Numerical Linear Algebra Techniques in Reduced-Order Modeling of Large-Scale Linear Dynamical Systems

Outline

1. Linear dynamical systems
2. Transfer function calculation
3. Reduced-order modeling
4. Padé approximation and moment-matching
5. Padé and Lanczos connection (PVL method)
6. Error estimation (convergence analysis)

Linear dynamical systems

Continuous, time-invariant, MIMO

$$\begin{aligned} \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{G}\mathbf{x}(t) &= \mathbf{B}\,\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{L}^T\mathbf{x}(t), \end{aligned}$$

with $\mathbf{x}(0) = \mathbf{x}_0$, where

$\mathbf{x}(t) \in \mathbf{R}^{N\times 1}$:	state variables N state-space dimension
$\mathbf{C}, \mathbf{G} \in \mathbf{R}^{N\times N}$	system matrices
$\mathbf{B} \in \mathbf{R}^{N\times m}$:	input influence arrays
$\mathbf{L} \in \mathbf{R}^{N\times p}$:	output influence arrays
$\mathbf{u}(t) \in \mathbf{R}^{m\times 1}$:	inputs (m)
$\mathbf{y}(t) \in \mathbf{R}^{p\times 1}$:	outputs (p)

Remarks

Assume that the pencil $\lambda \mathbf{C} + \mathbf{G}$ is regular, i.e.,

$$\det(\lambda \mathbf{C} + \mathbf{G}) \not\equiv 0.$$

Important special cases:

- $\mathbf{C}^T = \mathbf{C}$ and $\mathbf{G}^T = \mathbf{G}$, indefinite.
  e.g., RLC systems,
  linearization of symmetric quadratic systems
- $\mathbf{C}^T = \mathbf{C} \geq 0$ and $\mathbf{G}^T = \mathbf{G} \geq 0$.
  e.g., (extracted) RC systems

Other sources and applications of linear systems:

- Semi-discretization of PDEs
- Small-signal analysis of nonlinear systems
- Nonlinear systems with large linear subsystems

Linear dynamical systems in frequency domain

By taking Laplace transform

$$F(s) = \mathcal{L}\{f(t)\} = \int_{0+}^{\infty} f(t)e^{-st}dt,$$

we have

$$\begin{aligned} s\mathbf{C}\mathbf{X}(s) &= -\mathbf{G}\mathbf{X}(s) + \mathbf{B}\,\mathbf{U}(s) \\ \mathbf{Y}(s) &= \mathbf{L}^T\mathbf{X}(s) \end{aligned}$$

where

$$\{\mathbf{X}(s), \mathbf{Y}(s), \mathbf{U}(s)\} = \mathcal{L}\{\mathbf{x}(t), \mathbf{y}(t), \mathbf{u}(t)\}$$

and assume $\mathbf{x}(0+) = \mathbf{x}_0 = 0$.

The ratio of the output $\mathbf{Y}(s)$ to the input $\mathbf{U}(s)$ is given by the *transfer (matrix) function*

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

Computational tasks and challenges

Tasks:

- Steady-state analysis

  the steady-state response of a system to a sinusoidal input signal (or periodic excitation).

- Transient analysis

  the response of a system as a function of time

- Sensitivity analysis

  the proportional change of a system to a proportional change in the system parameters.

Challenges:

- Large state-space dimension $N = \mathcal{O}(10^3) \sim \mathcal{O}(10^6)$
- Stiffness (multi-energy domain, multi-scaling, ...)

Eigensystem analysis of transfer function

For simplicity, let $p = m = 1$ (SISO system).
To compute $H(s)$ around an expansion point s_0, let $s = s_0 + \sigma$. Then

$$H(s) = H(s_0 + \sigma) = \mathbf{l}^T(\mathbf{I} - \sigma\mathbf{A})^{-1}\mathbf{r}$$

where

$$\mathbf{A} = -(\mathbf{G} + s_0\mathbf{C})^{-1}\mathbf{C}, \qquad \mathbf{r} = (\mathbf{G} + s_0\mathbf{C})^{-1}\mathbf{b}.$$

Eigendecomposition of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{S}\,\Lambda\,\mathbf{S}^{-1} = \mathbf{S} \cdot \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N) \cdot \mathbf{S}^{-1}$$

Pole-residue representation

$$H(s_0+\sigma) = \mathbf{f}^T(\mathbf{I}-\sigma\Lambda)^{-1}\mathbf{g} = \rho_\infty + \sum_{\lambda_j\neq 0} \frac{\kappa_j}{\sigma - p_j}.$$

where

$p_j = s_0 + \frac{1}{\lambda_j}$, **poles** $\qquad \kappa_j = -\frac{f_j g_j}{\lambda_j}$, **residue**

$\mathbf{f} = \mathbf{S}^T\mathbf{l} = (f_j)$, $\qquad \mathbf{g} = \mathbf{S}^{-1}\mathbf{r} = (g_j)$.

In practice, A is either too ill-conditioned or too large to compute its full eigendecomposition!

Eigensystem analysis of transfer function

Alternative "direct" methods (for example):

- Using Schur decomposition.

  Let $\mathbf{A} = \mathbf{QTQ}^T$. Then

$$H(s) = \mathbf{l}^T(\mathbf{I} - \sigma\mathbf{A})^{-1}\mathbf{r} = (\mathbf{Q}^T\mathbf{l})^T(\mathbf{I} - \sigma\mathbf{T})^{-1}(\mathbf{Q}^T\mathbf{r})$$

Approximation (for example):

- Use partial eigen-decomposition

  With harmonic excitation $\mathbf{Bu}(t) = \mathbf{P}(\omega)e^{i\omega t}$, one may assume a harmonic solution of the form

$$\mathbf{x}(t) = \mathbf{Q}_k\mathbf{v}(\omega)e^{i\omega t}$$

  Then solve

$$\left[i\omega\mathbf{Q}_k^T\mathbf{C}\mathbf{Q}_k + \mathbf{Q}_k^T\mathbf{G}\mathbf{Q}_k\right]\mathbf{v}(\omega) = \mathbf{Q}_k^T\mathbf{P}(\omega)$$

  for $\mathbf{v}(\omega)$.

  $\mathbf{Q}_k$ are selected mode shapes (eigenvectors), lie within the range of forcing frequencies (**dominant modes**)

  $\Rightarrow$ *Modal superposition method*

Reudced-order modeling (ROM)

Goals:

(a) To replace the original system by a system of the same type but with *much smaller* state-space dimension

(b) To represent *a meaningful approximation* of the original system

(c) To preserve *essential properties* of the original system.

Efficient analysis and synthesis of a large scale dynamical system.

Reudced-order modeling (ROM), cont'

Specifically, given a linear system, we want to

Find a new linear system of the same form

$$\begin{aligned} \mathbf{C}_n \dot{\mathbf{z}}(t) + \mathbf{G}_n \mathbf{z}(t) &= \mathbf{B}_n \mathbf{u}(t), \\ \tilde{\mathbf{y}}(t) &= \mathbf{L}_n^T \mathbf{z}(t), \end{aligned}$$

where

$\mathbf{C}_n, \mathbf{G}_n \in \mathbf{R}^{n \times n}$

$\mathbf{B}_n \in \mathbf{R}^{n \times m}$

$\mathbf{L}_n \in \mathbf{R}^{n \times p}$

$\mathbf{u}(t) \in \mathbf{R}^{m \times 1}$: *inputs* ($m$)

$\mathbf{z}(t) \in \mathbf{R}^{n \times 1}$

$\tilde{\mathbf{y}}(t) \in \mathbf{R}^{p \times 1}$: *outputs* ($p$)

such that

- $n \ll N$
- $\|\mathbf{y}(t) - \tilde{\mathbf{y}}(t)\| \leq \epsilon$ *for* $\mathbf{u} \in \mathcal{F}, t \in [t_0, t_1]$

ROM in frequency domain

The transfer function of the reduced-order model is of the form:

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

Therefore, in the frequency domain, our *objective* is to ask

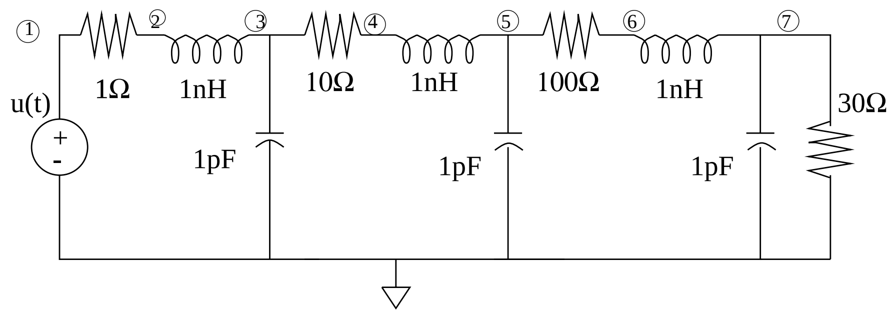
$$\|\mathbf{H}(s) - \mathbf{H}_n(s)\| \leq \epsilon$$

over the frequency range of interest, and

$$n \ll N$$

An illustrative simple example

A simple RLC network `[Chiprout & Nakhla'94]`



Modified nodal admittance formulation (Kirchoff's laws) yields

$$\begin{aligned} \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{G}\mathbf{x}(t) &= \mathbf{B}\,\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{L}^T\mathbf{x}(t), \end{aligned}$$

where

- $\mathbf{x}(t)$: 11×1,
- $\mathbf{C}$ and $\mathbf{G}$: 11×11,
- $\mathbf{B} = \mathbf{e}_8$,
- $\mathbf{L} = \mathbf{e}_7$.

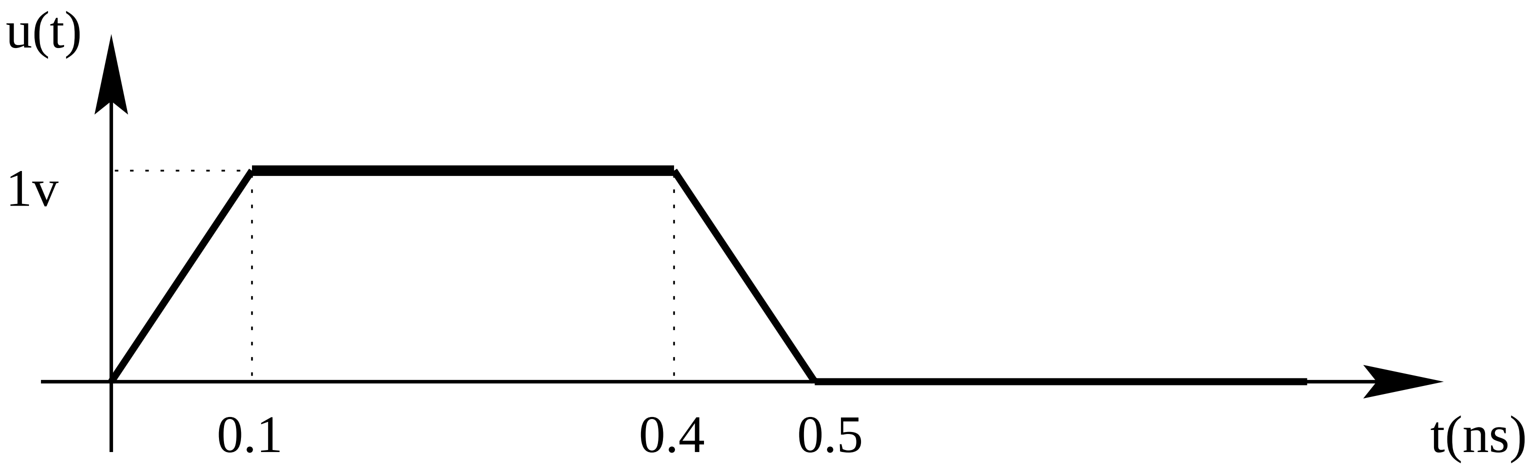
```
L=[0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 1 ; 0 ; 0 ; 0 ; 0];

B=[0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 1 ; 0 ; 0 ; 0];

G=[ 1 -1     0     0      0      0      0 1  0  0  0;
   -1  1     0     0      0      0      0 0  1  0  0;
    0  0   0.1  -0.1      0      0      0 0 -1  0  0;
    0  0  -0.1   0.1      0      0      0 0  0  1  0;
    0  0     0     0   0.01  -0.01      0 0  0 -1  0;
    0  0     0     0  -0.01   0.01      0 0  0  0  1;
    0  0     0     0      0      0 0.0333 0  0  0 -1;
    1  0     0     0      0      0      0 0  0  0  0;
    0  1    -1     0      0      0      0 0  0  0  0;
    0  0     0     1     -1      0      0 0  0  0  0;
    0  0     0     0      0      1     -1 0  0  0  0];
```

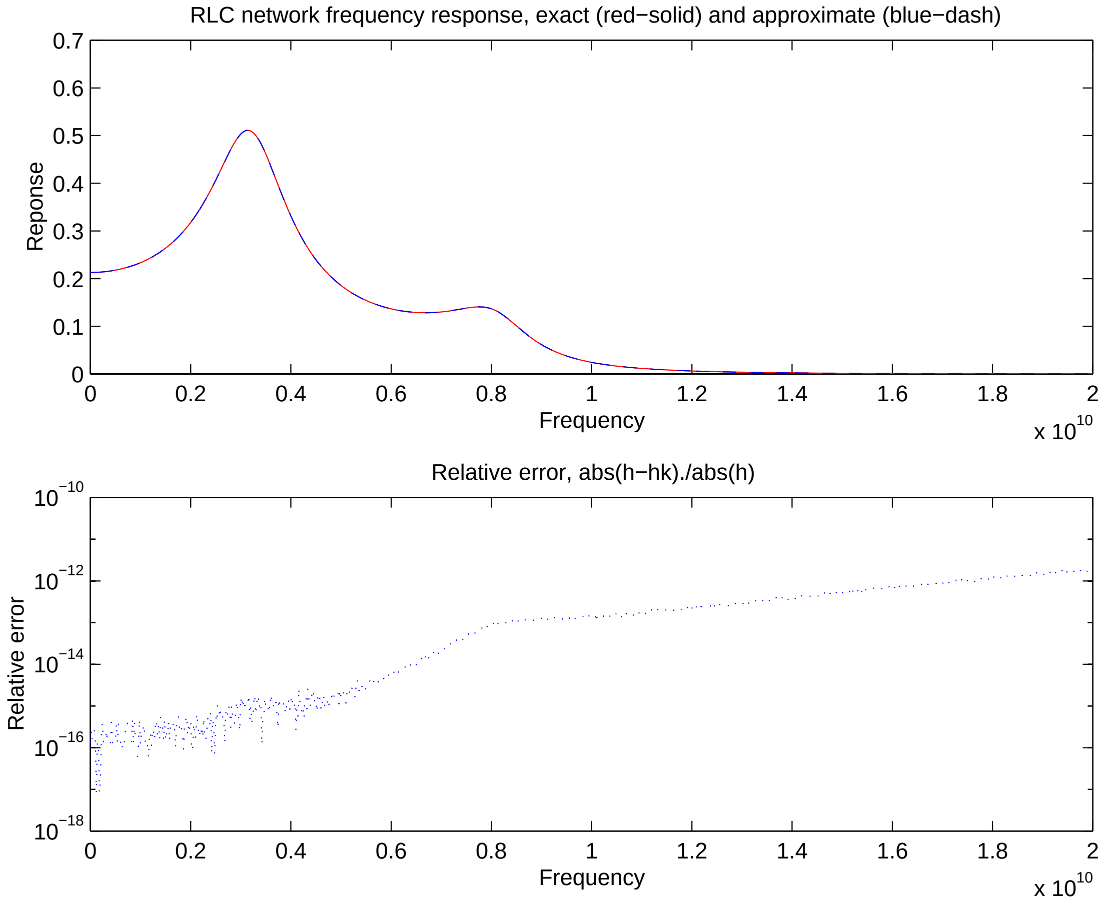
```
C=[ 0 0      0 0      0 0      0 0     0     0     0;
    0 0      0 0      0 0      0 0     0     0     0;
    0 0 1e-12 0      0 0      0 0     0     0     0;
    0 0      0 0      0 0      0 0     0     0     0;
    0 0      0 0 1e-12 0      0 0     0     0     0;
    0 0      0 0      0 0      0 0     0     0     0;
    0 0      0 0      0 0 1e-12 0     0     0     0;
    0 0      0 0      0 0      0 0     0     0     0;
    0 0      0 0      0 0      0 0 -1e-9     0     0;
    0 0      0 0      0 0      0 0     0 -1e-9     0;
    0 0      0 0      0 0      0 0     0     0 -1e-9];
}
```

and $\mathbf{u}(t)$:

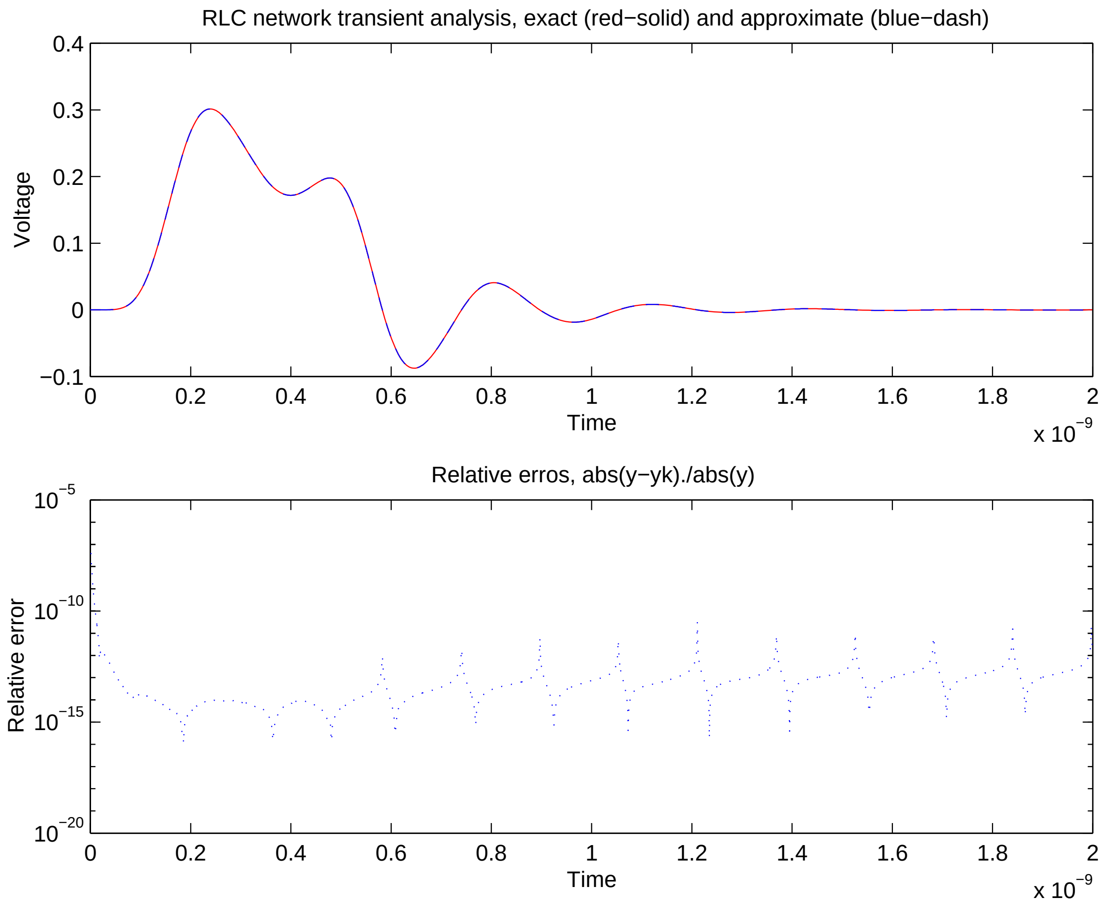


An illustrative simple example, cont'

Frequency responses $H(s)$ and $H_6(s)$:



Transient responses $y(t)$ and $\tilde{y}_6(t)$:



Padé approximation and moment-matching

Since $H(s_0+\sigma)$ is a rational function, it may be best approximated by Padé approximation:

$$H_n(s_0+\sigma) = \frac{P_{n-1}(\sigma)}{Q_n(\sigma)} \equiv \frac{a_{n-1}\sigma^{n-1} + \cdots + a_1\sigma + a_0}{b_n\sigma^n + b_{n-1}\sigma^{n-1} + \cdots + b_1\sigma + 1}$$

The coefficients $\{a_j\}$ and $\{b_j\}$ are uniquely determined by the first $2n$ Taylor coefficients of $H(s_0+\sigma)$:

$$\begin{aligned} H(s_0+\sigma) &= \mathbf{l}^T(\mathbf{I}-\sigma\mathbf{A})^{-1}\mathbf{r} \\ &= \mathbf{l}^T\mathbf{r} + (\mathbf{l}^T\mathbf{A}\mathbf{r})\sigma + (\mathbf{l}^T\mathbf{A}^2\mathbf{r})\sigma^2 + \cdots \\ &\equiv m_0 + m_1\sigma + m_2\sigma^2 + \cdots \end{aligned}$$

where

$$m_j = \mathbf{l}^T\mathbf{A}^j\mathbf{r} \quad \text{for} \quad j = 0, 1, 2, \ldots$$

are called moments (or markov parameters).

Padé approximation

$$H(s_0 + \sigma) = H_n(s_0 + \sigma) + \mathcal{O}(\sigma^{2n})$$

Padé approximation and moment-matching, cont'

Multiplying $Q_n(\sigma)$ on the both side of

$$H(s_0+\sigma) \cong H_n(s_0+\sigma) = \frac{P_{n-1}(\sigma)}{Q_n(\sigma)}$$

yields

$$H(s_0+\sigma)Q_n(\sigma) \cong P_{n-1}(\sigma).$$

Comparing the coefficient of σ^k terms for $k = 0, 1, \ldots, 2n-1$, we have

$$\begin{bmatrix} m_0 & m_1 & \cdots & m_{n-1} \\ m_1 & m_2 & \cdots & m_n \\ \vdots & \vdots & \cdots & \vdots \\ m_{n-1} & m_n & \cdots & m_{2n-2} \end{bmatrix} \begin{bmatrix} b_n \\ b_{n-1} \\ \vdots \\ b_1 \end{bmatrix} = - \begin{bmatrix} m_n \\ m_{n+1} \\ \vdots \\ m_{2n-1} \end{bmatrix}$$

where the coefficient matrix is the **Hankel matrix** M_n, and

$$\begin{cases} a_0 &= m_0 \\ a_1 &= m_0 b_1 + m_1 \\ & \vdots \\ a_{n-1} &= m_0 b_{n-1} + m_1 b_{n-2} + \cdots + m_{n-2} b_1 + m_{n-1} \end{cases}$$

Padé approximation and moment-macthing, cont’

$\Rightarrow$ Asymptotic Waveform Evaluations (AWE) techniques.

`[Pillage & Rohrer,’90]`
`[Chiprout & Nakhla,’94]`

However, $\mathbf{M}_n$ is generally very ill-conditioned!

For example

$$\begin{array}{ll} n = 2 & \text{cond}(\mathbf{M}_2) = 2.62 \\ n = 3 & \text{cond}(\mathbf{M}_3) = 5.4 \times 10^3 \\ n = 4 & \text{cond}(\mathbf{M}_4) = 1.13 \times 10^9 \\ n = 6 & \text{cond}(\mathbf{M}_6) = 1.21 \times 10^{17} \end{array}$$

Krylov subspaces

- *Right* Krylov subspace

$$\mathcal{K}_n(\mathbf{A}, \mathbf{r}) = \text{span}\{\, \mathbf{r}, \mathbf{A}\mathbf{r}, \mathbf{A}^2\mathbf{r}, \dots, \mathbf{A}^{n-1}\mathbf{r}\,\}$$

- *Left* Krylov subspace

$$\mathcal{K}_n(\mathbf{A}^T, \mathbf{l}) = \text{span}\{\, \mathbf{l}, \mathbf{A}^T\mathbf{l}, (\mathbf{A}^T)^2\mathbf{l}, \dots, (\mathbf{A}^T)^{n-1}\mathbf{l}\,\}$$

- Krylov subspace and Hankel matrix $\mathbf{M}_n$

$$
\begin{aligned}
& \begin{bmatrix} \mathbf{l}^T \\ \mathbf{l}^T\mathbf{A} \\ \vdots \\ \mathbf{l}^T\mathbf{A}^{n-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r} & \mathbf{A}\mathbf{r} & \cdots & \mathbf{A}^{n-1}\mathbf{r} \end{bmatrix} \\
= & \begin{bmatrix} \mathbf{l}^T\mathbf{r} & \mathbf{l}^T\mathbf{A}\mathbf{r} & \cdots & \mathbf{l}^T\mathbf{A}^{n-1}\mathbf{r} \\ \mathbf{l}^T\mathbf{A}\mathbf{r} & \mathbf{l}^T\mathbf{A}^2\mathbf{r} & \cdots & \mathbf{l}^T\mathbf{A}^{n}\mathbf{r} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{l}^T\mathbf{A}^{n-1}\mathbf{r} & \mathbf{l}^T\mathbf{A}^{n}\mathbf{r} & \cdots & \mathbf{l}^T\mathbf{A}^{2n-2}\mathbf{r} \end{bmatrix} \\
= & \begin{bmatrix} m_0 & m_1 & \cdots & m_{n-1} \\ m_1 & m_2 & \cdots & m_n \\ \vdots & \vdots & \cdots & \vdots \\ m_{n-1} & m_n & \cdots & m_{2n-2} \end{bmatrix} \\
= & \ \mathbf{M}_n
\end{aligned}
$$

Basis Vectors of Krylov subspaces

Key issue: Krylov subspaces provide the desired information, but the vectors

$$\{\, \mathbf{A}^j \mathbf{r} \,\} \qquad \text{and} \qquad \{\, (\mathbf{A}^T)^j \mathbf{l} \,\}$$

are unstable as basis vectors.

Question: How to construct more stable basis vectors

$$\{\, \mathbf{v}_k \,\} \qquad \text{and} \qquad \{\, \mathbf{w}_j \,\}$$

such that

$$\mathcal{K}_n(\mathbf{A}, \mathbf{r}) = \text{span}\{\, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \,\}$$

and

$$\mathcal{K}_n(\mathbf{A}^T, \mathbf{l}) = \text{span}\{\, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n \,\}$$

Solution: Lanczos process is an elegant way to generate the desired basis vectors `[Lanczos,'50]`.

Lanczos procedure = Krylov subspace + TSMGS

Algorithm template

(1) $\rho_1 = \|\mathbf{r}\|_2$
(2) $\eta_1 = \|\mathbf{l}\|_2$
(3) $\mathbf{v}_1 = \mathbf{r}/\rho_1$
(4) $\mathbf{w}_1 = \mathbf{l}/\eta_1$
(5) **for** $k = 1, 2, \ldots, n$ **do**
(6) $\quad \delta_k = \mathbf{w}_k^{\mathrm{T}} \mathbf{v}_k$
(7) $\quad \alpha_k = \mathbf{w}_k^{\mathrm{T}} \mathbf{A} \mathbf{v}_k / \delta_k$
(8) $\quad \beta_k = (\delta_k / \delta_{k-1}) \eta_k$
(9) $\quad \gamma_k = (\delta_k / \delta_{k-1}) \rho_k$
(10) $\quad \mathbf{v} = \mathbf{A}\mathbf{v}_k - \mathbf{v}_k \alpha_k - \mathbf{v}_{k-1} \beta_k$
(11) $\quad \mathbf{w} = \mathbf{A}^{\mathrm{T}} \mathbf{w}_k - \mathbf{w}_k \alpha_k - \mathbf{w}_{k-1} \gamma_k$
(12) $\quad \rho_{k+1} = \|\mathbf{v}\|_2$
(13) $\quad \eta_{k+1} = \|\mathbf{w}\|_2$
(14) $\quad \mathbf{v}_{k+1} = \mathbf{v}/\rho_{k+1}$
(15) $\quad \mathbf{w}_{k+1} = \mathbf{w}/\eta_{k+1}$
(16) **end for**

A fundamental tool in matrix computations.

Lanczos: governing equations

- Lanczos process in matrix-vector form

$$\begin{aligned} \mathbf{A}\mathbf{V}_n &= \mathbf{V}_n\mathbf{T}_n + \rho_{n+1}\mathbf{v}_{n+1}\mathbf{e}_n^T, \\ \mathbf{A}^T\mathbf{W}_n &= \mathbf{W}_n\tilde{\mathbf{T}}_n + \eta_{n+1}\mathbf{w}_{n+1}\mathbf{e}_n^T, \end{aligned}$$

where $\mathbf{T}_n$ and $\tilde{\mathbf{T}}_n$ are the tridiagonal matrices

$$\mathbf{T}_n = \begin{bmatrix} \alpha_1 & \beta_2 & & \\ \rho_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_n \\ & & \rho_n & \alpha_n, \end{bmatrix} \quad \tilde{\mathbf{T}}_n^T = \Delta_n \mathbf{T}_n \Delta_n^{-1} = \begin{bmatrix} \alpha_1 & \gamma_2 & & \\ \eta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \gamma_n \\ & & \eta_n & \alpha_n \end{bmatrix},$$

- Lanczos vectors

$$\begin{aligned} \mathbf{V}_n &= \{\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n\} &\longrightarrow \ \text{span}\{\mathbf{V}_n\} = \mathcal{K}_n(\mathbf{A}, \mathbf{r}) \\ \mathbf{W}_n &= \{\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_n\} &\longrightarrow \ \text{span}\{\mathbf{W}_n\} = \mathcal{K}_n(\mathbf{A}^T, \mathbf{l}) \end{aligned}$$

- Biorthogonality

$$\mathbf{W}_n^T\mathbf{V}_n = \Delta_n = \mathrm{diag}(\delta_k),$$

and $\mathbf{W}_n^T\mathbf{v}_{n+1} = 0$ and $\mathbf{w}_{n+1}^T\mathbf{V}_n = 0$

- Projection

$$\mathbf{W}_n^T \mathbf{A} \mathbf{V}_n = \Delta_n \mathbf{T}_n$$

If the Lanczos process is carried to the end, then

$$\mathbf{V}_N^{-1} \mathbf{A} \mathbf{V}_N = \mathbf{T}_N,$$

Reduced-order modeling by Lanczos

- Let $\mathbf{r} = \mathbf{v}_1 \rho_1$ and $\mathbf{l}^T = \mathbf{w}_1^T \eta_1$

$$\begin{aligned} H(s_0 + \sigma) &= \mathbf{l}^T (\mathbf{I} - \sigma \mathbf{A})^{-1} \mathbf{r} \\ &= (\mathbf{l}^T \mathbf{r})\, \mathbf{e}_1^T (\mathbf{I} - \sigma \mathbf{T}_N)^{-1} \mathbf{e}_1 \\ &= (\mathbf{l}^T \mathbf{r})\, \frac{\det(\mathbf{I} - \sigma \mathbf{T}'_N)}{\det(\mathbf{I} - \sigma \mathbf{T}_N)} \\ &= \frac{\text{zero}}{\text{pole}} \text{ representation} \end{aligned}$$

where $\mathbf{T}'_N$ is an $(N-1) \times (N-1)$ matrix obtained by deleting the first row and column of $\mathbf{T}_N$. Note that

$$\mathbf{X}^{-1} = \frac{\text{adj}(\mathbf{X})}{\det(\mathbf{X})}$$

- Define the n-th ROM of transfer function

$$H_n(s_0 + \sigma) = (\mathbf{l}^T \mathbf{r})\, \mathbf{e}_1^T (\mathbf{I} - \sigma \mathbf{T}_n)^{-1} \mathbf{e}_1$$

where

$$\mathbf{T}_N = \begin{array}{cc} & \begin{array}{cc} n & \end{array} \\ n & \left(\begin{array}{cc} \mathbf{T}_n & \\ & \ddots \end{array}\right) \end{array}$$

- Question: *What is $H_n(s_0 + \sigma)$?*

What is $H_n(s_0 + \sigma)$?

- Lanczos-Padé connection

  Theorem. $H_n(s_0 + \sigma)$ is the Padé approximation of $H(s_0 + \sigma)$.

- Lanczos-Padé connection

  [Gragg,'74, Gragg & Lindquist,'83]

- Krylov subspace, moments matching and applications to ROM

  [De Villemagen & Skelton,'87]

  [Craig, Hale & Su,'88--'92]

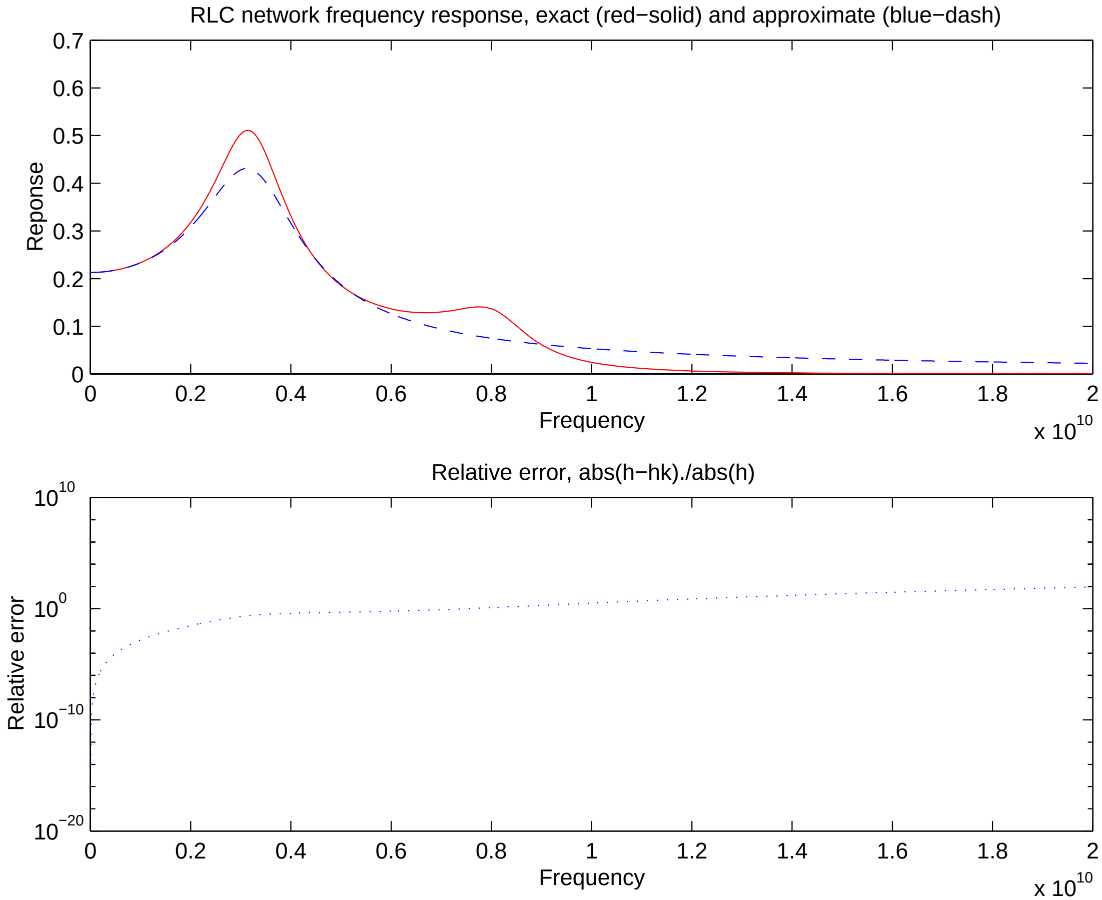
  [Feldman & Freund,'94] (PVL)

  [Gallivan, Grimme & Van Dooren,'94]

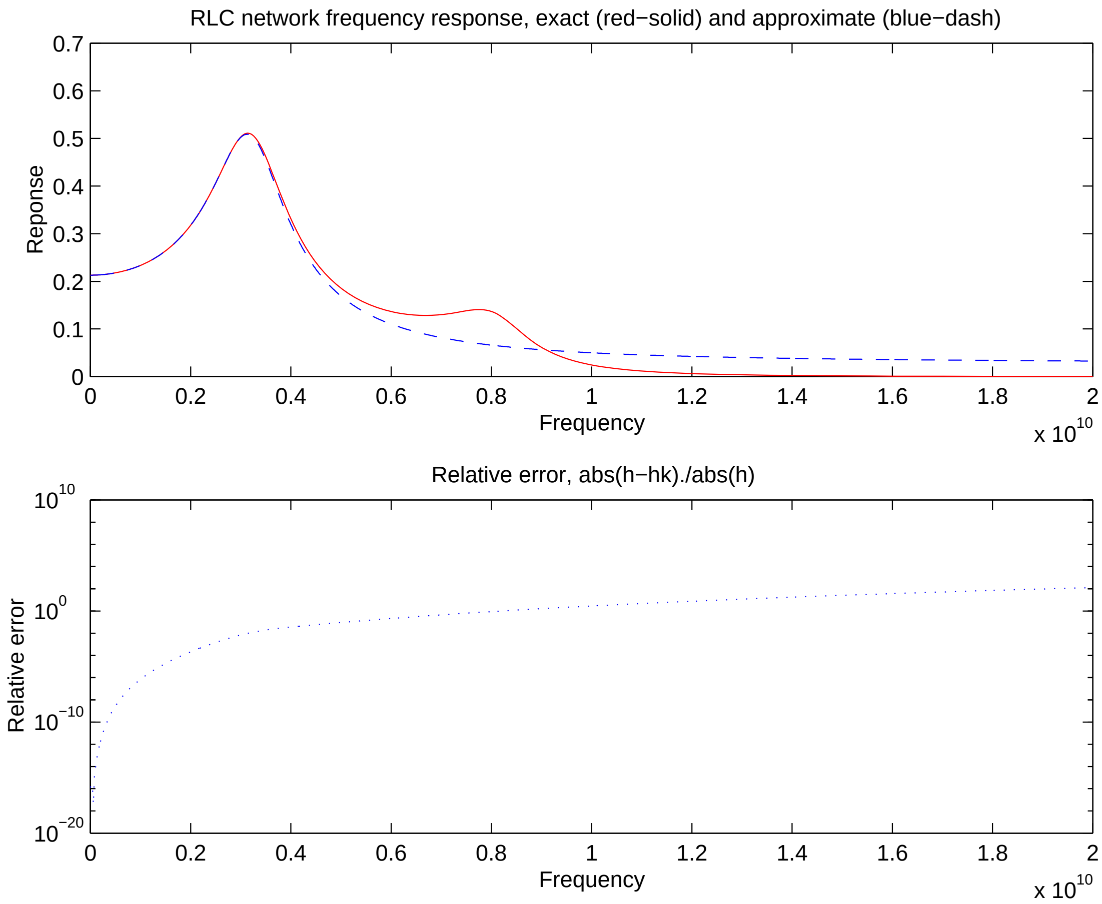
  [Bai & Ye,'97]

Back to the illustrative simple example

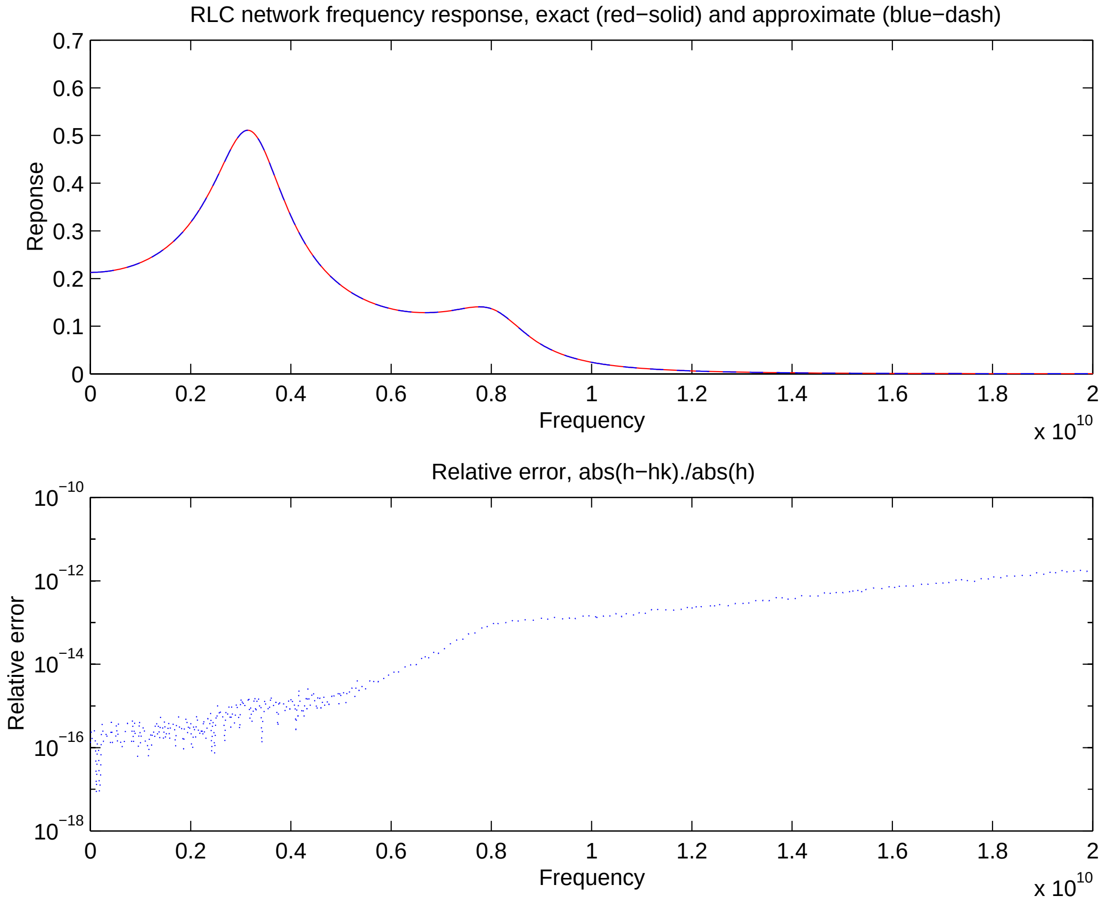
Frequency responses $H(s)$, $H_2(s)$ and errors



Frequency responses $H(s)$, $H_4(s)$ and errors



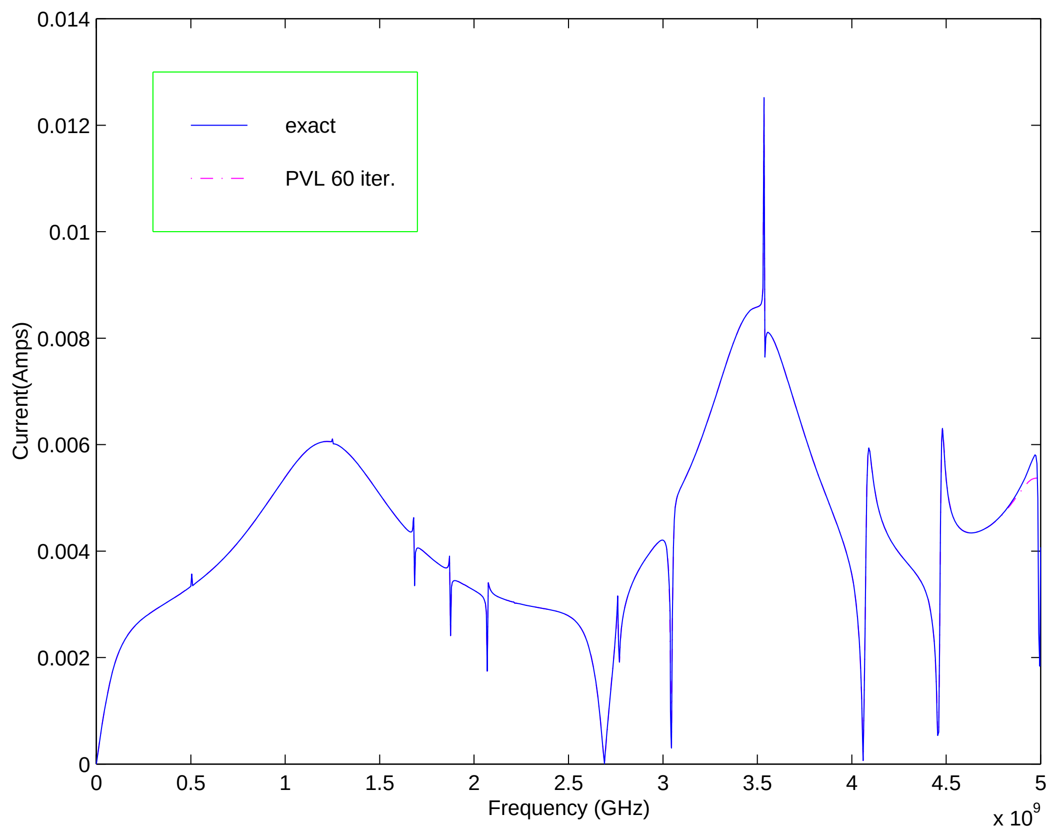
Frequency responses $H(s)$, $H_6(s)$ and errors



Example: the PEEC circuit

A 3D electromagnetic problem modeled via partial element equivalent circuit (PEEC) simulation `[Ruehli,'94]`

$|H(s)|$ and $|H_{60}(s)|$:



ROM in time domain

Linear dynamical system:

$$\begin{cases} \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{G}\mathbf{x}(t) &= \mathbf{B}\,\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{L}^T\mathbf{x}(t), \end{cases}$$

Let $\mathbf{A} = -(\mathbf{G} + s_0\mathbf{C})^{-1}\mathbf{C}$ and $\mathbf{R} = (\mathbf{G} + s_0\mathbf{C})^{-1}\mathbf{B}$, then yield the “shift-and-invert” system

$$\begin{cases} -\mathbf{A}\dot{\mathbf{x}}(t) + (\mathbf{I} + s_0\mathbf{A})\mathbf{x}(t) &= \mathbf{R}\,\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{L}^T\mathbf{x}(t), \end{cases}$$

Let

$$\mathbf{x}(t) \cong \mathbf{V}_n\mathbf{z}(t),$$

then an approximate system is given by

$$\begin{cases} -\mathbf{A}\mathbf{V}_n\dot{\mathbf{z}}(t) + (\mathbf{I} + s_0\mathbf{A})\mathbf{V}_n\mathbf{z}(t) &= \mathbf{R}\,\mathbf{u}(t), \\ \tilde{\mathbf{y}}(t) &= \mathbf{L}^T\mathbf{V}_n\mathbf{z}(t). \end{cases}$$

Multiplying $\mathbf{W}_n^T$ from the left, we have

$$\begin{cases} -\mathbf{W}_n^T\mathbf{A}\mathbf{V}_n\dot{\mathbf{z}}(t) + \mathbf{W}_n^T(\mathbf{I} + s_0\mathbf{A})\mathbf{V}_n\mathbf{z}(t) &= \mathbf{W}_n^T\mathbf{R}\,\mathbf{u}(t), \\ \tilde{\mathbf{y}}(t) &= \mathbf{L}^T\mathbf{V}_n\mathbf{z}(t). \end{cases}$$

By the Lanczos governing equations, we have

$$\begin{cases} -\mathbf{T}_n\dot{\mathbf{z}}(t) + (\mathbf{I}_n + s_0\mathbf{T}_n)\mathbf{z}(t) & = & \mathbf{R}_n\mathbf{u}(t), \\ \tilde{\mathbf{y}}(t) & = & \mathbf{L}_n^T\mathbf{z}(t). \end{cases}$$

Therefore, an n-th reduced-order model is given by

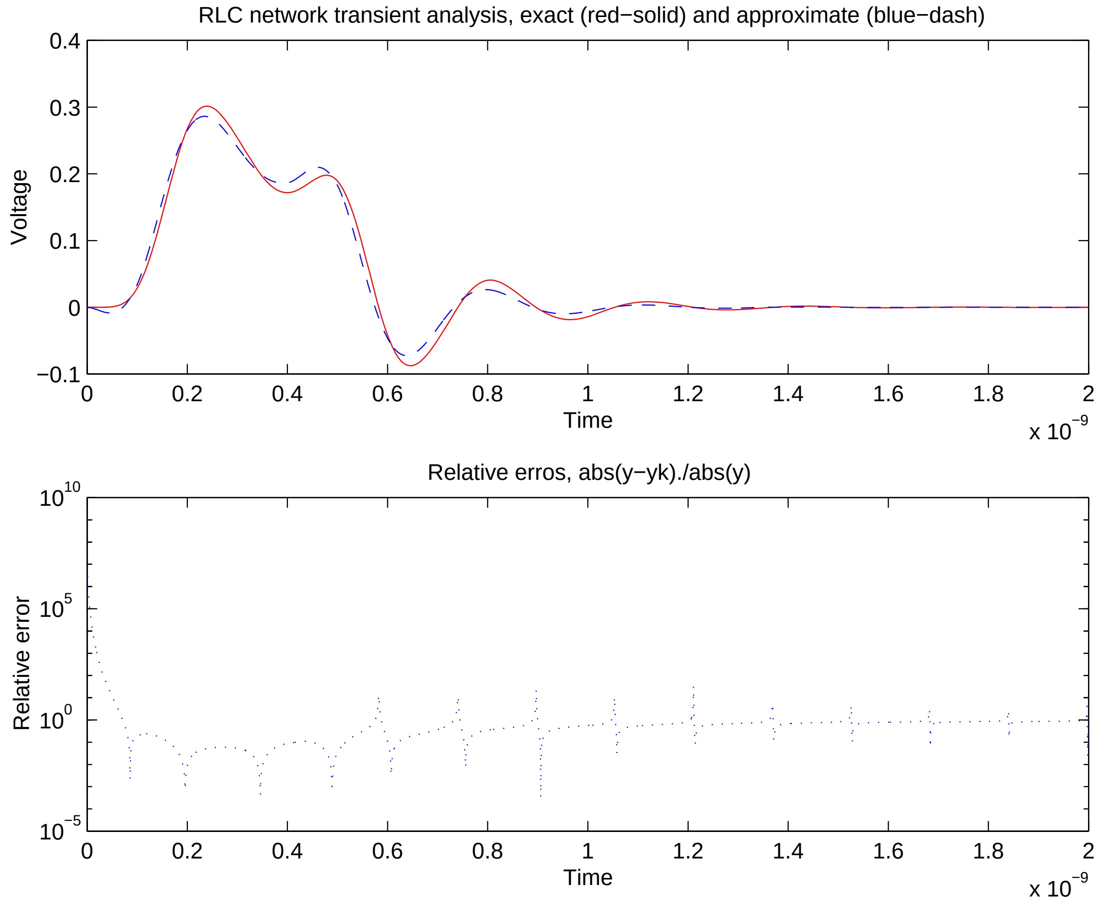
$$\begin{cases} \mathbf{C}_n\dot{\mathbf{z}}(t) + \mathbf{G}_n\mathbf{z}(t) & = & \mathbf{R}_n\mathbf{u}(t), \\ \tilde{\mathbf{y}}(t) & = & \mathbf{L}_n^T\mathbf{z}(t). \end{cases}$$

where

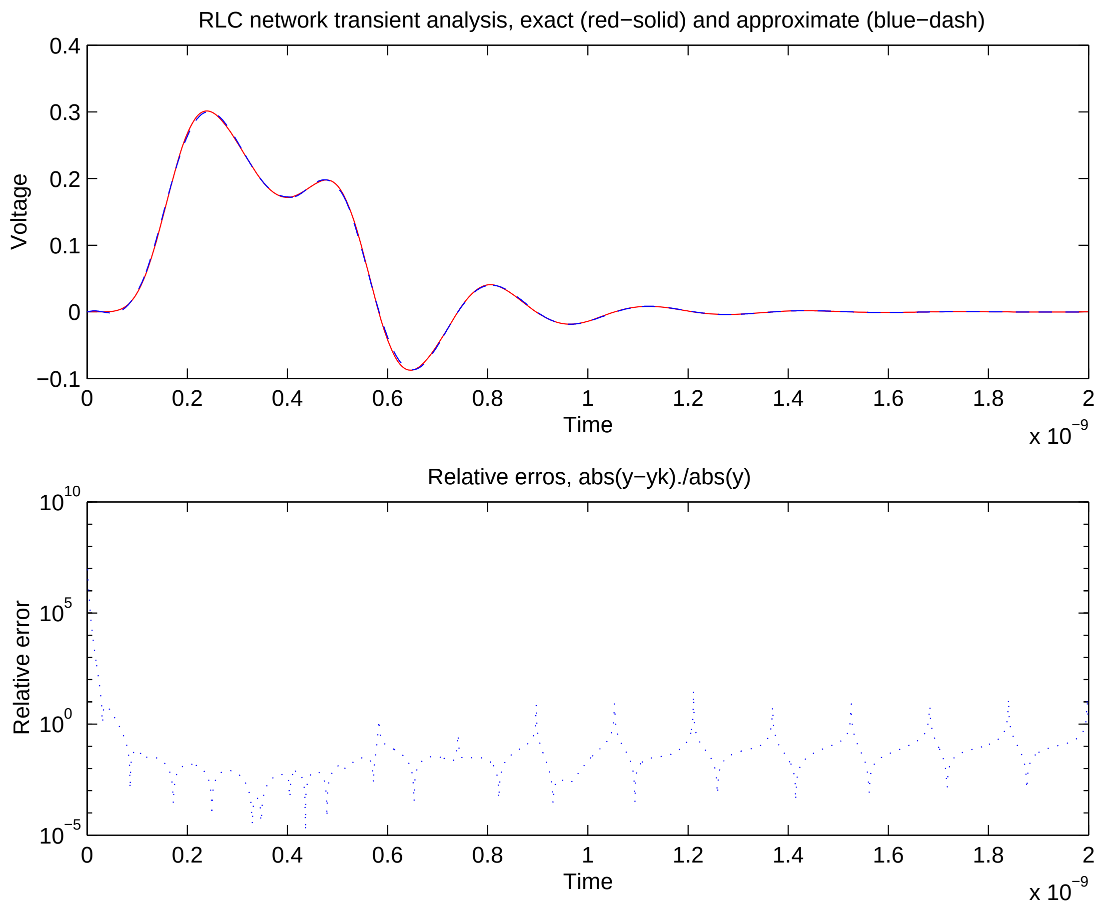
$$\mathbf{C}_n = -\mathbf{T}_n, \qquad \mathbf{G}_n\mathbf{z}(t) = \mathbf{I}_n + s_0\mathbf{T}_n$$

Back to the illustrative simple example

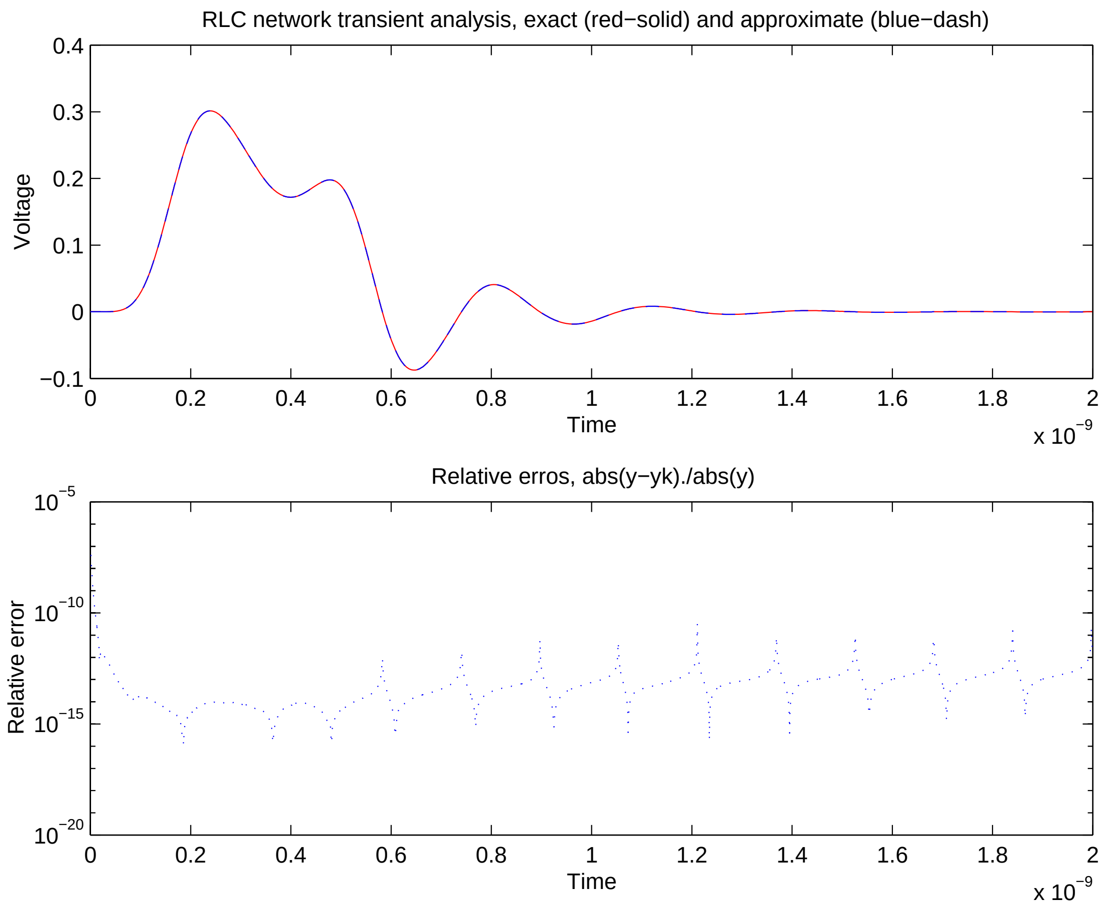
Transient responses $y(t)$, $\tilde{y}_2(t)$ and errors



Transient responses $y(t)$, $\tilde{y}_4(t)$ and errors



Transient responses $y(t)$, $\tilde{y}_6(t)$ and errors



Moments, Padé and error estimation

Objectives of error estimation:

1. Understanding convergence behaviors
2. Developing an adaptive stopping criterion
3. By-product: a simplified derivation of Padé-Lanczos connection

Tridiagonal form

Let $\mathbf{T}_n$ be the leading $n \times n$ principal submatrix of $\mathbf{T}_N$ for any $n \leq N$:

$$\mathbf{T}_N = \begin{array}{cc} & \begin{array}{cc} n & \end{array} \\ n & \left(\begin{array}{cc} \mathbf{T}_n & \\ & \ddots \end{array}\right) \end{array}$$

Then we have the following

Lemma. For any $0 \leq j \leq 2n-1$,

$$\mathbf{e}_1^{\mathrm{T}} \mathbf{T}_N^j \mathbf{e}_1 = \mathbf{e}_1^{\mathrm{T}} \mathbf{T}_n^j \mathbf{e}_1$$

and for $j = 2n$,

$$\mathbf{e}_1^{\mathrm{T}} \mathbf{T}_N^{2n} \mathbf{e}_1 = \mathbf{e}_1^{\mathrm{T}} \mathbf{T}_n^{2n} \mathbf{e}_1 + \beta_2 \beta_3 \cdots \beta_n \beta_{n+1} \cdot \rho_2 \rho_3 \cdots \rho_n \rho_{n+1}.$$

Furthermore,

$$\mathbf{e}_1^{\mathrm{T}} \mathbf{T}_n^j \mathbf{e}_n = \begin{cases} 0, & 0 \leq j < n-1 \\ \beta_2 \beta_3 \cdots \beta_n, & j = n-1 \end{cases}$$

and

$$\mathbf{e}_n^{\mathrm{T}} \mathbf{T}_n^j \mathbf{e}_1 = \begin{cases} 0, & 0 \leq j < n-1 \\ \rho_2 \rho_3 \cdots \rho_n, & j = n-1, \end{cases}$$

Ref. [Parlett'80] and [Ye'91]

Moment matching and Padé approximation

Transfer function

$$
\begin{aligned}
H(s) &= \mathbf{l}^{\mathrm{T}}(\mathbf{I}-\sigma\mathbf{A})^{-1}\mathbf{r} = \left(\mathbf{l}^{\mathrm{T}}\mathbf{r}\right)\mathbf{e}_1^{\mathrm{T}}(\mathbf{I}-\sigma\mathbf{T}_N)^{-1}\mathbf{e}_1 \\
&= \sum_{k=0}^{\infty} m_k\sigma^k
\end{aligned}
$$

The n-th reduced-order approximant

$$
\begin{aligned}
H_n(s) &= \left(\mathbf{l}^{\mathrm{T}}\mathbf{r}\right)\mathbf{e}_1^{\mathrm{T}}(\mathbf{I}-\sigma\mathbf{T}_n)^{-1}\mathbf{e}_1 \\
&= \sum_{k=0}^{\infty} \hat{m}_k\sigma^k
\end{aligned}
$$

We have

(1) By the Lemma, we have $2n$ moment-matching:

$$
\begin{aligned}
m_k = \mathbf{l}^{\mathrm{T}}\mathbf{A}^k\mathbf{r} &= (\mathbf{l}^{\mathrm{T}}\mathbf{r})\mathbf{e}_1^{\mathrm{T}}\mathbf{T}_N^k\mathbf{e}_1 \\
&= (\mathbf{l}^{\mathrm{T}}\mathbf{r})\mathbf{e}_1^{\mathrm{T}}\mathbf{T}_n^k\mathbf{e}_1 = \hat{m}_k
\end{aligned}
$$

for $k = 0, 1, \ldots, 2n-1$.

(2) By Neumann series expansion and (1), we have

$$H(s) - H_n(s) = (\mathbf{l}^{\mathrm{T}}\mathbf{r}) \left(\prod_{j=2}^{n+1} \beta_j \prod_{j=2}^{n+1} \rho_j \right) \sigma^{2n} + \mathcal{O}(\sigma^{2n+1})$$

(3) From (1) and (2), we conclude that

<u>$H_n(s)$ is a Padé approximation</u>

Forward error

Through an algebraic derivation, we have the following expression for the exact error:

$$H(s) - H_n(s) = (\mathbf{l}^{\mathrm{T}}\mathbf{r}) \left(\frac{\rho_{n+1}\eta_{n+1}}{\delta_n} \right) \left[\sigma^2 \tau_{n1}(\sigma) \tau_{1n}(\sigma) \right] \gamma_{n+1}(\sigma)$$

where

$$\tau_{1n}(\sigma) = \mathbf{e}_1^{\mathrm{T}}(\mathbf{I} - \sigma \mathbf{T}_n)^{-1}\mathbf{e}_n = \sigma^{n-1} \frac{\beta_2 \beta_3 \cdots \beta_n}{\det(\mathbf{I} - \sigma \mathbf{T}_n)}$$

$$\tau_{n1}(\sigma) = \mathbf{e}_n^{\mathrm{T}}(\mathbf{I} - \sigma \mathbf{T}_n)^{-1}\mathbf{e}_1 = \sigma^{n-1} \frac{\rho_2 \rho_3 \cdots \rho_n}{\det(\mathbf{I} - \sigma \mathbf{T}_n)}$$

and

$$\gamma_{n+1}(\sigma) = \mathbf{w}_{n+1}^{\mathrm{T}}(\mathbf{I} - \sigma \mathbf{A})^{-1}\mathbf{v}_{n+1}$$

The formulation can be simplified for symmetric transfer functions (RLC and RC systems)

Adaptive stopping criterion

1. Error bound

$$|H(s) - H_n(s)| \le |\mathbf{l}^{\mathrm{T}}\mathbf{r}| \left| \frac{\rho_{n+1}\eta_{n+1}}{\delta_n} \right| \left(\frac{\sigma^2 |\tau_{n1}(\sigma)\tau_{1n}(\sigma)|}{1 - |\sigma| \, \|A\|} \right)$$

   provided that $|\sigma| < 1/\|A\|$.

2. When "$|\sigma| < 1/\|A\|$" is violated, the error bound is not applicable (often too pessimistic). Heuristically, we observe γ_{n+1} tends to be steady,

$$\gamma_{n+1} \cong \mathbf{w}_{n+1}^{\mathrm{T}} \mathbf{A} \mathbf{v}_{n+1} \quad \text{for large } n.$$

   A plausible forward error estimation:

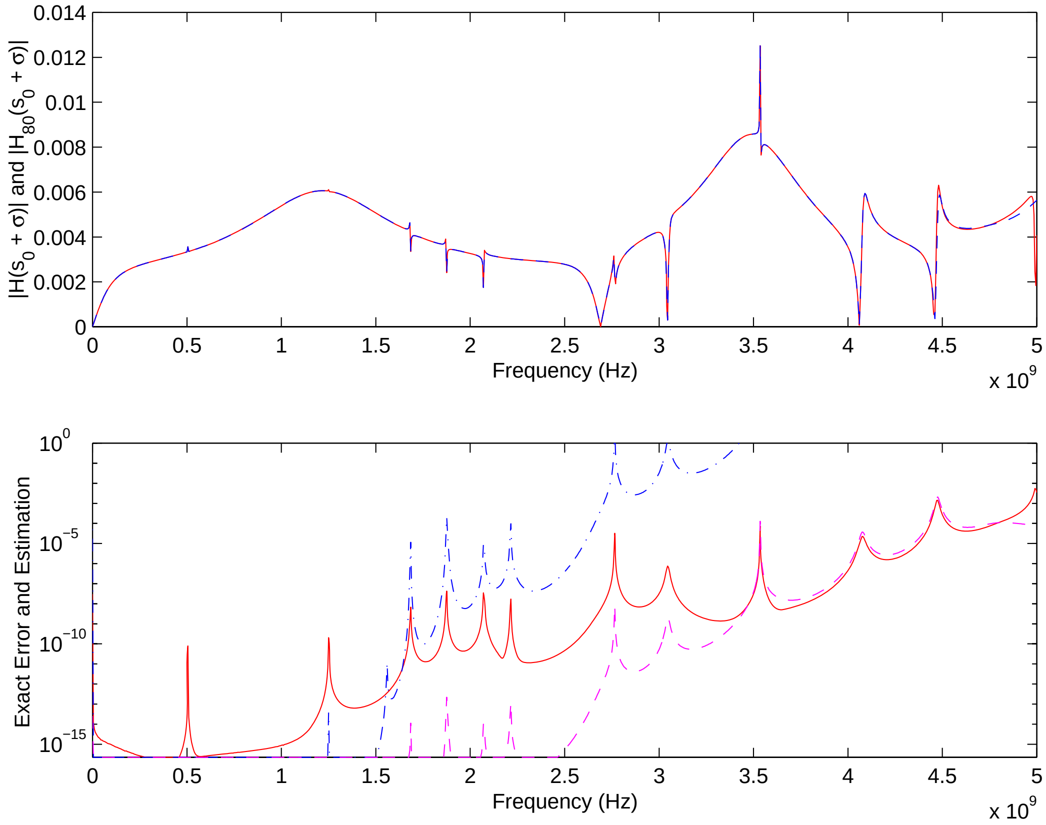
$$H(s) - H_n(s) \cong (\mathbf{l}^{\mathrm{T}}\mathbf{r}) \left(\frac{\rho_{n+1}\eta_{n+1}}{\delta_n} \right) \left(\sigma^2 \tau_{n1}\tau_{1n} \right) \left(\mathbf{w}_{n+1}^{\mathrm{T}} \mathbf{A} \mathbf{v}_{n+1} \right)$$

3. **Remaining issue:** more efficient estimation of

$$\gamma_{n+1}(\sigma) = \mathbf{w}_{n+1}^{\mathrm{T}} (\mathbf{I} - \sigma \mathbf{A})^{-1} \mathbf{v}_{n+1}$$

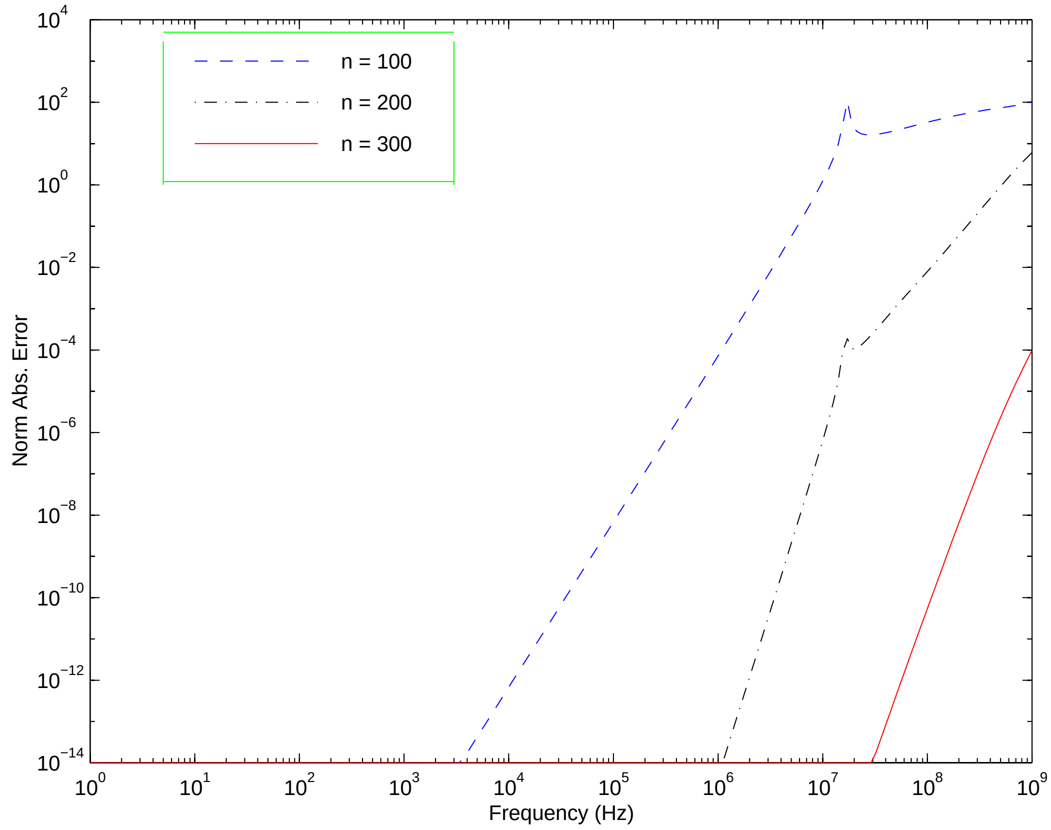
Examples

The PEEC circuit:



Examples, cont.

The “Clock” circuit ($N = 13,875$): *error bound only*

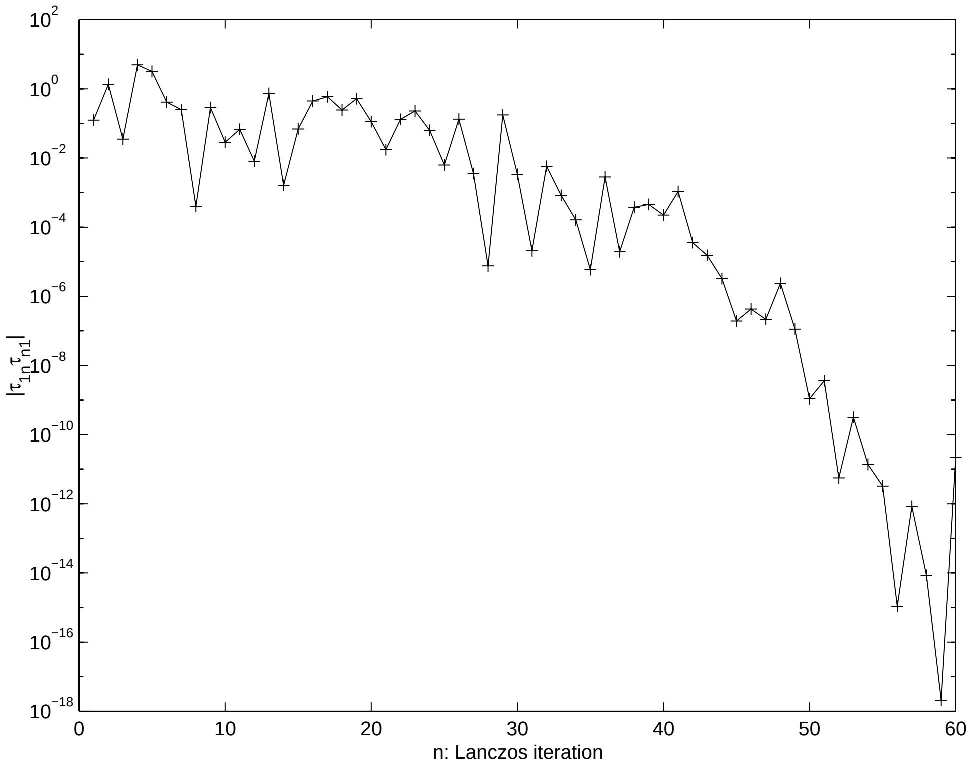


Understanding the Convergence

Major contributing factor:

$$\tau_{n1}(\sigma)\tau_{1n}(\sigma) = (\mathbf{e}_1^{\mathrm{T}}(\mathbf{I} - \sigma\mathbf{T}_n)^{-1}\mathbf{e}_n)(\mathbf{e}_n^{\mathrm{T}}(\mathbf{I} - \sigma\mathbf{T}_n)^{-1}\mathbf{e}_1)$$

The PEEC example for a fixed σ:



In fact: we can show that

$$\frac{|\tau_{n1}(\sigma)|}{\|\mathbf{e}_n^{\mathrm{T}}(\mathbf{I}-\sigma\mathbf{T}_n)^{-1}\|_2} \leq \sqrt{N} \min_{p\in\Pi_{n-1},p(0)=1} \|p(\mathbf{I}-\sigma\mathbf{A})\mathbf{r}\|_2.$$

$\Rightarrow$ The effect of distribution of eigenvalues (poles) on the rate of convergence (*Approximation Theory*).

Backward Error and Optimality

It can be shown that

$$\mathbf{l}^{\mathrm{T}}\left[\mathbf{I}-\sigma(\mathbf{A}+\mathbf{F})\right]^{-1}\mathbf{r}=(\mathbf{l}^{\mathrm{T}}\mathbf{r})\cdot\mathbf{e}_1^{\mathrm{T}}(\mathbf{I}-\sigma\mathbf{T}_n)^{-1}\mathbf{e}_1$$

for

$$\mathbf{F}=-\frac{1}{\delta_n}\begin{bmatrix}\mathbf{v}_n & \mathbf{v}_{n+1}\end{bmatrix}\begin{bmatrix}0 & \eta_{n+1}\\ \rho_{n+1} & 0\end{bmatrix}\begin{bmatrix}\mathbf{w}_n^{\mathrm{T}}\\ \mathbf{w}_{n+1}^{\mathrm{T}}\end{bmatrix}$$

Hence, one may also monitor

$$\|\mathbf{F}\|_{\mathrm{F}}^2=\rho_{n+1}^2+\eta_{n+1}^2+2\rho_{n+1}\eta_{n+1}(\mathbf{v}_{n+1}^{\mathrm{T}}\mathbf{v}_n)(\mathbf{w}_{n+1}^{\mathrm{T}}\mathbf{w}_n)$$

for the test of convergence.

However, this is often a very conservative estimator!

Open problem: optimal backward error

$$\min\left\{\|\mathbf{F}\|;\ \mathbf{l}^{\mathrm{T}}\left[\mathbf{I}-\sigma(\mathbf{A}+\mathbf{F})\right]^{-1}\mathbf{r}=(\mathbf{l}^{\mathrm{T}}\mathbf{r})\cdot\mathbf{e}_1^{\mathrm{T}}(\mathbf{I}-\sigma\mathbf{T}_n)^{-1}\mathbf{e}_1\right\}=?$$