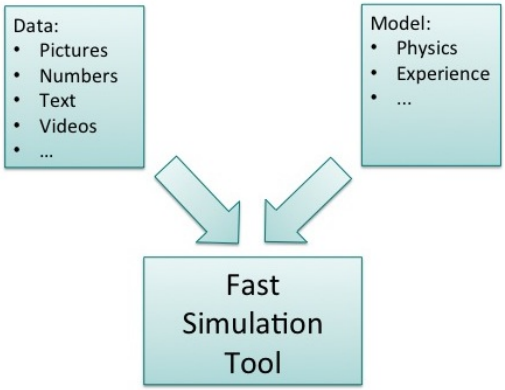
MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG

Data Driven Model Order Reduction

Peter Benner, Melina Freitag, Sara Grundel, Nils Hornung, Matthias Voigt

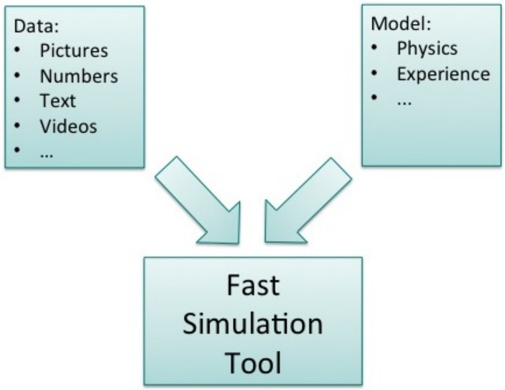
August 9, 2017

Data and/or Model



Questions

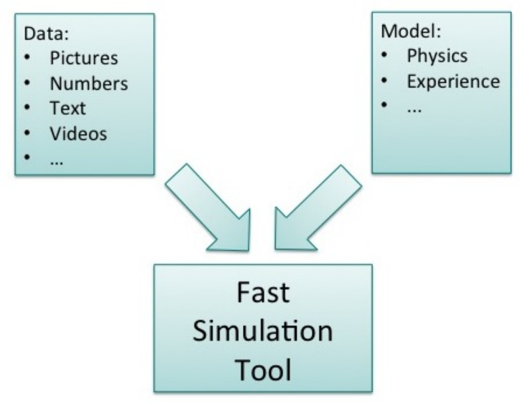
Data and/or Model



Questions

1. What kind of Data?

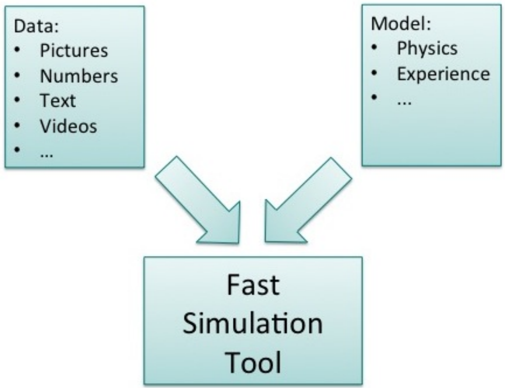
Data and/or Model



Questions

1. What kind of Data?
2. What kind of Model?

Data and/or Model

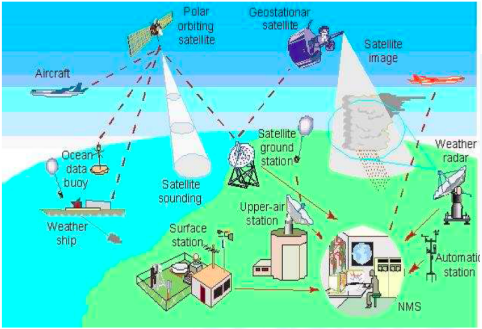


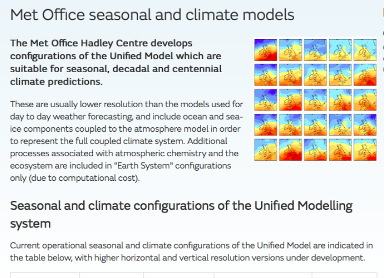
Questions

1. What kind of Data?
2. What kind of Model?
3. What kind of Simulation Tool?

THE EXAMPLE

Data Assimilation - Weather Forecast




Met Office seasonal and climate models

The Met Office Hadley Centre develops configurations of the Unified Model which are suitable for seasonal, decadal and centennial climate predictions.

These are usually lower resolution than the models used for day to day weather forecasting, and include ocean and sea-ice components coupled to the atmosphere model in order to represent the full coupled climate system. Additional processes associated with atmospheric chemistry and the ecosystem are included in "Earth System" configurations only (due to computational cost).

Seasonal and climate configurations of the Unified Modelling system

Current operational seasonal and climate configurations of the Unified Model are indicated in the table below, with higher horizontal and vertical resolution versions under development.

Motivation

"Big Data"

"In 2010, Google CEO Eric Schmidt observed that we now generate as much data in two days as we did from the start of human history through 2003."

Motivation

"Big Data"

"In 2010, Google CEO Eric Schmidt observed that we now generate as much data in two days as we did from the start of human history through 2003."

MOR

- PMOR and UQ
- Data Driven Methods
- nonlinear MOR

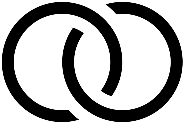
Motivation

"Big Data"

"In 2010, Google CEO Eric Schmidt observed that we now generate as much data in two days as we did from the start of human history through 2003."

MOR

- PMOR and UQ
- Data Driven Methods
- nonlinear MOR



Potential

Data-Driven Methods for Reduced-Order Modeling

Review Classical MOR

Input-Output-System

$$\dot{x} = f(x,u)$$
$$y = h(x,u)$$
$$u(t) \in \mathbb{R}^m$$
$$y(t) \in \mathbb{R}^p$$
$$x(t) \in \mathbb{R}^n$$
$$m, p \ll n$$

f, h known functions
Find:
f_r, h_r such that $\|y - y_r\|$ small for a given u

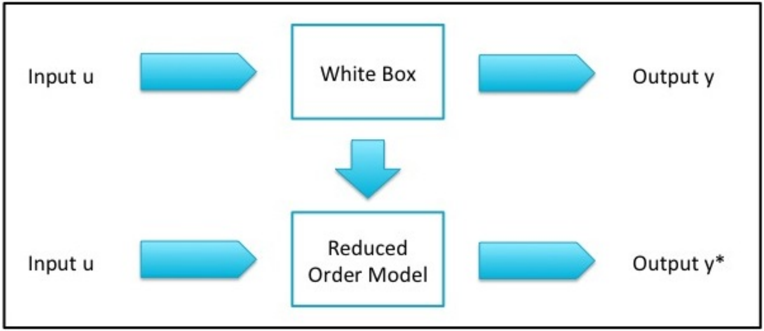
ROM

$$\dot{x}_r = f_r(x_r,u)$$
$$y_r = h_r(x_r,u)$$
$$u(t) \in \mathbb{R}^m$$
$$y_r(t) \in \mathbb{R}^p$$
$$x_r(t) \in \mathbb{R}^r$$
$$r \ll n$$

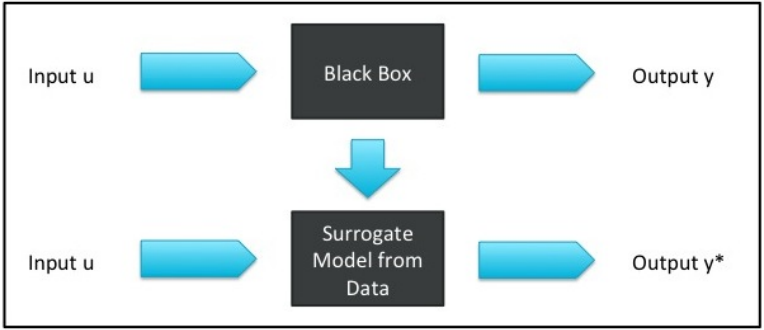
Techniques

1. IRKA,BT and Versions [Beattie, Benner, Breiten, Damm, Stykel,...]
2. POD, RB, ... [Fehr, Haasdonk, Himpe, Urban,...]
3. Löwner [Antoulas, Lefteriu, ...]

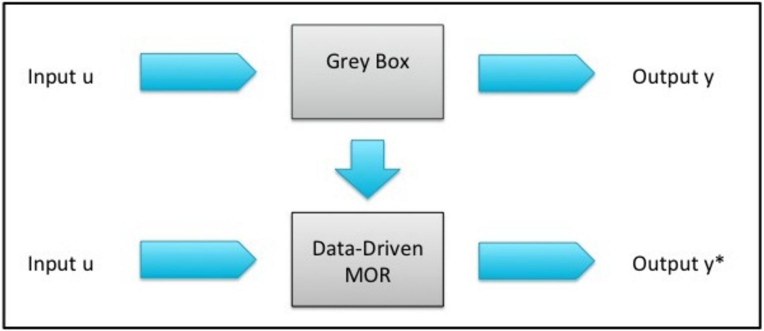
Grey Box


Input u
White Box
Output y
Input u
Reduced
Order Model
Output y*

Grey Box


Input u
Black Box
Output y
Input u
Surrogate Model from Data
Output y*

Grey Box



Overview Black Box Models

1. Interpolation
2. Response Surfaces
3. Kriging
4. Gradient-Enhanced Kriging (GEK),
5. Support Vector Machines,
6. Space Mapping,
7. Artificial Neural Networks
8. Dynamic Mode Decomposition (DMD)

DMD

Data in Time Series

$x_i \in \mathbb{R}^N$, $i = 0, \dots, n$

Assume there is an A such that

$x_{i+1} \approx Ax_i$

$$X = \begin{bmatrix} x_0 & x_1 & \dots & x_{n-1} \end{bmatrix}$$

$$Y = \begin{bmatrix} x_1 & \dots & x_{n-1} & x_n \end{bmatrix}$$

$$Y \approx AX$$

Want the eigendecomposition of A to compute time series solution fast and study behaviour

Keep in Mind

$Y = AX$

Algorithm

1. $X = U\Sigma V^*$
2. $Y = AX = AU\Sigma V^*$
   $\tilde{A} := U^* AU = U^* YV\Sigma^{-1}$
3. $\tilde{A}W = W\Lambda$
4. $\Phi = YV\Sigma^{-1}W$

$$x(t) = A^t x_0$$

$$x(t+1) \approx A\hat{x}(t) = \Phi\Lambda^t z_0$$

DMD-Variants

- Optimized DMD [K.K. Chen, J.H. Tu, and C.W. Rowley,] 2012
- Optimal Mode Decomposition [A. Wynn, D. S. Pearson, B. Ganapathisubramani and P. J. Goulart,] 2013
- Exact DMD: [Tu, Rowley, Luchtenburg, Brunton, and Kutz] 2014
- Sparsity Promoting DMD: [M.R. Jovanovic, P.J. Schmid, and J.W. Nichols] 2014
- Multi-Resolution DMD: [J.N. Kutz, X. Fu, and S.L. Brunton] 2015
- Extended DMD: [M.O. Williams , I.G. Kevrekidis, C.W. Rowley,] 2015
- DMD with Control [J.L. Proctor, S.L. Brunton, and J.N. Kutz] 2014:
- Total Least Squares DMD: [M.S. Hemati, C.W. Rowley, E.A. Deem, and L.N. Cattafesta] 2015

DMD Discussion

Pros

- Time Series Data
- Efficient and Easy
- No assumption on the model

Cons

- Linear (local)
- No error evaluation

Data Assimilation

Given

Guess of Initial State: $\hat{x}_0$
Observation: $y_0, \ldots, y_n$
Model $x_{i+1} = \mathcal{M}(x_i)$
State to Observation map $y_i = \mathcal{H}(x_i)$

Trying to find a minimum of

$$J(x) = \|x_0 - \hat{x}_0\| + \sum_i \|y_i - H(x_i)\| + \text{others}$$

such that $x_{i+1} = \mathcal{M}(x_i) + \text{others}$.

Goal

Find best starting vector and possibly update model to make the best possible forecast.

Toy Data Assimilation

Linear model

$$x_{k+1} = Mx_k$$
$$y_k = Hx_k$$

- Assumption: model is accurate
- Data: noisy measurements $y_1, \ldots, y_k$ (covariance R)
- startvalue distribution $x_0 \sim \mathcal{N}(x^b, B_0)$.

The best unbiased linear estimate for the start value is then given by minimizing the following cost functional:

$$J(x_0) = \frac{1}{2}(x_0 - x^b)^T B_0^{-1}(x_0 - x^b) + \frac{1}{2}\sum_{k=1}^{N}(Hx_k - y_k)^T R^{-1}(Hx_k - y_k)$$

Reduced Model

We assume that two projection matrices W and V are given such that we can reduce the entire system to

$$\hat{x}_{k+1} = \hat{M}\hat{x}_k \tag{1}$$

$$\hat{y}_k = \hat{H}\hat{x}_k \tag{2}$$

where $\hat{M} = W^T M V$, $\hat{H} = HV$ and $\hat{x}_0 \sim \mathcal{N}(W^T x^b, W^T B_0 W)$ and the noisy measurements stay the same with covariance R.
As before we find the best linear unbiased estimate by mimizing

$$\hat{J}(\hat{x}_0) = \frac{1}{2}(\hat{x}_0 - W^T x^b)^T \hat{B}_0^{-1}(\hat{x}_0 - W^T x^b) + \frac{1}{2}\sum_{k=1}^{N}(\hat{H}\hat{x}_k - y_k)^T R^{-1}(\hat{H}\hat{x}_k - y_k)$$

How to pick V, W Optimal Solution:

$$(B_0^{-1} + \sum_{k=1}^{N}(M^k)^T H^T R^{-1} H M^k)x_0 = (B_0^{-1}x^b + \sum_{k=1}^{N}(M^k)^T H^T R^{-1} y_k)$$

Comparison

The same would be true for the reduced system and therefore we want to compare $\|x_0 - V\hat{x}_0\|$. In order to write things more consice we define $\mathcal{H}$:

$$\mathcal{H} = \begin{bmatrix} HM & HM^2 & \dots & HM^N \end{bmatrix}$$

This means

$$\begin{aligned}
\|x_o - V\hat{x}_0\| = \|(B_0^{-1} + \mathcal{H}^T R^{-1}\mathcal{H})^{-1}(B_0^{-1}x^b + \mathcal{H}^T R^{-1} y) \\
-V(\hat{B}_0^{-1} + \hat{\mathcal{H}}^T R^{-1}\hat{\mathcal{H}})^{-1}(\hat{B}_0^{-1} W^T x^b + \hat{\mathcal{H}}^T R^{-1} y)\| \\
\leq \|(B_0^{-1} + \mathcal{H}^T R^{-1}\mathcal{H})^{-1} B_0^{-1} x^b - V(\hat{B}_0^{-1} + \hat{\mathcal{H}}^T R^{-1}\hat{\mathcal{H}})^{-1}\hat{B}_0^{-1} W^T x^b\| \\
+\|(B_0^{-1} + \mathcal{H}^T R^{-1}\mathcal{H})^{-1}\mathcal{H}^T R^{-1} y - V(\hat{B}_0^{-1} + \hat{\mathcal{H}}^T R^{-1}\hat{\mathcal{H}})^{-1}\hat{\mathcal{H}}^T R^{-1} y\|
\end{aligned}$$

Conditions

To summarize in order to get a good reduced order estimate the reduced order model needs to satisfy the following conditions:

$$VW^T x^b \approx x^b \tag{3}$$

$$\mathcal{H}V \approx \hat{\mathcal{H}} \tag{4}$$

$$B_0^{-1} V \hat{B}_0 \hat{\mathcal{H}}^T \approx \mathcal{H}^T \tag{5}$$

One could write this conditions also as

$$VW^T x^b \approx x^b \tag{6}$$

$$\mathcal{H}V \approx \hat{\mathcal{H}} \tag{7}$$

$$V\hat{B}_0 V^T = W^T B_0 W V^T \approx B_0 \tag{8}$$

$$\hat{M} = W^T M V \quad \hat{H} = HV \quad \mathcal{H} = \begin{bmatrix} HM & HM^2 & \dots & HM^N \end{bmatrix}$$

Linear Models and Notation

Problem type

Linear Input-output system:

$$\dot{x}(t) = A(\theta)x(t) + B(\theta)u(t)$$
$$y(t) = C(\theta)x(t)$$

Data:

$$y_1, \ldots, y_n$$
$$\theta_1, \ldots, \theta_n \quad u_1, \ldots, u_n$$

The world is mostly nonlinear!

Use of linear models

- Local Analysis
- Control
- Optimization

Data Driven Method

A modified challenge

Data

We assume that the data is given in form of frequencies $\xi_1, \dots, \xi_N$ and frequency response data $H_1, \dots, H_N$.

Let Y and U be the Laplace Transforms of y and u.

$$Y(\xi) = H(\xi)U(\xi) = C(\xi I - A)^{-1}BU(\xi)$$

- H_i is from measurements $(H(\xi_i) \neq H_i)$
- Matrices A, B, C, D are given but may have a model error
- Create a reduced linear system $\hat{A}, \hat{B}, \hat{C}, \hat{D}$ that is a good approximation of the true system

$\mathcal{H}_2$ Norm

$\mathcal{H}_2$ optimal MOR

Error between the true and the reduced output,

$$\operatorname{ess}\sup_{t>0} |y(t) - \hat{y}(t)| \leq \|H - \hat{H}\|_{\mathcal{H}_2} \|u\|_{\mathcal{L}_2},$$

where the $\mathcal{H}_2$-norm is defined as

$$\|H - \hat{H}\|^2_{\mathcal{H}_2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\imath\omega) - \hat{H}(\imath\omega)|^2 d\omega. \tag{9}$$

Theorem

Given a stable dynamical system. For a reduced order system of order r to minimize (9), it is necessary that at its mirror poles the reduced system be a Hermite interpolant of the original system.

$\mathcal{H}_2$ optimal MOR

Reduction via Projection

$$\hat{A} = W^T A V \quad \hat{B} = W^T B \quad \hat{C} = CV$$

If $(\sigma_i I - A)^{-1} B \in \mathrm{im}(V), \quad (\sigma_i I - A)^{-T} C^T \in \mathrm{im}(W)$, for $i = 1, \ldots, r$ the reduced transfer function interpolates the full transfer function at $\sigma_1, \ldots, \sigma_r$

Solution $\Rightarrow$ Iterative Rational Krylov Algorithm (IRKA)

IRKA

$\mathcal{H}_2$ optimal MOR

Reduction via Projection

$$\hat{A} = W^T A V \quad \hat{B} = W^T B \quad \hat{C} = CV$$

If $(\sigma_i I - A)^{-1} B \in \operatorname{im}(V), \quad (\sigma_i I - A)^{-T} C^T \in \operatorname{im}(W)$, for $i = 1, \ldots, r$ the reduced transfer function interpolates the full transfer function at $\sigma_1, \ldots, \sigma_r$

Solution $\Rightarrow$ Iterative Rational Krylov Algorithm (IRKA)

- given optimal interpolation points we can create the projection matrix

IRKA

$\mathcal{H}_2$ optimal MOR

Reduction via Projection

$$\hat{A} = W^T A V \quad \hat{B} = W^T B \quad \hat{C} = CV$$

If $(\sigma_i I - A)^{-1} B \in \mathrm{im}(V), \quad (\sigma_i I - A)^{-T} C^T \in \mathrm{im}(W)$, for $i = 1, \ldots, r$ the reduced transfer function interpolates the full transfer function at $\sigma_1, \ldots, \sigma_r$

Solution $\Rightarrow$ Iterative Rational Krylov Algorithm (IRKA)

- given optimal interpolation points we can create the projection matrix
- optimal interpolation points are not known

IRKA

$\mathcal{H}_2$ optimal MOR

Reduction via Projection

$$\hat{A} = W^T A V \quad \hat{B} = W^T B \quad \hat{C} = CV$$

If $(\sigma_i I - A)^{-1} B \in \operatorname{im}(V), \quad (\sigma_i I - A)^{-T} C^T \in \operatorname{im}(W)$, for $i = 1, \ldots, r$ the reduced transfer function interpolates the full transfer function at $\sigma_1, \ldots, \sigma_r$

Solution $\Rightarrow$ Iterative Rational Krylov Algorithm (IRKA)

- given optimal interpolation points we can create the projection matrix
- optimal interpolation points are not known
- fixed point iteration is used

IRKA

$\mathcal{H}_2$ optimal MOR

Reduction via Projection

$$\hat{A} = W^T A V \quad \hat{B} = W^T B \quad \hat{C} = CV$$

If $(\sigma_i I - A)^{-1} B \in \operatorname{im}(V), \quad (\sigma_i I - A)^{-T} C^T \in \operatorname{im}(W)$, for $i = 1, \ldots, r$ the reduced transfer function interpolates the full transfer function at $\sigma_1, \ldots, \sigma_r$

Solution $\Rightarrow$ Iterative Rational Krylov Algorithm (IRKA)

- given optimal interpolation points we can create the projection matrix
- optimal interpolation points are not known
- fixed point iteration is used
- upon convergence a local minimizer is found

Loewner

Loewner

Given frequencies together with the value of the transfer function at those frequencies, a data driven approach to MOR is to create a state space system which interpolates there. Given interpolation points $(\xi_1, \ldots, \xi_N, \sigma_1, \ldots, \sigma_r)$, and its transfer function values $W = [H(\sigma_1), \ldots, H(\sigma_r)]$ and $V^T = [H(\xi_1), \ldots, H(\xi_N)]$, we can define the Loewner matrices $\mathbb{L}, \sigma\mathbb{L}$ and the symmetric Loewner matrices $\mathbb{L}^s, \sigma\mathbb{L}^s$

$$\mathbb{L}_{ij} = \frac{V_i - W_j}{\xi_i - \sigma_j}, \qquad \sigma\mathbb{L}_{ij} = \frac{\xi_i V_i - \sigma_j W_j}{\xi_i - \sigma_j},$$

$$\mathbb{L}^s_{ij} = \begin{cases} \frac{W_i - W_j}{\sigma_i - \sigma_j} & \text{if } i \neq j \\ H'(\sigma_i) & \text{if } i = j \end{cases} \quad \sigma\mathbb{L}^s_{ij} = \begin{cases} \frac{\sigma_i W_i - \sigma_j W_j}{\sigma_i - \sigma_j} & \text{if } i \neq j \\ H(\sigma_i) + \sigma_i H'(\sigma_i) & \text{if } i = j \end{cases}.$$

Data and Model

Loewner

If $N = r$ the order r reduced state space system that interpolates:

$$\tilde{H}(s) = W(\sigma\mathbb{L} - s\mathbb{L})^{-1}V. \tag{10}$$

We will now look at this problem as a rational interpolation problem from the setup of barycentric interpolation. Here we know that the function

$$\tilde{G}(s) = \frac{\sum_{k=1}^{r} \frac{\alpha_k W_k}{s-\sigma_k}}{\sum_{k=1}^{r} \frac{\alpha_k}{s-\sigma_k} + 1}. \tag{11}$$

is a strictly proper rational function that interpolates H at σ_k for all $\alpha_1, \ldots, \alpha_k$, as long as they are not all zero.

Lemma

The two transfer functions $\tilde{H}$ and $\tilde{G}$ as in (10) and (11) are identical exactly when $\mathbb{L}\alpha + V = 0$.

Loewner

Let $N > r$, the strictly proper rational function interpolating σ_k:

$$\tilde{G}(s) = \frac{\sum_{k=1}^{r} \frac{\alpha_k W_k}{s-\sigma_k}}{\sum_{k=1}^{r} \frac{\alpha_k}{s-\sigma_k} + 1}. \tag{12}$$

We want to pick α such that $\tilde{G}(\xi_i) \approx H(\xi_i)$. Let $\mathbb{L}$ be the Loewner matrix:

$$\tilde{G}(\xi_i) - H(\xi_i) = \frac{\sum_{k=1}^{r} \frac{\alpha_k W_k}{\xi_i-\sigma_k}}{\sum_{k=1}^{r} \frac{\alpha_k}{\xi_i-\sigma_k} + 1} - V_i = \frac{(-\mathbb{L}\alpha - V)_i}{\sum_{k=1}^{r} \frac{\alpha_k}{\xi_i-\sigma_k} + 1}, \tag{13}$$

Lemma

A state space system that has the transfer function $\tilde{G}$ as in (12) with $\mathbb{L}^*\mathbb{L}\alpha + \mathbb{L}^* V = 0$ is given by: (Z denote the first r singular vectors of $\mathbb{L}$)

$$\tilde{E} = -Z^*\mathbb{L}, \quad \tilde{A} = -Z^*\sigma\mathbb{L}, \quad \tilde{C} = W, \quad \tilde{B} = Z^* V$$

Data and Model

Loewner and IRKA

- Compute σ from the inaccurate model by IRKA
- Use $\xi_1, \ldots, \xi_n, H_i, \ldots, H_n$ to determine α
- Used reduced rational function as your model

Data and Model

Offline Phase

- Compute optimal $\mathcal{H}_2$ interpolation points $\sigma_1, \ldots, \sigma_r$ from the model

- INPUT: reduced order r, interpolation points $\sigma_1, \ldots, \sigma_r$,
- Data: $\xi_1, \ldots, \xi_N, H_i, \ldots, H_N$
- Define $V_i = H_i$, Compute $W_j = H(\sigma_j)$,
- Setup $\mathbb{L}_{ij} = \frac{V_i - W_j}{\xi_i - \sigma_j}$, $\sigma\mathbb{L}_{ij} = \frac{\xi_i V_i - \sigma_j W_j}{\xi_i - \sigma_j}$
- Compute the SVD of $\mathbb{L} = U \Sigma V^T$
- $Z = U(:, 1:r)$
- $\hat{A} = -Z^* \mathbb{L}$, $\hat{E} = -Z^* \sigma \mathbb{L}$, $\hat{B} = W$, $\hat{C} = Z^* V$

Numerical Results

UQ example

Matlab Code

```
function [A,B,C,D,E]=UQexample(Q)
n=100;
A=Q*diag(-10*rand(n,1))*Q'; B=ones(n,1); C=ones(1,n); D=0; E=eye(n);
```

r	N	IRKA		DataMOR		Hermite		Loewner		
		err	#	err	#	err	#	err	#	#
4	8	0.01	142	0.02	12	0.02	8	10	(5)	8
4	16	0.01	142	0.01	20	0.02	8	0.2	(5)	16
6	12	2E-5	134	5E-5	18	2E-4	12	15	(4)	12
6	24	2E-5	134	3E-5	30	2E-4	12	4E-3	(5)	24
10	20	2E-11	151	2E-10	30	2E-5	20	–	(0)	20
10	40	2E-11	151	1E-10	50	2E-5	20	2E-6	(1)	40

Table : Average over 5 runs

Numerical Results

Synthetic Model

In benchmark collection `http://www.modelreduction.org`,

Matlab Code

```
function [A,B,C,D,E]=ParaModel(p)
n = 100;
...
Ae = spdiags(aa,0,n,n);
A0 = spdiags([0;bb],1,n,n) + spdiags(-bb,-1,n,n);
B = 2*sparse(mod([1:n],2)).';
C(1:2:n-1) = c.'; C(2:2:n) = d.'; C = sparse(C);
A=A0+p*Ae; D=0; E=eye(n);
```

Setup

Take model at $p = 1$ as the known but inexact model. Take measurements from different parameter models

Numerical Results

Synthetic Model

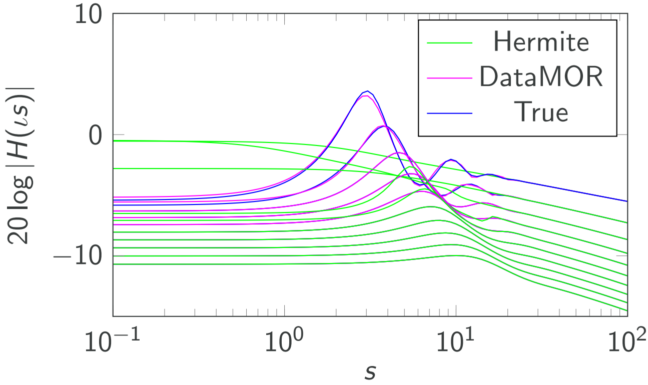


Figure : Bode plot for different values of the parameter

Numerical Results

Synthetic Model

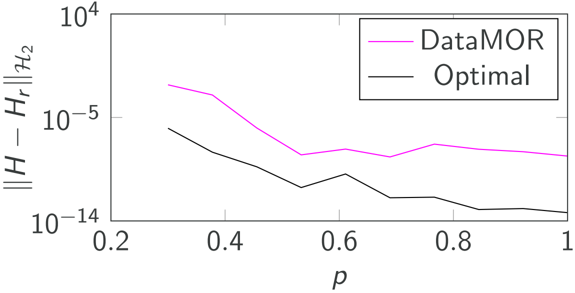


Figure : Error plot (left)

Iteration Times DataMOR: $r + N = 120$ large systems.
IRKA directly: $2rIRKAiteration = 2r20 = 800$,

Bringing it together

What are we solving?

$$\min_{\alpha} \|H_{\sigma}^{\alpha}(\xi_i) - H_i\|$$

DMD as model adjustment technique, linear correction technique

Bringing it together

What are we solving?

$$\min_{\alpha} \| H_{\sigma}^{\alpha}(\xi_i) - H_i \|$$

- Not even that exactly

DMD as model adjustment technique, linear correction technique

Bringing it together

What are we solving?

$$\min_{\alpha} \|H_{\sigma}^{\alpha}(\xi_i) - H_i\|$$

- Not even that exactly
- Does is make sense to pick the given σ?

DMD as model adjustment technique, linear correction technique

Bringing it together

What are we solving?

$$\min_{\alpha} \| H_{\sigma}^{\alpha}(\xi_i) - H_i \|$$

- Not even that exactly
- Does is make sense to pick the given σ?

$$\min_{\alpha,\sigma} \| H_{\sigma}^{\alpha}(\xi_i) - H_i \|$$

DMD as model adjustment technique, linear correction technique

Bringing it together

What are we solving?

$$\min_{\alpha} \|H_\sigma^\alpha(\xi_i) - H_i\|$$

- Not even that exactly
- Does is make sense to pick the given σ?

$$\min_{\alpha,\sigma} \|H_\sigma^\alpha(\xi_i) - H_i\|$$

$$\min_{\alpha,\sigma} \|H_\sigma^\alpha(\xi_i) - H_i + e_i\| + \|e_i\| + \|H - H_\sigma^\alpha\|_{\mathcal{H}_2}$$

DMD as model adjustment technique, linear correction technique

Questions and Remarks

Thank you for your attention.

- Peter Benner, Sara Grundel. *Model Order Reduction for a family of linear systems with applications in parametric and uncertain systems.* Applied Mathematics Letters,2015