A Unified Krylov Projection Framework for Structure-Preserving Model Reduction

Zhaojun Bai[1], Ren-cang Li[2], and Yangfeng Su[3]

[1] Department of Computer Science and Department of Mathematics, University of California, Davis, CA 95616, USA, bai@cs.ucdavis.edu

[2] Department of Mathematics, University of Texas, Arlington, TX 76019, rcli@uta.edu

[3] School of Mathematical Science, Fudan University, Shanghai 200433, China, yfsu@fudan.edu.cn

1 Introduction

Physical systems often have certain characteristics that are critical in determining the system behaviors. Often these characteristics appear in the form of the system matrices being naturally blocked with each sub-block having its own physical relevance. For example, the system matrices from linearizing a second order dynamical system admit a natural 2-by-2 block partitioning. General purpose subspace projection techniques for model order reduction usually destroy any block structure and thus the reduced systems may not be of the same type as the original system. Other things being equal, we would like to preserve the block structure and hence some of the important characteristics so that the reduced systems are much like the original system but only at much smaller scales.

Structure-preserving Krylov subspace projection methods have received much attention in recent years. In this chapter, we discuss the advance of the structure-preserving methods under a unified Krylov projection formulation. We shall start by building a mathematical foundation and a general paradigm to preserve important block structures under subspace projections. The general paradigm provides a unified projection formulation. When necessary, the technique can be used to preserve certain blocks in the system matrices. We then go on to study in detail model order reductions of RCL and RCS systems.

The rest of this paper is organized as the follows. In Section 2, we discuss a unified Krylov subspace projection formulation for model order reduction with properties of structure-preserving and moment-matching, and present a generic algorithm for constructing structure-preserving projection matrices. The inherent structural properties of Krylov subspaces for certain block matrices are presented in Section 3. Section 4 examines structure-preserving

Appeared in Model Order Reduction: Theory, Research Aspects and Applications, Edited by W. H. A. Schilders, H. A. van der Vorst and J. Rommers, Math. in Industry 13, Springer, pp.75-93, 2008

model order reduction of RCL and RCS equations including the objective of developing synthesized RCL and RCS equations.

Throughout the paper, $\mathbb{R}^{k\times\ell}$ is the set of $k \times \ell$ real matrices. I is the identity matrix and its dimension can be judged from the context. Unless otherwise explicitly stated, capital letters are matrices, while lower case letters are vectors or scalars. X^{T} is the transpose of the matrix X, $\mathrm{span}\{X\}$ is the subspace spanned by the columns of X.

Let A be $N \times N$, and let B be $N \times p$. The *kth Krylov subspace* generated by A on B is defined to be

$$\mathcal{K}_k(A,B) = \mathrm{span}\{B, AB, \ldots, A^{k-1}B\}. \tag{1.1}$$

For convenience, when $k = 0$, define $\mathcal{K}_0(A,B) = \{0\}$, a subspace of the zero vector.

2 A unified Krylov projection structure-preserving model order reduction framework

Consider the *matrix-valued transfer function* of the first-order multi-input multi-output (MIMO) linear dynamical system

$$H(s) = \mathbf{L}^{\mathrm{T}}(s\mathbf{C} + \mathbf{G})^{-1}\mathbf{B}, \tag{2.1}$$

where $\mathbf{C}$ and $\mathbf{G}$ are $N \times N$, $\mathbf{B}$ is $N \times m$ and $\mathbf{L}$ is $N \times p$. Often $p \ll N$ and $m \ll N$.

Assume that $\mathbf{G}$ is nonsingular. The transfer function can be expanded around $s = 0$ as

$$\begin{aligned} H(s) &= \sum_{\ell=0}^{\infty}(-1)^{\ell}s^{\ell}\mathbf{L}^{\mathrm{T}}(\mathbf{G}^{-1}\mathbf{C})^{\ell}\mathbf{G}^{-1}\mathbf{B} \\ &\equiv \sum_{\ell=0}^{\infty}(-1)^{\ell}s^{\ell}M_{\ell}, \end{aligned}$$

where

$$M_{\ell} = \mathbf{L}^{\mathrm{T}}(\mathbf{G}^{-1}\mathbf{C})^{\ell}\mathbf{G}^{-1}\mathbf{B}$$

are referred to as the *moments* at $s = 0$. In the case when $\mathbf{G}$ is singular or approximations to $H(s)$ around a selected point $s_0 \neq 0$ are sought[4], a shift

$$s = (s - s_0) + s_0 \equiv \sigma + s_0 \tag{2.2}$$

can be performed and then

[4] It is assumed that the matrix pencil $s\mathbf{C} + \mathbf{G}$ is regular, meaning that there are at most N values of s at which $s\mathbf{C} + \mathbf{G}$ is singular.

$$s\mathbf{C} + \mathbf{G} = (s - s_0)\mathbf{C} + s_0\mathbf{C} + \mathbf{G} \equiv \sigma\mathbf{C} + (s_0\mathbf{C} + \mathbf{G}).$$

Upon substitutions (i.e., renaming)

$$\mathbf{G} \leftarrow s_0\mathbf{C} + \mathbf{G}, \quad s \leftarrow \sigma,$$

the problem of approximating $H(s)$ around $s = s_0$ becomes equivalent to approximate the substituted $H(\sigma)$ around $\sigma = 0$. For this reason, without loss of generality we shall focus approximation mostly around $s = 0$ in this chapter unless some care has to be taken for computational efficiency when a shift like (2.2) has to be performed.

Many transfer functions appearing in different forms can be re-formulated in the first order form (2.1).

Example 1. Consider a simple *integro-differential-algebraic equations* (Integro-DAEs) arising from the MNA formulation of circuits, such as the ones described in the chapters by Freund and by Gad, Nakhla and Achar of this book:

$$\begin{cases} C\frac{d}{dt}z(t) + Gz(t) + \Gamma\int_0^t z(\tau)d\tau = Bu(t), \\ \qquad\qquad\qquad\qquad\qquad\qquad y(t) = B^{\mathrm{T}}z(\tau). \end{cases}$$

The transfer function of the Integro-DAEs is given by

$$H(s) = B^{\mathrm{T}}\left(sC + G + \frac{1}{s}\Gamma\right)^{-1}B. \tag{2.3}$$

If one defines

$$\mathbf{C} = \begin{bmatrix} C & 0 \\ 0 & -W \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} G & \Gamma \\ W & 0 \end{bmatrix}, \quad \mathbf{L} = \mathbf{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \tag{2.4}$$

for any *nonsingular* matrix W. Then the transfer function is of the form (2.1), namely

$$H(s) = \mathbf{B}^{\mathrm{T}}(s\mathbf{C} + \mathbf{G})^{-1}\mathbf{B}. \tag{2.5}$$

In (2.4), the matrix W is usually taken to be Γ (if it is nonsingular) or simply the identity matrix.

Alternatively, if one defines

$$\mathbf{C} = \begin{bmatrix} G & C \\ W & 0 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \Gamma & 0 \\ 0 & -W \end{bmatrix}, \quad \mathbf{L} = \mathbf{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \tag{2.6}$$

again for any *nonsingular* matrix W (usually taken to be C if it is nonsingular, or simply the identity matrix). Then the transfer function is turned into the form

$$H(s) = s\,\mathbf{B}^{\mathrm{T}}(s\mathbf{C} + \mathbf{G})^{-1}\mathbf{B}. \tag{2.7}$$

Leaving out the front factor s, (2.7) is in the form of (2.1). In the second linearization (2.7), matrix-vector products with the matrices $\mathbf{G}^{-1}\mathbf{C}$ and $\mathbf{G}^{-\mathrm{T}}\mathbf{C}^{\mathrm{T}}$

are much easier to do than the first linearization (2.4) and (2.5). These two types of matrix-vector products are needed in forming Krylov subspaces for calculating approximations to $H(s)$ around $s = 0$. In this respect, the first linearization (2.4) and (2.5) favors approximations around $s = \infty$. In the case when approximations near a finite point $s_0 \neq 0$ are sought, a shift like (2.2) must be performed and then neither linearization has cost advantage over the other because the $s_0\mathbf{C}+\mathbf{G}$ is no longer block diagonal for both (2.4) and (2.6). But we point out that if the shift is performed before linearization, the same advantage as the second linearization over the first one for approximations near $s = 0$ is retained. Detail for this shift-*before*-linearization is discussed in Section 4. ◇

Example 2. The interconnected (coupled) system described in [12] and [18] (see also the chapters by Reis and Stykel and by Vanderndorpe and Van Dooren of this book) gives rise to the following transfer function

$$H(s) = L_0^{\mathrm{T}} \left(I - W(s)\mathcal{E}\right)^{-1} W(s)B_0,$$

where $\mathcal{E}$ is the subsystem incidence matrix as a glue for connecting all subsystems $H_1(s), \ldots, H_k(s)$ together, and

$$\begin{aligned} W(s) &= \mathrm{diag}(\ H_1(s),\ \ldots,\ H_k(s)\ ) \\ &= \mathrm{diag}(\ L_1^{\mathrm{T}}(sI - A_1)^{-1}B_1,\ \ldots, L_k^{\mathrm{T}}(sI - A_k)^{-1}B_k\ ). \end{aligned}$$

Let $A = \mathrm{diag}(A_1, \ldots, A_k)$, $B = \mathrm{diag}(B_1, \ldots, B_k)$, and $L = \mathrm{diag}(L_1, \ldots, L_k)$, then the transfer function $H(s)$ can be turned into the form (2.1), namely

$$H(s) = \mathbf{L}^{\mathrm{T}}(s\mathbf{C} + \mathbf{G})^{-1}\mathbf{B},$$

where $\mathbf{C} = I$, $\mathbf{G} = -A - B\mathcal{E}L$, $\mathbf{B} = BB_0$ and $\mathbf{L} = LL_0$. ◇

Model order reduction of the transfer function $H(s)$ defined by (2.1) via subspace projection starts by computing matrices

$$\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{N\times n} \quad \text{such that} \quad \mathcal{Y}^{\mathrm{T}}\mathbf{G}\mathcal{X} \text{ is nonsingular,}$$

then defines a *reduced-order transfer function*

$$H_{\mathrm{r}}(s) = \mathbf{L}_{\mathrm{r}}^{\mathrm{T}}(s\mathbf{C}_{\mathrm{r}} + \mathbf{G}_{\mathrm{r}})^{-1}\mathbf{B}_{\mathrm{r}}, \tag{2.8}$$

where

$$\mathbf{C}_{\mathrm{r}} = \mathcal{Y}^{\mathrm{T}}\mathbf{C}\mathcal{X}, \quad \mathbf{G}_{\mathrm{r}} = \mathcal{Y}^{\mathrm{T}}\mathbf{G}\mathcal{X}, \quad \mathbf{B}_{\mathrm{r}} = \mathcal{Y}^{\mathrm{T}}\mathbf{B}, \quad \mathbf{L}_{\mathrm{r}} = \mathcal{X}^{\mathrm{T}}\mathbf{L}. \tag{2.9}$$

Similarly, the reduced transfer function $H\mathrm{r}(s)$ can be expanded around $s = 0$:

$$\begin{aligned} H_{\mathrm{r}}(s) &= \sum_{\ell=0}^{\infty}(-1)^{\ell}s^{\ell}\mathbf{L}_{\mathrm{r}}^{\mathrm{T}}(\mathbf{G}_{\mathrm{r}}^{-1}\mathbf{C}_{\mathrm{r}})^{\ell}\mathbf{G}_{\mathrm{r}}^{-1}\mathbf{B}_{\mathrm{r}} \\ &= \sum_{\ell=0}^{\infty}(-1)^{\ell}s^{\ell}M_{\mathrm{r},\ell}, \end{aligned}$$

where

$$M_{\mathrm{r},\ell} = \mathbf{L}_\mathrm{r}^\mathrm{T}(\mathbf{G}_\mathrm{r}^{-1}\mathbf{C}_\mathrm{r})^\ell \mathbf{G}_\mathrm{r}^{-1}\mathbf{B}_\mathrm{r}$$

are referred to as the *moments* of the reduced system.

In practice it is often that $n \ll N$. This makes the reduced system matrices $\mathbf{G}_\mathrm{r}$, $\mathbf{C}_\mathrm{r}$, $\mathbf{L}_\mathrm{r}$, and $\mathbf{B}_\mathrm{r}$ much smaller. By choosing $\mathcal{X}$ and $\mathcal{Y}$ right, the reduced system associated with the reduced transfer function can be made to resemble the original system enough to have practical relevance.

The following theorem dictates how good a reduced transfer function approximates the original transfer function. For the case when $\mathbf{G}$ is the identity matrix, it is due to [19]. The general form as stated in the following theorem was proved by [6]; a new proof in the projection language was given later in [9]. Its implication to structure-preserving model reduction was also first realized in [9].

Theorem 1. *Suppose that* $\mathbf{G}$ *and* $\mathbf{G}_\mathrm{r}$ *are nonsingular. If*

$$\mathcal{K}_k(\mathbf{G}^{-1}\mathbf{C}, \mathbf{G}^{-1}\mathbf{B}) \subseteq \mathrm{span}\{\mathcal{X}\}$$

and

$$\mathcal{K}_j(\mathbf{G}^{-\mathrm{T}}\mathbf{C}^\mathrm{T}, \mathbf{G}^{-\mathrm{T}}\mathbf{L}) \subseteq \mathrm{span}\{\mathcal{Y}\},$$

then the moments of $H(s)$ *and of its reduced function* $H_\mathrm{r}(s)$ *satisfy*

$$M_\ell = M_{\mathrm{r},\ell} \quad \textit{for} \quad 0 \le \ell \le k + j - 1,$$

which imply

$$H_\mathrm{r}(s) = H(s) + \mathcal{O}(s^{k+j}).$$

Remark 1. The conditions suggest that by enforcing $\mathrm{span}\{\mathcal{X}\}$ and/or $\mathrm{span}\{\mathcal{Y}\}$ to contain more appropriate Krylov subspaces associated with multiple points, $H_\mathrm{r}(s)$ can be made to approximate $H(s)$ well near all those points. See [6] and [13, 14] for more detail.

Let us now discuss the objectives of structure-preserving model order reduction. For the simplicity of exposition, consider system matrices $\mathbf{G}$, $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{L}$ having the following 2×2 block structure

$$\mathbf{C} = \begin{array}{cc} & \begin{array}{cc} N_1 & N_2 \end{array} \\ \begin{array}{c} N_1' \\ N_2' \end{array} & \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \end{array}, \quad \mathbf{G} = \begin{array}{cc} & \begin{array}{cc} N_1 & N_2 \end{array} \\ \begin{array}{c} N_1' \\ N_2' \end{array} & \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & 0 \end{bmatrix} \end{array},$$
$$\mathbf{B} = \begin{array}{cc} & p \\ \begin{array}{c} N_1' \\ N_2' \end{array} & \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \end{array}, \quad \mathbf{L} = \begin{array}{cc} & m \\ \begin{array}{c} N_1 \\ N_2 \end{array} & \begin{bmatrix} L_1 \\ 0 \end{bmatrix} \end{array}, \tag{2.10}$$

where $N_1 + N_2 = N_1' + N_2' = N$. System matrices from the *time-domain modified nodal analysis* (MNA) circuit equations take such forms (see Section 4).

A structure-preserving model order reduction technique generates a reduced-order system that structurally preserves the block structure, namely, to have

$$\begin{aligned}
\mathbf{C}_\mathrm{r} &= \begin{array}{cc} & \begin{array}{cc} n_1 & n_2 \end{array} \\ \begin{array}{c} n_1' \\ n_2' \end{array} & \begin{bmatrix} C_{\mathrm{r},11} & 0 \\ 0 & C_{\mathrm{r},22} \end{bmatrix} \end{array}, \quad \mathbf{G}_\mathrm{r} = \begin{array}{cc} & \begin{array}{cc} n_1 & n_2 \end{array} \\ \begin{array}{c} n_1' \\ n_2' \end{array} & \begin{bmatrix} G_{\mathrm{r},11} & G_{\mathrm{r},12} \\ G_{\mathrm{r},21} & 0 \end{bmatrix} \end{array}, \\
\mathbf{B}_\mathrm{r} &= \begin{array}{cc} & p \\ \begin{array}{c} n_1' \\ n_2' \end{array} & \begin{bmatrix} B_{\mathrm{r},1} \\ 0 \end{bmatrix} \end{array}, \quad \mathbf{L}_\mathrm{r} = \begin{array}{cc} & m \\ \begin{array}{c} n_1 \\ n_2 \end{array} & \begin{bmatrix} L_{\mathrm{r},1} \\ 0 \end{bmatrix} \end{array},
\end{aligned} \tag{2.11}$$

where $n_1+n_2 = n_1'+n_2' = n$. Furthermore, each sub-block is a direct reduction from the corresponding sub-block in the original system.

In the formulation of subspace projection, this objective of structure-preserving model order reduction can be accomplished by picking the projection matrices

$$\mathcal{X} = \begin{array}{cc} & \begin{array}{cc} n_1 & n_2 \end{array} \\ \begin{array}{c} N_1 \\ N_2 \end{array} & \begin{bmatrix} X_1 & \\ & X_2 \end{bmatrix} \end{array}, \quad \mathcal{Y} = \begin{array}{cc} & \begin{array}{cc} n_1' & n_2' \end{array} \\ \begin{array}{c} N_1' \\ N_2' \end{array} & \begin{bmatrix} Y_1 & \\ & Y_2 \end{bmatrix} \end{array}. \tag{2.12}$$

Then

$$\mathcal{Y}^\mathrm{T}\mathbf{C}\mathcal{X} = \begin{bmatrix} Y_1^\mathrm{T} & \\ & Y_2^\mathrm{T} \end{bmatrix} \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \begin{bmatrix} X_1 & \\ & X_2 \end{bmatrix} = \begin{bmatrix} C_{\mathrm{r},11} & 0 \\ 0 & C_{\mathrm{r},22} \end{bmatrix} = \mathbf{C}_\mathrm{r},$$

$$\mathcal{Y}^\mathrm{T}\mathbf{G}\mathcal{X} = \begin{bmatrix} Y_1^\mathrm{T} & \\ & Y_2^\mathrm{T} \end{bmatrix} \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & 0 \end{bmatrix} \begin{bmatrix} X_1 & \\ & X_2 \end{bmatrix} = \begin{bmatrix} G_{\mathrm{r},11} & G_{\mathrm{r},12} \\ G_{\mathrm{r},21} & 0 \end{bmatrix} = \mathbf{G}_\mathrm{r},$$

$$\mathcal{Y}^\mathrm{T}\mathbf{B} = \begin{bmatrix} Y_1^\mathrm{T} & \\ & Y_2^\mathrm{T} \end{bmatrix} \begin{bmatrix} B_1 \\ 0 \end{bmatrix} = \begin{bmatrix} B_{\mathrm{r},1} \\ 0 \end{bmatrix} = \mathbf{B}_\mathrm{r},$$

$$\mathcal{X}^\mathrm{T}\mathbf{L} = \begin{bmatrix} X_1^\mathrm{T} & \\ & X_2^\mathrm{T} \end{bmatrix} \begin{bmatrix} L_1 \\ 0 \end{bmatrix} = \begin{bmatrix} L_{\mathrm{r},1} \\ 0 \end{bmatrix} = \mathbf{L}_\mathrm{r}.$$

For the case when $\mathcal{Y}$ is taken to be the same as $\mathcal{X}$, this idea is exactly the so-called "*split congruence transformations*" in [8]. A discussion of this idea in a general framework was described in [9].

We now discuss a generic algorithm to generate the desired projection matrices $\mathcal{X}$ and $\mathcal{Y}$ as in (2.12). Suppose that we have computed the basis matrices $\widetilde{X}$ and $\widetilde{Y}$ by, e.g., a block Arnoldi procedure [15], such that

$$\mathcal{K}_k(\mathbf{G}^{-1}\mathbf{C}, \mathbf{G}^{-1}\mathbf{B}) \subseteq \operatorname{span}\left\{\widetilde{X}\right\}$$

and

$$\mathcal{K}_j(\mathbf{G}^{-\mathrm{T}}\mathbf{C}^\mathrm{T}, \mathbf{G}^{-\mathrm{T}}\mathbf{L}) \subseteq \operatorname{span}\left\{\widetilde{Y}\right\}.$$

In general, $\widetilde{X}$ and $\widetilde{Y}$ generated by an Arnoldi process do not have the shape as the desired $\mathcal{X}$ and $\mathcal{Y}$ in, e.g., (2.12), and thus taking $\mathcal{X} = \widetilde{X}$ and $\mathcal{Y} = \widetilde{Y}$ will not preserve the 2-by-2 block structure presented in the matrices as in (2.10). So instead of simply taking $\mathcal{X} = \widetilde{X}$ and $\mathcal{Y} = \widetilde{Y}$, we need to seek $\mathcal{X}$ and $\mathcal{Y}$ having the form as in (2.12) and in the meantime satisfying

$$\text{span}\left\{\widetilde{X}\right\} \subseteq \text{span}\left\{\mathcal{X}\right\} \text{ and } \text{span}\left\{\widetilde{Y}\right\} \subseteq \text{span}\left\{\mathcal{Y}\right\} \tag{2.13}$$

so that the first $k+j$ moments of $H(s)$ and its reduced function $H_{\mathrm{r}}(s)$ match as claimed by Theorem 1.

This task can be accomplished by the following algorithm that for a given $\widetilde{Z} = \begin{bmatrix} \widetilde{Z}_1 \\ \widetilde{Z}_2 \end{bmatrix}$, it computes $\mathcal{Z} = \begin{bmatrix} Z_1 & \\ & Z_2 \end{bmatrix}$ satisfying

$$\text{span}\{\widetilde{Z}\} \subseteq \text{span}\{\mathcal{Z}\}.$$

Algorithm 1

```
1. Compute $Z_i$ having full column rank such that
   span$\{\widetilde{Z}_i\} \subseteq$ span$\{Z_i\}$;
2. Output $\mathcal{Z} = \begin{bmatrix} Z_1 & \\ & Z_2 \end{bmatrix}$.
```

Remark 2. There are a variety of ways to realize Step 1: Rank revealing QR decompositions, modified Gram-Schmidt process, or singular value decompositions [3, 4, 7]. For maximum efficiency, one should make Z_i have as fewer columns as one can. Notice the smallest possible number is rank$(\widetilde{Z}_i)$, but one may have to add a few extra columns to make sure the total number of columns in all X_i and that in all Y_i are the same when constructing $\mathcal{X}$ and $\mathcal{Y}$ below in (2.14).

For convenience, we introduce notation $\rightsquigarrow$ that transforms $\widetilde{Z}$ to $\mathcal{Z}$, i.e.,

$$\widetilde{Z} = \begin{bmatrix} \widetilde{Z}_1 \\ \widetilde{Z}_2 \end{bmatrix} \rightsquigarrow \mathcal{Z} = \begin{bmatrix} Z_1 & \\ & Z_2 \end{bmatrix} \text{ satisfying span}\{\widetilde{Z}\} \subseteq \text{span}\{\mathcal{Z}\}.$$

Return to the subspace embedding objective (2.13), by Algorithm 1, we partition $\widetilde{X}$ and $\widetilde{Y}$ as

$$\widetilde{X} = \begin{bmatrix} \widetilde{X}_1 \\ \widetilde{X}_2 \end{bmatrix} \text{ and } \widetilde{Y} = \begin{bmatrix} \widetilde{Y}_1 \\ \widetilde{Y}_2 \end{bmatrix}$$

consistently with the block structures in $\mathbf{G}$, $\mathbf{C}$, $\mathbf{L}$, and $\mathbf{B}$, and then perform

$$\begin{bmatrix} \widetilde{X}_1 \\ \widetilde{X}_2 \end{bmatrix} \rightsquigarrow \mathcal{X} = \begin{bmatrix} X_1 & \\ & X_2 \end{bmatrix} \text{ and } \begin{bmatrix} \widetilde{Y}_1 \\ \widetilde{Y}_2 \end{bmatrix} \rightsquigarrow \mathcal{Y} = \begin{bmatrix} Y_1 & \\ & Y_2 \end{bmatrix}. \tag{2.14}$$

There are numerically more efficient alternatives when further characteristics in the sub-blocks in $\mathbf{G}$ and $\mathbf{C}$ is known. For example, when $\mathbf{G}$ and $\mathbf{C}$ are as in (2.6) from linearizing a transfer function like (2.3), X_1 and Y_1 can be computed directly via the Second-Order Arnoldi process (SOAR) [1, 2]. More detail is in the next section.

3 Structure of Krylov subspace and Arnoldi process

The generic Algorithm 1 presents a way to preserve the sub-block structure in the reduced systems by first computing the bases of the related Krylov subspaces and then splitting the basis matrices. In this section, we discuss a situation when this first-computing-then-splitting can be combined into one to generate the desired $\mathcal{X}$ and $\mathcal{Y}$ directly. This is made possible by taking advantage of a structural property of Krylov subspaces for certain block matrix. The next theorem was implicitly implied in [2, 16] (see also [9]).

Theorem 2. *Suppose that A and B admit the following partitioning*

$$A = \begin{array}{c c} & \begin{array}{cc} {\scriptstyle N} & {\scriptstyle N} \end{array} \\ \begin{array}{c} {\scriptstyle N} \\ {\scriptstyle N} \end{array} & \begin{bmatrix} A_{11} & A_{12} \\ \alpha I & 0 \end{bmatrix} \end{array}, \quad B = \begin{array}{c c} & {\scriptstyle p} \\ \begin{array}{c} {\scriptstyle N} \\ {\scriptstyle N} \end{array} & \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \end{array}, \tag{3.1}$$

where α is a scalar. Let a basis matrix $\widetilde{X}$ of the Krylov subspace $\mathcal{K}_k(A, B)$ be partitioned as

$$\widetilde{X} = \begin{array}{c c} \begin{array}{c} {\scriptstyle N} \\ {\scriptstyle N} \end{array} & \begin{bmatrix} \widetilde{X}_1 \\ \widetilde{X}_2 \end{bmatrix} \end{array}.$$

Then

$$\operatorname{span}\{\widetilde{X}_2\} \subseteq \operatorname{span}\{B_2, \widetilde{X}_1\}.$$

In particular if also $B_2 = 0$, then $\operatorname{span}\{\widetilde{X}_2\} \subseteq \operatorname{span}\{\widetilde{X}_1\}$.

This theorem provides a theoretical foundation to simply compute $\widetilde{X}_1$, then expand $\widetilde{X}_1$ to X_1 so that $\operatorname{span}\{X_1\} = \operatorname{span}\{B_2, \widetilde{X}_1\}$ (by orthogonalizing B_2's columns against $\widetilde{X}_1$'columns), and finally set

$$\mathcal{X} = \begin{bmatrix} X_1 & \\ & X_1 \end{bmatrix}.$$

In practice, X_1 can be computed directly by a structured Arnoldi process, referred to as the *second-order Arnoldi process* (SOAR) in [1, 2, 16], as given below.

Algorithm 2 Structured Arnoldi process (framework)

Input: A **and** B **as in (3.1).**
Output: $\widetilde{X}_1$ **as in Theorem 2 and** X_1 **with** $\mathrm{span}\{X_1\} = \mathrm{span}\{B_2, \widetilde{X}_1\}$**.**

```
$B_1 = Q_1 R$ (QR decomposition)
$P_1 = \alpha B_2 R_2^{-1}$
for $j = 1, 2, \ldots, k$ do
    $T = A_{11} Q_j + A_{12} P_j$
    $S = \alpha Q_j$
    for $i = 1, 2, \ldots, j$ do
        $Z = Q_i^{\mathrm{T}} T$
        $T = T - Q_i Z$
        $S = S - P_i Z$
    enddo
    $T = Q_j R$ (QR decomposition)
    $P_j = S R^{-1}$
enddo
$\widetilde{X}_1 = [Q_1, Q_2, \ldots, Q_k]$
$T = B_2;$
for $j = 1, 2, \ldots, k$ do
    $Z = Q_j^{\mathrm{T}} T$
    $T = T - Q_j Z$
enddo
$T = QR$ (QR decomposition)
$X_1 = [\widetilde{X}_1, Q]$
```

Remark 3. Algorithm 2 is a simple-minded version to illustrate the key ingredients. Practical implementation will have to incorporate the possibility when various QR decompositions produce (nearly) singular upper triangular matrices R.

4 RCL and RCS systems

4.1 Basic equations

The MNA (*modified nodal analysis*) formulation [20] of an RCL circuit network in frequency domain is of the form

$$\begin{cases} \left(s \begin{bmatrix} C & 0 \\ 0 & L \end{bmatrix} + \begin{bmatrix} G & E \\ -E^{\mathrm{T}} & 0 \end{bmatrix} \right) \begin{bmatrix} v(s) \\ i(s) \end{bmatrix} = \begin{bmatrix} B_v \\ 0 \end{bmatrix} u(s), \\ \qquad\qquad\qquad\qquad y(s) = \begin{bmatrix} D_v^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} v(s) \\ i(s) \end{bmatrix}, \end{cases} \tag{4.1}$$

where $v(s)$ and $i(s)$ denote N_1 nodal voltage and N_2 auxiliary branch currents, respectively; u and y are the input current sources and output voltages; B_v and D_v denote the incidence matrices for the input current sources and output node voltages; C, L and G represent the contributions of the capacitors, inductors and resistors, respectively; and E is the incidence matrix for the inductances.

When an RCL network is modeled with a 3-D extraction method for interconnection analysis, the resulted inductance matrix L is usually very large and dense [10]. This may cause great difficulties to the subsequent simulation process. As an alternative approach, we can use the susceptance matrix $S = L^{-1}$, which is sparse after dropping small entries [5, 22]. The resulting equations are called *the RCS equations*:

$$\left\{\begin{aligned} \left(s\begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} G & E \\ -SE^{\mathrm{T}} & 0 \end{bmatrix}\right)\begin{bmatrix} v(s) \\ i(s) \end{bmatrix} &= \begin{bmatrix} B_v \\ 0 \end{bmatrix} u(s), \\ y(s) &= \begin{bmatrix} D_v^{\mathrm{T}} & 0 \end{bmatrix}\begin{bmatrix} v(s) \\ i(s) \end{bmatrix}. \end{aligned}\right. \tag{4.2}$$

Accordingly, the equations in (4.1) are called *the RCL equations.*

Eliminating the branch current variable $i(s)$ of the RCL and RCS equations in (4.1) and (4.2), we have the so-called *second-order form*

$$\left\{\begin{aligned} \left(sC + G + \tfrac{1}{s}\Gamma\right) v(s) &= B_v u(s), \\ y(s) &= D_v^{\mathrm{T}} v(s), \end{aligned}\right. \tag{4.3}$$

where

$$\Gamma = EL^{-1}E^{\mathrm{T}} = ESE^{\mathrm{T}}.$$

The transfer function $H(s)$ of the RCL and RCS equations in (4.1) and (4.2) can thus be written as

$$H(s) = D_v^{\mathrm{T}} \left(sC + G + \frac{1}{s}\Gamma\right)^{-1} B_v. \tag{4.4}$$

Perform the shift (2.2) to get

$$\begin{aligned} H(s) &= s\, D_v^{\mathrm{T}} (s^2 C + sG + \Gamma)^{-1} B_v \\ &= (s_0 + \sigma)\, D_v^{\mathrm{T}} [\sigma^2 C + \sigma(2s_0 C + G) + (s_0^2 C + s_0 G + \Gamma)]^{-1} B_v \\ &= (s_0 + \sigma)\, \mathbf{L}^{\mathrm{T}} (\sigma \mathbf{C} + \mathbf{G})^{-1} \mathbf{B}, \end{aligned}$$

where

$$\mathbf{C} = \begin{bmatrix} G_0 & C \\ W & 0 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \Gamma_0 & 0 \\ 0 & -W \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} D_v \\ 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} B_v \\ 0 \end{bmatrix}, \tag{4.5}$$

and $G_0 = 2s_0 C + G$, $\Gamma_0 = s_0^2 C + s_0 G + \Gamma$ and W is any *nonsingular* matrix.

4.2 Model order reduction

The SPRIM method described in the chapter by Freund of this book provides a structure-preserving model order reduction method for the RCL equations in (4.1). In this section, we discuss an alternative structure-preserving method for the RCL equations in (4.1) and the the RCS equations in (4.2) using the framework presented in Sections 2 and 3. The method is referred to as *the SAPOR method*, initially published in [17, 11]. The SAPOR method exploits their second-order form (4.3).

For the system matrices $\mathbf{C}$, $\mathbf{G}$ and $\mathbf{B}$ in (4.5), we have[5]

$$\mathbf{G}^{-1}\mathbf{C} = \begin{bmatrix} \Gamma_0^{-1}G_0 & \Gamma_0^{-1}C \\ -I & 0 \end{bmatrix}, \quad \mathbf{G}^{-1}\mathbf{B} = \begin{bmatrix} \Gamma_0^{-1}B_v \\ 0 \end{bmatrix}.$$

They have the same block structures that Theorem 2 requires. Apply Algorithm 2 to compute $X_{\rm r}$ with orthonormal columns such that

$$\mathcal{K}_k(\mathbf{G}^{-1}\mathbf{C}, \mathbf{G}^{-1}\mathbf{B}) \subset \operatorname{span}\left\{ \begin{bmatrix} X_{\rm r} & \\ & X_{\rm r} \end{bmatrix} \right\} \tag{4.6}$$

which is needed by Theorem 1 for matching the first k moments. The framework of the projection technique in Section 2 can also be viewed as a change-of-variables

$$v(s) \approx X_{\rm r} v_{\rm r}(s), \tag{4.7}$$

where $v_{\rm r}(s)$ is a vector of dimension n. Substituting (4.7) into (4.3), and multiplying the first equation in (4.3) by $X_{\rm r}^{\rm T}$ from the left yields the reduced-order model of the second-order equations in (4.3):

$$\begin{cases} \left(sC_{\rm r} + G_{\rm r} + \dfrac{1}{s}\Gamma_{\rm r}\right) v_{\rm r}(s) = B_{{\rm r},v} u(s), \\ \qquad\qquad\qquad\qquad \widetilde{y}(s) = D_{{\rm r},v}^{\rm T} v_{\rm r}(s), \end{cases} \tag{4.8}$$

where

$$C_{\rm r} = X_{\rm r}^{\rm T} C X_{\rm r}, \; G_{\rm r} = X_{\rm r}^{\rm T} G X_{\rm r}, \; \Gamma_{\rm r} = E_{\rm r}^{\rm T} \Gamma E_{\rm r}, \; E_{\rm r} = X_{\rm r}^{\rm T} E, \tag{4.9}$$

and

$$B_{{\rm r},v} = X_{\rm r}^{\rm T} B_v, \; D_{{\rm r},v} = X_{\rm r}^{\rm T} D_v.$$

The transfer function of the reduced system (4.8) is given by

$$H_{\rm r}(s) = D_{{\rm r},v}^{\rm T} \left(sC_{\rm r} + G_{\rm r} + \frac{1}{s}\Gamma_{\rm r}\right)^{-1} B_{{\rm r},v}. \tag{4.10}$$

The reduced second-order form (4.8) corresponds to a reduced order system of the original RCS equations in (4.2). This can be seen by setting

[5] To preserve the symmetry in C, G, and Γ as by (4.9), we do not need a Krylov subspace of $\mathbf{G}^{-\rm T}\mathbf{C}^{\rm T}$ on $\mathbf{G}^{-\rm T}\mathbf{L}$.

$$\mathcal{X}=\mathcal{Y}=\begin{array}{cc} & \begin{array}{cc} {\scriptstyle n} & {\scriptstyle N_2} \end{array} \\ \begin{array}{c} {\scriptstyle N_1} \\ {\scriptstyle N_2} \end{array} & \begin{bmatrix} X_\mathrm{r} & \\ & I \end{bmatrix} \end{array},$$

and then projecting the original RCS equations in (4.2) as in Section 2 to get the reduced order equations

$$\begin{cases} \left(s\begin{bmatrix} C_\mathrm{r} & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} G_\mathrm{r} & E_\mathrm{r} \\ -SE_\mathrm{r}^\mathrm{T} & 0 \end{bmatrix}\right)\begin{bmatrix} v_\mathrm{r}(s) \\ \tilde{i}(s) \end{bmatrix} = \begin{bmatrix} B_{\mathrm{r},v} \\ 0 \end{bmatrix} u(s), \\ \qquad\qquad\qquad\qquad \tilde{y}(s) = \begin{bmatrix} D_{\mathrm{r},v}^\mathrm{T} & 0 \end{bmatrix}\begin{bmatrix} v_\mathrm{r}(s) \\ \tilde{i}(s) \end{bmatrix}. \end{cases} \tag{4.11}$$

Note that $\tilde{i}(s)$ is a vector of N_2 components, the same as the original auxiliary branch currents $i(s)$.

4.3 Towards a synthesized system

The reduced system (4.11) preserves the block structures and the symmetry of system data matrices of the original RCS system (4.2). However, the matrix E_r in the reduced-order RCS system (4.11) cannot be interpreted as an incidence matrix. Towards the objective of synthesis based on the reduced-order model, we shall reformulate the projection (4.7) and the reduced-order system (4.8). This work was first published in [21].

We begin with the original RCS equations in (4.2). Let

$$\widehat{i}(s) = E\, i(s). \tag{4.12}$$

Then the RCS equations in (4.2) can be written as

$$\begin{cases} \left(s\begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} G & I \\ -\Gamma & 0 \end{bmatrix}\right)\begin{bmatrix} v(s) \\ \widehat{i}(s) \end{bmatrix} = \begin{bmatrix} B_v \\ 0 \end{bmatrix} u(s), \\ \qquad\qquad\qquad\qquad y(s) = \begin{bmatrix} D_v^\mathrm{T} & 0 \end{bmatrix}\begin{bmatrix} v(s) \\ \widehat{i}(s) \end{bmatrix}. \end{cases} \tag{4.13}$$

With the change-of-variables (4.12), the incidence matrix E in the original RCS equations in (4.2) is now the identity matrix I in (4.13). The matrix Γ plays the role of the susceptance matrix. An identity incidence matrix can be interpreted as "*self-inductance*", although the susceptance matrix Γ is not diagonal yet. We will discuss how to do so later in this subsection.

Note that the new current vector $\widehat{i}(s)$ is of the size N_1, typically $N_1 \geq N_2$. The order of the new RCS equations in (4.13) is $2N_1$. The equations in (4.2) and (4.13) have the same voltage variables and the same output. However, they are not equivalent since the current variables $i(s)$ cannot be recovered from $\widehat{i}(s)$. The reformulated equations in (4.13) are referred to as *the expanded RCS equations*, or RCSe for short.

In the first-order form, the transfer function $H(s)$ of the RCSe equations in (4.13) are given by

$$H(s) = \mathbf{L}^{\mathrm{T}}(s\mathbf{C} + \mathbf{G})^{-1}\mathbf{B}, \tag{4.14}$$

where $\mathbf{G}$ and $\mathbf{C}$ are $2N_1 \times 2N_1$:

$$\mathbf{C} = \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}, \; \mathbf{G} = \begin{bmatrix} G & I \\ -\mathit{\Gamma} & 0 \end{bmatrix},$$

and

$$\mathbf{B} = \begin{bmatrix} B_v \\ 0 \end{bmatrix}, \; \mathbf{L} = \begin{bmatrix} D_v \\ 0 \end{bmatrix}.$$

For the reduced-order model of the RCSe equations in (4.13), let us define

$$\mathcal{X} = \mathcal{Y} = \begin{array}{c} \\ {\scriptstyle N_1} \\ {\scriptstyle N_1} \end{array} \!\! \begin{array}{c} \begin{array}{cc} {\scriptstyle n} & {\scriptstyle n} \end{array} \\ \begin{bmatrix} X_{\mathrm{r}} & \\ & X_{\mathrm{r}} \end{bmatrix} \end{array}. \tag{4.15}$$

Then by the change-of-variables

$$v(s) \approx X_{\mathrm{r}}^{\mathrm{T}} v_{\mathrm{r}}(s) \quad \text{and} \quad \widehat{i}(s) \approx X_{\mathrm{r}}^{\mathrm{T}} i_{\mathrm{r}}(s), \tag{4.16}$$

and using the projection procedure in Section 2, we have the reduced-order RCSe equations

$$\left\{ \begin{array}{r} \left(s \begin{bmatrix} C_{\mathrm{r}} & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} G_{\mathrm{r}} & I \\ -\mathit{\Gamma}_{\mathrm{r}} & 0 \end{bmatrix} \right) \begin{bmatrix} v_{\mathrm{r}}(s) \\ i_{\mathrm{r}}(s) \end{bmatrix} = \begin{bmatrix} B_{\mathrm{r},v} \\ 0 \end{bmatrix} u(s), \\ \widetilde{y}(s) = \begin{bmatrix} D_{\mathrm{r},v}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} v_{\mathrm{r}}(s) \\ i_{\mathrm{r}}(s) \end{bmatrix}. \end{array} \right. \tag{4.17}$$

Compared with the RCSe equations in (4.13), the reduced equations in (4.17) not only preserve the 2-by-2 block structure of the system data matrices $\mathbf{G}$ and $\mathbf{C}$, but also preserve the identity of the incidence matrix.

For the objective of synthesis of the original RCL and RCS equations in (4.1) and (4.2), let us further consider the structures of the input and output matrices and the incidence matrix. Without loss of generality, we assume that the sub-blocks B_v and D_v in the input and output of the RCS equations in (4.2) are of the forms:

$$B_v = \begin{array}{c} \\ {\scriptstyle p_1} \\ {\scriptstyle N_1 - p_1} \end{array} \!\! \begin{array}{c} {\scriptstyle p} \\ \begin{bmatrix} B_{v1} \\ 0 \end{bmatrix} \end{array}, \quad D_v = \begin{array}{c} \\ {\scriptstyle p_1} \\ {\scriptstyle N_1 - p_1} \end{array} \!\! \begin{array}{c} {\scriptstyle m} \\ \begin{bmatrix} D_{v1} \\ 0 \end{bmatrix} \end{array}. \tag{4.18}$$

It indicates that there are totally p_1 different input and output nodes. Otherwise we can reorder the nodes in the RLC/RCS circuit network such that B_v and D_v are in the desired forms.

Furthermore, we assume that the incidence matrix E in (4.2) has the zero block on the top, conformal with the partition of the input and output matrices in (4.18):

$$E = \begin{array}{cc} & N_2 \\ \begin{array}{c} p_1 \\ N_1-p_1 \end{array} & \left[\begin{array}{c} 0 \\ \tilde{E} \end{array}\right] \end{array}. \tag{4.19}$$

This assumption means that there is no susceptance (inductor) directly connecting to the input and output nodes [21].

With the assumptions in (4.18) and (4.19), let $X_{\rm r}$ be an orthonormal basis for the projection subspace in (4.6). Using partitioning-and-embedding steps in Algorithm 1 of Section 2, we have

$$X_{\rm r} = \begin{array}{cc} & n \\ \begin{array}{c} p_1 \\ N_1-p_1 \end{array} & \left[\begin{array}{c} X_{\rm r}^{(1)} \\ X_{\rm r}^{(2)} \end{array}\right] \end{array} \rightsquigarrow \widehat{X}_{\rm r} = \begin{array}{cc} & \begin{array}{cc} p_1 & n \end{array} \\ \begin{array}{c} p_1 \\ N_1-p_1 \end{array} & \left[\begin{array}{cc} I & \\ & X_2 \end{array}\right] \end{array},$$

where the columns of X_2 form an orthonormal basis for the range of $X_{\rm r}^{(2)}$. For simplicity, we assume that there is no deflation, namely, $\mathrm{rank}(X_{\rm r}^{(2)}) = \mathrm{rank}(X_2) = n$.

Similarly to (4.15) and (4.16), using the subspace projection with

$$\mathcal{X} = \mathcal{Y} = \begin{array}{cc} & \begin{array}{cc} p_1+n & p_1+n \end{array} \\ \begin{array}{c} N_1 \\ N_1 \end{array} & \left[\begin{array}{cc} \widehat{X}_{\rm r} & \\ & \widehat{X}_{\rm r} \end{array}\right] \end{array},$$

we have the reduced-order RCSe equations

$$\left\{ \begin{aligned} \left(s \begin{bmatrix} C_{\rm r} & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} G_{\rm r} & I \\ -\mathit{\Gamma}_{\rm r} & 0 \end{bmatrix} \right) \begin{bmatrix} v_{\rm r}(s) \\ i_{\rm r}(s) \end{bmatrix} &= \begin{bmatrix} B_{{\rm r},v} \\ 0 \end{bmatrix} u(s), \\ \widetilde{y}(s) &= \begin{bmatrix} D_{{\rm r},v}^{\rm T} & 0 \end{bmatrix} \begin{bmatrix} v_{\rm r}(s) \\ i_{\rm r}(s) \end{bmatrix}, \end{aligned} \right. \tag{4.20}$$

where $C_{\rm r}$, $G_{\rm r}$ and $\mathit{\Gamma}_{\rm r}$ are $(p_1+n) \times (p_1+n)$ matrices:

$$C_{\rm r} = \widehat{X}_{\rm r}^{\rm T} C \widehat{X}_{\rm r}, \; G_{\rm r} = \widehat{X}_{\rm r}^{\rm T} G \widehat{X}_{\rm r}, \; \mathit{\Gamma}_{\rm r} = \widehat{X}_{\rm r}^{\rm T} \mathit{\Gamma} \widehat{X}_{\rm r},$$

and the input and output sub-block matrices $B_{{\rm r},v}$ and $D_{{\rm r},v}$ preserve the structure in (4.18):

$$B_{{\rm r},v} = \widehat{X}_{\rm r}^{\rm T} \begin{bmatrix} B_{v1} \\ 0 \end{bmatrix} = \begin{array}{cc} & p \\ \begin{array}{c} p_1 \\ n \end{array} & \left[\begin{array}{c} B_{v1} \\ 0 \end{array}\right] \end{array}, \quad D_{{\rm r},v} = \widehat{X}_{\rm r}^{\rm T} \begin{bmatrix} D_{v1} \\ 0 \end{bmatrix} = \begin{array}{cc} & m \\ \begin{array}{c} p_1 \\ n \end{array} & \left[\begin{array}{c} D_{v1} \\ 0 \end{array}\right] \end{array}.$$

Note that

$$\operatorname{span}\left\{\begin{bmatrix} X_{\mathrm{r}} & \\ & X_{\mathrm{r}} \end{bmatrix}\right\} \subseteq \operatorname{span}\left\{\begin{bmatrix} \widehat{X}_{\mathrm{r}} & \\ & \widehat{X}_{\mathrm{r}} \end{bmatrix}\right\}.$$

The reduced RCSe system (4.20) also preserves the moment-matching property for the system (4.17) by Theorem 1.

Now we turn to the diagonalization of $\mathit{\Gamma}$ in the RCSe equations in (4.13) for the objective of synthesis. The assumption (4.19) for the incidence matrix E implies that $\mathit{\Gamma}$ is of the form

$$\mathit{\Gamma} = EL^{-1}E^{\mathrm{T}} = \begin{array}{c|c} & \begin{array}{cc} p_1 & N_1-p_1 \end{array} \\ \begin{array}{c} p_1 \\ N_1-p_1 \end{array} & \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{\mathit{\Gamma}} \end{bmatrix} \end{array}.$$

It can be seen that in the reduced RCSe equations in (4.20), $\mathit{\Gamma}_{\mathrm{r}}$ has the same form

$$\mathit{\Gamma}_{\mathrm{r}} = \begin{array}{c|c} & \begin{array}{cc} p_1 & n \end{array} \\ \begin{array}{c} p_1 \\ n \end{array} & \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{\mathit{\Gamma}}_{\mathrm{r}} \end{bmatrix} \end{array},$$

where $\widetilde{\mathit{\Gamma}}_{\mathrm{r}} = Q_2^{\mathrm{T}}\widetilde{\mathit{\Gamma}}Q_2$. Note that $\widetilde{\mathit{\Gamma}}$ is symmetric semi-positive definite, so is $\widetilde{\mathit{\Gamma}}_{\mathrm{r}}$. Let

$$\widetilde{\mathit{\Gamma}}_{\mathrm{r}} = \widetilde{V}\mathit{\Lambda}\widetilde{V}^{\mathrm{T}}$$

be the eigen-decomposition of $\widetilde{\mathit{\Gamma}}_{\mathrm{r}}$, where V is orthogonal and $\mathit{\Lambda}$ is diagonal. Define

$$V = \begin{array}{c|c} & \begin{array}{cc} p_1+n & p_1+n \end{array} \\ \begin{array}{c} p_1+n \\ p_1+n \end{array} & \begin{bmatrix} \widehat{V} & \\ & \widehat{V} \end{bmatrix} \end{array},$$

where

$$\widehat{V} = \begin{array}{c|c} & \begin{array}{cc} p_1 & n \end{array} \\ \begin{array}{c} p_1 \\ n \end{array} & \begin{bmatrix} I & \\ & \widetilde{V} \end{bmatrix} \end{array}.$$

Then by a congruence transformation using the matrix V, the reduced-order RCSe equations in (4.20) is equivalent to the equations

$$\begin{cases} \left(s\begin{bmatrix} \widehat{C}_{\mathrm{r}} & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} \widehat{G}_{\mathrm{r}} & I \\ -\widehat{\mathit{\Gamma}}_{\mathrm{r}} & 0 \end{bmatrix}\right)\begin{bmatrix} \widehat{v}_{\mathrm{r}}(s) \\ \widehat{i}_{\mathrm{r}}(s) \end{bmatrix} = \begin{bmatrix} \widehat{B}_{\mathrm{r},v} \\ 0 \end{bmatrix} u(s), \\ \qquad\qquad\qquad \widehat{y}(s) = \begin{bmatrix} \widehat{D}_{\mathrm{r},v}^{\mathrm{T}} & 0 \end{bmatrix}\begin{bmatrix} \widehat{v}_{\mathrm{r}}(s) \\ \widehat{i}_{\mathrm{r}}(s) \end{bmatrix}, \end{cases} \tag{4.21}$$

where $\widehat{v}_{\mathrm{r}}(s) = \widehat{V}^{\mathrm{T}}v_{\mathrm{r}}(s)$ and $\widehat{i}_{\mathrm{r}}(s) = \widehat{V}^{\mathrm{T}}i_{\mathrm{r}}(s)$. $\widehat{C}_{\mathrm{r}}$, $\widehat{G}_{\mathrm{r}}$ and $\widehat{\mathit{\Gamma}}_{\mathrm{r}}$ are $(p_1+n)\times(p_1+n)$ matrices:

$$\widehat{C}_{\mathrm{r}} = \widehat{V}^{\mathrm{T}}C_{\mathrm{r}}\widehat{V}, \quad \widehat{G}_{\mathrm{r}} = \widehat{V}^{\mathrm{T}}G_{\mathrm{r}}\widehat{V}, \quad \widehat{\mathit{\Gamma}}_{\mathrm{r}} = \widehat{V}^{\mathrm{T}}\mathit{\Gamma}_{\mathrm{r}}\widehat{V}.$$

Moreover with V being block diagonal, the input and output structures are preserved, too:

$$\widehat{B}_{\mathrm{r},v} = \widehat{V}^{\mathrm{T}} B_{\mathrm{r},v} = \begin{array}{cc} & p \\ \begin{matrix} p_1 \\ n \end{matrix} & \begin{bmatrix} B_{v1} \\ 0 \end{bmatrix} \end{array}, \quad \widehat{D}_{\mathrm{r},v} = \widehat{V}^{\mathrm{T}} D_{\mathrm{r},v} = \begin{array}{cc} & p \\ \begin{matrix} p_1 \\ n \end{matrix} & \begin{bmatrix} D_{v1} \\ 0 \end{bmatrix} \end{array}.$$

We note that after the congruence transformation, $\widehat{\mathit{\Gamma}}_r$ is diagonal

$$\widehat{\mathit{\Gamma}}_{\mathrm{r}} = \begin{array}{cc} & \begin{matrix} p_1 & n \end{matrix} \\ \begin{matrix} p_1 \\ n \end{matrix} & \begin{bmatrix} 0 & 0 \\ 0 & \mathit{\Lambda} \end{bmatrix} \end{array}$$

Therefore, to avoid large entries in the synthesized inductors for synthesized RCL equations, we partition the eigenvalue matrix $\mathit{\Lambda}$ of $\widetilde{\mathit{\Gamma}}_{\mathrm{r}}$ into

$$\mathit{\Lambda} = \begin{array}{cc} & \begin{matrix} \ell & n-\ell \end{matrix} \\ \begin{matrix} \ell \\ n-\ell \end{matrix} & \begin{bmatrix} \mathit{\Lambda}_1 & \\ & \mathit{\Lambda}_2 \end{bmatrix} \end{array},$$

where $\mathit{\Lambda}_2$ contains the $n-\ell$ smallest eigenvalues that are smaller than a given threshold ϵ in magnitude. Setting $\mathit{\Lambda}_2 = 0$, we derive reduced RCSe equations of the same form as in (4.21) with the "susceptance" matrix

$$\widehat{\mathit{\Gamma}}_{\mathrm{r}} = \begin{array}{cc} & \begin{matrix} p_1 & \ell & n-\ell \end{matrix} \\ \begin{matrix} p_1 \\ \ell \\ n-\ell \end{matrix} & \begin{bmatrix} 0 & & \\ & \mathit{\Lambda}_1 & \\ & & 0 \end{bmatrix} \end{array}.$$

Subsequently, we can define reduced-order equations to resemble the RCL form (4.1):

$$\begin{cases} \left(s \begin{bmatrix} \widehat{C}_{\mathrm{r}} & 0 \\ 0 & \widehat{L}_{\mathrm{r}} \end{bmatrix} + \begin{bmatrix} \widehat{G}_{\mathrm{r}} & I \\ -I & 0 \end{bmatrix} \right) \begin{bmatrix} \widehat{v}_{\mathrm{r}}(s) \\ \widehat{i}_{\mathrm{r}}(s) \end{bmatrix} = \begin{bmatrix} B_{\mathrm{r},v} \\ 0 \end{bmatrix} u(s), \\ \qquad\qquad \widetilde{y}(s) = \begin{bmatrix} D_{\mathrm{r},v}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \widehat{v}_{\mathrm{r}}(s) \\ \widehat{i}_{\mathrm{r}}(s) \end{bmatrix}, \end{cases} \tag{4.22}$$

where the inductance matrix $\widehat{L}_{\mathrm{r}}$ is given by

$$\widehat{L}_{\mathrm{r}} = \begin{array}{cc} & \begin{matrix} p_1 & \ell & n-\ell \end{matrix} \\ \begin{matrix} p_1 \\ \ell \\ n-\ell \end{matrix} & \begin{bmatrix} 0 & & \\ & \mathit{\Lambda}_1^{-1} & \\ & & 0 \end{bmatrix} \end{array}.$$

Since L_{r} is diagonal, there is no inductance loop in synthesized network. We refer to the equations in (4.22) as *the synthesized RCL equations* of the original RCL equations in (4.1). The synthesized RCL equations in (4.22) are described in detail by an RCLSYN (RCL equivalent circuit synthesis) tool presented in [21].

Example 3. We consider a 64-bit bus circuit network with 8 inputs and 8 outputs. The order N of the corresponding RCL model is $N = 16963$. By the structure-preserving model order reduction described in this section, we obtain a reduced-order RLC equations of the form (4.22), with order $n = 640$.

SPICE transient analysis are performed on both the original RLC circuit and the synthesized circuit (4.22) with excitations of pulse current sources at eight inputs. The transient simulation results are shown in Figure 1. The transient response of the synthesized circuit is visually indistinguishable from that of the original RLC circuit. SPICE AC analysis was also performed on both the original RLC circuit and the synthesized RLC circuit with current excitation at the near end of the first line. The voltage at the far end of the first line is considered as the observing point. The AC simulation results are shown in Figure 2. We see that two curves are visually indistinguishable. The CPU elapsed time for the transient and AC analysis are shown in the following table

	Full RCL	Synthesized RCL
Dimensionality	16963	640
Transient analysis	5007.59 (sec.)	90.16 (sec.)
AC analysis	29693.02 (sec.)	739.29 (sec.)

From the table, we see that with the reduced RCL equations a factor of 50 of speedup for the transient analysis and a factor of 40 of speedup for the AC analysis have been achieved. ◇

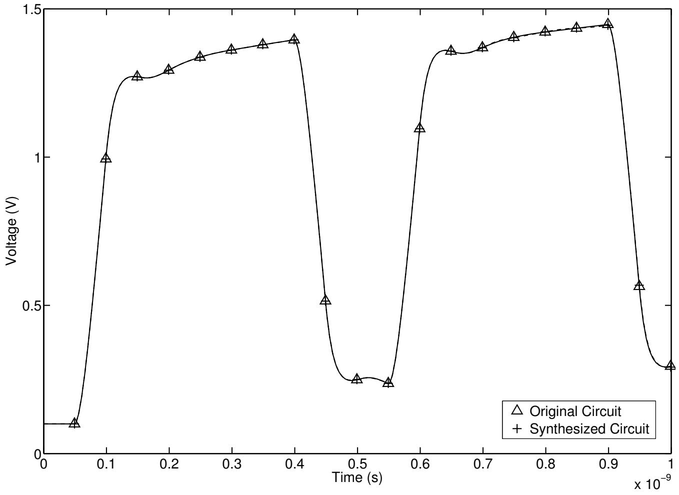

Fig. 1. Transient analysis of the bus circuit

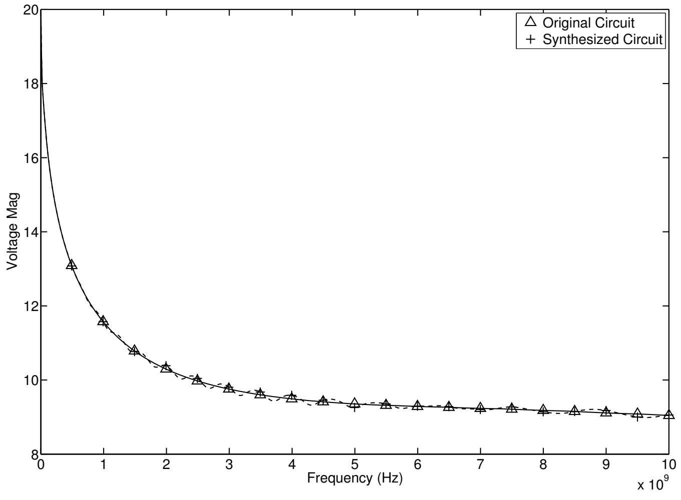

Fig. 2. AC analysis of the bus circuit

Acknowledgment

Bai is supported in part by NSF grant DMS-0611548. Li is supported in part by NSF Grant DMS-0702335. Su is supported in part by Shanghai Dawn Project 200601 and China NSF project 90307017.

References

1. Z. Bai and Y. Su. Dimension reduction of large-scale second-order dynamical systems via a second-order arnoldi method. *SIAM J. Sci. Comput.*, 25(5):1692–1709, 2005.
2. Z. Bai and Y. Su. SOAR: A second-order arnoldi method for the solution of the quadratic eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 26(3):640–659, Mar. 2005.
3. Ake Björck. *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, 1996.
4. J. Demmel. *Applied Numerical Linear Algebra.* SIAM, Philadelphia, 1997.
5. A. Devgan, H. Ji, and W. Dai. How to efficiently capture on-chip inductance effects: Introducing a new circuit element K. In *Proceedings of IEEE/ACM ICCAD 2000*, pages 150–155, 2000.
6. E.J.Grimme. *Krylov Projection Methods for model reduction.* PhD thesis, Univ. Illinois, Urbana-Champaign, 1997.
7. G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, Maryland, 3rd edition, 1996.
8. Kevin J. Kerns and Andrew T. Yang. Preservation of passivity during RLC network reduction via split congruence transformations. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, 17(7):582–591, July 1998.

9. Ren-Cang Li and Zhaojun Bai. Structure-preserving model reduction. *Comm. Math. Sci.*, 3(2):179–199, 2005.
10. D. Ling and A. Ruehli. *Circuit Analysis, Simulation and Design - Advances in CAD for VLSI*, volume 3, chapter 11. Elsevier Science Publisher, 1987.
11. B. Liu, X. Zeng, and Y. Su. Block SAPOR: Block second-order arnoldi method for passive order reduction of multi-input multi-output RCS interconnect circuits. In *Proc. IEEE/ACM ASP-DAC*, pages 244–249, 2005.
12. T. Reis and T. Stykel. Stability analysis and model order reduction of coupled systems. *Math. Comput. Model. Dyn. Syst.*, 13:413–436, 2007.
13. Axel Ruhe. Rational Krylov sequence methods for eigenvalue computation. *Linear Algebra Appl.*, 58:391–405, 1984.
14. Axel Ruhe. Rational Krylov algorithms for nonsymmetric eigenvalue problems. II. matrix pairs. *Linear Algebra Appl.*, 197-198:283–295, 1994.
15. Yousef Saad. *Iterative Methods for Sparse Linear Systems.* SIAM, Philadelphia, 2nd edition, 2003.
16. T.-J. Su and R. R. Craig. Model reduction and control of flexible structures using Krylov vectors. *J. Guidance, Control, and Dynamics*, 14(2):260–267, 1991.
17. 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. IEEE/ACM Int. Conf. Computer Aided Design*, pages 74–79, Nov. 2004.
18. A. Vandendorpe and P. Van Dooren. On model reduction of interconnected systems. Proceedings of the 16th International Symposium of Mathematical Theory of Networks and Systems, Leuven, Belgium, 2004.
19. C. E. Villemagne and R. E. Skelton. Model reduction using a projection formulation. *Internat. J. Control*, 46(6):2141–2169, 1987.
20. J. Vlach and K. Singhal. *Computer Methods for Circuit Analysis and Design.* Van Nostrand Reinhold, New York, second edition, 1994.
21. F. Yang, X. Zeng, Y. Su, and D. Zhou. RLC equivalent circuit synthesis method for structure-preserved reduced-order model of interconnect in VLSI. *Commun. Comput. Phys.*, 3:376–396, February 2008.
22. H. Zheng, B. Krauter, M. Beattie, and L. Pileggi. Window-based susceptance models for large-scale RLC circuit analyses. In *Proceedings of IEEE/ACM DATE 2002*, pages 628–633, 2002.