Model Order Reduction

(for ODE-Constrained Optimization)

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

WWU Münster
Institute for Computational and Applied Mathematics

DCM Workshop
24.04.15

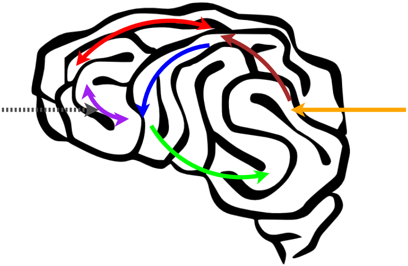
Disclaimer

- The presented methods are subject of ongoing research.
- I am a mathematician; there will be math!

Outline

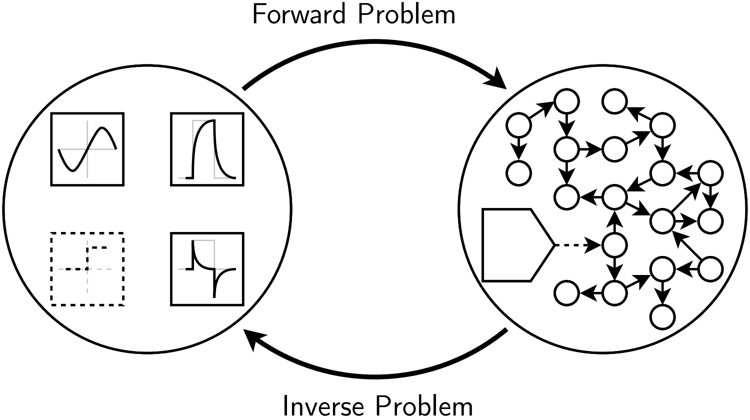
Motivation

- How are regions of the brain connected?
- How does (sensory) input disperse?
- How does connectivity change under input?
- How does the brain learn and unlearn?



Procedure

1. Experimental Data
2. Forward Model
3. Inverse Problem



Notation

- $x(t)$ State Trajectory
  aka Neuronal Activity
- $y(t)$ Output Trajectory
  aka Measured Response
- $u(t)$ Input / Control
  aka External Stimulus
- θ Parameters
  aka Connectivity Strength

Act I

Network Models

Ordinary Differential Equations

Initial Value Problem (IVP) w Ordinary Differential Equation[1] (ODE):

$$\dot{x}(t) = ax(t)$$
$$x(0) = x_0$$

Components:

- $x : \mathbb{R}^+ \to \mathbb{R}$ Solution Trajectory
- $\dot{x} = \frac{dx}{dt}$ Newton Notation for a Time Derivative
- $x_0 \in \mathbb{R}$ Initial Value
- $a \in \mathbb{R}$

Solution:

$$x(t) = e^{at}x_0$$

[1] Here: Autonomous & Linear

Systems of ODEs

IVP with a System of ODEs:

$$\dot{x}(t) = Ax(t)$$
$$x(0) = x_0$$

Components:

- $x : \mathbb{R}^+ \rightarrow \mathbb{R}^N$ Solution Trajectory
- $\dot{x}_i = \frac{dx_i}{dt}$ Component-Wise Derivative
- $x_0 \in \mathbb{R}^N$ Initial Value
- $A \in \mathbb{R}^{N \times N}$

Solution:

$$x(t) = e^{At}x_0$$

Control System

Linear Control System[2]:

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

Components:

- $y : \mathbb{R}^N \to \mathbb{R}^O$ Output Trajectory
- $A \in \mathbb{R}^{N \times N}$ System Matrix
- $B \in \mathbb{R}^{N \times M}$ Input Matrix
- $C \in \mathbb{R}^{O \times N}$ Output Matrix

Solution:

$$y(t) = (x * u)(t) = Ce^{At}x_0 + \int_0^t Ce^{A\tau}Bu(\tau)d\tau$$

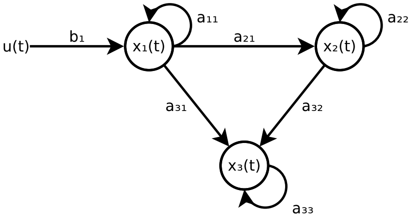
[2] in State-Space Form

Network Interpretation

Linear Dynamical System:

$$\dot{x}(t) = Ax(t) + Bu(t)$$

Example (single input, three region network):



$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} u(t)$$

Connectivity Parametrization

Connectivity Strength:

$$\theta_{iN+j} = a_{ij}$$

Parametrized Linear Dynamical System:

$$\dot{x}(t) = A(\theta)x(t) + Bu(t)$$

with Nonlinear Parameter Mapping.

General Control System

Possibly Nonlinear Control System:

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = g(x(t), u(t), \theta)$$
$$x(0) = x_0$$

(Special Case) Linear Control System:

$$f(x(t), u(t), \theta) = A(\theta)x(t) + Bu(t)$$
$$g(x(t), u(t), \theta) = Cx(t)$$

(Nonlinear Example) Hyperbolic Network Model:

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

Connectivity Inference

Data Model

Model-Data Relation:

$$y_d := y_\theta + \varepsilon$$

Components:

- y_d Measured Output
- y_θ Parametrized Model Output
- ε Noise

Inverse Problem

ODE Constrained Optimization:

$$
\begin{aligned}
\theta_d &= \text{argmin}_\theta \|y_\theta - y_d\|_2^2 \\
\text{s.t.:} & \\
\dot{x}(t) &= f(x(t), u(t), \theta) \\
y_\theta(t) &= g(x(t), u(t), \theta) \\
x(0) &= x_0
\end{aligned}
$$

- $\|\cdot\|_2$ means L_2 (Euclidian Distance).
- This is a least-squares minimization.

Difficulties

Well-Posed Problem:

- A solution exists
- The solution is unique
- Solution is stable

A problem not well-posed is ill-posed.

Ill-Conditioned Problem:

- Small Perturbation result in large errors

Regularization

Tikhonov Regularization:

$$\theta_d = \text{argmin}_\theta \|y_\theta - y_d\|_2^2 + \beta\|\theta\|_2^2$$

New Components:

- β Regularization Coefficient
- $\|\theta\|_2^2$ Regularization Operator

Dynamic Causal Modelling

Two-Part Model [Friston et al'03]

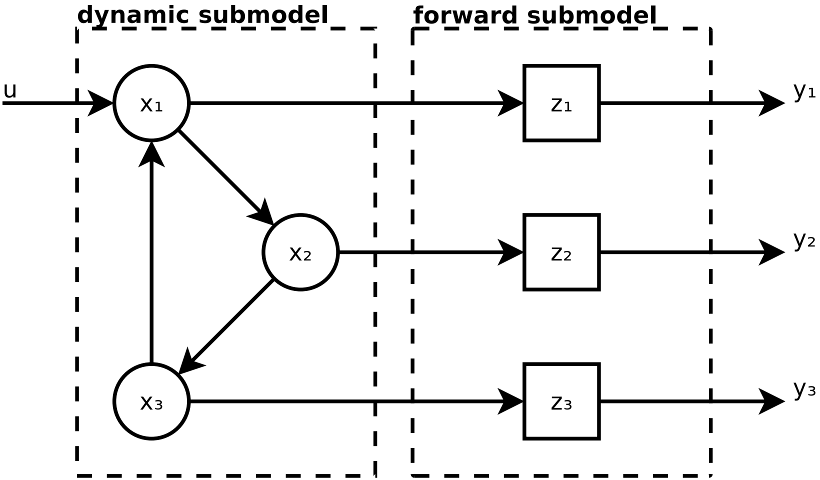
Dynamic Sub-Model:

- Models neuronal activity,
- and coupling between different brain regions
- Multiple-Input-Multiple-Output

Forward Sub-Model:

- Transforms neuronal activity to measurable output
- Physiologically motivated
- Single-Input-Single-Output

Schematic Two-Part Model



Model Properties

Models differ for

- EEG / MEG
- fMRI / fNIRS

but have commonalities:

Both ...

- exhibit stable behaviour,
- (originally) contain nonlinearities,
- encode connectivity in parameters,
- are assumed to be deterministic.

A Closer Look at ...

... the fMRI Model,

because:

- the dynamic sub-model is easier to understand,
- and has less physiological assumptions.

Dynamic Sub-Model

- Models neuronal activity
- Hidden from (direct) measurement
- Most likely nonlinear

- Encodes coupling
- Input (External Stimulus)
- (Connectivity) Parameters

$$\dot{x}(t) = f(x(t), u(t), \theta)$$

Linear(ized) Model

Using a Taylor series approximation:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t), \theta) \\ &\approx f(0, 0, \theta) + \frac{\mathrm{d}f}{\mathrm{d}x}x(t) + \frac{\mathrm{d}f}{\mathrm{d}u}u(t)\end{aligned}$$

$$\Rightarrow \dot{\tilde{x}}(t) = A(\theta)\tilde{x}(t) + B(\theta)u(t)$$

- Models effective connectivity
- Parameters θ are the components of A, B
- Stability constraints apply to θ

Bilinear Extension

Bilinear approximation [Friston et al'03]:

$$
\begin{aligned}
\dot{x}(t) &= f(x(t), u(t), \theta) \\
&\approx f(0, 0, \theta) + \frac{\mathrm{d}f}{\mathrm{d}x} x(t) + \frac{\mathrm{d}f}{\mathrm{d}u} u(t) + \frac{\mathrm{d}^2 f}{\mathrm{d}x \mathrm{d}u} x(t) u(t)
\end{aligned}
$$

$$
\Rightarrow \dot{\tilde{x}}(t) = A(\theta)\tilde{x}(t) + B(\theta)u(t) + \sum_i u_i(t) G_i(\theta)\tilde{x}(t)
$$

$G_i(\theta)$ describes influence of i-th external input on coupling strength (lateral connectivity).

Quadratic Extension

Quadratic approximation [Stephan et al'08]:

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$\approx f(0,0,\theta) + \frac{\mathrm{d}f}{\mathrm{d}x}x(t) + \frac{\mathrm{d}f}{\mathrm{d}u}u(t) + \frac{\mathrm{d}^2 f}{\mathrm{d}x\mathrm{d}u}x(t)u(t) + \frac{\mathrm{d}^2 f}{\mathrm{d}x\mathrm{d}x}x(t)x(t)$$

$$\Rightarrow \dot{\tilde{x}}(t) = A(\theta)\tilde{x}(t) + Bu(t) + \sum_i u_i(t)G_i(\theta)\tilde{x}(t) + \sum_j \tilde{x}_j(t)H_j(\theta)\tilde{x}(t)$$

$H_j(\theta)$ describes influence of j-th state on coupling strength.

Forward-Submodel [Friston'02]

- Transforms neuronal activity to an observable BOLD signal
- Is a nonlinear SISO system

$$\dot{s}_i(t) = \kappa_x x_i(t) - \kappa_s s_i(t) - \kappa_f (1 - f_i(t))$$

$$\dot{f}_i(t) = s_i(t)$$

$$\dot{v}_i(t) = \frac{1}{\kappa_0}(f_i(t) - v_i(t)^{\frac{1}{\alpha}})$$

$$\dot{q}_i(t) = \frac{1}{\kappa_0}(f_i(t)\frac{E(f_i(t),\kappa_\rho)}{\kappa_\rho} - v_i(t)^{\frac{1}{\alpha}}\frac{q_i(t)}{v_i(t)})$$

$$y_i(t) = k_1(1 - v_i(t)) + k_2(1 - q_i(t)) + k_3(1 - \frac{q_i(t)}{v_i(t)}))$$

- Activity induced signal
- Inflow
- Venous volume (inflow - outlow)
- Deoxy. Content (intake - release)
- BOLD Output (volume + content + concentration)

Joint Model

Combining the

- MIMO dynamic sub-model
- SISO forward sub-model

yields a joint nonlinear state-space system:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{z}_1(t) \\ \vdots \\ \dot{z}_n(t) \end{pmatrix} = \begin{pmatrix} F_{dyn}(x(t), u(t), \theta) \\ F_{out,1}(z_1(t), x_1(t)) \\ \vdots \\ F_{out,n}(z_n(t), x_n(t)) \end{pmatrix}$$

$$y = g(z_1(t), \dots, z_n(t))$$

Bayesian Inference

Bayes' Rule:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

- $P(\theta|y_d)$ Posterior (Probability of θ given y_d)
- $P(y_d|\theta)$ Likelihood (Probability of y_d given θ)
- $P(\theta)$ Prior (Probability of θ)
- $P(y_d)$ Evidence (Probability of y_d)

Proportionality:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

Gaussian Setting

In case the prior and the noise are gaussian, $P(\theta) = \mathcal{N}(\kappa, K),\ P(\varepsilon) = \mathcal{N}(0, \Lambda)$:

$$\begin{aligned} y_d &= y_\theta + \varepsilon \\ \Rightarrow \epsilon &= y_d - y_\theta \\ \Rightarrow P(\epsilon) &= P(y_d - y_\theta) \\ \Rightarrow P(\theta|y_d) &\propto P(y_d - y_\theta)P(\theta) \end{aligned}$$

$$\begin{aligned} P(y_d|\theta) &\propto \exp(-\frac{1}{2}\|y_\theta - y_d\|^2_{\Lambda^{-1}}) \\ P(\theta) &\propto \exp(-\frac{1}{2}\|\theta - \kappa\|^2_{K^{-1}}) \\ \Rightarrow P(\theta|y_d) &\propto \exp(-\frac{1}{2}\|y_\theta - y_d\|^2_{\Lambda^{-1}} - \frac{1}{2}\|\theta - \kappa\|^2_{K^{-1}}), \end{aligned}$$

Maximum-A-Posteriori

MAP Estimator (Prior incorporated maximum likelihood estimator):

$$\begin{aligned}\theta_{\mathsf{MAP}} &= \mathsf{argmax}_{\theta\in\mathbb{R}^P}\ \exp\left(-\frac{1}{2}\|y_\theta - y_d\|^2_{\Lambda^{-1}} - \frac{1}{2}\|\theta-\kappa\|^2_{K^{-1}}\right)\\ &= \mathsf{argmin}_{\theta\in\mathbb{R}^P}\left(\frac{1}{2}\|y_\theta - y_d\|^2_{\Lambda^{-1}} + \frac{1}{2}\|\theta-\kappa\|^2_{K^{-1}}\right).\end{aligned}$$

At second glance this is a regularized least-square problem!

Large-Scale Models

The dynamic sub-model (dynamical system):

$$\dot{x}(t) = A(\theta)x(t) + Bu(t)$$

dimensions determine the paramter space dimension: $P = N^2$.

The minimization algorithm for such a nonlinear problem computes many simulations of the system, due to the necessary perturbations of many directions in the parameter space.

$\Rightarrow$ A few more nodes in the network may prolong the inversion procedure significantly.

Model Reduction

Model Reduction

System:

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = g(x(t), u(t), \theta)$$
$$x(0) = x_0$$

Setting:

- $\dim(x(t)) \gg 1$ (Many Network Nodes / Brain Regions)
- $\dim(u(t)) \ll \dim(x(t))$ (Significantly Less Inputs)
- $\dim(y(t)) \ll \dim(x(t))$ (Significantly Less Outputs)
- $\dim(\theta) \gg 1$ (Many Parameters)

Aim:

- $\dim(x_r(t)) \ll \dim(x(t))$
- $\dim(\theta_r) \ll \dim(\theta)$
- $\|y_\theta - y_{r,\theta_r}\| \ll 1$

Projection-Based Model Reduction

Trajectory Projection:

$$x_r(t) := Ux(t)$$
$$x(t) \approx Vx_r(t)$$

Petrov-Galerkin Projection:

$$U \in \mathbb{R}^{n \times N}, \quad V \in \mathbb{R}^{N \times n}, \quad V^T U = \mathbb{1}, \quad n \ll N$$

Galerkin Projection:

$$U \in \mathbb{R}^{n \times N}, \quad V := U^T, \quad U^T U = \mathbb{1}, \quad n \ll N$$

- We will only be concerned with Galerkin projection.
- Petrov-Galerkin can pose issues with stability.

State-Space Reduction

General Control System:

$$
\begin{aligned}
\dot{x}(t) &= f(x(t), u(t), \theta) \\
y(t) &= g(x(t), u(t), \theta) \\
x(0) &= x_0
\end{aligned}
$$

Projection-Based Reduced Order Model (ROM):

$$
\begin{aligned}
\dot{x}_r(t) &= Vf(Ux_r(t), u(t), \theta) \\
y_r(t) &= g(Ux_r(t), u(t), \theta) \\
x_r(0) &= Vx_0
\end{aligned}
$$

Aim: $\|y - y_r\| \ll 1$

Linear State-Space Reduction

Linear Control System:

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

Projection-Based ROM:

$$\begin{aligned} \dot{x}_r(t) &= UAVx_r(t) + UBu(t) \\ y_r(t) &= CVx_r(t) \\ x_r(0) &= Vx_0 \end{aligned}$$

(Side Note) PCA / POD / SVD

You may already have done model reduction!

Assume:

- given a discrete time series,
- to which a PCA is applied.

This is more or less a centered POD method of snapshots.

For finite dimensional operators PCA and POD are essentially a (sparse) SVD.

Parameter Identification

Parameter Space:

$$\theta \in \mathbb{R}^P, \quad P \gg 1$$

i.e. $P = N^2$.

- Which (linear combination of the) parameter is influencing the behaviour of the system the most?
- This is also related to sensitivity analysis.

Parameter (Galerkin) Projection:

$$\theta_r := \Pi\theta$$
$$\theta \approx \Pi^T \theta_r$$
$$\Pi^T \Pi = \mathbb{1}$$

Parameter-Space Reduction

General Control System:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t), \theta)\\ y(t) &= g(x(t), u(t), \theta)\\ x(0) &= x_0\end{aligned}$$

Projection-Based ROM:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t), \Pi^T \theta_r)\\ y(t) &= g(x(t), u(t), \Pi^T \theta_r)\end{aligned}$$

Aim: $\|y_\theta - y_{\theta_r}\| \ll 1$

Combined State and Parameter Reduction

General Control System:

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = g(x(t), u(t), \theta)$$

Reduced Order Model:

$$\dot{x}_r(t) = f_r(x_r(t), u(t), \theta_r)$$
$$y_r(t) = g_r(x_r(t), u(t), \theta_r)$$

Projection-Based ROM:

$$\dot{x}_r(t) = U^T f(Ux_r(t), u(t), \Pi^T \theta_r)$$
$$y_r(t) = g(Ux_r(t), u(t), \Pi^T \theta_r)$$

Aim: $\|y_\theta - y_{r,\theta_r}\| \ll 1$

Challenge

- How to find U? (MOR)
- How to find Π? (SYSID)
- Since $y(\theta)$, y_r has to be valid for all admissable θ! (pMOR)
- Since $y(\theta)$, θ_r has to approximate θ well! (COMRED)
- FYI: My Models are nonlinear! (nMOR)
- BTW: I have non-affine parameter dependencies!

Dual Approach

1. Gramian-Based Combined Reduction
2. Optimization-Based Combined Reduction

Gramian-Based Combined Reduction

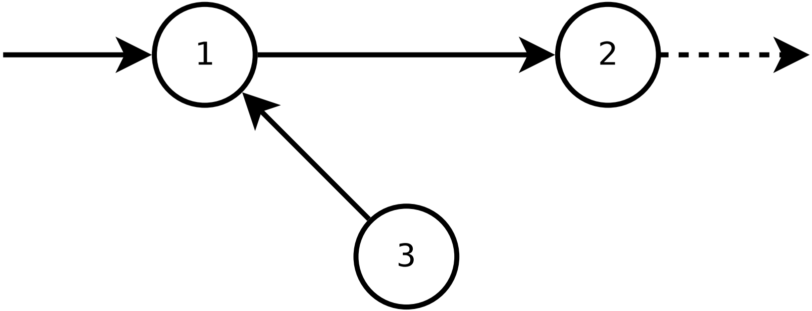
Based on ...

- System Theory / Control Theory
- Linear Control Systems and their encoded properties

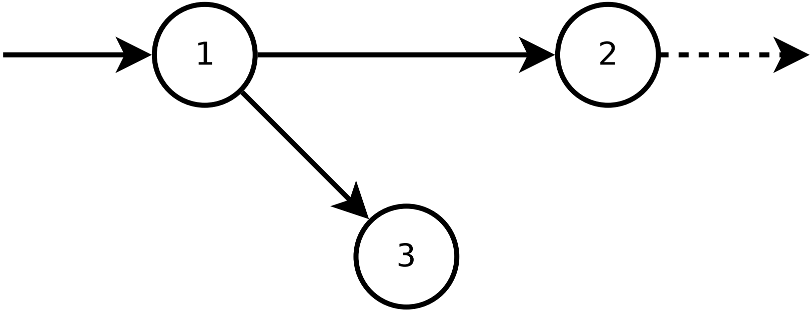
Features:

- For Nonlinear Systems: Empirical Gramians [Lall et al'99]
- Combined Reduction: Empirical Cross Gramian and Joint Gramian [H. & Ohlberger'14]

Controllability



Observability



Balanced Truncation [Moore'81]

For Linear Control Systems

- Controllability and Observability can be computed
- as singular values of the System's Gramian Matrices.
- Balancing these two matrices yields the so called Hankel Singular Values[3].

Why HSVs?

- A state component that is neither controllable nor observable
- is not contributing to the input-to-output energy transfer.
- The smaller the HSV, the less important the associated (balanced) state is.

---

[3] Singular values of the Hankel operator mapping inputs to outputs.

Parameter Observability

Parameter Augmented General Control System:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{\theta}(t) \end{pmatrix} = \begin{pmatrix} f(x(t), u(t), \theta(t)) \\ 0 \end{pmatrix}$$

$$y(t) = g(x(t), u(t), \theta(t))$$

$$\begin{pmatrix} x(0) \\ \theta(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ \theta_0 \end{pmatrix}$$

Optimization-Based Combined Reduction

Based on:

- Greedy Algorithm
- Large-Scale Inverse Problems

Features:

- Combined Reduction: [Lieberman et al'12]
- Data-Driven: [H. & Ohlberger (submitted)]

Greedy Algorithm

Minimize Maximal Error:

$$\begin{aligned}\theta_{i+1} &= \text{argmax}_{\theta \perp \theta_{0,\dots i}} \|y_\theta - y_{\theta_r}\|_2^2 + \gamma\|\theta\|_2^2 \\ &= \text{argmin}_{\theta \perp \theta_{0,\dots i}} -\|y_\theta - y_{\theta_r}\|_2^2 - \gamma\|\theta\|_2^2\end{aligned}$$

Parameter Projection:

$$\Pi = [\theta_0, \dots, \theta_p]$$

Enhanced Greedy Algorithm

Monte-Carlo Parameter Base & Data-Driven Regularization:

$$M = [P(\theta)_0, \ldots, P(\theta)_p]$$
$$\tilde{\theta}_{i+1} = \text{argmax}_{M\theta \perp M\theta_{0,\ldots i}} \|y_{\tilde{\theta}} - y_{\tilde{\theta}_r}\|_2^2 + \gamma \|\tilde{\theta}\|_2^2 + \delta \|y_d - y_{\tilde{\theta}_r}\|_2^2$$

Parameter Projection:

$$\Pi = [M_0^{-1}\tilde{\theta}_0, \ldots, M_p^{-1}\tilde{\theta}_p]$$

Combined Reduction

Parameter Greedy:

$$\theta_{i+1} = \text{argmax}_{\theta \perp \theta_{0,\ldots i}} \|y_\theta - y_{r,\theta_r}\|_2^2 + \gamma \|\theta\|_2^2$$
$$\bar{x}_i = \text{pod}_1(x(\theta_i))$$

State Projection:

$$U = [\bar{x}_0, \ldots, \bar{x}_n]$$

Alltogether

Back to the Beginning

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = g(x(t), u(t), \theta)$$
$$x(0) = x_0$$

Dimensions:

- $\dim(x(t)) \gg 1$
- $\dim(u(t)) \ll \dim(x(t))$
- $\dim(y(t)) \ll \dim(x(t))$
- $\dim(\theta) \gg 1$

Inverse Problem

ODE Constrained Optimization:

$$
\begin{aligned}
\theta_d &= \operatorname{argmin}_\theta \|y_\theta - y_d\|_2^2 \\
\text{s.t.:} & \\
\dot{x}(t) &= f(x(t), u(t), \theta) \\
y_\theta(t) &= g(x(t), u(t), \theta) \\
x(0) &= x_0
\end{aligned}
$$

Remember:

- $\dim(x(t)) \gg 1$
- $\dim(u(t)) \ll \dim(x(t))$
- $\dim(y(t)) \ll \dim(x(t))$
- $\dim(\theta) \gg 1$

Reduced Order Inverse Problem

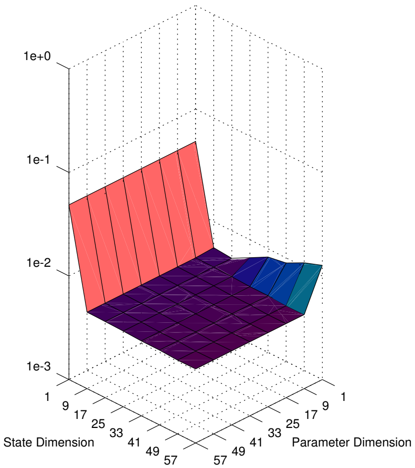
ODE Constrained Optimization:

$$
\begin{aligned}
\theta_d &= \operatorname{argmin}_{\theta_r} \|y_{\theta_r} - y_d\|_2^2 \\
\text{s.t.:} & \\
\dot{x}_r(t) &= f_r(x_r(t), u(t), \theta_r) \\
y_{r,\theta}(t) &= g_r(x_r(t), u(t), \theta_r) \\
x_r(0) &= x_{r,0}
\end{aligned}
$$

Remember:

- $\dim(x_r(t)) \ll \dim(x(t))$
- $\dim(\theta_r) \ll \dim(\theta)$
- $\|y_\theta - y_{r,\theta_r}\| \ll 1$

Numerical Results (HNM)



Gramian-Based and Optimization-Based Combined Reduction

What does this mean?

- State-Space dynamics can be bound to low-dimensional sub-spaces of the high-dimensional state-space.
- Identifiable parameters can be restricted to small sub-spaces of the high-dimensional parameter-space.
- State- and parameter-spaces can be reduced jointly,
- also for a nonlinear system.
- The inverse problem can be solved on the reduced spaces.
- Open issue: accurate parameter reconstruction

tl;dl

- Networks can be modelled by control systems
- In this sense, the parameter inference is an ODE constrained inverse problem
- DCM is a flavor of such inverse problem in a bayesian setting
- Model Reduction approximates large models with smaller surrogate models
- and thus accelerates the inversion / optimization

More Info:

- Me: `http://wwwmath.uni-muenster.de/u/himpe`
- M. Ohlberger: `http://wwwmath.uni-muenster.de/u/ohlberger`
- MoRePaS: `http://morepas.org`
- MORwiki: `http://modelreduction.org`

Thanks!