Westfälische Wilhelms-Universität Münster

Applied Mathematics Münster

Combined Reduction for Neural Networks

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

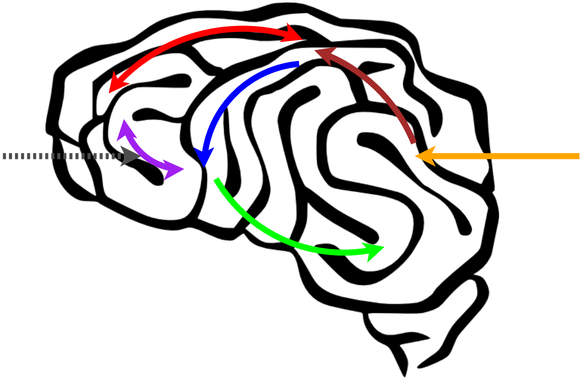
WWU Münster
Institute for Computational and Applied Mathematics

12.11.2013

Overview

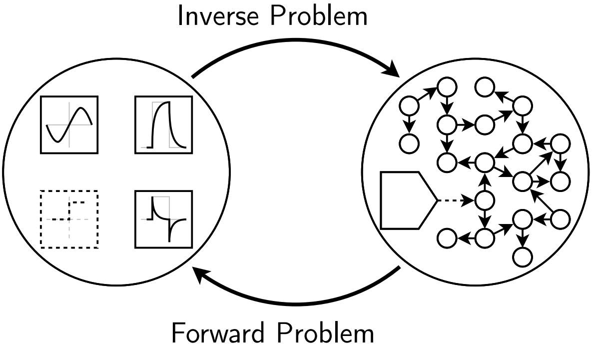
Motivation

- How are brain regions connected?
- How does sensory input disperse?
- How is connectivity altered under external influence?
- How does the brain learn and unlearn?



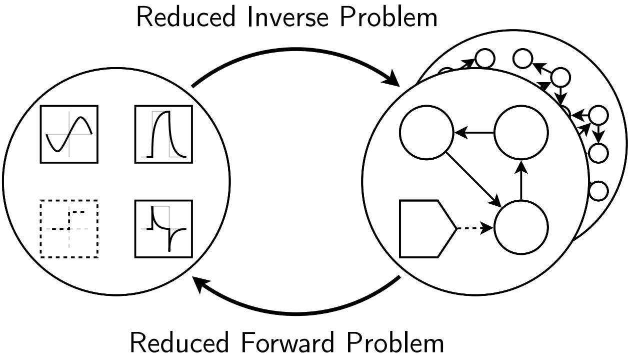
Application

- Experiments with Controlled Input
  - EEG / MEG
  - fMRI / fNIRS
- Inverse Problem



Application

- Experiments with Controlled Input
  - EEG / MEG
  - fMRI / fNIRS
- Model Reduction
- Inverse Problem



Model

General control system:

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

with:

- Input $u \in \mathbb{R}^m$
- State $x \in \mathbb{R}^n$
- Output $y \in \mathbb{R}^o$
- Parameters $\theta \in \mathbb{R}^p$

Model

Linear control system:

$$\dot{x} = Ax + Bu$$
$$y = Cx + Du$$

with:

- Input $u \in \mathbb{R}^m$
- State $x \in \mathbb{R}^n$
- Output $y \in \mathbb{R}^o$
- Parameters $\theta \in \mathbb{R}^p$

Model

(exemplary) Parametrized linear control system:

$$\dot{x} = A_\theta x + Bu$$
$$y = Cx$$

with:

- Input $u \in \mathbb{R}^m$
- State $x \in \mathbb{R}^n$
- Output $y \in \mathbb{R}^o$
- Parameters $\theta \in \mathbb{R}^p$

Model

(exemplary) Nonlinear control system:

$$\dot{x} = Af(x) + Bu$$
$$y = Cx$$

with:

- Input $u \in \mathbb{R}^m$
- State $x \in \mathbb{R}^n$
- Output $y \in \mathbb{R}^o$
- Parameters $\theta \in \mathbb{R}^p$

Reduced Order Model

For a parametrized linear system:

$$\begin{aligned}\dot{x} &= A(\theta)x + Bu \\ y &= Cx,\end{aligned}$$

a state reduced system is given by:

$$\begin{aligned}\dot{\tilde{x}} &= \tilde{A}(\theta)\tilde{x} + \tilde{B}u \\ \tilde{y} &= \tilde{C}\tilde{x}.\end{aligned}$$

Reduced Order Model

For a parametrized linear system:

$$\dot{x} = A(\theta)x + Bu$$
$$y = Cx,$$

a parameter reduced system is given by:

$$\dot{x} = A(\tilde{\theta})x + Bu$$
$$\tilde{y} = Cx.$$

Reduced Order Model

For a parametrized linear system:

$$\dot{x} = A(\theta)x + Bu$$
$$y = Cx,$$

a combined reduced system is given by:

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

Combined (State and Parameter) Reduction

- Situation: $\dim(u) \ll \dim(x) \wedge \dim(y) \ll \dim(x) \wedge \dim(x) \gg 1$
- Sufficient: $y = h(t, u)$,
- Available: $y = g(x(t, u))$,
- Aim: $\dim(\tilde{x}) \ll \dim(x) \wedge \dim(\tilde{\theta}) \ll \dim(\theta) \wedge \|y - \tilde{y}\| \ll 1$,
- How: project state & parameter spaces to dominant subspaces,
- Using: Galerkin or Petrov-Galerkin projections.

Challenge: Find a $\begin{cases} \text{state} \\ \text{parameter} \end{cases}$ projection efficiently.

Combined (State and Parameter) Reduction

- Situation: $\dim(u) \ll \dim(x) \wedge \dim(y) \ll \dim(x) \wedge \dim(x) \gg 1$
- Sufficient: $y = h(t, u)$,
- Available: $y = g(x(t, u))$,
- Aim: $\dim(\tilde{x}) \ll \dim(x) \wedge \dim(\tilde{\theta}) \ll \dim(\theta) \wedge \|y - \tilde{y}\| \ll 1$,
- How: project state & parameter spaces to dominant subspaces,
- Using: Galerkin or Petrov-Galerkin projections.

Challenge: Find a $\begin{cases} \text{state} \\ \text{parameter} \end{cases}$ projection efficiently. **Accepted!**

Optimization-Based Combined Reduction

Concept:

1. Iterative computation of projection base.
2. Each parameter base vector is determined by optimization.
3. From this the state projection base is constructed.

Orthogonal Projections:

- Parameter projection P,
- State projection V.

Prior Information can be used as initial value.

Greedy Sampling[2]

Maximize error between full-order and reduced-order model output:

$$\theta_I = \operatorname{argmax} J(\theta) = \frac{1}{2}\|\tilde{y}(\tilde{\theta}) - y(\theta)\|_2^2 + \frac{\alpha}{2}\|\tilde{\theta}\|_{S^{-1}}^2,$$

with:

$$\|\theta\|_{S^{-1}}^2 = \theta^T S^{-1} \theta.$$

θ_I is next parameter base vector also used to integrate $\dot{x}$:

$$x(t, \theta_I) = \int_0^t e^{A(\theta_I)\tau} B u(\tau) d\tau$$

Next state base vector is computed from state time series[1] $x(t, \theta_I)$:

$$\overline{x(\theta_I)} = \begin{cases} \sigma(x(t, \theta_I)) \\ \frac{1}{t}\int_0^t x(t, \theta_I) dt \end{cases}$$

---

[1] see [Bashir'08] and [Lall'99]

[2] Introduced by [Lieberman'10]

Algorithm[3]

$\theta_0 \leftarrow \theta_{\text{prior}}$

$P \leftarrow \theta_0$

$V \leftarrow \overline{x(\theta_0)}$

for $I = 1 : q$

- $\theta_I \leftarrow \text{argmax}(J(\theta_{I-1}))$
- $a \leftarrow V^T A(\theta_{I-1}) V$
- $b \leftarrow V^T B$
- $c \leftarrow CV$
- $P \leftarrow \text{orth}([P, \theta_I])$
- $V \leftarrow \text{orth}([V, \overline{x(\theta_I)}])$

end for

[3] see [Bashir'08], [Lieberman'10]

Data-Driven Reduction[4]

- Full-order integration $y(\theta)$ is expensive.
- As an inverse problem, data y_d is available:

$$\theta_I = \operatorname{argmax} \frac{1}{2}\|\tilde{y}(\tilde{\theta}) - y_d\|_2^2 + \frac{\alpha}{2}\|\tilde{\theta}\|_{S^{-1}}^2$$

- PRO: Offline time is reduced.
- CON: Reduced model is only valid for specific data.

[4] [Himpe (In Preparation)]

Trust-Region Reduction[5]

- Each iterations optimization is a costly operation.
- The parameter space can be expanded iteratively,
- and trust-region-like: $\dim(\theta_0) = 1, \ \dim(\theta_I) = I + 1$;

together with a projection:

$$\varphi : \mathbb{R}^{I+1} \to \mathbb{R}^p,$$

for the incorporation of new parameter base vectors:

$$P \leftarrow \mathrm{orth}([P, \varphi(\theta_I)]).$$

- PRO: Massive offline time reduction.
- CON: Higher reduction error.

---

5 [Himpe (In Preparation)]

Further Enhancements[6]

Greedy Sampling using:

- 1-Norm
- 2-Norm
- ∞-Norm

Orthogonalization by:

- Singular Value Decomposition
- QR-Decomposition

[6] [Himpe (In Preparation)]

Implementation (`optmor`)

optmor - Optimization-Based Model Order Reduction

Attributes:

- Optional Data-Driven Reduction
- Optional Trust-Region Reduction
- Configurable State Direction
- Configurable Objective Function
- Configurable Orthogonalization
- Arbitrary Parametrization

Features:

- Lazy Arguments
- Compatible with MATLAB & OCTAVE
- Implicit Parallelization
- Open-Source licensed

Prototype!

Gramian-Based Combined Reduction[8]

Split Input-To-Output map into:

1. Input-To-State map $\sim$ Controllability,
2. State-To-Output map $\sim$ Observability.

System Gramians:

- Controllability gramian: W_C
- Observability gramian: W_O
- Cross gramian[7]: W_X

Balanced Truncation:

$$\sigma_i = \sqrt{\lambda(W_C W_O)}$$

$$\exists U, V : V W_C V^* = U^* W_O U = \text{diag}(\sigma_i)$$

Direct Truncation:
(Approximate Balancing)

$$W_X = UDV \Rightarrow D \approx \text{diag}(\sigma_i)$$

[7] Symmetric systems only!

[8] Review in [Antoulas'05]

Empirical Gramians

- $W_C = \underset{u \in U}{mean}(\int_0^\infty x(t)x^*(t)dt)$
- $W_O = \underset{x_0 \in X}{mean}(\int_0^\infty \rho(y^*(t)y(t))dt)$
- $W_X = \underset{u \in U, x_0 \in X}{mean}(\int_0^\infty \varphi(x(t), y(t))dt)$

with U some perturbation to the input u: $U = E_u \times R_u \times Q_u$

- $E_u = \{e_i \in \mathbb{R}^j; \|e_i\| = 1; e_i e_{j \neq i} = 0; i = 1, \dots, m\}$
- $R_u = \{S_i \in \mathbb{R}^{j \times j}; S_i^* S_i = \mathbb{1}; i = 1, \dots, s\}$
- $Q_u = \{c_i \in \mathbb{R}; c_i > 0; i = 1, \dots, q\}$

and X some perturbation to the initial state x_0: $X = E_x \times R_x \times Q_x$

- $E_x = \{f_i \in \mathbb{R}^n; \|f_i\| = 1; f_i f_{j \neq i} = 0; i = 1, \dots, n\}$
- $R_x = \{T_i \in \mathbb{R}^{n \times n}; T_i^* T_i = \mathbb{1}; i = 1, \dots, t\}$
- $Q_x = \{d_i \in \mathbb{R}; d_i > 0; i = 1, \dots, r\}$

Empirical Controllability Gramian[9]

For sets E_u, R_u, Q_u, input $u(t)$ and input during the steady state $\bar{x}$, $\bar{u}$, the **empirical controllability gramian** is given by:

$$W_C = \frac{1}{|Q_u||R_u|} \sum_{h=1}^{|Q_u|} \sum_{i=1}^{|R_u|} \sum_{j=1}^{m} \frac{1}{c_h^2} \int_0^\infty \Psi^{hij}(t) dt,$$
$$\Psi^{hij}(t) = (x^{hij}(t) - \bar{x})(x^{hij}(t) - \bar{x})^* \in \mathbb{R}^{n \times n}.$$

With x^{hij} being the states for the input configuration $u^{hij}(t) = c_h S_i e_j u(t) + \bar{u}$.

For linear systems the empirical controllability gramian equals the analytic controllability gramian [Lall'99].

9 Introduced by [Lall'99]

Empirical Observability Gramian[10]

For sets E_x, R_x, Q_x and output y during the steady state $\bar{x}$, $\bar{y}$, the **empirical observability gramian** is given by:

$$W_O = \frac{1}{|Q_x||R_x|} \sum_{k=1}^{|Q_x|} \sum_{l=1}^{|R_x|} \frac{1}{d_k^2} T_l \int_0^\infty \Psi^{kl}(t) dt \ T_l^*,$$

$$\Psi_{ab}^{kl} = (y^{kla}(t) - \bar{y})^* (y^{klb}(t) - \bar{y}) \in \mathbb{R}.$$

With y^{kla} being the systems output for the initial state configuration $x_0^{kla} = d_k T_l f_a + \bar{x}$.

For linear systems the empirical observability gramian equals the analytic observability gramian [Lall'99].

[10] Introduced by [Lall'99]

Empirical Cross Gramian[11]

For sets E_u, E_x, R_u, R_x, Q_u, Q_x, input $\bar{u}$ during steady state $\bar{x}$ with output $\bar{y}$, the **empirical cross gramian** is given by:

$$W_X = \frac{1}{|Q_u||R_u|m|Q_x||R_x|} \sum_{h=1}^{|Q_u|} \sum_{i=1}^{|R_u|} \sum_{j=1}^{m} \sum_{k=1}^{|Q_x|} \sum_{l=1}^{|R_x|} \frac{1}{c_h d_k} \int_0^\infty T_l \Psi^{hijkl}(t) T_l^* dt,$$

$$\Psi_{ab}^{hijkl}(t) = f_b^* T_k^* \Delta x^{hij}(t) e_i^* S_h^* \Delta y^{kla}(t),$$

$$\Delta x^{hij}(t) = (x^{hij}(t) - \bar{x}),$$

$$\Delta y^{kla}(t) = (y^{kla}(t) - \bar{y}).$$

Where x^{hij} and y^{kla} being states and output for the input $u^{hij}(t) = c_h S_i e_j u(t) + \bar{u}$ and initial state $x_0^{kla} = d_k T_l f_a + \bar{x}$ respectively.

For linear systems the empirical cross gramian equals the analytic cross gramian [Himpe'13].

[11] Introduced for SISO by [Streif'09], for MIMO by [Himpe'13]

Empirical Sensitivity Gramian[12]

Treating the parameters as additional inputs of $dim(\theta)$ with steady input θ (if possible) gives:

$$\dot{x} = f(x,u) + \sum_{k=1}^{P} f(x,\theta_k) \Rightarrow W_C = W_{C,0} + \sum_{k=1}^{P} W_{C,k}$$

Sensitivity Gramian W_S:

$$W_{S,ii} = \text{tr}(W_{C,i}).$$

Combined Reduction:

- controllability gramian is a byproduct
- requires additional observability gramian

[12] Based on [Sun'06], introduced by [Himpe'13a]

Empirical Identifiability Gramian[13]

Augmenting the system by $\dim(\theta)$ constant states with initial value θ yields:

$$\dot{x}_a = \begin{pmatrix} \dot{x} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} f(x,u,\theta) \\ 0 \end{pmatrix}, x_a(0) = \begin{pmatrix} x_0 \\ \theta \end{pmatrix} \Rightarrow W_{O,a} = \left(\begin{array}{c|c} W_O & W_M \\ \hline W_M^* & W_P \end{array}\right).$$

Identifiability Gramian W_I:

$$W_I = W_P - W_M^* {W_O}^{-1} W_M \approx W_P.$$

Combined Reduction:

- observability gramian is a byproduct
- requires additional controllability gramian

[13] Introduced by [Geffen'08]

Empirical Joint Gramian[14]

Augmenting the system by $\dim(\theta)$ constant states with initial value θ yields:

$$\dot{x}_a = \begin{pmatrix} \dot{x} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} f(x,u,\theta) \\ 0 \end{pmatrix}, x_a(0) = \begin{pmatrix} x_0 \\ \theta \end{pmatrix} \Rightarrow W_{J,a} = \left(\begin{array}{c|c} W_X & W_M \\ \hline 0 & 0 \end{array}\right).$$

Cross-Identifiability Gramian W_{II}:

$$W_{II} = -W_M^* (W_X + W_X^T)^{-1} W_M \approx -W_M^* \operatorname{diag}(W_X + W_X^T)^{-1} W_M.$$

Combined Reduction:

- cross gramian is a byproduct
- requires NO additional gramian

[14] Introduced by [Himpe'13]

Implementation (`emgr`)

emgr - Empirical Gramian Framework

Gramians:

- Empirical Controllability Gramian
- Empirical Observability Gramian
- Empirical Cross Gramian
- Empirical Sensitivity Gramian
- Empirical Identifiability Gramian
- Empirical Joint Gramian

Features:

- Uniform Interface
- Compatible with MATLAB & OCTAVE
- Vectorized & Parallelizable
- Open-Source licensed

More info at: `http://gramian.de`

Numerical Results (Parametrized Linear System)

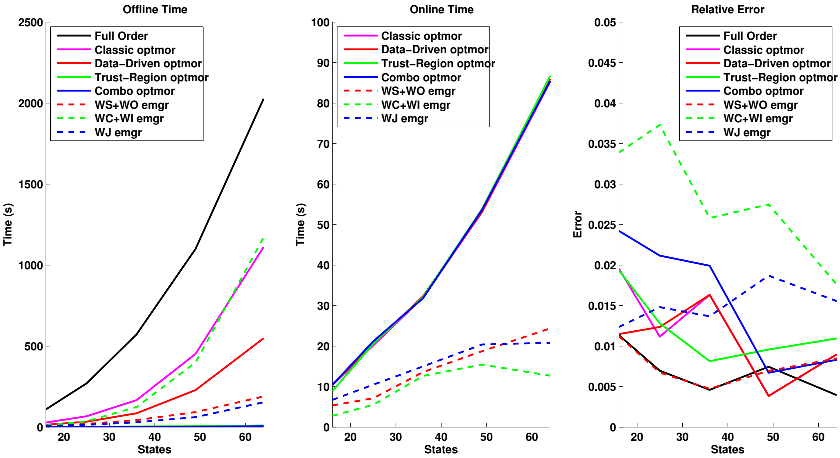
$$\dot{x} = A(\theta)x + Bu$$
$$y = Cx$$

- $x \in \mathbb{R}^{m^2}$
- $u \in \mathbb{R}^{m}$
- $y \in \mathbb{R}^{m}$
- $\theta \in \mathbb{R}^{m^4}$
- $A \in \mathbb{R}^{m^2 \times m^2}$
- $B \in \mathbb{R}^{m^2 \times m}$
- $C \in \mathbb{R}^{m \times m^2}$
- $A(\theta) : \mathbb{R}^{m^4} \to \mathbb{R}^{m^2} \bigotimes \mathbb{R}^{m^2} = \mathbb{R}^{m^2 \times m^2}, \theta \mapsto A$

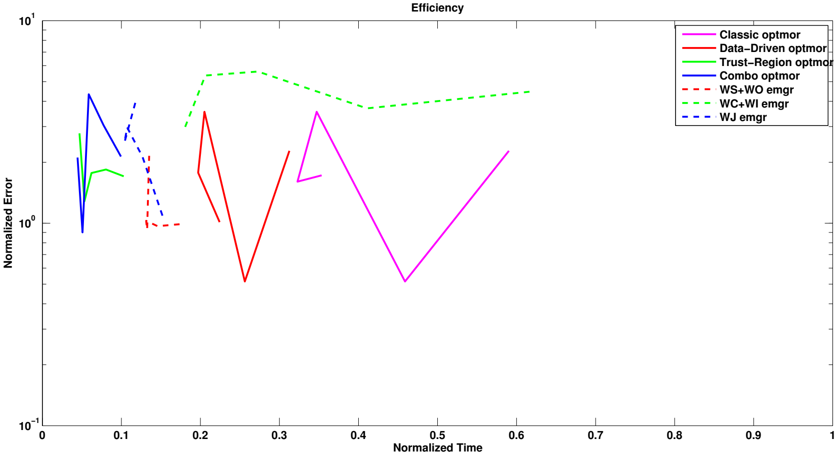
Notes:

$$\theta \in U(0,1), \quad \lambda(A(\theta)) < 0$$
$$A = A^T \wedge C = B^T \Rightarrow CA^{-1}B = B^T A^{-T} C^T$$
$$\theta_{\text{prior}} = \text{vec}(-\mathbb{1}_{m^2}), \quad S_{\text{prior}} = \mathbb{1}_{m^4}$$

Numerical Results (Offline, Online, Error)



Numerical Results (Effectivity)



Comparison

	Optimization-Based	Gramian-Based
Problems	Linear	Nonlinear
Sparsity	Explicit	Implicit
Parallelization	Implicit	Explicit
Offline Time	Faster	Fast
Online Time	Good	Better
Relative Error	Acceptable	Acceptable
Scale	Extreme	Large
Issues	Nonlinear	Nonsymmetric

tl;dl

- Combined Reduction: Reduction of States and Parameters.
- Optimization-Based: Extreme-Scale Linear Models.
- Empirical Gramian-Based: Large-Scale Nonlinear Models.
- Get the Source Code: `http://j.mp/comred13` .

Thanks!