Geometric adaptation and parameterization of subspace-based reduced models

Ralf Zimmermann, AG Numerik, ICM, TU Braunschweig
Workshop: MOR 4 MEMS, KIT, Nov. 17/18, 2015
Geometric adaptation and parameterization of ROMs

- Motivation: projection-based nonlinear model reduction
- Geometric subspace adaptation
- Parametric model reduction via exploiting derivatives
- Outlook
Motivation: projection-based nonlinear model reduction

Spatio-temporal dynamical system, semi-discrete

- \( \frac{d}{dt} y(t) = F(y(t)) \in \mathbb{R}^N \), state \( y(t) \in \mathbb{R}^N \), \( F: \mathbb{R}^N \rightarrow \mathbb{R}^N \) nonlinear

- Subspace of solution candidates \( \mathcal{V} = \text{colspan}(V), V \in \mathbb{R}^{N \times n} \)

- Projection: \( y(t) \approx Vy_r(t), \ V \in \mathbb{R}^{N \times n} \)
  \[
  \frac{d}{dt} y_r(t) = V^T F(Vy_r(t)) \in \mathbb{R}^n
  \]

- Observation: evaluation of nonlinear term is as costly as before!
Motivation: projection-based nonlinear model reduction

Discrete Empirical Interpolation Method [1]

- DEIM ansatz: represent candidates for nonlinear part by another subspace

\[ F(Vy_r(t)) \approx U\nu, \quad U \in \mathbb{R}^{N \times p} \]

- Try and fit \( F(Vy_r(t)) \) as good as possible using \( U \in \mathbb{R}^{N \times p} \)

\[ \min_{\nu} \|P^T U\nu - P^T F(Vy_r(t))\| \]

Motivation: projection-based nonlinear model reduction

Discrete Empirical Interpolation Method \cite{1}

- DEIM ansatz: represent candidates for nonlinear part by another subspace

\[ F(Vy_r(t)) \approx Uv, \quad U \in \mathbb{R}^{N \times p} \]

- Try and fit \( F(Vy_r(t)) \) as good as possible using \( U \in \mathbb{R}^{N \times p} \)

\[
\min_v \| P^T U v - P^T F(Vy_r(t)) \|.
\]

\[
(P = (e_{i,1}, \ldots, e_{i,r}) \in \mathbb{R}^{N \times r}: \text{point selection matrix / mask matrix})
\]
Motivation: projection-based nonlinear model reduction

Similar ideas appear in

- Missing Point Estimation (MPE), [Astrid et al., 2008]
- Masked Projection, [Galbally et al., 2010]
- Gappy POD, [Everson and Sirovich, 1995, Bui-Thanh et al., 2004]
Geometric adaptation and parameterization of ROMs

- Motivation: projection-based nonlinear model reduction

- **Geometric subspace adaptation**

- Parametric model reduction via exploiting derivatives

- Outlook
Geometric adaptation and parameterization of ROMs

- Motivation: projection-based nonlinear model reduction

- Geometric subspace adaptation\(^2\)

- Parametric model reduction via exploiting derivatives

- Outlook

\(^2\) RZ, Peherstorfer, Willcox, Geometric subspace optimization with applications to online adaptive nonlinear model reduction, preprint 2015
A key idea of model reduction

Summary

- Subspace construction, \( \text{dim } p \ll n \)

\[
W(q) \approx U \nu(q), \quad U \in \mathbb{R}^{n \times p}
\]

Idea: twist the subspace

- Determine coefficient vector \( \nu(q) \in \mathbb{R}^p \) via your method of choice

‘offline’, very expensive

‘online’, very cheap

What if you need to adapt your ROM in the online stage?
Least-squares problem restricted to subspace

- \( \text{argmin}_{\nu \in \mathbb{R}^p} \| b - A U \nu \| \iff (U^T A^T A U) \nu = U^T A^T b \)

- Given \( \mathcal{U} = \text{colspan}(U) \) the minimum residual w.r.t. \( \mathcal{U} \) is

\[
F(\mathcal{U}) = b^T b - b^T A U (U^T A^T A U)^{-1} U^T A^T b \in \mathbb{R}
\]

\( \Rightarrow \) Invariant under changes of basis

\( \Rightarrow \) Well-defined differentiable objective function on Grassmann manifold

- \( \text{Gr}(n, p) = \{ [U] \subset \mathbb{R}^n | \dim([U]) = p \} \)

- \( \mathcal{U} = [U] = \{ U \in \mathbb{R}^{n \times p} | \text{colspan}(U) = \mathcal{U} \} \) = **equivalence class** of matrices spanning the same subspace
Least-squares problem restricted to subspace

- Double minimization objective

\[(*) \quad \min_{U \in Gr(n,p)} \min_{x \in U} \| b - Ax \| \iff \min_{U \in Gr(n,p)} F(U) \]

- Exact optimum not unique: any subspace that contains the unrestricted least-squares solution
- MOR setting: too expensive to compute directly

⇒ optimize (*) numerically
Conjugate Gradients for non-quadratic function

Given: $F : \mathbb{R}^d \rightarrow \mathbb{R}$ differentiable

- **Start:** $u_0 \in \mathbb{R}^d, g_0 = \nabla F(u_0), h_0 = -g_0$
- For $i = 0, 1, 2, \ldots$
  1. IF $\|g_0\| < \varepsilon$ : STOP
  2. ELSE (line search)
     
     $s_i = \arg\min_{s>0} F(u_i + sh_i)$
     
     $u_{i+1} = u_i + s_i h_i$
     
     $g_{i+1} = \nabla F(u_{i+1}), h_{i+1} = -g_{i+1} + \gamma_i h_i,$

where $\gamma_i = \frac{\langle g_{i+1} - g_i, g_{i+1} \rangle}{\langle g_i, g_i \rangle}$ (Polak-Ribière approx. of Hessian)
Conjugate Gradients for non-quadratic functions

Given: $F: \mathbb{R}^d \rightarrow \mathbb{R}$ differentiable

- **Start**: $u_0 \in \mathbb{R}^d$, $g_0 = \nabla F(u_0)$, $h_0 = -g_0$
- For $i=0,1,2,…$
  1. IF $\|g_0\| < \varepsilon$ : STOP
  2. ELSE (line search)
     \[ s_i = \text{argmin}_{s \geq 0} E(u_i + sh_i) \]
     \[ u_{i+1} = u_i + s_i h_i \]
  3. \[ g_{i+1} = \nabla F(u_{i+1}), h_{i+1} = -g_{i+1} + \gamma_i h_i, \]

where \[ \gamma_i = \frac{\langle g_{i+1} - g_i, g_{i+1} \rangle}{\langle g_i, g_i \rangle} \] (Polak-Ribière approx. of Hessian)

On curved manifolds:
- No vector space structure
- Tangent vectors and points not compatible
- No “+” operator for vectors in different tangent spaces
Manifold analogues of required operations

- Replace directional line search along a geodesic line.
Manifold analogues of required operations

Use concept of **parallel transport** to update search direction

- Move $H_0 \in T_{U_0} Gr(n, p)$ to $\tau(H_0)(t) \in T_{U_t} Gr(n, p)$
  along **geodesic** starting in $U_0$ with velocity $\Delta_0$
Given: $F: \text{Gr}(n, p) \rightarrow \mathbb{R}$ differentiable

- **Start:** $[U_0] \in \text{Gr}(n, p), G_0 = \nabla F([U_0]), H_0 = -G_0$
- **For** $i=0,1,2,...$
  1. **IF** $\|G_0\| < \varepsilon$ : **STOP**
  2. **ELSE** (geodesic line search)
     
     $$s_i = \arg\min_{s \geq 0} F(\text{Exp}_{[U_i]}(sH_i))$$
     $$[U_{i+1}] = \text{Exp}_{[U_i]}(s_iH_i)$$
     
     3. $G_{i+1} = \nabla F([U_{i+1}]), H_{i+1} = -G_{i+1} + \gamma_i \tau(H_i)$,

where $\gamma_i = \frac{\langle G_{i+1} - \tau(G_i), G_{i+1}\rangle}{\langle G_i, G_{i+1}\rangle}$ (Polak-Ribière approx. of Hessian)

**Move along geodesic = evaluate Exp mapping**
**Conjugate Gradients on the Grassmannian**

Given: \( F : Gr(n, p) \rightarrow \mathbb{R} \) differentiable

- **Start:** \([U_0] \in Gr(n, p), G_0 = \nabla F([U_0]), H_0 = -G_0\)
- **For** \(i = 0, 1, 2, \ldots\)
  1. IF \( \|G_0\| < \varepsilon \) : STOP
  2. ELSE (geodesic line search)

\[
s_i = \text{argmin}_{s > 0} F(\text{Exp}_{[U_i]}(sH_i))
\]

\[
[U_{i+1}] = \text{Exp}_{[U_i]}(s_iH_i)
\]

3. \( G_{i+1} = \nabla F([U_{i+1}]), H_{i+1} = -G_{i+1} + \gamma_i \tau(H_i), \)

where \( \gamma_i = \frac{\langle G_{i+1} - \tau(G_i), G_{i+1} \rangle}{\langle G_i, G_i \rangle} \) (Polak-Ribière approx. of Hessian)

“adding” tangent vectors by parallel transport along geodesics

---

[16] Absil, Mahony, Sepulchre, 2008

---

Edelman, Arias, Smith, 1998
Absil, Mahony, Sepulchre, 2008
Example A
Subspace adaptation in the DEIM ROM context

Discrete Empirical Interpolation Method

- Try and fit $F(Vy_r(t))$ as good as possible using $U \in \mathbb{R}^{N \times p}$

$$\min_v \|P^T U v - \left[ P^T F(Vy_r(t)) \right]_b \|$$

($P \in \mathbb{R}^{p \times r}$ point selection matrix, DEIM oversampling required)

- Fits to our setting: optimum is

$$(U^T P P^T U) \quad v = U^T P b$$

- Residual depending on subspace

$$\rho(U) = b^T b - b^T P^T U (U^T P P^T U)^{-1} U^T P b \in \mathbb{R}$$
Example A
Subspace adaptation in the DEIM ROM context

Application to FitzHugh-Nagumo system

- Model of electrical activity in a neuron
- Time-dependent nonlinear PDE system
- Benchmark problem in original DEIM paper
- Spatial dimensions: 2,048 DoFs
- Time discretization: $K = 10^6$ steps

- Subspace optimization every 50, 100, 200 time steps using 200, 400, 600 filter points
- Subspace optimization is very fast, fits the selected DEIM equations in 1 CG iteration!
**Example A**  
**Subspace adaptation in the DEIM ROM context**

**Application to FitzHugh-Nagumo System**

![Graphs showing the averaged relative $L_2$ error of state vs DEIM dimension](chart.png)

- **Left graph:** (adapt every 50 steps)
  - red: static
  - blue: adapt, samples 200
  - green: adapt, samples 400
  - purple: adapt, samples 600

- **Right graph:** (adapt with 200 samples)
  - red: static
  - blue: adapt, every 200
  - green: adapt, every 100
  - purple: adapt, every 50
Example B
Subspace adaptation in the Gappy POD context

Gappy POD in a nutshell\cite{3,4}:

Given

- set of snapshots \( y^1, \ldots, y^m \in \mathbb{R}^n \)
- (reduced-order) subspace \( U \approx \text{span}(y^1, \ldots, y^m) \)
- vector of gappy data \( y^g \in \mathbb{R}^n \) (defective entries set to zero)

- POD ansatz:
  \[
  y \approx Uv, \quad U \in \mathbb{R}^{n \times p}, \ v \in \mathbb{R}^p
  \]

- Gappy POD least-squares problem

\[
\min_v \| P^T U v - P^T y^g \|
\]

\cite{3} Everson, Sirovich, J. Opt. Soc. Am., 1995,
\cite{4} Bui-Thanh, Damadoran, Willcox, AIAA Journal, 2004
Example B
Subspace adaptation in the Gappy POD context

Training the subspace to produce an unsampled feature
Snapshot ensemble: ten (64x64)-pictures represented by vectors of 4096 brightness values (Yale Faces data base)

Use Grassmann-optimization in order to “learn” the “glasses-on” feature.
Example B
Subspace adaptation in the Gappy POD context

Subspace training

Reference

Gappy POD reconstruction

Training set

Reference projection onto Init. subspace
Example B
Subspace adaptation in the Gappy POD context

Subspace training:
Example B
Subspace adaptation in the Gappy POD context

Initial data set projected onto adapted subspaces

Initial data set projected onto optimized subspace

Initial data set projected onto subspace with only the last column optimized

Initial data set
Geometric adaptation and parameterization of ROMs

- Motivation: projection-based nonlinear model reduction
- Geometric subspace adaptation
- Parametric model reduction via exploiting derivatives
- Outlook
Parametric model reduction via exploiting derivatives

- **Standard method for local approximations to non-linear functions:**
  - Linearisation via 1st order Taylor series

- **Parameter-dependent ROMs**
  - Given a (projection-based) ROM at operating point $q_0$
  - Reduced subspace spanned by left-hand singular vectors of solution snapshots
    
    $Y(q_0) = U(q_0)\Sigma(q_0)V^T(q_0)$
    
    $\rightarrow$ parameter-dependent (orth.) basis $U(q_0)$, $U(q_0)^TU(q_0) = I$
    
    $\rightarrow$ parameter-dependent subspace $\mathcal{U} = \mathcal{U}(q_0) = \text{colspan}(U(q_0))$
  - **Objective:**
    
    compute $\{U(q_0 + \Delta q)\}$ without computing additional snapshots
In Euclidean spaces: Starting velocity is first-order derivative

\[ Y(q + \Delta q) = Y(q) + \dot{Y}\Delta q + O(\Delta q^2) \]

Transfer to manifold setting
Parametric model reduction via exploiting derivatives

\[ \exp_{U(q_0)}(\Delta q \dot{U}(q_0)) \]

\[ U(q_0) \]

\[ \Delta q \dot{U}(q_0) \]
 Lemma (folklore): Derivative of SVD problem
Let $Y: q \mapsto Y(q) = U(q)\Sigma(q)V(q)^T \in \mathbb{R}^{n \times p}$ be differentiable and let $\mathcal{C} = Y^T Y$. For simplicity, assume mutually distinct non-zero singular values:

- $\dot{\sigma}_j = (u_j^T \dot{Y} v_j$)
- $\dot{V} = V \Gamma$, where $\Gamma_{ij} = \begin{cases} (v_i^T \dot{\mathcal{C}} v_j) / (\sigma_j + \sigma_i)(\sigma_j - \sigma_i), & i \neq j \\ 0, & i = j \end{cases}$ (skew-symmetric)
- $\dot{U} = (\dot{Y} V + U(\Sigma \dot{\Gamma} - \dot{\Sigma}))\Sigma^{-1}$

Similar for truncated SVD
Remark: Consider \( U(q) \) as a curve in the **Stiefel manifold**

\[
U: [q - \Delta q, q + \Delta q] \to St(n, p)
\]

Differential:

\[
d_qU: T_q \mathbb{R} \cong \mathbb{R} \to T_{U(q)}St(n, p) \cong \{\Delta \in \mathbb{R}^{n \times p} | \Delta^T U(q) = -U(q)^T \Delta\}
\]

Observation:

• Lemma gives derivative in accordance with differential geometric **Stiefel derivative**

• **Grassmann derivative** is orth. component \( \dot{U}^\perp := (I - UU^T)\dot{U} \)
Parametric model reduction via exploiting derivatives

- Constructing ROM subspaces for small parameter changes

Given:
- $U(q_0) \in St(n, p) = \{ U \in R^{n \times p}, U^T U = I_p \}$
- $\dot{U}(q_0) \in T_{U(q_0)}St$

Valid approximation: (for Grassmann: analogous)

$$U(q_0 + \Delta q) \approx \exp_{U(q_0)}(\Delta q \dot{U}(q_0)) \quad (\in St(n, p))$$

- Observation:
  - $\exp_{St}^{St}(\Delta q \dot{U}(q_0)) = U(q_0) + \Delta q \dot{U}(q_0) + \mathcal{O}(\Delta q^2)$
  - $\exp_{Gr}^{Gr}(\Delta q \dot{U}^\perp(q_0)) = U(q_0) + \Delta q \dot{U}^\perp(q_0) + \mathcal{O}(\Delta q^2)$
### Taylor-like approaches to local pMOR

1. \( U(q + \Delta q) \approx \text{svd}(Y(q) + \Delta q \hat{Y}(q)) \)  
   \( \text{(Taylor, brute force)} \)  
2. \( U(q + \Delta q) \approx U(q) + \Delta q \hat{U}(q) \)  
   \( \text{([Hay et al. 2009])} \)  
3. \( U(q + \Delta q) \approx qr \left( U(q) + \Delta q \hat{U}(q) \right) \)  
   \( \text{(reorth of [Hay et al. 2009])} \)  
4. \( U(q + \Delta q) \approx \exp^{st} \left( \Delta q \hat{U}(q) \right) \)  
   \( \text{(Stiefel geodesic)} \)  
5. \( U(q + \Delta q) \approx \exp^{Gr} \left( \Delta q \hat{U}(q)^\perp \right) \)  
   \( \text{(Grassmann geodesic)} \)

#### Associated FLOP count for each \( \Delta q \)

1. \( O(nm^2) \)  
2. \( O(np) \)  
3. \( O(np^2) \)  
4. \( O(np^2) \)  
5. \( O(np) \)
Example: Linear parameterized convection-diffusion problem

- Parametric PDE with known closed form solution
- Explicit expression for snapshot derivatives
Example: Linear parameterized convection-diffusion problem

- Parametric PDE with known closed form solution
- Explicit expression for snapshot derivatives
Example: Linear parameterized convection-diffusion problem

- Parametric PDE with known closed form solution
- Explicit expression for snapshot derivatives
Geometric adaptation and parameterization of ROMs

- Motivation: projection-based nonlinear model reduction
- Geometric subspace adaptation
- Parametric model reduction via exploiting derivatives
- Outlook
Conclusion & Outlook

Towards a nonlinear parametric model reduction scheme...

- Stiefel geodesic: feasible subspace representative that matches Taylor up to terms of second order
- FLOP count: $O(np^2)$, n=full dim., p = reduced dim.
- Grassmann geodesic: feasible subspace representative that matches essential part of Taylor up to terms of second order
- FLOP count: $O(np)$, n=full dim., p = reduced dim.
- Great savings expected for highly compressed models

Next steps:

- Apply to real-life problem
- Error estimation: second-order term in Taylor series vs. second-order term in geodesic
- How to handle multiple parameters?
Thank you for your attention! Questions?