Model reduction of large-scale systems

An overview and some new results

Thanos Antoulas

Rice University and Jacobs University

email: aca@rice.edu
URL: www.ece.rice.edu/˜aca

Electrical and Computer Engineering Control Seminar
The University of Michigan, Ann Arbor
29 January 2010

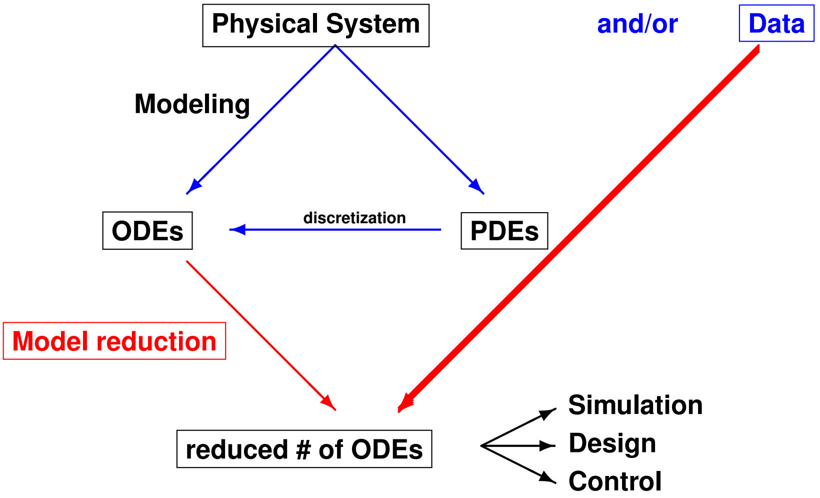
Outline

Collaborators

- Chris Beattie
- Serkan Gugercin
- Sanda Lefteriu
- Cosmin Ionita
- Roxana Ionutiu
- Joost Rommes
- Andrew Mayo
- Dan Sorensen
- Timo Reis
- Matthias Heinkenschloss

Outline

The big picture



Model Reduction via Projection

Given is $\mathbf{f}(\dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}(t)) = \mathbf{0}$, $\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t))$ or
$\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$, $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)$.

Common framework for (most) model reduction methods: **Petrov-Galerkin projective approximation**

Find $\mathbf{x}(t)$ contained in $\mathbb{C}^n$ such that

$$\mathbf{E}\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t) - \mathbf{B}\,\mathbf{u}(t) \;\perp\; \mathbb{C}^n \text{ (i.e., } = 0) \Rightarrow \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t).$$

Choose r-dimensional trial and test subspaces, $\mathcal{V}_r$, $\mathcal{W}_r \subset \mathbb{C}^n$:

Find $\mathbf{v}(t)$ contained in $\mathcal{V}_r$ such that

$$\mathbf{E}\dot{\mathbf{v}}(t) - \mathbf{A}\mathbf{v}(t) - \mathbf{B}\,\mathbf{u}(t) \;\perp\; \mathcal{W}_r \Rightarrow \mathbf{y}_r(t) = \mathbf{C}\mathbf{v}(t) + \mathbf{D}\mathbf{u}(t).$$

Model Reduction via Projection

Let $\mathcal{V}_r = \text{Range}(\mathbf{V}_r)$ and $\mathcal{W}_r = \text{Range}(\overline{\mathbf{W}}_r)$. Then the reduced system trajectories are $\mathbf{v}(t) = \mathbf{V}_r \mathbf{x}_r(t)$ with $\mathbf{x}_r(t) \in \mathbb{C}^r$ for each t, and the Petrov-Galerkin approximation can be rewritten as

$$\mathbf{W}_r^* \left(\mathbf{E}\mathbf{V}_r \dot{\mathbf{x}}_r(t) - \mathbf{A}\mathbf{V}_r \mathbf{x}_r(t) - \mathbf{B}\,\mathbf{u}(t)\right) = \mathbf{0} \quad \text{and} \quad \mathbf{y}_r(t) = \mathbf{C}\mathbf{V}_r \mathbf{x}_r(t) + \mathbf{D}\mathbf{u}(t),$$

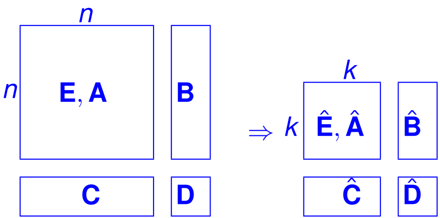
leading to the reduced order state-space representation with

Reduced order system

$$\mathbf{E}_r = \mathbf{W}_r^* \mathbf{E} \mathbf{V}_r, \;\; \mathbf{A}_r = \mathbf{W}_r^* \mathbf{A} \mathbf{V}_r, \;\; \mathbf{B}_r = \mathbf{W}_r^* \mathbf{B}, \;\; \mathbf{C}_r = \mathbf{C}\mathbf{V}_r \;\; \text{and} \;\; \mathbf{D}_r = \mathbf{D}.$$

The quality of the reduced system depends on the choice of $\mathcal{V}_r$ and $\mathcal{W}_r$.

$$\left(\begin{array}{c|c} \mathbf{E}, \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right) \Rightarrow \left(\begin{array}{c|c} \hat{\mathbf{E}}, \hat{\mathbf{A}} & \hat{\mathbf{B}} \\ \hline \hat{\mathbf{C}} & \hat{\mathbf{D}} \end{array}\right) = \left(\begin{array}{c|c} \mathbf{W}^*\mathbf{E}\mathbf{V}, \mathbf{W}^*\mathbf{A}\mathbf{V} & \mathbf{W}^*\mathbf{B} \\ \hline \mathbf{C}\mathbf{V} & \mathbf{D} \end{array}\right), \; k \ll n$$



Norms:

- $\mathcal{H}_\infty$-norm: worst output error $\|\mathbf{y}(t) - \hat{\mathbf{y}}(t)\|$ for $\|\mathbf{u}(t)\| = 1$.
- $\mathcal{H}_2$-norm: $\|\mathbf{h}(t) - \hat{\mathbf{h}}(t)\|$

Nonlinear systems

Consider system described by implicit equations:

$$\mathbf{f}\left(\frac{d}{dt}\mathbf{x}(t), \mathbf{x}(t), \mathbf{u}(t)\right) = \mathbf{0},\ \ \mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t)),$$

with: $\mathbf{u}(t) \in \mathbb{R}^m$, $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{y}(t) \in \mathbb{R}^p$.

Problem: Approximate this system with:

$$(\hat{\mathbf{f}}, \hat{\mathbf{h}}) :\ \ \mathbf{u}(t) \in \mathbb{R}^m,\ \ \hat{\mathbf{x}}(t) \in \mathbb{R}^k,\ \ \hat{\mathbf{y}}(t) \in \mathbb{R}^p,\ \ k \ll n$$

by means of a Petrov-Galerkin projection: $\Rightarrow \Pi = \mathbf{V}\mathbf{W}^*$, $\mathbf{V},\ \mathbf{W} \in \mathbb{R}^{n\times k}$,

$$\mathbf{W}^*\mathbf{f}\left(\frac{d}{dt}\mathbf{V}\hat{\mathbf{x}}(t), \mathbf{V}\hat{\mathbf{x}}(t),\ \mathbf{u}(t)\right) = \mathbf{0},\ \ \hat{\mathbf{y}}(t) = \mathbf{h}(\mathbf{V}\hat{\mathbf{x}}(t),\ \mathbf{u}(t))$$

The approximation is "good" if $\mathbf{x} - \Pi\mathbf{x}$ is "small".

Issues and requirements

Issues with large-scale systems

1. Storage
2. Computational speed
3. Accuracy
4. System theoretic properties

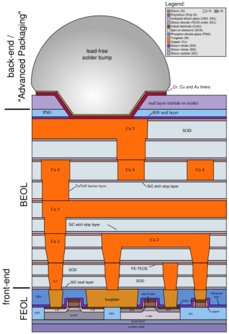
Requirements for model reduction

1. Approximation error small - global error bound
2. Preservation of stability/passivity
3. Procedure must be computationally efficient
4. Procedure must be automatic
5. In addition: many ports, parameters, nonlinearities, ...

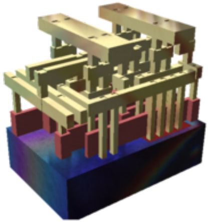
Motivating Examples: Simulation/Control

1. **Passive devices**	• VLSI circuits • Thermal issues • Power delivery networks
2. **Data assimilation**	• North sea forecast • Air quality forecast
3. **Molecular systems**	• MD simulations • Heat capacity
4. **CVD reactor**	• Bifurcations
5. **Mechanical systems**:	•Windscreen vibrations • Buildings
6. **Optimal cooling**	• Steel profile
7. **MEMS**: **Micro Electro- -Mechanical Systems**	• Elf sensor
8. **Nano-Electronics**	• Plasmonics

VLSI circuits



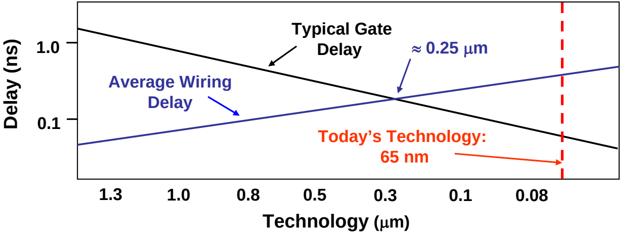
CMOS Chip



3D Silicon Chip

nanometer details	10^8 components
several GHz speed	several km interconnect
$\approx$ 10 layers	

VLSI circuits



65nm technology: **gate delay $<$ interconnect delay!**

Conclusion: Simulations are required to verify that internal electromagnetic fields do not significantly delay or distort circuit signals. Therefore interconnections must be modeled.

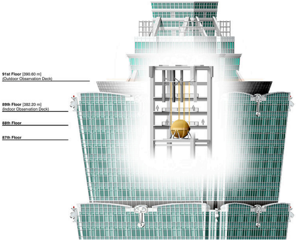
$\Rightarrow$ Electromagnetic modeling of packages and interconnects $\Rightarrow$ resulting models very complex: using PEEC methods (discretization of Maxwell's equations): $n \approx 10^5 \cdots 10^6 \Rightarrow$ SPICE: inadequate

Mechanical Systems: Buildings

Earthquake prevention

Taipei 101: 508m

Damper between 87-91 floors



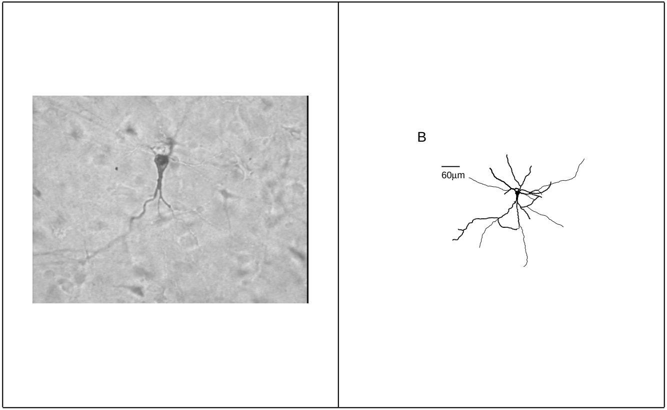
730 ton damper

Building	Height	Control mechanism	Damping frequency Damping mass
CN Tower, Toronto	533 m	Passive tuned mass damper	
Hancock building, Boston	244 m	Two passive tuned dampers	0.14Hz, 2x300t
Sydney tower	305 m	Passive tuned pendulum	0.1,0.5z, 220t
Rokko Island P&G, Kobe	117 m	Passive tuned pendulum	0.33-0.62Hz, 270t
Yokohama Landmark tower	296 m	Active tuned mass dampers (2)	0.185Hz, 340t
Shinjuku Park Tower	296 m	Active tuned mass dampers (3)	330t
TYG Building, Atsugi	159 m	Tuned liquid dampers (720)	0.53Hz, 18.2t

Neural models

How a cell distinguishes between inputs

Image from Neuromart (Rice-Baylor archive of neural morphology)



Goal: • simulation of systems containing a few million neurons

• simulations currently limited to systems with $\approx$10K neurons

Outline

Approximation methods: Overview

Krylov

- Realization
- Interpolation
- Lanczos
- Arnoldi

SVD

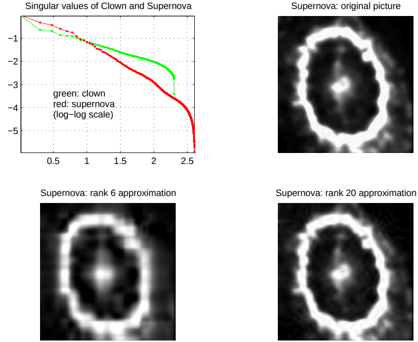
Nonlinear systems	Linear systems
• POD methods • Empirical Gramians	• Balanced truncation • Hankel approximation

Krylov/SVD Methods

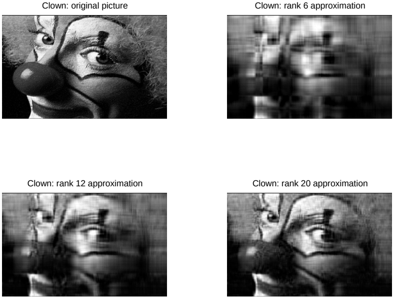
SVD Approximation methods

A prototype approximation problem – the **SVD**
(Singular Value Decomposition): $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$.

Supernova



Clown



Singular values provide trade-off between accuracy and complexity

POD: Proper Orthogonal Decomposition

Consider: $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$, $\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t))$.

Snapshots of the state:

$$\mathcal{X} = [\mathbf{x}(t_1) \;\; \mathbf{x}(t_2) \;\; \cdots \;\; \mathbf{x}(t_N)] \in \mathbb{R}^{n\times N}$$

SVD: $\mathcal{X} = \mathbf{U}\Sigma\mathbf{V}^* \approx \mathbf{U}_k\Sigma_k\mathbf{V}_k^*$, $k \ll n$. Approximate the state:

$$\hat{\mathbf{x}}(t) = \mathbf{U}_k^*\mathbf{x}(t) \;\;\Rightarrow\;\; \mathbf{x}(t) \approx \mathbf{U}_k\hat{\mathbf{x}}(t), \;\; \hat{\mathbf{x}}(t) \in \mathbb{R}^k$$

Project state and output equations. Reduced order system:

$$\boxed{\dot{\hat{\mathbf{x}}}(t) = \mathbf{U}_k^*\mathbf{f}(\mathbf{U}_k\hat{\mathbf{x}}(t), \mathbf{u}(t)), \;\; \mathbf{y}(t) = \mathbf{h}(\mathbf{U}_k\hat{\mathbf{x}}(t), \mathbf{u}(t))}$$

$\Rightarrow$ $\hat{\mathbf{x}}(t)$ evolves in a **low-dimensional** space.

Issues with POD:

(a) Choice of snapshots, (b) singular values not I/O invariants.

SVD methods: the Hankel singular values

Trade-off between accuracy and complexity for linear dynamical systems is provided by the **Hankel Singular Values**, which are computed (for stable systems) as follows:

Define the **gramians**

$$\mathbf{P} = \int_0^\infty e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^* e^{\mathbf{A}^* t}\, dt, \quad \mathbf{Q} = \int_0^\infty e^{\mathbf{A}^* t}\mathbf{C}^*\mathbf{C} e^{\mathbf{A}t}\, dt$$

To compute the gramians we need to solve **2 Lyapunov equations**:

$$\left.\begin{array}{l} \mathbf{AP} + \mathbf{PA}^* + \mathbf{BB}^* = \mathbf{0}, \ \ \mathbf{P} > \mathbf{0} \\ \mathbf{A}^*\mathbf{Q} + \mathbf{QA} + \mathbf{C}^*\mathbf{C} = \mathbf{0}, \ \ \mathbf{Q} > \mathbf{0} \end{array}\right\} \Rightarrow \boxed{\sigma_i = \sqrt{\lambda_i(\mathbf{PQ})}}$$

σ_i: **Hankel singular values** of system Σ.

SVD methods: approximation by balanced truncation

There exists basis where $\mathbf{P} = \mathbf{Q} = \mathbf{S} = \mathrm{diag}\,(\sigma_1, \cdots, \sigma_n)$. This is the balanced basis of the system.

In this basis partition:

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right),\ \mathbf{B} = \left(\begin{array}{c} \mathbf{B}_1 \\ \hline \mathbf{B}_2 \end{array}\right),\ \mathbf{C} = (\mathbf{C}_1 \mid \mathbf{C}_2),\ \mathbf{S} = \left(\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & \Sigma_2 \end{array}\right).$$

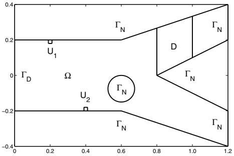
The reduced system is obtained by balanced truncation

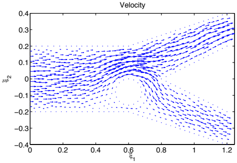
$$\hat{\Sigma} = \left(\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{B}_1 \\ \hline \mathbf{C}_1 & \end{array}\right).$$

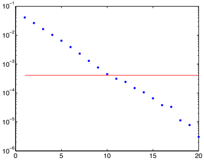
Σ_2 contains the small Hankel singular values.

Projector: $\Pi = \mathbf{V}\mathbf{W}^* $ where $\mathbf{V} = \mathbf{W} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{0} \end{pmatrix}$.

Example







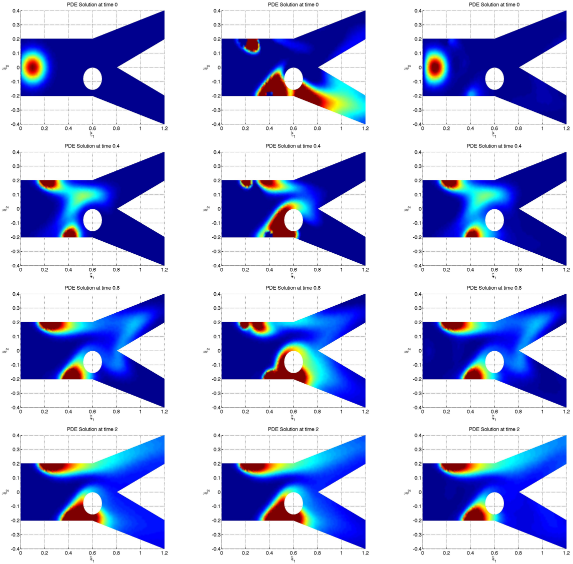
We consider model reduction of a semidiscretized advection diffusion equation. The problem is motivated by the optimal control of pollutants. The advection diffuson equation for the concentration c of the pollutant. is

$$
\begin{aligned}
&\frac{\partial}{\partial t}c(\xi,t) - \nabla(\kappa\nabla c(\xi,t)) + \mathbf{v}(\xi)\cdot\nabla c(\xi,t) \\
&= u(\xi,t)\chi_{U_1}(\xi) + u(\xi,t)\chi_{U_2}(\xi) \qquad \text{in } \Omega,
\end{aligned}
$$

with boundary and initial conditions $c(\xi,t)=0$ in Γ_D, and $\frac{\partial}{\partial n}c(\xi,t)=0$, in Γ_N, $c(\xi,0)=2\exp(-100((\xi_1-0.1)^2+\xi_2^2))$, in Ω. The domain Ω and the boundary segments Γ_D, Γ_N are depicted below. χ_S is the characteristic function corresponding to the subdomain $S\subset\Omega$. The inputs are u_i defined on $U_i\times(0,T)$, $i=1,2$, where U_1, U_2 are he subdomains shown below. In our numerical experiment we set $u_1=u_2=50$. The diffusivity is $\kappa=0.005$ and the advection $\mathbf{v}$ is the solution of a steady state Stokes equation on Ω wit inflow condition at $\xi_1=0$ and “do nothing” outflow condition at $\xi_2=1.2$.

Original	Reduced
$m=16$	$m=16$
$n=2673$	$k=10$
$p=283$	$p=283$

Example



$t = 0.0$

$t = 0.4$

$t = 0.8$

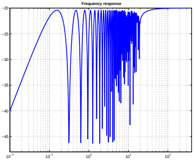
$t = 2.0$

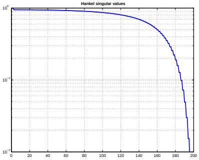
Properties of balanced reduction

1. Stability is preserved
2. Global error bound: $\sigma_{k+1} \leq \| \Sigma - \hat{\Sigma} \|_{\infty} \leq 2(\sigma_{k+1} + \cdots + \sigma_n)$

Drawbacks

1. Dense computations, matrix factorizations and inversions $\Rightarrow$ may be ill-conditioned; number of operations $\mathcal{O}(n^3)$
2. Bottleneck: solution of two Lyapunov equations
3. Slow decay of HSV: a transmission line





Frequency response and Hankel singular values

Approximation methods: Krylov methods

Krylov

- Realization
- Interpolation
- Lanczos
- Arnoldi

SVD

Nonlinear systems	Linear systems
• POD methods	• Balanced truncation
• Empirical Gramians	• Hankel approximation

Krylov/SVD Methods

The basic Krylov iteration

Given $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$, let $\mathbf{v}_1 = \frac{\mathbf{b}}{\|\mathbf{b}\|}$. At the k^{th} step:

$$\boxed{\mathbf{A}\mathbf{V}_k = \mathbf{V}_k\mathbf{H}_k + \mathbf{f}_k\mathbf{e}_k^*} \qquad \text{where}$$

$\mathbf{e}_k \in \mathbb{R}^k$: canonical unit vector
$\mathbf{V}_k = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_k] \in \mathbb{R}^{k\times k}$, $\mathbf{V}_v^*\mathbf{V}_k = \mathbf{I}_k$
$\mathbf{H}_k = \mathbf{V}_k^*\mathbf{A}\mathbf{V}_k \in \mathbb{R}^{k\times k}$

Three uses of the Krylov iteration

(1) Iterative solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$: approximate the solution $\mathbf{x}$ iteratively.

(2) Iterative approximation of the eigenvalues of $\mathbf{A}$. In this case $\mathbf{b}$ is not fixed apriori. The eigenvalues of the projected $\mathbf{H}_k$ approximate the dominant eigenvalues of $\mathbf{A}$.

(3) Approximation of linear systems by moment matriching.

$\Rightarrow$ **Item (3) is of interest in the present context.**

Krylov methods: Approximation by moment matching

Given $\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t),\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)$, expand transfer function around s_0:

$$\mathbf{H}(s) = \eta_0 + \eta_1(s - s_0) + \eta_2(s - s_0)^2 + \eta_3(s - s_0)^3 + \cdots$$

Moments at s_0: η_j.

Find $\hat{\mathbf{E}}\dot{\hat{\mathbf{x}}}(t) = \hat{\mathbf{A}}\hat{\mathbf{x}}(t) + \hat{\mathbf{B}}\mathbf{u}(t),\ \mathbf{y}(t) = \hat{\mathbf{C}}\hat{\mathbf{x}}(t) + \hat{\mathbf{D}}\mathbf{u}(t)$, with

$$\hat{\mathbf{H}}(s) = \hat{\eta}_0 + \hat{\eta}_1(s - s_0) + \hat{\eta}_2(s - s_0)^2 + \hat{\eta}_3(s - s_0)^3 + \cdots$$

such that for appropriate s_0 and ℓ:

$$\boxed{\eta_j = \hat{\eta}_j,\ \ j = 1, 2, \cdots, \ell}$$

The general interpolation framework

- Expansion around infinity: η_j are the Markov parameters $\Rightarrow$ **partial realization**.

  Expansion around arbitrary $s_0 \in \mathbb{C}$: η_j moments $\Rightarrow$ **rational interpolation**.
- **Goal**: produce $\mathbf{H}_r(s)$, that approximates with high fidelity a very large order transfer function, $\mathbf{H}(s)$: $\mathbf{H}_r(s) \approx \mathbf{H}(s)$, by means of interpolation at a set of points $\{\sigma_i\}_{i=1}^r \subset \mathbb{C}$: $\mathbf{H}_r(\sigma_i) = \mathbf{H}(\sigma_i)$ for $i = 1, \ldots, r$.
- This is a good starting place for SISO systems but is overly restrictive for MIMO systems, since the condition $\mathbf{H}_r(\sigma_i) = \mathbf{H}(\sigma_i)$ imposes $m \cdot p$ scalar conditions at each point. Instead consider interpolation conditions that are imposed in specified directions: **tangential interpolation**.
- **Remark**. The Krylov and rational Krylov algorithms match moments **without** computing them. Thus **moment matching** methods can be implemented in a numerically efficient way.

Problem 1: Model Reduction given state space system data

Given a full-order model **E**, **A**, **B**, **C**, and **D** and given

left interpolation points:		*right interpolation points*:
$\{\mu_i\}_{i=1}^{q} \subset \mathbb{C}$,		$\{\sigma_i\}_{i=1}^{r} \subset \mathbb{C}$
with corresponding	and	with corresponding
left tangent directions:		*right tangent directions*:
$\{\ell_i\}_{i=1}^{q} \subset \mathbb{C}^p$,		$\{\mathbf{r}_i\}_{i=1}^{r} \subset \mathbb{C}^m$.

Find a reduced-order model $\mathbf{E}_r$, $\mathbf{A}_r$, $\mathbf{B}_r$, $\mathbf{C}_r$, and $\mathbf{D}_r$ such that the associated transfer function, $\mathbf{H}_r(s)$, is a *tangential interpolant* to $\mathbf{H}(s)$:

$$\ell_i^* \mathbf{H}_r(\mu_i) = \ell_i^* \mathbf{H}(\mu_i) \quad \text{and} \quad \mathbf{H}_r(\sigma_j)\mathbf{r}_j = \mathbf{H}(\sigma_j)\mathbf{r}_j,$$
$$\text{for } i = 1, \cdots, q, \qquad\qquad \text{for } j = 1, \cdots, r,$$

Problem 2: Model reduction given input-output data

Given a set of input-output response measurements on a system specified by

left driving frequencies: $\{\mu_i\}_{i=1}^{q} \subset \mathbb{C}$, using *left input directions*: $\{\ell_i\}_{i=1}^{q} \subset \mathbb{C}^p$, producing *left responses*: $\{\mathbf{v}_i\}_{i=1}^{q} \subset \mathbb{C}^m$,

and

right driving frequencies: $\{\sigma_i\}_{i=1}^{r} \subset \mathbb{C}$ using *right input directions*: $\{\mathbf{r}_i\}_{i=1}^{r} \subset \mathbb{C}^m$ producing *right responses*: $\{\mathbf{w}_i\}_{i=1}^{r} \subset \mathbb{C}^p$

Find a system model by specifying (reduced) system matrices $\mathbf{E}_r$, $\mathbf{A}_r$, $\mathbf{B}_r$, $\mathbf{C}_r$, $\mathbf{D}_r$ such that $\mathbf{H}_r(s)$, is a *tangential interpolant* to the given data:

$$\ell_i^* \mathbf{H}_r(\mu_i) = \mathbf{v}_i^* \quad \text{for } i = 1, \cdots, q, \qquad \text{and} \qquad \mathbf{H}_r(\sigma_j)\mathbf{r}_j = \mathbf{w}_j, \quad \text{for } j = 1, \cdots, r.$$

Interpolation points and tangent directions are determined (typically) by the availability of experimental data.

Interpolatory Projections

We seek $\mathbf{H}_r(s)$ so that

$$\left.\begin{array}{ll} \mathbf{H}(\sigma_i)\mathbf{r}_i = \mathbf{H}_r(\sigma_i)\mathbf{r}_i, & \text{for } i = 1, \cdots, r, \\ \ell_j^* \mathbf{H}(\mu_j) = \ell_j^* \mathbf{H}_r(\mu_j), & \text{for } j = 1, \cdots, r, \end{array}\right\}$$

The goal is to interpolate $\mathbf{H}(s)$ without ever computing the quantities to be matched since these are numerically ill-conditioned. This is achieved by the projection framework.

Interpolatory projections

$$\mathbf{V}_r = \left[(\sigma_1 \mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{r}_1, \; \cdots, \; (\sigma_r \mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{r}_r\right],$$

$$\mathbf{W}_r^* = \begin{bmatrix} \ell_1^* \mathbf{C}(\mu_1 \mathbf{E} - \mathbf{A})^{-1} \\ \vdots \\ \ell_r^* \mathbf{C}(\mu_r \mathbf{E} - \mathbf{A})^{-1} \end{bmatrix}.$$

Properties of Krylov (interpolatory) methods

- Number of operations: $\mathcal{O}(kn^2)$ or $\mathcal{O}(k^2 n)$ vs. $\mathcal{O}(n^3)$ $\Rightarrow$ efficiency

Q: **How to choose the projection points?**

We will discuss two cases:

1. **Passivity preserving model reduction.**
2. **Optimal $\mathcal{H}_2$ model reduction.**

Outline

Choice of Krylov projection points: Passivity preserving model reduction

Passive systems:
$\mathcal{R}e \int_{-\infty}^{t} \mathbf{u}(\tau)^* \mathbf{y}(\tau) \mathrm{d}\tau \geq 0, \forall\, t \in \mathbb{R}, \forall\, \mathbf{u} \in \mathcal{L}_2(\mathbb{R})$.

Positive real rational functions:
(1) $\mathbf{H}(s) = \mathbf{D} + \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}$, is analytic for $\mathcal{R}e(s) > 0$,
(2) $\mathcal{R}e\, \mathbf{H}(s) \geq 0$ for $\mathcal{R}e(s) \geq 0$, s not a pole of $H(s)$.

Theorem: $\Sigma = \left(\begin{array}{c|c} \mathbf{E}, \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right)$ is passive $\Leftrightarrow$ $\mathbf{H}(s)$ is positive real.

Conclusion: Positive realness of $\mathbf{H}(s)$ implies the existence of a **spectral factorization** $\mathbf{H}(s) + \mathbf{H}^*(-s) = \Phi(s)\Phi^*(-s)$, where the poles and (finite) zeros of $\Phi(s)$ are stable. The *spectral zeros* are λ such that: $\Phi(\lambda)$, loses rank. Hence there is a right spectral zero direction, $\mathbf{r}$, such that
$(\mathbf{H}(\lambda) + \mathbf{H}^*(-\lambda))\mathbf{r} = \mathbf{0} \Rightarrow \mathbf{r}^*(\mathbf{H}(-\overline{\lambda}) + \mathbf{H}^*(\overline{\lambda})) = \mathbf{0}$.

Passivity preserving model reduction: New result

- **Method**: Rational Krylov
- **Solution**: projection points = spectral zeros

Passivity preserving tangential interpolation

Suppose the dynamical system represented by $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$, is stable and passive. Suppose that for some $r \geq 1$, $\lambda_1, \cdots, \lambda_r$ are stable spectral zeros of $\mathbf{H}$ with corresponding right spectral zero directions $\mathbf{r}_1, \ldots, \mathbf{r}_r$.

If a reduced order system $\mathbf{H}_r(s)$ tangentially interpolates $\mathbf{H}(s)$ with $\sigma_i = \lambda_i$, along $\mathbf{r}_i$, $\mu_i = -\overline{\lambda_i}$, along $\mathbf{r}^*$ for $i = 1, \ldots, r$, then $\mathbf{H}_r(s)$ is stable and passive.

The computation of the spectral zeros of the system can be formulated as a structured eigenvalue problem. Let

$$\mathcal{H} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & -\mathbf{A}^* & -\mathbf{C}^* \\ \mathbf{C} & \mathbf{B}^* & \mathbf{D}+\mathbf{D}^* \end{bmatrix}, \ \mathcal{E} = \begin{bmatrix} \mathbf{E} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Spectral zero interpolation preserving passivity

Hamiltonian EVD & projection

- **Hamiltonian eigenvalue problem**

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & -\mathbf{A}^* & -\mathbf{C}^* \\ \mathbf{C} & \mathbf{B}^* & \mathbf{D}+\mathbf{D}^* \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \mathbf{E} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} \Lambda$$

The generalized eigenvalues Λ are the spectral zeros of Σ

- **Partition eigenvectors**

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_- & \mathbf{X}_+ \\ \mathbf{Y}_- & \mathbf{Y}_+ \\ \mathbf{Z}_- & \mathbf{Z}_+ \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_- & & \\ & \Lambda_+ & \\ & & \pm\infty \end{bmatrix}$$

Λ_- are the stable spectral zeros

- **Projection**
  - $\mathbf{V} = \mathbf{X}_-, \ \mathbf{W} = \mathbf{Y}_-$
  - $\hat{\mathbf{E}} = \mathbf{W}^*\mathbf{E}\mathbf{V}, \ \hat{\mathbf{A}} = \mathbf{W}^*\mathbf{A}\mathbf{V}, \ \hat{\mathbf{B}} = \mathbf{W}^*\mathbf{B}, \ \hat{\mathbf{C}} = \mathbf{C}\mathbf{V}, \ \hat{\mathbf{D}} = \mathbf{D}$

Dominant spectral zeros – SADPA

What is a good choice of k spectral zeros out of n ?

- **Dominance criterion**: Spectral zero s_j is **dominant** if: $\frac{|R_j|}{|\Re(s_j)|}$, is large.
- Efficient computation for large scale systems: we compute the $k \ll n$ **most dominant** eigenmodes of the Hamiltonian pencil.
- **SADPA** (Subspace Accelerated Dominant Pole Algorithm) solves this **iteratively**.

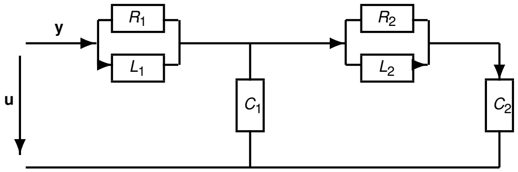
Conclusion:

Passivity preserving model reduction becomes a

structured eigenvalue problem

An example

Consider the following RLC circuit:



Using the voltages across C_1, C_2, and the currents through L_1, L_2, as state variables, $\mathbf{x}_i$, $= 1, 2, 3, 4$, respectively, we end up with equations of the form $\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$, $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)$, where

$$\mathbf{E} = \begin{pmatrix} C_1 & 0 & -G_1 L_1 & G_2 L_2 \\ 0 & C_2 & 0 & -G_2 L_2 \\ 0 & 0 & L_1 & 0 \\ 0 & 0 & 0 & L_2 \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \ \mathbf{C} = [-G_1, \ 0, \ 1, \ 0], \ \mathbf{D} = G_1,$$

$G_i = \frac{1}{R_i}$, $i = 1, 2$, are the corresponding conductances.

This system has a double spectral zero at $s = 0$. The Hamiltonian pair is:

$$\mathcal{H} = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & G_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -G_1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 2G_1 \end{pmatrix},$$

and

$$\mathcal{E} = \begin{pmatrix} C_1 & 0 & -G_1 L_1 & G_2 L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_2 & 0 & -G_2 L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & L_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & C_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -G_1 L_1 & 0 & L_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & G_2 L_2 & -G_2 L_2 & 0 & L_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It follows that the desired (Jordan chain of) eigenvectors $\mathbf{x}_1$, $\mathbf{x}_2$, corresponding to $\mathbf{0}$ are:

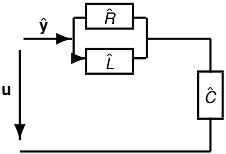
$$[\mathbf{x}_1, \ \ \mathbf{x}_2] = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \\ \hline -1 & 0 \\ -1 & 0 \\ -G_1 & -1 \\ 0 & \frac{-C_2}{C_1+C_2} \\ \hline 1 & 0 \end{array}\right) \Rightarrow \mathbf{V}_r = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \end{array}\right), \ \mathbf{W}_r = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ G_1 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \end{array}\right).$$

The reduced quantities are:

$$\mathbf{E}_r = \mathbf{W}_r^* \mathbf{E} \mathbf{V}_r = \begin{pmatrix} C_1 + C_2 & 0 \\ 0 & L_1 + \frac{{C_2}^2}{(C_1+C_2)^2} L_2 \end{pmatrix}, \mathbf{A}_r = \mathbf{W}_r^* \mathbf{A} \mathbf{V}_r = \begin{pmatrix} -G_1 & 1 \\ -1 & 0 \end{pmatrix},$$

$$\mathbf{B}_r = \mathbf{W}_r^* \mathbf{B} = \begin{pmatrix} G_1 \\ 1 \end{pmatrix}, \ \mathbf{C}_r = \mathbf{C}\mathbf{V}_r = \begin{pmatrix} -G_1 & 1 \end{pmatrix}, \ \mathbf{D}_r = \mathbf{D}.$$

From these matrices we can read-off an RLC realization:



where $\hat{C} = C_1 + C_2$, $\hat{L} = \frac{1}{\kappa} = L_1 + \frac{{C_2}^2}{(C_1+C_2)^2} L_2$, and $\hat{R} = R_1$.

Choice of Krylov projection points:
Optimal H_2 model reduction

Recall: the H_2 norm of a stable (SISO) system is:

$$\|\Sigma\|_{H_2} = \left(\int_{-\infty}^{+\infty} \mathbf{h}^2(t)dt \right)^{1/2}$$

where $\mathbf{h}(t) = \mathbf{C}e^{\mathbf{A}t}\mathbf{B}$, $t \geq 0$, is the impulse response of Σ.

Goal: construct a *Krylov projector* such that

$$\Sigma_k = \arg \min_{\substack{\deg(\hat{\Sigma}) = r \\ \hat{\Sigma}\ :\ \text{stable}}} \left\| \Sigma - \hat{\Sigma} \right\|_{H_2} = \left(\int_{-\infty}^{+\infty} (\mathbf{h} - \hat{\mathbf{h}})^2(t)dt \right)^{1/2}$$

The optimization problem is **nonconvex**. The common approach is to find reduced order models that satisfy first-order necessary optimality conditions.

First-order necessary optimality conditions

Let $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}})$ solve the optimal H_2 problem and let $\hat{\lambda}_i$ denote the eigenvalues of $\hat{\mathbf{A}}$ (we assume for simplicity that $\mathbf{E} = \mathbf{I}$ and $m = p = 1$). The necessary conditions are:

$$\mathbf{H}(-\hat{\lambda}_i^*) = \hat{\mathbf{H}}(-\hat{\lambda}_i^*) \ \text{ and } \ \frac{d}{ds}\mathbf{H}(s)\Big|_{s=-\hat{\lambda}_i^*} = \frac{d}{ds}\hat{\mathbf{H}}(s)\Big|_{s=-\hat{\lambda}_i^*}$$

Thus the reduced system has to match the first two moments of the original system at the *mirror images* of the eigenvalues of $\hat{\mathbf{A}}$.

1. Make an initial selection of σ_i, for $i = 1, \ldots, k$
2. $\mathbf{W} = [(\sigma_1\mathbf{I} - \mathbf{A}^*)^{-1}\mathbf{C}^*, \ \cdots, \ (\sigma_k\mathbf{I} - \mathbf{A}^*)^{-1}\mathbf{C}^*]$
3. $\mathbf{V} = [(\sigma_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}, \ \cdots, \ (\sigma_k\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}]$
4. while (not converged)
   - $\hat{\mathbf{A}} = (\mathbf{W}^*\mathbf{V})^{-1}\mathbf{W}^*\mathbf{A}\mathbf{V}$,
   - $\sigma_i \longleftarrow -\lambda_i(\hat{\mathbf{A}}) + \text{Newton correction}, \ i = 1, \ldots, k$
   - $\mathbf{W} = [(\sigma_1\mathbf{I} - \mathbf{A}^*)^{-1}\mathbf{C}^*, \ \cdots, \ (\sigma_k\mathbf{I} - \mathbf{A}^*)^{-1}\mathbf{C}^*]$
   - $\mathbf{V} = [(\sigma_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}, \ \cdots, \ (\sigma_k\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}]$
5. $\hat{\mathbf{A}} = (\mathbf{W}^*\mathbf{V})^{-1}\mathbf{W}^*\mathbf{A}\mathbf{V}$, $\hat{\mathbf{B}} = (\mathbf{W}^*\mathbf{V})^{-1}\mathbf{W}^*\mathbf{B}$, $\hat{\mathbf{C}} = \mathbf{C}\mathbf{V}$

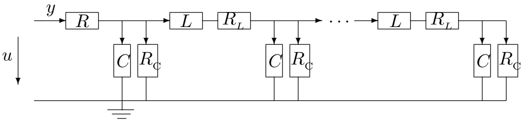
A numerical algorithm for optimal H_2 model reduction

• Global minimizers are difficult to obtain with certainty; current approaches favor seeking reduced order models that satisfy a local (first-order) necessary condition for optimality. Even though such strategies do not guarantee global minimizers, they often produce effective reduced order models.

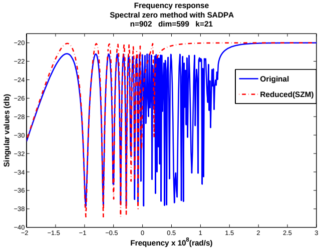
• The main computational cost of this algorithm involves solving $2r$ linear systems to generate $\mathbf{V}_r$ and $\mathbf{W}_r$. Computing the eigenvectors $\mathbf{Y}$ and $\mathbf{X}$, and the eigenvalues of the reduced pencil $\lambda\mathbf{E}_r - \mathbf{A}_r$ are cheap since the dimension r is small.

• **IRKA** has been successfully applied to finding H_2-optimal reduced models for systems of order $n > 160,000$.

Moderate-dimensional example



- total system variables $\mathrm{n} = 902$, independent variables $\mathrm{dim} = 599$, reduced dimension $\mathrm{k} = 21$
- reduced model captures dominant modes



$\mathcal{H}_\infty$ and $\mathcal{H}_2$ error norms

Relative norms of the error systems

Reduction Method $n = 902, dim = 599, k = 21$	$\mathcal{H}_\infty$	$\mathcal{H}_2$
PRIMA	1.477	-
Spectral Zero Method with SADPA	0.962	0.841
Optimal $\mathcal{H}_2$	**0.594**	**0.462**
Balanced truncation (BT)	0.939	0.646
Riccati Balanced Truncation (PRBT)	0.961	0.816

Choice of points: summary

•	Passive model reduction	mirror image of **spectral zeros**
•	Optimal H_2 model reduction	mirror image of **reduced system poles**
•	Hankel norm model reduction	mirror image of **system poles**
•	Systems with proportional damping	mirror image of **'focal' point**

Question: What is the deeper meaning of mirror image points?

Approximation methods: Summary

Krylov

- Realization
- Interpolation
- Lanczos
- Arnoldi

SVD

Nonlinear systems	**Linear systems**
• POD methods • Empirical Gramians	• Balanced truncation • Hankel approximation

Properties

- **numerical efficiency**
- **$n \gg 10^3$**

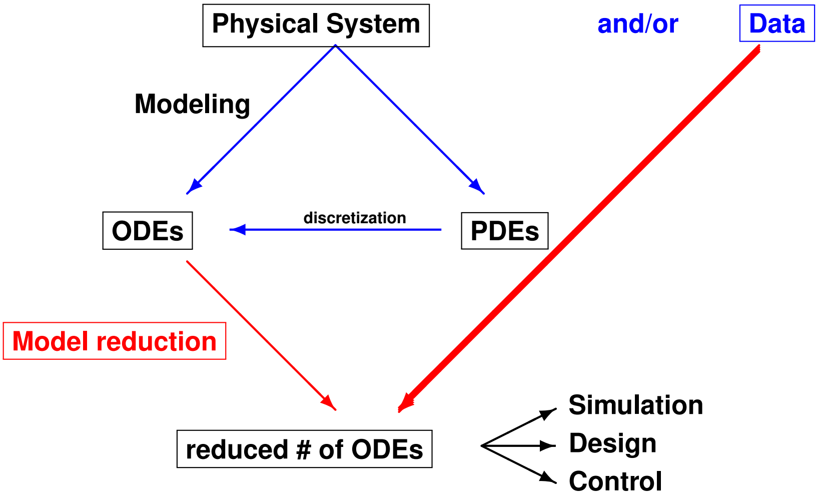
Krylov/SVD Methods

Properties

- **Stability**
- **Error bound**
- **$n \approx 10^3$**

Outline

Recall: the big picture



A motivation: electronic systems

- Growth in communications and networking and demand for high data bandwidth requires streamlining of the simulation of entire complex systems from chips to packages to boards, etc.
- Thus in circuit simulation, signal intergrity (lack of signal distortion) of high speed electronic designs require that interconnect models be valid over a wide bandwidth.

An important tool: **S-parameters**

- They represent a component as a black box. Accurate simulations require accurate component models.
- In high frequencies S-parameters are important because wave phenomena become dominant.
- Advantages: $0 \leq |\mathbf{S}| \leq 1$ and can be measured using VNAs (Vector Network Analyzers).

Scattering or S-parameters

Given a system in I/O representation: $\mathbf{y}(s) = \mathbf{H}(s)\mathbf{u}(s)$,

the associated **S-paremeter representation** is

$$\bar{\mathbf{y}}(s) = \mathbf{S}(s)\bar{\mathbf{u}}(s) = \underbrace{[\mathbf{H}(s) + \mathbf{I}][\mathbf{H}(s) - \mathbf{I}]^{-1}}_{\mathbf{S}(s)}\, \bar{\mathbf{u}}(s),$$

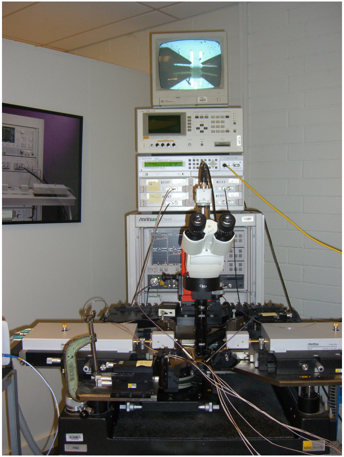
where

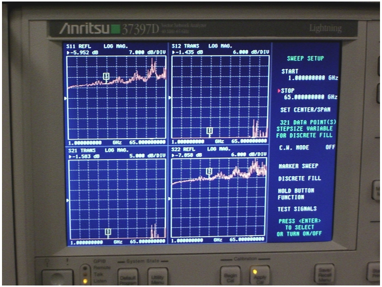
$\bar{\mathbf{y}} = \frac{1}{2}(\mathbf{y} + \mathbf{u})$ are the *transmitted waves* and,
$\bar{\mathbf{u}} = \frac{1}{2}(\mathbf{y} - \mathbf{u})$ are the *reflected waves*.

S-parameter measurements.

$\mathbf{S}(j\omega_k)$: samples of the frequency response of the **S**-parameter system representation.

Measurement of S-parameters





VNA (Vector Network Analyzer) – Magnitude of S-parameters for 2 ports

Model construction from data:

Interpolation

Assume for simplicity that the given data are scalar:

$$(s_i, \phi_i),\ i = 1, 2, \cdots, N,\ \ s_i \neq s_j,\ i \neq j$$

Find $\mathbf{H}(s) = \frac{\mathbf{n}(s)}{\mathbf{d}(s)}$ such that $\mathbf{H}(s_i) = \phi_i$, $i = 1, 2, \cdots, N$, and $\mathbf{n}, \mathbf{d}$: coprime polynomials.

Main tool: **Loewner matrix**. Divide the data in disjoint sets: (λ_i, w_i), $i = 1, 2, \cdots, k$, (μ_j, v_j), $j = 1, 2, \cdots, q$, $k + q = N$:

$$\mathbb{L} = \begin{bmatrix} \frac{v_1 - w_1}{\mu_1 - \lambda_1} & \cdots & \frac{v_1 - w_k}{\mu_1 - \lambda_k} \\ \vdots & \ddots & \vdots \\ \frac{v_q - w_1}{\mu_q - \lambda_1} & \cdots & \frac{v_q - w_k}{\mu_q - \lambda_k} \end{bmatrix} \in \mathbb{C}^{q \times k}$$

Model construction from data:

Interpolation

Main results (1986). • The rank of $\mathbb{L}$ encodes the information about the minimal degree interpolants: $\mathrm{rank}\,\mathbb{L}$ or $N - \mathrm{rank}\,\mathbb{L}$.

• If $\mathbf{H}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$, then

$$\mathbb{L} = -\underbrace{\begin{bmatrix} \mathbf{C}(\lambda_1\mathbf{I} - \mathbf{A})^{-1} \\ \mathbf{C}(\lambda_2\mathbf{I} - \mathbf{A})^{-1} \\ \vdots \\ \mathbf{C}(\lambda_k\mathbf{I} - \mathbf{A})^{-1} \end{bmatrix}}_{\mathcal{O}} \underbrace{\begin{bmatrix} (\mu_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} & \cdots & (\mu_q\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \end{bmatrix}}_{\mathcal{R}}$$

• **Special case.** single point with multiplicity: $(s_0; \phi_0, \phi_1, \cdots, \phi_{N-1})$. Then

$$\mathbb{L} = \begin{bmatrix} \frac{\phi_1}{1!} & \frac{\phi_2}{2!} & \frac{\phi_3}{3!} & \frac{\phi_4}{4!} & \cdots \\ \frac{\phi_2}{2!} & \frac{\phi_3}{3!} & \frac{\phi_4}{4!} & \cdots & \\ \frac{\phi_3}{3!} & \frac{\phi_4}{4!} & & & \\ \frac{\phi_4}{4!} & \vdots & & \ddots & \\ \vdots & & & & \end{bmatrix} \Rightarrow \textbf{Hankel structure}$$

Thus the **Loewner matrix** **generalizes** the **Hankel matrix** when general interpolation replaces realization.

General framework – tangential interpolation

Given: • right data: $(\lambda_i; \mathbf{r}_i, \mathbf{w}_i)$, $i = 1, \cdots, k$

• left data: $(\mu_j; \ell_j, \mathbf{v}_j)$, $j = 1, \cdots, q$.

We assume for simplicity that all points are distinct.

Problem: Find rational $p \times m$ matrices $\mathbf{H}(s)$, such that

$$\boxed{\mathbf{H}(\lambda_i)\mathbf{r}_i = \mathbf{w}_i} \qquad \boxed{\ell_j\mathbf{H}(\mu_j) = \mathbf{v}_j}$$

Right data:

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{bmatrix} \in \mathbb{C}^{k\times k}, \quad \begin{array}{l} \mathbf{R} = [\mathbf{r}_1 \;\; \mathbf{r}_2, \;\; \cdots \;\; \mathbf{r}_k] \in \mathbb{C}^{m\times k}, \\ \mathbf{W} = [\mathbf{w}_1 \;\; \mathbf{w}_2 \;\; \cdots \;\; \mathbf{w}_k] \in \mathbb{C}^{p\times k} \end{array}$$

Left data:

$$M = \begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_q \end{bmatrix} \in \mathbb{C}^{q\times q}, \mathbf{L} = \begin{bmatrix} \ell_1 \\ \vdots \\ \ell_q \end{bmatrix} \in \mathbb{C}^{q\times p}, \mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_q \end{bmatrix} \in \mathbb{C}^{q\times m}$$

General framework – tangential interpolation

Input-output data: $\mathbf{H}(\lambda_i)\mathbf{r}_i = \mathbf{w}_i$, $\ell_j\mathbf{H}(\mu_j) = \mathbf{v}_j$. The Loewner matrix is:

$$\mathbb{L} = \begin{bmatrix} \frac{\mathbf{v}_1\mathbf{r}_1 - \ell_1\mathbf{w}_1}{\mu_1 - \lambda_1} & \cdots & \frac{\mathbf{v}_1\mathbf{r}_k - \ell_1\mathbf{w}_k}{\mu_1 - \lambda_k} \\ \vdots & \ddots & \vdots \\ \frac{\mathbf{v}_q\mathbf{r}_1 - \ell_q\mathbf{w}_1}{\mu_q - \lambda_1} & \cdots & \frac{\mathbf{v}_q\mathbf{r}_k - \ell_q\mathbf{w}_k}{\mu_q - \lambda_k} \end{bmatrix} \in \mathbb{C}^{q\times k}$$

Therefore $\mathbb{L}$ satisfies the Sylvester equation

$$\mathbb{L}\Lambda - M\mathbb{L} = \mathbf{VR} - \mathbf{LW}$$

Given a realization $\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}$: $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}$, let $\mathbf{X}, \mathbf{Y}$ be the generalized reachability/obervability matrices:

$\mathbf{x}_i = (\lambda_i\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{r}_i \;\Rightarrow \mathbf{X}$: *generalized* reachability matrix
$\mathbf{y}_j = \ell_j\mathbf{C}(\mu_j\mathbf{E} - \mathbf{A})^{-1} \;\Rightarrow \mathbf{Y}$: *generalized* observability matrix.

$$\Rightarrow \boxed{\mathbb{L} = -\mathbf{YEX}}$$

The shifted Loewner matrix

- The shifted Loewner matrix, $\sigma\mathbb{L}$, is the Loewner matrix associated to $s\mathbf{H}(s)$.

$$\sigma\mathbb{L} = \begin{bmatrix} \frac{\mu_1 \mathbf{v}_1 \mathbf{r}_1 - \ell_1 \mathbf{w}_1 \lambda_1}{\mu_1 - \lambda_1} & \cdots & \frac{\mu_1 \mathbf{v}_1 \mathbf{r}_k - \ell_1 \mathbf{w}_k \lambda_k}{\mu_1 - \lambda_k} \\ \vdots & \ddots & \vdots \\ \frac{\mu_q \mathbf{v}_q \mathbf{r}_1 - \ell_q \mathbf{w}_1 \lambda_1}{\mu_q - \lambda_1} & \cdots & \frac{\mu_q \mathbf{v}_q \mathbf{r}_k - \ell_q \mathbf{w}_k \lambda_k}{\mu_q - \lambda_k} \end{bmatrix} \in \mathbb{C}^{q \times k}$$

- $\sigma\mathbb{L}$ satisfies the Sylvester equation

$$\boxed{\sigma\mathbb{L}\Lambda - M\sigma\mathbb{L} = \mathbf{VR}\Lambda - M\mathbf{LW}}$$

- $\sigma\mathbb{L}$ can be factored as

$$\Rightarrow \quad \boxed{\sigma\mathbb{L} = -\mathbf{YAX}}$$

Construction of Interpolants (Models)

Assume that $k = \ell$, and let

$$\det(x\mathbb{L} - \sigma\mathbb{L}) \neq 0, \quad x \in \{\lambda_i\} \cup \{\mu_j\}$$

Then

$$\mathbf{E} = -\mathbb{L}, \;\; \mathbf{A} = -\sigma\mathbb{L}, \;\; \mathbf{B} = \mathbf{V}, \;\; \mathbf{C} = \mathbf{W}$$

is a minimal realization of an interpolant of the data, i.e., the function

$$\boxed{\mathbf{H}(s) = \mathbf{W}(\sigma\mathbb{L} - s\mathbb{L})^{-1}\mathbf{V}}$$

interpolates the data.

Proof. Multiplying the first equation by s and subtracting it from the second we get

$$(\sigma\mathbb{L} - s\mathbb{L})\Lambda - M(\sigma\mathbb{L} - s\mathbb{L}) = \mathbf{L}\mathbf{W}(\Lambda - s\mathbf{I}) - (M - s\mathbf{I})\mathbf{V}\mathbf{R}.$$

Multiplying this equation by $\mathbf{e}_i$ on the right and setting $s = \lambda_i$, we obtain

$$(\lambda_i\mathbf{I} - M)(\sigma\mathbb{L} - \lambda_i\mathbb{L})\mathbf{e}_i = (\lambda_i\mathbf{I} - M)\mathbf{V}\mathbf{r}_i \;\; \Rightarrow \;\; (\lambda_i\mathbb{L} - \sigma\mathbb{L})\mathbf{e}_i = \mathbf{V}\mathbf{r}_i \;\; \Rightarrow \;\; \mathbf{W}\mathbf{e}_i = \mathbf{W}(\lambda_i\mathbb{L} - \sigma\mathbb{L})^{-1}\mathbf{V}$$

Therefore $\mathbf{w}_i = \mathbf{H}(\lambda_i)\mathbf{r}_i$. This proves the right tangential interpolation property. To prove the left tangential interpolation property, we multiply the above equation by $\mathbf{e}_j^*$ on the left and set $s = \mu_j$:

$$\mathbf{e}_j^*(\sigma\mathbb{L} - \mu_j\mathbb{L})(\Lambda - \mu_j\mathbf{I}) = \mathbf{e}_j^*\mathbf{L}\mathbf{W}(\mu_j\mathbf{I} - \Lambda) \;\; \Rightarrow \;\; \mathbf{e}_j^*(\sigma\mathbb{L} - \mu_j\mathbb{L}) = \ell_j\mathbf{W} \;\; \Rightarrow \;\; \mathbf{e}_j^*\mathbf{V} = \ell_j\mathbf{W}(\sigma\mathbb{L} - \mu_j\mathbb{L})^{-1}\mathbf{V}$$

Therefore $\mathbf{v}_j = \ell_j\mathbf{H}(\mu_j)$.

Construction of interpolants:

New procedure

Main assumption:

$$\boxed{\text{rank}\,(x\mathbb{L}-\sigma\mathbb{L}) = \text{rank}\,\begin{pmatrix} \mathbb{L} & \sigma\mathbb{L} \end{pmatrix} = \text{rank}\begin{pmatrix} \mathbb{L} \\ \sigma\mathbb{L} \end{pmatrix} =: k,\ \ x \in \{\lambda_i\} \cup \{\mu_j\}}$$

Then for some $x \in \{\lambda_i\} \cup \{\mu_j\}$, we compute the *SVD*

$$x\mathbb{L}-\sigma\mathbb{L} = \mathbf{Y}\Sigma\mathbf{X}$$

with $\text{rank}\,(x\mathbb{L}-\sigma\mathbb{L}) = \text{rank}\,(\Sigma) = \text{size}\,(\Sigma) =: k$, $\mathbf{Y} \in \mathbb{C}^{\nu\times k}$, $\mathbf{X} \in \mathbb{C}^{k\times\rho}$.

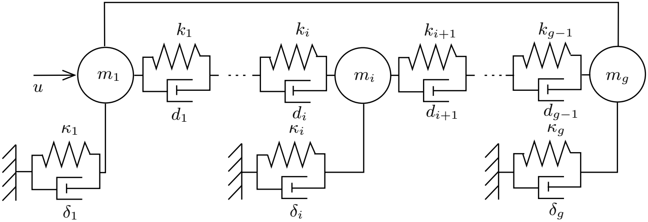
Theorem. A realization $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$, of an interpolant is given as follows:

$\mathbf{E} = -\mathbf{Y}^*\mathbb{L}\mathbf{X}^*$	$\mathbf{B} = \mathbf{Y}^*\mathbf{V}$
$\mathbf{A} = -\mathbf{Y}^*\sigma\mathbb{L}\mathbf{X}^*$	$\mathbf{C} = \mathbf{W}\mathbf{X}^*$

Remark. The system $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$ can now be further reduced using any of the usual reduction methods.

Example: mechanical system

Mechanical example: Stykel, Mehrmann

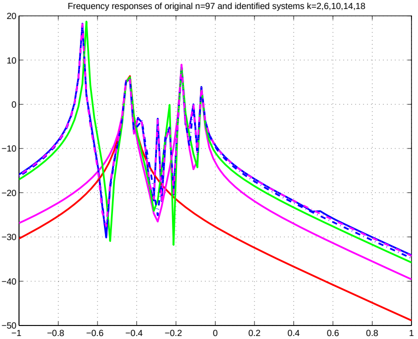


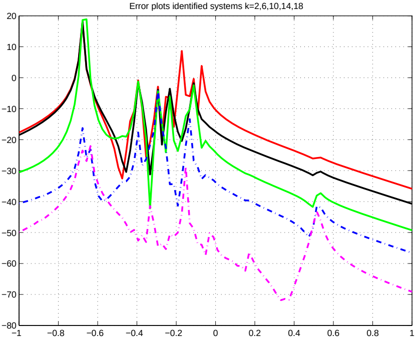
The vibration of this system is described in generalized state space form as:

$$\begin{aligned} \dot{\mathbf{p}}(t) &= \mathbf{v}(t) \\ \mathbf{M}\dot{\mathbf{v}}(t) &= \mathbf{K}\mathbf{p}(t) + \mathbf{D}\mathbf{v}(t) - \mathbf{G}^*\lambda(t) + \mathbf{B}_2\mathbf{u}(t) \\ \mathbf{0} &= \mathbf{G}\mathbf{p}(t) \\ \mathbf{y}(t) &= \mathbf{C}_1\mathbf{p}(t) \end{aligned}$$

Measurements: 500 frequency response data between $[-2i, +2i]$.

Mechanical system: plots



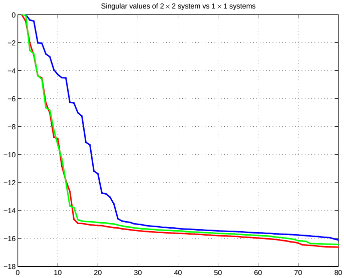


Left: Frequency responses of original system and approximants (orders 2, 6, 10, 14, 18)

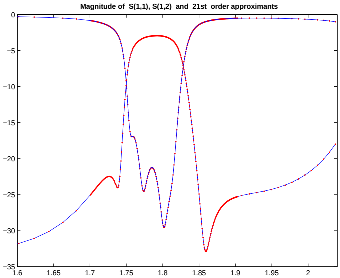
Right: Frequecy responses of error systems

Example: Four-pole band-pass filter

•1000 measurements between 40 and 120 GHz; S-parameters 2×2**,**
MIMO (approximate) interpolation $\Rightarrow \mathbb{L}, \sigma\mathbb{L} \in \mathbb{R}^{2000\times 2000}$**.**



The singular values of $\mathbb{L}$, $\sigma\mathbb{L}$



The $S(1,1)$ and $S(1,2)$ parameter data

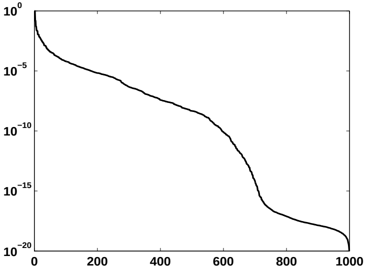
17-th order model

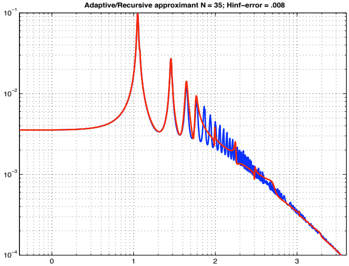
Example: delay system

$$\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}_0\mathbf{x}(t) + \mathbf{A}_1\mathbf{x}(t-\tau) + \mathbf{B}\mathbf{u}(t),\ \ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t),$$

where $\mathbf{E}$, $\mathbf{A}_0$, $\mathbf{A}_1$ are 500×500 and $\mathbf{B}$, $\mathbf{C}^*$ are 500-vectors.

- **Procedure**: compute 1000 frequency response samples.
- Left figure: Singular values of $\mathbb{L}$.
- Then apply **recursive/adaptive** Loewner-framework procedure.
- Right figure: (Blue: original, red: 35th order approximant.)





Outline

Summary

We discussed **SVD** and **Krylov** or **Interpolatory** methods. The latter amount to the choice of interpolation data.

1. Passivity preserving reduction
   Key quantities: spectral zeros
   Problem reduced to a Hamiltonian eigenvalues problem
2. Optimal H_2 reduction
   Key quantities: mirror image of reduced system poles
   Algorithm: Iterative Krylov
   Only model reduction algorithm with error optimization
3. Reduction from data (e.g. S-parameters)
   Key tool tangential interpolation
   Can deal with many input/output ports
   Key tool: Loewner matrix pair
   Natural way to construct full and reduced models
   $\Rightarrow$ does not force inversion of $\mathbf{E}$

(Some) Challenges in model reduction

- Uncertain systems, i.e. systems depending on parameters
- Non-linear systems (POD, TPWL, EIM, etc.)
- Non-linear differential-algebraic (DAE) systems
- Sparsity and parallelization
- Domain decomposition methods
- MEMS and multi-physics problems (micro-fluidic bio-chips)
- CAD tools for VLSI/nanoelectronics
- Molecular Dynamics (MD) simulations
- Model reduction and data assimilation (weather prediction)
- Active control of high-rise buildings
- $\cdots$

References

- Balanced truncation with non-zero initial conditions
  - Reis, Heinkenschloss, Antoulas, CAAM TR 09-29, Rice University (2009)
- Passivity preserving model reduction
  - Antoulas: Systems and Control Letters (2005)
  - Sorensen: Systems and Control Letters (2005)
  - Ionutiu, Rommes, Antoulas: Passivity-Preserving Model Reduction Using Dominant Spectral-Zero Interpolation, IEEE Trans. CAD (Computer-Aided Design of Integrated Circuits and Systems), vol 27, pages: 2250 - 2263 (2008).
- Optimal $\mathcal{H}_2$ model reduction
  - Gugercin, Antoulas, Beattie: SIAM J. Matrix Anal. Appl. (2008)
  - Kellems, Roos, Xiao, Cox: Low-dimensional, morphologically accurate models of subthreshold membrane potential, J. Comput. Neuroscience, 27:161-176 (2009).

References

- Model reduction from data
  - A.J. Mayo and A.C. Antoulas, *A framework for the solution of the generalized realization problem*, Linear Algebra and Its Applications, vol. 425, pages 634-662 (2007).
  - Lefteriu, Antoulas: A New Approach to Modeling Multiport Systems from Frequency-Domain Data, IEEE Trans. CAD, vol. 29, pages 14-27 (2010).
- Interpolatory model reduction
  - A.C. Antoulas, C.A. Beattie, and S. Gugercin, *Interpolatory model reduction of large-scale systems*, Report, September 2009.
  - C.A. Beattie and S. Gugercin, *Interpolatory projection methods for structure preserving model reduction*, Systems and Control Letters, vol. 58, 2009.
- General reference: Antoulas, SIAM (2005)