Efficient PDE Optimization under Uncertainty using Adaptive Model Reduction and Sparse Grids

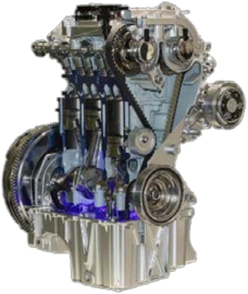
Matthew J. Zahr

Advisor: Charbel Farhat
Computational and Mathematical Engineering
Stanford University

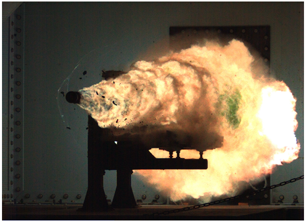
Joint work with: Kevin Carlberg (Sandia CA), Drew Kouri (Sandia NM)

CME 500 Seminar
Stanford University, Stanford, CA
April 11, 2016

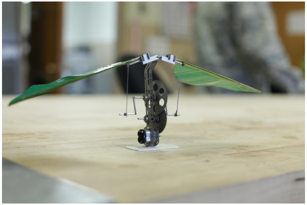
Current interest in computational physics reaches far beyond analysis of a single configuration of a physical system into ***design*** *(shape and topology) and* ***control*** *in an* ***uncertain*** *setting*



Engine System



EM Launcher



Micro-Aerial Vehicle

PDE-constrained optimization formulation

Goal: Rapidly solve PDE-constrained optimization problem of the form

$$\begin{aligned} \underset{\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathcal{J}(\boldsymbol{u}, \boldsymbol{\mu}) \\ \text{subject to} \quad & \boldsymbol{r}(\boldsymbol{u}; \boldsymbol{\mu}) = 0 \end{aligned}$$

- $\boldsymbol{r} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \to \mathbb{R}^{n_{\boldsymbol{u}}}$ — discretized PDE
- $\mathcal{J} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \to \mathbb{R}$ — quantity of interest
- $\boldsymbol{u} \in \mathbb{R}^{n_{\boldsymbol{u}}}$ — PDE state vector
- $\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}$ — optimization parameters

Virtually all expense emanates from primal/dual PDE solvers

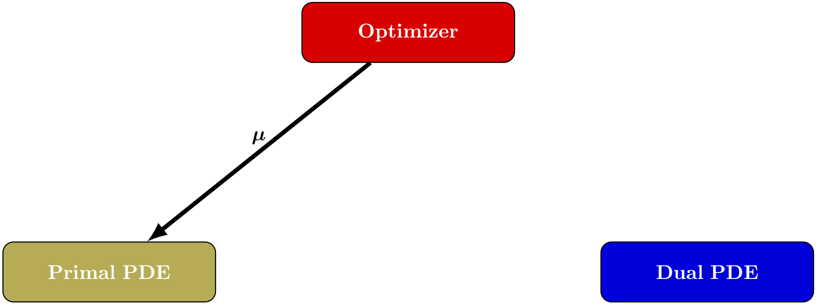
Optimizer

Primal PDE

Dual PDE

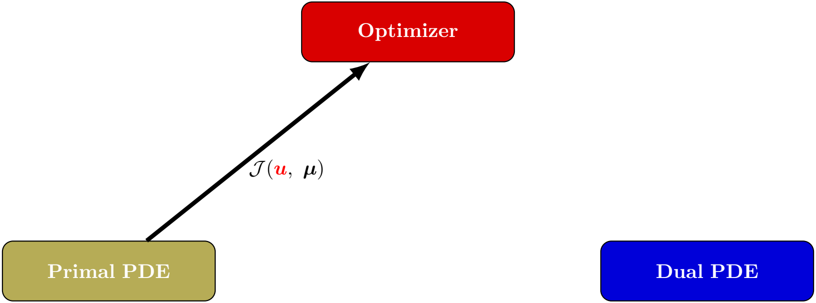
Nested approach to PDE-constrained optimization

Virtually all expense emanates from primal/dual PDE solvers



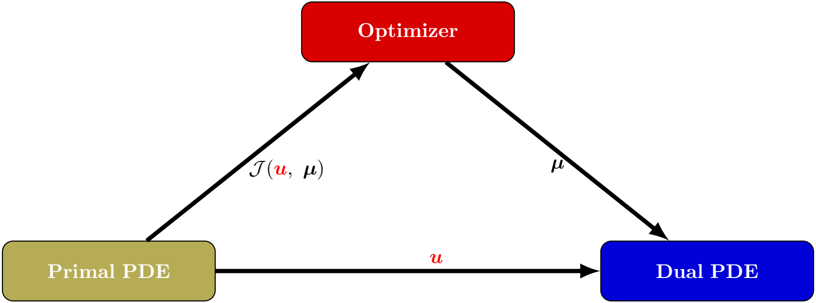
Nested approach to PDE-constrained optimization

Virtually all expense emanates from primal/dual PDE solvers



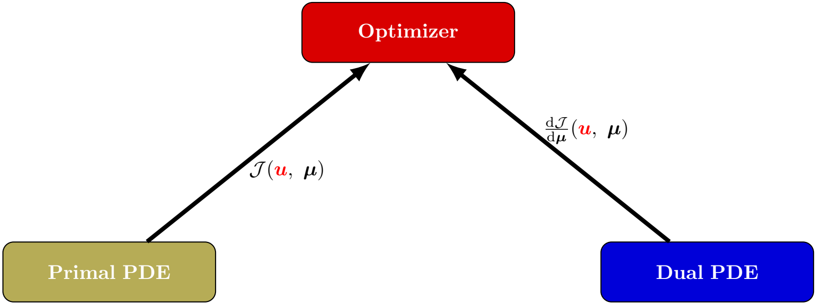
Nested approach to PDE-constrained optimization

Virtually all expense emanates from primal/dual PDE solvers



Nested approach to PDE-constrained optimization

Virtually all expense emanates from primal/dual PDE solvers



Maximum lift-to-drag airfoil configuration

Maximum stiffness lacrosse head

Optimal control based on compressible Navier-Stokes

Energy = 9.4096e+00	Energy = 4.9476e+00	Energy = 4.6110e+00
Thrust = 1.7660e-01	Thrust = 2.5000e+00	Thrust = 2.5000e+00

Initial	Optimal Control	Optimal Shape/Control

[Zahr and Persson, 2016], [Zahr et al., 2016]

PDE-constrained optimization under uncertainty

Goal: Efficiently solve stochastic PDE-constrained optimization problems

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

- $\boldsymbol{r} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \to \mathbb{R}^{n_{\boldsymbol{u}}}$ discretized stochastic PDE
- $\mathcal{J} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \to \mathbb{R}$ quantity of interest
- $\boldsymbol{u} \in \mathbb{R}^{n_{\boldsymbol{u}}}$ PDE state vector
- $\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}$ (deterministic) optimization parameters
- $\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ stochastic parameters
- $\mathbb{E}[\mathcal{F}] \equiv \displaystyle\int_{\boldsymbol{\Xi}} \mathcal{F}(\boldsymbol{\xi})\rho(\boldsymbol{\xi})\, d\boldsymbol{\xi}$

*Each function evaluation requires integration over stochastic space – **expensive***

Proposed approach: managed two-level inexactness

Two levels of inexactness to obtain an inexpensive, approximate version of the stochastic optimization problem

- **Anisotropic sparse grids** used for *inexact integration* of risk measures
- **Reduced-order models** used for *inexact evaluations* at collocation nodes

Manage inexactness with trust-region-like method

- Embedded in globally convergent **trust-region-like** algorithm with a strong connection to error indicators and refinement mechanism
- **Error indicators** 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

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: Define the index set $\mathcal{I} \subset \mathbb{N}^{n_{\boldsymbol{\xi}}}$ and

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

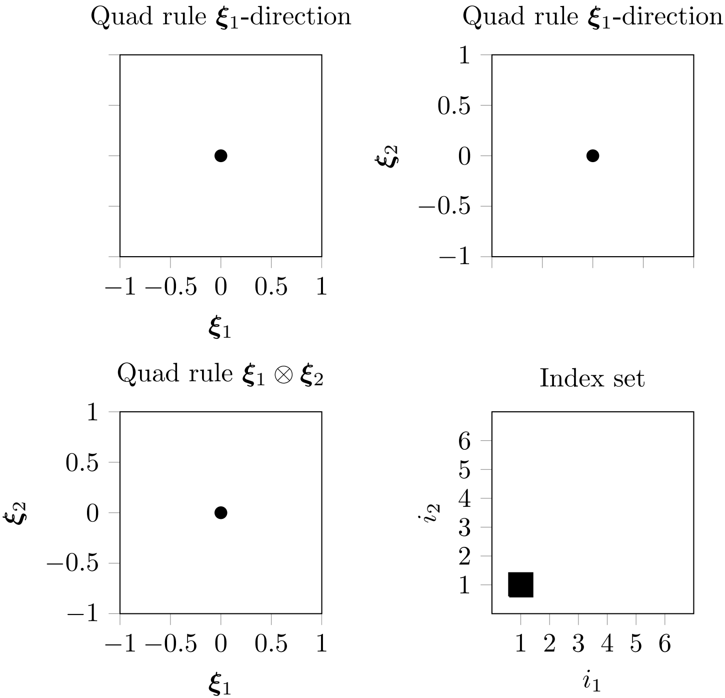
Forward Neighbors:

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

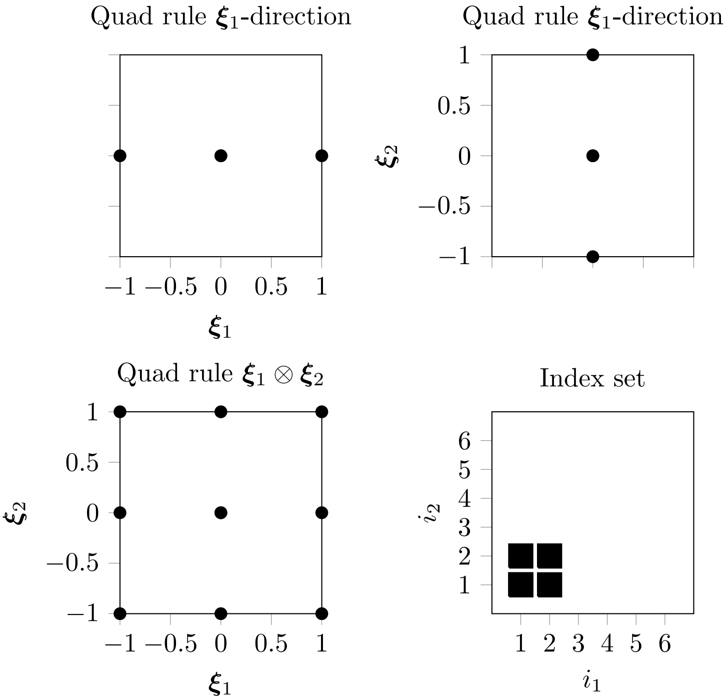
Truncation Error: [Gerstner and Griebel, 2003, Kouri et al., 2013]

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

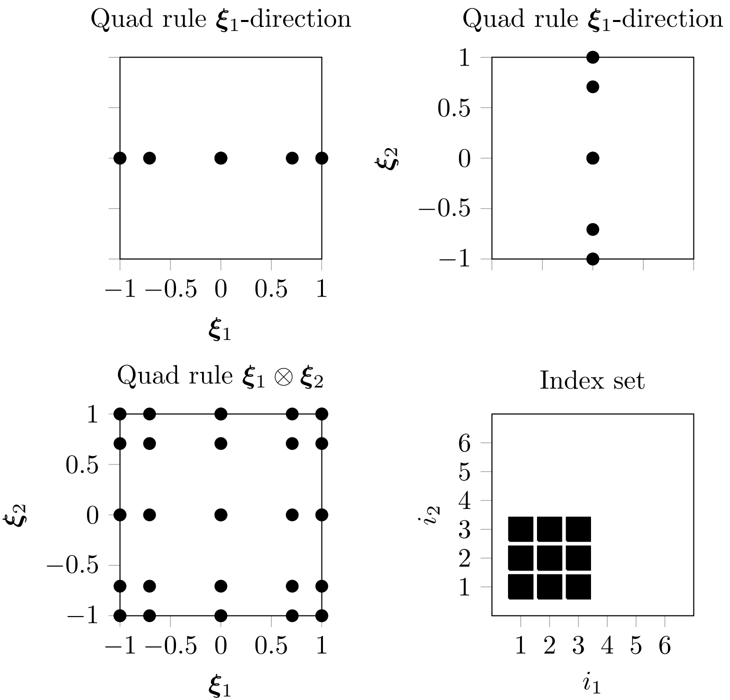
Tensor product quadrature



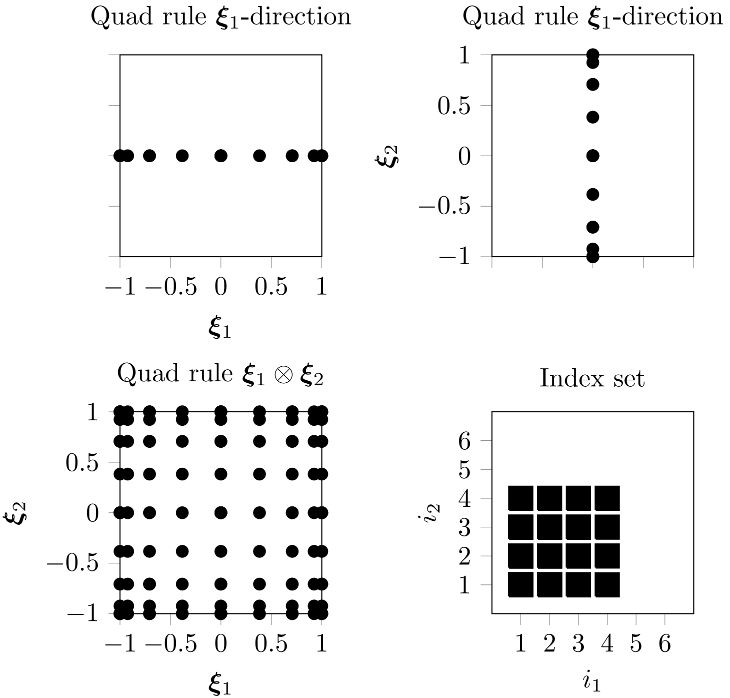
Tensor product quadrature



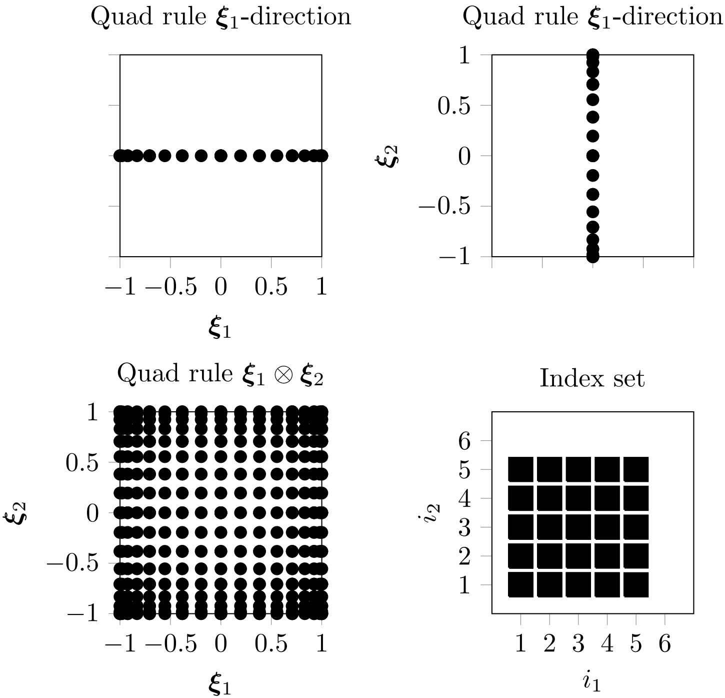
Tensor product quadrature



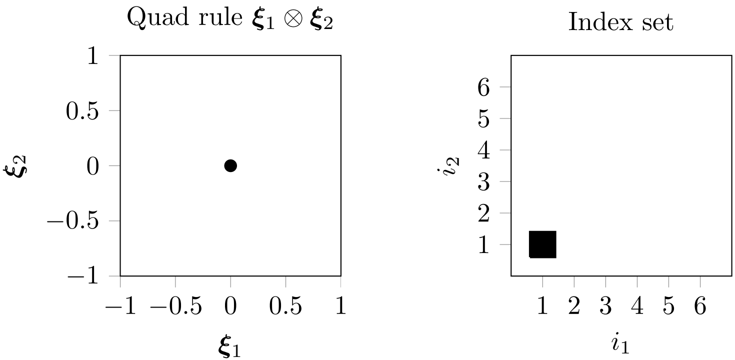
Tensor product quadrature



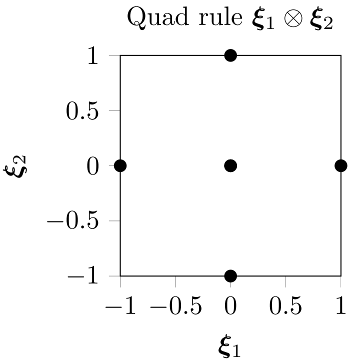
Tensor product quadrature

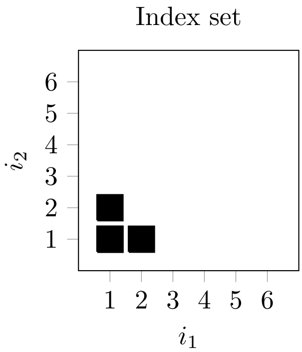


Isotropic sparse grid quadrature

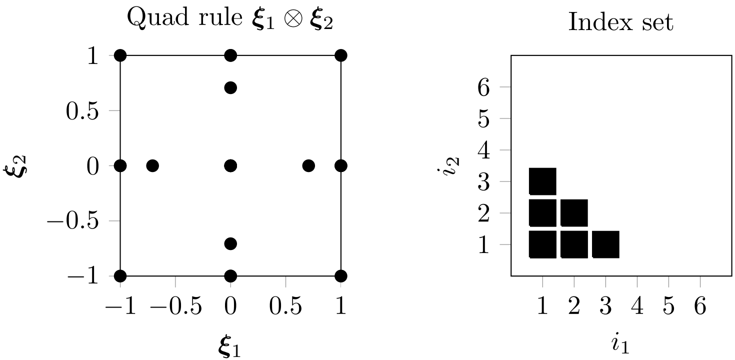


Isotropic sparse grid quadrature

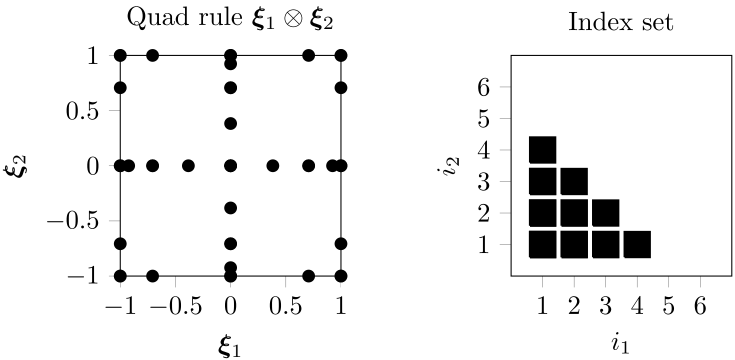




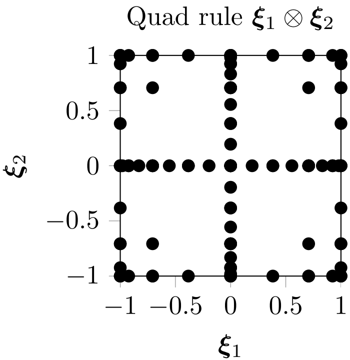
Isotropic sparse grid quadrature

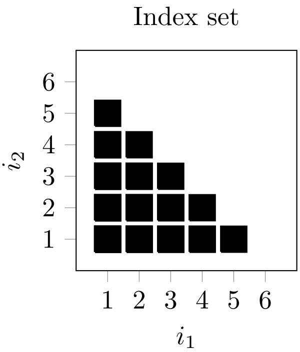


Isotropic sparse grid quadrature

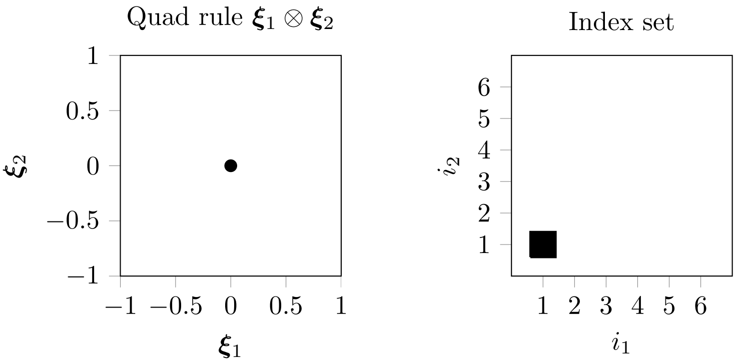


Isotropic sparse grid quadrature

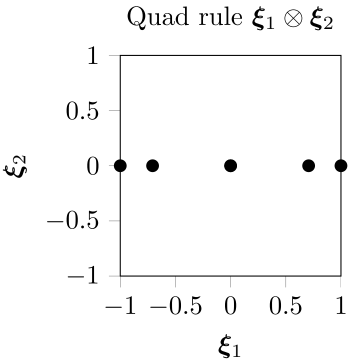


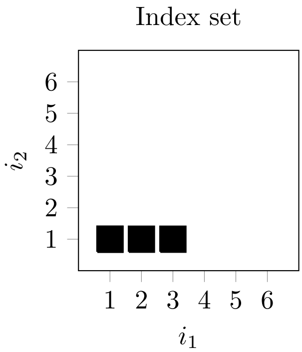


Anisotropic sparse grid quadrature

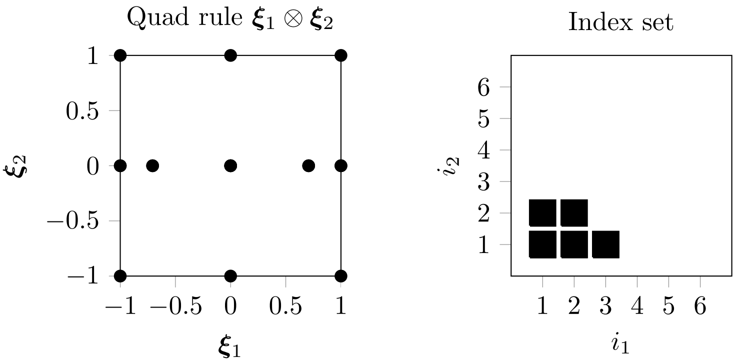


Anisotropic sparse grid quadrature

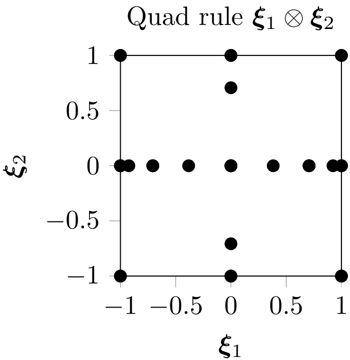


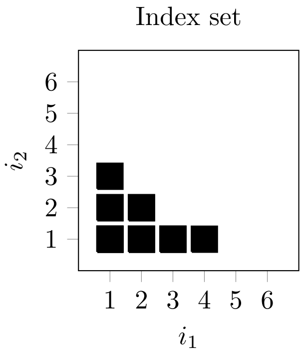


Anisotropic sparse grid quadrature

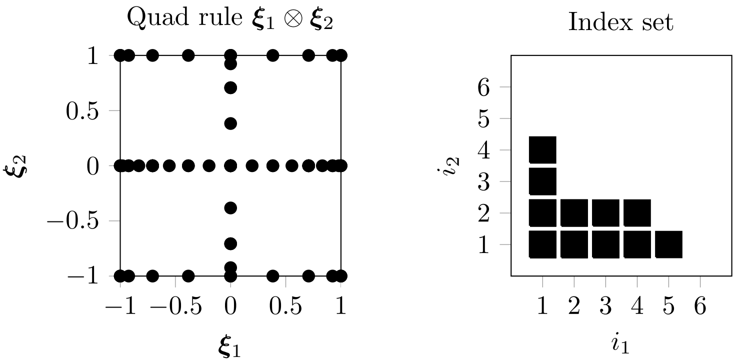


Anisotropic sparse grid quadrature

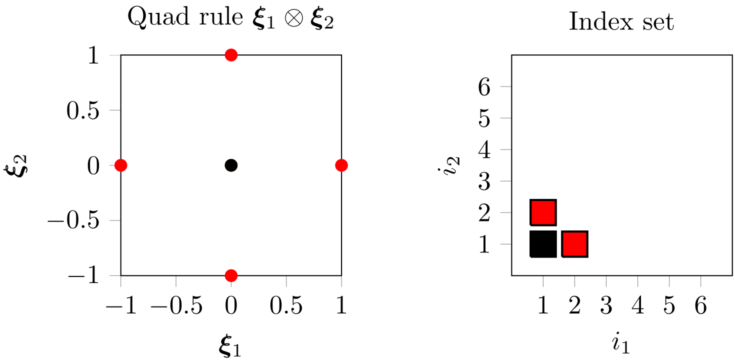




Anisotropic sparse grid quadrature

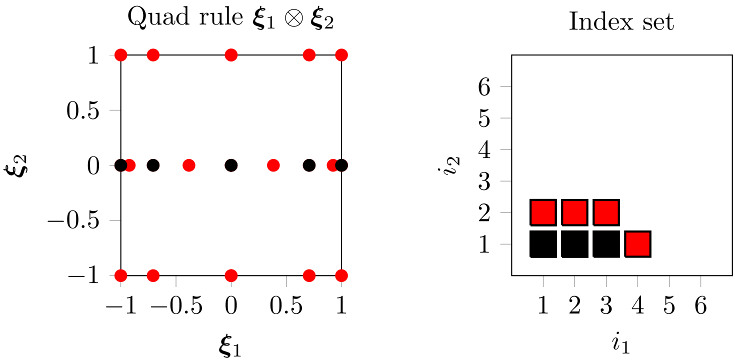


Anisotropic sparse grid quadrature: neighbors

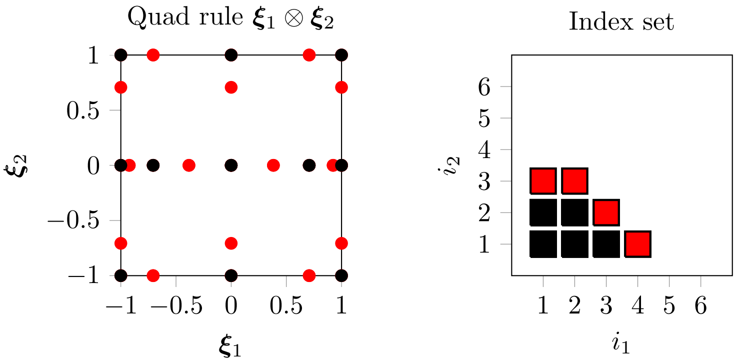


Index set ($\mathcal{I}$) – ● Neighbors ($\mathcal{N}(\mathcal{I})$) – ●

Anisotropic sparse grid quadrature: neighbors

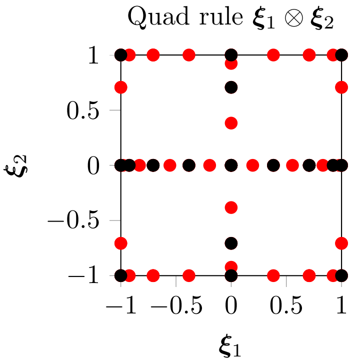


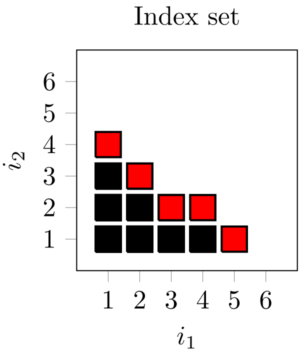
Index set ($\mathcal{I}$) – ●   Neighbors ($\mathcal{N}(\mathcal{I})$) – ●



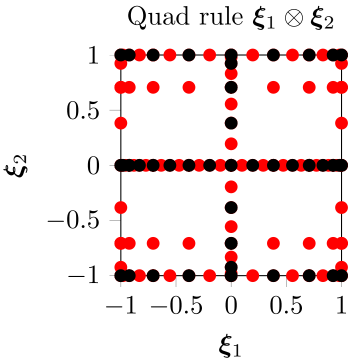
Index set $(\mathcal{I})$ – ● Neighbors $(\mathcal{N}(\mathcal{I}))$ – ●

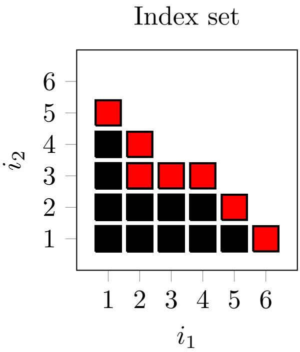
Anisotropic sparse grid quadrature: neighbors





Index set ($\mathcal{I}$) – ●   Neighbors ($\mathcal{N}(\mathcal{I})$) – ●





Index set ($\mathcal{I}$) – ● Neighbors ($\mathcal{N}(\mathcal{I})$) – ●

Stochastic collocation via anisotropic sparse grids

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

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

$$\Downarrow$$

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

Projection-based model reduction to reduce PDE size

- Model reduction ansatz: *state vector lies in low-dimensional subspace*

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

  - $\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_{\boldsymbol{u}}^1 & \cdots & \boldsymbol{\phi}_{\boldsymbol{u}}^{k_{\boldsymbol{u}}} \end{bmatrix} \in \mathbb{R}^{n_{\boldsymbol{u}} \times k_{\boldsymbol{u}}}$ is the reduced basis ($n_{\boldsymbol{u}} \gg k_{\boldsymbol{u}}$)
  - $\boldsymbol{y} \in \mathbb{R}^{k_{\boldsymbol{u}}}$ are the reduced coordinates of $\boldsymbol{u}$

- Substitute assumption into high-dimensional model $\boldsymbol{r}(\boldsymbol{u},\, \boldsymbol{\mu},\, \boldsymbol{\xi}) = 0$ and use Galerkin projection to obtain a 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

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

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

$$\begin{aligned} \boldsymbol{\Phi}_{\boldsymbol{X}} &= \mathrm{POD}(\boldsymbol{X}) \\ \boldsymbol{\Phi}_{\boldsymbol{Y}} &= \mathrm{POD}(\boldsymbol{Y}) \end{aligned}$$

- Concatenate and orthogonalize to get reduced-order basis

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

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

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

$$\Downarrow$$

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

$$\Downarrow$$

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

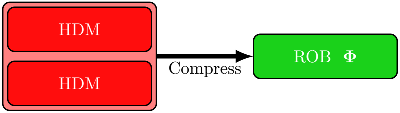
Trust-region framework for optimization

Schematic

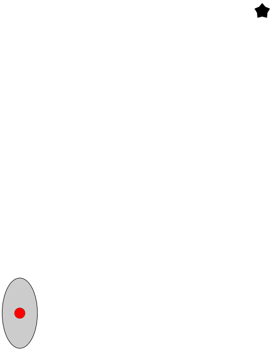
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust-region framework for optimization



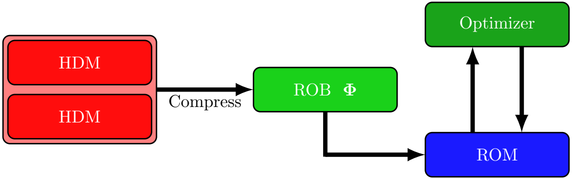
Schematic



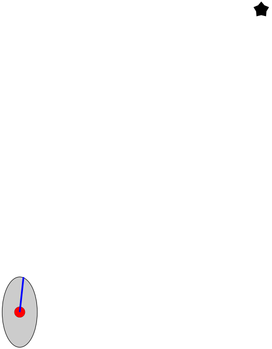
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust-region framework for optimization



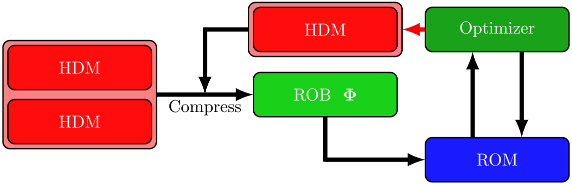
Schematic



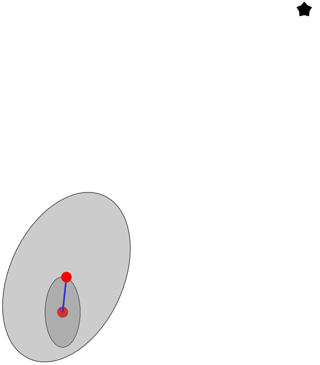
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust-region framework for optimization



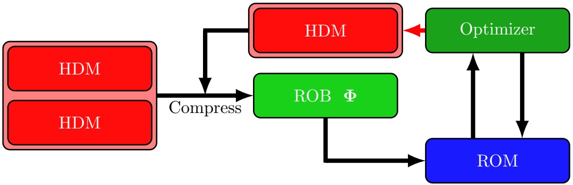
Schematic



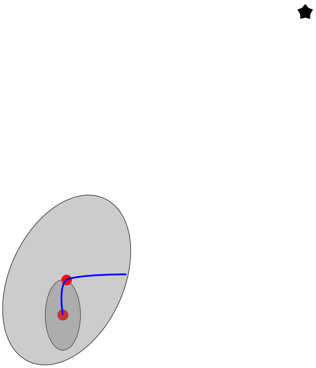
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust-region framework for optimization



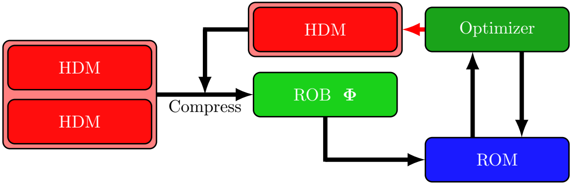
Schematic



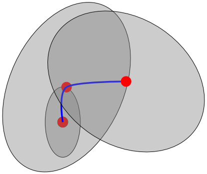
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust-region framework for optimization



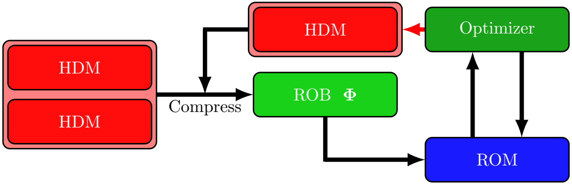
Schematic



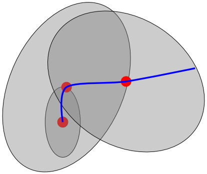
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust-region framework for optimization

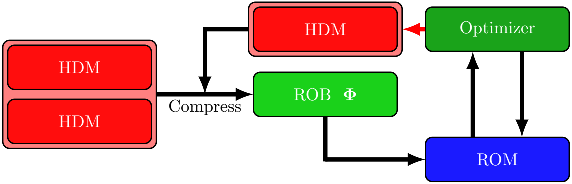


Schematic

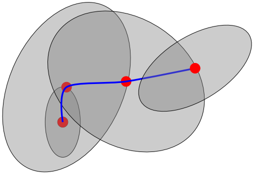


$\boldsymbol{\mu}$-space

Breakdown of Computational Effort



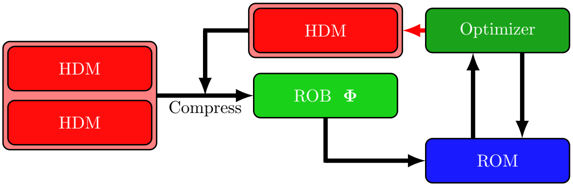
Schematic



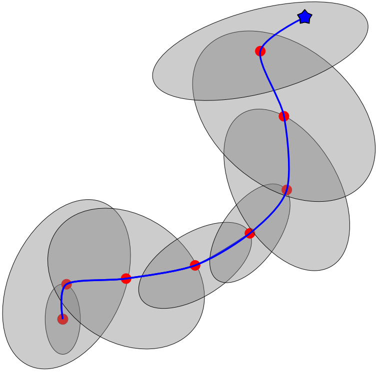
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust-region framework for optimization



Schematic



$\boldsymbol{\mu}$-space

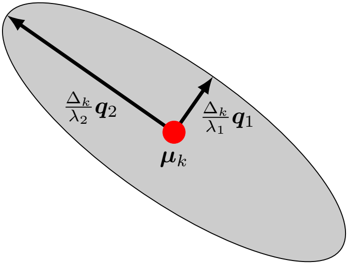
Breakdown of Computational Effort

Trust-regions based on error indicators: motivation

Let $\boldsymbol{\vartheta}_k(\boldsymbol{\mu})$ be a vector-valued error indicator, then to first order[1]

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

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



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 $\boldsymbol{\vartheta}_k(\boldsymbol{\mu}_k) = 0$, i.e., exactness at trust-region center

A trust-region method based on error indicators

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

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad F(\boldsymbol{\mu})$$

that leverages trust-region subproblems of the form

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

where $\vartheta_k(\boldsymbol{\mu})$ is an *error bound*, 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

Trust-region model: two-level approximation based on reduced-order models and sparse grids

The trust-region subproblem

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

is defined as

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

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

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

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

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

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

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

Error indicator for 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

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

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

Trust-region method based on error indicators: algorithm

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

$$\vartheta_k(\boldsymbol{\mu}_k) \leq \kappa_1 \Delta_k \qquad \varphi_k \quad \leq \kappa_2 \, ||\nabla m_k(\boldsymbol{\mu}_k)||$$

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

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

3: **Step acceptance**: Compute

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

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

4: **Trust-region update**:

if $\rho_k \leq \eta_1$ **then** $\Delta_{k+1} \in (0, \gamma\vartheta_k(\hat{\boldsymbol{\mu}}_k)]$ **end if**

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

if $\rho_k \geq \eta_2$ **then** $\Delta_{k+1} \in [\Delta_k, \Delta_{\max}]$ **end if**

Trust-region method based on error indicators: convergence

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

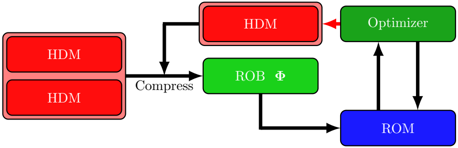
The trust-region method guarantees

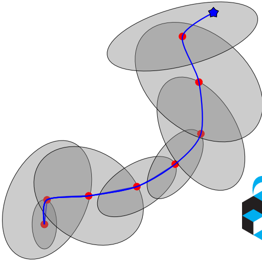
$$\boxed{\liminf_{k\to\infty} ||\nabla_{\boldsymbol{\mu}}\mathbb{E}\left[\mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu}_k,\,\cdot\,),\,\boldsymbol{\mu}_k,\,\cdot\,)\right]|| = 0}$$

if $\exists \zeta,\,\tau > 0$ such that

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

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





Control two-level inexactness: dimension-adaptive sparse grids and greedy sampling

while $\left|\mathbb{E}_{\mathcal{N}(\mathcal{I}_k)}\left[\mathcal{J}(\boldsymbol{\Phi}\boldsymbol{y}(\boldsymbol{\mu},\,\cdot\,),\,\boldsymbol{\mu},\,\cdot\,)\right]\right| > \frac{1}{2\alpha_2}\kappa_1\Delta_k$ **do**

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

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

while $\mathbb{E}_{\mathcal{I}_k\cup\mathcal{N}(\mathcal{I}_k)}\left[||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{y}(\boldsymbol{\mu}_k,\,\cdot\,),\,\boldsymbol{\mu}_k,\,\cdot\,)||\right] > \frac{1}{2\alpha_1}\kappa_1\Delta_k$ **do**

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

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

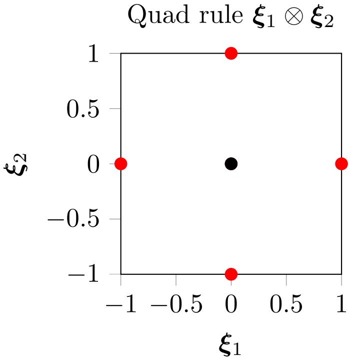
Sample high-dimensional model: Let $j^* = \arg\max r_j$ and compute

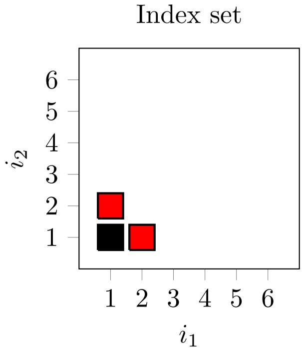
$$\boldsymbol{u}(\boldsymbol{\mu}_k,\boldsymbol{\xi}_{j^*}),\;\; \frac{\partial\boldsymbol{u}}{\partial\boldsymbol{\mu}}(\boldsymbol{\mu}_k,\boldsymbol{\xi}_{j^*})$$

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

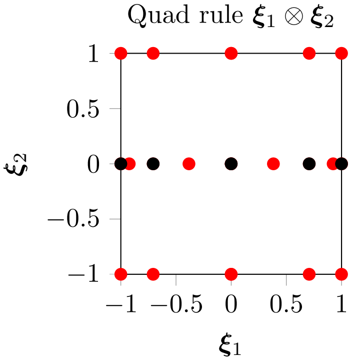
end while

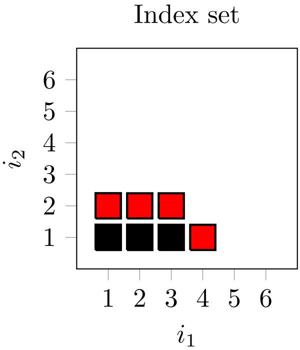
end while



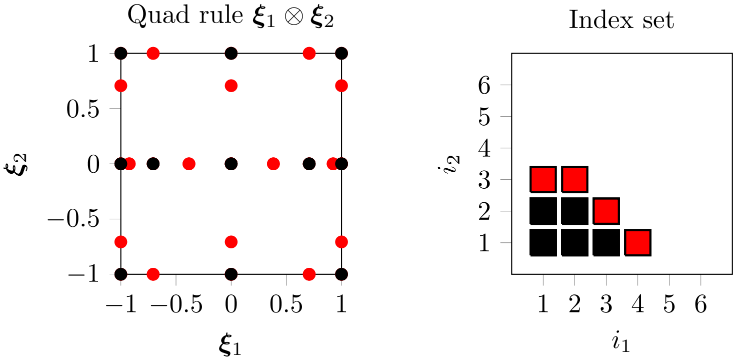


Index set $(\mathcal{I})$ – ● Neighbors $(\mathcal{N}(\mathcal{I}))$ – ●

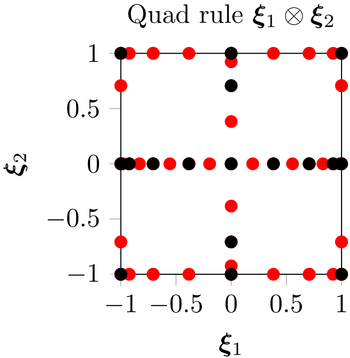


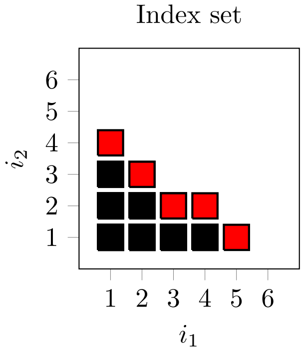


Index set $(\mathcal{I})$ – ●   Neighbors $(\mathcal{N}(\mathcal{I}))$ – ●

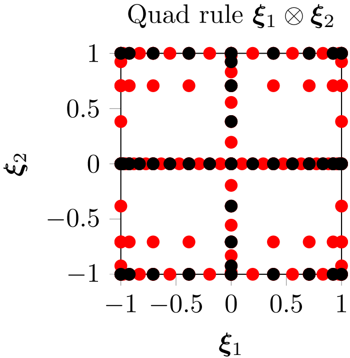


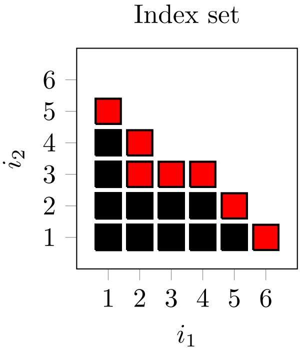
Index set ($\mathcal{I}$) – ●   Neighbors ($\mathcal{N}(\mathcal{I})$) – ●





Index set $(\mathcal{I})$ – ●   Neighbors $(\mathcal{N}(\mathcal{I}))$ – ●





Index set ($\mathcal{I}$) – ●

Neighbors ($\mathcal{N}(\mathcal{I})$) – ●

Optimal control of steady Burgers' equation

- **Optimization problem:**

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

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

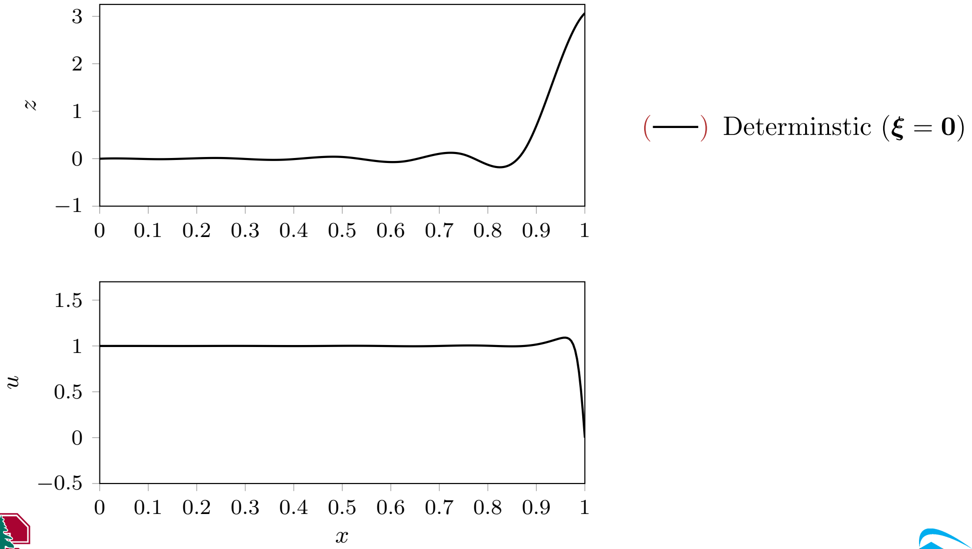
$$-\nu(\boldsymbol{\xi})\partial_{xx}u(\boldsymbol{\mu},\, \boldsymbol{\xi},\, x) + u(\boldsymbol{\mu},\, \boldsymbol{\xi},\, x)\partial_x u(\boldsymbol{\mu},\, \boldsymbol{\xi},\, x) = z(\boldsymbol{\mu},\, x) \quad x \in (0,\, 1), \quad \boldsymbol{\xi} \in \boldsymbol{\Xi}$$
$$u(\boldsymbol{\mu},\,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

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

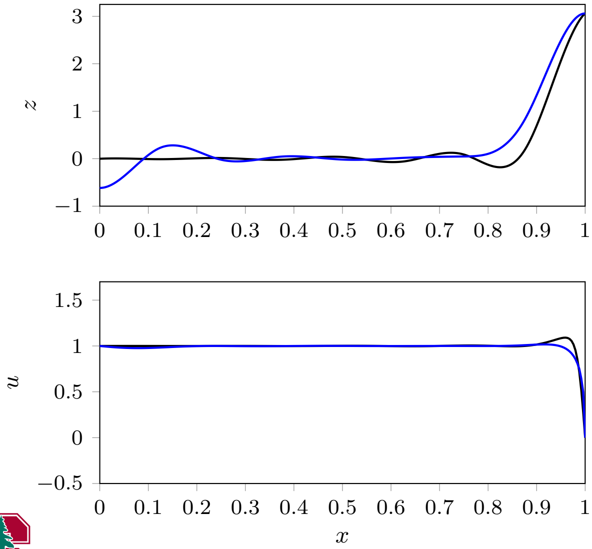
$$\nu(\boldsymbol{\xi}) = 10^{\xi_1 - 2} \qquad d_0(\boldsymbol{\xi}) = 1 + \frac{\xi_2}{1000} \qquad d_1(\boldsymbol{\xi}) = \frac{\xi_3}{1000}$$

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

Trust-region method recovers optimal control and statistics



Trust-region method recovers optimal control and statistics

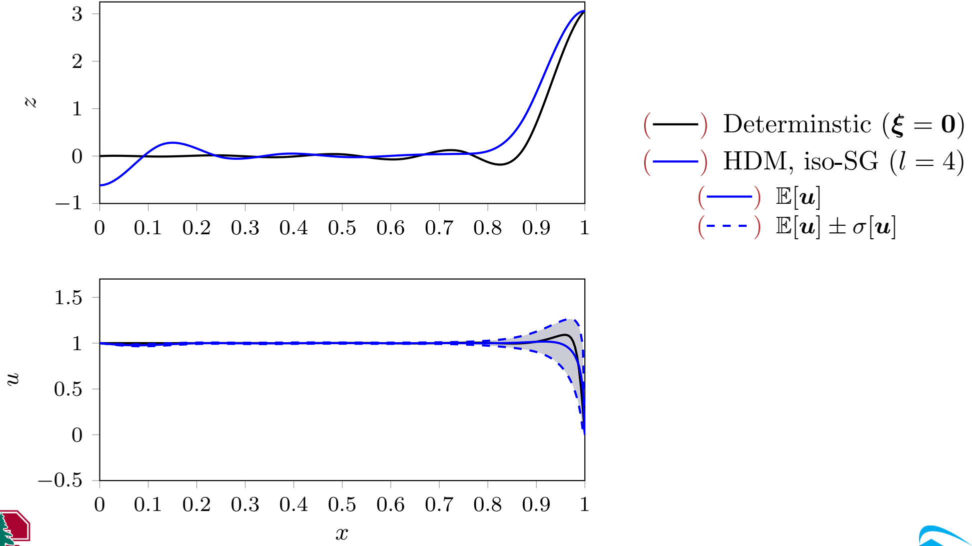


(——) Determinstic ($\boldsymbol{\xi} = \mathbf{0}$)

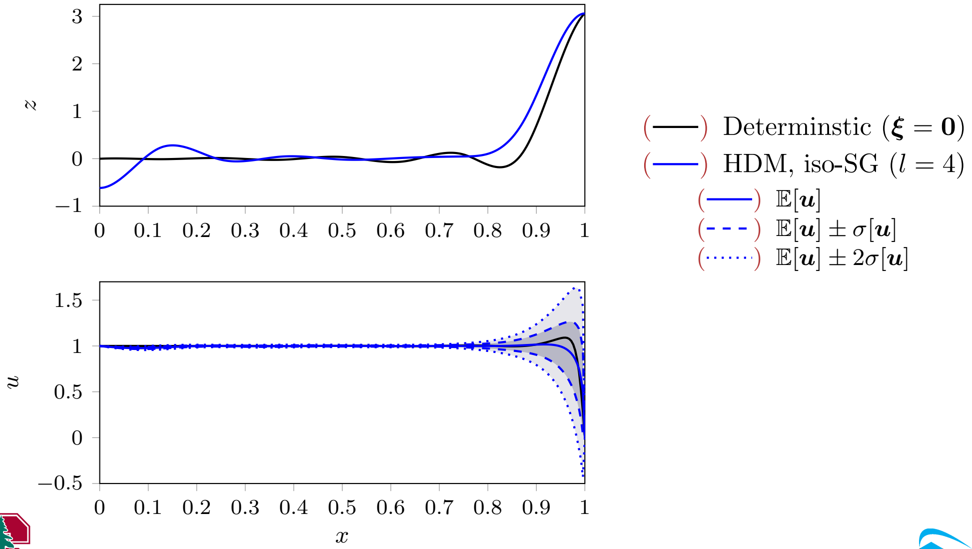
(——) HDM, iso-SG ($l = 4$)

(——) $\mathbb{E}[\boldsymbol{u}]$

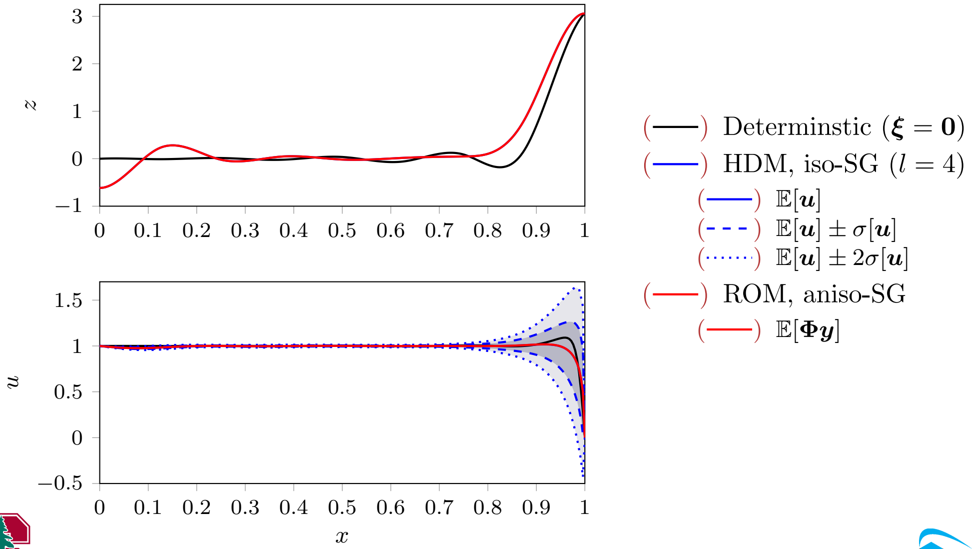
Trust-region method recovers optimal control and statistics



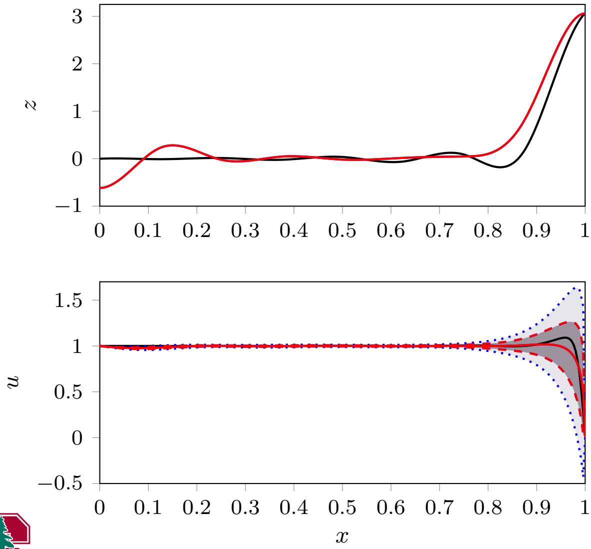
Trust-region method recovers optimal control and statistics



Trust-region method recovers optimal control and statistics



Trust-region method recovers optimal control and statistics



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

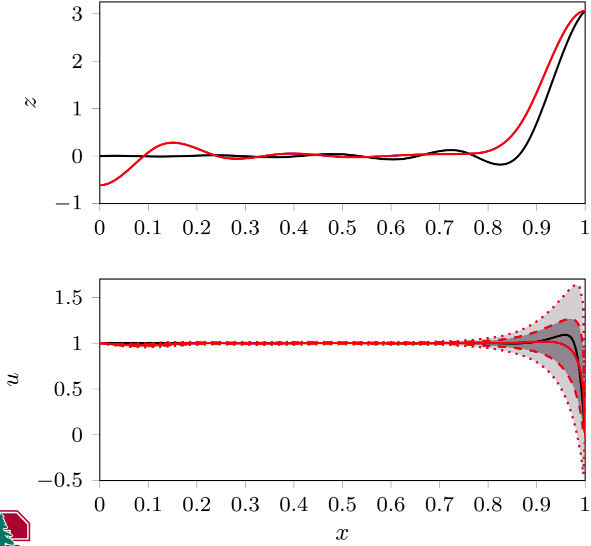
(——) HDM, iso-SG ($l = 4$)

- (——) $\mathbb{E}[\boldsymbol{u}]$
- (- - -) $\mathbb{E}[\boldsymbol{u}] \pm \sigma[\boldsymbol{u}]$
- (·······) $\mathbb{E}[\boldsymbol{u}] \pm 2\sigma[\boldsymbol{u}]$

(——) ROM, aniso-SG

- (——) $\mathbb{E}[\boldsymbol{\Phi y}]$
- (- - -) $\mathbb{E}[\boldsymbol{\Phi y}] \pm \sigma[\boldsymbol{\Phi y}]$

Trust-region method recovers optimal control and statistics



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

( ) HDM, iso-SG ($l = 4$)

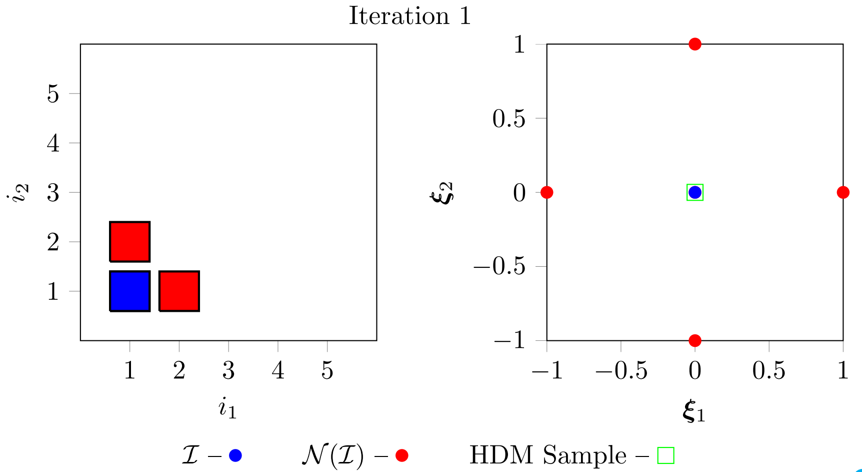
- ( ) $\mathbb{E}[\boldsymbol{u}]$
- ( ) $\mathbb{E}[\boldsymbol{u}] \pm \sigma[\boldsymbol{u}]$
- ( ) $\mathbb{E}[\boldsymbol{u}] \pm 2\sigma[\boldsymbol{u}]$

( ) ROM, aniso-SG

- ( ) $\mathbb{E}[\boldsymbol{\Phi y}]$
- ( ) $\mathbb{E}[\boldsymbol{\Phi y}] \pm \sigma[\boldsymbol{\Phi y}]$
- ( ) $\mathbb{E}[\boldsymbol{\Phi y}] \pm 2\sigma[\boldsymbol{\Phi y}]$

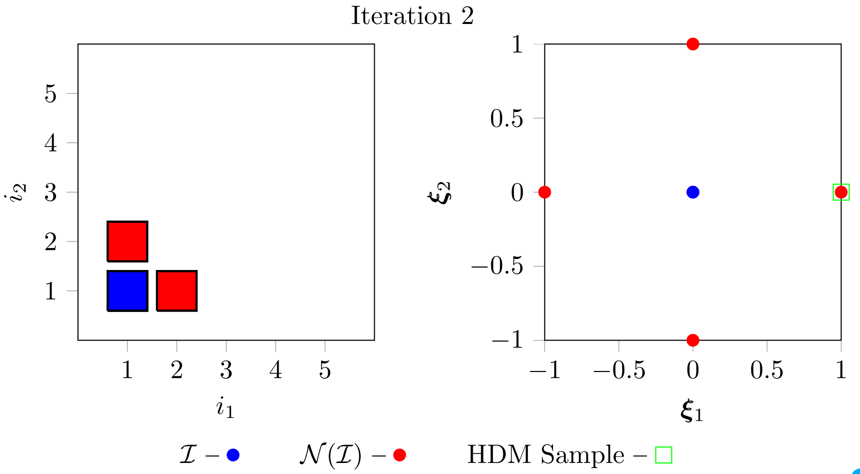
Trust-region method requires very few HDM queries to converge

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



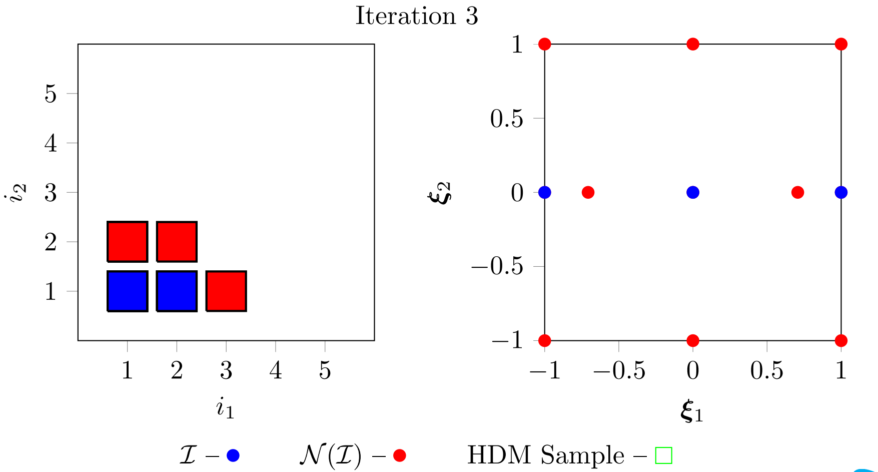
Trust-region method requires very few HDM queries to converge

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



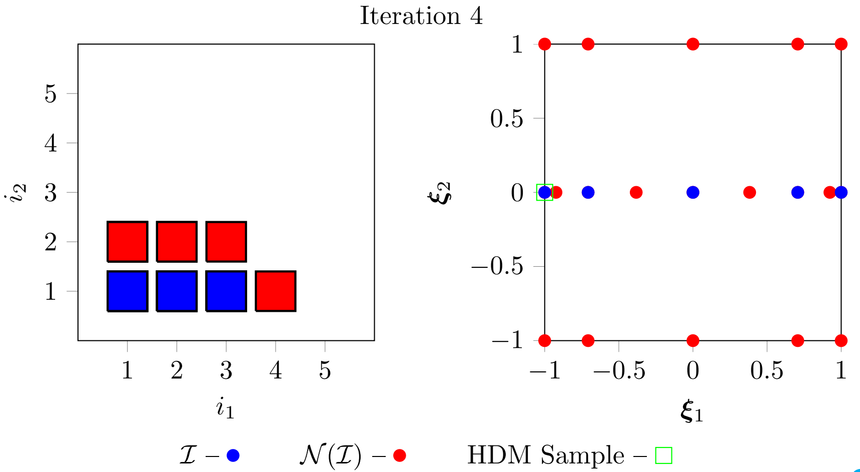
Trust-region method requires very few HDM queries to converge

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



Trust-region method requires very few HDM queries to converge

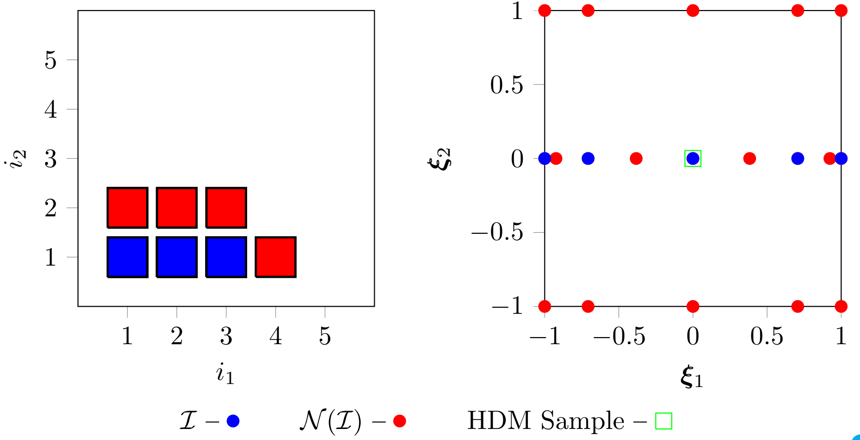
Prior to each trust-region subproblem, the model (sparse grid, $\mathcal{I}_k$, and basis, $\mathbf{\Phi}_k$) must be constructed such that error indicators are below a tolerance



Trust-region method requires very few HDM queries to converge

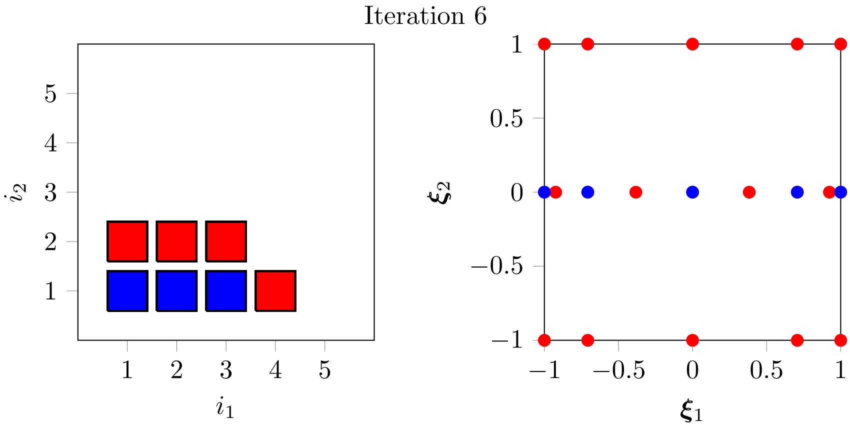
Prior to each trust-region subproblem, the model (sparse grid, $\mathcal{I}_k$, and basis, $\mathbf{\Phi}_k$) must be constructed such that error indicators are below a tolerance

Iteration 5



Trust-region method requires very few HDM queries to converge

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



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

Global convergence of trust-region method

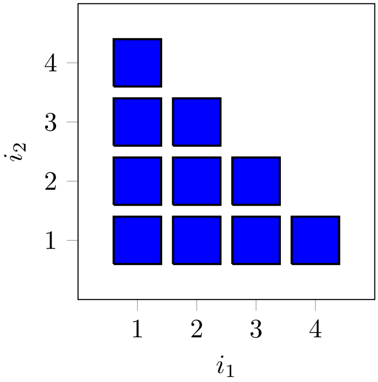
The trust-region method finds a sequence of parameters $\boldsymbol{\mu}_k$ *such that the gradient of HDM* $(||\nabla \mathcal{J}(\boldsymbol{\mu})||)$ *converges to* 0

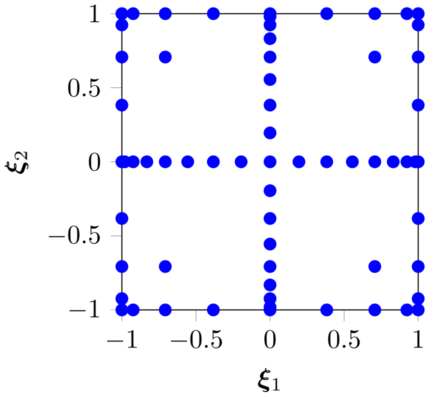
$m_k(\hat{\boldsymbol{\mu}}_k)$	$\|\|\nabla m_k(\hat{\boldsymbol{\mu}}_k)\|\|$	$\mathcal{J}(\hat{\boldsymbol{\mu}}_k)$	$\|\|\nabla \mathcal{J}(\hat{\boldsymbol{\mu}}_k)\|\|$	ρ_k	Δ_k	Success?
3.8783e-03	3.3779e-03	8.3351e-03	6.8542e-03	-	-	-
3.1121e-03	2.0393e-04	7.2687e-03	7.0676e-03	1.3918e+00	1.0000e+02	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

Few HDM queries are required for convergence at the cost of many ROM queries

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

1000× reduction in HDM queries vs. stochastic collocation on 4-level isotropic sparse grid





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

Summary

- *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
- 1000$\times$ decrease in number of HDM queries

Future work

- Incorporate nonlinear constraints
- Local reduced-order models for improved efficiency
- Less expensive error indicators for cheaper trust-region subproblems [Drohmann and Carlberg, 2014]

References I

De Santis, D., Zahr, M. J., and Farhat, C. (January 4 - 8 (2016)).
Gradient-based aerodynamic shape optimization using the FIVER embedded boundary method.
In *AIAA Science and Technology Forum and Exposition (SciTech 2016)*, San Diego, California.

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.

Zahr, M. J. and Farhat, C. (in preparation, 2016).
Roms for topology optimization.
International Journal for Numerical Methods in Engineering.

References II

Zahr, M. J. and Persson, P.-O. (In review, 2016).
An adjoint method for a high-order discretization of deforming domain conservation laws for optimization of flow problems.
Journal of Computational Physics.

Zahr, M. J., Persson, P.-O., and Wilkening, J. (In review, 2016).
A fully discrete adjoint method for optimization of flow problems on deforming domains with time-periodicity constraints.
Computers and Fluids.