II.2 Structure Preserving Model Order Reduction

via Krylov Subspace Projection

Outline

A. A unified theory for SPMOR

B. Case study: RCL/RCS circuits

Outline of Part A – a unified SPMOR theory

1. Transfer function in the first-order form
2. Model order reduction
3. The moment-matching theorem
4. SPMOR
   (a) Basic formulation
   (b) A generic algorithm
   (c) Structure of Krylov subspace and structured Arnoldi procedure (framework)

joint with R. C. Li of Univ. of Texas, Arlington

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

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

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

- Assume that $\mathbf{G}$ is nonsingular.

- The transfer function can be expanded around $s = 0$ as

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

  where $M_{\ell} = \mathbf{L}^{\mathsf{T}}(\mathbf{G}^{-1}\mathbf{C})^{\ell}\mathbf{G}^{-1}\mathbf{B}$ 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, a shift

$$s = (s - s_0) + s_0 \equiv \sigma + s_0$$

can be performed and then

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

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 $\sigma = 0$.

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

Example: RCL circuits

- The MNA formulation of RCL circuits in the Integro-DAEs form:

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

- The transfer function of the Integro-DAEs is given by

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

- References: [Freund], [Gad *et al*] in [S-vdV-R]

- Linearization #1:

  Define

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

  for any *nonsingular* matrix W.

  Then the transfer function:

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

- Linearization #2:

  Define

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

  for any *nonsingular* matrix W

  Then the transfer function:

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

- Remarks:
  1. In the linearization #2, the matrix-vector products with the matrices $\mathbf{G}^{-1}\mathbf{C}$ and $\mathbf{G}^{-\mathsf{T}}\mathbf{C}^{\mathsf{T}}$ are much easier to do than the linearization #1.
  2. The linearization #1 favors approximations around $s = \infty$.
  3. In the case when approximations near a finite point $s_0 \neq 0$ are sought, a shift 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.
  4. If the shift is performed *before* linearization, the same advantage as the linearization #2 over the linearization #1 for approximations near $s = 0$ is retained.

Example: Coupled systems

- The transfer function of interconnected (coupled) systems:

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

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

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

- Let $A = \text{diag}(A_1, \ldots, A_k)$, $B = \text{diag}(B_1, \ldots, B_k)$, $L = \text{diag}(L_1, \ldots, L_k)$, Then $H(s)$ can be turned into the first-order form

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

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

- References: [Reis and Stykel] and [Vandendorpe and van Dooren] in [S-vdV-R]

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

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

- Then defines a *reduced-order transfer function*

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

where

$$\mathbf{C}_{\mathrm{r}} = \mathcal{Y}^{\mathsf{T}} \mathbf{C} \mathcal{X}, \quad \mathbf{G}_{\mathrm{r}} = \mathcal{Y}^{\mathsf{T}} \mathbf{G} \mathcal{X}, \quad \mathbf{B}_{\mathrm{r}} = \mathcal{Y}^{\mathsf{T}} \mathbf{B}, \quad \mathbf{L}_{\mathrm{r}} = \mathcal{X}^{\mathsf{T}} \mathbf{L}. \tag{1}$$

- The reduced transfer function $H_{\mathrm{r}}(s)$ can be expanded around $s = 0$:

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

where $M_{\mathrm{r},\ell} = \mathbf{L}_{\mathrm{r}}^{\mathsf{T}}(\mathbf{G}_{\mathrm{r}}^{-1}\mathbf{C}_{\mathrm{r}})^{\ell}\mathbf{G}_{\mathrm{r}}^{-1}\mathbf{B}_{\mathrm{r}}$ are referred to as the *moments* of the reduced system.

- Desired properties:
  1. $n \ll N$.
  2. 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: *Moment matching, stability, passivity, ...*

Krylov subspaces associated with $H(s)$

- Transfer function

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

- Two associated Krylov subspace

  1. *Right* Krylov subspace:

$$\mathcal{K}_k(\mathbf{G}^{-1}\mathbf{C}, \mathbf{G}^{-1}\mathbf{B}) = \\ \text{span}\{\mathbf{G}^{-1}\mathbf{B}, (\mathbf{G}^{-1}\mathbf{C})\mathbf{G}^{-1}\mathbf{B}, \ldots (\mathbf{G}^{-1}\mathbf{C})^{k-1}\mathbf{G}^{-1}\mathbf{B}, \}$$

  2. *Left* Krylov subspace:

$$\mathcal{K}_k(\mathbf{G}^{-\mathsf{T}}\mathbf{C}^{\mathsf{T}}, \mathbf{G}^{-\mathsf{T}}\mathbf{L}) = \\ \text{span}\{\mathbf{G}^{-\mathsf{T}}\mathbf{L}, (\mathbf{G}^{-\mathsf{T}}\mathbf{C}^{\mathsf{T}})\mathbf{G}^{-\mathsf{T}}\mathbf{L}, \ldots (\mathbf{G}^{-\mathsf{T}}\mathbf{C}^{\mathsf{T}})^{k-1}\mathbf{G}^{-\mathsf{T}}\mathbf{L}, \}$$

- Numerical stable computation of the bases of these Krylov subspaces are nontrivial tasks

The following theorem dictates how good a reduced transfer function $H_{\rm r}(s)$ approximates the original transfer function $H(s)$.

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

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

and

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

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

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

which imply

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

Remarks:

- The conditions suggest that by enforcing span$\{\mathcal{X}\}$ and/or span$\{\mathcal{Y}\}$ to contain more appropriate Krylov subspaces associated with multiple points, $H_{\rm r}(s)$ can be made to approximate $H(s)$ well near all those points – *multi-point approximation*.
- When $\mathbf{G} = \mathbf{I}$, it is due to [Villemagne and Skelton'87]
- The general form as stated above was proved by [Grimme'97]
- A different proof is available in [Freund'05]
- A proof using the projection language was given in [Li and B.'05].
- Its implication to structure-preserving model reduction was also realized in [Li and B.'05] and [Freund'05]

- System structure:

  For the simplicity of exposition, consider system matrices $\mathbf{G}$, $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{L}$ having the following 2×2 block structure

$$\mathbf{C} = \begin{array}{c} \\ N_1' \\ N_2' \end{array} \begin{array}{c} \begin{array}{cc} N_1 & N_2 \end{array} \\ \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \end{array}, \quad \mathbf{G} = \begin{array}{c} \\ N_1' \\ N_2' \end{array} \begin{array}{c} \begin{array}{cc} N_1 & N_2 \end{array} \\ \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & 0 \end{bmatrix} \end{array}, \tag{2}$$

$$\mathbf{B} = \begin{array}{c} \\ N_1' \\ N_2' \end{array} \begin{array}{c} \begin{array}{c} p \end{array} \\ \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \end{array}, \quad \mathbf{L} = \begin{array}{c} \\ N_1 \\ N_2 \end{array} \begin{array}{c} \begin{array}{c} m \end{array} \\ \begin{bmatrix} L_1 \\ 0 \end{bmatrix} \end{array},$$

  where $N_1 + N_2 = N_1' + N_2' = N$.

  System matrices from the *time-domain modified nodal analysis* (MNA) circuit equations of RCL circuits take such forms. (*more later*)

- The objectives of SPMOR:
  1. structurally preserves the block structure:

$$\mathbf{C}_{\mathrm{r}} = \begin{array}{c} \\ n_1' \\ n_2' \end{array} \begin{array}{c} \begin{array}{cc} n_1 & n_2 \end{array} \\ \begin{bmatrix} C_{\mathrm{r},11} & 0 \\ 0 & C_{\mathrm{r},22} \end{bmatrix} \end{array}, \quad \mathbf{G}_{\mathrm{r}} = \begin{array}{c} \\ n_1' \\ n_2' \end{array} \begin{array}{c} \begin{array}{cc} n_1 & n_2 \end{array} \\ \begin{bmatrix} G_{\mathrm{r},11} & G_{\mathrm{r},12} \\ G_{\mathrm{r},21} & 0 \end{bmatrix} \end{array}, \tag{3}$$

$$\mathbf{B}_{\mathrm{r}} = \begin{array}{c} \\ n_1' \\ n_2' \end{array} \begin{array}{c} \begin{array}{c} p \end{array} \\ \begin{bmatrix} B_{\mathrm{r},1} \\ 0 \end{bmatrix} \end{array}, \quad \mathbf{L}_{\mathrm{r}} = \begin{array}{c} \\ n_1 \\ n_2 \end{array} \begin{array}{c} \begin{array}{c} m \end{array} \\ \begin{bmatrix} L_{\mathrm{r},1} \\ 0 \end{bmatrix} \end{array},$$

     where $n_1 + n_2 = n_1' + n_2' = n$.
  2. Each sub-block is a direct reduction from the corresponding sub-block in the original system.
- Advantages of SPMOR:
  1. Easily provable preservation of the original system properties, such as stability, passivity, ...
  2. Better numerical stability and accuracy

In the formulation of subspace projection, SPMOR objectives can be accomplished by picking the projection matrices

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

Then

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

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

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

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

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*" [Kerns and Yang'97].

A generic algorithm to generate the desired projection matrices $\mathcal{X}$ and $\mathcal{Y}$:

- Compute the basis matrices $\widetilde{X}$ and $\widetilde{Y}$ such that

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

  and

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

- 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

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

satisfying

$$\text{span}\left\{\widetilde{X}\right\} \subseteq \text{span}\left\{\mathcal{X}\right\} \text{ and } \text{span}\left\{\widetilde{Y}\right\} \subseteq \text{span}\left\{\mathcal{Y}\right\} \qquad (4)$$

Remarks

- The *subspace embedding* task "$\rightsquigarrow$" can be accomplished as follows:
  1. Compute Z_i having full column rank such that $\text{span}\{\widetilde{Z}_i\} \subseteq \text{span}\{Z_i\}$;
  2. Output $\mathcal{Z} = \begin{bmatrix} Z_1 & \\ & Z_2 \end{bmatrix}$.
- There are a variety of ways to realize Step 1: Rank revealing QR decompositions, modified Gram-Schmidt process, or SVD.

- 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 $\mathcal{Y}$.
- There are numerically more efficient alternatives when further characteristics in the sub-blocks in $\mathbf{G}$ and $\mathbf{C}$ is known – (*more later*)
- The first $k + j$ moments of $H(s)$ and the SPMOR transfer function $H_{\mathrm{r}}(s)$ match.

  Proof: a direct consequence of the moment-matching theorem and the generic algorithm.

The *first-computing-then-splitting* can be combined into one to generate the desired $\mathcal{X}$ and $\mathcal{Y}$ directly, by taking advantage of a structural property of Krylov subspaces for certain block matrix.

Theorem. Suppose that A and B admit the following partitioning

$$A = \begin{matrix} \\ N \\ N \end{matrix} \begin{matrix} \begin{matrix} N & N \end{matrix} \\ \begin{bmatrix} A_{11} & A_{12} \\ \alpha I & 0 \end{bmatrix} \end{matrix}, \quad B = \begin{matrix} \\ N \\ N \end{matrix} \begin{matrix} \begin{matrix} p \end{matrix} \\ \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \end{matrix},$$

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

$$\widetilde{X} = \begin{matrix} \\ N \\ N \end{matrix} \begin{bmatrix} \widetilde{X}_1 \\ \widetilde{X}_2 \end{bmatrix}.$$

Then

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

Remarks:

1. This theorem provides a theoretical foundation to simply compute $\widetilde{X}_1$, then expand $\widetilde{X}_1$ to X_1 so that span$\{X_1\}$ = span$\{B_2, \widetilde{X}_1\}$ and finally set
$$\mathcal{X} = \begin{bmatrix} X_1 & \\ & X_1 \end{bmatrix}.$$
2. The theorem was implicitly implied in [Su and Craig'91, Bai and Su'05] and explicitly stated in [Li and Bai'05] for more general cases.
3. For $A = \begin{bmatrix} A_{11} & A_{12} \\ \alpha I & \end{bmatrix}$, X_1 can be computed directly by a structured Arnoldi procedure, such as SOAR.

Outline of Part B – Case study: RCL circuits

- RCL and RCS circuit equations
- Transfer functions
- SPMOR – version 1
- Towards a synthesizable reduced-order RCL system
  - Expanded RCL (RCS) equations
  - Transfer function
  - SPMOR – version 2
  - Preserving I/O ports
  - Diagonalization
  - Reduced-order RCL equation – synthesizable yet?
  - An example

mostly due to Y. Su and X. Zeng group at Fudan Univ., China

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

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

where

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

RCS circuit equations

- 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.
- As an alternative approach, we can use the susceptance matrix $S = L^{-1}$, which is sparse after dropping small entries:
- RCS circuit equations:

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

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

$$\begin{cases} \left(sC + G + \frac{1}{s}\Gamma\right) v(s) = B_v u(s), \\ y(s) = D_v^{\mathsf{T}} v(s), \end{cases}$$

  where

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

- The transfer function $H(s)$:

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

- Perform the shift "$s \rightarrow s_0 + \sigma$" to get

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

where

$$\mathbf{C} = \begin{bmatrix} G_0 & C \\ W & 0 \end{bmatrix}, \ \mathbf{G} = \begin{bmatrix} \Gamma_0 & 0 \\ 0 & -W \end{bmatrix}, \ \mathbf{L} = \begin{bmatrix} D_v \\ 0 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} B_v \\ 0 \end{bmatrix}, \tag{5}$$

and $G_0 = 2s_0C + G$, $\Gamma_0 = s_0^2C + s_0G + \Gamma$ and W is any *nonsingular* matrix.

- The SPRIM method [Freund'04] provides a SPMOR model for the RCL equations.
- The following is an alternative SPMOR model, referred to as *the SAPOR method* [Yang et al'04].
- For the system matrices $\mathbf{C}$, $\mathbf{G}$ and $\mathbf{B}$,

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

- By using the block structure of $\mathbf{G}$, and applying the SOAR, we can generate X_{r} with orthonormal columns such that

$$\mathcal{K}_k(\mathbf{G}^{-1}\mathbf{C}, \mathbf{G}^{-1}\mathbf{B}) \subseteq \text{span} \left\{ \begin{bmatrix} X_{\mathrm{r}} & \\ & X_{\mathrm{r}} \end{bmatrix} \right\}$$

- The subspace projection technique can be viewed as a

change-of-variable:

$$v(s) \approx X_{\mathrm{r}} v_{\mathrm{r}}(s),$$

where $v_{\mathrm{r}}(s)$ is a vector of dimension n.

- Substituting into the RCS equation, yields

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

where

$$C_{\mathrm{r}} = X_{\mathrm{r}}^{\mathsf{T}} C X_{\mathrm{r}},\ G_{\mathrm{r}} = X_{\mathrm{r}}^{\mathsf{T}} G X_{\mathrm{r}},\ \Gamma_{\mathrm{r}} = E_{\mathrm{r}}^{\mathsf{T}} \Gamma E_{\mathrm{r}},\ E_{\mathrm{r}} = X_{\mathrm{r}}^{\mathsf{T}} E,$$

and

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

- The transfer function of the reduced system is given by

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

- By setting

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

The reduced second-order form corresponds to a reduced order SAPOR system of the original RCS equations:

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

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

- The SAPOR system preserves the block structures and the symmetry of system data matrices of the original RCS system.
- However, the matrix E_{r} in the SAPOR cannot be interpreted as an incidence matrix.
- Towards synthesis based on the reduced-order model, we shall reformulate the projection and the SAPOR system [Yang et al'08]

- Let

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

  Then the original RCS equations can be written as as an *expanded RCS (RCSe) equations*:

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

- Note that the incidence matrix E in the original RCS equations is now the identity matrix I.
- The new current vector $\widehat{i}(s)$ is of the size N_1, typically $N_1 \geq N_2$. The order of RCSe equations is $2N_1$.

In the first-order form, the transfer function $H(s)$ of the RCSe equations:

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

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

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

and

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

Let

$$\mathcal{X} = \mathcal{Y} = \begin{matrix} \\ N_1 \\ N_1 \end{matrix} \overset{\begin{matrix} n & n \end{matrix}}{\begin{bmatrix} X_{\mathrm{r}} & \\ & X_{\mathrm{r}} \end{bmatrix}}.$$

Then by the change-of-variables

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

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

$$\begin{cases} \left(s \begin{bmatrix} C_{\mathrm{r}} & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} G_{\mathrm{r}} & I \\ -\Gamma_{\mathrm{r}} & 0 \end{bmatrix} \right) \begin{bmatrix} v_{\mathrm{r}}(s) \\ i_{\mathrm{r}}(s) \end{bmatrix} = \begin{bmatrix} B_{\mathrm{r},v} \\ 0 \end{bmatrix} u(s), \\ \widetilde{y}(s) = \begin{bmatrix} D_{\mathrm{r},v}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} v_{\mathrm{r}}(s) \\ i_{\mathrm{r}}(s) \end{bmatrix}. \end{cases}$$

Note that the reduced equations 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.

- Assume that the sub-blocks B_v and D_v in the input and output of the RCS equations are of the forms:

$$B_v = \begin{matrix} \\ p_1 \\ N_1-p_1 \end{matrix} \overset{p}{\begin{bmatrix} B_{v1} \\ 0 \end{bmatrix}}, \quad D_v = \begin{matrix} \\ p_1 \\ N_1-p_1 \end{matrix} \overset{m}{\begin{bmatrix} D_{v1} \\ 0 \end{bmatrix}}.$$

- Furthermore, assume that the incidence matrix E has the zero block on the top, conformal with the partition of the input and output matrices:

$$E = \begin{matrix} \\ p_1 \\ N_1-p_1 \end{matrix} \overset{N_2}{\begin{bmatrix} 0 \\ \tilde{E} \end{bmatrix}}.$$

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

- let $X_{\rm r}$ be an orthonormal basis for the projection subspace
  Using partitioning-and-embedding steps, we have

$$X_{\rm r} = \begin{matrix} \\ p_1 \\ N_1-p_1 \end{matrix} \overset{n}{\begin{bmatrix} X_{\rm r}^{(1)} \\ X_{\rm r}^{(2)} \end{bmatrix}} \rightsquigarrow \widehat{X}_{\rm r} = \begin{matrix} \\ p_1 \\ N_1-p_1 \end{matrix} \overset{p_1 \quad n}{\begin{bmatrix} I & \\ & X_2 \end{bmatrix}},$$

  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, rank$(X_{\rm r}^{(2)})$ = rank$(X_2) = n$.

  Using the subspace projection with

$$\mathcal{X} = \mathcal{Y} = \begin{matrix} \\ N_1 \\ N_1 \end{matrix} \overset{p_1+n \quad p_1+n}{\begin{bmatrix} \widehat{X}_{\rm r} & \\ & \widehat{X}_{\rm r} \end{bmatrix}},$$

we have the reduced-order RCSe equations

$$\begin{cases} \left(s \begin{bmatrix} C_{\rm r} & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} G_{\rm r} & I \\ -\Gamma_{\rm r} & 0 \end{bmatrix} \right) \begin{bmatrix} v_{\rm r}(s) \\ i_{\rm r}(s) \end{bmatrix} = \begin{bmatrix} B_{{\rm r},v} \\ 0 \end{bmatrix} u(s), \\ \widetilde{y}(s) = \begin{bmatrix} D_{{\rm r},v}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} v_{\rm r}(s) \\ i_{\rm r}(s) \end{bmatrix}, \end{cases}$$

where $C_{\rm r} = \widehat{X}_{\rm r}^{\mathsf{T}} C \widehat{X}_{\rm r}$, $G_{\rm r} = \widehat{X}_{\rm r}^{\mathsf{T}} G \widehat{X}_{\rm r}$, $\Gamma_{\rm r} = \widehat{X}_{\rm r}^{\mathsf{T}} \Gamma \widehat{X}_{\rm r}$, and $B_{{\rm r},v}$ and $D_{{\rm r},v}$ preserve the original I/O structure:

$$B_{{\rm r},v} = \widehat{X}_{\rm r}^{\mathsf{T}} \begin{bmatrix} B_{v1} \\ 0 \end{bmatrix} = \begin{matrix} p_1 \\ n \end{matrix} \overset{p}{\begin{bmatrix} B_{v1} \\ 0 \end{bmatrix}}, \quad D_{{\rm r},v} = \widehat{X}_{\rm r}^{\mathsf{T}} \begin{bmatrix} D_{v1} \\ 0 \end{bmatrix} = \begin{matrix} p_1 \\ n \end{matrix} \overset{m}{\begin{bmatrix} D_{v1} \\ 0 \end{bmatrix}}.$$

Note that

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

- The reduced RCSe system has the same moment-matching property!

- Again for synthesis, consider the diagonalization of Γ in the RCSe equations.
- The "zero-block" assumption of the incidence matrix E implies that Γ is of the form

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

- Let $\widetilde{\Gamma}_{\mathrm{r}} = Q_2^{\mathsf{T}}\widetilde{\Gamma}Q_2$, then the reduced RCSe equations Γ_{r} has the same form

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

  Note that $\widetilde{\Gamma}$ is symmetric semi-positive definite, so is $\widetilde{\Gamma}_{\mathrm{r}}$.

- Let
$$\widetilde{\Gamma}_{\mathrm{r}} = \widetilde{V}\Lambda\widetilde{V}^{\mathsf{T}}$$
be the eigen-decomposition of $\widetilde{\Gamma}_{\mathrm{r}}$, where V is orthogonal and Λ is diagonal.
- Define
$$V = \begin{array}{c} \\ p_1+n \\ p_1+n \end{array} \begin{array}{c} \begin{array}{cc} p_1+n & p_1+n \end{array} \\ \begin{bmatrix} \widehat{V} & \\ & \widehat{V} \end{bmatrix} \end{array},$$
where
$$\widehat{V} = \begin{array}{c} \\ p_1 \\ n \end{array} \begin{array}{c} \begin{array}{cc} p_1 & n \end{array} \\ \begin{bmatrix} I & \\ & \widetilde{V} \end{bmatrix} \end{array}.$$
- Then by a congruence transformation using the matrix V, the

reduced-order RCSe equations is equivalent to the equations

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

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

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

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

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

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

diagonal

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

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

$$\Lambda = \begin{array}{c} \\ \ell \\ n-\ell \end{array} \begin{array}{c} \begin{array}{cc} \ell & n-\ell \end{array} \\ \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \end{array},$$

where Λ_2 contains the $n - \ell$ smallest eigenvalues that are smaller than a given threshold ϵ in magnitude. and therefore set $\Lambda_2 = 0$.

The "susceptance" matrix is $\widehat{\Gamma}_{\mathrm{r}} = \begin{array}{c} \\ p_1 \\ \ell \\ n-\ell \end{array} \begin{array}{c} \begin{array}{ccc} p_1 & \ell & n-\ell \end{array} \\ \begin{bmatrix} 0 & & \\ & \Lambda_1 & \\ & & 0 \end{bmatrix} \end{array}$.

- In summary, we derived the following *the synthesized RCL equations*:

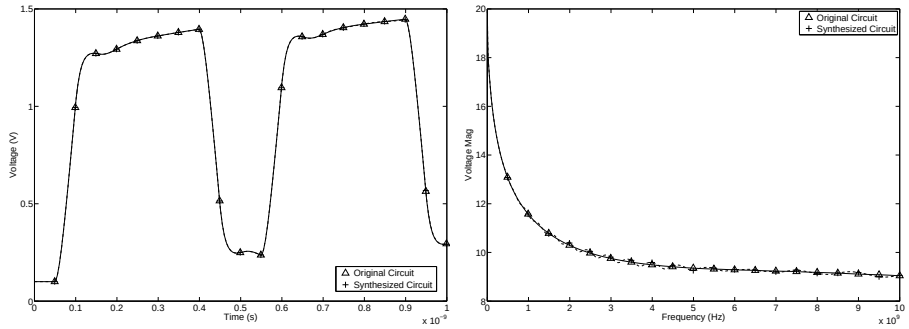
$$\begin{cases} \left(s \begin{bmatrix} \widehat{C}_{\rm r} & 0 \\ 0 & \widehat{L}_{\rm r} \end{bmatrix} + \begin{bmatrix} \widehat{G}_{\rm r} & I \\ -I & 0 \end{bmatrix} \right) \begin{bmatrix} \widehat{v}_{\rm r}(s) \\ \widehat{i}_{\rm r}(s) \end{bmatrix} = \begin{bmatrix} B_{{\rm r},v} \\ 0 \end{bmatrix}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

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

$$\widehat{L}_{\rm r} = \begin{array}{c} \\ p_1 \\ \ell \\ n-\ell \end{array} \begin{array}{c} \begin{array}{ccc} p_1 & \ell & n-\ell \end{array} \\ \begin{bmatrix} 0 & & \\ & \Lambda_1^{-1} & \\ & & 0 \end{bmatrix} \end{array}.$$

- **RCLSYN** (RCL equivalent circuit synthesis) tool [Yang et al'08]

- A 64-bit bus circuit network with 8 inputs and 8 outputs. The order of RCL equation $N = 16963$, the reduced order $n = 640$.
- SPICE transient and AC analysis:



- The CPU elapsed time for the transient and AC analysis are shown in the following table:

	Full RCL	Synthesized RCL	Speedup
Transient analysis	5007.59 (sec.)	90.16 (sec.)	50×
AC analysis	29693.02 (sec.)	739.29 (sec.)	40×

Further reading

The materials presented in this lecture are based on the following papers, and references therein:

- R.-C. Li and Z. Bai, *Structure-preserving model reduction using a Krylov subspace projection formulation*, Comm. Math. Sci. 3:179-199, 2005
- Z. Bai, R-C. Li and Y. Su, *A Unified Krylov Projection Framework for Structure-Preserving Model Reduction*. In "Model Order Reduction: Theory, Research Aspects and Applications", Springer Series of Mathematics in Industry, Vol.13. Schilders, Wilhelmus H.A.; van der Vorst, Henk A.; Rommes, Joost (Eds.) pp.75-93, 2008

Both papers are available on the class website.