, \Gamma_r = V_1^T \Gamma V_1, B_r = V_1^T B$

Figure 4.1: The SBPOR algorithm.

where

$$\mathcal{C} = \begin{bmatrix} -L^{-1} & 0 \\ 0 & C \end{bmatrix}, \mathcal{G} = \begin{bmatrix} 0 & E_l L^{-1} \\ L^{-1} E_l{}^T & G \end{bmatrix}, \mathcal{B} = \begin{bmatrix} 0 \\ B \end{bmatrix} \tag{4.14}$$

Here E_l is the incidence matrix for inductor matrix L in the modified nodal analysis (MNA) formulation and $\Gamma = E_l L^{-1} E_l{}^T$. We remark that $E_l L^{-1}$ will not have zero rows for a physical system as $E_l q$ is actually a vector of a branch vector potential [61]. So $\mathcal{G}$ will not be singular, required by our new SOGA algorithm, for any physical system that has DC paths to ground for any node.

Since C, G, L are all symmetric, such a first-order realization is also symmetric and the second-order gramian measures the contribution of the node voltages $\dot{q} = v$ with respect to the transfer function.

For first-order system in descriptor form (4.13), the gramian $\mathcal{X}$ can be computed from

the expression in frequency domain [49]

$$\mathcal{X} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}(j\omega\mathcal{C}+\mathcal{G})^{-1}\mathcal{B}\mathcal{B}^T(j\omega\mathcal{C}+\mathcal{G})^{-H}dw \tag{4.15}$$

Let ω_k be kth sampling point over the frequency range of interests. If we define

$$z_k = (j\omega_k\mathcal{C}+\mathcal{G}^{-1})\mathcal{B} \tag{4.16}$$

then $\hat{\mathcal{X}}$ can be approximately computed as

$$\hat{\mathcal{X}} = \frac{1}{2\pi}\sum z_k z_k^H = \mathcal{Z}\mathcal{Z}^H \tag{4.17}$$

where $\mathcal{Z}$ is a matrix whose columns are z_k. If we partition $\mathcal{Z}^H = \begin{bmatrix} Z_1^H & Z_2^H \end{bmatrix}$ and compatibly partition the approximated gramian as

$$\hat{\mathcal{X}} = \begin{bmatrix} \hat{R} & \hat{S} \\ \hat{S}^T & \hat{F} \end{bmatrix} = \begin{bmatrix} Z_1Z_1^H & Z_1Z_2^H \\ Z_2Z_1^H & Z_2Z_2^H \end{bmatrix} \tag{4.18}$$

then the approximated second-order gramian is $\hat{F}$, which can be diagonalized as

$$\begin{aligned} &\hat{F} = Z_2Z_2^H = (\hat{U}\hat{\Sigma}\hat{V})(\hat{U}\hat{\Sigma}\hat{V})^T = \hat{U}\hat{\Sigma}^2\hat{U}^T = \\ &\begin{bmatrix} \hat{U}_1 & \hat{U}_2 \end{bmatrix}\begin{bmatrix} \hat{\Sigma}_1^2 & 0 \\ 0 & \hat{\Sigma}_2^2 \end{bmatrix}\begin{bmatrix} \hat{U}_1^T \\ \hat{U}_2^T \end{bmatrix} \end{aligned} \tag{4.19}$$

Therefore, $\hat{U}_1$ will be used to perform the reduction as in the SBPOR method. The SOGA algorithm is presented in Fig. 4.2.

ALGORITHM 2: SOGA
Input: C, G, Γ, B
Output: C_r, G_r, Γ_r, B_r

1. Start from the symmetric first-order realization (4.13)
2. Do until satisfied:
3. Select a frequency points s_k
4. Compute $z_k = (s_k\mathcal{C} + \mathcal{G})^{-1}\mathcal{B}$
5. Form $\mathcal{Z}_k = [z_1, z_2, \ldots, z_k]$ and partition $\mathcal{Z}_k = \begin{bmatrix} Z_{k1} \\ Z_{k2} \end{bmatrix}$
6. Compute the SVD of the matrix Z_{k2}. If the error is satisfactory, go to Step 7. Otherwise, go to Step 2.
7. Form the projection matrix $\hat{U}_1$ from the singular vectors of Z_k, dropping ones corresponding to small singular values below a desired tolerance, and form the reduced model as
   $C_r = \hat{U}_1^T C \hat{U}_1, G_r = \hat{U}_1^T G \hat{U}_1, \Gamma_r = \hat{U}_1^T \Gamma \hat{U}_1, B_r = \hat{U}_1^T B$

Figure 4.2: The SOGA algorithm.

4.2 Experimental results

In this section, we show examples that illustrate the effectiveness of proposed SBPOR method and compare it with existing relevant MOR approaches.

4.2.1 Comparison with first-order TBR

Given a circuit in the form (4.3), we first compare SBPOR with the first-order TBR method. Note that the order q in the reduced models reduced by SBPOR on (4.3) will correspond to the order of $2q$ in the reduced models by the first-order TBR method performed on equivalent first-order realization (4.2). We choose a small circuit for the purpose of illustration so that both impedances and real parts can be compared at all possible reduced

orders. The RLCK circuit has 4 nodal voltages and thus has a dimension of 4 in a second-order formulation. The equivalent first-order realization has a dimension of 8. As shown in

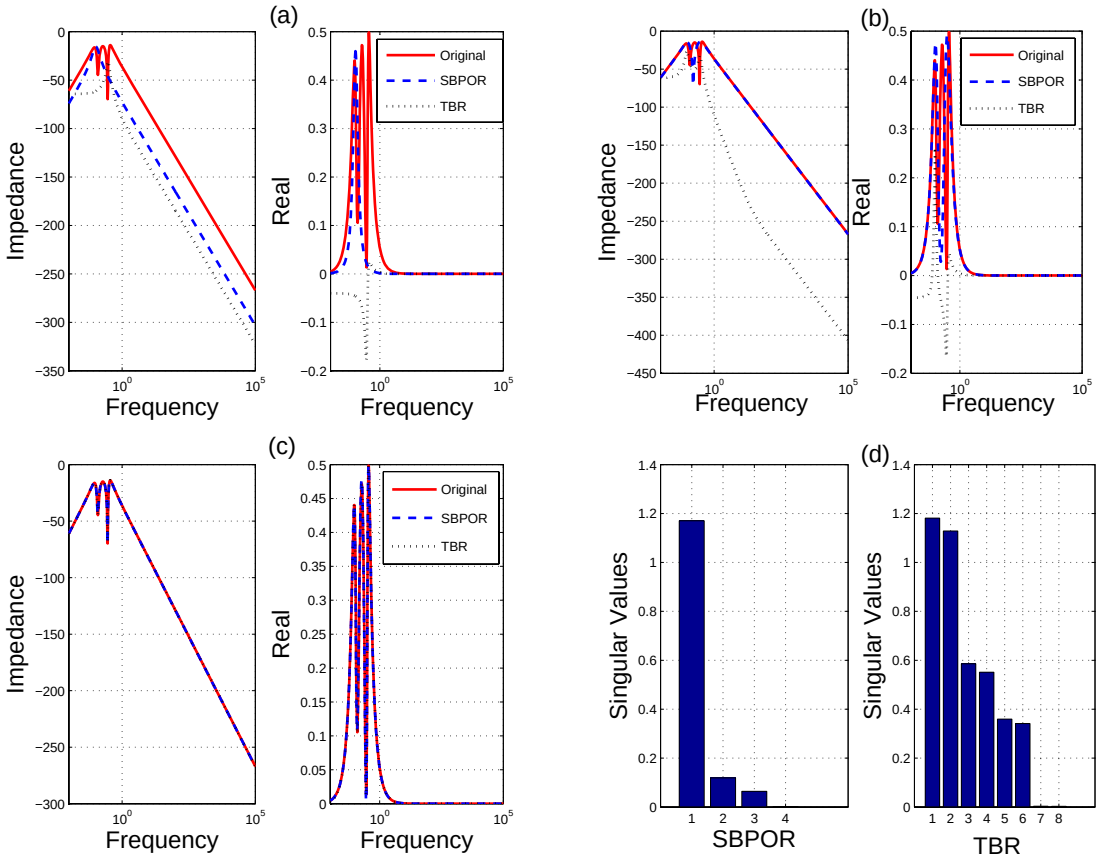

Figure 4.3: Comparison with the first-order TBR method (performed on linearied first-order system).

Fig. 4.3(a),(b),(c), SBPOR outperforms standard TBR at each reduced order (q=1,2,3). This can be explained from the 'energy' distribution of singular values as shown in Fig. 4.3(d), where the second-order singular values decay much faster than the first-order ones. The passivity of reduced models can be tested from the real parts. As expected, SBPOR can guarantee the passivity of reduced models while standard TBR cannot. As shown in Fig. 4.3(a),(b),(c), only in Fig. 4.3(c), the real part of TBR reduced model is positive at all frequencies and thus the reduced model is passive. Note that standard TBR applied to

the equivalent first-order realization(2.54) also results in a first-order reduced model and thus is not a second-order MOR approach available. We just use it as a criterion to show the accuracy of our new approach.

4.2.2 Comparison with SAPOR

In the second example, we want to compare our method with moment-matching based second-order MOR approach SAPOR [53]. The example is an RLCK circuit, which has 100 nodal voltages. The reduced second-order model has a dimension of 2. As shown in

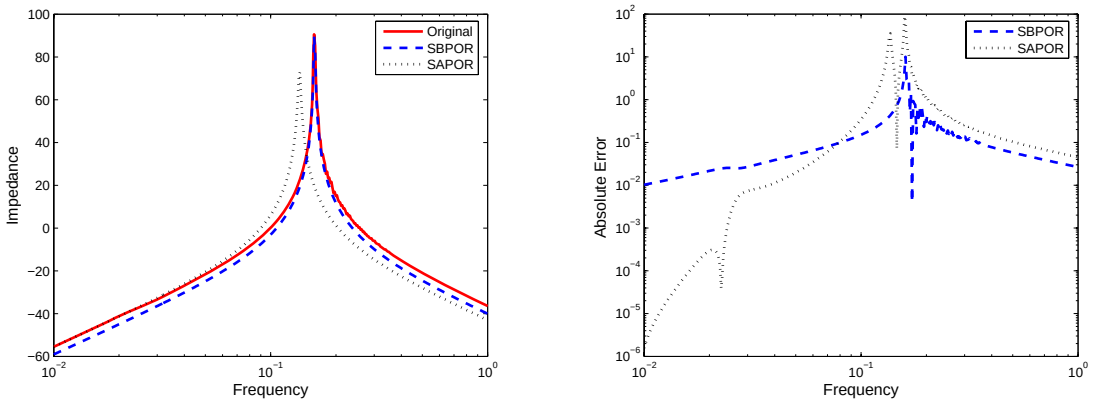

Figure 4.4: Comparison with Krylov-based second-order MOR method SAPOR [53].

Fig. 4.4(a), SBPOR is globally accurate at all frequencies while SAPOR has very good local behavior around DC (the expansion point of SAPOR is 0.01 Hz) but behaves so bad at other frequencies. The error is shown in Fig. 4.4(b), where the maximum absolute error for SBPOR is about 10 but for SAPOR is almost 100.

4.2.3 Comparison with existing second-order TBR

In this part, we want to compare the new method, SBPOR, with existing technique [37] in the control literature, which we name TBR2. The example is an RLCK circuit with 100

nodal voltages and the reduced dimension is 10. In Fig. 4.5(a), we can see that SBPOR outperforms TBR2 obviously. As shown in Fig. 4.5(b), the maximum absolute error for SBPOR is smaller than 10 while it is almost 100 for TBR2. The reason is that the system in TBR2 is not really balanced and thus the accuracy is sacrificed.

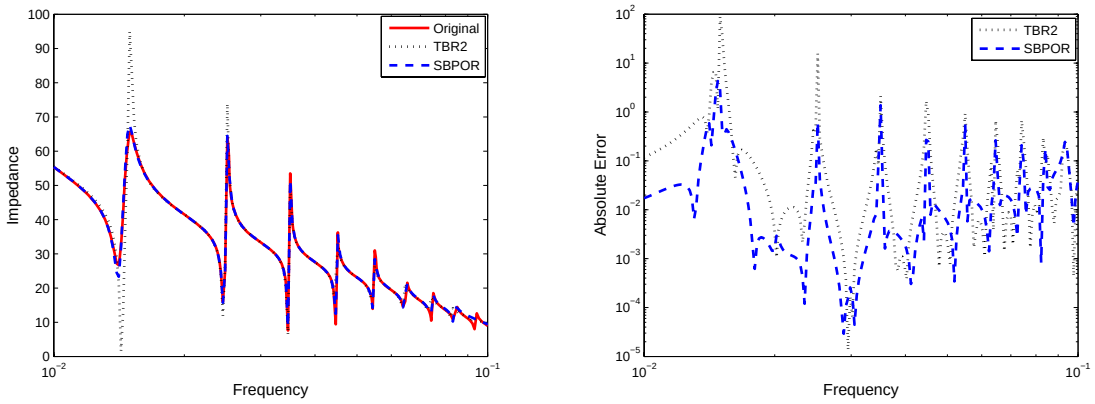

Figure 4.5: Comparison with the existing second-order TBR method [37].

4.2.4 Comparison of SOGA with SAPOR

The original model is an RLCK circuit with 1000 nodes in a second-order formulation. The reduced model has an order of 11 ($q = 11$). As shown in Fig. 4.6, SOGA produces a better approximation than SAPOR over a wide frequency band (the expansion point of SAPOR is 1 Hz). The computational cost of SOGA is almost the same as that of SAPOR given the same reduction order. The reduction CPU times of several mesh-structured RLC examples are shown in Table 4.1, where the n is the number of nodes and the reduced order is 10.

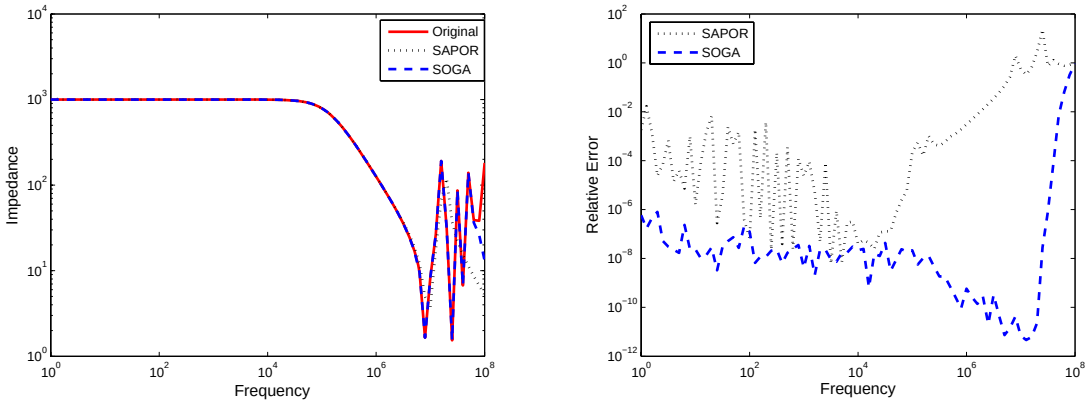

Figure 4.6: Accuracy comparison between SOGA and SAPOR.

Table 4.1: Reduction CPU time comparison of SOGA and SAPOR (seconds).

	n=640	n=1000	n=2680	n=4380
SOGA	3.257	6.875	25.24	602.63
SAPOR	1.438	3.420	21.42	580.57

Chapter 5

Positive-real rational interpolation based reduction via Carathéodory extension

Model order reduction (MOR) by Krylov subspace methods have been proved to be an efficient technique to reduce the complexity of interconnects [45, 17, 50, 40, 19]. For passive reduction of the interconnect circuits, existing approaches, however, require that the system must be formulated into the passive form, a state-space representation with positive semi-definite system matrices and the same input and output mapping matrices. The passivity is ensured by the congruency transformation, which can preserve the passive form in the reduced model. Those restrictions are generally satisfied for interconnects modeled as RLC circuits. However, many passive systems are not conveniently put into such a form [42].

In fact, passivity is more generally characterized by the positive realness of the transfer function [6]. The system is passive if and only if its transfer function $H(s)$ is positive

real [6], which means

$$\begin{aligned}
&(1) \quad H(s) \text{ is analytic for } Re(s) > 0 \\
&(2) \quad \overline{H(s)} = H(\bar{s}) \text{ for } s \in \mathbb{C} \\
&(3) \quad H(s) + H(s)^H \geq 0 \text{ for } Re(s) > 0
\end{aligned}$$

where $\overline{H}$ denotes complex conjugate, H^H denotes Hermitian (complex conjugate and transpose), and ≥ 0 denotes positive semi-definiteness in a matrix context.

In order to preserve passivity for system with arbitrary internal structure, the positive realness of the transfer function of reduced system should be enforced. Toward this goal, several methods have been proposed in the past several years, which can be divided into two categories.

The first class is represented by positive-real truncated balanced realization (PRTBR) algorithms [42, 56] Those algorithms are based on truncated balanced stochastic realization developed in the control community [13, 21]. Different from classical TBR [38], Lur'e equations or algebraic Riccati equations are needed to be solved in those algorithms instead of Lyapunov equations. However, their high computational cost $O(n^3)$ (n is the order of the system to be reduced) makes them infeasible for large-scale systems in practice.

More recently, the second class methods were proposed by Antoulas [2] and Sorensen [51]. The methods are based on an observation that if the transfer function of reduced model preserves a subset of the spectral zeros of the original system and admits the same values as the original system in the mirror points of the preserved spectral zeros, the transfer function of reduced system is also positive real. However, to obtain spectral zeros or the corresponding invariant subspace of the original system, generalized eigenvalue problem needs to be solved, which is also very expensive for large-scale applications. In addition, it is not clear

how to select spectral zeros to be preserved.

Over the last several years, a new theory of positive real analytic interpolation with complexity constraint has been developed for both scalar and matrix interpolating function (or interpolant for short) in discrete-time setting [8, 7, 15, 4, 3]. The problem is to interpolate prescribed values and successive derivatives on a given set of points in the unit disc by means of a strictly *positive real* rational function in the unit disc. One special case is the *Carathéodory extension*, which specifies the interpolation conditions at the origin up to a number of derivatives.

Usually, there are an infinite number of interpolants fulfilling the interpolation conditions, which are parameterized by the spectral zeros of the interpolants. Given a set of spectral zeros, the interpolant can be determined by an optimization problem. Actually, any interpolant satisfying the interpolation conditions with derivatives meets our needs well, which means we do not need to preserve the spectral zeros of the original system. As a result, a special case, *central or maximum entropy solution*, is of particular interest to us, which can be determined by solving a linear system of equations instead of the optimization problem.

In this section, we propose a novel Carathéodory extension based model reduction scheme. The new method, called *CEMOR*, can generate guaranteed passive reduced models of dynamic systems with arbitrary internal structure and formulations. The reduced model will agree with the original model up to a number of moments at an expansion point. In the proposed method, we first choose an expansion point and compute the moments of the original system at that point as the interpolation conditions with derivatives. Then we transform the interpolation conditions to the discrete-time domain, obtain the central solution of Carathéodory extension, and transform the interpolant back to the continuous-time domain

as the reduced system. The proposed rational interpolation method is as efficient as Krylov subspace methods but can generate guaranteed passive reduced models for systems with arbitrary internal structure.

5.1 Carathéodory extension

In this section, we present the classical Carathéodory extension problem, which is derived in discrete-time domain, and how it is related to our reduction problem.

5.1.1 Problem statement

Given a scalar sequence $(w_0, w_1, \ldots, w_m)$, the Carathéodory extension problem with degree constraint amounts to determinating any function $f(z)$ satisfying the following three conditions:

(1) $f(z)$ fulfills the interpolation constraints:

$$\frac{f^{(k)}(0)}{k!} = w_k (k = 0, 1, \ldots, m) \tag{5.1}$$

(2) $f(z)$ is strictly positive real, i.e., f is analytic in the closed unit disc $\bar{\mathbb{D}}$, where $(\mathbb{D} = \{z : |z| < 1\})$, and $Ref(z) > 0$ for all $z \in \bar{\mathbb{D}}$,

(3) f is rational and the degree $(degf(z)) \leq m$.

There exists an interpolant for the interpolation problem with derivative constraints if

and only if a symmetric Toeplitz matrix

$$P = \begin{bmatrix} 2w_0 & w_1 & \dots & w_m \\ w_1 & 2w_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & w_1 \\ w_m & \dots & w_1 & 2w_0 \end{bmatrix} \tag{5.2}$$

is nonnegative definite.

Note that, if $(w_0, w_1, \dots, w_m)$ are the moments of a high order system in discrete-time domain, $f(z)$ can be the transfer function of a reduced system of order m and the reduced system must be passive as $f(z)$ is positive real.

5.1.2 Determination of interpolating function

The complete parameterization of the set of interpolating functions was developed by Byrnes and Lindquist [8, 7] Assume that there exists a bijective (one-to-one and onto) map between the set of pairs of real polynomials

$$\{(\alpha(z), \beta(z)) : \deg(\alpha(z)) \leq m, \deg(\beta(z)) \leq m\} \tag{5.3}$$
$$\alpha(0) \neq 0, \beta(0) \neq 0$$

where

$$f(z) = \frac{\beta(z)}{\alpha(z)} = \frac{\beta_0 + \beta_1 z + \dots + \beta_m z^m}{\alpha_0 + \alpha_1 z + \dots + \alpha_m z^m} \tag{5.4}$$

and the set of real stable polynomials

$$\{\rho(z) : \deg(\rho(z)) = m, \rho(z) \neq 0, \forall z \in \bar{\mathbb{D}}\} \tag{5.5}$$

where

$$\rho(z) = \rho_0 + \rho_1 z + \cdots + \rho_m z^m \tag{5.6}$$

where deg() gives the degree of a polynomial. A stable polynomial here means that all the roots are outside the unit circle (similar to the right-hand plane in the continuous-time domain). In fact, $\Phi(z)$, the *spectral density* of $f(z)$, is given by

$$\begin{aligned} &\Phi(z) = f(z) + f(z^{-1}) = \tfrac{\alpha(z)\beta(z^{-1})+\alpha(z^{-1})\beta(z)}{\alpha(z)\alpha(z^{-1})} \\ &= \tfrac{\rho(z)\rho(z^{-1})}{\alpha(z)\alpha(z^{-1})} \end{aligned} \tag{5.7}$$

The bijectivity implies that the roots of $\rho(z)$, which are the zeros of spectral density, so-called *spectral zeros*, are the characterizing factor. In other words, if a set of spectral zeros is assigned, $f(z)$ can be uniquely determined satisfying the three conditions at the same time. Specifically, the computation of an interpolant $f(z)$ from $\rho(z)$ amounts to an optimization problem $min_{\alpha \in S_m} J_\rho(\alpha)$

$$J_\rho(\alpha) = \boldsymbol{\alpha}^T P \boldsymbol{\alpha} - 2 < \log(\alpha(z)), \rho(z)\rho(z^{-1}) > \tag{5.8}$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_m \end{bmatrix} \in \mathbb{R}^{(m+1)\times 1} \tag{5.9}$$

and $\alpha(z)$ should be in the region S_m:

$$S_m = \{\alpha(z) : \alpha_0 + \alpha_1 z + \cdots + \alpha_m z^m \neq 0,\ \alpha_0 > 0, \forall z \in \bar{\mathbb{D}}\} \tag{5.10}$$

where $< f(z), g(z) >= \frac{1}{2\pi}\int_{-\pi}^{\pi} f^*(e^{i\theta})g(e^{i\theta})d\theta$ defines the inner product of two functions $f(z)$ and $g(z)$. If the coefficients are real, $f^*(z) = f(z^{-1})$, which is the case for our problem.

5.1.3 Maximum entropy solution

Now, let us consider the *central or maximum entropy solution*, i.e. the special case of the problem in which all the spectral zeros are assigned at infinity. In this special case, $\rho(z)\rho(z^{-1}) = 1$, and hence $< \log(\alpha(z)), \rho(z)\rho(z^{-1}) >= \log(\alpha_0)$. Consequently, the objective function becomes

$$J_1(\alpha) = \boldsymbol{\alpha}^T P \boldsymbol{\alpha} - 2\log(\alpha_0) \tag{5.11}$$

Since the Toeplitz matrix P is positive definite, J_1 is strictly convex. Hence, there is at most one minimum. To determine this possible minimum, set the gradient equal to zero to obtain

$$\nabla J_1 = \begin{bmatrix} 2w_0 & w_1 & \dots & w_m \\ w_1 & 2w_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & w_1 \\ w_m & \dots & w_1 & 2w_0 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} - \begin{bmatrix} 1/\alpha_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0 \tag{5.12}$$

which can be written as $P\boldsymbol{\alpha} = \boldsymbol{e}\frac{1}{\alpha_0}$, where $\boldsymbol{e} = [1, 0, \dots, 0]^T$. Note that, since $\alpha_0 = \boldsymbol{e}^T\boldsymbol{\alpha} = \boldsymbol{e}^T P^{-1}\boldsymbol{e}\frac{1}{\alpha_0}$, we have

$$\alpha_0 = \sqrt{\boldsymbol{e}^T P^{-1}\boldsymbol{e}} \tag{5.13}$$

With α_0, $\boldsymbol{\alpha}$ (defined in (5.9)) is the unique solution of a linear system of equations (5.12) because P is positive definite.

In fact, there is an alternative way to compute $\alpha(z)$, which is less expensive. Defining $\varphi_i = \alpha_i/\alpha_0$ for $i = 1, 2, \ldots, m$, we obtain

$$\begin{bmatrix} 2w_0 & w_1 & \ldots & w_{m-1} \\ w_1 & 2w_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & w_1 \\ w_{m-1} & \ldots & w_1 & 2w_0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_m \end{bmatrix} = - \begin{bmatrix} w_1/\alpha_0 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \tag{5.14}$$
$$\begin{bmatrix} w_0 & w_1 & \ldots & w_m \end{bmatrix} \begin{bmatrix} 1 & \varphi_1 & \ldots & \varphi_m \end{bmatrix}^T = \frac{1}{a_0^2}$$

Hence, we obtain the well-known *normal equation*, which can be solved quickly using the *Leyinson algorithm* [46] at the cost of $O(m^2)$.

Finally, given $\alpha(z)$, $\beta(z)$ can be solved by

$$\alpha(z)\beta(z^{-1}) + \alpha(z^{-1})\beta(z) = \rho(z)\rho(z^{-1}) \tag{5.15}$$

where $\rho(z)\rho(z^{-1}) = 1$ due to the maximum entropy solution. Identifying coefficients of the same power in z, we can come up with the following linear equations

$$\left(\begin{bmatrix} \alpha_0 & \alpha_1 & \ldots & \alpha_m \\ \alpha_1 & \ldots & \alpha_m & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \alpha_m & 0 & \ldots & 0 \end{bmatrix} + \begin{bmatrix} \alpha_0 & \alpha_1 & \ldots & \alpha_m \\ 0 & \alpha_0 & \ldots & \alpha_{m-1} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \alpha_0 \end{bmatrix} \right) \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5.16}$$

which can also be written as $(H_\alpha + T_\alpha)\boldsymbol{\beta} = \boldsymbol{e}$. The first matrix H_α in the left-hand side is a Hankel matrix whose first column is $\boldsymbol{\alpha}$ and whose elements are zero below the first anti-diagonal. The second matrix T_α in the left-hand side is a Toeplitz matrix whose first row is $\boldsymbol{\alpha}$ and whose elements are zero below the diagonal. By solving the linear equations, $\beta(z)$

can be obtained from $\alpha(z)$ and we obtain the interpolant $f(z)$, which can be the transfer function of a reduced model in the discrete-time domain.

5.2 New reduction method: CEMOR

We present the new model reduction method based on single-input single-output system $h(s)$. However, the results can be generalized to multi-input multi-output case as shown in the next section. We first give overall flow of the algorithm and then present the important steps in detail.

5.2.1 Algorithm flow

In CEMOR, steps 1 to 3 compute the moments from the original circuits and perform the scaling as required by the new method. Step 4 and Step 10 transform the information between continuous and discrete-time domains. Steps 6 to 8 compute the reduced model $f(z)$ in the discrete-time domain.

5.2.2 Frequency scaling

We choose a positive real expansion point $\sigma \in R^+$. Typically, the point chosen in higher dynamic frequency range will result in a more compact model. Now we normalize the expansion point to 1 by frequency scaling

$$h(s) = D + C(sE - A)^{-1}B = D + C(\frac{s}{\sigma}(\sigma E) - A)^{-1}B \tag{5.17}$$

ALGORITHM: CEMOR
Input: $h : (E, A, B, C, D)$
Output: $h_r : (E_r, A_r, B_r, C_r D_r)$

1. Choose an expansion point $\sigma \in R^+$
2. Perform frequency scaling to normalize the expansion point to 1 by (5.19)
3. Generate moments at normalized expansion point 1 by (5.21)
4. Transform moments to discrete-time domain by (5.25)
5. Compute Toeplitz matrix using (5.2)
6. Check P for nonnegative-definite property. If not, go back to step 1 with a new expansion point σ or go back to step 5 to get a Toeplitz matrix with less moments
7. Solve (5.12) or (5.14) to obtain $\alpha(z)$
8. Solve (5.16) to obtain $\beta(z)$
9. Realize $f(z)$ by any canonical form
10. Transform interpolant back to continuous-time domain by (5.28)
11. Perform reverse frequency scaling by (5.29)

Figure 5.1: The CEMOR algorithm flow.

which results in the following system with expansion point $\tilde{\sigma} = 1$

$$h(\tilde{s}) = D + C(\tilde{s}\tilde{E} - A)^{-1}B \tag{5.18}$$

where $\tilde{s} = s/\sigma$ and

$$\tilde{E} = \sigma E \tag{5.19}$$

Scaling the expansion point to 1 will ensure a good numeral condition in the bilinear transformation process.

5.2.3 Moment generation and passive condition

Given the dynamic system (5.18) and the expansion point $\tilde{\sigma} = 1$, defining

$$\mathcal{A} = (A - \tilde{\sigma}\tilde{E})^{-1}\tilde{E} \quad \mathcal{R} = -(A - \tilde{\sigma}\tilde{E})^{-1}B \tag{5.20}$$

the moments at the expansion point have the following formula

$$m_0 = C\mathcal{R} + D \quad m_i = C\mathcal{A}^i\mathcal{R}(i \geq 1) \tag{5.21}$$

For the reduced model $h_r(s)$, we require

(1) $h_r(s)$ fulfills the interpolation constraints:

$$\frac{h_r^{(k)}(\tilde{\sigma})}{k!} = m_k(k = 0, 1, \ldots, r) \tag{5.22}$$

(2) $h_r(s)$ is strictly positive real, i.e., $h(s)$ is analytic in the right-hand plane $\{s : Re(s) \geq 0\}$ and $Re(h_r(s)) > 0$ for all $\{s : Re(s) \geq 0\}$,

(3) $h_r(s)$ is rational and deg$(h_r(s)) \leq r$.

If $h_r(s)$ satisfies the three conditions, it is the transfer function of the desired reduced model, which is passive and accurate to the rth moment of the original transfer function.

5.2.4 Transformation to discrete-time domain

The Carathéodory extension problem considered in the previous section is assumed to find the mapping $f(z)$ from the unit disc onto the right half-plane. In the continuous-time domain, we need to find a transfer function $h_r(s)$ from the right half-plane to right half-plane. In this case, we need to transform the interpolation data (interpolation point and

moments) from right-hand plane to unit disc first.

So we need a bilinear transformation, which maps the right-half plane $\{s : Re(s) > 0\}$ to the unit disc $\{z : |z| < 1\}$ and maps interpolation point $s = 1$ to $z = 0$. The following bilinear transformation will achieve this [4].

$$z(s) = \frac{-s+1}{s+1} \quad s(z) = \frac{-z+1}{z+1} \tag{5.23}$$

Under the bilinear transformation, we have $f(z) = h_r(s(z))$ and $f(0) = h_r(1)$. Given

$$\frac{f^{(k)}(0)}{k!} = w_k \quad \frac{h_r^{(k)}(1)}{k!} = m_k \tag{5.24}$$

for $(k = 0, 1, \ldots, r)$, the derivatives $f^{(k)}(0)$ is related to the derivatives $h_r^{(k)}(1)$ as follows

$$\begin{aligned} w_0 &= m_0 \\ w_k &= \frac{1}{k!} \sum_{l=1}^{k} \binom{k}{l} m_{k-l+1}(k-l+1)!(s^{(1)}(0))^{k-l} s^{(l)}(0) \end{aligned} \tag{5.25}$$

The coefficients $\binom{k}{l}$ are binomial coefficients, which fulfill the recursive formula

$$\begin{aligned} \binom{k}{l} &= 1(l = 1, k) \\ \binom{k}{l} &= \frac{2k-l}{l}\binom{k-1}{l-1} + \binom{k-1}{l}(1 < l < k) \end{aligned} \tag{5.26}$$

The term $s^{(l)}(0)$ is obtained by

$$s^{(l)}(0) = 2(-1)^1 l! \tag{5.27}$$

Now we can use the Carathéodory extension method in the previous section to obtain $f(z)$

from $w_0, \ldots, w_r$.

5.2.5 Transformation back to continuous-time domain

Given the transfer function $f(z)$, it can be realized by any canonical state-space form (A_f, B_f, C_f, D_f). which can be transformed back to continuous-time domain by the following transformation derived from (5.23) [41]

$$\begin{aligned} A_r &= (I - A_f)(I + A_f)^{-1} \quad & B_r &= -2(I + A_f)^{-1} B_f \\ C_r &= C_f (I + A_f)^{-1} & D_r &= -C_f (I + A_f)^{-1} B_f + D_f \end{aligned} \tag{5.28}$$

So the reduced model is given by $(E_r, A_r, B_r, C_r, D_r)$, where

$$E_r = \frac{1}{\sigma} I_r \tag{5.29}$$

is the inverse process of the frequency scaling in (5.19).

5.2.6 Complexity analysis

Note that, the cost to generate moments is $O(n^\alpha)$ which is similar to Krylov subspace methods (α depends on the sparsity of the system and $1 < \alpha < 2$ for most cases of interest). The rest procedures only require $O(r^3)$, where r is the order of reduced system. Since $r << n$, the cost is dominated by $O(n^\alpha)$.

5.3 Extension to MIMO systems

Given a interpolation point $z_0 = 0$ and a set of m matrix-valued interpolation values $(W_0, W_1, \ldots, W_m) \subset \mathbb{R}^{p\times p}$, where W_0 is assumed to be symmetric (nonsymmetric case can be transformed to be symmetric as shown at the end of the section), the matrix-valued Carathéodory extension problem with degree constraint amounts to determinating any function $F(z)$ of order $r = mp$ fulfilling the interpolation constraints

$$\frac{1}{k!}F^{(k)}(0) = W_k(k = 0, 1, \ldots, m) \tag{5.30}$$

$F(z)$ is strictly positive real, i.e., $F(z)$ is analytic in the closed unit disc $\bar{\mathbb{D}}$ and

$$Re(F(z)) = \frac{1}{2}(F(z) + F(z^{-1})^T) > 0 \tag{5.31}$$

for all $z \in \bar{\mathbb{D}}$. There exists an interpolant for the interpolation problem with derivative constraints if and only if a symmetric block Toeplitz matrix

$$\mathbf{\Pi} = \begin{bmatrix} W_0 + W_0^T & W_1^T & \ldots & W_m^T \\ W_1 & W_0 + W_0^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & W_1^T \\ W_m & \ldots & W_1 & W_0 + W_0^T \end{bmatrix} \tag{5.32}$$

is nonnegative definite. In the MIMO case, the cost function (5.8) is generalized as

$$J_P(\alpha) = trace(\mathbf{R}^T\mathbf{\Pi}\mathbf{R}) - 2 < \log(\det(R(z))), \rho(z)\rho(z^{-1}) > \tag{5.33}$$

where

$$\mathbf{R} = \begin{bmatrix} R_0 \\ \vdots \\ R_n \end{bmatrix} \in \mathbb{R}^{p(m+1)\times p} \tag{5.34}$$

and R_0 is assumed to be upper triangular matrix. $R_0, \ldots, R_n$ are coefficients of the $p \times p$ matrix polynomial

$$R(z) = R_0 + R_1 z + \cdots + R_n z^n \tag{5.35}$$

which is the generalization of $\alpha(z)$ in scalar case. Similarly, as for central solution, where $\rho(z)\rho(z^{-1}) = 1$, the objective function becomes

$$J_1(\alpha) = trace(\mathbf{R}^T \mathbf{\Pi} \mathbf{R}) - 2\log(\det(R_0)) \tag{5.36}$$

Setting the gradient of J_1 equal to zero, we obtain

$$\mathbf{\Pi R} = \mathbf{E} R_0^{-T} \tag{5.37}$$

where $\mathbf{E} = [I, 0, \ldots, 0]^T$. Note that, since $R_0 = \mathbf{E}^T \mathbf{R} = \mathbf{E}^T \mathbf{\Pi}^{-1} \mathbf{E} R_0^{-T}$, we have

$$R_0 R_0^T = \mathbf{E}^T \mathbf{\Pi}^{-1} \mathbf{E} \tag{5.38}$$

By performing Cholesky factorization, we can obtain R_0. With R_0, $\mathbf{R}$ is the unique solution of a linear system of equations (5.37) because $\mathbf{\Pi}$ is positive definite. Similar to the scalar case (5.14), the coefficients of $R(z)$ can also be solved by a matrix-version of the Levinson algorithm.

The spectral density of $F(z)$ can be factorized as follows

$$\Phi(z) = F(z) + F(z^{-1}) = V(z^{-1})^T V(z) \tag{5.39}$$

which is the generalization of (5.7) in scalar case. $V(z) = \rho(z)R(z)^{-1}$ is the *spectral factor* of $\Phi(z)$. For maximum-entropy solution, we have $\rho(z) = 1$ and $V(z) = R(z)^{-1}$, which can be realized by canonical form. Given any minimal realization of $V(z)$

$$V(z) = zC_v(I - zA_v)^{-1}B_v + D_v \tag{5.40}$$

there is a unique $F(z)$ satisfying (5.39) [3]

$$F(z) = 2z(B_v^T X A_v + D_v^T C_v)(I - zA_v)^{-1}B_v + B_v^T X B_v + D_v^T D_v \tag{5.41}$$

where X is the unique solution to the Lyapunov equation

$$A_v^T X A_v - X + C_v^T C_v = 0 \tag{5.42}$$

Note that instead of $O(n^3)$, the cost of the Lyapunov equation here is $O((mp)^3)$, which is not expensive because $pm \ll n$.

For nonsymmetric W_0, perform SVD on W_0 as $W_0 = USZ^T$ and transform W_i to $\tilde{W}_i$ by $\tilde{W}_i = U^T W_i Z$ such that $\tilde{W}_0$ is symmetric. After obtaining the interpolant $\tilde{F}(z)$ from $\tilde{W}_i$, the interpolant $F(z)$ from W_i can be obtained as $F(z) = U\tilde{F}(z)Z^T$.

5.4 Experimental results

The proposed method has been implemented in Matlab 7.0 and tested on an Intel quad-core workstation with 16GB memory. The example is from [51], which is an RLC ladder network of order $n = 201$. The state variables are as follows: In general, n is odd and x_{2i-1} is the voltage across capacitor C_i for $i = 1, 2, \ldots, (n+1)/2$, while x_{2i} is the current through inductor L_i for $i = 1, 2, \ldots, (n-1)/2$. Two resistors R_1 and R_2 are placed at either end of the ladder as shown in Fig. 5.2. for an order $n = 5$ example. The input is the voltage source ant the output is the port current. All the capacitors and inductors have unit vale while $R_1 = 1/2$, $R_2 = 1/5$. For general model of order n, the state-space equation has the following form

$$A = \begin{bmatrix} -2 & 1 & 0 & \ldots & 0 & 0 \\ -1 & 0 & 1 & \ldots & 0 & 0 \\ 0 & -1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & \ldots & -1 & -5 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 2 \end{bmatrix} \tag{5.43}$$

$$C = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & -2 \end{bmatrix}, D = 1$$

As in [51], the reduced order is also chosen to be 20 and the expansion point is $1Hz$.

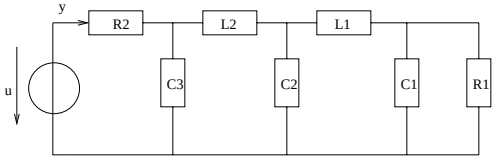

Figure 5.2: RLC ladder of order 5

The performance is compared with positive-real TBR (PRTBR) reduced model of the same order. As shown in Fig. 5.3, CEMOR is much more accurate than PRTBR in the frequency range of interest from $0.01Hz$ to $10Hz$. Although PRTBR is more accurate beyond $10Hz$, the error for CEMOR is less than $-100dB$ beyond $10Hz$, which can be ignored. In fact, compared with the results shown in [51], CEMOR is also better than the methods via interpolation of spectral zeros [51] in this example.

We remark that PRTBR in theory has global error bound and can be applied to system with any internal structure as the positive real property is explicitly enforced in the Lur'e equations. But practically, in addition to the high cost, this method has very stringent demands on the numerical condition of system matrices and are tricky to implement in a stable way [44]. As a result, the performance in practice may not be as good as expected. CEMOR can be viewed as a PRIMA-like reduction technique for general structure systems. It is extremely efficient and much less demanding on the numerical condition of the examples at the cost of lacking the global error bound.

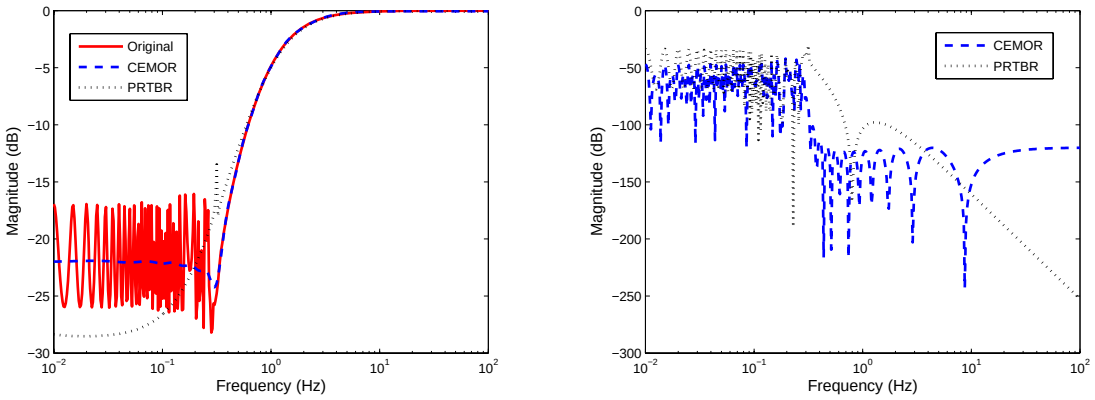

Figure 5.3: Frequency responses.

Finally, we compare the CPU time of CEMOR and PRTBR. The reduction CPU times are shown in Table 5.1, where the n is the order of the ladder and the reduced order is 20.

Table 5.1: Reduction CPU time comparison of CEMOR and PRTBR (seconds).

n	CEMOR	PRTBR	Speedup
201	0.438	3.844	8.776
401	0.687	31.688	46.125
1001	3.556	560.988	157.745
2001	7.031	11717.136	1666.494

From Table 5.1, we can see, PRTBR is very slow and infeasible for large-scale applications while CEMOR is very efficient. The speedup will go up for larger circuits. Actually for circuits larger than 4K, PRTBR can't finish in reasonable time.

Chapter 6

Passive modeling of interconnects by waveform shaping

Although projection based MOR methods are very successful, their applications are mainly limited to RLC circuits. Many high-speed circuits, like RF surface acoustic wave (SAW) filters, spiral inductors, high-speed transmission lines, are still modeled by using measured data (like Scattering parameters) due to many high frequency effects and the freqency dependency of circuit parameters. Another issue with projection based MOR methods is that they become very inefficient for reducing circuits with many terminals in terms of both computational costs and reduced model sizes [26]. The main reason for lost efficiency lies in the fact that with more terminals, more transfer functions are needed to compute and more poles will be used for each increased order of block moments, which is not necessary as polese are system information and should not depend on the terminals.

For generating general-purpose compact models from many measured and simulated data, fitting methods based on least square rational approximation in frequency domain are still widely used [23]. One critical issue in such a modeling process is to preserve the pas-

sivity of the original system in the reduced models. Existing approaches like PRIME [39] enforce the passivity by physically realizing each pole/residue (conjugate pole pair) term in the fractional form using Foster's synthesis method. If a pole/residue term can't be realized, it is discarded. As a result, PRIME can lead to very large errors. The latest approach to this general passivity enforcement problem is based on the convex programming (CP) approach by using the state-space representation of the system [12]. The passivity is enforced by using semi-definite constraints during a semi-definite (convex) optimization. But the CP based method suffers very high computational costs and can optimize circuits with less than about 20 poles and 20 terminals in a typical computation setting (on latest Intel Pentium 4 CPU with 1GB memory).

In this section, we propose a new passivity enforcement approach for general purpose modeling of passive linear circuits. Our new method is based on the observation that most of interconnect circuits like clock trees, substrate, packing, RF passives, and transmission lines are lossy and their frequency responses behave like a band-pass or low-pass filter in general. As a result, the models for those passive systems need not to be passive for all frequencies, as required by traditional passivity enforcement methods. Practically they need only to be passive for a limited bandwidth in which most of the signal energy is concentrated.

Instead of making the reduced models passive for all frequencies, the new method works on the signals going into the reduced models to enforce the passivity. The idea is to slightly shape the waveforms of the signals such that the resulting spectra are bandlimited to the frequency range in which the reduced system is passive. As a result, the reduced models only need to be band-limited passive, which we call conditionally passive in this paper and can be achieved much easier than traditional passivity for a reduced sys-

tem. We propose two approaches to band-limit (shape) the waveforms. The first method is based on frequency domain fast Fourier transform (FFT) and inverse FFT to explicitly shape the waveforms. The second method is based on insertion of passive low-pass filters (LPF) into the reduced models to implicitly shape the waveforms. For the second method, we analyze the delay and distortion effects introduced by using low-pass filters and propose methods to mitigate the delay effects. Experimental results on several interconnect circuits demonstrate the effectiveness of the proposed methods.

6.1 Conditional passivity and conditional positive-realness

In this section, we analyze the relationship of a system's transient responses and its input signals in terms of passivity. We show that a non-passive system can still behave like a passive system when its input signals are band-limited. Such systems can be defined as the conditionally passive and its network functions are conditionally positive-real.

Passivity is an important property of many physical systems. A passive network does not generate energy. If the reduced order model (ROM) loses its passivity, it may lead to unbounded responses in transient simulation, which means new energy has been generated in this network.

Fig. 6.1 shows a transient simulation result of a non-passive circuit under a sinusoidal excitation.

O. Brune [6] has proved that the admittance and impedance matrix of an electrical circuit consisting of an interconnection of a finite number of positive R, positive C, positive L, and transformers are passive if and only if their rational functions are positive real. A

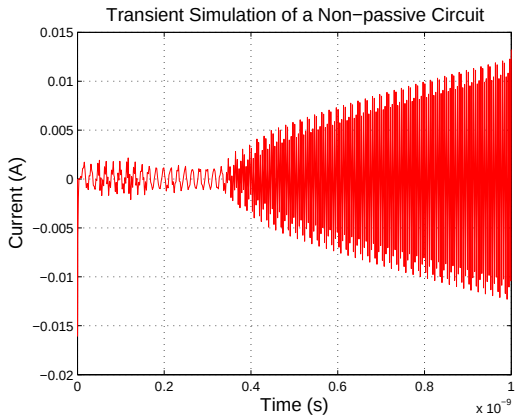

Figure 6.1: Transient response of a non-passive circuit.

network with admittance matrix function $\mathbf{Y}(s)$ is said to be positive real iff

(1) $\mathbf{Y}(s)$ is analytic, for $Re(s) > 0$

(2) $\overline{\mathbf{Y}(s)} = \mathbf{Y}(\bar{s})$, for $Re(s) > 0$

(3) $\mathbf{Y}(s) + \mathbf{Y}(s)^H \geq 0$, for $Re(s) > 0$

Condition (1) means that there are no unstable poles (poles lying on right-half-plane (RHP) in s-domain). Condition (2) refers to system that has real response. And condition (3) is equivalent to the real part of $\mathbf{Y}(s)$ having a positive semi-definite matrix at all frequencies. In other words, the real parts of all the eigenvalues of the $H(s)$ must be equal to or larger than zero. But condition (3) is difficult to satisfy as it requires the checking of frequency responses from DC to infinity.

We know for a passive system, its admittance matrix $Y(s)$ needs to be positive real (Re$\{Y(s)\}$ is positive definite) for all frequencies. However, when the $Y(s)$ is not positive real for some frequency ranges, will the system always exhibit non-passive behavior in the time domain as shown in Fig. 6.1? Actually the answer depends on the spectrum (energy) of the input signal. If the input signal is band-limited to the frequency range where the

reduced model is positive real, then the system will still behave passively as the original system.

This can be illustrated by the following example. Fig. 6.2 shows the frequency responses of a RLC circuit and its reduced model. The two circuits match well below 15Ghz. Above 60Ghz, the real part of the transfer function of the reduced model becomes negative as shown in Fig. 6.2(a), which means the system becomes non passive. When a sinusoidal input signal of 10Ghz is applied to both systems, we get the exact responses in the time domain as shown in Fig. 6.3. However, if we apply a sinusoidal signal of 60Ghz, the time-domain responses of the original system and the reduced system will be dramatically different as shown in Fig. 6.4. The response of the reduced system actually explodes.

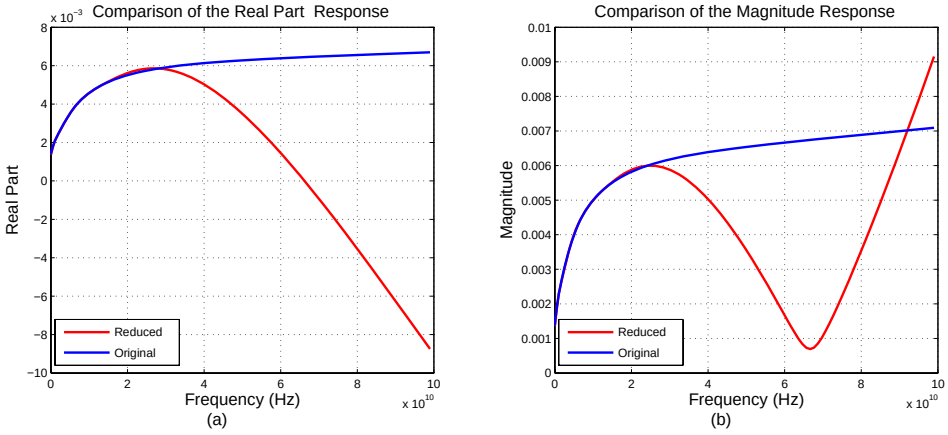

Figure 6.2: Frequency responses of a reduced model and its original RC circuit.

We note that many ideal input signals like Dirac delta function $\delta(t)$, unit step function $u(t) = 1, t \geq 0, u(t) = 0, t < 0$, and unit ramp function $f(t) = t, t \geq 0, f(t) = 0, t < 0$, have an infinite spectrum of frequencies. For example, for Dirac delta function, $L(\delta(t)) = 1$, where $L(X)$ means taking the Laplace transform of function X. So $\delta(t)$ has a constant spectrum for all frequencies and it can easily make any non-passive system to exhibit the non-passive behavior as shown in [28]. The Laplace transform of unit step function and unit

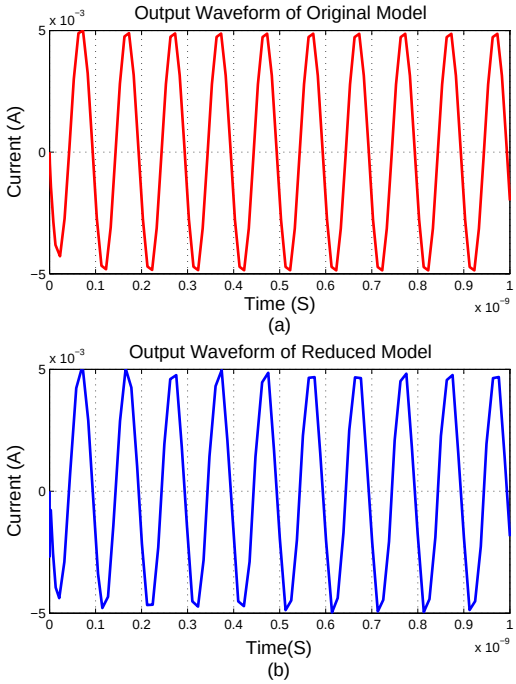

Figure 6.3: Transient responses of a reduced model and its original RC circuit.

ramp function is $1/s$ and $1/s^2$ respectively. When those signals are applied to a non-passive system, non-passive behavior can be easily observed as shown in [40].

However, such ideal signals do not exist in the real world. Most of the active transistors, passive interconnects, RF passive components, and transmission lines exhibit limited bandwidth due to unavoidable capacitive loss, which implies that signal generated by and propagated through those systems will bear limited bandwidth. This situation will become worse as we move to the deep sub 100nm technology. So for realistic signals, we can build a reduced system which is only passive for the given frequency range and the resulting system will still be passive as far as the simulation is concerned. For this purpose, we introduce the conditional passivity and conditional positive-realness.

A network with admittance matrix function $\mathbf{Y}(s)$ is said to be conditionally positive

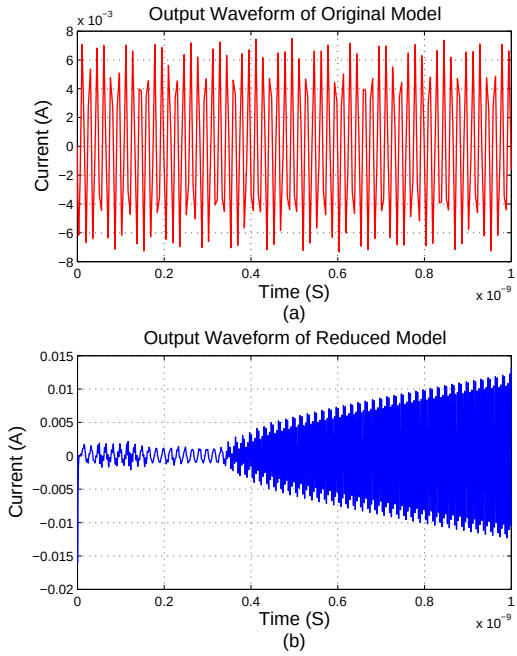

Figure 6.4: Transient responses of a reduced model and its original RC circuit.

real iff

(cpr1) $\mathbf{Y}(s)$ is analytic, for $Re(s) > 0$

(cpr2) $\overline{\mathbf{Y}(s)} = \mathbf{Y}(\bar{s})$, for $Re(s) > 0$ $\qquad 0 \leq Im(s) \leq 2\pi f_{max}$

(cpr3) $\mathbf{Y}(s) + \mathbf{Y}(s)^H \geq 0$, for $Re(s) > 0$

In other words, $\mathbf{Y}(s)$ will be positive real for the given frequency range $[0, f_{max}]$.

The main benefit for a reduced system to be conditionally passive is that conditional passivity can be much easier to achieve than strict passivity. Many existing frequency domain rational fitting methods [23, 36] can be used to do this with much more scalable computational costs than the convex programming method [12]. On the other hand, we put more constraints on the signals driving the conditionally passive systems: we need to make sure that the signal spectrum is band limited such that its bandwidth is within the positive

real bandwidth of the reduced system. In the following section, we present two methods to achieve this requirement.

6.2 Passivity enforcement by waveform shaping

In this section, we discuss two methods to band-limit a signal by slightly shaping its waveform. Note that based on the Fourier transform, if a signal is finite in time, its spectrum extends to infinity frequency, and if its bandwidth is finite, its duration is infinite in time. For a practical non-periodic time-limited signal like switching currents in the signal lines due to transistor switching, one can never band limit such a signal from a strictly mathematical point of view. But practically we can make the out-of-band frequency energy sufficiently small compared to the in-band frequency energy such that the out-of-band energy will not stimulate the non-passive behavior of the system.

6.2.1 FFT and IFFT based waveform shaping

The first method is based on the fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT). The idea is to first transform the original transient signal into the frequency domain. Since in FFT (or discrete Fourier transform, DFT), we treat the non-periodic signal as a periodic signal, the resulting signal's spectrum becomes discrete. Then we truncate those frequencies beyond f_{max}, which is given. After this, we perform the inverse FFT on the truncated spectrum to get the time domain waveform of the shaped signal (we only take the waveform in one period). The whole process is illustrated in Fig. 6.5 and the algorithm is outlined in Fig. 6.6.

Fig. 6.7(a)and Fig. 6.7(b) show a ramp signal and its spectrum. The shaped waveform

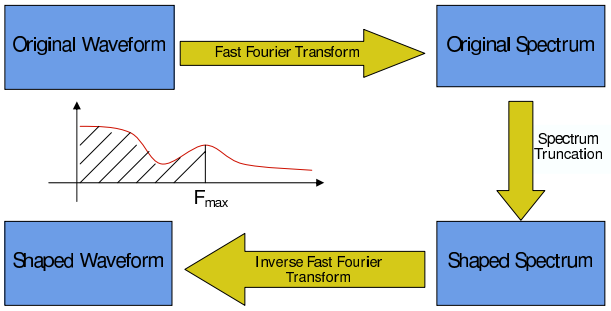

Figure 6.5: The algorithm flow of FFT and IFFT based waveform shaping.

FFT_IFFT_WAVEFORMSHAPING( ) {
Sample input data with F_s;
Fast Fourier Transform
$X_0(k) = \sum_{j=1}^{N} x_0(j) w_N^{(j-1)(k-1)}$;
Spectrum Truncation
If $f_k < f_{max}$ or $f_k > F_s - f_{max}$,
$X_1(k) = X_0(k)$;
If $f_{max} < f_k < F_s - f_{max}$,
$X_1(k) = 0$;
Inverse Fast Fourier Transform
$x_1(j) = (1/N) \sum_{k=1}^{N} X_1(k) w_N^{-(j-1)(k-1)}$;
return vector: x_1 of length N;
}

Figure 6.6: The algorithm of FFT and IFFT based waveform shaping.

with the cut-off frequency $f_{max} = 10Ghz$ and the corresponding truncated spectrum are shown in Fig. 6.7(c) and Fig. 6.7(d). The shaped waveform with the cut-off frequency $f_{max} = 2Ghz$ and the corresponding truncated spectrum are shown in Fig. 6.7(e) and Fig. 6.7(f). In general, the spectrum truncation does not change significantly the waveform characteristics like delay and slew etc. As we truncate high frequency components, the shaped waveform shows some undershoots and overshoots in Fig. 6.7(e). Those small undershoots and overshoots do not affect delay and timing of the shaped waveform when it propagates through the reduced model. If we truncate the spectrum at a higher frequency such as 10GHz, we find that the resulting waveform is almost the same as the original

one, which is shown in Fig. 6.7(c). This demonstrates that if the cutoff frequency is high enough, the distortion caused by truncating can be tolerated.

The drawback of the explicit waveform shaping method using FFT and IFFT is that it takes extra computational costs to process the signals. The computational costs are $O(nlog_2(n))$, where n is the number of sampling points.

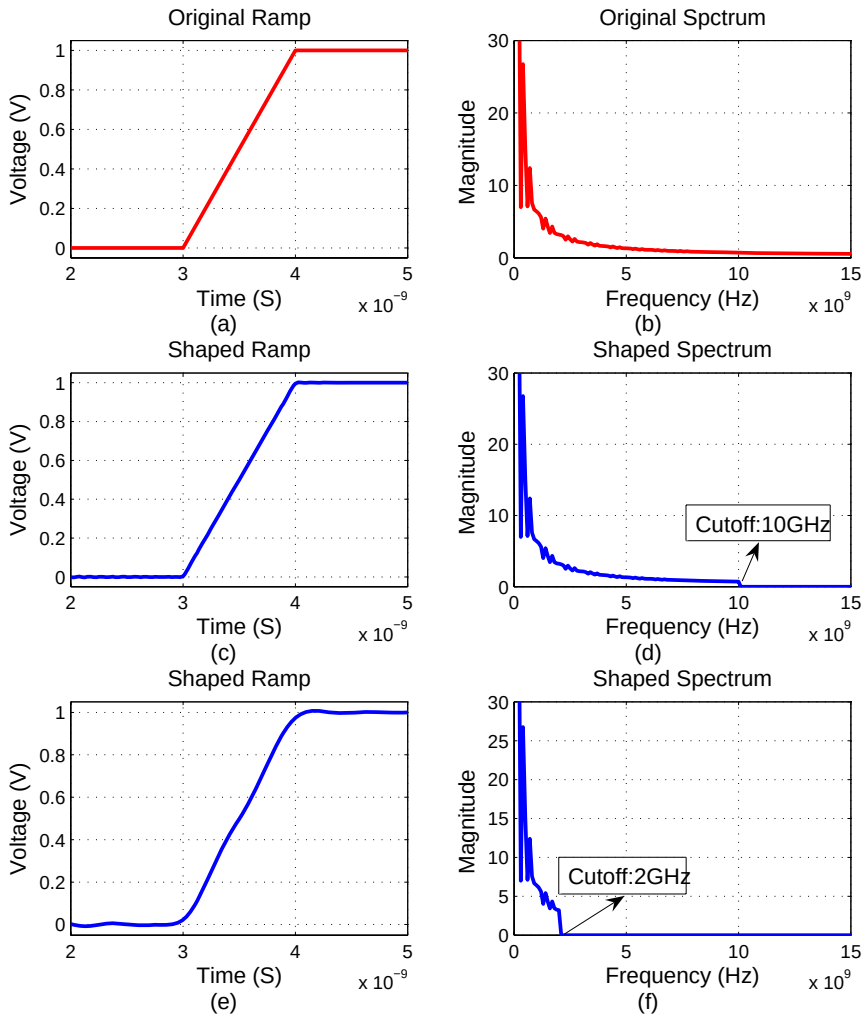

Figure 6.7: A ramp signal shaped at different frequencies.

6.2.2 Low-pass filter based waveform shaping

The second method is based on the implicit waveform shaping by adding passive low-pass filters between the input signal and the reduced system as shown in Fig. 6.8. In this way, we guarantee that the signals through the reduced system are band-limited.

Notice if we have a few input terminals (as for many interconnect circuits like clock trees or clock meshes), adding a few filters at those terminals will not increase the sizes of the reduced models significantly.

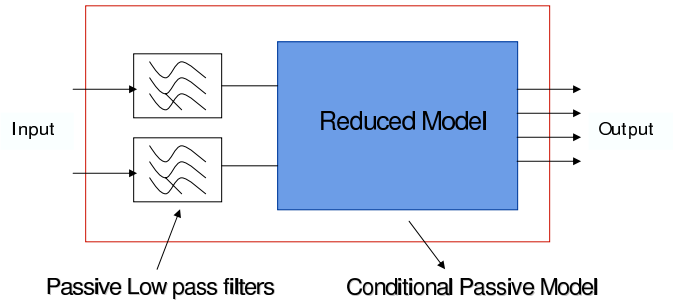

Figure 6.8: Low pass filter based waveform shaping.

Since the filter can be passively realized by LC ladder, it can be combined with the reduced model to function as a passive model. Therefore, we can conveniently use this new model in current simulation software such as Spice.

However, we need to look at several issues associated with this method before we use it. First, the low-pass filter can also distort the input signals as different frequency components may be delayed differently. Second, the introduction of low-pass filter can introduce delay. In the following, we discuss methods to mitigate those two problems.

Mitigation of distortion problems

The phase function and the resulting group delay function of a filter have profound time domain ramifications as they have a direct effect on the waveform shape of the output

signals. As a result, we choose Bessel filter family due to its good time domain property. A Bessel filter has a linear phase characteristic over the pass-band of the filter, which implies a constant time delay over the pass-band of the filter (see Fig. 6.9) so that the phase distortion in the filtering process can be avoided. From Fig. 6.9(a), we can see a constant time delay from DC to the normalized frequency 1 when the order (n) of filter is higher than 3. In addition, its step response exhibits negligible overshoot and ringing.

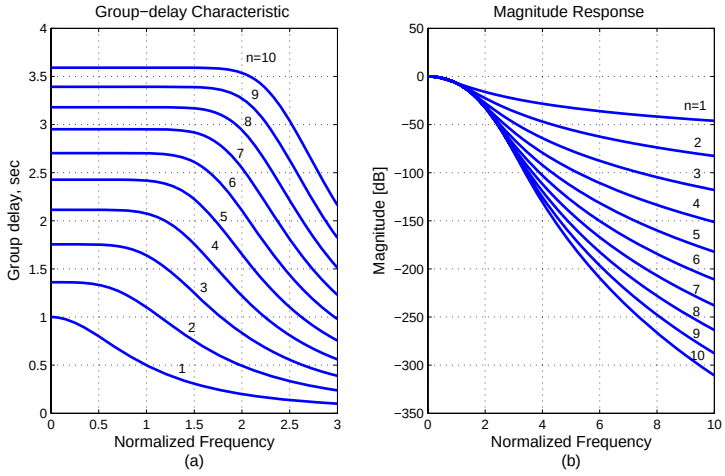

Figure 6.9: Group-delay characteristic and magnitude response for different order Bessel filters (normalized frequency).

However, a gradual roll-off (longer decay range) is the price we have to pay for a good time domain property. Fortunately, we can compromise it by increasing the order of the filter (see Fig. 6.9(b)) at the cost of larger reduced models. Another way is to increase the passive frequency range so that the filter has sufficient reduction of spectrum (again at the cost of larger reduced models).

Mitigation of delay problems

Another issue we have to take into consideration is the time delay caused by the filter. Three factors can influence the time delay: the prototype, the order, and the cutoff frequency of

a filter. Among them, the cutoff frequency is the dominant factor because the delay is inversely proportional to the cutoff frequency of a filter.

$$\text{Actual Delay} = \frac{\text{Normalized Delay}}{\text{Actual Corner Frequency (fc)}} \tag{6.1}$$

For example, for the eighth order Bessel filter, the normalized delay is 2.703s. If the cutoff frequency is as high as 20GHz, the actual delay could be as small as 0.135ns.

Hence, if the cutoff frequency is sufficiently high, the group delay caused by the filter can be made sufficiently small compared to the delay of the original circuit so that such a delay can be ignored.

6.3 Experimental results

In this section, we present some experimental results on two interconnect circuits from our industry partner. All the experimental results are conducted on a computer with AMD Athlon(64) 3800+ 2.41Ghz CPU and 500MB DDR memory. The conditional passivity is achieved by using the minimum square fitting method on the required transfer functions with poles computed from projection based methods like PRIMA. This fitting method can make the reduced models accurate to the given maximum frequency and ensure the passivity of the models in the given frequency range.

The first example is a RC circuit with 210 nodes and 3 terminals. In this experiment, we use a steep square waveform as the input signal, as shown in Fig. 6.10(a). We apply this signal to the original model, the reduced model, and the LPF (low-pass filter) based reduced model. The output waveforms of these three models are shown in Fig. 6.10(b)(c)(d), respectively.

The reduced model is conditionally passive: the passivity of the model can only be preserved at the frequency range from DC to 15GHz. Since a steep square waveform contains high frequency components beyond this range, we can observe the erratic time-domain behavior caused by energy generated at high frequencies, as shown in Fig. 6.10(c).

However, by eliminating those high frequency components by LPF, the output waveform of the LPF based reduced model (Fig. 6.10(d)) matches the output waveform of the original model (Fig. 6.10(b)) with little discrepancy. Therefore, the LPF based reduced model can function as a passive model at all frequencies.

In addition, we compare the qualities of Bessel LPF based reduced model and Ellipse LPF based reduced model in Fig. 6.11. The Fig. 6.11(a) and Fig. 6.11(b) show the transient responses from the original circuit due to the square input waveform. Fig. 6.11(c) and Fig. 6.11(d) show the transient responses from Bessel LPF based reduced model while Fig. 6.11(e) and Fig. 6.11(f) are the transient responses from Ellipse based reduced models using the same filter order. So the results clearly show that the Bessel LPF reduced model is superior to the Ellipse based models. As shown in (Fig. 6.11(d)(f)), Bessel LPF based reduced model can effectively avoid the overshoots and ringings. This result further demonstrates the rational of our choice for Bessel LPF over other types of LPFs.

The second example is a RC circuit (168 nodes) with 132 terminals (14 drivers and 118 receivers). This circuit does not have much to reduce due to large number of terminals compared to the number of nodes. But it serve as an example that the convex programming method fails to optimize due to the large terminal count. Still we use fitting method to do the frequency domain reduction and make the reduced models accurate to 50Ghz. We use a steep square waveform as shown in Fig. 6.12(a) as the inputs. We apply this signal to the original model, the reduced model, and the LPF based reduced model. The output

waveforms of these three models are shown in Fig. 6.12(b)(c)(d), respectively.

By eliminating high frequency components by LPF. The results are similar: output waveforms from the LPF based reduced model (Fig. 6.12(d)) match well the original model (Fig. 6.12(b)). But the simple reduced models lead to erratic time-domain behavior due to its non-passivity at high frequencies as shown in Fig. 6.12(c).

The experimental results also show that the output of LPF based reduced model exhibits less ringing than the output of original model. This is because the ringing is caused by high frequency components of input signal and many of those components are eliminated by LPF in the reduced model. If the those ringings are of interests, we can observe more of them by increasing the frequency range of the reduced models.

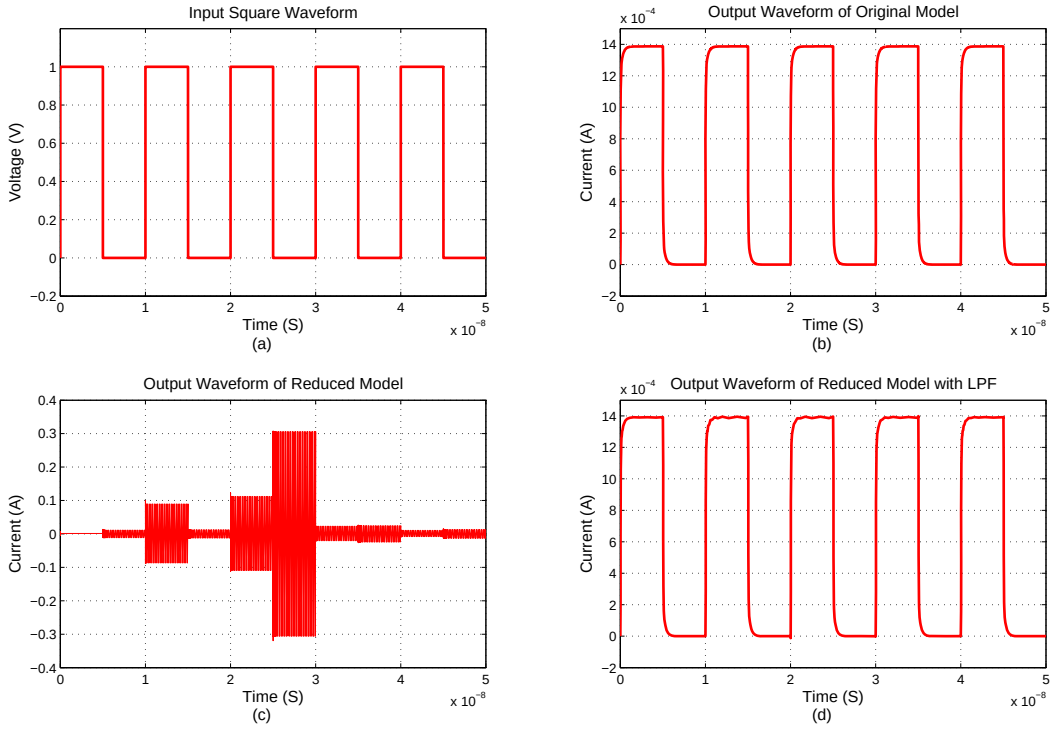

Figure 6.10: The comparison of responses of different models in time domain for the first example.

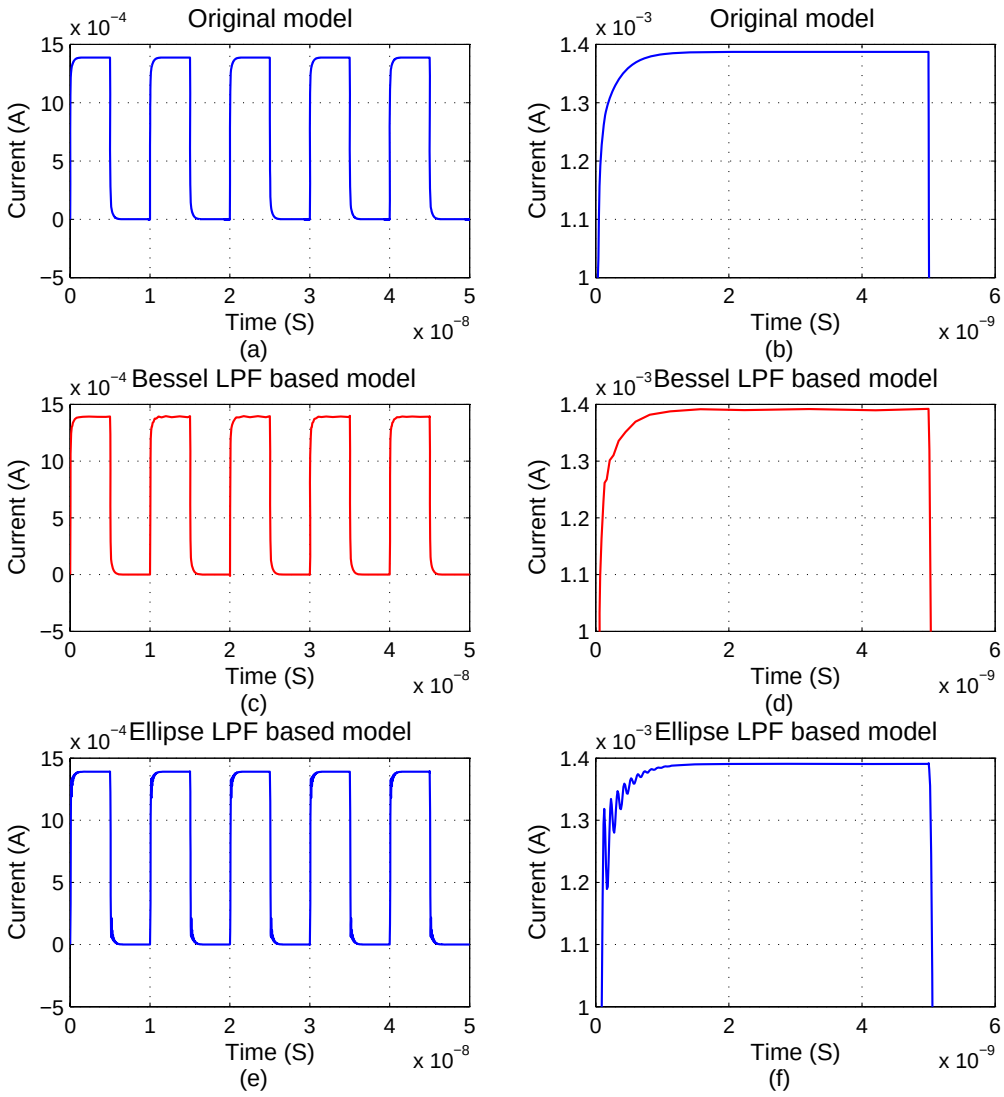

Figure 6.11: Comparison in time domian between reduced models based on Bessel filters and Ellipse filters.

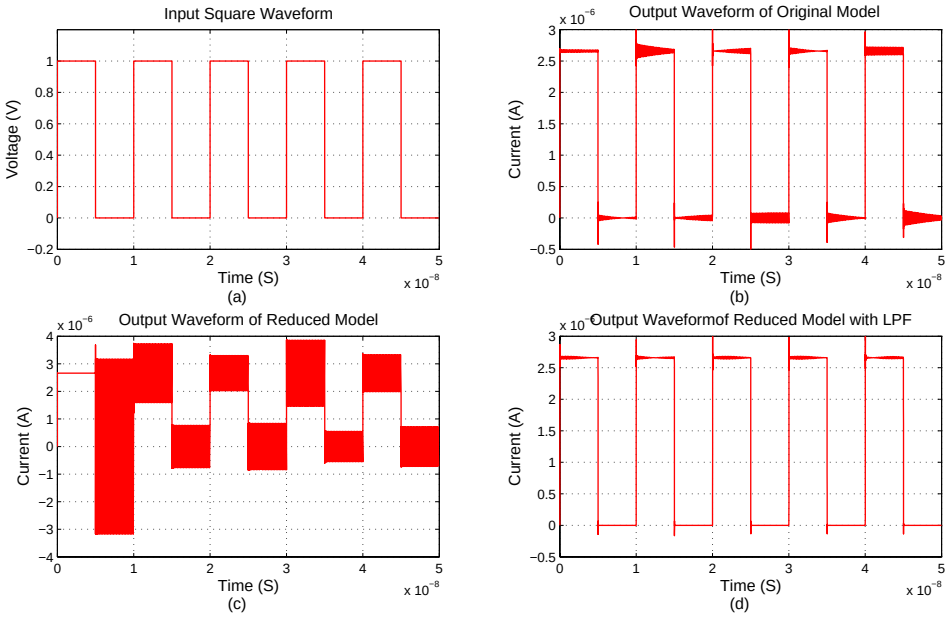

Figure 6.12: The comparison of responses of different models in time domain for second example.

Chapter 7

Decentralized model order reduction of linear networks with massive ports

The efficiency of model order reduction degrades as the number of ports increases. The reason for the degradation is fundamental and does not depend on any particular reduction algorithm [18]. For Krylov-subspace based algorithms, the cost associated with model computation is directly proportional to the number of inputs, i.e. to the number of columns in the transfer function matrix. For example, in the PRIMA algorithm [40], if only two (block) moments are to be matched at each port, and the network has 1000 ports, the resulting reduced model will have 2000 states. Similarly, in the TBR algorithm, for systems with many inputs, many states may be needed because of the high dimension of the controllable subspace.

In this section, we propose a *decentralized* model order reduction scheme where a whole MIMO circuit is decoupled into a number of MISO circuits based on the input-output interactions and each circuit is reduced individually. The decoupling process is guided using the *relative gain array* (RGA) [5], which measures the degree of interaction

of each input-output pair. Our method is based on the observation that for an output terminal, not all the input terminals are relevant, and this relevance is determined by their *relative gains*. As a result, an MIMO system can be naturally partitioned into many MISO systems and the traditional passivity-preserving model order reduction can be performed on these MISO systems. The new reduction algorithm, termed *DeMOR*, can perform very efficient reduction on MIMO systems.

7.1 Measurement of interaction

Relative Gain Array (RGA) is a matrix of interaction measures for all possible single-input single-output (SISO) pairings in an MIMO LTI system [5]. This concept has found widespread utility in process control, and as a system robustness measure. The RGA thus indicates the preferable variable pairings in a decentralized control system based on interaction considerations.

For a system $H(s)$ with p inputs and p outputs, there will be $p \times p$ relative gain elements

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1p} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2p} \\ \dots & \dots & & \dots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pp} \end{bmatrix} \tag{7.1}$$

and the relative gains between an output y_i and an input u_j are given by

$$\lambda_{ij} = \frac{g_{ij} \mid_u}{g_{ij} \mid_y} = \frac{(\triangle y_i / \triangle u_j) \mid_u}{(\triangle y_i / \triangle u_j) \mid_y} \tag{7.2}$$

where g_{ij} is the gain of the respective transfer function h_{ij}. A simple 2×2 coupled system

is shown in Fig. 7.1.

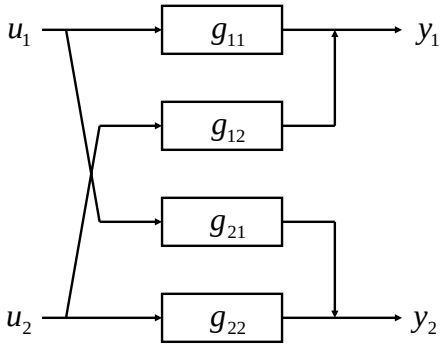

Figure 7.1: A coupled 2×2 system.

First, assume that all inputs except u_j remain constant, a step change in input u_j of magnitude $\triangle u_j$ will produce a change $\triangle y_i$ in output y_i. Thus, the gain between u_j and y_i when the other inputs are kept constant is given by

$$g_{ij}|_u = \frac{\triangle y_i}{\triangle u_j}|_u \tag{7.3}$$

which can be viewed as an open loop gain with respect to other inputs.

Second, when keeping all the outputs except y_i constant, a step change in input u_j of magnitude $\triangle u_j$ will result in another change in y_i. In this process, other outputs will also be affected due to cross-coupling. In order to keep them constant, we need to adjust other inputs correspondingly, which will also contribute to the change in y_i. The gain under the new set of conditions is denoted by

$$g_{ij}|_y = \frac{\triangle y_i}{\triangle u_j}|_y \tag{7.4}$$

which can be viewed as a closed loop gain with respect to other inputs.

Although the above gains are between the same pair of variables, they may have different values because they have been obtained under different conditions. If interaction exists, the change in y_i due to a change in u_j for the two cases (when other inputs and when other outputs are kept constant), will be different. The ratio,

$$\lambda_{ij} = \frac{g_{ij}|_u}{g_{ij}|_y} \tag{7.5}$$

defines the relative gain between the output y_i and input u_j.

There are two extreme cases: first, if $\lambda_{ij} = 0$, y_i is NOT influenced by u_j at all; second, if $\lambda_{ij} = 1$, closed loop gain is equal to open loop gain, which means the interaction from other inputs is zero and y_i is influenced by u_j ONLY.

In fact, by taking the absolute value of each RGA element and taking the inverse for those larger than 1, the scaled elements will fall into the range of $[0, 1]$

$$\begin{aligned} \lambda_{ij} &= |\lambda_{ij}|(|\lambda_{ij}| \leq 1) \\ \lambda_{ij} &= \tfrac{1}{|\lambda_{ij}|}(|\lambda_{ij}| > 1) \end{aligned} \tag{7.6}$$

The larger the scaled number is, the more important the corresponding input will be. Usually, most input-output pairs are magnitude-wise insignificant and their corresponding values are close to zero. For a given output i, the contribution of each input can be easily compared and those inputs can be arranged in a descending order in terms of their contribution. Usually, one output is only predominately influenced by a small number of inputs only.

The steady-state relative gain array of the system $H(s)$ at DC can be computed as

follows

$$\Lambda(H) = H(0) \circ H(0)^{-T} \tag{7.7}$$

where $\circ$ denotes element-by-element multiplication (often called the Hadamard or Schur product), and H^{-T} is the transpose of H^{-1}. For systems with non-square tranfer matrix, we can use *peudoinverse* instead

$$\Lambda(H) = H(0) \circ (H(0)^{T})^{+} \tag{7.8}$$

7.2 Decentralized model order reduction

An interconnect circuit can be formulated as the following state-space form using modified nodal analysis (MNA)

$$\begin{aligned} C\dot{x}(t) &= -Gx(t) + Bu(t) \\ y(t) &= L^T x(t) \end{aligned} \tag{7.9}$$

where $C, G \in R^{n \times n}$, $B, L \in R^{n \times p}$, and in which $x(t)$ is the state vector, and $u(t)$ and $y(t)$ represent the input and output, respectively. Typically, we have $p \ll n$. Model reduction algorithms seek to produce a smaller system

$$\begin{aligned} \tilde{C}\dot{\tilde{x}}(t) &= -\tilde{G}\tilde{x}(t) + \tilde{B}u(t) \\ \tilde{y}(t) &= \tilde{L}^T \tilde{x}(t) \end{aligned} \tag{7.10}$$

where $\tilde{C}, \tilde{G} \in R^{r \times r}$, $\tilde{B}, \tilde{L}, \in R^{r \times p}$. Order r is much smaller than the original order n, i.e. $r \ll n$, but the output $y(t)$ and $\tilde{y}(t)$ are approximately equal for inputs $u(t)$ of interest.

For the interconnect circuit in (7.9), the transfer function is

$$H(s) = L^T(Cs + G)^{-1}B \tag{7.11}$$

and the steady-state gain $H(0)$ is the DC gain H_{DC}

$$H_{DC} = L^T G^{-1} B \tag{7.12}$$

The RGA can be computed as

$$\Lambda(H) = H_{DC} \circ {H_{DC}}^{-T} \tag{7.13}$$

with RGA, we can decompose the whole system into a set of subsystems, each of which corresponds to one output.

For the ith decentralized model, the projection matrices V_i is constructed so that the columns span a *spatial dominant Krylov subspace* $K_m(A, R_i)$, where

$$A = (G + sC)^{-1}C \quad R_i = (G + sC)^{-1}B_i \tag{7.14}$$

In this approach, instead of all the inputs, B_i is only composed of the dominant inputs corresponding to the ith output. The ith reduced model is obtained by

$$\tilde{C}_i = V_i^T C V_i, \tilde{G}_i = V_i^T G V_i, \tilde{B}_i = V_i^T B, \tilde{L}_i = V_i^T L \tag{7.15}$$

Note that, for ith reduced model, only the ith output (the output corresponding to the ith row of the output matrix L_i^T) is valid.

In this projection framework, although more emphasis has been placed on energy transfers from dominant inputs, the energy transfers from other inputs are still coarsely preserved. Different from existing Krylov subspace methods, where only principle components in terms of frequency is considered, the new method tales into consideration principal components in terms of both frequency (temporal) and spatial information to reduce the system complexity. The DeMOR algorithm is shown in Fig. 7.2. When modeling passive

DECEMTRALIZED MODEL ORDER REDUCTION (DEMOR):
Input: $H : (G, C, B, L)$
Output: $\tilde{H}_i : (\tilde{G}_i, \tilde{C}_i, \tilde{B}_i, \tilde{L}_i)(i = 1, \ldots, p)$

1. Solve $GM = B$ for M_0
2. Compute $H_{DC} = L^T M_0$
3. Compute relative gain array $\Lambda(H) = H_{DC} \circ {H_{DC}}^{-T}$
4. Scale the RGA values to the range of $[0, 1]$
5. Set the threshold ϵ
6. For output i $(i = 1, \ldots, p)$
   Determine the corresponding dominant input matrix B_i
   Model order reduction using PRIMA to obtain $\tilde{H}_i$
   $\tilde{C}_i = V_i^T C V_i, \tilde{G}_i = V_i^T G V_i, \tilde{B}_i = V_i^T B, \tilde{L}_i = V_i^T L$
   where $colspan(V_i) = K_m((G + sC)^{-1}C, (G + sC)^{-1}B_i)$

Figure 7.2: Decemtralized model order reduction (DeMOR).

systems which cannot produce energy internally, it is desired that the reduced models also be passive. Otherwise, the reduced models may cause nonphysical behavior when used in later simulations, such as by generating energy at high frequencies that causes erratic or unstable time-domain behavior. Now we show such nonphysical behavior can be avoided in each DeMOR reduced model.

For example, for the ith reduced model, we have

$$\tilde{C}_i + \tilde{C}_i^T \geq 0 \quad \tilde{G}_i + \tilde{G}_i^T \geq 0 \quad L_i = B_i \tag{7.16}$$

Therefore, the ith reduced model is provably passive [40], which means nonphysical behavior will not be observed in any output of the ith reduced model when used in later simulations. Since the output to be used in the ith reduced model in DeMOR is just one of such outputs (the ith output), there will be no nonphysical behavior in simulations with DeMOR reduced models.

7.3 Localized modeling scheme for power grid analysis

Fast analysis of power grid networks has been a challenging problem for many years. The huge size renders circuit simulation inefficient and the large number of inputs further limits the application of existing Krylov-subspace macromodeling algorithms. However, strong locality has been observed that two nodes geometrically far have very small electrical impact on each other because of the exponential attenuation. However, no systematic approaches have been proposed to exploit such locality.

In this section, we propose a novel modeling and simulation scheme, which can automatically identify the dominant inputs for a given observed node in a power grid network. This enables us to build extremely compact models by projecting the system onto the locally dominant Krylov subspace corresponding to those dominant inputs only. The resulting simulation can be very fast with the compact models if we only need to view the responses of a few nodes under many different inputs. Experimental results show that the proposed method can have at least 100X speedup over SPICE-like simulations on a number of large power grid networks up to 1M nodes.

7.3.1 Locality of an RC mesh

We first show that RC meshes modeling power grid networks have strong local property. The *locality* of RC mesh can be evaluated from the relative gain array. Given a 21×21 RC mesh, each node is viewed as an input and the relative gains from each node to the central node coordinate $(11, 11)$ are plotted in Fig. 7.3. It is easy to see the nodes closer to the central node will have more impact. In addition, the node is only predominately influenced by a small number of inputs nearby, which validates the observation in [11]. As a result, RGA is a valid locality indicator to identify the most dominant inputs for a given output. With *locality*, we build a locally dominant reduction subspace (to be explained below) for a few observing nodes of interest, regardless of the size of the network, the number of inputs, and the patterns of input signals.

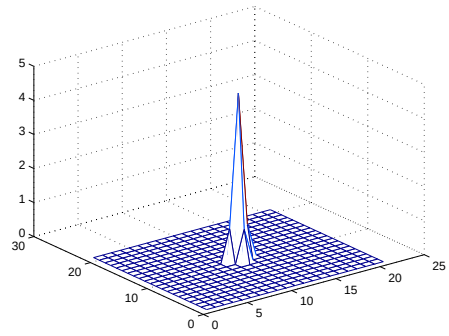

Figure 7.3: Locality illustration of an RC mesh.

7.3.2 Localized compact models at DC and wide frequency range

A *localized* model of a power grid network can be formulated as follows

$$\begin{aligned} C\dot{v}(t) &= -Gv(t) + Bi(t) \\ y(t) &= L_o^T v(t) \end{aligned} \tag{7.17}$$

where $C, G \in R^{n\times n}$ are capacitance and conductance matrices respectively, $B \in R^{n\times p}$ is the input matrix, $v(t) \in R^n$ and $i(t) \in R^p$ are the node voltage vector and input current vector.

Note that $L_o \in R^{n\times q}$ is the output matrix corresponding to q nodes we are measuring. In our problem, we are only interested in a few nodes. The corresponding transfer function is

$$H(s) = L_o^T(Cs + G)^{-1}B \tag{7.18}$$

and the steady-state gain $H(0)$ becomes the DC gain H_{DC}

$$H_{DC} = L_o^T G^{-1} B \tag{7.19}$$

In this case, H_{DC} is not a square matrix, then pseudoinverse of H_{DC}^T, $(H_{DC}^T)^+$ is used to compute the RGA

$$\Lambda(H) = H_{DC} \circ (H_{DC}{}^T)^+ \tag{7.20}$$

Now, we show that the RGA evaluated at DC $(s = 0)$ is sufficient for other frequencies. The reason is that a power grid network can be deemed as a cascaded low-pass RC filter and for the low-pass filter, the attenuation of high frequency components is much faster than the attenuation of low frequency components, which means high frequency components tend to be more localized and a decision based on DC is conservative. So the results of RGA at DC are actually valid for all the frequency range for those RC networks.

In Fig. 7.4, we show that as the frequency increases (DC (top), $1G$(middle), $100G$(bottom)), RGA values become more locally concentrated around a few nodes for each output node.

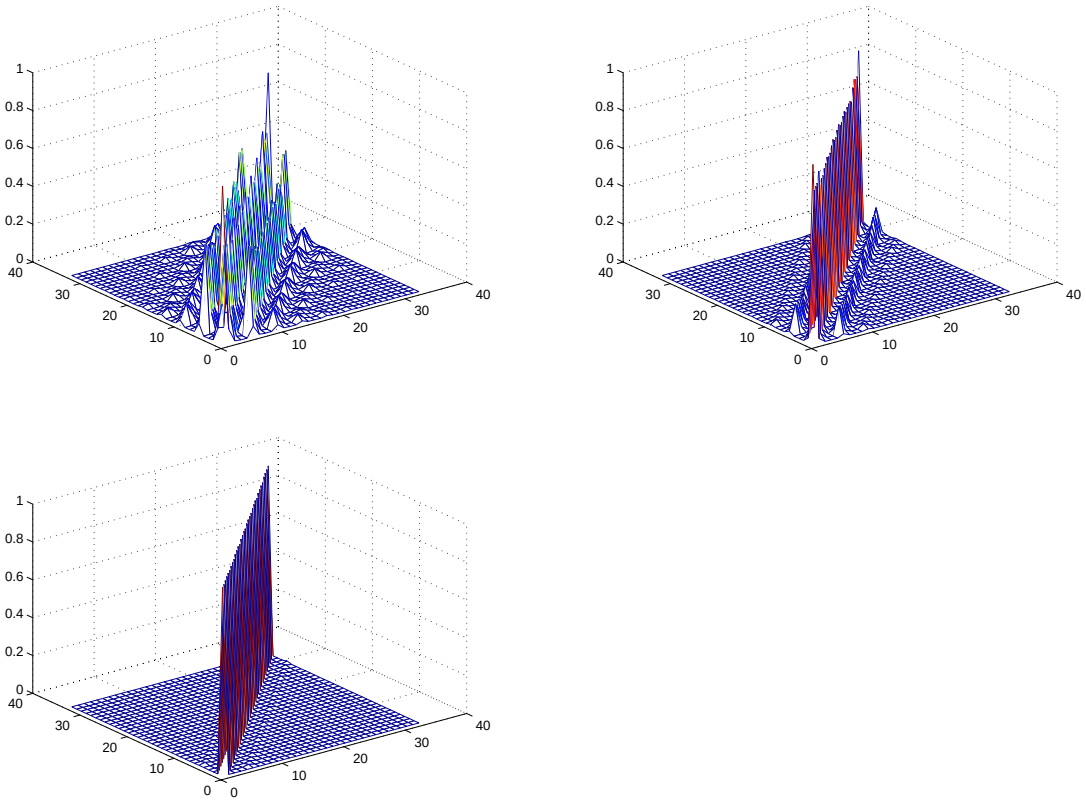

Figure 7.4: RGA computed at different frequencies.

7.3.3 Locally dominant Krylov subspace method

For the localized model, the projection matrix V_{domi} is constructed so that the columns span a Krylov subspace $K_m(A, R_{domi})$, where

$$A = G^{-1}C, R_{domi} = G^{-1}B_{domi} \tag{7.21}$$

In our approach, instead of all the inputs, B_{domi} is only composed of a small number of *dominant* inputs corresponding to those outputs of interest. This leads to the newly pro-

posed locally dominant Krylov subspace reduction:

$$\begin{aligned} \tilde{C}_{loc} &= V_{domi}^T C V_{domi}, \tilde{G}_{loc} = V_{domi}^T G V_{domi} \\ \tilde{B}_{loc} &= V_{domi}^T B, \tilde{L}_{loc} = V_{domi}^T L_o \end{aligned} \tag{7.22}$$

where $colspan(V_{domi}) \subseteq K_m(A, R_{domi})$. If we are interested in a number of nodes, then the system is projected onto a subspace which is the union of the locally dominant subspaces of those nodes. We remark that if the nodes of interests are limited to one local region, the order of reduced model may not increase even if more nodes are to be measured. The reason is that those nodes share many inputs (thus their subspaces). The proposed locally dominant Krylov subspace algorithm is shown in Fig. 7.5.

Our new method can be viewed by exploiting both temporal and spatial information to reduce the system complexity. Existing Krylov-subspace methods only take into consideration *frequency* (*thus temporal*) information. A reduction can be achieved because the frequency components are not viewed as equally important and only the dominant frequency subspace is preserved. However, those methods fail to consider *spatial* information and all the inputs are implicitly assumed to be equally important and fully preserved. In fact, if we focus on a few nodes or a local region, most inputs are insignificant owning to strong locality. As a result, the new approach can generate much more compact models than the existing temporal-only reduction methods.

7.3.4 Computational complexity analysis

For an RC circuit of order n and with p ports, it will take $O(n^\beta)$ to compute the DC moment H_{DC} as matrix G is very sparse in general, where, typically, $1 \leq \beta \leq 1.5$ for a $n \times n$ sparse matrix.

THE NEW LOCALLY DOMINANT KRYLOV SUBSPACE SIMULATION ALGORITHM:
Input: $G, C, B, u(t)$, set of observation nodes
Output: transient waveform at the observation nodes

1. Solve $GM_0 = B$ for M_0 (DC moment)
2. Compute $H_{DC} = {L_o}^T M_0$
3. Compute RGA $\Lambda(H)$=$H_{DC} \circ (H_{DC}^T)^+$
4. Scale the RGA values to the range of $[0, 1]$ and arrange them in a descending order in terms of the contribution to each output
5. For those outputs of interests, determine the corresponding dominant input matrix B_{domi} based on RGA
6. Compute localized reduced models by projection
7. Compute transient response on the reduced models for $u(t)$

Figure 7.5: The locally dominant Krylov subspace simulation method for power grid network analysis.

The computation of pseudoinverse is based on the singular value decomposition (SVD) of the matrix ${H_{DC}}^T$, which is a $p \times q$ matrix, where q is the number of nodes we are interested in. Since we are only interested in a small number of nodes, we have $q \ll p$ and the cost is $O(q^2 p)$. Assume that $p \ll n$ as this is the typical case, the reduction process is still dominated by $O(n^\beta)$.

The localized model will take about $O(nr^2 + rn^\alpha + n^\beta)$ to reduce using Krylov-subspace method, where r is the reduced order and $1 \leq \alpha \leq 1.2$ for sparse matrices. n^β is the same cost as the computation of the DC moment. The transient simulation of the reduced system takes about $O(r^3 + r^2 m)$ where m is the number of time steps in time domain.

Since the reduced order r is a very small number, the total cost is still dominated by $O(n^\beta)$, which is one DC solution of the original network. The reduced models can be used for many inputs without further solving the network again.

7.3.5 Partitions of input signals for RGA computation

To efficient compute RGA values for an RC mesh network, we find that we will be better off if we can compute RGA for current source inputs only (without voltage source inputs) as the two types of the signals are quite different in terms of magnitude.

Given a power grid network, as shown in Fig. 7.6, there are a small number of voltage sources and a huge number of current sources. The voltages supplies are DC with a constant value and the current sources are pulse currents generated when the gates are switching.

Assume that we are interested in a particular node, the voltage responses at that node can be decomposed into two parts owning to superposition:

$$V = V_{voltage} + V_{current} \tag{7.23}$$

where the first part is the responses of DC voltage sources, which is *static*, and the second part is the responses of independent current sources, which is *dynamic*. And the power grid network can be decomposed into two parts correspondingly, which are shown in Fig. 7.7 and Fig. 7.8.

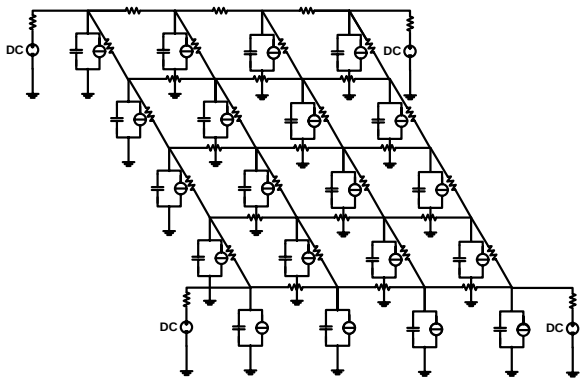

Figure 7.6: Power grid model

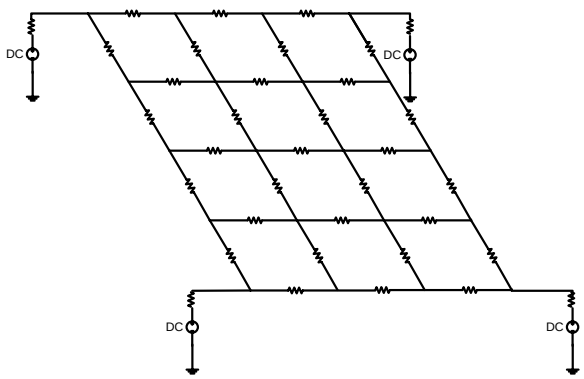

Figure 7.7: Power grid model: static part

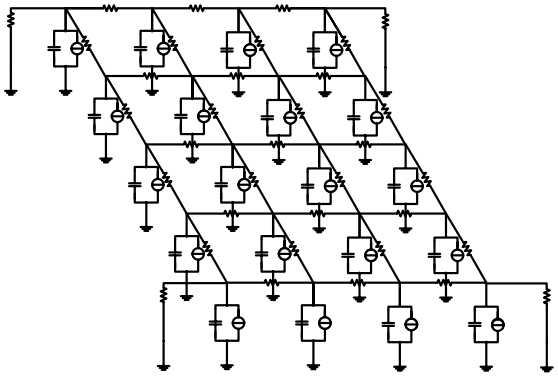

Figure 7.8: Power grid model: dynamic part

$V_{voltage}$ is static and it can be easily obtained via one DC analysis as shown in Fig. 7.7 because the number of voltage inputs is small and the conductance matrix is sparse. To compute the RGA, we only use the dynamic part shown in Fig. 7.8, which can give better indication of the RGA among all the current inputs.

7.4 Experimental results

The proposed method has been implemented in Matlab 7.0 and tested on an Intel quad-core workstation with 16GB memory.

The first example is a simple RC mesh ($R = 1\Omega$ and $C = 1pf$) with 1600 nodes. We verify the reduction accuracy in the time domain and frequency domain. 33 current sources are applied to the circuit, each of which generates a series of pulses of unit magnitude. The voltage responses at those input nodes are to be observed. The RGA value is shown in Fig. 7.9. We can see that the most input-output pairs are magnitude-wise insignificant and their corresponding values are close to zero.

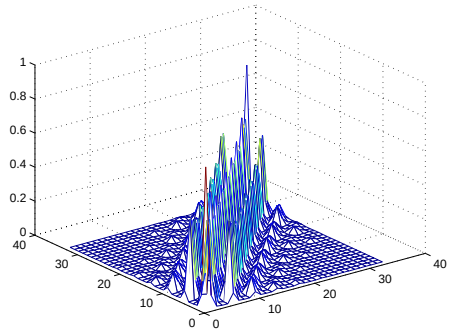

Figure 7.9: Degree of interaction measured by RGA.

First, we take a look at the transient responses at port 1. From Fig. 7.10, we see that port 1 is only dominately interacted with itself. The RGA values of other inputs are lower than the threshold value, 0.1. In this case, a reduced model of order 7 can match the original output well (we only build one reduced model for port 1). The results of PRIMA, SVDMOR, and DeMOR are shown in Fig. 7.10 by using the same order. We notice that SVDMOR dose not work well. The reason is that DC matrix has full rank, which is usually the case for a complete matrix-valued transfer function. The frequency response at port 1 from all inputs (i.e. $|\sum_{j=1}^{p} h_{1j}(s)|$) is also shown in Fig. 7.10 (the bottom one).

Now, we take a look at another port, the port 12, which is in the center of the circuit. From the RGA values for port 12, there are three dominant inputs: input 8, input 12, and input 16. A reduced model of order 12 is needed for a good match, where four moments

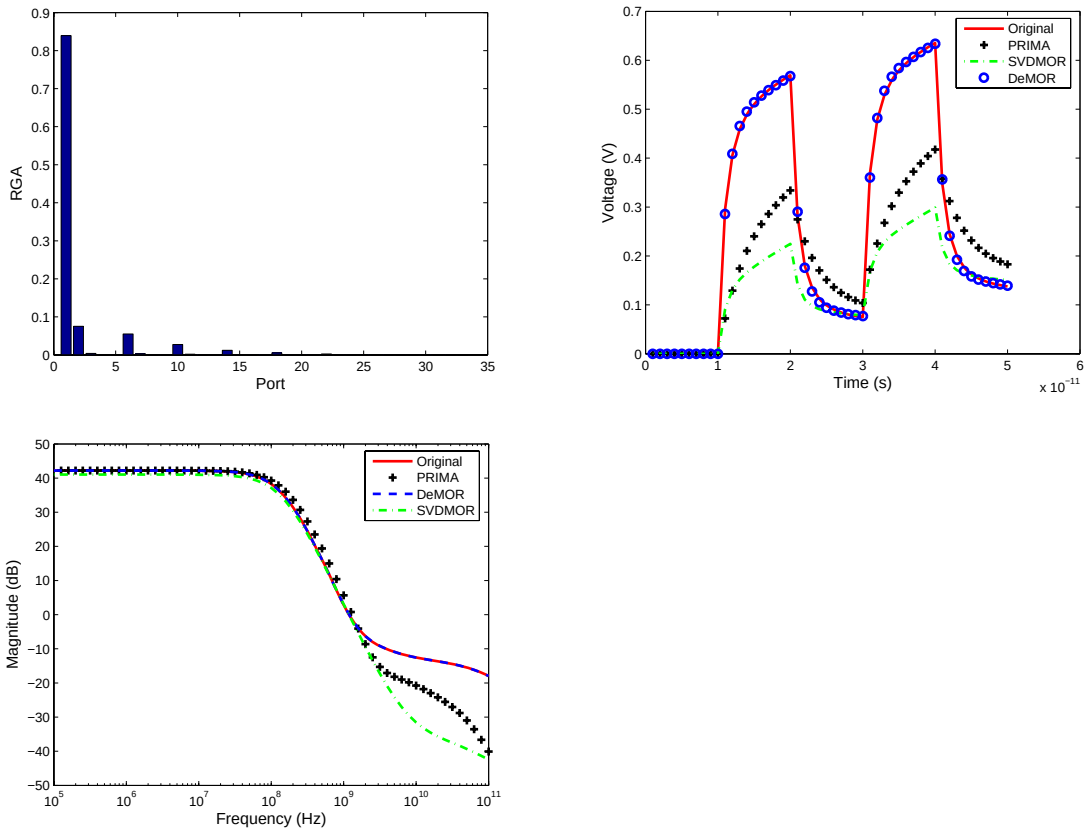

Figure 7.10: Relative gains (top), time-domain (middle) and frequency-domain responses at port 1 (bottom) for an RC circuit.

of the corresponding inputs are matched. The reduction results of PRIMA, SVDMOR, and DeMOR are compared in Fig. 7.11.

DeMOR is quite suitable for analyzing a number of nodes in a local region. We can perform the RGA analysis for all those nodes and find their dominant inputs. Typically, those nodes to be observed may share a very small number of dominant inputs, which is the case for power grid networks where input sources are not attached to every node to be observed.

In the second example, we have a power grid network with 10000 nodes and 1000 even distributed current sources. Now we are interested in the transient responses for 500 nodes

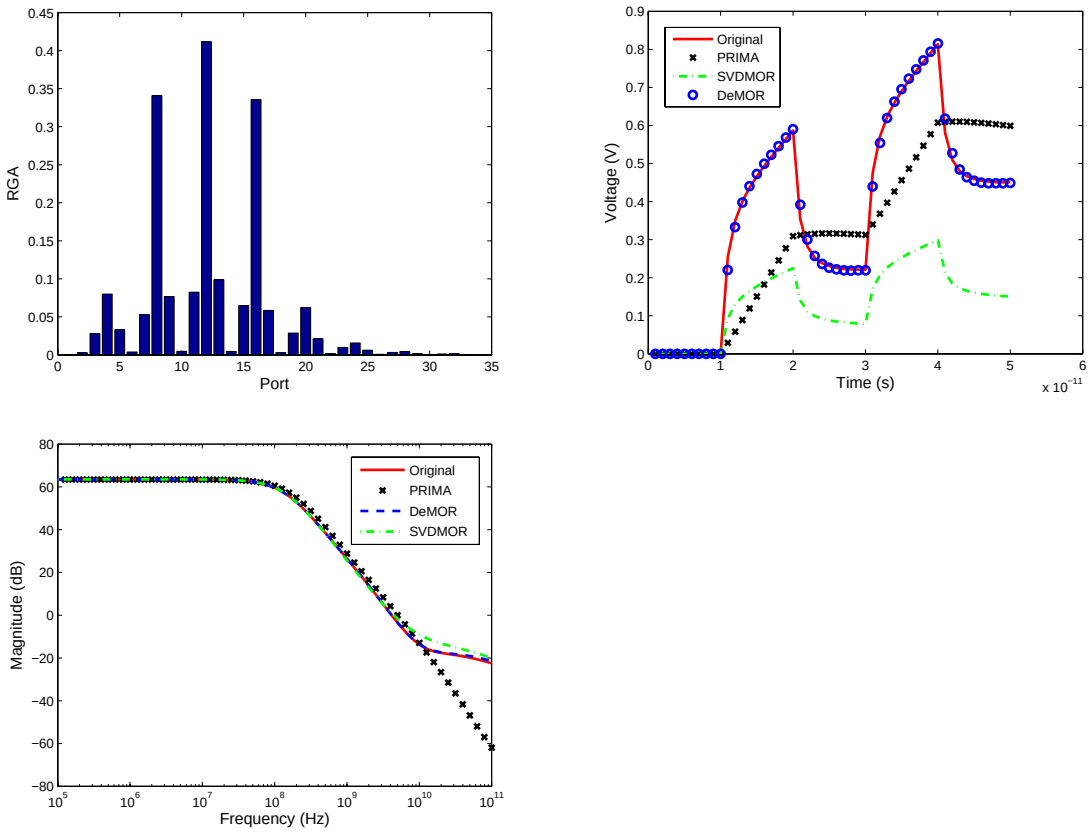

Figure 7.11: Relative gains (top), time-domain (middle) and frequency-domain responses at port 12 (bottom) for an RC circuit.

in a local region. The results of RGA of the circuit are shown in Fig. 7.12. We can see, the distribution of those nodes in terms of dominant inputs are very concentrated, which means a large number of nodes share only a small number of dominant inputs. For each node, we choose the most dominant input. Since many inputs are shared, the redundant ones are eliminated. As a result, 25 representative inputs are identified for the 500 nodes and only 2 moments are matched for each input, which results in a reduced model of order 50. The transient responses of the 500 nodes, however, can be well approximated by the localized reduced model.

Fig. 7.13 shows the transient responses at one of the 500 nodes. Given the same reduced

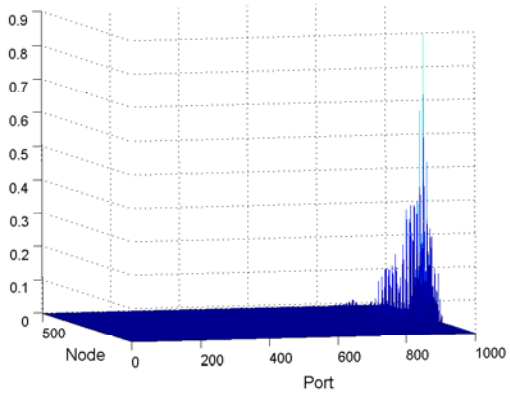

Figure 7.12: Degree of interaction measured by RGA.

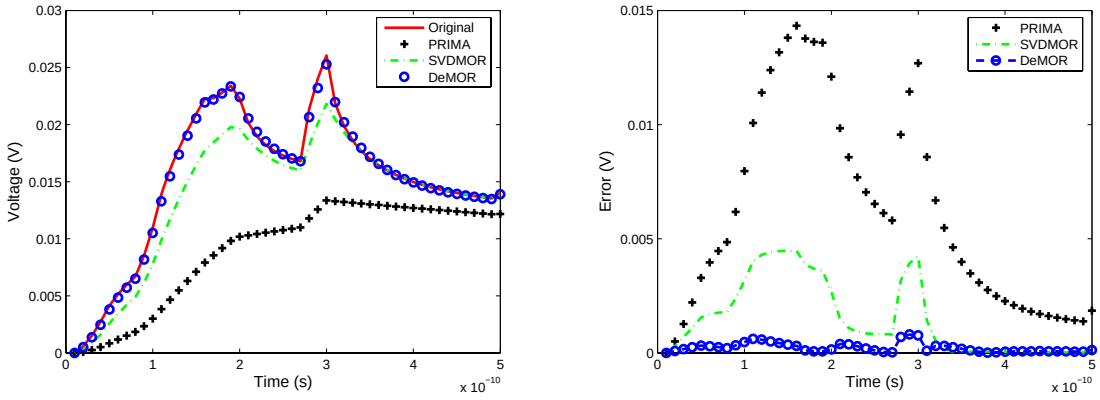

Figure 7.13: The simulation results of the part of the grid.

order, DeMOR can match the original response well, but there is still noticeable error for SVDMOR.

Chapter 8

Conclusion

In this thesis, we have proposed several non-Krylov subspace model order reduction techniques to mitigate three existing problems of Krylov subspace methods.

First, Krylov subspace methods can not generate models as compact as desired. To mitigate this problem, we have proposed two solutions based on fast balanced truncation via gramian approximation at the similar cost as the Krylov subspace methods. Different from existing single gramian approximation, our methods take into consideration of both gramians, which results in compact models with global accuracy. The first-order solution is very accurate for general structure systems and the second-order solution can preserve both passivity and structure information inherent to RLC circuit formulation.

Second, Kryov subspace methods are lack of passivity guarantees for general structure systems. In this study, a novel rational interpolation has been proposed to generate guaranteed passive reduced models for general structure dynamic systems. The proposed method is based on the maximum entropy solution of Carathéodory extension problem, which is as efficient as Krylov subspace methods with similar moment-matching property. Experimental results have demonstrated that CEMOR can be orders of magnitude faster

than the positive-real TBR approach for reducing general structure systems with comparable accurate reduced models. We have also proposed the concept of conditional passivity and a method to generate frequency band-limited passive reduced models. Such relaxation makes the circuit passive modeling work much easy using fitting based methods for general structure systems.

Finally, Kryov subspace methods degrade dramatically as the number of inputs is increased. In this study, we have proposed a novel approach resolve the long-standing problem of model order reduction of linear networks with many ports. The new method, termed *DeMOR*, adopts a *decentralized* reduction scheme, where a whole MIMO circuit is decoupled into a number of MISO circuits based on the input-output interactions and the order of each circuit is reduced separately. But different from existing approaches, a terminal reduction process is carried out with the aid of the relative gain array (RGA), which measures the degree of interaction of each input-output pair. As a result, efficient passive reduction of each subsystems become possible and so does the whole system. The proposed method is suitable for resistance-dominant interconnects like on-chip power grid, substrate planes. DeMOR can lead to extremely compact models for those systems with massive ports compared with the traditional MOR methods. Experimental results have demonstrated the advantage of the proposed method compared to existing approaches.

Bibliography

[1] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. The Society for Industrial and Applied Mathematics (SIAM), 2005.

[2] A. C. Antoulas. A new result on passivity preserving model reduction. *Systems & Control Letters*, 2005.

[3] A. Blomqvist, A. Lindquist, and R. Nagamune. Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: an optimization approach. *IEEE Trans. Automat. Contr.*, 48:2172–2190, 2003.

[4] A. Blomqvist and R. Nagamune. An extension of a Nevanlinna-Pick interpolation solver to cases including derivative constraints. In *IEEE Conference on Decision and Control*, pages 2552–2557, 2002.

[5] Edgar H. Bristol. On a new measure of interaction for multivariable process control. *IEEE Trans. Automat. Contr.*, AC-11:133–134, 1966.

[6] O. Brune. Synthesis of a finite two-terminal network whose driving point impedance is a prescribed function of frequency. *Journal of Math. and Phys.*, 10:191–236, 1931.

[7] C. I. Byrnes, S. V. Gusev, and A. Lindquist. A convex optimization approach to the rational covariance extension problem. *SLAM J. Matrix Anal. Optim.*, 37(1):211–229, 1999.

[8] C. I. Byrnes, A. Lindquist, S. V. Gusev, and A. S. Matveev. A complete parameterization of all positive rational extensions of a covariance sequence. *IEEE Trans. Automat. Contr.*, 40:1841–1857, 1995.

[9] Youns Chahlaoui and Paul Van Dooren. A collection of benchmark examples for model reduction of linear time invariant dynamical systems. SLICOT Working Note 2002-2, February 2002.

[10] X. Chen and J. T. Wen. Positive realness preserving model reduction with h_∞ norm error bounds. *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, 42(1):23–29, Jan 1995.

[11] E. Chiprout. Fast flip-chip power grid analysis via locality and grid shells. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 485–488, Nov. 2004.

[12] C. P. Coelho, J. Phillips, and L. M. Silveira. A convex programming approach for generating guaranteed passive approximations to tabulated frequency-data. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 23(2):293–301, Feb. 2004.

[13] U. Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Trans. Automat. Contr.*, 29:1097–1100, 1984.

[14] A. Devgan, H. Ji, and W. Dai. How to efficiently capture on-chip inductance effects: introducing a new circuit element K. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 150–155, 2000.

[15] P. Enqvist. A homotopy approach to rational covariance extension with degree constraint. *Int. J. Appl. Math. Comput. Sci.*, 11:1173–1201, 2001.

[16] P. Feldmann. Model order reduction techniques for linear systems with large numbers of terminals. In *Proc. European Design and Test Conf. (DATE)*, pages 944–947, 2004.

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

[18] P. Feldmann and F. Liu. Sparse and efficient reduced order modeling of linear subcircuits with large number of terminals. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 88–92, 2004.

[19] R. W. Freund. SPRIM: structure-preserving reduced-order interconnect macromodeling. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 80–87, 2004.

[20] K. Glover. All optimal Hankel-norm approximations of linear multi-variable systems and their L_∞ error bounds". *Int. J. Control*, 36:1115–1193, 1984.

[21] M. Green. Balanced stochastic realization. *Linear Algebra and its Application*, 1988.

[22] E. J. Grimme. *Krylov projection methods for model reduction (Ph.D. Thesis)*. University of Illinois at Urbana-Champaign, 1997.

[23] B. Gustavsen and A. Semlyen. Enforcing passivity for admittance matrices approximated by rational functions. *IEEE Trans. on Power Systems*, 16(1):97–104, Feb. 2001.

[24] Z. He, M. Celik, and L. Pillegi. SPIE: Sparse partial inductance extraction. In *Proc. Design Automation Conf. (DAC)*, pages 137–140, 1997.

[25] H. Ji, A. Devgan, and W. Dai. Ksim: A stable and efficient RKC simulator for capturing on-chip inductance effect. In *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, pages 379–384, 2001.

[26] M. Kamon, F. Wang, and J. White. Generating nearly optimally compact models from Krylov-subspace based reduced-order models. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 47(4):239–248, 2000.

[27] K. J. Kerns and A. T. Yang. Stable and efficient reduction of large, multiport RC network by pole analysis via congruence transformations. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 16(7):734–744, July 1998.

[28] K. J. Kerns and A. T. Yang. Stable and efficient reduction of large, multiport rc network by pole analysis via congruence transformations. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 16(7):734–744, July 1998.

[29] A. J. Laub and W. F. Arnold. Controllability and observability criteria for multivariable linear second-order models. *IEEE Trans. Automat. Contr.*, 29:163–165, 1987.

[30] A. J. Laub, M. T. Heath, C. C. Paige, and R. C. Ward. Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Trans. Automat. Contr.*, 32:115–122, 1987.

[31] J. R. Li. *Model reduction of large linear systems via low rank system gramians (Ph.D. Thesis)*. MIT, 2002.

[32] J. R. Li, F. Wang, and J. White. An efficient Lyapunov equation-based approach for generating reduced-order models of interconnect. In *Proc. Design Automation Conf. (DAC)*, pages 1–6, 1999.

[33] J. R. Li and J. White. Efficient model reduction of interconnect via approximate system gramians. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 380–383, 1999.

[34] Peng Li and Weiping Shi. Model order reduction of linear networks with massive ports via frequency-dependent port packing. In *Proc. Design Automation Conf. (DAC)*, pages 267–272, 2006.

[35] P. Liu, S. X.-D. Tan, H. Li, Z. Qi, J. Kong, B. McGaughy, and L. He. An efficient method for terminal reduction of interconnect circuits considering delay variations. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 821–826, 2005.

[36] T. Mangold and P. Russer. Full-wave modeling and automatic equivalent-circuit generation of millimeter-wave planar and multilayer structures. *IEEE Trans. on Microwave Theory and Techniques*, 47(6):851–858, June 1999.

[37] David. G. Meyer and Sriram Srinivasan. Balancing and model reduction for second-order form linear systems. *IEEE Trans. Automat. Contr.*, AC-41:1632–1644, 1996.

[38] B. Moore. Principal component analysis in linear systems: Controllability, and observability, and model reduction. *IEEE Trans. Automat. Contr.*, 26(1):17–32, 1981.

[39] Jason Morsey and Andreas C. Cangellaris. PRIME: passive realization of interconnect models from measured data. *Electrical Performance of Electronic Packaging*, pages 47–50, Oct. 2001.

[40] A. Odabasioglu, M. Celik, and L.T. Pileggi. PRIMA: Passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pages 645–654, 1998.

[41] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall, 1999.

[42] J. R. Phillips, L. Daniel, and L. M. Silveira. Guaranteed passive balancing transformation for model order reduction. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 22(8):1027–1041, 2003.

[43] J. R. Phillips and L. M. Silveira. Poor man's TBR: a simple model reduction scheme. In *Proc. European Design and Test Conf. (DATE)*, pages 938–943, 2004.

[44] J. R. Phillips and L. M. Silveira. Poor man's TBR: a simple model reduction scheme. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 24(1):43– 55, 2005.

[45] L. T. Pillage and R. A. Rohrer. Asymptotic waveform evaluation for timing analysis. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pages 352–366, April 1990.

[46] B. Porat. *Digital Processing of Random Signals, Theory & Methods*. Prentice Hall, 1994.

[47] M. G. Safonov and R. Y. Chiang. A Schur method for balanced truncation model reduction. *IEEE Trans. Automat. Contr.*, 34:729–733, 1989.

[48] Bernard N. Sheehan. ENOR: model order reduction of RLC circuits using nodal equations for efficient factorization. In *Proc. Design Automation Conf. (DAC)*, pages 17–21, 1999.

[49] L. M. Silveira and J. R. Phillips. Exploiting input information in a model reduction algorithm for massively coupled parasitic networks. In *Proc. Design Automation Conf. (DAC)*, pages 385–388, 2004.

[50] M. Silveira, M. Kamon, I. Elfadel, and J. White. A coordinate-transformed Arnoldi algorithm for generating guaranteed stable reduced-order models of RLC circuits. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 288–294, 1996.

[51] D. C. Sorensen. Passivity preserving model reduction via interpolation of spectral zeros. *Systems & Control Letters*, 2005.

[52] T. Stykel. Grammian-based model order reduction for descriptor systems. *Math. Control Signals Systems*, 16:297–319, 2004.

[53] Y. Su, J. Wang, X. Zeng, Z. Bai, C. Chiang, and D. Zhou. SAPOR: second-order Arnoldi method for passive order reduction of RCS circuits. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 74–79, 2004.

[54] D. Vasilyev and J. White. A more reliable reduction algorithm for behavioral model extraction. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 813–820, 2005.

[55] J. M. Wang and T. V. Nguyen. Extended Krylov subspace method for reduced order analysis of linear circuit with multiple sources. In *Proc. Design Automation Conf. (DAC)*, pages 247–252, 2000.

[56] N. Wang and V. Balakrishnan. Fast balanced stochastic truncation via a quadratic extension of the alternating direction implicit iteration. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 801–805, 2005.

[57] N. Wang, V. Balakrishnan, and C.-K. Koh. Passivity-preserving model reduction via a computationally efficient projection-and-balance scheme. In *Proc. Design Automation Conf. (DAC)*, pages 369–374, 2004.

[58] K. Willcox and J. Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal*, 2002.

[59] B. Yan, S. X-.D. Tan, P. Liu, and B. McGaughy. Passive interconnect macromodeling via balanced truncation of linear systems in descriptor form. In *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, pages 355–360, 2007.

[60] B. Yan, S. X-.D. Tan, and B. McGaughy. Second-order balanced truncation for passive-model order reduction of RLCK circuits. *IEEE Trans. on Circuits and Systems II: Express Briefs*, 55(9):942–946, 2008.

[61] H. Yu, Y. Shi, L. He, and D. Smart. A fast block structure preserving model order reduction in inverse inductance circuits. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 7–11, Nov. 2006.

[62] Yunkai Zhou. *Numerical methods for large scale matrix equations with applications in LTI system model reduction (Ph.D. Thesis)*. Rice University, 2002.