Model Reduction of Dynamical Systems

P. Van Dooren,
ICTEAM, Université catholique de Louvain

Spring School on Matrix Functions
Lille, May 2013

Motivation

Predict a storm surge in the North sea (Verlaan-Heemink ’97)

60.000 variables, 15 inputs (buoys and radars)

Problem

Using measurements predict the state of the North Sea variables in order to operate the sluices in due time (6h.)

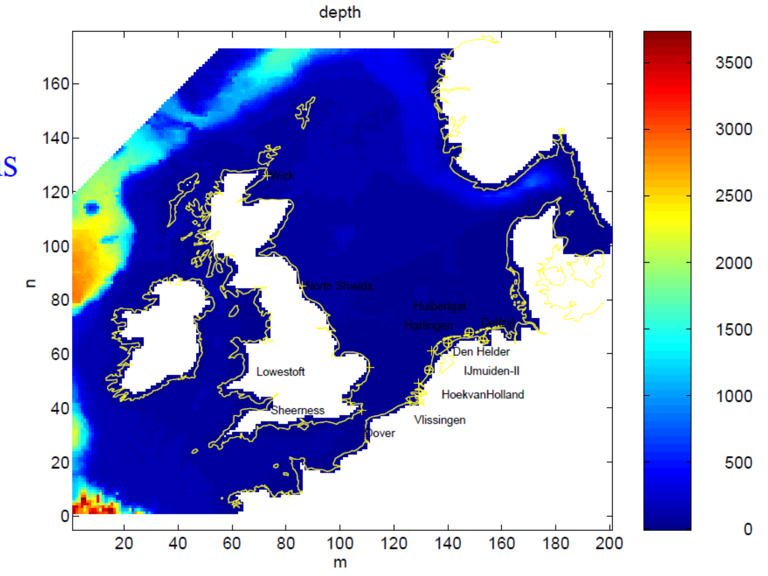
Solution

$x(t) \doteq [h(t), v_x(t), v_y(t)]$ satisfies the shallow water equations

$$\begin{aligned} \partial x(t)/\partial t &= F(x(t), w(t)) \\ y(t) &= G(x(t), v(t)) \end{aligned}$$

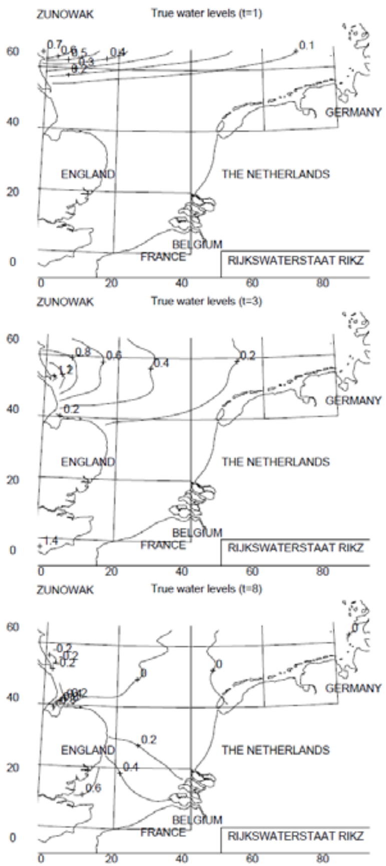
with measurements $y(t)$ and noise processes $v(.), w(.)$

$\Longrightarrow$ estimate and predict $\hat{x}(t)$ using Kalman filtering

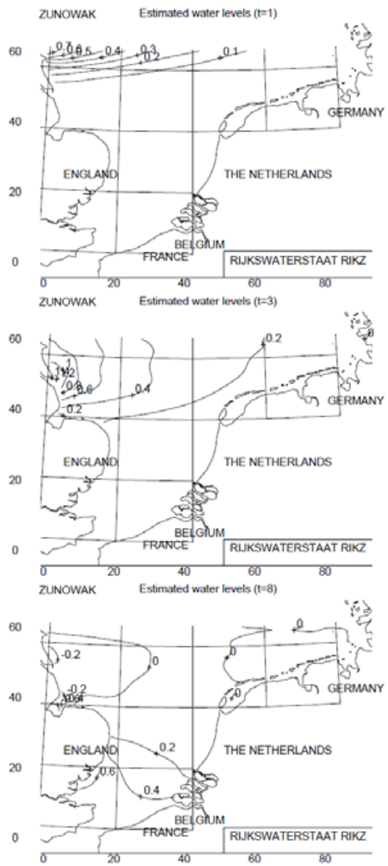


Motivation

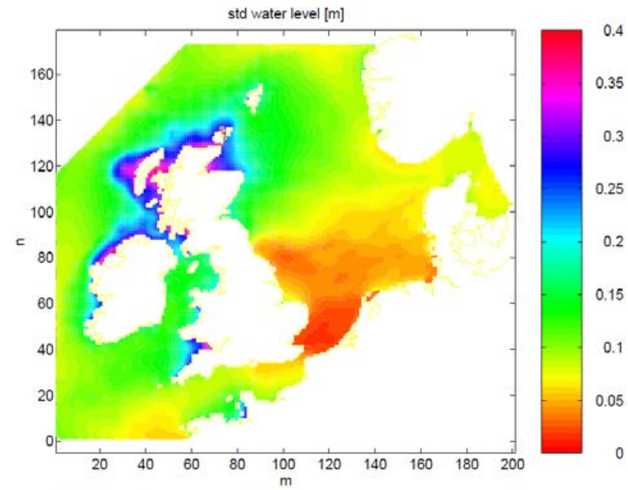
True water levels



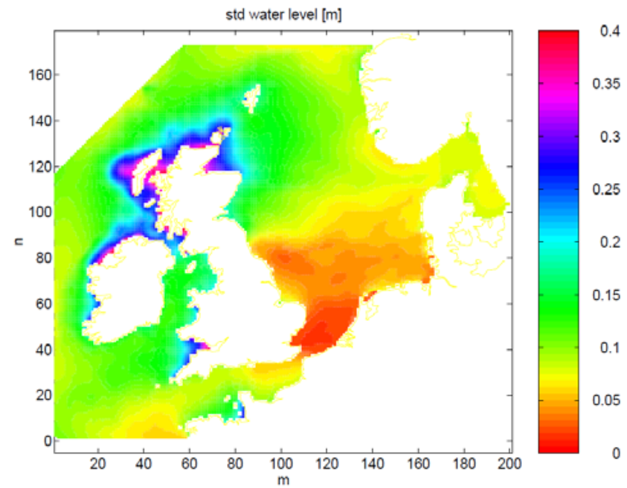
Estimated water levels



Reconstruction works well around estuarium



Standard deviation of fi lter using 8 measurement locations



Standard deviation of fi lter using measurement locations
(1 used for validation)

What models ?

General implicit dynamical systems are modeled via

$$\mathcal{S} \begin{cases} F(\dot{x}(\cdot), x(\cdot), u(\cdot)) = 0 \\ y(\cdot) = G(x(\cdot), u(\cdot)), \end{cases}$$

using "state" $x(\cdot)$ of dimension $N >> m, p.$

We focus on *explicit* state equations because derivations are simpler

What models ?

continuous-time

$$\begin{cases} \dot{x}(t) = F(x(t), u(t)) \\ y(t) = G(x(t), u(t)) \end{cases}$$

$$\Downarrow$$

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ y(t) = C(t)x(t) + D(t)u(t) \end{cases}$$

$$\Downarrow$$

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases}$$

discrete-time

$$\begin{cases} x(k+1) = F(x(k), u(k)) \\ y(k) = G(x(k), u(k)) \end{cases}$$

$$\Downarrow$$

$$\begin{cases} x(k+1) = A(k)x(k) + B(k)u(k) \\ y(k) = C(k)x(k) + D(k)u(k) \end{cases}$$

$$\Downarrow$$

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) \\ y(k) = Cx(k) + Du(k) \end{cases}$$

The strongest results are for the linear cases

Many results also extend to the implicit case

Some results extend to the time-varying case

Explicit Discrete Linear Time Invariant Systems

$$\begin{cases} x_{k+1} = Ax_k + Bu_k \\ y_k = Cx_k + Du_k, \end{cases}$$

$u_k \in \Re^m$, $y_k \in \Re^p$, $x_k \in \Re^N$, $N >> m, p$

Find another system driven with the same input $u_k \in \Re^m$

$$\begin{cases} \hat{x}_{k+1} = \hat{A}\hat{x}_k + \hat{B}u_k \\ \hat{y}_k = \hat{C}\hat{x}_k + \hat{D}u_k, \end{cases}$$

but with different $\hat{y}_k \in \Re^p$, $\hat{x}_k \in \Re^n$.

Model reduction = find a model of order $n << N$ with small output error $\|y_k - \hat{y}_k\|_{l_2}$ (nearby time reponses !)

What norm ?

Transfer functions and norms

$$H(z) = C(zI_N - A)^{-1}B + D, \quad \hat{H}(z) = \hat{C}(zI_n - \hat{A})^{-1}\hat{B} + \hat{D},$$

Take the Fourier transform of the time series (preserves energy)

$$u_f(e^{j\omega}) = \mathcal{F}\{u_k\}, \quad y_f(e^{j\omega}) = \mathcal{F}\{y_k\}, \quad \hat{y}_f(e^{j\omega}) = \mathcal{F}\{\hat{y}_k\}$$

which yields

$$y_f(e^{j\omega}) = H(e^{j\omega})u_f(e^{j\omega}), \quad \hat{y}_f(e^{j\omega}) = \hat{H}(e^{j\omega})u_f(e^{j\omega}).$$

What norm ?

... and hence a bound for $e_k \doteq [y_k - \hat{y}_k]$:

$$\mathcal{F}\{e_k\} = e_f(e^{j\omega}) = [H(e^{j\omega}) - \hat{H}(e^{j\omega})]u_f(e^{j\omega}).$$

Worst case relative error $\frac{\|e_f(e^{j\omega})\|_{l_2}}{\|u_f(e^{j\omega})\|_{l_2}} = \frac{\|y_k - \hat{y}_k\|_{l_2}}{\|u_k\|_{l_2}}$ is bounded by

$$\|H(\cdot) - \hat{H}(\cdot)\|_\infty \doteq \sup_\omega \|H(e^{j\omega}) - \hat{H}(e^{j\omega})\|_2,$$

Problem : *find the best (stable) approximation $\hat{H}(\cdot)$ of degree n for $H(\cdot)$ in $\| \, . \, \|_\infty$ norm.*

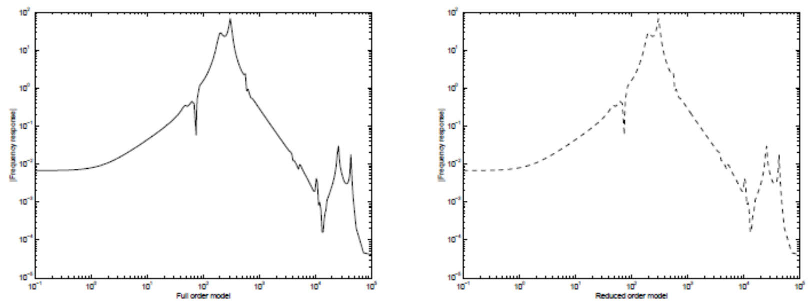
In continuous-time :

Transfer functions and norms

$$H(s) = C(sI_N - A)^{-1}B + D, \quad \hat{H}(s) = \hat{C}(sI_n - \hat{A})^{-1}\hat{B} + \hat{D},$$

are $p \times m$ rational matrices

try to match frequency responses



by minimizing their difference using

$$\|H(.) - \hat{H}(.)\|_\infty \doteq \sup_{\omega} \sigma_{max}\{H(j\omega) - \hat{H}(j\omega)\}$$

Convolution map S from inputs to outputs

Define $H(z) = H_0 + H_1 z^{-1} + H_2 z^{-2} + \ldots$ to find

$$\begin{bmatrix} \vdots \\ y_{-2} \\ y_{-1} \\ y_0 \\ y_1 \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} \ddots & & & & & \\ & H_0 & & & & \\ & H_1 & H_0 & & & \\ & H_2 & H_1 & H_0 & & \\ & H_3 & H_2 & H_1 & H_0 & \\ & & & & & \ddots \end{bmatrix}}_{\mathcal{S}} \begin{bmatrix} \vdots \\ u_{-2} \\ u_{-1} \\ u_0 \\ u_1 \\ \vdots \end{bmatrix}$$

$$\|H(\cdot) - \hat{H}(\cdot)\|_\infty = \|\mathcal{S} - \hat{\mathcal{S}}\|_2,$$

$\mathcal{S}$ does not have a discrete set of singular values !

Hankel map *H* : past inputs to future outputs

Put $u_0, u_1, u_2 ... = 0$ to see the constrained map

$$
\left[\begin{array}{c} \vdots \\ y_{-2} \\ y_{-1} \\ \hline y_0 \\ y_1 \\ \vdots \end{array}\right]
=
\left[\begin{array}{ccc|ccc}
\ddots & & & & & \\
 & H_0 & & & & \\
 & H_1 & H_0 & & & \\ \hline
 & H_2 & H_1 & H_0 & & \\
 & H_3 & H_2 & H_1 & H_0 & \\
 & & & & & \ddots
\end{array}\right]
\left[\begin{array}{c} \vdots \\ u_{-2} \\ u_{-1} \\ \hline 0 \\ 0 \\ \vdots \end{array}\right]
$$

$$\|\mathcal{S} - \hat{\mathcal{S}}\|_{\mathcal{H}} = \|\mathcal{H} - \hat{\mathcal{H}}\|_2,$$

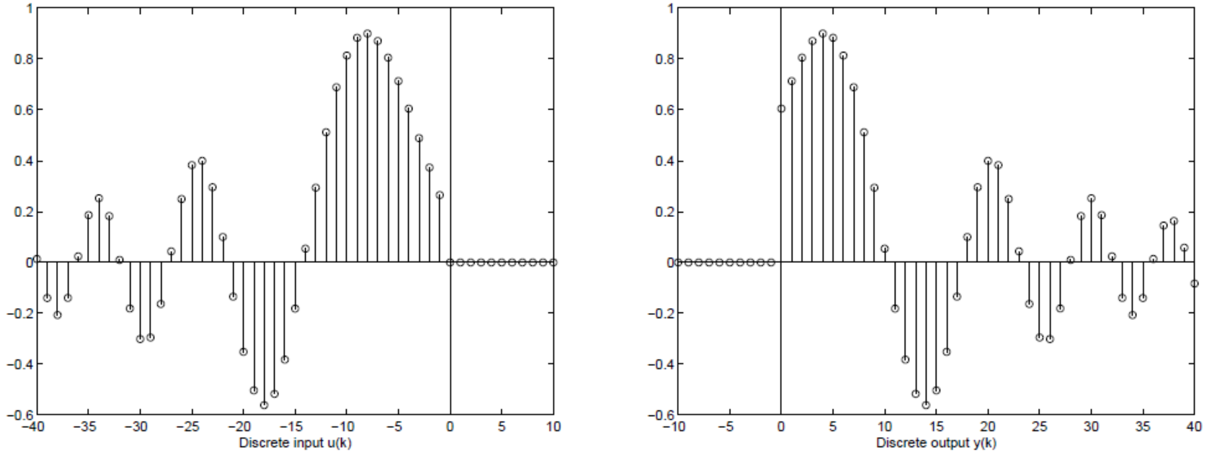
$\mathcal{H}$ has rank N and has discrete singular values !

Hankel map *H* : past inputs to future outputs

Assume $H_0 \doteq D = 0$ and use identities $H_k \doteq CA^{k-1}B$

$$y_k = \sum_{-\infty}^{0} CA^{(k-j)}Bu(j-1) = CA^k \cdot \sum_{0}^{\infty} A^j Bu(-j-1),$$

$$y_k = CA^k x(0), \quad x(0) = \sum_{0}^{\infty} A^j Bu(-j-1).$$



$$u_k,\ k \in (-\infty, 0) \Longrightarrow x(0) \Longrightarrow y_k,\ k \in [0, \infty)$$

Hankel map factorization

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix}}_{\mathcal{O}} \underbrace{\underbrace{\begin{bmatrix} B & AB & A^2B & \dots \end{bmatrix}}_{\mathcal{C}} \begin{bmatrix} u_{-1} \\ u_{-2} \\ u_{-3} \\ \vdots \end{bmatrix}}_{x_0}$$

Factorization $\mathcal{H} = \mathcal{O}\mathcal{C}$ is not unique, and $\mathcal{H} = (\mathcal{O}T)(T^{-1}\mathcal{C})$ corresponds to the transformed system $\{T^{-1}AT, T^{-1}B, CT\}$

Gramians derived from the Hankel map

Define the dual maps

$$\mathcal{O}^* : y([0,\infty)) \mapsto x(0), \quad \mathcal{C}^* : x(0) \mapsto u((-\infty,0))$$

and the (observability and controllability) *Gramians*

$$G_o \doteq \mathcal{O}^*\mathcal{O}, \quad G_c \doteq \mathcal{C}\mathcal{C}^*$$

$$G_o = \sum_0^{+\infty}(CA^k)^T(CA^k), \quad G_c = \sum_0^{+\infty}(A^kB)(A^kB)^T,$$

solved by

$$A^TG_oA - G_o + C^TC = 0 \quad \text{and} \quad AG_cA^T - G_c + BB^T = 0.$$

Since $\mathcal{H} = \mathcal{O}\mathcal{C}$ it follows that

$$\mathcal{H}^*\mathcal{H} = \mathcal{C}^*\mathcal{O}^*\mathcal{O}\mathcal{C}, \quad \text{and} \quad G_c G_o \doteq \mathcal{C}\mathcal{C}^*\mathcal{O}^*\mathcal{O}$$

have the same nonzero eigenvalues $\Rightarrow \sigma_i^2(\mathcal{H}) = \lambda_i(G_c G_o)$.

Balancing transformation T

$$\hat{G}_o \doteq T^T G_o T, \quad \hat{G}_c \doteq T^{-1} G_c T^{-T},$$

$$\left\{\hat{A}, \hat{B}, \hat{C}\right\} \doteq \{T^{-1}AT, T^{-1}B, CT\} \stackrel{T}{\Longleftarrow} \{A, B, C\}$$

diagonalizes both Gramians and makes them equal :

$$\hat{\mathcal{O}} = \mathcal{O}T, \quad \hat{\mathcal{C}} = T^{-1}\mathcal{C}, \quad \hat{G}_o = \hat{\mathcal{O}}^*\hat{\mathcal{O}} = \Sigma = \hat{\mathcal{C}}\hat{\mathcal{C}}^* = \hat{G}_c$$

H_∞ and Hankel norms are often close to each other $\Rightarrow$

New problem : *find the best (stable) approximation* $\hat{H}(\cdot)$ *of degree* n *for* $H(\cdot)$ *in the Hankel norm.*

Approximation via balanced truncation

Partition a balanced system as follows

$$\Sigma \doteq \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}, \quad \sigma_{min}(\Sigma_{11}) >> \sigma_{max}(\Sigma_{22}),$$

and define

$$\hat{A} \doteq \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix}, \hat{B} \doteq \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix}, \hat{C} \doteq \begin{bmatrix} \hat{C}_1 & \hat{C}_2 \end{bmatrix}.$$

Define also

$$\hat{Y}^T \doteq \begin{bmatrix} I_n & 0 \end{bmatrix}, \quad \hat{X} \doteq \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \quad \hat{P} \doteq \hat{X}\hat{Y}^T,$$

Then "truncated" system can be written as

$$\left\{\hat{A}_{11}, \hat{B}_1, \hat{C}_1\right\} = \left\{\hat{Y}^T\hat{A}\hat{X}, \hat{Y}^T\hat{B}, \hat{C}\hat{X}\right\} \tag{1}$$

Since

$$\hat{\mathcal{O}} = \left[\begin{array}{cc} U_1\Sigma_{11}^{\frac{1}{2}} & U_2\Sigma_{22}^{\frac{1}{2}} \end{array}\right], \; \hat{\mathcal{C}} = \left[\begin{array}{c} \Sigma_{11}^{\frac{1}{2}}V_1^T \\ \Sigma_{22}^{\frac{1}{2}}V_2^T \end{array}\right],$$

then $\hat{\mathcal{O}}\hat{P}\hat{\mathcal{C}} \doteq U_1\Sigma_{11}V_1^T$ is an "optimal" rank n approximation of $\mathcal{H}$:

$$\|\mathcal{H} - \hat{\mathcal{O}}\hat{P}\hat{\mathcal{C}}\|_2 = \|\hat{\mathcal{O}}\hat{\mathcal{C}} - \hat{\mathcal{O}}\hat{P}\hat{\mathcal{C}}\|_2 = \|U_2\Sigma_{22}V_2^T\|_2 = \sigma_{n+1}$$

but it is not Hankel !

If $\sigma_n >> \sigma_{n+1}$ then (1) has a Hankel map $\hat{\mathcal{H}}$ close to $\hat{\mathcal{O}}\hat{P}\hat{\mathcal{C}}$ and

$$\sigma_{n+1}^2 \leq \|\mathcal{H} - \hat{\mathcal{H}}\|_2^2 \leq \sum_{i=n+1}^{N} \sigma_i^2$$

Balanced truncation is thus "near optimal" and stable !

In the original system we can also write

$$\left\{\hat{A}_{11}, \hat{B}_1, \hat{C}_1\right\} = \{Y^T AX, Y^T B, CX\}$$

where the projector P is given by $P = XY^T$:

$$Y^T X = I_n, \quad Y^T G_c G_o X = \Sigma^2$$

Also X and Y are "dominant" eigenspaces of $G_c G_o$

Numerical procedure

Solve for the Gramians from

$$A^T G_o A - G_o + C^T C = 0 \quad \text{and} \quad A G_c A^T - G_c + B B^T = 0,$$

then find "dominant spaces" X and Y such that

$$Y^T X = I_n, \quad G_c G_o X = X \Sigma^2, \quad Y^T G_c G_o = \Sigma^2 Y^T$$

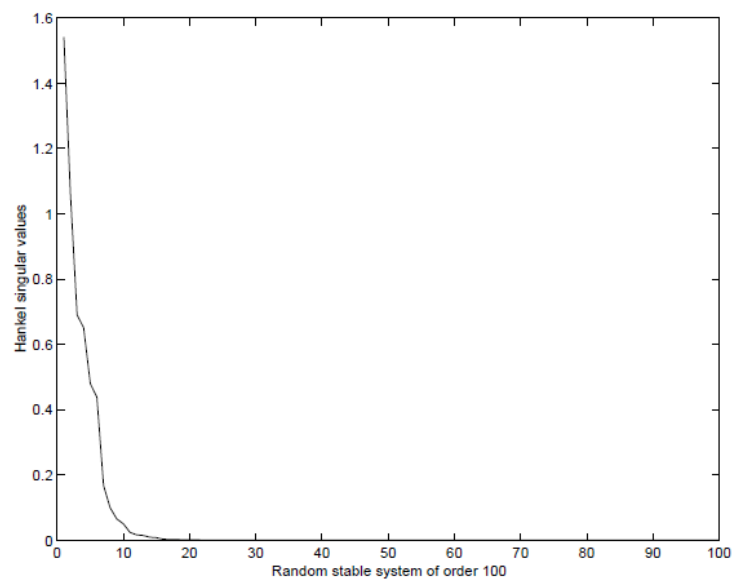
Notice that $\quad U_1 \Sigma_{11}^{\frac{1}{2}} = \mathcal{O} X, \quad Y^T \mathcal{C} = \Sigma_{11}^{\frac{1}{2}} V_1^T$

Problem : behaviour of Hankel singular values

It is known that for $\mathcal{H} > 0$ $\frac{\sigma_1(\mathcal{H})}{\sigma_n(\mathcal{H})} \approx 4^n$!

But for all pass transfer function $\Sigma = I_N$!

Typical behaviour for stable system :



One has to be careful with numerical calculation of Gramians (ill conditioning) and construction of projectors

Square root approach

From

$$G_c = S^T S, \quad G_o = R^T R$$

compute the singular value decomposition

$$SR^T = \left[\begin{array}{c|c} U_1 & U_2 \end{array}\right] \left[\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & \Sigma_2 \end{array}\right] \left[\begin{array}{c|c} V_1 & V_2 \end{array}\right]^T$$

where

$$\Sigma_1 = \mathrm{diag}\{\sigma_1, \ldots, \sigma_n\}, \quad \Sigma_2 = \mathrm{diag}\{\sigma_{n+1}, \ldots, \sigma_N\},$$

Then define

$$Y = R^T V_1 \Sigma_1^{-\frac{1}{2}}, \quad X = S^T U_1 \Sigma_1^{-\frac{1}{2}},$$

This implies

$$Y^T X = \Sigma_1^{-\frac{1}{2}} V_1^T R S^T U_1 \Sigma_1^{-\frac{1}{2}} = I_n,$$

$$Y^T G_c G_o X = Y^T S^T S R^T R X = \Sigma_1^{\frac{1}{2}} U_1^T S R^T V_1 \Sigma_1^{\frac{1}{2}} = \Sigma_1^2,$$

It follows also that the singular values σ_i of SR^T are the (nonzero) Hankel singular values.

The Gramians $G_c = S^T S$ and $G_o = R^T R$ are not needed to construct the projector $P = Y^T X$, but only the factors S and R !

Other bases for X and Y can be chosen without diagonalizing

$$Y^T G_c G_o X$$

[Varga] computes orthonormal bases via the QR decompositions :

$$S^T U_1 = XW, \quad X^T X = I_n, \quad W \text{ upper triangular}$$

$$R^T V_1 = YZ, \quad Y^T Y = I_n, \quad Z \text{ upper triangular.}$$

Reduced order model is then obtained as

$$\left\{\hat{A}_{11}, \hat{B}_1, \hat{C}_1\right\} = \{(Y^T X)^{-1} Y^T AX, (Y^T X)^{-1} Y^T B, CX\}$$

Dense Stein solvers (exact)

Bartels-Stewart/Hammarling solver ($O(N^3)$ flops)

1. Compute upper triangular Schur form $A_u \doteq U^T A U$ and put $G_u \doteq U^T G_o U$, $C_u \doteq CU$

2. Put $G_u \doteq R^T R$ and solve for columns of R in

$$A_u^T R^T R A_u - R^T R = C_u^T C_u$$

3. Return to original coordinates $G_o = U R^T R U^T$ if needed

Advantages

better conditioned $\kappa(R) = \sqrt{\kappa(G_o)}$
back transformation is not always needed
works for both Gramians in continuous and discrete time

Dense Stein solvers (approximate)

The Stein equation

$$A^T X A + M = X$$

is a special Riccati equation solved via the Disc function.

The corresponding recursive algorithm is the Smith iteration

$$\boxed{\begin{array}{l} A_0 \leftarrow A, \quad M_0 \leftarrow M, \quad \text{for} \;\; k = 0, 1, 2, \ldots \\ A_{k+1} \leftarrow A_k^2, \quad M_{k+1} \leftarrow M_k + A_k^T M_k A_k \end{array}}$$

Taking the Cholesky factorization $M_k = R_k^T . R_k$ one obtains the next factor R_{k+1} from a QR factorization [Benner et al.] :

$$\left[\begin{array}{c} R_k \\ R_k A_k \end{array} \right] \rightarrow Q_{k+1} . R_{k+1} .$$

Interpolation approach (continuous-time)

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \qquad H(s) = C(sI_N - A)^{-1}B + D$$

$u_k \in \Re^m$, $y_k \in \Re^p$, $x_k \in \Re^N$, $N >> m, p$

Find another system driven with the same input $u_k \in \Re^m$

$$\begin{cases} \dot{\hat{x}}(t) = \hat{A}\hat{x}(t) + \hat{B}u(t) \\ \hat{y}(t) = \hat{C}\hat{x}(t) + \hat{D}u(t) \end{cases} \qquad \hat{H}(s) = \hat{C}(sI_n - \hat{A})^{-1}\hat{B} + \hat{D}$$

but with different $\hat{y}(t) \in \Re^p$, $\hat{x}(t) \in \Re^n$.

Problem : *find the best (stable) approximation* $\hat{H}(\cdot)$ *of degree* n *for* $H(\cdot)$ *in* $\|\ .\ \|_\infty$ *norm.*

Should also approximate Gramians

Interpolating $H(\omega)$ is a good idea because of Parseval's theorem

$$G_o = \frac{1}{2\pi}\int_{-\infty}^{+\infty} (-j\omega I - A^T)^{-1} C^T C (j\omega I - A)^{-1} d\omega,$$

$$G_c = \frac{1}{2\pi}\int_{-\infty}^{+\infty} (j\omega I - A)^{-1} B B^T (-j\omega I - A^T)^{-1} d\omega$$

or fitting the exponential since

$$G_o = \int_0^{+\infty} e^{A^T t} C^T C e^{At} dt, \quad G_c = \int_0^{+\infty} e^{At} B B^T e^{A^T t} dt$$

or approximate Lyapunov solvers (ADI)

$$A^T G_o + G_o A = C^T C, \quad A G_c + G_c A^T = B B^T$$

Krylov subspaces

What technique to use ? The discrete-time case :

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \end{bmatrix} = \mathcal{O}\mathcal{C} \begin{bmatrix} u_{-1} \\ u_{-2} \\ u_{-3} \\ \vdots \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \begin{bmatrix} B & AB & A^2B & \dots \end{bmatrix} \begin{bmatrix} u_{-1} \\ u_{-2} \\ u_{-3} \\ \vdots \end{bmatrix}$$

... suggests to use Krylov sequences !

$$\mathcal{K}_j(M,R) = \mathrm{Im}\left\{R, MR, M^2R, \dots, M^{j-1}R\right\}$$

Rational interpolation and moment matching

Let X and Y define a projector ($Y^TX = I_n$) and

$$\left\{\hat{A}, \hat{B}, \hat{C}, D\right\} = \{Y^TAX, Y^TB, CX, D\}$$

Taylor series of $H(s) \doteq C(sI - A)^{-1}B + D$ around $s = \infty$

$$H(s) = H_0 + H_1 s^{-1} + H_2 s^{-2} + \cdots ,$$

where the moments H_i are equal to :

$$H_0 = D, \quad H_i = CA^{i-1}B, \; i = 1, 2, \ldots$$

The reduced order model $\hat{H}(s) \doteq \hat{C}(sI - \hat{A})^{-1}\hat{B} + \hat{D}$

Rational interpolation and moment matching

... has a similar expansion

$$\hat{H}(s) = \hat{H}_0 + \hat{H}_1 s^{-1} + \hat{H}_2 s^{-2} + \cdots ,$$

with moments $\hat{H}_i$:

$$\hat{H}_0 = \hat{D}, \quad \hat{H}_i = \hat{C}\hat{A}^{i-1}\hat{B}, \; i = 1, 2, ...$$

Theorem : Let $m = p$, $Y^T X = I_n$ and assume

$$\mathrm{Im} X = \mathrm{Im}\left[B, AB, A^2 B, \ldots A^{k-1} B\right],$$

$$\mathrm{Im} Y = \mathrm{Im}\left[C^T, A^T C^T, A^{2T} C^T, \ldots A^{(k-1)T} C^T\right]$$

then the first $2k$ moments match :

$$H_j = \hat{H}_j, \quad j = 1, \ldots, 2k.$$

Rational interpolation and moment matching

Lanczos algorithm and Padé approximation

One shows for $p = m = 1$ that $\hat{H}(s)$ is the (Laurent)-Padé approximation to $H(s)$ since the first $2k$ moments are matched.

Is the basis of the Asymptotic Waveform Expansion (AWE) :
1) compute the moments H_i
2) construct the Padé approximation $\hat{H}(s)$

A more reliable method to construct the reduced order model $\hat{H}(s)$ is to construct directly a realization $\left\{\hat{A}, \hat{b}, \hat{c}, d\right\}$ via Lanczos

Rational interpolation and moment matching

Lanczos is numerically better than Padé

Example : stiff RC ladder circuit $C_1 = 10^{-3}$, $C_2 = 10^{-6}$, $C_3 = 10^{-9}$

$$\left[\begin{array}{c|c} A & b \\ \hline c & d \end{array}\right] = \left[\begin{array}{ccc|c} -2C_1^{-1} & C_1^{-1} & 0 & C_1^{-1} \\ C_2^{-1} & -2C_2^{-1} & C_2^{-1} & 0 \\ 0 & C_3^{-1} & -C_3^{-1} & 0 \\ \hline 1 & -1 & 0 & 0 \end{array}\right]$$

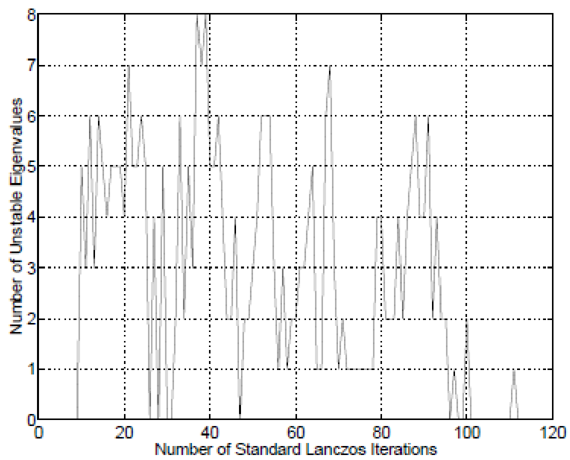
Reconstructed eigenvalues of A from full order model

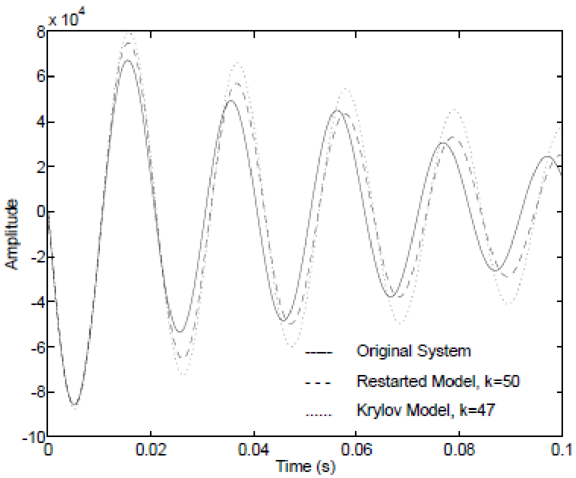
	eig 1	eig 2	eig 3
Exact	-9.98999000e2	-1.00000100e6	-1.00100100e9
Moment Match	-9.98999000e2	-1.00000078e6	-5.45486876e6
Lanczos	-9.98999000e2	-1.00000100e6	-1.00100100e9

Rational interpolation and moment matching

Interpolated models may be unstable !

Unstable eigenvalues of Padé approximants of CD-player





Few model orders are stable (lightly damped system)

Solution (?): Implicitly Restarted Lanczos fits modified moments

$$C\phi(A)A^{i-1}B = \hat{C}\phi(\hat{A})\hat{A}^{i-1}\hat{B}, \quad i = 1, \ldots, 2k$$

Rational interpolation and moment matching

Extension to other bases

Any basis X, Y for the same (Krylov) spaces yields the **same** transfer function $\hat{H}(s)$ provided the reduced model is defined as

$$\left\{\hat{A}, \hat{B}, \hat{C}, D\right\} = \{(Y^T X)^{-1} Y^T A X, (Y^T X)^{-1} Y^T B, C X, D\}$$

Use two Arnoldi processes instead of one unsymmetric Lanczos

Block Arnoldi process

An orthogonal basis V for the (block) Krylov subspace

$$\mathcal{K}_\ell(A, B) = \operatorname{Im}\left\{B, AB, A^2 B, \ldots, A^{\ell-1} B\right\}$$

Better numerical properties but slower

Rational interpolation and moment matching

Need for multi-point approximations

We can move the approximation problem to another point

$$H(s) = H_0 + H_1(s-\sigma)^1 + H_2(s-\sigma)^2 + \cdots ,$$

where the **moments** H_i are equal to :

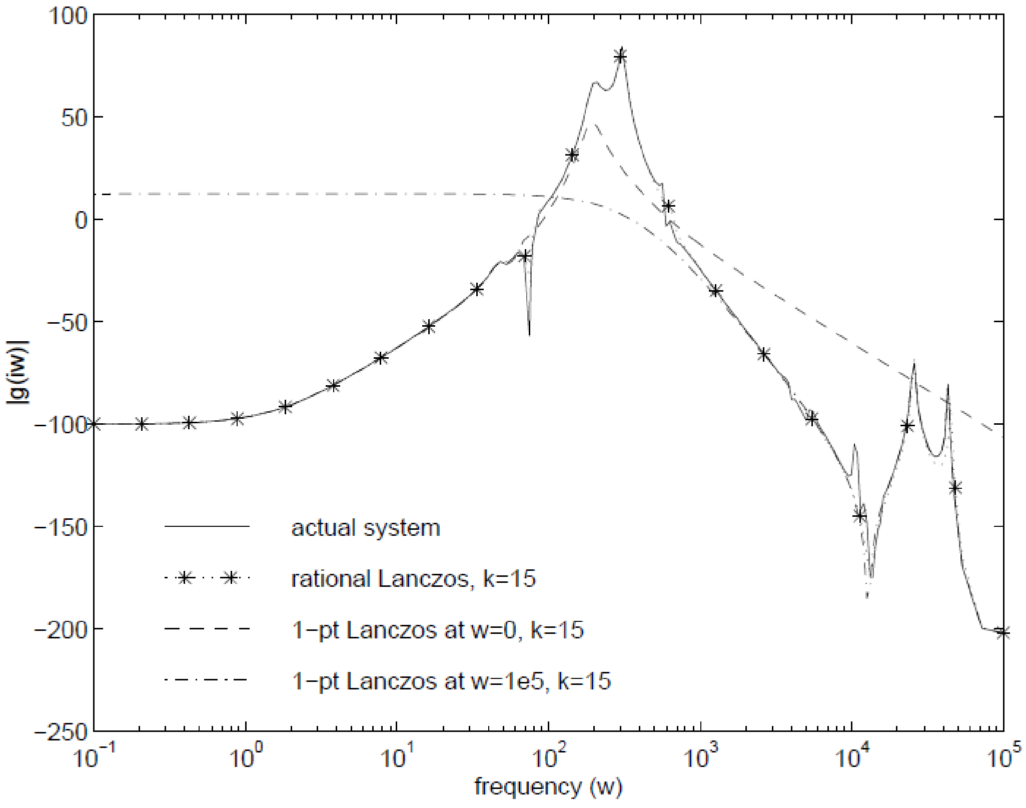
$$H_0 \doteq H(\sigma) = D - C(A - \sigma I)^{-1}B,$$

$$H_i \doteq \frac{1}{i!}\frac{\partial^i H(s)}{\partial s^i}\bigg|_{s=\sigma} = -C(A-\sigma I)^{-(i+1)}B,\ i > 0.$$

Interpolate $H(s)$ and its derivatives at σ can be transformed to the first problem.

Rational interpolation and moment matching

Yet



indicates that multi-point approximations are better

General theorem (state-space)

Consider multiple points $\{\sigma^{(1)}, \ldots, \sigma^{(K)}\}$ and the moments

$$H_i^{(k)} = -C(A - \sigma^{(k)} I)^{-(i-1)} B,$$

of the original model, and

$$\hat{H}_i^{(k)} = -\hat{C}(\hat{A} - \sigma^{(k)} I)^{-(i-1)} \hat{B}.$$

of the reduced order model.

Moment matching at the interpolation points $\sigma^{(k)}, k = 1, \ldots, K$ is completely described by the following theorem :

(Notice that **inclusion** is sufficient !)

Theorem : If

$$\bigcup_{k=1}^{K} \mathcal{K}_{J_{b_k}}\left((A-\sigma^{(k)}I)^{-1},(A-\sigma^{(k)}I)^{-1}B\right) \subseteq \mathrm{Im}X$$

and

$$\bigcup_{k=1}^{K} \mathcal{K}_{J_{c_k}}\left((A-\sigma^{(k)}I)^{-T},(A-\sigma^{(k)}I)^{-T}C^T\right) \subseteq \mathrm{Im}Y$$

then the moments $H_i^{(k)}$ of the system $\{A,B,C,D\}$ and the corresponding moments $\hat{H}_i^{(k)}$ of the reduced order model

$$\{(Y^TX)^{-1}Y^TAX,(Y^TX)^{-1}Y^TB,CX,D\}$$

match up to $J_{(k)} \doteq J_{b_k}+J_{c_k}$:

$$H_i^{(j_k)}=\hat{H}_i^{(j_k)}, \quad j_k=1,2,\ldots,J_{(k)}, \quad k=1,2,\ldots K$$

provided Y^TX is invertible.

Implicit continuous LTI systems

$$\begin{cases} E\dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \qquad H(s) = C(sE - A)^{-1}B$$

$u_k \in \Re^m,\ y_k \in \Re^p,\ x_k \in \Re^N,\ N >> m, p$

Find another system driven with the same input $u_k \in \Re^m$

$$\begin{cases} \hat{E}\dot{\hat{x}}(t) = \hat{A}\hat{x}(t) + \hat{B}u(t) \\ \hat{y}(t) = \hat{C}\hat{x}(t) \end{cases} \qquad \hat{H}(s) = \hat{C}(s\hat{E} - \hat{A})^{-1}\hat{B}$$

but with different $\hat{y}(t) \in \Re^p,\ \hat{x}(t) \in \Re^n$.

The nth order model is defined from $X, Y \in \Re^{N\times n}$:

$$\{\hat{E}, \hat{A}, \hat{B}, \hat{C}\} \doteq \{Y^T EX, Y^T AX, Y^T B, CX\}.$$

The transfer function moments

$$\left.\frac{\partial^i H(s)}{\partial s^i}\right|_{s=\sigma} \doteq i! H_i \ ,$$

yield an expansion around each interpolation point σ

$$H_i = -C(A-\sigma E)^{-1}\{E(A-\sigma E)^{-1}\}^i B,$$

$$H_i = -C\{(A-\sigma E)^{-1}E\}^i (A-\sigma E)^{-1} B,$$

on which we will base the Krylov spaces

$$\mathcal{K}_j(M,G) = \mathrm{Im}\,\{G, MG, M^2G, \ldots, M^{j-1}G\}$$

Generalized moment condition :

Theorem : If

$$\bigcup_{k=1}^{K} \mathcal{K}_{J_{b_k}} \left((A - \sigma^{(k)} E)^{-1} E, (A - \sigma^{(k)} E)^{-1} B \right) \subseteq \mathrm{Im} X$$

and

$$\bigcup_{k=1}^{K} \mathcal{K}_{J_{c_k}} \left((A - \sigma^{(k)} E)^{-T} E^T, (A - \sigma^{(k)} E)^{-T} C^T \right) \subseteq \mathrm{Im} Y$$

then the moments $H_i^{(k)}$ of the system $\{E, A, B, C\}$ and the corresponding moments $\hat{H}_i^{(k)}$ of the reduced order model

$$\{Y^T E X, Y^T A X, Y^T B, C X\}$$

match up to $J_{(k)} \doteq J_{b_k} + J_{c_k}$:

$$H_i^{(j_k)} = \hat{H}_i^{(j_k)}, \quad j_k = 1, 2, \ldots, J_{(k)}, \quad k = 1, 2, \ldots K$$

provided the pencil $sY^T E X - Y^T A X$ is regular.

This includes modal matching

Theorem Let $\mathrm{Im}X_i$, $\mathrm{Im}Y_i$ be left and right invariant subspaces of the regular pencil $(\lambda E - A)$ with given spectrum, then the reduced order pencil $(\lambda \hat{E} - \hat{A}) \doteq Y^T(\lambda E - A)V$ has a right invariant subspace with the same spectrum if

$$\mathrm{Im}X_i \subseteq \mathrm{Im}X$$

and has a left invariant subspace with the same spectrum if

$$\mathrm{Im}Y_i \subseteq \mathrm{Im}Y.$$

This extension allows to incorporate matching poles of the original system into the reduced order system.

Importance :

Approximation of DAE requires often that the reduced order system has the same algebraic conditions (these corresponds to an invariant subspace at ∞

Tangential interpolation

Consider the MIMO error function $E(s) := H(s) - \hat{H}(s)$

Standard interpolation at $s = \sigma$ implies $E(\sigma) = 0$

Tangential interpolation conditions at $s = \sigma$ are less restrictive :

$$E(\sigma)v(\sigma) = 0, \quad w(\sigma)E(\sigma) = 0, \quad w(\sigma)E(\sigma)v(\sigma),$$

where $w(s)$ and $v(s)$ are (polynomial) row and column vectors

There also exist higher order conditions :

$$E(s)v(s) = 0(s-\sigma)^k, \; w(s)E(s) = 0(s-\sigma)^k, \; w(s)E(s)v(s) = 0(s-\sigma)^k$$

which are obtained via generalized Krylov methods or via Sylvester equations

Tangential interpolation

The interpolation condition

$$\left[C(sI_N - A)^{-1}B - \hat{C}(sI_n - \hat{A})^{-1}\hat{B}\right] v(s) = O(s-\sigma)^k$$

is obtained via the generalized Krylov space inclusion

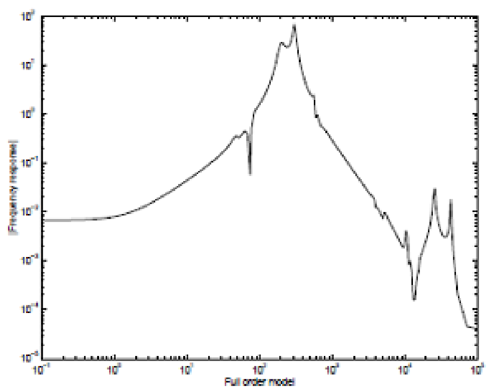
$$[(\sigma I_N - A)^{-1}B \ldots (\sigma I_N - A)^{-k}B] \begin{bmatrix} v_0 & \ldots & v_{k-1} \\ & \ddots & \vdots \\ & & v_0 \end{bmatrix} \subseteq \operatorname{Im}(X)$$

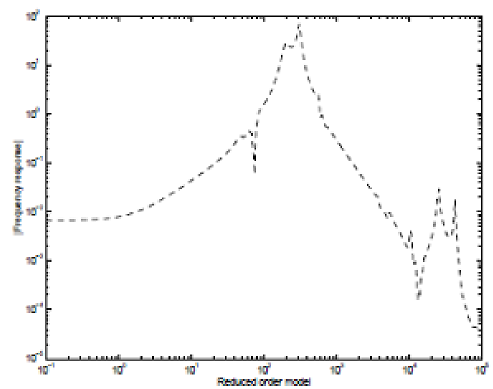
or by solving the Sylvester equation $AX - X\tilde{A} = B\tilde{B}$ with

$$\tilde{A} = \sigma I_k - \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}, \tilde{B} = \begin{bmatrix} v_0 & \ldots & v_{k-1} \end{bmatrix}$$

H_2 optimal approximations

Minimizing the cost $\mathcal{J} := \|E(z)\|_{\mathcal{H}_2} := tr \int_{-\infty}^{\infty} E(e^{j\omega})E(e^{j\omega})^H \frac{d\omega}{2\pi}$ ensures the frequency response to match





and the time responses to match if $H(z)$ and $\hat{H}(z)$ are stable since

$$\mathcal{J} = \mathrm{tr} \sum_{k=0}^{\infty} (C_e A_e^k B_e)(C_e A_e^k B_e)^T$$

with

$$(A_e, B_e, C_e) := \left(\begin{bmatrix} A & \\ & \hat{A} \end{bmatrix}, \begin{bmatrix} B \\ \hat{B} \end{bmatrix}, \begin{bmatrix} C & -\hat{C} \end{bmatrix} \right),$$

How evaluate this norm ?

$$\mathcal{J} = \operatorname{tr}\left(C_e P_e C_e^T\right) = \operatorname{tr}\left(B_e^T Q_e B_e\right)$$

where P_e and Q_e solve the Stein equations

$$A_e P_e A_e^T + B_e B_e^T = P_e, \quad A_e^T Q_e A_e + C_e^T C_e = Q_e$$

One can also partition

$$P_e := \begin{bmatrix} P & X \\ X^T & \hat{P} \end{bmatrix}, \quad Q_e := \begin{bmatrix} Q & Y \\ Y^T & \hat{Q} \end{bmatrix}$$

and solve

$$\begin{bmatrix} A & \\ & \hat{A} \end{bmatrix} \begin{bmatrix} P & X \\ X^T & \hat{P} \end{bmatrix} \begin{bmatrix} A^T & \\ & \hat{A}^T \end{bmatrix} + \begin{bmatrix} B \\ \hat{B} \end{bmatrix} \begin{bmatrix} B^T & \hat{B}^T \end{bmatrix} = \begin{bmatrix} P & X \\ X^T & \hat{P} \end{bmatrix},$$

$$\begin{bmatrix} A^T & \\ & \hat{A}^T \end{bmatrix} \begin{bmatrix} Q & Y \\ Y^T & \hat{Q} \end{bmatrix} \begin{bmatrix} A & \\ & \hat{A} \end{bmatrix} + \begin{bmatrix} C^T \\ -\hat{C}^T \end{bmatrix} \begin{bmatrix} C & -\hat{C} \end{bmatrix} = \begin{bmatrix} Q & Y \\ Y^T & \hat{Q} \end{bmatrix}$$

Gradients are easier

Let us define the gradient of a scalar function $f(X)$ as

$$[\nabla_X f(X)]_{i,j} = \frac{d}{dX_{i,j}} f(X), \quad i = 1, \ldots, n, \quad j = 1, \ldots, p$$

then the gradients $\nabla_{\hat{A}} \mathcal{J}, \nabla_{\hat{B}} \mathcal{J}, \nabla_{\hat{C}} \mathcal{J}$ satisfy the equations

$$\boxed{\tfrac{1}{2}\nabla_{\hat{A}} \mathcal{J} = \hat{Q}\hat{A}\hat{P} + Y^T AX, \quad \tfrac{1}{2}\nabla_{\hat{B}} \mathcal{J} = \hat{Q}\hat{B} + Y^T B, \quad \tfrac{1}{2}\nabla_{\hat{C}} \mathcal{J} = \hat{C}\hat{P} - CX}$$

where

$$A^T Y \hat{A} - C^T \hat{C} = Y, \quad \hat{A}^T \hat{Q} \hat{A} + \hat{C}^T \hat{C} = \hat{Q},$$
$$\hat{A} X^T A^T + \hat{B} B^T = X^T, \quad \hat{A}\hat{P}\hat{A}^T + \hat{B}\hat{B}^T = \hat{P}$$

This is quite cheap to solve (sparse matrix techniques)!

This result is due to Wilson (1970) and was «revisited» by several others (Hyland Bernstein, Halevi, Gugercin-Antoulas-Beattie, VD-Gallivan-Absil, Bunse-Gerstner-Kubalinska-Vossen-Wilczek)

Leads to a fixed point iteration

Define $(X, Y, \hat{P}, \hat{Q}) = F(\hat{A}, \hat{B}, \hat{C})$ where

$$A^T Y \hat{A} - C^T \hat{C} = Y, \quad \hat{A}^T \hat{Q} \hat{A} + \hat{C}^T \hat{C} = \hat{Q},$$

$$\hat{A} X^T A^T + \hat{B} B^T = X^T, \quad \hat{A} \hat{P} \hat{A}^T + \hat{B} \hat{B}^T = \hat{P}$$

and then compute$(\hat{A}, \hat{B}, \hat{C}) = G(X, Y, \hat{P}, \hat{Q})$ from

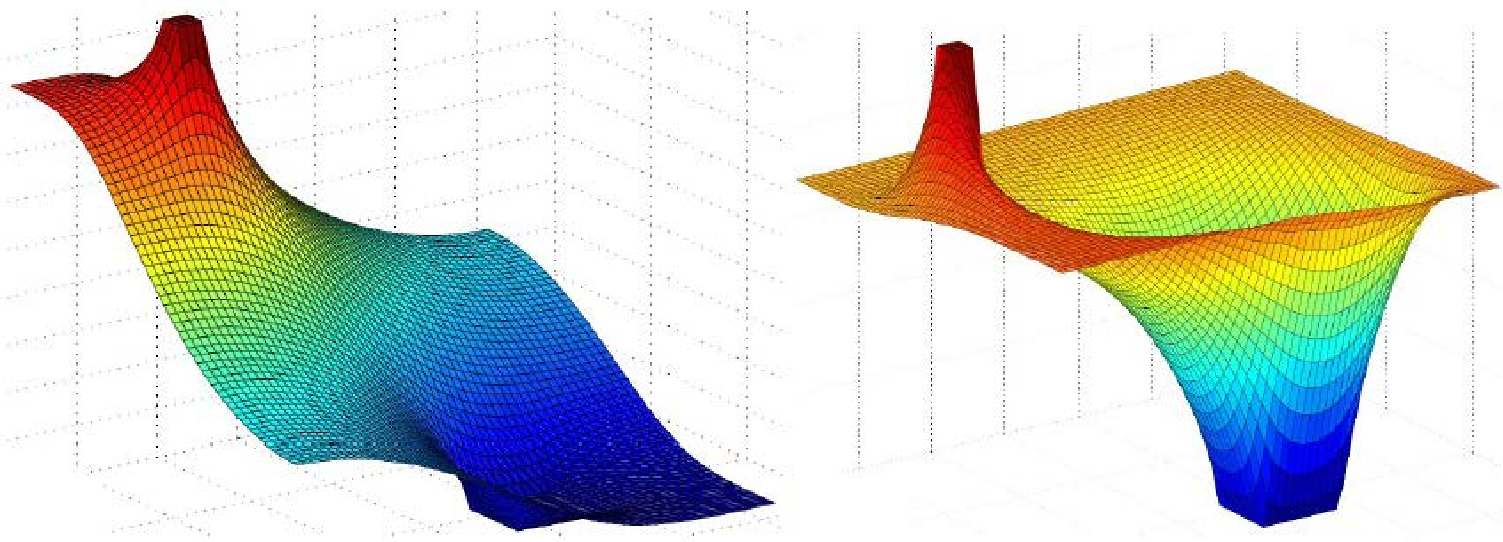
$$W := -Y\hat{Q}^{-1}, V := X\hat{P}^{-1} \ \hat{A} = W^T A V, \hat{B} = W^T B, \hat{C} = CV,$$

The fixed point of $(\hat{A}, \hat{B}, \hat{C}) = G(F(\hat{A}, \hat{B}, \hat{C}))$ are also stationary points of $\|E(z)\|_{\mathcal{H}_2}$

The basic idea is in essence due to Sorensen-Antoulas and made more formal in Gugercin-Antoulas-Beattie (IRKA)

A PDE example on a FE mesh

Approximating the dynamic of boundary value problem

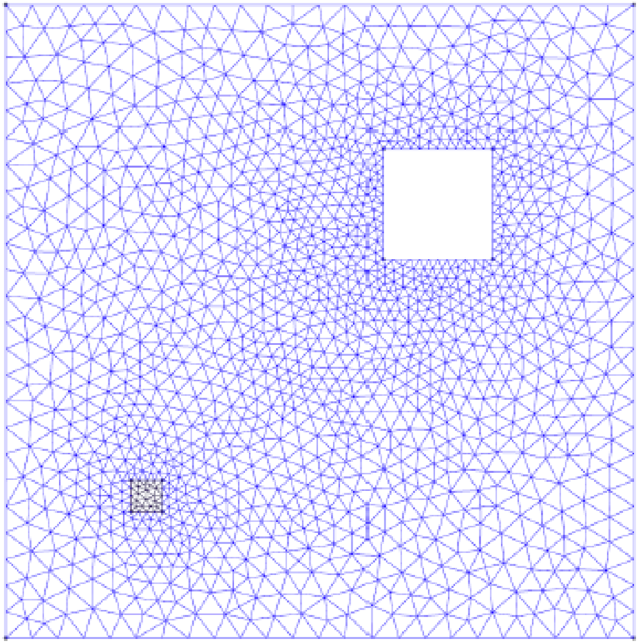


Examples : diffusion and convection/diffusion equation

We will assume

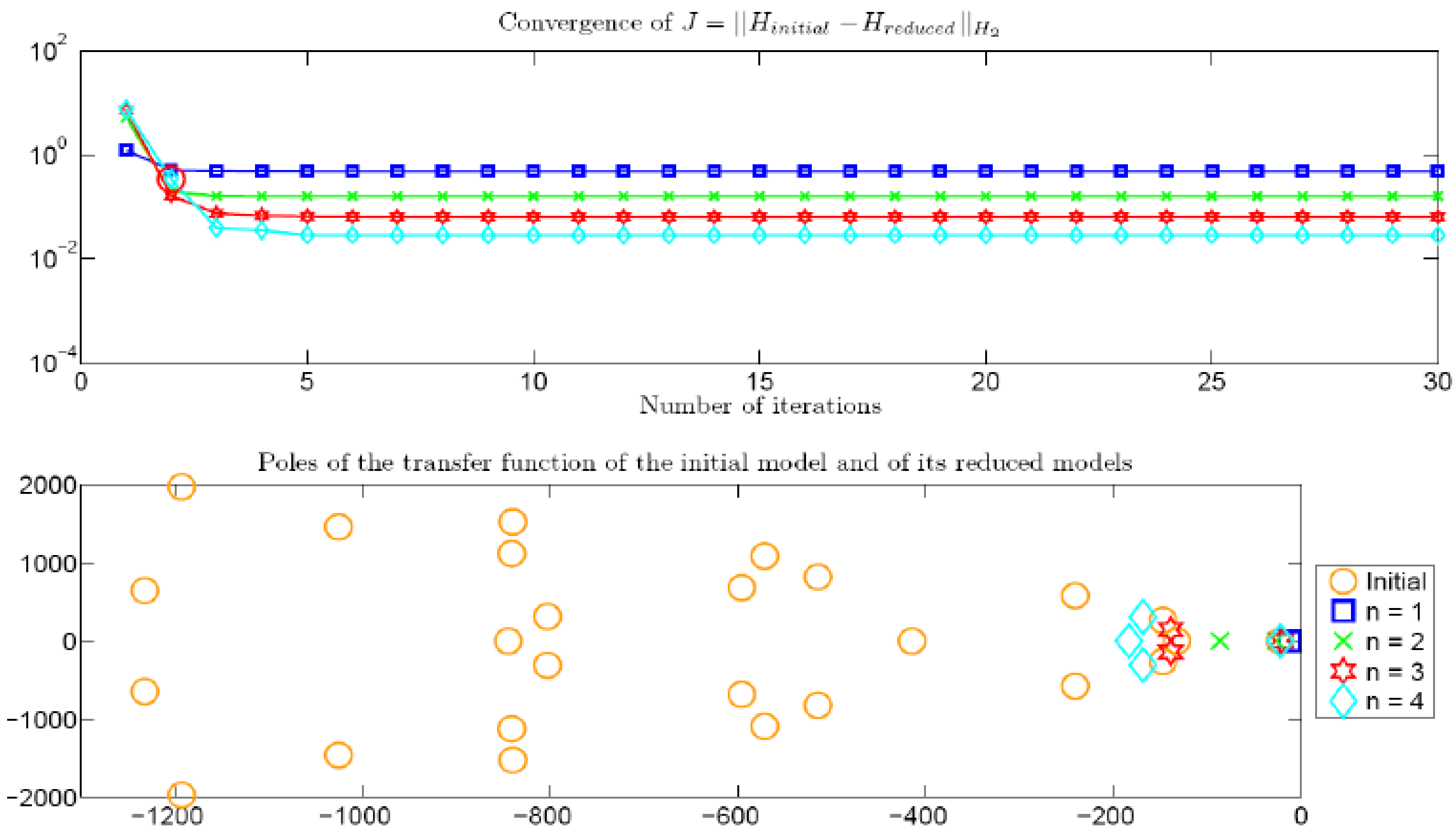
Finite element grids
are cheap to generate
for every resolution

The models are
- Linear time-invariant or
- Linear time-varying



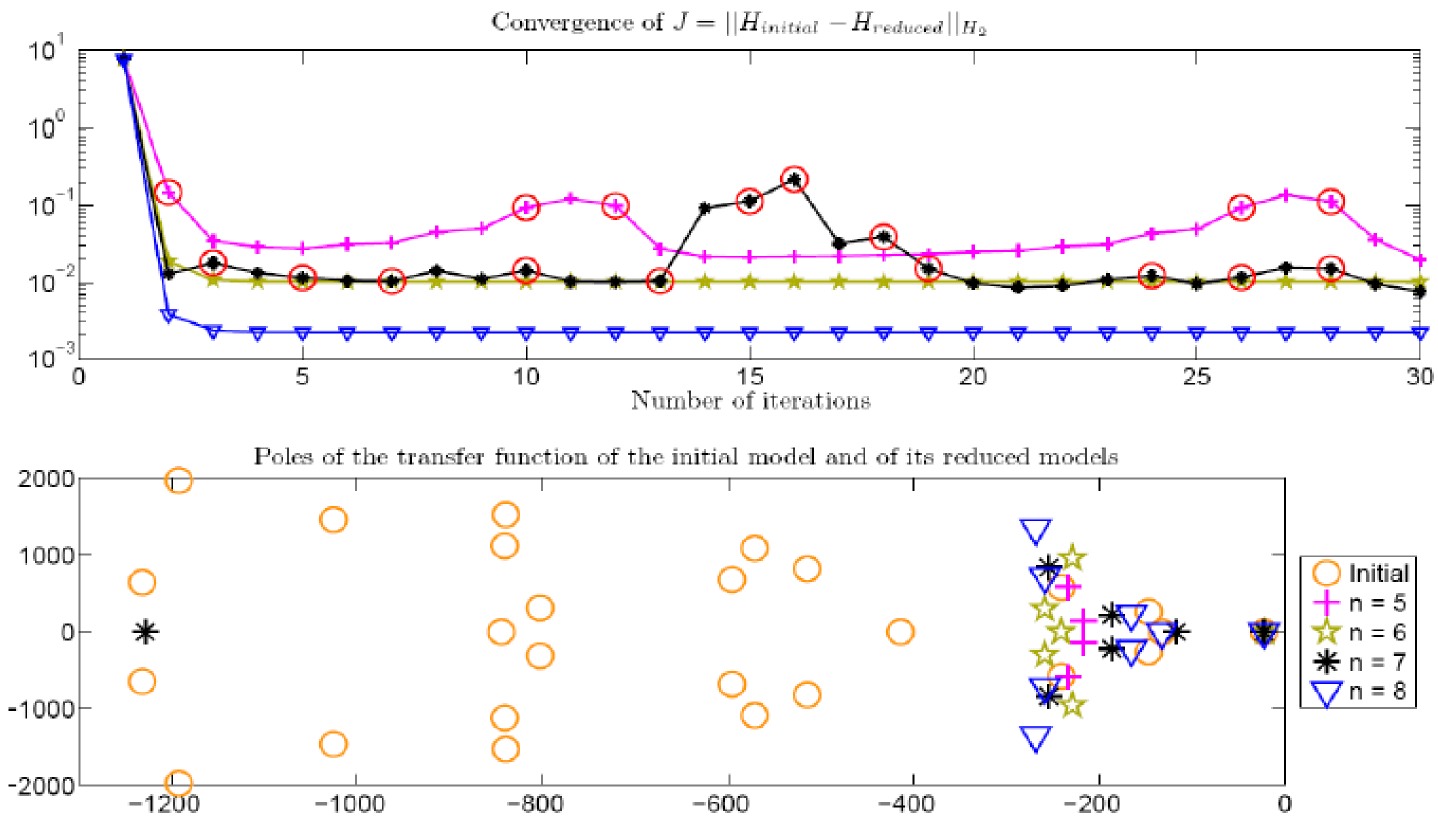
H_2 approximations give good results

Fixed point iteration often converges ...

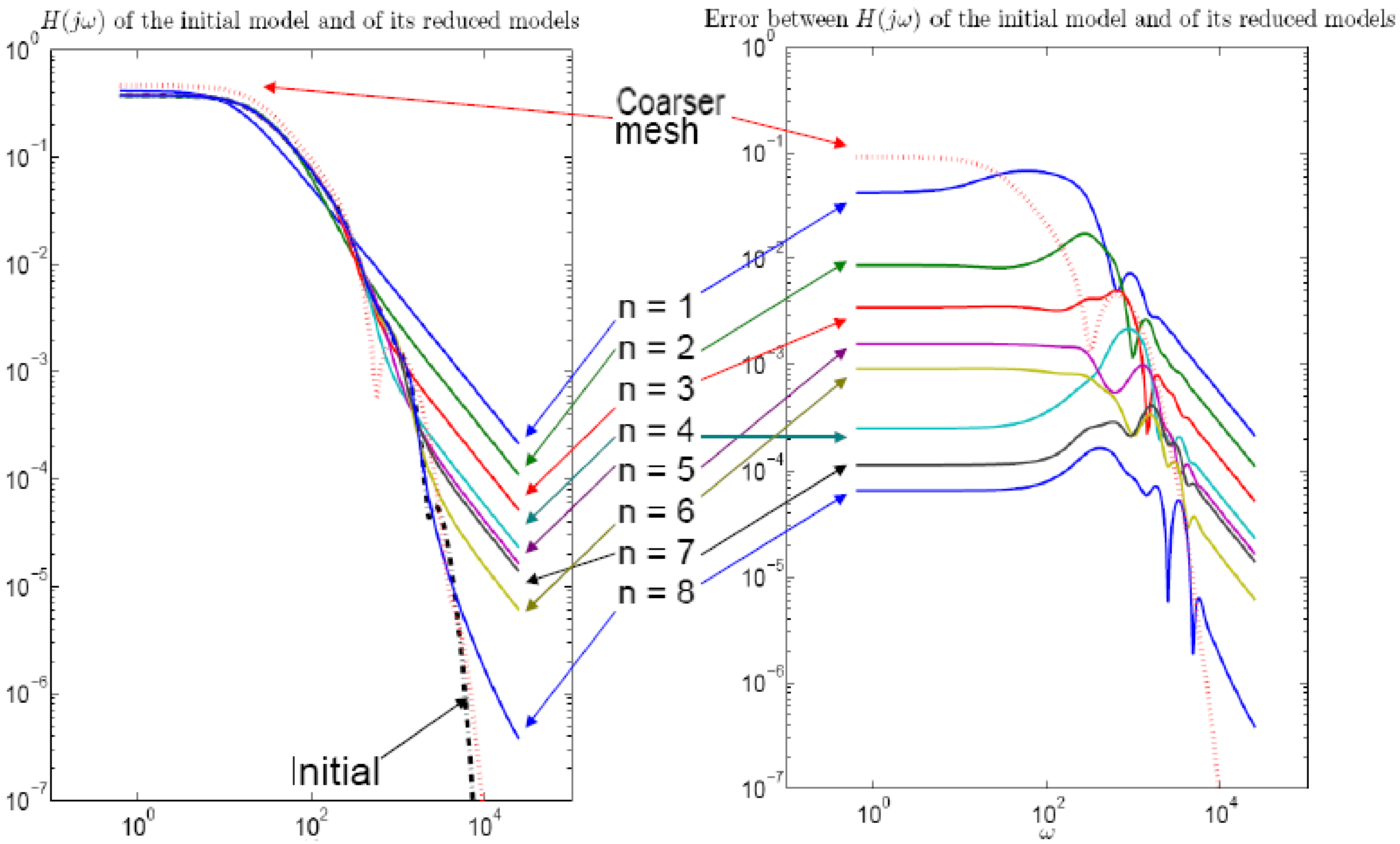


but can also be erratic

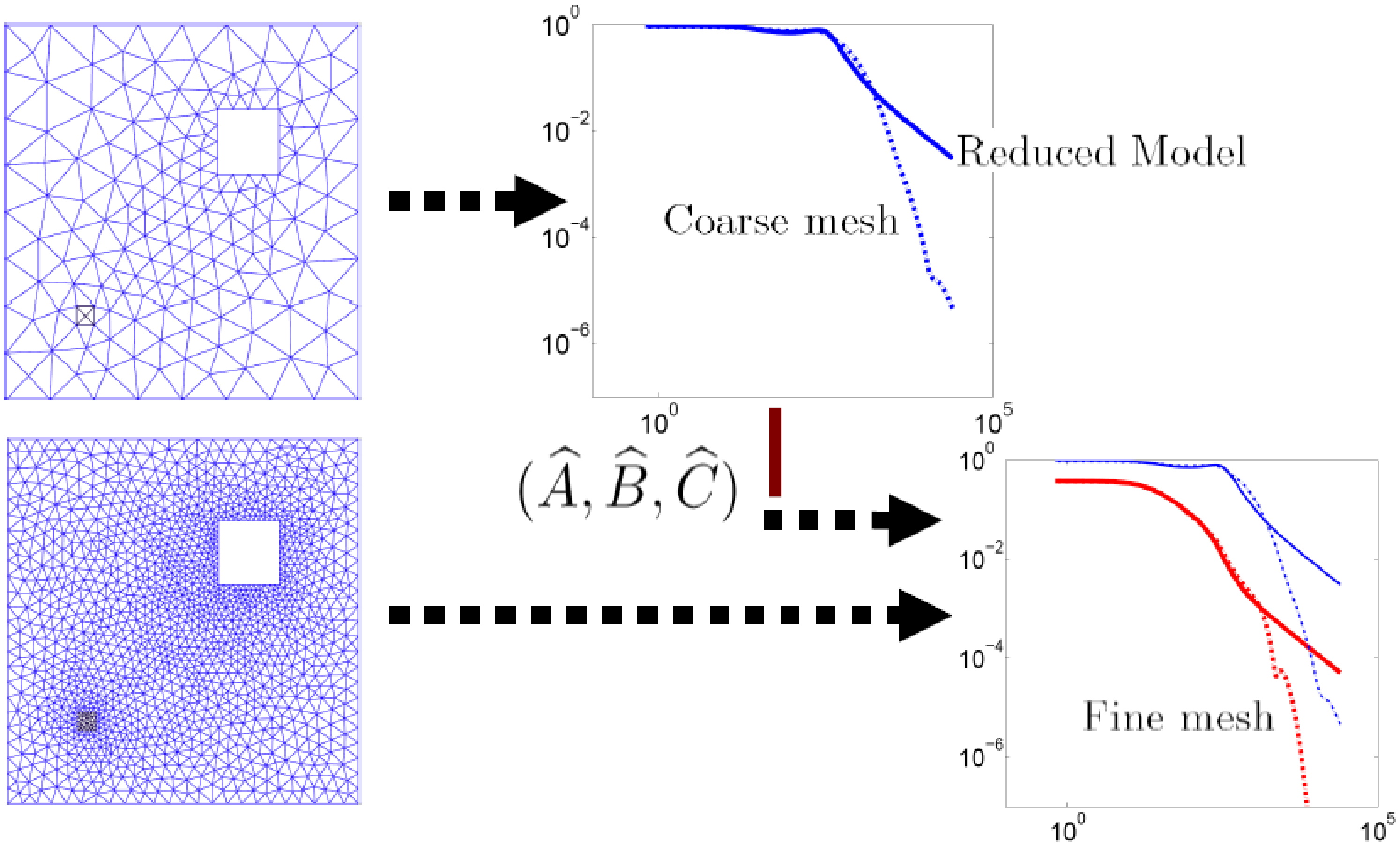
Because intermediate steps can yield unstable systems



Approximation errors

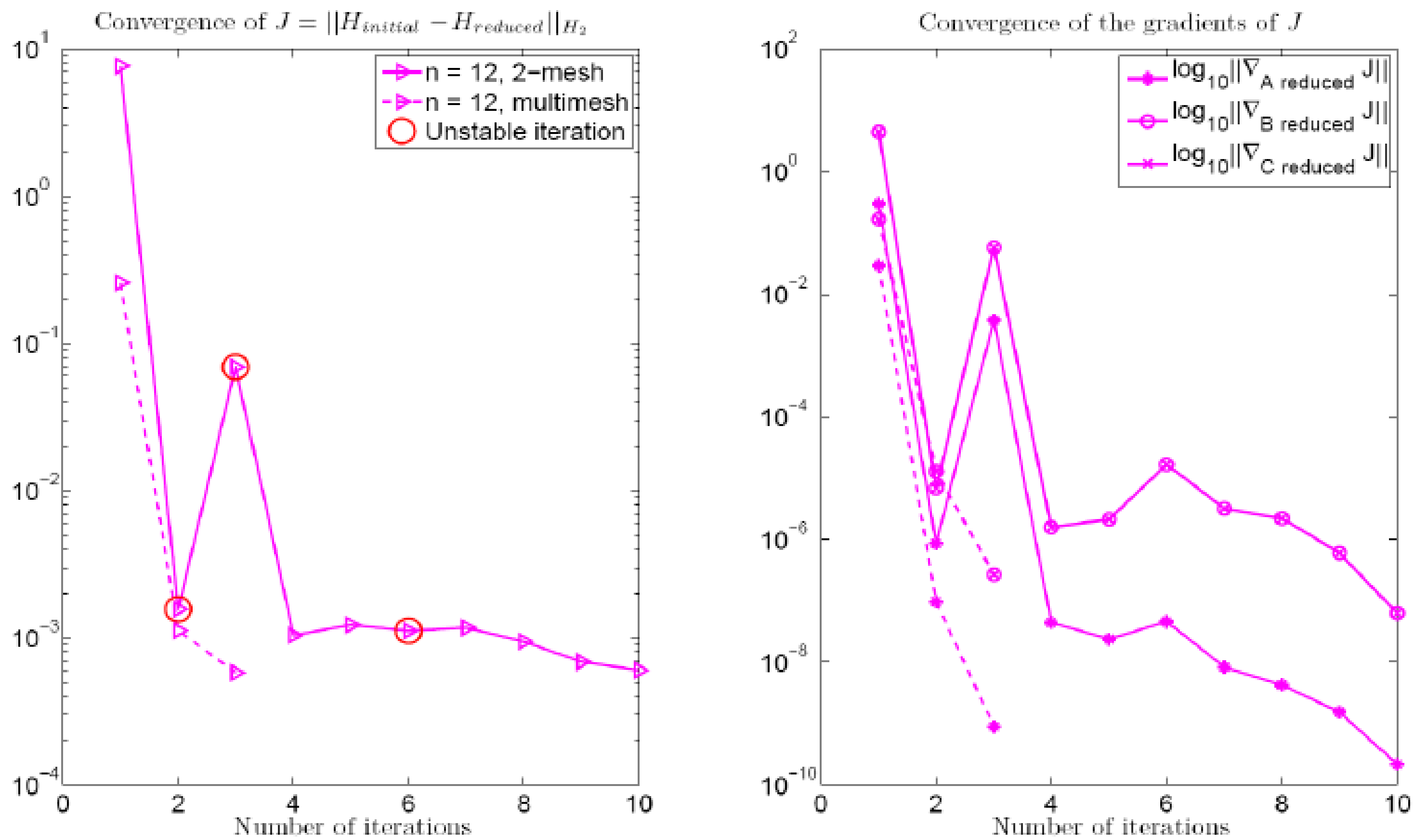


Multilevel idea



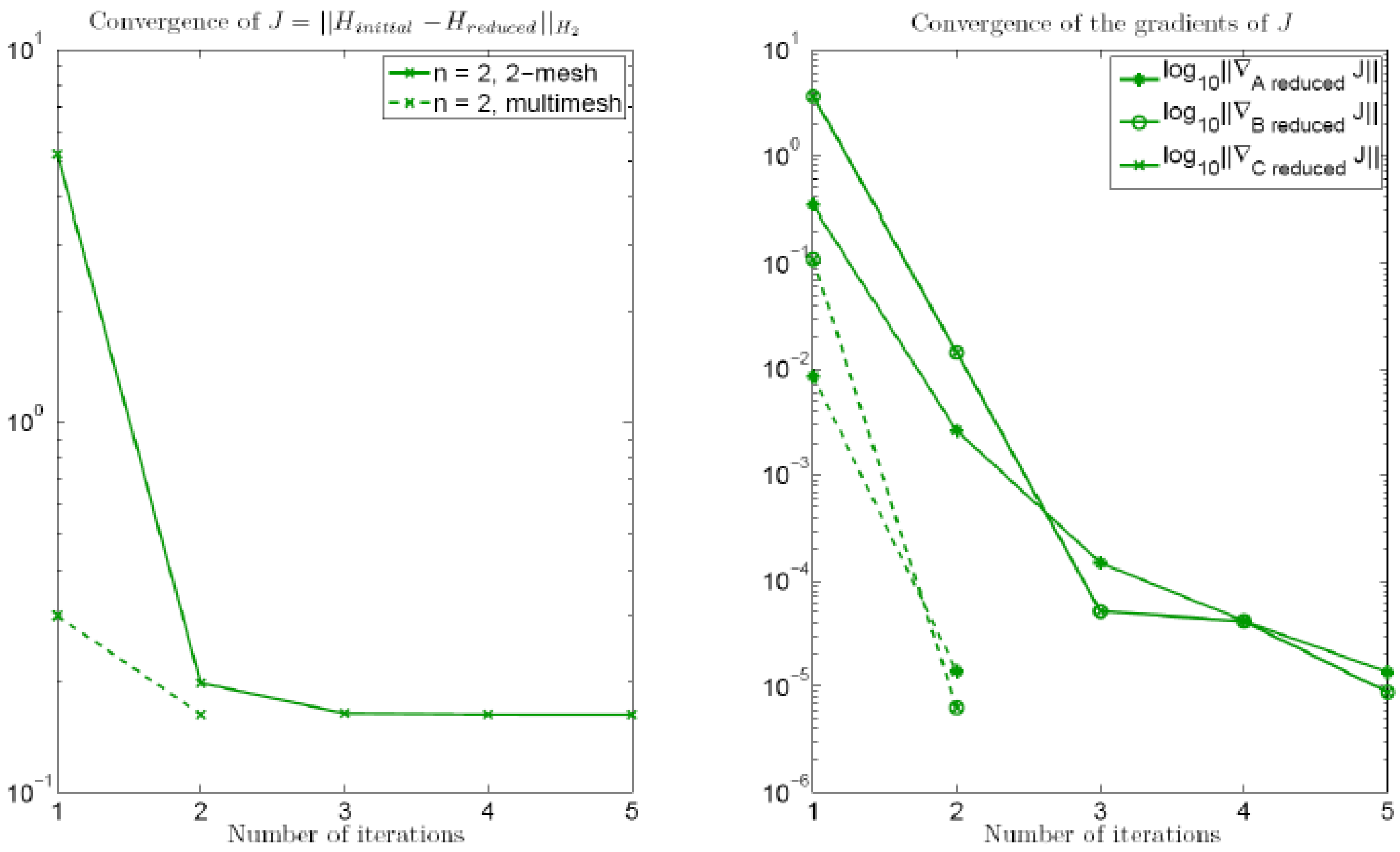
Experiments

Unstable intermediate systems are now avoided



Experiments

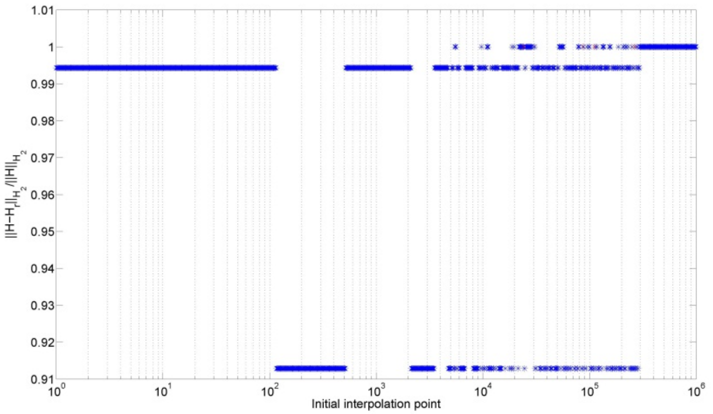
Multilevel method requires only one or two (final) steps



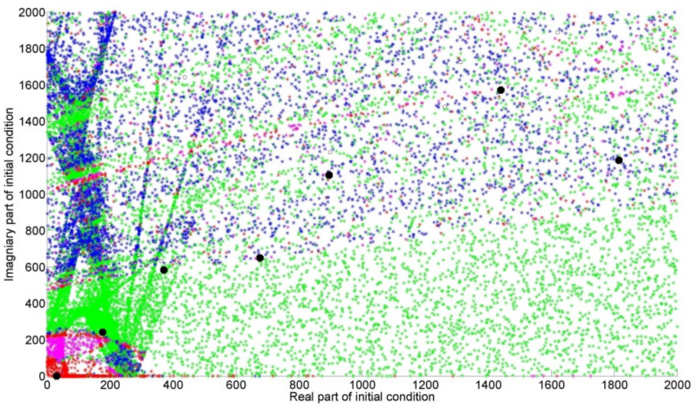
Convergence is delicate

Basins of attraction of different local minima of low order error function

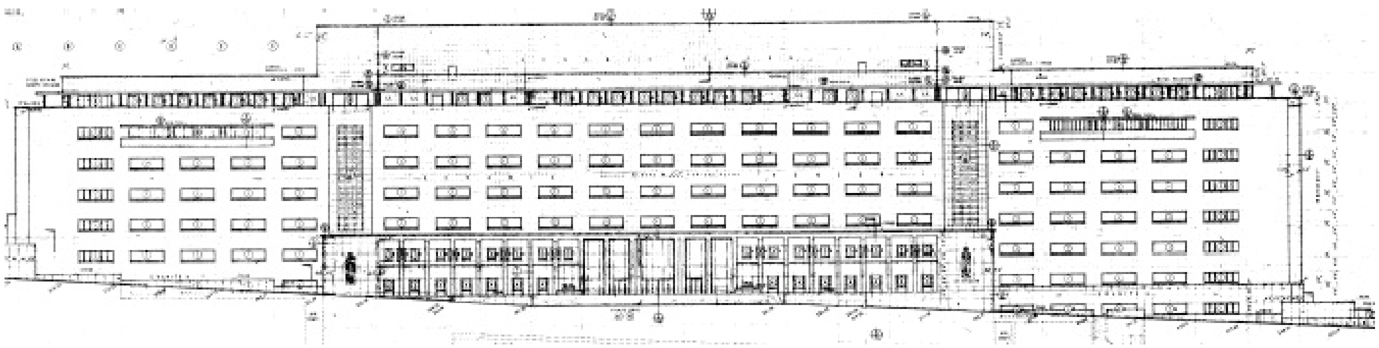
Order 1 approximation
Approximation error vs
Initial interpolation point



Order 2 approximation
Basin of attraction vs
Initial interpolation points



A mechanical application



Modeling of mechanical structures

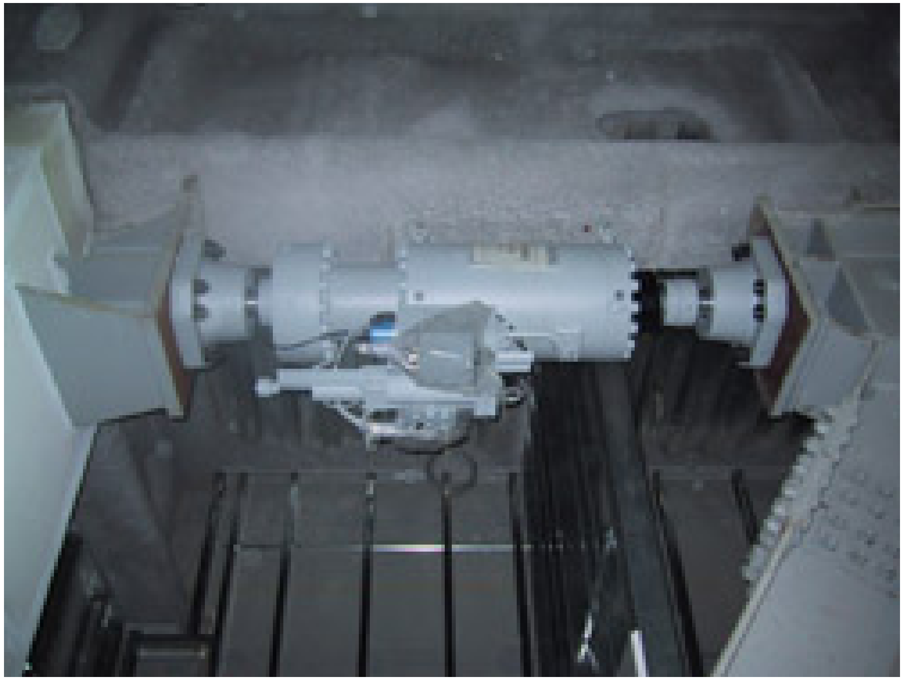
Identification/calibration (cheap sensors)

Simulation/validation (prognosis)

Model reduction

Control (earthquakes, large flexible structures)

Passive / Semi-Active Fluid Dampers



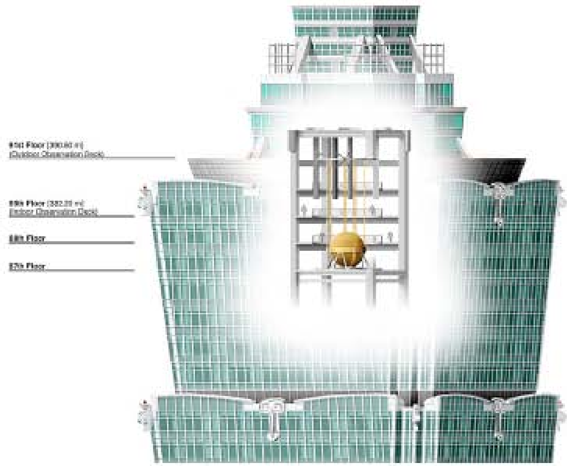
Passive fluid dampers contain bearings and oil absorbing seismic energy. Semi-active dampers work with variable orifice damping.

(Picture courtesy Steven Williams)

More examples of Control Mechanisms

Taipei 101: 508m

Damper between 87-91 floors

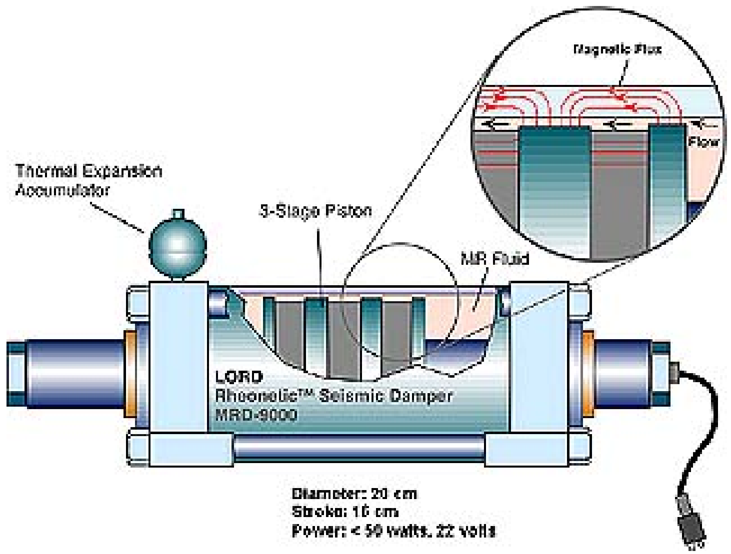


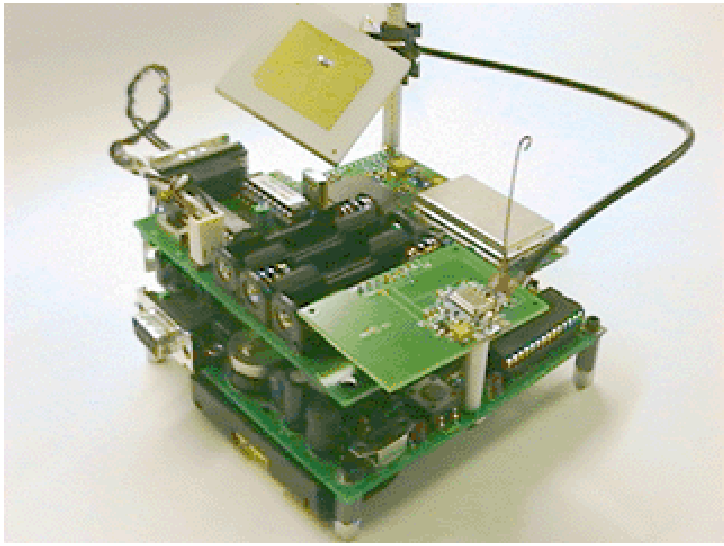
730 ton damper

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

Engineering Structures, Vol. 17, No. 9, Nov. 1995.

The Future: Fine-Grained Semi-Active Control





Dampers are based on Magneto-Rheological fluids with viscosity that changes in milliseconds, when exposed to a magnetic field.

New sensing and networking technology allows to do fine-grained real-time control of structures subjected to winds, earthquakes or hazards.

(Pictures courtesy Lord Corp.)

This technology starts to be applied…

Dongting Lake Bridge has now MR dampers to control wind-induced vibration

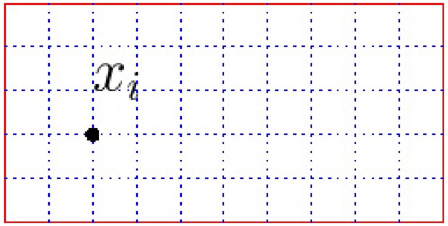
(Pictures courtesy of Prof. Y. L. Xu, Hong Kong Poly.)

Second order system models

Model derived from finite element discretization yield systems of the type

$$M\ddot{x}(t) + D\dot{x}(t) + Sx(t) = Bu(t), \quad y(t) = Cx(t)$$

whose solutions describe "vibrations" in the structure



where

M is the mass matrix $(M = M^T \succ 0)$

S is the stiffness matrix $(S = S^T \succ 0)$

$D(\omega)$ is the damping matrix (frequency dependent)

Reduced order model

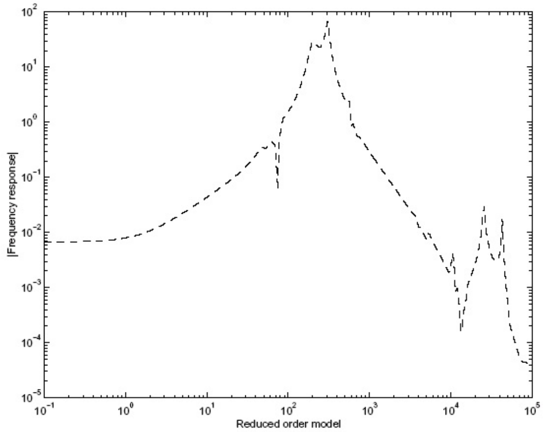
look for smaller model

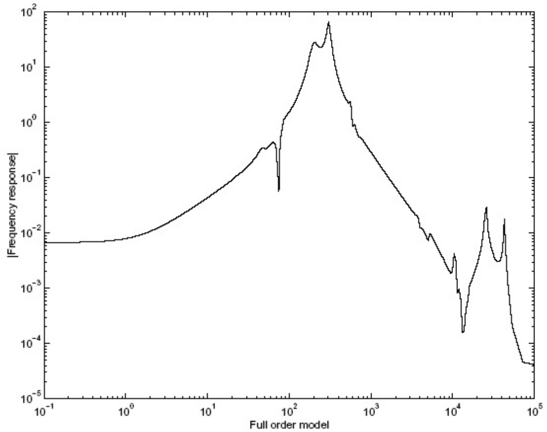
$$\hat{M}\ddot{\hat{x}}(t) + \hat{D}\dot{\hat{x}}(t) + \hat{S}\hat{x}(t) = \hat{B}u(t), \quad \hat{y}(t) = \hat{C}\hat{x}(t)$$

where $u(t) \in \mathbb{R}^m, y(t), \hat{y}(t) \in \mathbb{R}^p,\ x(t) \in \mathbb{R}^N,\ \hat{x}(t) \in \mathbb{R}^n,\ n \ll N$

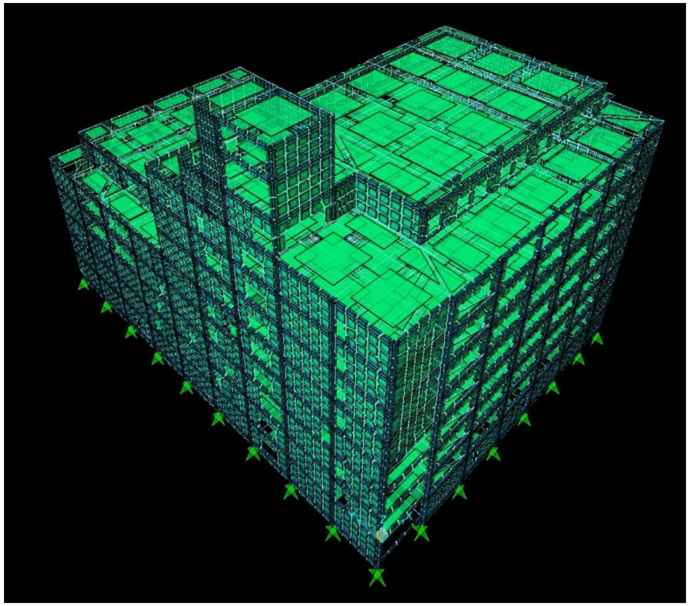
and transfer functions (frequency responses) are close

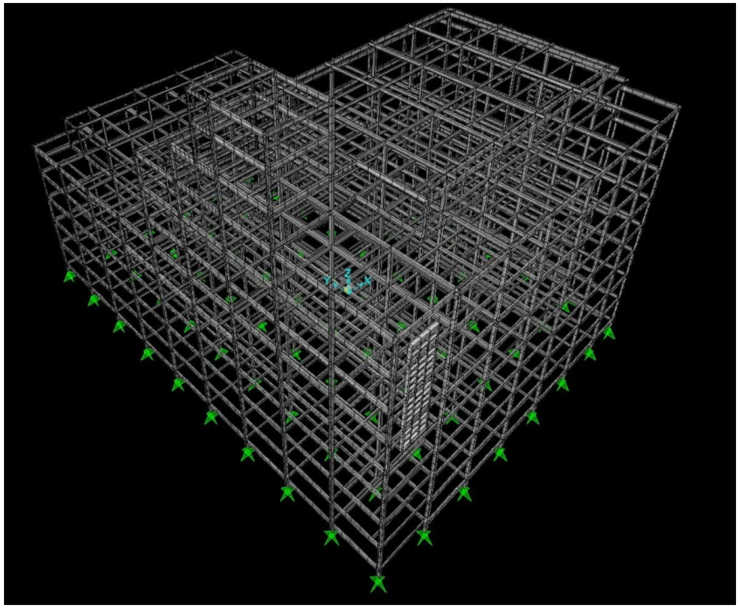
$$\hat{C}(\hat{M}s^2 + \hat{D}s + \hat{S})^{-1}\hat{B} \simeq C(Ms^2 + Ds + S)^{-1}B$$





Start by simplifying the model …

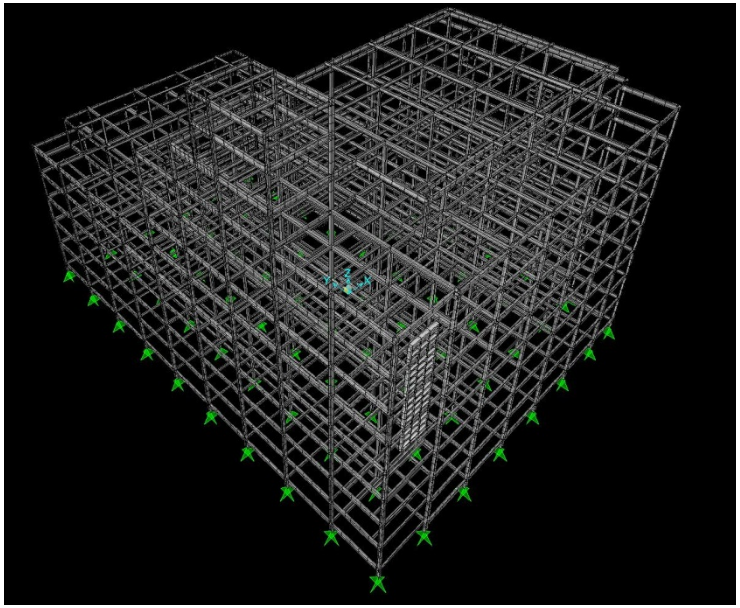




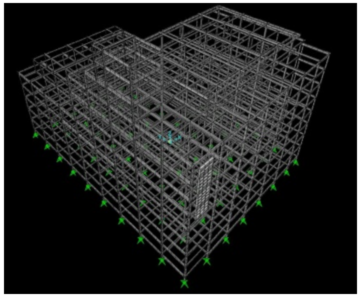
Simplify by
keeping only concrete
substructure

and then reduce the state dimension …

26400 2nd order eqs



20 2nd order eqs



i.e. reduce the number of equations describing the "state" of the system

Use of 2nd order models

Generalized state space model with $\xi^T(t) = [\, x^T(t) \quad \dot{x}^T(t) \,]$

$$\left\{ \begin{array}{rcl} \underbrace{\begin{bmatrix} D & M \\ M & 0 \end{bmatrix}}_{\mathcal{E}} \dot{\xi}(t) & = & \underbrace{\begin{bmatrix} -S & 0 \\ 0 & M \end{bmatrix}}_{\mathcal{A}} \xi(t) + \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{\mathcal{B}} u(t), \\ y(t) & = & \underbrace{\begin{bmatrix} C & 0 \end{bmatrix}}_{\mathcal{C}} \xi(t) \end{array} \right.$$

Reduced order model $\{W^*\mathcal{E}V, W^*\mathcal{A}V, W^*\mathcal{B}, \mathcal{C}V\}$ is second order if

$$W = \begin{bmatrix} W_{11} & 0 \\ 0 & W_{22} \end{bmatrix}, \; V = \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix}$$

Structure constraint on projection spaces !

Clamped beam example

n	k	m	p	$\frac{\|\mathcal{H}-\hat{\mathcal{H}}_{BT}\|_2}{\|\mathcal{H}\|_2}$	$\frac{\|\mathcal{H}-\hat{\mathcal{H}}_{SOBT}\|_2}{\|\mathcal{H}\|_2}$
174	17	1	1	2.88e-05	1.83e-04

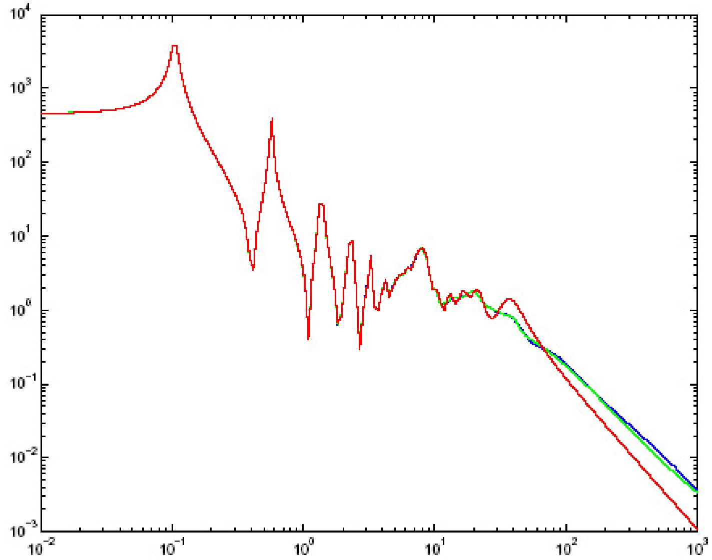
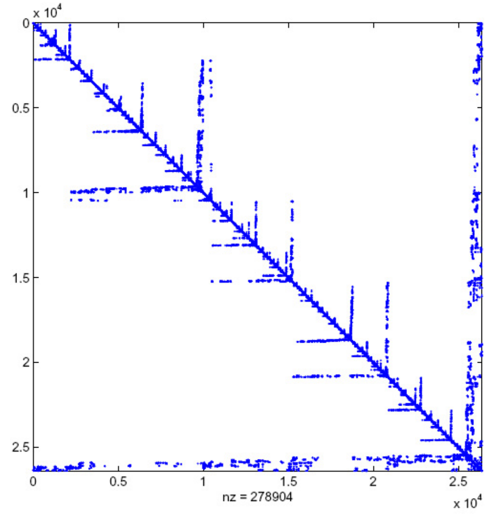


Figure 1: Amplitude of the frequency response.

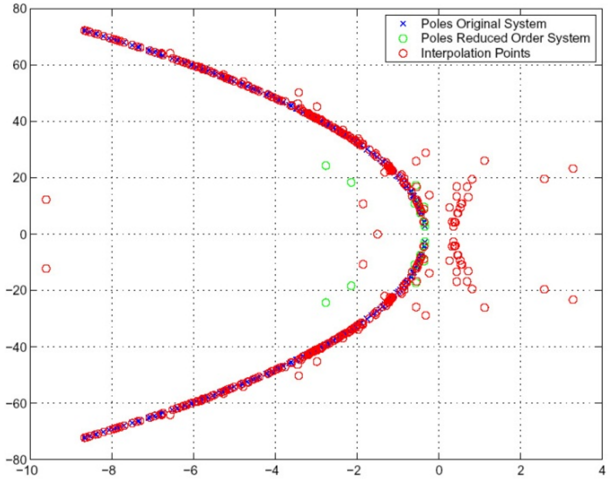
(— Original model, — BT, — SOBT)

Interpolation of large scale systems

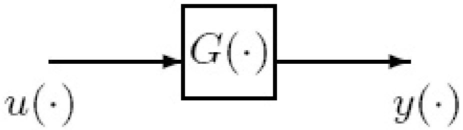
LA hospital building
has 2x26400 variables
but model is sparse

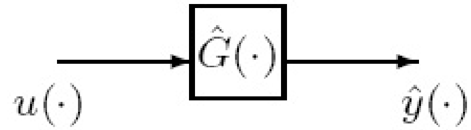


Eigenfrequencies closest
to the origin are typically
good interpolation points

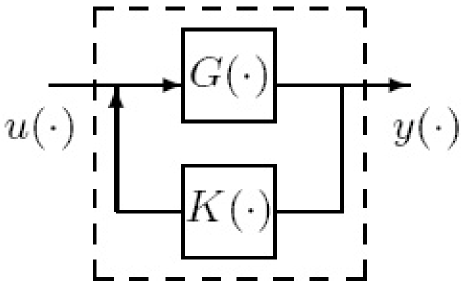


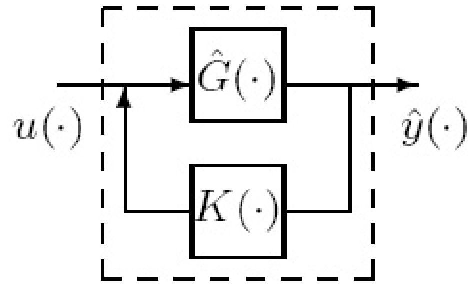
Interconnected systems





Open loop and closed loop approximations can be very different





The transfer function is now structured :

$$\min \left\| (I - G(s)K(s))^{-1}G(s) - (I - \hat{G}(s)K(s))^{-1}\hat{G}(s) \right\|_{H_\infty}$$

Several examples

Second order systems

$$\mathcal{S} \; : \; \begin{cases} (s^2M + sD + S)x(s) &= Bu(s) \\ y(s) &= Cx(s) \end{cases}$$

Cascaded systems

$$\mathcal{S} \; : \quad y(s) = [W_{out} G(s) W_{in}(s)] u(s)$$

Closed loop systems

$$\mathcal{S} \; : \quad \begin{bmatrix} y(s) \\ w(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \begin{bmatrix} u(s) \\ v(s) \end{bmatrix}$$

$$v(s) = C(s)w(s)$$

General interconnected systems

Interconnected systems

$$\begin{bmatrix} w_1(s) \\ \vdots \\ w_k(s) \end{bmatrix} = \begin{bmatrix} G_1(s) & & \\ & \ddots & \\ & & G_k(s) \end{bmatrix} \begin{bmatrix} v_1(s) \\ \vdots \\ v_k(s) \end{bmatrix}$$

Interconnect the components to each other and to the input and output $u(s), y(s)$

$$\begin{bmatrix} v_1(s) \\ \vdots \\ v_k(s) \end{bmatrix} = \begin{bmatrix} K_{11} & \dots & K_{1k} \\ \vdots & \ddots & \vdots \\ K_{k1} & \dots & K_{kk} \end{bmatrix} \begin{bmatrix} w_1(s) \\ \vdots \\ w_k(s) \end{bmatrix} + \begin{bmatrix} H_1 \\ \vdots \\ H_k \end{bmatrix} u(s)$$

$$y(s) = \begin{bmatrix} F_1 & \dots & F_k \end{bmatrix} \begin{bmatrix} w_1(s) \\ \vdots \\ w_k(s) \end{bmatrix}$$

Realize interconnected systems

Realize $G_i(s)$ of McMillan degree n_i as $C_i(sI_{n_i} - A_i)^{-1}B_i + D_i$ yielding :

$$A := diag\{A_i\}, \quad B := diag\{B_i\}, \quad C := diag\{C_i\}, \quad D := diag\{D_i\},$$

The interconnected system $T(s)$ is then realized by :

$$A_T := A + BK(I - DK)^{-1}C, \quad B_T := B(I - KD)^{-1}H,$$
$$C_T := F(I - DK)^{-1}C, \quad D_T := FD(I - KD)^{-1}H$$

It is not as bad as it looks since K is typically sparse, which is maintained in the realization of $T(s)$

Closed loop Gramians

Idea: For each subsystem $G_i(s)$, define $n_i \times n_i$ Gramians containing information about the energy distribution of the I/O map in x_i only.

Let us consider the controllability and observability Gramians of $T(s)$:

$$A_T P_T + P_T A_T^T + B_T B_T^T = 0 \quad , \quad A_T^T Q_T + Q_T A_T + C_T^T C_T = 0.$$

Decompose, with $P_{ij}, Q_{ij} \in \mathbb{C}^{n_i \times n_j}$

$$P_T = \begin{bmatrix} P_{11} & \dots & P_{1k} \\ \vdots & \ddots & \vdots \\ P_{k1} & \dots & P_{kk} \end{bmatrix} \quad , \quad Q_T = \begin{bmatrix} Q_{11} & \dots & Q_{1k} \\ \vdots & \ddots & \vdots \\ Q_{k1} & \dots & Q_{kk} \end{bmatrix},$$

Constrained Gramians

Energy interpretation of diagonal blocks $[P_{ii}]^{-1}$ and $[P^{-1}]_{ii}$

$$\min_{u,\bigcup_{j\neq i} x_j(0)} \|u(-\infty,0)\|^2_{\mathcal{L}_2} = x_i(0)^T [P_{ii}]^{-1} x_i(0)$$

$$\min_{u,\forall_{j\neq i} x_j(0)=0} \|u(-\infty,0)\|^2_{\mathcal{L}_2} = x_i(0)^T [P^{-1}]_{ii} x_i(0)$$

Similar for diagonal blocks $[Q_{ii}]^{-1}$ and $[Q^{-1}]_{ii}$

(see Enns, Liu, Varga, VV...)

Constrained Krylov spaces

If

$$(\hat{A}, \hat{B}, \hat{C}) = (Z^T A V, Z^T B, C V), \quad Z^T V = I$$

where

$$\mathcal{K}_k((\sigma I - A)^{-1}, (\sigma I - A)^{-1} B) \subseteq Im(V)$$

Then

$$(\hat{A}_T, \hat{B}_T, \hat{C}_T) = (Z^T A_T V, Z^T B_T, C_T V), \quad Z^T V = I$$

and

$$\mathcal{K}_k((\sigma I - A_T)^{-1}, (\sigma I - A_T)^{-1} B_T) \subseteq \mathcal{K}_k((\sigma I - A)^{-1}, (\sigma I - A)^{-1} B)$$

Open loop equals closed loop interpolation unless you change points !?

Time-varying linear systems

Approximate the (discrete) time-varying systems

$$\begin{cases} x(k+1) = A(k)x(k) + B(k)u(k) \\ y(k) = C(k)x(k) + D(k)u(k) \end{cases}$$

by a lower order models of same type. We notice that

$$\begin{bmatrix} y(k) \\ y(k+1) \\ y(k+2) \\ \vdots \end{bmatrix} = \begin{bmatrix} C(k) \\ C(k+1)A(k) \\ C(k+2)A(k+1)A(k) \\ \vdots \end{bmatrix} x(k),$$

$x(k) =$

$$\begin{bmatrix} B_{(k-1)} & A_{(k-1)}B_{(k-2)} & A_{(k-1)}A_{(k-2)}B_{(k-3)} & \dots \end{bmatrix} \begin{bmatrix} u(k-1) \\ u(k-2) \\ u(k-3) \\ \vdots \end{bmatrix}$$

H_2 approximation

Systems now look like

$$\begin{cases} x_{k+1} & = A_k x_k + B_k u_k \\ y_k & = C_k x_k \end{cases} \qquad \begin{cases} \hat{x}_{k+1} & = \hat{A}_k \hat{x}_k + \hat{B}_k u_k \\ \hat{y}_k & = \hat{C}_k \hat{x}_k \end{cases}$$

with an error system where $e_k := y_k - \hat{y}_k$

$$\mathcal{E} := \begin{cases} x^e_{k+1} & = A^e_k x^e_k + B^e_k u_k \\ e_k & = C^e_k x^e_k \end{cases}$$

where

$$A^e_k := \begin{bmatrix} A_k & \\ & \hat{A}_k \end{bmatrix}, \quad B^e_k = \begin{bmatrix} B_k \\ \hat{B}_k \end{bmatrix}, \quad C^e_k = \begin{bmatrix} C_k - \hat{C}_k \end{bmatrix}$$

Its state for initial condition $x^e_{k_0} = 0$ is given by

$$x^e_k = \sum_{i=k_0}^{k-1} \Phi^e_{k,i+1} B^e_i u_i, \quad \Phi^e_{k+1,i} = A^e_k \Phi^e_{k,i} \ (k \geq i), \quad \Phi^e_{k,k} = I$$

Error function is a linear map

Error system response satisfies

$$\tilde{e} = E\tilde{u}, \quad \tilde{e} := \begin{bmatrix} e_{k_0+1} \\ \vdots \\ e_{k_f+1} \end{bmatrix}, \quad \tilde{u} := \begin{bmatrix} u_{k_0} \\ \vdots \\ u_{k_f} \end{bmatrix}, \quad E = D_C H D_B$$

and

$$D_C = \begin{bmatrix} C^e_{k_0+1} & & 0 \\ & \ddots & \\ 0 & & C^e_{k_f+1} \end{bmatrix}, \quad D_B = \begin{bmatrix} B^e_{k_0} & & 0 \\ & \ddots & \\ 0 & & B^e_{k_f} \end{bmatrix}$$

$$H = \begin{bmatrix} \Phi^e_{k_0,k_0} & & 0 \\ \vdots & \ddots & \\ \Phi^e_{k_f,k_0} & \cdots & \Phi^e_{k_f,k_f} \end{bmatrix}$$

depends on the reduced order model

$$\|\mathcal{E}\|^2_{\mathcal{H}_2} := \mathcal{J}(k_0, k_f) := \mathrm{tr}(E^T E) = tr(E E^T)$$

One shows that

$$\mathcal{J}(k_0, k_f) := \mathrm{tr} \sum_{k=k_0+1}^{k_f+1} C_k^e P_k^e {C_k^e}^T = \mathrm{tr} \sum_{k=k_0}^{k_f} {B_k^e}^T Q_k^e B_k^e$$

where

$$P_{k+1}^e = \begin{bmatrix} A_k & \\ & \hat{A}_k \end{bmatrix} P_k^e \begin{bmatrix} A_k^T & \\ & \hat{A}_k^T \end{bmatrix} + \begin{bmatrix} B_k \\ \hat{B}_k \end{bmatrix} \begin{bmatrix} B_k^T & \hat{B}_k^T \end{bmatrix}$$

$$Q_{k-1}^e = \begin{bmatrix} A_k^T & \\ & \hat{A}_k^T \end{bmatrix} Q_k^e \begin{bmatrix} A_k & \\ & \hat{A}_k \end{bmatrix} + \begin{bmatrix} C_k^T \\ \hat{C}_k^T \end{bmatrix} \begin{bmatrix} C_k & \hat{C}_k \end{bmatrix},$$

[VD-Gallivan-Absil, CESAME rept., Univ. Cath. Louvain, 2009]

Gradients are given by

$$\nabla_{\hat{A}_k}\mathcal{J} = 2(\hat{Q}_k\hat{A}_k\hat{P}_k + Y_k^T A_k X_k),$$

$$\nabla_{\hat{B}_k}\mathcal{J} = 2(\hat{Q}_k\hat{B}_k + Y_k^T B_k),$$

$$\nabla_{\hat{C}_k}\mathcal{J} = 2(\hat{C}_k\hat{P}_k - C_k X_k)$$

Updating rules and fixed point results are as before

$$W_k := Y_k\hat{Q}_k^{-1}, V_k = X_k\hat{P}_k^{-1}$$

$$(A_k^e, B_k^e, C_k^e) := (W_k^T A_k V_k, W_k^T B_k, C_k V_k).$$

where X_k, Y_k, $\tilde{P}_k$, $\tilde{Q}_k$ satisfy Stein like recurrences

$$X_{k+1} = A_k X_k \hat{A}_k^T + B_k\hat{B}_k^T,$$

$$\hat{P}_{k+1} = \hat{A}_k\hat{P}_k\hat{A}_k^T + \hat{B}_k\hat{B}_k^T,$$

$$Y_{k-1} = A_k^T Y_k \hat{A}_k^T - C_k^T\hat{C}_k,$$

$$\hat{Q}_{k-1} = \hat{A}_k^T Q_k \hat{A}_k + \hat{C}_k^T\hat{C}_k,$$

Nonlinear systems

Look for a simple energy function

Consider the discrete-time system

$$\begin{cases} x(k+1) = G(x(k), u(k)) \\ y(k) = H(x(k), u(k)). \end{cases}$$

One could linearize along a "nominal" trajectory $(x(k), u(k))$
and get $A(.), B(.), C(.), D(.)$ from Taylor expansion of $G(.,.), H(.,.)$

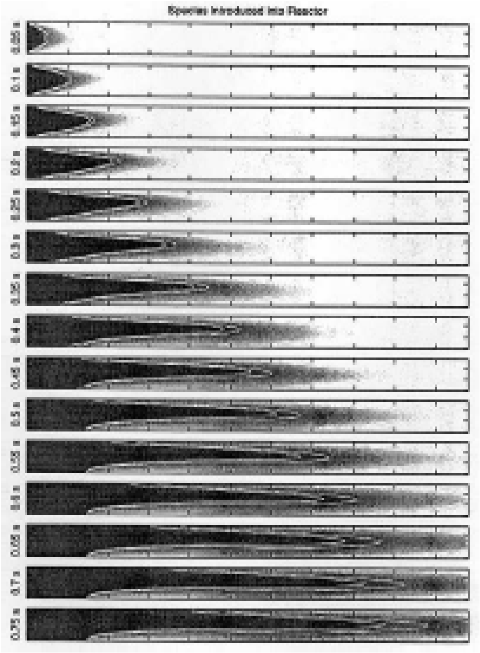
Simpler idea (POD) : (Holmes-Lumley-Berkooz '96)

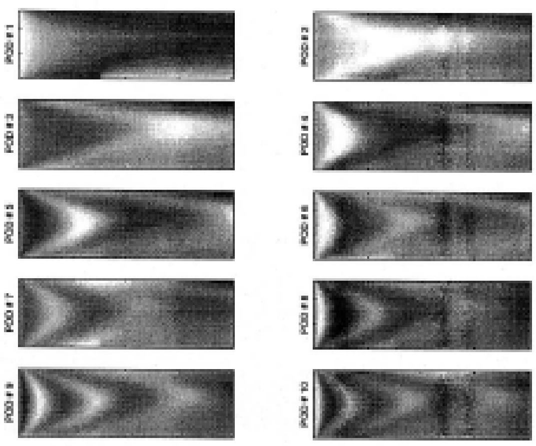
Use the "energy function" $G \doteq \sum_{k=k_i}^{k_f} x(k)x(k)^T$.

Example

Chemical vapor deposition reactor

Compute state trajectories for one "typical" input :

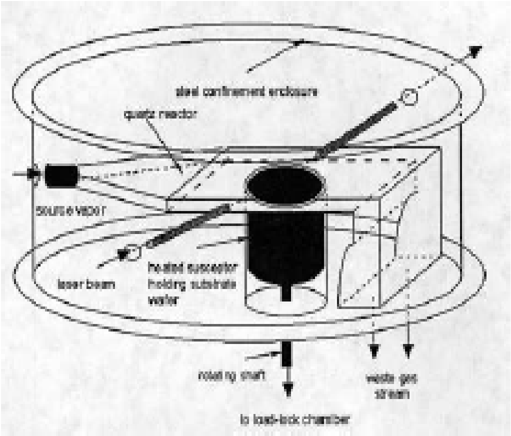




Snap shots of "typical" states

Ten dominant "states"

Use POD in CVD reactor (Ly-Tran '99)



Schematic representation of a horizontal quartz reactor in a steel confi nement shell

Conclusions

- Model reduction of linear time invariant systems is quite sophisticated and efficient these days
- Algorithmic aspects are the issue right now
- Time-varying extensions exist (for discrete-time case)
- Nonlinear extensions exist (several approaches)
- There are many successful test cases
- Model reduction is stil quite hot …