Westfälische
Wilhelms-Universität
Münster

Data-Driven Parameter Reduction Methods for Bayesian Inverse Problems

Christian Himpe
christian.himpe@wwu.de
Mario Ohlberger
mario.ohlberger@uni-muenster.de

WWU Münster
Institute for Computational and Applied Mathematics

04.10.2012

Introduction: LTIC System

Linear Time-Invariant Control System:

$$\dot{x}(t) = A\ x(t) + Bu(t)$$
$$y(t) = Cx(t)$$
$$x(0) = x_0$$

with:

- States $x(t) \in \mathbb{R}^n$,
- system matrix $A \ \in \mathbb{R}^{n \times n}$,
- input (control) $u(t) \in \mathbb{R}^m$,
- input matrix $B \in \mathbb{R}^{n \times m}$,
- output $y(t) \in \mathbb{R}^o$,
- output matrix $C \in \mathbb{R}^{o \times n}$.

Introduction: LTIC System

Parametrized Linear Time-Invariant Control System:

$$
\begin{aligned}
\dot{x}(t) &= A_\theta x(t) + Bu(t) \\
y(t) &= Cx(t) \\
x(0) &= x_0
\end{aligned}
$$

with:

- States $x(t) \in \mathbb{R}^n$,
- system matrix $A_\theta \in \mathbb{R}^{n \times n}, \quad \theta \in \mathbb{R}^{n^2}$,
- input (control) $u(t) \in \mathbb{R}^m$,
- input matrix $B \in \mathbb{R}^{n \times m}$,
- output $y(t) \in \mathbb{R}^o$,
- output matrix $C \in \mathbb{R}^{o \times n}$.

Introduction: Bayesian Inversion

- Bayes theorem: $p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$.
- Inverse problem: $D = h(\theta, u) + \epsilon$.
- Noise ϵ assumed to be gaussian.
- Likelihood $p(D|\theta) = p_\epsilon(h(\theta, u) - D)$.
- Prior $p(\theta)$ beforehand assumption.
- Posterior $p(\theta|D) \propto p(D|\theta)p(\theta)$.

About

Setting:

- Bayesian inverse problem,
- with high-dimensional parameter space;
- underlying model being a linear dynamic system.
- Parametrization restricted to system matrix.

Aim:

- Reduce number of parameters,
- and thereby number of states.

Overview

Contents:

Optimization-Based Reduction: Concept

- Projection-based combined state and parameter reduction.
- Iteratively assemble parameter and state basis.
- Minimize maximum error between full and reduced model,
- with objective function: $\mathrm{J} = \alpha \| Y(P) - y(p) \|_2^2 + \beta (PSP^T)$
- from [1]

[1] Lieberman, C. and Willcox, K. and Ghattas, O.
Parameter and State Model Reduction for Large-Scale Statistical Inverse Problems.
SIAM Journal on Scientific Computing, 32(5): 2523-2542, 2010.

Optimization-Based Reduction: New Contribution

- Data-driven term included into objective function:
  $J = \alpha \| Y(P) - y(p) \|_2^2 + \beta(PSP^T) + \gamma \| D - y(p) \|_2^2$.
- Basis Extension by PCA ([2]).

[2] Haasdonk, B. and Ohlberger, M.
Reduced Basis Method for Finite Volume Approximations of Parametrized Linear Evolution Equations.
M2AN. Mathematical Modelling and Numerical Analysis, 42(2): 277–302, 2008.

Optimization-Based Reduction: Algorithm

1. Initialize parameter- and state-basis.
2. Integrate the full model.
3. Do
    1. Solve optimization problems.
    2. Integrate current reduced model.
    3. Orthogonalize parameters into parameter-basis.
    4. Compute PCA and orthogonalize into state-basis.

Optimization-Based Reduction: Implementation

- generic optimization,
- vectorized PCA,
- encapsulated system matrix,
- Octave and Matlab compatible.

Observability-Based Reduction: Empirical Gramian

- Based on state-to-output map: $W_o = \int_0^\infty e^{A^T\tau} C^T C e^{A\tau} d\tau$.
- Define perturbations:

$$E = \{f_i \in \mathbb{R}^n; \|f_i\| = 1; f_i f_j = 0; i = 1, \dots, n\}$$
$$K = \{l_i \in \mathbb{R}; l_i > 0; i = 1, \dots, L\}$$

- Integrate using the initial states: $x(0) = -1^j l_h f_i + \bar{x}$.
- Empirical observability gramian is given by:

$$\hat{W}_o = \frac{1}{L} \sum_{j=1}^{2} \sum_{h=1}^{L} \frac{1}{l_h^2} \int_0^\infty \Psi^{hij}(t) dt$$
$$\Psi_{ab}^{hij} = (y^{hij}(t) - \tilde{y})^* (y^{hij}(t) - \tilde{y}) \in \mathbb{R}.$$

- Introduced in [3].

[3] Lall, S. and Marsden, J.E. and Glavaski, S.
Empirical model reduction of controlled nonlinear systems.
Proceedings of the IFAC World Congress, F: 473–478, 1999.

Observability-Based Reduction: Identifiability Gramian

- Augmented system: $\begin{pmatrix} \dot{x} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} f(x,\theta) \\ 0 \end{pmatrix}$.
- Resulting observability gramian: $\tilde{W}_o = \left(\begin{array}{c|c} W_o & W_a \\ \hline W_a^* & W_b \end{array}\right)$.
- Identifiability gramian: $W_i = W_b - W_a^* W_o^{-1} W_a$.
- Introduced in [4].

[4] Geffen, D. and Findeisen, R. and Schliemann, M. and Allgower, F. and Guay, M.
Observability based parameter identifiability for biochemical reaction networks.
American Control Conference, 2130–2135, 2008.

Observability-Based Reduction: Bayesian Setting

- Initialize parameter-states in augmented system with prior mean.
- Select perturbation scales utilizing prior covariance.
- Average against PCA of experimental data.

Observability-Based Reduction: Algorithm

1. Augment system and exploit prior distribution.
2. Compute empirical observability gramian of augmented system.
3. Extract identifiability gramian.
4. Determine truncation projection utilizing the SVD.

Observability-Based Reduction: Implementation

- Vectorized matrix assembly,
- flexible scale subdivision,
- vectorized PCA,
- Octave and Matlab compatible.

Method Comparison: Theoretic Properties

	Optimization-Based	Observability-Based
Problems	Linear	Nonlinear
Computation	Complex	Simple
Parallelization	Possible	Easy
Parameter	Detached	Preserved

Numerical Experiments: Parametrization

$$
\begin{aligned}
\dot{x} &= Ax + Bu \\
y &= Cx \\
x(0) &= 0
\end{aligned}
$$

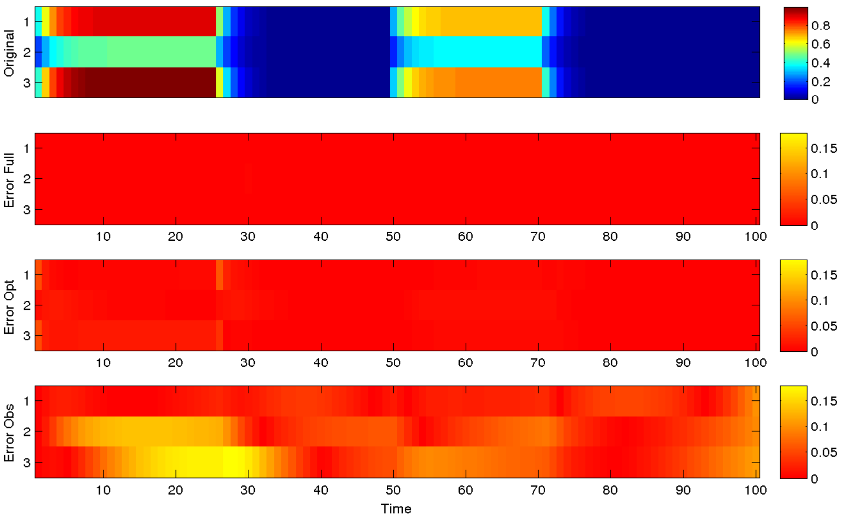
- System matrix generated randomly with ensured stability.
- Prior Distribution for diagonal elements: $N(-1, 1)$.
- Prior Distribution for off-diagonal elements: $N(0, 1)$.
- Input matrix B and output matrix C also generated randomly.

Numerical Experiments: Setup

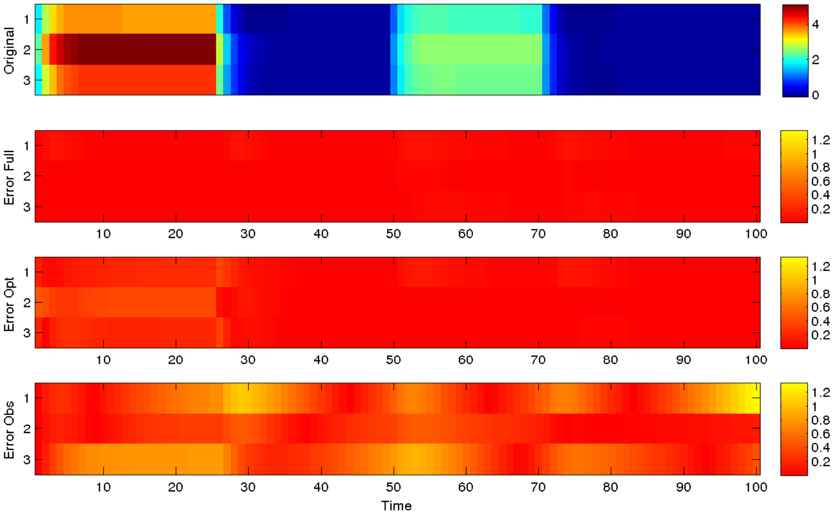
$$
\begin{aligned}
\dot{x} &= Ax + Bu \\
y &= Cx \\
x(0) &= 0
\end{aligned}
$$

- N states,
- N^2 parameters,
- 2 input sources,
- 3 outputs,
- reduced to 9 parameters (constituting 3 states),
- with: $N = 10, 40, 100$.

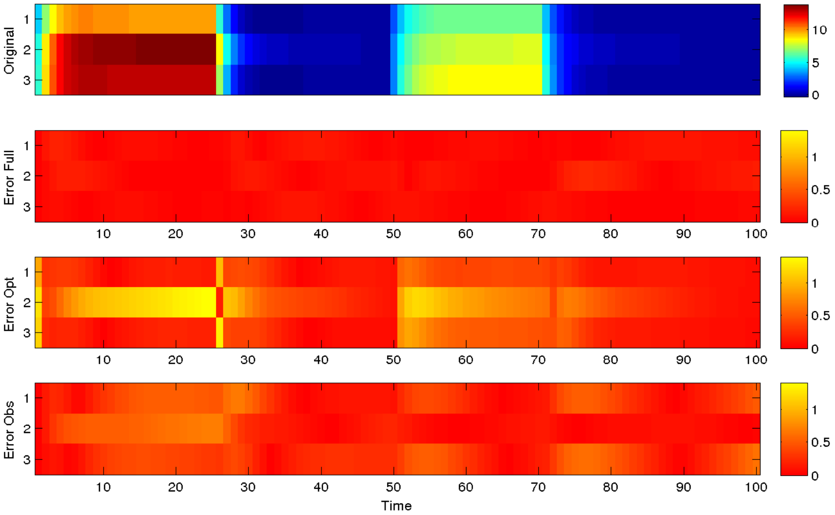
Numerical Experiments: 10 States (100 Parameters)



Numerical Experiments: 40 States (1600 Parameters)



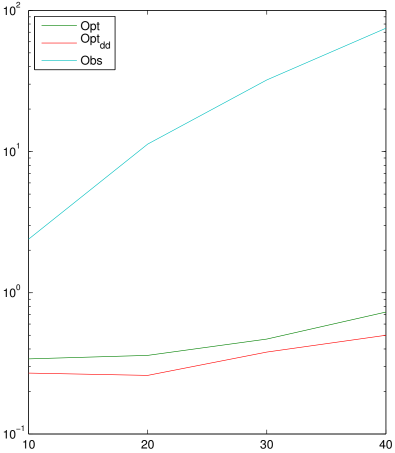
Numerical Experiments: 100 States (10000 Parameters)



Numerical Experiments: Offline Time

N	Opt	Opt_{dd}	Obs
10	0.34s	0.26s	2.40s
20	0.36s	0.26s	11.30s
30	0.47s	0.38s	32.16s
40	0.73s	0.50s	74.76s

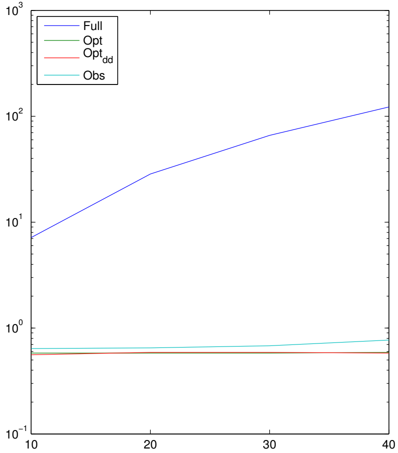
Averaged over 10 runs



Numerical Experiments: Online Time

N	Full	Opt	Opt$_{dd}$	Obs
10	7.14s	0.58s	0.56s	0.64s
20	28.53s	0.58s	0.59s	0.65s
30	66.00s	0.58s	0.59s	0.68s
40	122.35s	0.59s	0.58s	0.77s

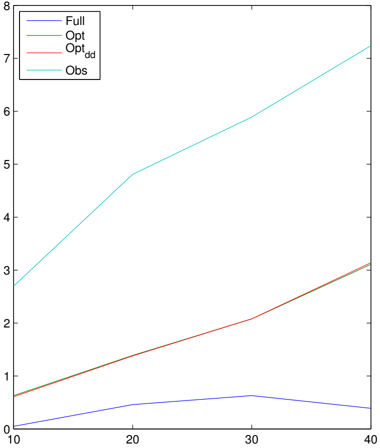
Averaged over 10 runs



Numerical Experiments: L2-Error

N	Full	Opt	Opt_{dd}	Obs
10	0.05	0.63	0.61	2.70
20	0.46	1.39	1.38	4.82
30	0.63	2.08	2.08	5.89
40	0.39	3.11	3.14	7.24

Averaged over 10 runs



Method Comparison: Applied Properties

	Optimization-Based	Observability-Based
Problems	Linear	Nonlinear
Computation	Complex	Simple
Parallelization	Possible	Easy
Parameter	Detached	Preserved
Offline Time	Low	Growing
Online Time	Constant Low	Low
Error	Low	Medium

tl;dl

Outlook:

- Optimized integration (Both).
- Nonlinear problems (Optimization-Based).
- Parameter independent state reduction (Observability-Based).

Summary:

- Optimization-Based method faster.
- Observation-Based method more flexible.

Sourcecode:

- `http://j.mp/morepas`

Thank you!