Adaptive Stochastic Collocation for PDE-Constrained Optimization under Uncertainty using Sparse Grids and Model Reduction

Matthew J. Zahr

Advisor: Charbel Farhat
Computational and Mathematical Engineering
Stanford University

Joint work with: Kevin Carlberg (Sandia CA), Drew Kouri (Sandia NM)

SIAM Conference on Uncertainty Quantification
MS104: Reduced-Order Modeling in Uncertainty Quantification
Lausanne, Switzerland
April 7, 2016

PDE-Constrained Optimization under Uncertainty

Goal: Efficiently solve stochastic PDE-constrained optimization problems

$$\begin{aligned} \underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathbb{E}[\mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu},\,\cdot\,),\,\boldsymbol{\mu},\,\cdot\,)] \\ \text{subject to} \quad & \boldsymbol{r}(\boldsymbol{u}(\boldsymbol{\mu},\,\boldsymbol{\xi}),\,\boldsymbol{\mu},\,\boldsymbol{\xi}) = 0 \quad \forall \boldsymbol{\xi} \in \boldsymbol{\Xi} \end{aligned}$$

- $\boldsymbol{r} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \rightarrow \mathbb{R}^{n_{\boldsymbol{u}}}$ is the discretized stochastic PDE
- $\mathcal{J} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \rightarrow \mathbb{R}$ is a quantity of interest
- $\boldsymbol{u} \in \mathbb{R}^{n_{\boldsymbol{u}}}$ is the PDE state vector
- $\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}$ is the vector of (deterministic) optimization parameters
- $\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ is the vector of stochastic parameters

Literature Review: Stochastic Optimal Control

- **Stochastic collocation**
  - Dimension-adaptive sparse grids – globalized with trust-region method [Kouri et al., 2013, Kouri et al., 2014]
  - Generalized polynomial chaos – sequential quadratic programming [Tiesler et al., 2012]
  - \+ *Orders of magnitude improvement over isotropic sparse grids*
  - \- *Still requires many PDE solves for even moderate dimensional problems*
- **Model order reduction**
  - Goal-oriented, dimension-adaptive, weighted greedy algorithm for training stochastic reduced-order model [Chen and Quarteroni, 2015]
  - Extension to optimal control [Chen and Quarteroni, 2014]
  - \+ *Reduction in number of PDE solves at cost of large number of ROM solves*
  - \- *Restriction to offline-online framework may lead to unnecessay PDE solves and large reduced bases*

Proposed Approach

Introduce two levels of inexactness to obtain an inexpensive, approximate version of the stochastic optimization problem; manage inexactness with trust-region-like method

- **Anisotropic sparse grids** used for *inexact integration* of risk measures
- **Reduced-order models** used for *inexact evaluations* at collocation nodes
- **Error indicators** introduced to account for *both* sources of inexactness
- **Refinement** of integral approximation and reduced-order model via *dimension-adaptive* sparse grids and a *greedy method* over collocation nodes
- Embedded in globally convergent **trust-region-like** algorithm with a strong connection to error indicators and refinement mechanism

Anisotropic Sparse Grids [Gerstner and Griebel, 2003]

1D Quadrature Rules: Define the difference operator

$$\Delta_k^j \equiv \mathbb{E}_k^j - \mathbb{E}_k^{j-1}$$

where $\mathbb{E}_k^0 \equiv 0$ and $\mathbb{E}_k^j$ as the level-j 1d quadrature rule for dimension k

Anisotropic Sparse Grid:

$$\mathbb{E}_{\mathcal{I}} \equiv \sum_{\mathbf{i} \in \mathcal{I}} \Delta_1^{i_1} \otimes \cdots \otimes \Delta_{n_{\boldsymbol{\xi}}}^{i_{n_{\boldsymbol{\xi}}}}$$

Forward Neighbors:

$$\mathcal{N}(\mathcal{I}) = \{\mathbf{k} + \mathbf{e}_j \mid \mathbf{k} \in \mathcal{I}\} \setminus \mathcal{I}$$

Truncation Error: [Gerstner and Griebel, 2003, Kouri et al., 2013]

$$\mathbb{E} - \mathbb{E}_{\mathcal{I}} = \sum_{\mathbf{i} \notin \mathcal{I}} \Delta_1^{i_1} \otimes \cdots \otimes \Delta_{n_{\boldsymbol{\xi}}}^{i_{n_{\boldsymbol{\xi}}}} \approx \sum_{\mathbf{i} \in \mathcal{N}(\mathcal{I})} \Delta_1^{i_1} \otimes \cdots \otimes \Delta_{n_{\boldsymbol{\xi}}}^{i_{n_{\boldsymbol{\xi}}}} = \mathbb{E}_{\mathcal{N}(\mathcal{I})}$$

Stochastic Collocation via Anisotropic Sparse Grids

Stochastic collocation using anisotropic sparse grid nodes used to approximate integral with summation

$$\begin{aligned} \underset{\boldsymbol{u}\in\mathbb{R}^{n_{\boldsymbol{u}}},\ \boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathbb{E}[\mathcal{J}(\boldsymbol{u},\, \boldsymbol{\mu},\, \cdot\,)] \\ \text{subject to} \quad & \boldsymbol{r}(\boldsymbol{u},\, \boldsymbol{\mu},\, \boldsymbol{\xi}) = 0 \quad \forall \boldsymbol{\xi} \in \boldsymbol{\Xi} \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} \underset{\boldsymbol{u}\in\mathbb{R}^{n_{\boldsymbol{u}}},\ \boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathbb{E}_{\mathcal{I}}[\mathcal{J}(\boldsymbol{u},\, \boldsymbol{\mu},\, \cdot\,)] \\ \text{subject to} \quad & \boldsymbol{r}(\boldsymbol{u},\, \boldsymbol{\mu},\, \boldsymbol{\xi}) = 0 \quad \forall \boldsymbol{\xi} \in \boldsymbol{\Xi}_{\mathcal{I}} \end{aligned}$$

Projection-Based Model Reduction to Reduce PDE Size

- Model Order Reduction (MOR) assumption: *state vector lies in low-dimensional subspace*

$$\boldsymbol{u} \approx \boldsymbol{\Phi} \boldsymbol{y} \qquad \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\mu}} \approx \boldsymbol{\Phi} \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{\mu}}$$

  where
  - $\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_{\boldsymbol{u}}^1 & \cdots & \boldsymbol{\phi}_{\boldsymbol{u}}^{k_{\boldsymbol{u}}} \end{bmatrix} \in \mathbb{R}^{n_{\boldsymbol{u}} \times k_{\boldsymbol{u}}}$ is the reduced basis
  - $\boldsymbol{y} \in \mathbb{R}^{k_{\boldsymbol{u}}}$ are the reduced coordinates of $\boldsymbol{u}$
  - $n_{\boldsymbol{u}} \gg k_{\boldsymbol{u}}$
- Substitute assumption into High-Dimensional Model (HDM), $\boldsymbol{r}(\boldsymbol{u}, \boldsymbol{\mu}, \boldsymbol{\xi}) = 0$, and use a Galerkin projection to obtain the square system

$$\boldsymbol{\Phi}^T \boldsymbol{r}(\boldsymbol{\Phi} \boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\xi}) = 0$$

Definition of $\boldsymbol{\Phi}$: Data-Driven Reduction

State-Sensitivity Proper Orthogonal Decomposition (SSPOD)

- Collect state and sensitivity snapshots by sampling HDM

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{u}(\boldsymbol{\mu}_1, \boldsymbol{\xi}_1) & \boldsymbol{u}(\boldsymbol{\mu}_2, \boldsymbol{\xi}_2) & \cdots & \boldsymbol{u}(\boldsymbol{\mu}_n, \boldsymbol{\xi}_n) \end{bmatrix}$$
$$\boldsymbol{Y} = \begin{bmatrix} \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_1, \boldsymbol{\xi}_1) & \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_2, \boldsymbol{\xi}_2) & \cdots & \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_n, \boldsymbol{\xi}_n) \end{bmatrix}$$

- Use Proper Orthogonal Decomposition to generate reduced basis for each individually

$$\boldsymbol{\Phi}_{\boldsymbol{X}} = \mathrm{POD}(\boldsymbol{X})$$
$$\boldsymbol{\Phi}_{\boldsymbol{Y}} = \mathrm{POD}(\boldsymbol{Y})$$

- Concatenate and orthogonalize to get reduced-order basis

$$\boldsymbol{\Phi} = \mathrm{QR}\left(\begin{bmatrix} \boldsymbol{\Phi}_{\boldsymbol{X}} & \boldsymbol{\Phi}_{\boldsymbol{Y}} \end{bmatrix}\right)$$

Reduced-Order Stochastic Collocation via Anisotropic Sparse Grids

Stochastic collocation of the reduced-order model over anisotropic sparse grid nodes used to approximate integral with cheap summation

$$\begin{aligned} \underset{\boldsymbol{u}\in\mathbb{R}^{n_{\boldsymbol{u}}},\ \boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathbb{E}[\mathcal{J}(\boldsymbol{u},\ \boldsymbol{\mu},\ \cdot\ )] \\ \text{subject to} \quad & \boldsymbol{r}(\boldsymbol{u},\ \boldsymbol{\mu},\ \boldsymbol{\xi}) = 0 \quad \forall \boldsymbol{\xi} \in \boldsymbol{\Xi} \end{aligned}$$

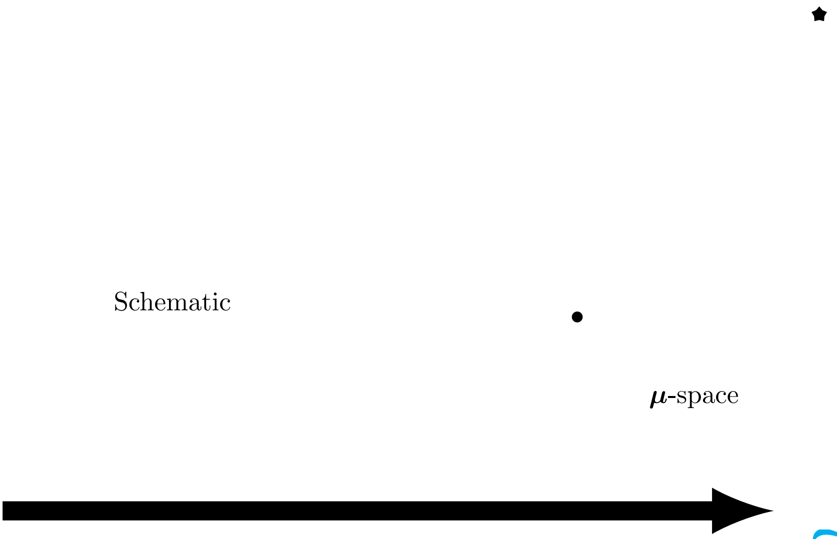
$$\Downarrow$$

$$\begin{aligned} \underset{\boldsymbol{u}\in\mathbb{R}^{n_{\boldsymbol{u}}},\ \boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathbb{E}_{\mathcal{I}}[\mathcal{J}(\boldsymbol{u},\ \boldsymbol{\mu},\ \cdot\ )] \\ \text{subject to} \quad & \boldsymbol{r}(\boldsymbol{u},\ \boldsymbol{\mu},\ \boldsymbol{\xi}) = 0 \quad \forall \boldsymbol{\xi} \in \boldsymbol{\Xi}_{\mathcal{I}} \end{aligned}$$

$$\Downarrow$$

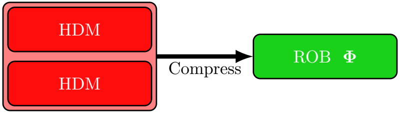
$$\begin{aligned} \underset{\boldsymbol{y}\in\mathbb{R}^{k_{\boldsymbol{u}}},\ \boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathbb{E}_{\mathcal{I}}[\mathcal{J}(\boldsymbol{\Phi}\boldsymbol{y},\ \boldsymbol{\mu},\ \cdot\ )] \\ \text{subject to} \quad & \boldsymbol{\Phi}^T\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{y},\ \boldsymbol{\mu},\ \boldsymbol{\xi}) = 0 \quad \forall \boldsymbol{\xi} \in \boldsymbol{\Xi}_{\mathcal{I}} \end{aligned}$$

ROM-Based Trust-Region Framework for Optimization

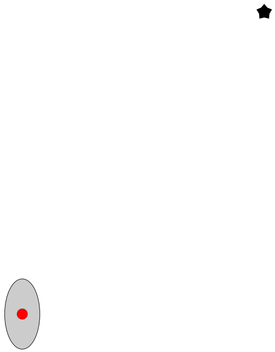


Breakdown of Computational Effort

ROM-Based Trust-Region Framework for Optimization



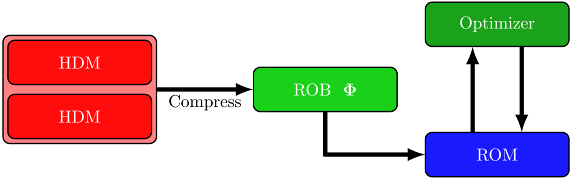
Schematic



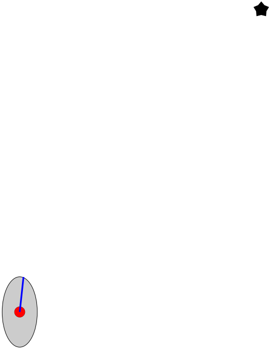
μ-space

Breakdown of Computational Effort

ROM-Based Trust-Region Framework for Optimization



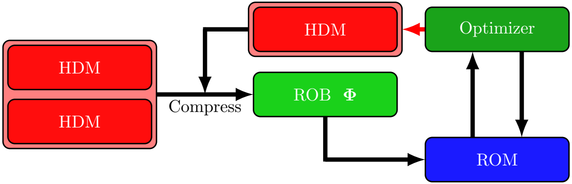
Schematic



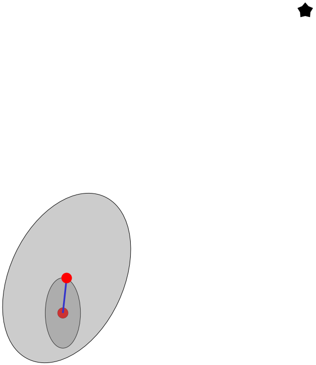
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

ROM-Based Trust-Region Framework for Optimization



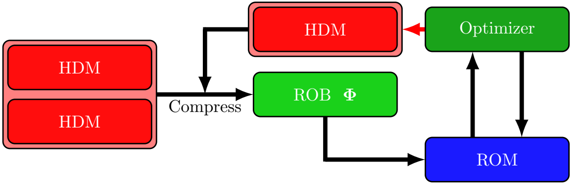
Schematic



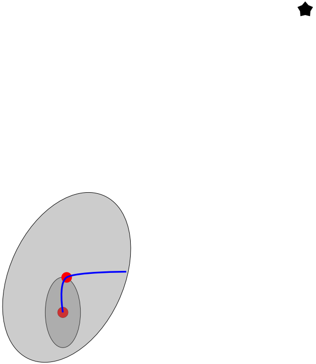
μ-space

Breakdown of Computational Effort

ROM-Based Trust-Region Framework for Optimization



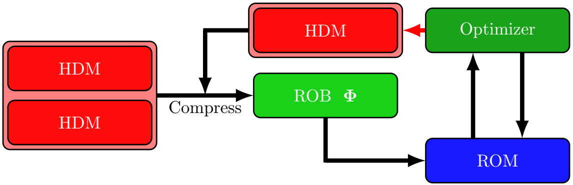
Schematic



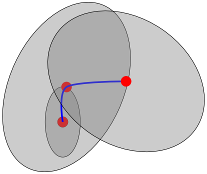
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

ROM-Based Trust-Region Framework for Optimization



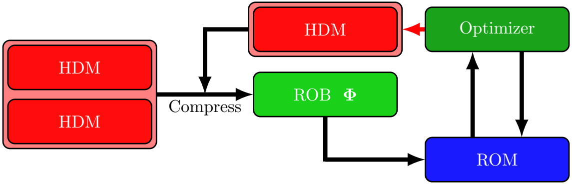
Schematic



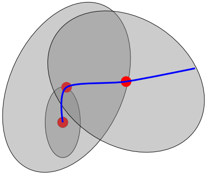
μ-space

Breakdown of Computational Effort

ROM-Based Trust-Region Framework for Optimization



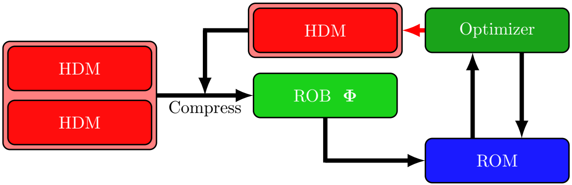
Schematic



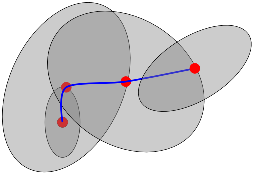
μ-space

Breakdown of Computational Effort

ROM-Based Trust-Region Framework for Optimization



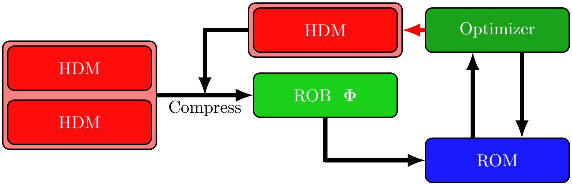
Schematic



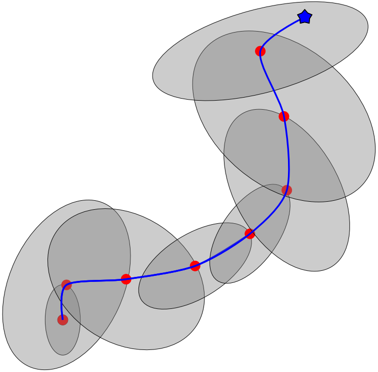
μ-space

Breakdown of Computational Effort

ROM-Based Trust-Region Framework for Optimization



Schematic



μ-space

Breakdown of Computational Effort

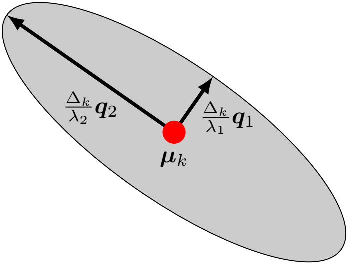
Trust-Regions Based on Error Indicators: Motivation

Let $\hat{\boldsymbol{\vartheta}}_k(\boldsymbol{\mu}) = \boldsymbol{\vartheta}_k(\boldsymbol{\Phi}\boldsymbol{y}(\boldsymbol{\mu}),\, \boldsymbol{\mu})$ be a vector-valued ROM error indicator and

$$\boldsymbol{A}_k \equiv \frac{\partial \hat{\boldsymbol{\vartheta}}_k}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_k)^T \frac{\partial \hat{\boldsymbol{\vartheta}}_k}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_k) = \boldsymbol{Q}_k \boldsymbol{\Lambda}_k^2 \boldsymbol{Q}_k^T$$

Then, to first order[1],

$$\vartheta_k(\boldsymbol{\mu}) \equiv \left|\left|\hat{\boldsymbol{\vartheta}}_k(\boldsymbol{\mu})\right|\right|_2 = \left|\left|\frac{\partial \hat{\boldsymbol{\vartheta}}_k}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_k)(\boldsymbol{\mu} - \boldsymbol{\mu}_k)\right|\right|_2 = ||\boldsymbol{\mu} - \boldsymbol{\mu}_k||_{\boldsymbol{A}_k} \leq \Delta_k$$



Annotated schematic of trust-region: $\boldsymbol{q}_i = \boldsymbol{Q}_k \boldsymbol{e}_i$ and $\lambda_i = \boldsymbol{e}_i^T \boldsymbol{\Lambda}_k \boldsymbol{e}_i$

[1] assuming $\hat{\boldsymbol{\vartheta}}_k(\boldsymbol{\mu}_k) = 0$, i.e., ROM exact at trust-region center

A trust-region method with a strong connection to error indicators

Propose a trust-region method to solve the unconstrained stochastic PDE optimization problem

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad F(\boldsymbol{\mu}) \equiv \mathbb{E}\left[\hat{\mathcal{J}}(\boldsymbol{\mu},\,\cdot\,)\right]$$

that leverages trust-region subproblems of the form

$$\begin{aligned}\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & m_k(\boldsymbol{x}) \\ \text{subject to} \quad & \vartheta_k(\boldsymbol{x}) \leq \Delta_k,\end{aligned}$$

where $\vartheta_k(\boldsymbol{\mu})$ is an *error indicator*, i.e.,

there exists a constant[2] $\zeta > 0$ such that

$$|F(\boldsymbol{\mu}) - m_k(\boldsymbol{\mu})| \leq \zeta \vartheta_k(\boldsymbol{\mu})$$

[2] arbitrary, i.e., not tied to algorithmic parameters, and does not need to be computed or estimated

Error indicator for function values must account for ROM inaccuracy and sparse grid truncation error

Error indicator for function values are based on collocation of the **residual norm** to account for ROM error and **forward neighbors** to account for truncation error

Define

$$\hat{\mathcal{J}}(\boldsymbol{\mu}, \boldsymbol{\xi}) \equiv \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu}, \boldsymbol{\xi}), \boldsymbol{\mu}, \boldsymbol{\xi}) \qquad \hat{\mathcal{J}}_r(\boldsymbol{\mu}, \boldsymbol{\xi}) \equiv \mathcal{J}(\boldsymbol{\Phi y}(\boldsymbol{\mu}, \boldsymbol{\xi}), \boldsymbol{\mu}, \boldsymbol{\xi})$$

$$\hat{\boldsymbol{r}}(\boldsymbol{\mu}, \boldsymbol{\xi}) \equiv \boldsymbol{r}(\boldsymbol{\Phi y}(\boldsymbol{\mu}, \boldsymbol{\xi}), \boldsymbol{\mu}, \boldsymbol{\xi})$$

$$\begin{aligned}\left|\mathbb{E}\left[\hat{\mathcal{J}}(\boldsymbol{\mu}, \cdot)\right] - \mathbb{E}_{\mathcal{I}}\left[\hat{\mathcal{J}}_r(\boldsymbol{\mu}, \cdot)\right]\right| &\leq \mathbb{E}\left[\left|\hat{\mathcal{J}}(\boldsymbol{\mu}, \cdot) - \hat{\mathcal{J}}_r(\boldsymbol{\mu}, \cdot)\right|\right] + \left|\mathbb{E}_{\mathcal{I}^c}\left[\hat{\mathcal{J}}_r(\boldsymbol{\mu}, \cdot)\right]\right| \\ &\lesssim \zeta_1 \mathbb{E}_{\mathcal{I}\cup\mathcal{N}(\mathcal{I})}\left[||\hat{\boldsymbol{r}}(\boldsymbol{\mu}, \boldsymbol{\xi})||\right] + \left|\mathbb{E}_{\mathcal{N}(\mathcal{I})}\left[\hat{\mathcal{J}}_r(\boldsymbol{\mu}, \cdot)\right]\right|\end{aligned}$$

Error indicator for function gradients must account for primal and dual ROM inaccuracy and sparse grid truncation error

Error indicator for function gradients are based on collocation of the **primal and dual residual norm** to account for ROM error and **forward neighbors** to account for truncation error

Define

$$\hat{\boldsymbol{r}}^{\partial}(\boldsymbol{\mu}, \boldsymbol{\xi}) \equiv \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}}(\boldsymbol{\Phi}\boldsymbol{y}(\boldsymbol{\mu}, \ \boldsymbol{\xi}))\boldsymbol{\Phi}\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}, \ \boldsymbol{\xi}) + \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\mu}}(\boldsymbol{\Phi}\boldsymbol{y}(\boldsymbol{\mu}, \ \boldsymbol{\xi}), \ \boldsymbol{\mu}, \boldsymbol{\xi})$$

$$\begin{aligned}
&\left|\left|\mathbb{E}\left[\nabla_{\boldsymbol{\mu}}\hat{\mathcal{J}}(\boldsymbol{\mu}, \cdot)\right] - \mathbb{E}_{\mathcal{I}}\left[\nabla_{\boldsymbol{\mu}}\hat{\mathcal{J}}(\boldsymbol{\mu}, \cdot)\right]\right|\right| \\
&\quad \leq \mathbb{E}\left[\left|\left|\nabla_{\boldsymbol{\mu}}\hat{\mathcal{J}}(\boldsymbol{\mu}, \cdot) - \nabla_{\boldsymbol{\mu}}\hat{\mathcal{J}}_r(\boldsymbol{\mu}, \cdot)\right|\right|\right] + \left|\left|\mathbb{E}_{\mathcal{I}^c}\left[\nabla_{\boldsymbol{\mu}}\hat{\mathcal{J}}_r(\boldsymbol{\mu}, \cdot)\right]\right|\right| \\
&\quad \lesssim \zeta_2\mathbb{E}_{\mathcal{I}\cup\mathcal{N}(\mathcal{I})}\left[||\hat{\boldsymbol{r}}(\boldsymbol{\mu}, \cdot)||\right] + \zeta_3\mathbb{E}_{\mathcal{I}\cup\mathcal{N}(\mathcal{I})}\left[||\hat{\boldsymbol{r}}^{\partial}(\boldsymbol{\mu}, \cdot)||\right] + \left|\left|\mathbb{E}_{\mathcal{N}(\mathcal{I})}\left[\nabla_{\boldsymbol{\mu}}\hat{\mathcal{J}}_r(\boldsymbol{\mu}, \cdot)\right]\right|\right|
\end{aligned}$$

Model problem is the two-level approximation of the objective using reduced-order models and anisotropic sparse grids

The trust-region subproblem

$$\begin{aligned} \underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & m_k(\boldsymbol{x}) \\ \text{subject to} \quad & \vartheta_k(\boldsymbol{x}) \leq \Delta_k, \end{aligned}$$

is defined[3] as

$$\begin{aligned} m_k(\boldsymbol{\mu}) &= \mathbb{E}_{\mathcal{I}_k}\left[\hat{\mathcal{J}}_k^r(\boldsymbol{\mu},\,\cdot\,)\right] \\ \vartheta_k(\boldsymbol{\mu}) &= \alpha_1 \mathbb{E}_{\mathcal{I}_k\cup\mathcal{N}(\mathcal{I}_k)}\left[||\hat{\boldsymbol{r}}_k(\boldsymbol{\mu},\,\cdot\,)||\right] + \alpha_2\left|\mathbb{E}_{\mathcal{N}(\mathcal{I}_k)}\left[\hat{\mathcal{J}}_k^r(\boldsymbol{\mu},\,\cdot\,)\right]\right| \end{aligned}$$

An error indicator for the model gradient is also required and chosen as

$$\begin{aligned} \varphi_k = {} & \beta_1 \mathbb{E}_{\mathcal{I}_k\cup\mathcal{N}(\mathcal{I}_k)}\left[||\hat{\boldsymbol{r}}(\boldsymbol{\mu}_k,\,\cdot\,)||\right] + \beta_2 \mathbb{E}_{\mathcal{I}_k\cup\mathcal{N}(\mathcal{I}_k)}\left[||\hat{\boldsymbol{r}}^\partial(\boldsymbol{\mu}_k,\,\cdot\,)||\right] \\ & + \beta_3\left|\left|\mathbb{E}_{\mathcal{N}(\mathcal{I}_k)}\left[\nabla_{\boldsymbol{\mu}}\hat{\mathcal{J}}_r(\boldsymbol{\mu}_k,\,\cdot\,)\right]\right|\right| \end{aligned}$$

[3] $\hat{\mathcal{J}}_k^r$, $\hat{\boldsymbol{r}}_k$, and $\hat{\boldsymbol{r}}_k^\partial$ or defined exactly as $\hat{\mathcal{J}}^r$, $\hat{\boldsymbol{r}}$, and $\hat{\boldsymbol{r}}^\partial$ with $\boldsymbol{\Phi}_k$ in place of $\boldsymbol{\Phi}$.

Trust-Region Method for Managing Two-Level Inexactness

1: **Model update**: Choose m_k, i.e., an index set, $\mathcal{I}_k$, and reduced basis, $\mathbf{\Phi}_k$, such that

$$\vartheta_k(\boldsymbol{x}_k) \leq \kappa_{ucv}\Delta_k \qquad \varphi_k \quad \leq \kappa_{ubp}\,||\nabla m_k(\boldsymbol{x}_k)||$$

2: **Step computation**: Approximately solve the trust-region subproblem

$$\hat{\boldsymbol{x}}_k \quad = \quad \underset{\boldsymbol{x}\in\mathbb{R}^N}{\arg\min}\; m_k(\boldsymbol{x}) \quad \text{subject to} \quad \vartheta_k(\boldsymbol{x}) \leq \Delta_k$$

3: **Step acceptance**: Compute

$$\rho_k = \frac{F(\boldsymbol{x}_k) - F(\hat{\boldsymbol{x}}_k)}{m_k(\boldsymbol{x}_k) - m_k(\hat{\boldsymbol{x}}_k)}$$

if $\rho_k \geq \eta_1$ **then** $\boldsymbol{x}_{k+1} = \hat{\boldsymbol{x}}_k$ **else** $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k$ **end if**

4: **Trust-region update**:

if $\rho_k \leq \eta_1$ **then** $\Delta_{k+1} \in (0, \gamma\vartheta_k(\hat{\boldsymbol{x}}_k)]$ **end if**
if $\rho_k \in (\eta_1, \eta_2)$ **then** $\Delta_{k+1} \in [\gamma\vartheta_k(\hat{\boldsymbol{x}}_k), \Delta_k]$ **end if**
if $\rho_k \geq \eta_2$ **then** $\Delta_{k+1} \in [\Delta_k, \Delta_{\max}]$ **end if**

Dimension-Adaptive Sparse Grids and Greedy Sampling to Control Two-Level Inexactness

while $\mathbb{E}_{\mathcal{N}(\mathcal{I}_k)}\left[\mathcal{J}_k^r(\boldsymbol{\mu},\,\cdot\,)\right] > \frac{1}{2\alpha_2}\kappa_{ucv}\Delta_k$ **do**

Refine index set: Let $\mathbf{j} = \underset{\mathbf{j}\in\mathcal{N}(\mathcal{I}_k)}{\arg\max}\left|\mathbb{E}_{\mathbf{j}}\left[\mathcal{J}_k^r(\boldsymbol{\mu},\,\cdot\,)\right]\right|$

$$\mathcal{I}_k \leftarrow \mathcal{I}_k \cup \{\mathbf{j}\}$$

while $\mathbb{E}_{\mathcal{I}_k\cup\mathcal{N}(\mathcal{I}_k)}\left[||\hat{\boldsymbol{r}}_k(\boldsymbol{\mu}_k,\,\cdot\,)||\right] > \frac{1}{2\alpha_1}\kappa_{ucv}\Delta_k$ **do**

Evaluate error indicator: For each $\boldsymbol{\xi}_j \in \boldsymbol{\Xi}_{\mathcal{I}_k\cup\mathcal{N}(\mathcal{I}_k)}$, compute

$$r_j = ||\hat{\boldsymbol{r}}_k(\boldsymbol{\mu}_k,\,\boldsymbol{\xi}_j\,)||$$

where ω_j is the quadrature weight associated with node $\boldsymbol{\xi}_j$

Sample high-dimensional model: Let $j^* = \arg\max r_j$ and compute

$$\boldsymbol{w}(\boldsymbol{\mu}_k,\boldsymbol{\xi}_{j^*}),\ \frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_k,\boldsymbol{\xi}_{j^*})$$

and update reduced-order basis, $\boldsymbol{\Phi}_k$, using SSPOD

end while

end while

Trust-Region Convergence Theory

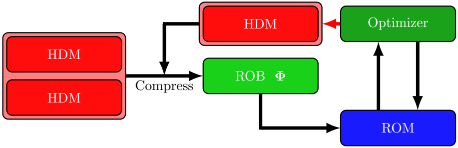
The trust-region method guarantees

$$\liminf_{k\to\infty} ||\nabla_{\boldsymbol{\mu}} \mathbb{E}\left[\mathcal{J}_k(\boldsymbol{\mu}_k,\,\cdot\,)\right]|| = 0$$

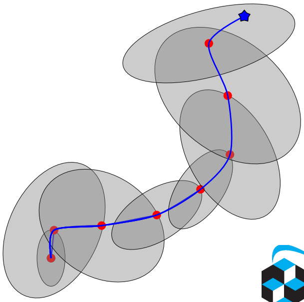
provided there exists constants ζ, $\tau > 0$ such that

$$\left|\mathbb{E}\left[\mathcal{J}_k(\boldsymbol{\mu},\,\cdot\,)\right] - m_k(\boldsymbol{\mu})\right| \leq \zeta \vartheta_k(\boldsymbol{\mu})$$

$$||\nabla_{\boldsymbol{\mu}} \mathbb{E}\left[\mathcal{J}_k(\boldsymbol{\mu}_k,\,\cdot\,)\right] - \nabla_{\boldsymbol{\mu}} m_k(\boldsymbol{\mu}_k)|| \leq \tau \varphi_k$$



Schematic



$\boldsymbol{\mu}$-space

Optimal control of steady Burgers' equation

- **Optimization problem:**

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad \frac{1}{2}\int_{\boldsymbol{\Xi}} \rho(\boldsymbol{\xi}) \int_0^1 (u(\boldsymbol{\xi},\, x) - \bar{w})^2 \, dx \, d\xi + \frac{\alpha}{2}\int_0^1 z(\boldsymbol{\mu},\, x)^2 \, dx$$

where $u(\boldsymbol{\xi},\, x)$ solves

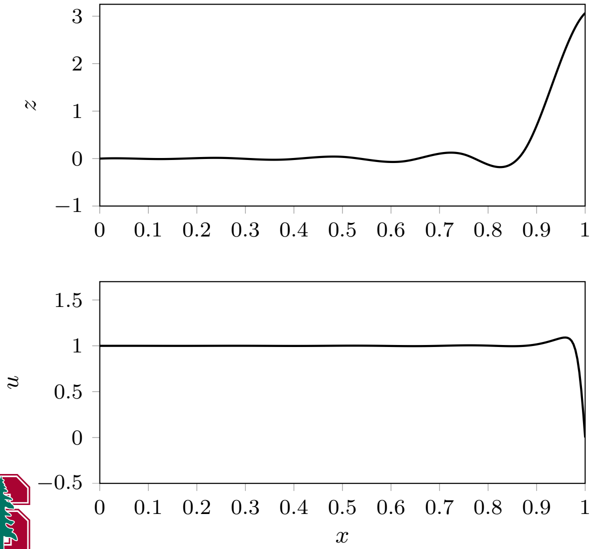
$$\begin{aligned} -\nu(\boldsymbol{\xi})\partial_{xx}u(\boldsymbol{\xi},\, x) + u(\boldsymbol{\xi},\, x)\partial_x u(\boldsymbol{\xi},\, x) &= z(\boldsymbol{\mu},\, x) \quad x \in (0,\, 1), \quad && \boldsymbol{\xi} \in \boldsymbol{\Xi} \\ u(\boldsymbol{\xi},\, 0) = d_0(\boldsymbol{\xi}) \qquad u(\boldsymbol{\xi},\, 1) &= d_1(\boldsymbol{\xi}) && \boldsymbol{\xi} \in \boldsymbol{\Xi} \end{aligned}$$

- **Desired state:** $\bar{\boldsymbol{w}} \equiv 1$
- **Stochastic Space:** $\boldsymbol{\Xi} = [-1,\, 1]^3$, $\rho(\boldsymbol{\xi})d\boldsymbol{\xi} = 2^{-3}d\boldsymbol{\xi}$

$$\nu(\boldsymbol{\xi}) = 10^{\boldsymbol{\xi}_1 - 2} \qquad d_0(\boldsymbol{\xi}) = 1 + \frac{\boldsymbol{\xi}_2}{1000} \qquad d_1(\boldsymbol{\xi}) = \frac{\boldsymbol{\xi}_3}{1000}$$

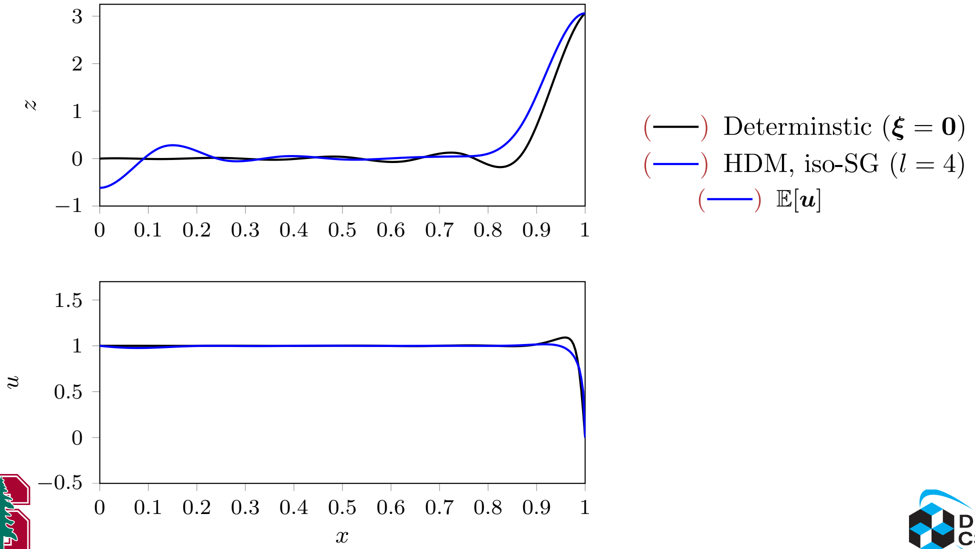
- **Parametrization:** $z(\boldsymbol{\mu},\, x)$ – cubic splines with 9 knots, $n_{\boldsymbol{\mu}} = 11$

Error-Based Trust-Region Method Recovers Optimal Control

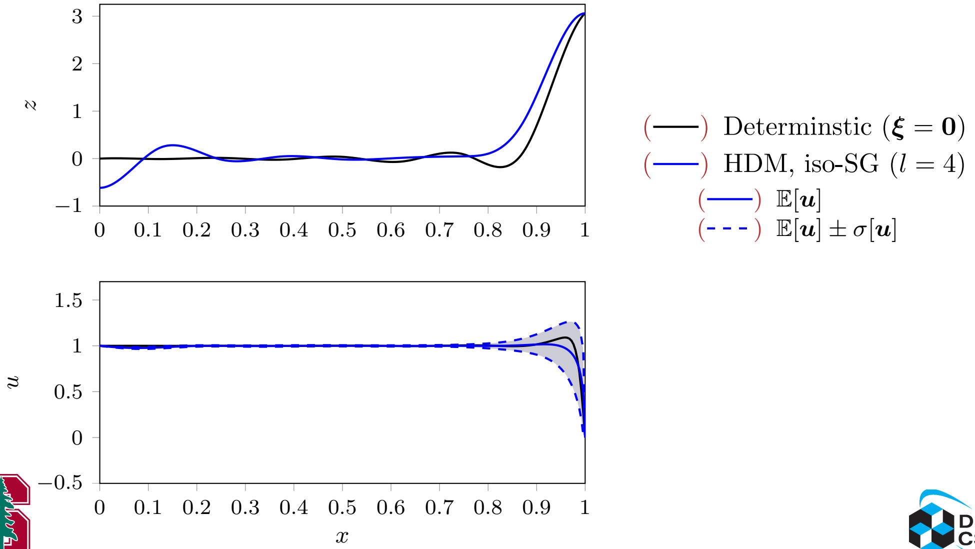


(——) Determinstic ($\boldsymbol{\xi} = \mathbf{0}$)

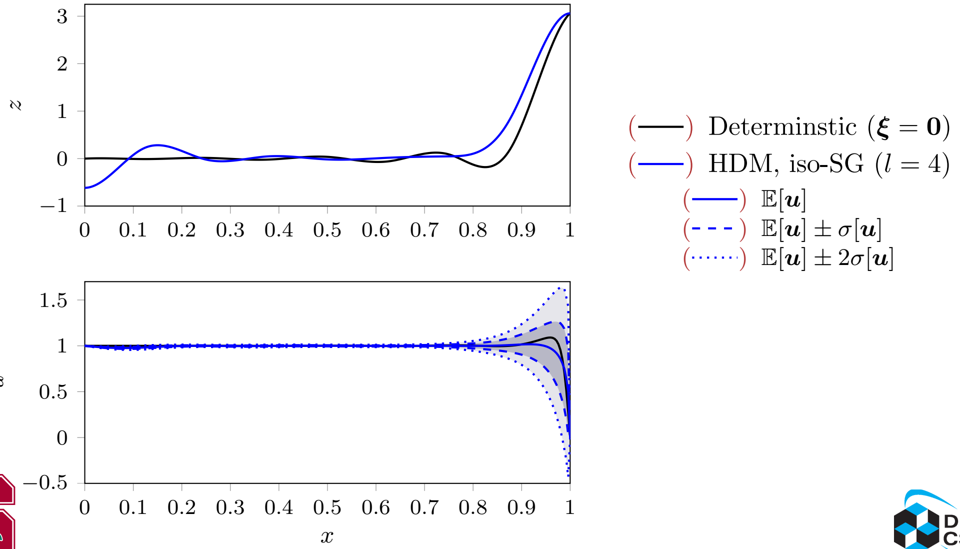
Error-Based Trust-Region Method Recovers Optimal Control



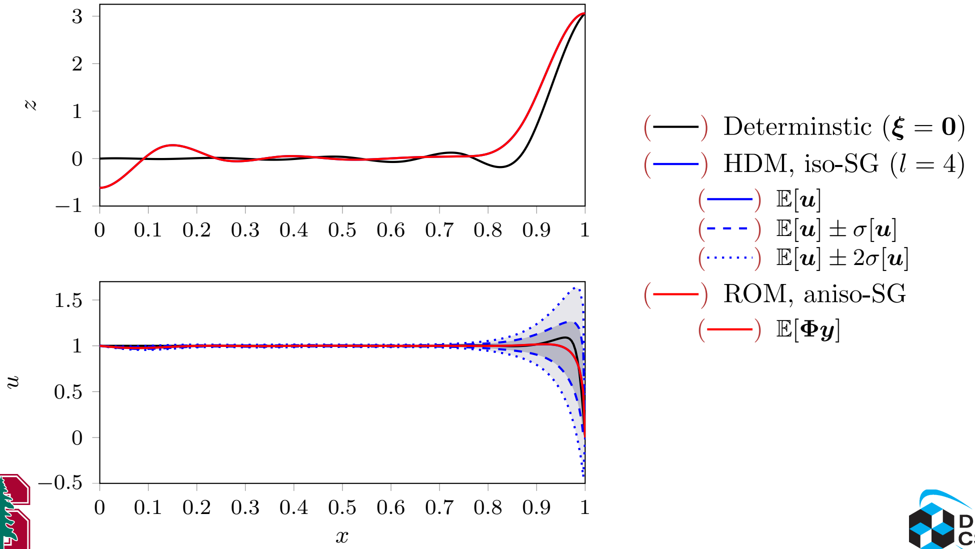
Error-Based Trust-Region Method Recovers Optimal Control



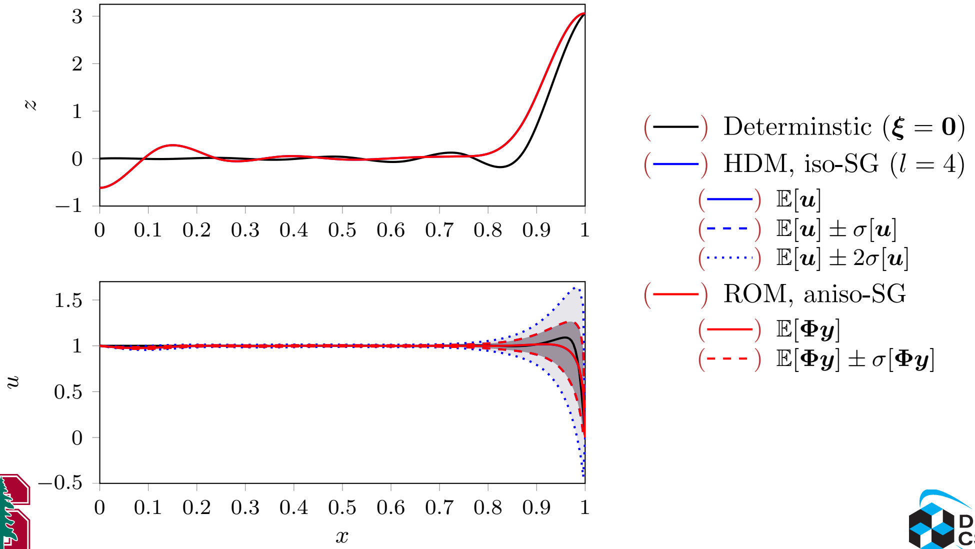
Error-Based Trust-Region Method Recovers Optimal Control



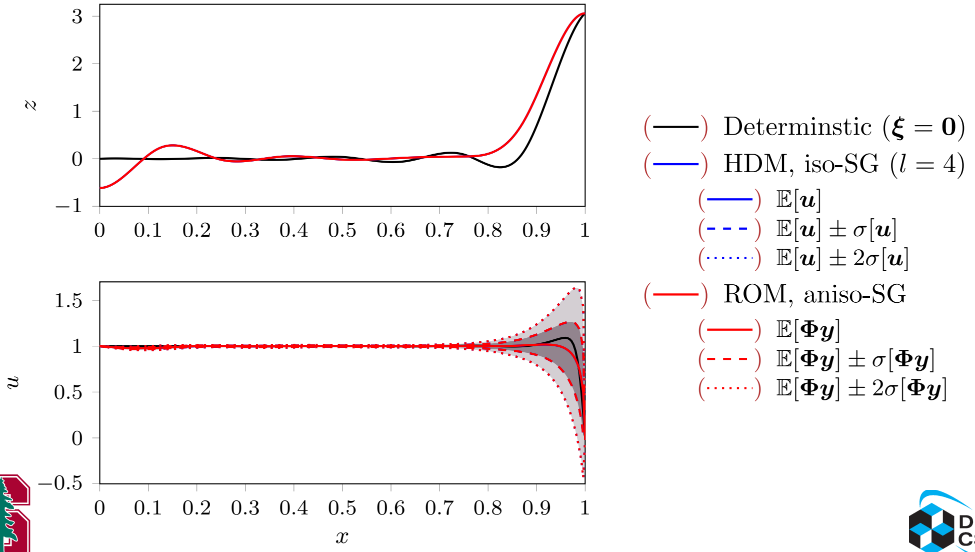
Error-Based Trust-Region Method Recovers Optimal Control



Error-Based Trust-Region Method Recovers Optimal Control



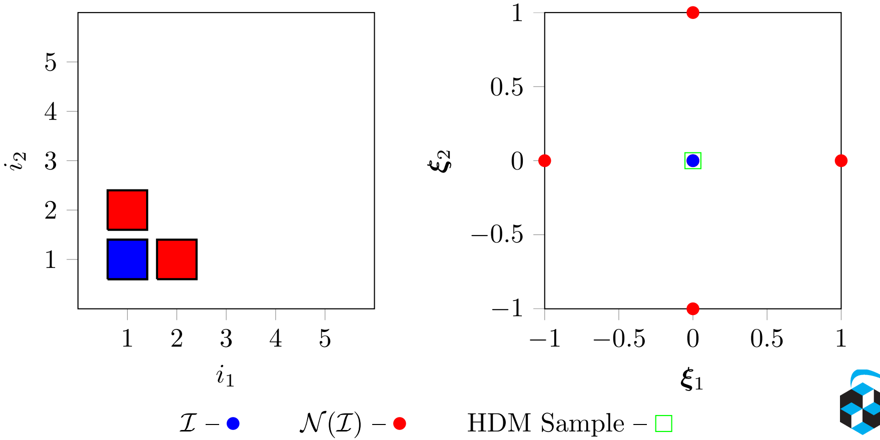
Error-Based Trust-Region Method Recovers Optimal Control



Error-Based Trust-Region Method Requires Very Few HDM Queries to Converges to Optimal Control

Prior to each trust-region subproblem, the model (sparse grid, $\mathcal{I}_k$, and basis, $\mathbf{\Phi}_k$) must be constructed such that error indicators are below a tolerance

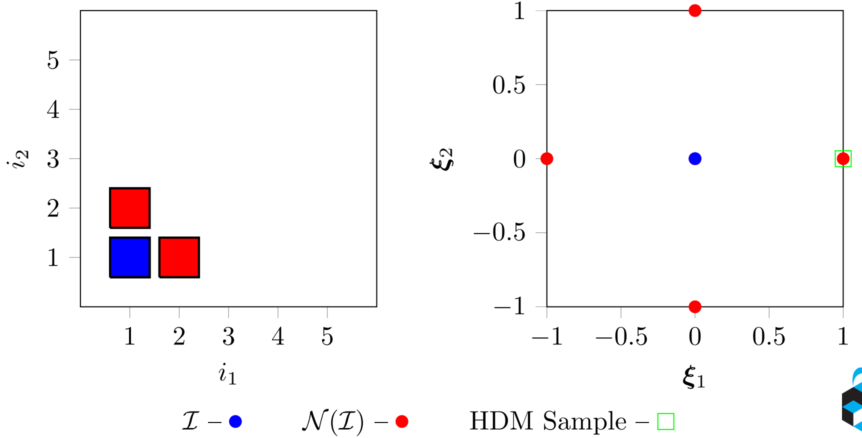
Iteration 1



Error-Based Trust-Region Method Requires Very Few HDM Queries to Converges to Optimal Control

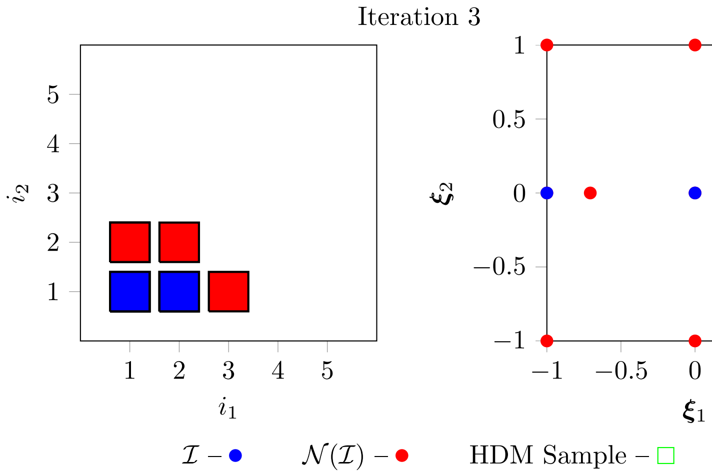
Prior to each trust-region subproblem, the model (sparse grid, $\mathcal{I}_k$, and basis, $\mathbf{\Phi}_k$) must be constructed such that error indicators are below a tolerance

Iteration 2



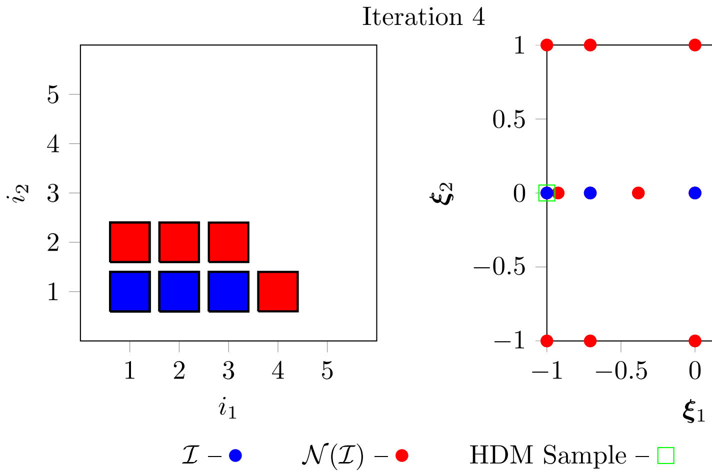
Error-Based Trust-Region Method Requires Very Few HDM Queries to Converges to Optimal Control

Prior to each trust-region subproblem, the model (sparse grid, $\mathcal{I}_k$, and basis, $\mathbf{\Phi}_k$) must be constructed such that error indicators are below a tolerance



Error-Based Trust-Region Method Requires Very Few HDM Queries to Converges to Optimal Control

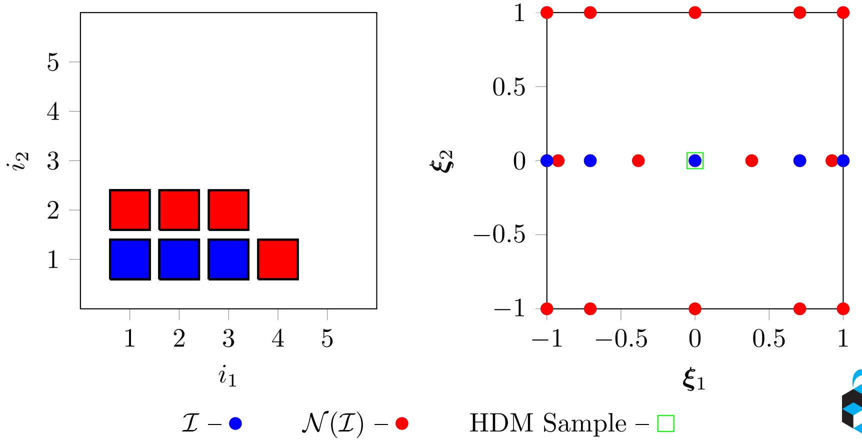
Prior to each trust-region subproblem, the model (sparse grid, $\mathcal{I}_k$, and basis, $\mathbf{\Phi}_k$) must be constructed such that error indicators are below a tolerance



Error-Based Trust-Region Method Requires Very Few HDM Queries to Converges to Optimal Control

Prior to each trust-region subproblem, the model (sparse grid, $\mathcal{I}_k$, and basis, $\mathbf{\Phi}_k$) must be constructed such that error indicators are below a tolerance

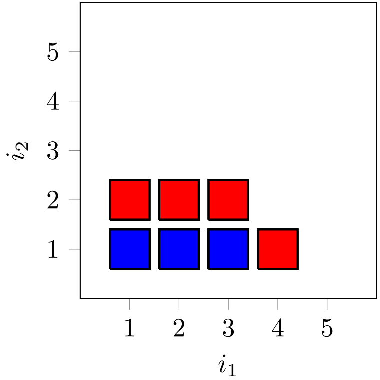
Iteration 5

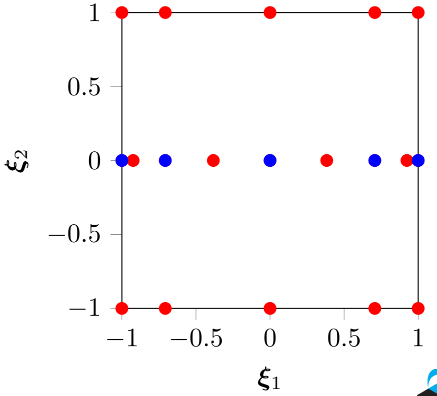


Error-Based Trust-Region Method Requires Very Few HDM Queries to Converges to Optimal Control

Prior to each trust-region subproblem, the model (sparse grid, $\mathcal{I}_k$, and basis, $\mathbf{\Phi}_k$) must be constructed such that error indicators are below a tolerance

Iteration 6





$\mathcal{I}$ – ● $\mathcal{N}(\mathcal{I})$ – ● HDM Sample – □

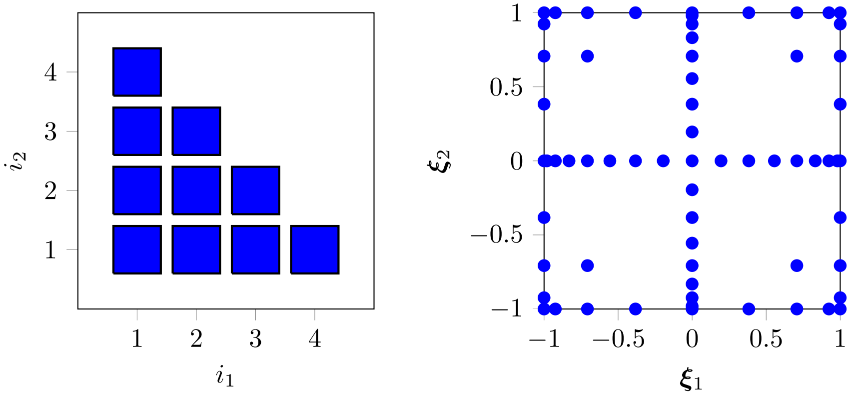
Global Convergence of Trust-Region Method

The trust-region method finds a sequence of parameters $\boldsymbol{\mu}_k$ such that the gradient of HDM ($||\nabla\mathcal{J}(\boldsymbol{\mu})||$) converges to 0 from arbitrary starting point – global convergence – while incurring few HDM queries

$m_k(\hat{\boldsymbol{\mu}}_k)$	$\|\|\nabla m_k(\hat{\boldsymbol{\mu}}_k)\|\|$	$\mathcal{J}(\hat{\boldsymbol{\mu}}_k)$	$\|\|\nabla\mathcal{J}(\hat{\boldsymbol{\mu}}_k)\|\|$	ρ_k	Δ_k	Success?
3.8783e-03	3.3779e-03	8.3351e-03	6.8542e-03	-	-	-
3.1121e-03	2.0393e-04	7.2687e-03	7.0676e-03	1.3918e+00	1.0000e+02	True
3.0474e-03	7.7900e-05	6.8352e-03	3.3518e-03	3.3943e-01	2.0000e+02	True
1.1910e-02	3.7019e-04	9.7269e-03	3.5655e-03	-2.6141e-01	1.0000e+02	False
6.3680e-03	9.6334e-06	6.3591e-03	8.6182e-05	1.0070e+00	2.8202e-03	True
6.3587e-03	7.2419e-07	6.3589e-03	7.2665e-07	1.0018e+00	5.6404e-03	True

HDM Queries	ROM Queries (max size)
4	3720 (48)

Comparison to Stochastic Optimization with Collocation on 4-Level Isotropic Sparse Grid: 1000-Fold Reduction in HDM Queries



Iterations (L-BFGS)	HDM Queries	$\|\|\nabla \mathcal{J}\|\|$
34	6372	6.6064e-07

Leveraging and Managing Two-Levels of Inexactness for Efficient Stochastic PDE-Constrained Optimization

Conclusions

- *Trust-region method* with strong connection to model error indicators
- Two-level approximation of moments of quantities of interest of SPDE
  - *Anisotropic sparse grids* - inexact integration
  - *Reduced-order models* - inexact evaluations
- Two-level inexactness managed through trust-region method

Future work

- Comparison with traditional trust-region method (TRPOD)
- Comparison with offline/online approaches
- Incorporate nonlinear constraints
- Local reduced-order models for improved efficiency
- Less expensive error indicators for cheaper trust-region subproblems [Drohmann and Carlberg, 2014]

Acknowledgement

References I

Chen, P. and Quarteroni, A. (2014).
Weighted reduced basis method for stochastic optimal control problems with elliptic pde constraint.
SIAM/ASA Journal on Uncertainty Quantification, 2(1):364–396.

Chen, P. and Quarteroni, A. (2015).
A new algorithm for high-dimensional uncertainty quantification based on dimension-adaptive sparse grid approximation and reduced basis methods.
Journal of Computational Physics, 298:176–193.

Drohmann, M. and Carlberg, K. (2014).
The romes method for statistical modeling of reduced-order-model error.
SIAM Journal on Uncertainty Quantification.

Gerstner, T. and Griebel, M. (2003).
Dimension–adaptive tensor–product quadrature.
Computing, 71(1):65–87.

Kouri, D. P., Heinkenschloss, M., Ridzal, D., and van Bloemen Waanders, B. G. (2013).
A trust-region algorithm with adaptive stochastic collocation for pde optimization under uncertainty.
SIAM Journal on Scientific Computing, 35(4):A1847–A1879.

References II

Kouri, D. P., Heinkenschloss, M., Ridzal, D., and van Bloemen Waanders, B. G. (2014).
Inexact objective function evaluations in a trust-region algorithm for pde-constrained optimization under uncertainty.
SIAM Journal on Scientific Computing, 36(6):A3011–A3029.

Tiesler, H., Kirby, R. M., Xiu, D., and Preusser, T. (2012).
Stochastic collocation for optimal control problems with stochastic pde constraints.
SIAM Journal on Control and Optimization, 50(5):2659–2682.