Adaptive model reduction to accelerate optimization problems governed by partial differential equations

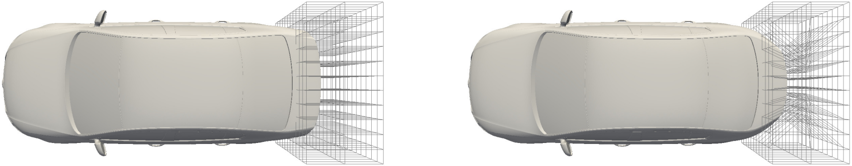
Matthew J. Zahr

Advisor: Charbel Farhat
Computational and Mathematical Engineering
Stanford University

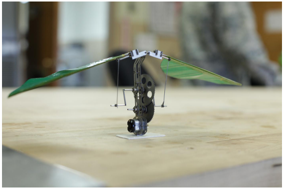
PhD Thesis Defense
Stanford University, Stanford, CA
August 03, 2016

PDE optimization is **ubiquitous** in science and engineering

Design: Find system that optimizes performance metric, satisfies constraints



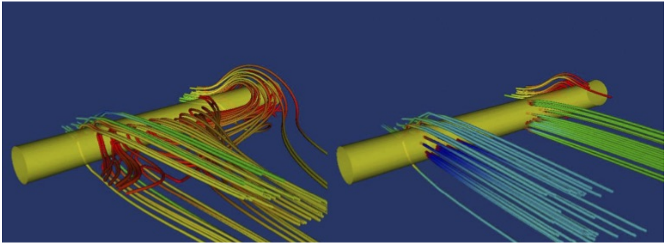
Aerodynamic shape design of automobile



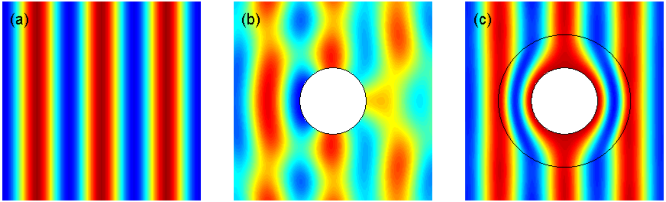
Optimal flapping motion of micro aerial vehicle

PDE optimization is **ubiquitous** in science and engineering

Control: Drive system to a desired state

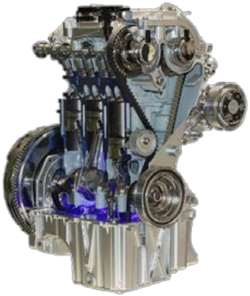


Boundary flow control

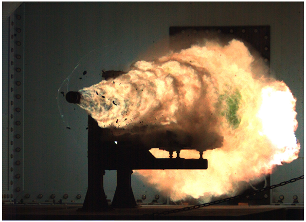


Metamaterial cloaking – electromagnetic invisibility

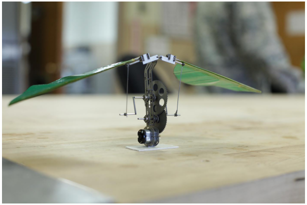
Current interest in ***computational physics*** *reaches far beyond analysis of a single configuration of a physical system into* ***design*** *(shape and topology) and* ***control*** *in an* ***uncertain*** *setting*



Engine System



EM Launcher



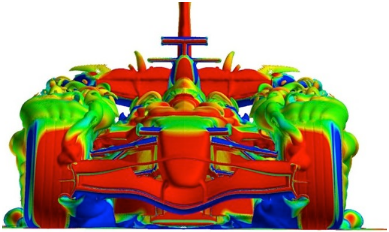
Micro-Aerial Vehicle

Repeated queries to **high-fidelity simulations** required by optimization and uncertainty quantification may be **prohibitively time-consuming**

Deterministic PDE-constrained optimization formulation

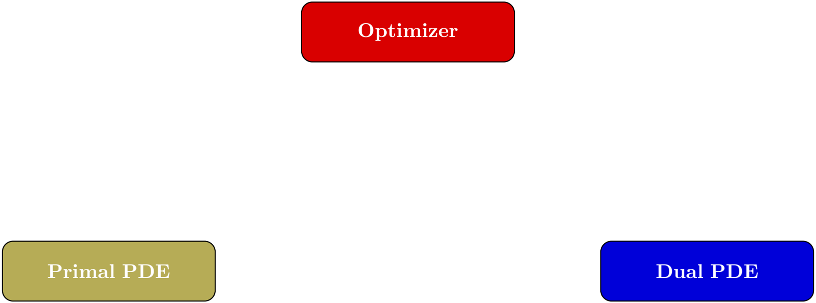
$$\begin{aligned} \underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathcal{J}(\boldsymbol{u},\, \boldsymbol{\mu}) \\ \text{subject to} \quad & \boldsymbol{r}(\boldsymbol{u};\, \boldsymbol{\mu}) = 0 \end{aligned}$$

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

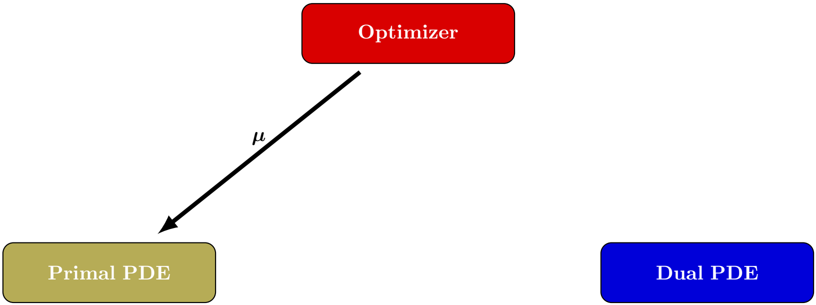


Nested approach to PDE-constrained optimization

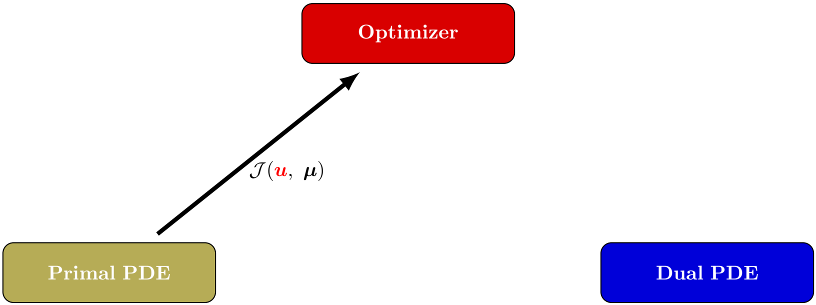
Virtually all expense emanates from primal/dual PDE solves



Virtually all expense emanates from primal/dual PDE solves

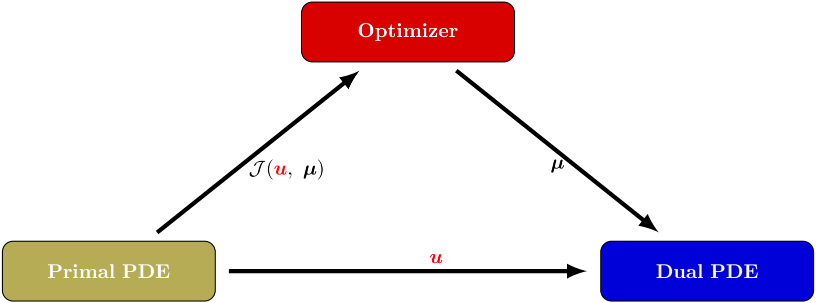


Virtually all expense emanates from primal/dual PDE solves



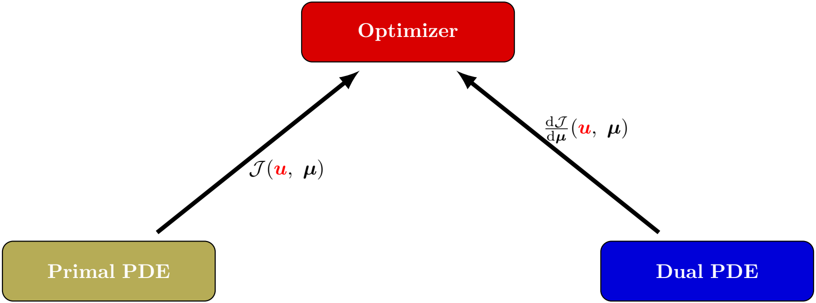
Nested approach to PDE-constrained optimization

Virtually all expense emanates from primal/dual PDE solves



Nested approach to PDE-constrained optimization

Virtually all expense emanates from primal/dual PDE solves



Applications in computational mechanics: static

Maximum lift-to-drag airfoil configuration

Stochastic PDE-constrained optimization formulation

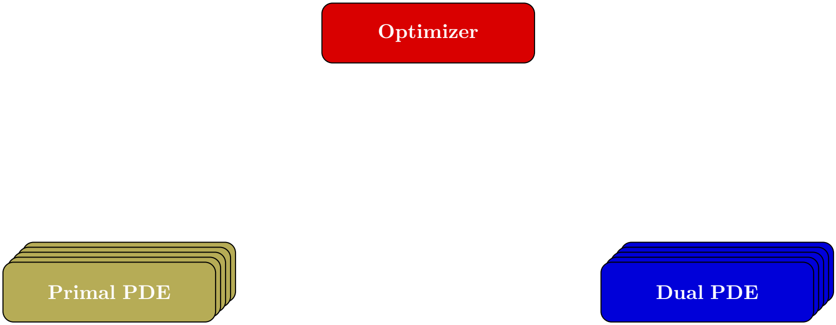
$$\begin{aligned} \underset{\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad & \mathbb{E}[\mathcal{J}(\boldsymbol{u}, \boldsymbol{\mu}, \cdot)] \\ \text{subject to} \quad & \boldsymbol{r}(\boldsymbol{u}; \boldsymbol{\mu}, \boldsymbol{\xi}) = 0 \quad \forall \boldsymbol{\xi}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

- $\boldsymbol{r} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \to \mathbb{R}^{n_{\boldsymbol{u}}}$ — 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
- $\mathbb{E}[\mathcal{F}] \equiv \int_{\boldsymbol{\Xi}} \mathcal{F}(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) \, d\boldsymbol{\xi}$

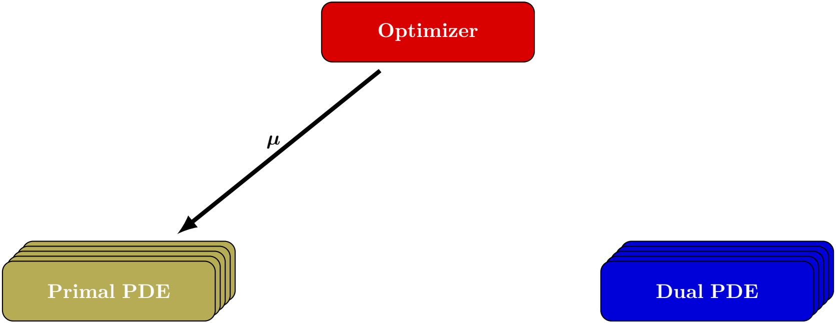
Each function evaluation requires integration over stochastic space – ***expensive***

Nested approach to stochastic PDE-constrained optimization

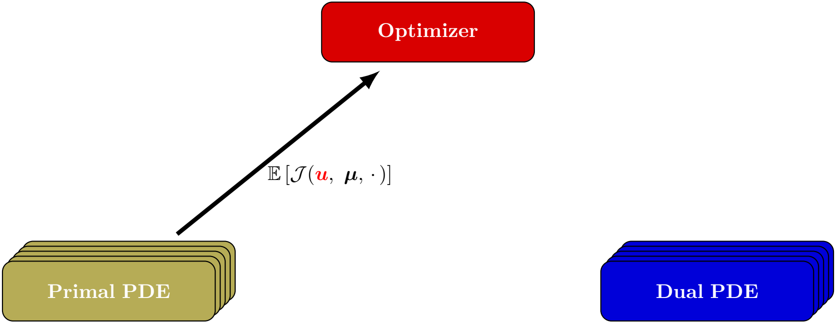
Ensemble *of primal/dual PDE solves increases cost by* ***orders of magnitude***



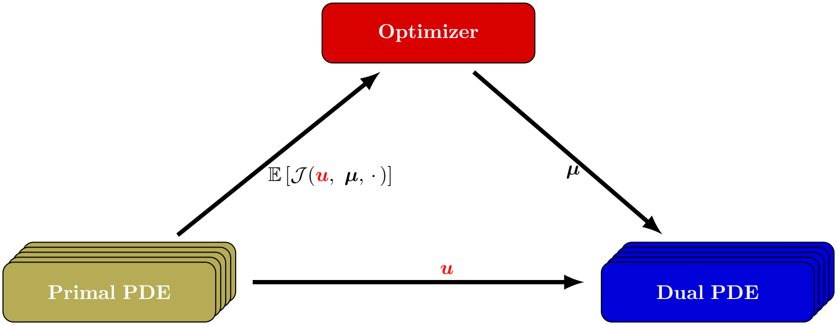
Ensemble *of primal/dual PDE solves increases cost by* ***orders of magnitude***



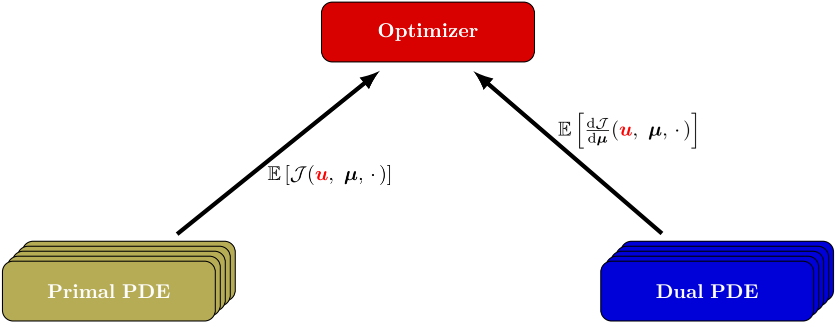
Ensemble *of primal/dual PDE solves increases cost by* ***orders of magnitude***



Ensemble *of primal/dual PDE solves increases cost by* ***orders of magnitude***



Ensemble** of primal/dual PDE solves increases cost by **orders of magnitude



Proposed approach: managed inexactness

Replace expensive PDE with inexpensive approximation model

- **Reduced-order models** used for *inexact PDE evaluations*
- **Partially converged solutions** used for *inexact PDE evaluations*
- **Anisotropic sparse grids** used for *inexact integration* of risk measures

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

[1] Must be *computable* and apply to general, nonlinear PDEs

Proposed approach: managed inexactness

Replace expensive PDE with inexpensive approximation model

- **Reduced-order models** used for *inexact PDE evaluations*
- **Partially converged solutions** used for *inexact PDE evaluations*
- **Anisotropic sparse grids** used for *inexact integration* of risk measures

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

Manage inexactness with trust region method

- Embedded in globally convergent **trust region** method
- **Error indicators**[1] to account for *all* sources of inexactness
- **Refinement** of approximation model using *greedy algorithms*

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad F(\boldsymbol{\mu}) \qquad \longrightarrow \qquad \begin{aligned} &\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} && m_k(\boldsymbol{\mu}) \\ &\text{subject to} && ||\boldsymbol{\mu}-\boldsymbol{\mu}_k|| \leq \Delta_k \end{aligned}$$

[1] Must be *computable* and apply to general, nonlinear PDEs

Proposed approach: managed inexactness

Replace expensive PDE with inexpensive approximation model

- **Reduced-order models** used for *inexact PDE evaluations*
- **Partially converged solutions** used for *inexact PDE evaluations*
- **Anisotropic sparse grids** used for *inexact integration* of risk measures

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

Manage inexactness with trust region method

- Embedded in globally convergent **trust region** method
- **Error indicators**[1] to account for *all* sources of inexactness
- **Refinement** of approximation model using *greedy algorithms*

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad F(\boldsymbol{\mu}) \qquad \longrightarrow \qquad \begin{aligned} &\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} && m_k(\boldsymbol{\mu}) \\ &\text{subject to} && ||\boldsymbol{\mu}-\boldsymbol{\mu}_k|| \leq \Delta_k \end{aligned}$$

[1] Must be *computable* and apply to general, nonlinear PDEs

- First-order consistency [Alexandrov et al., 1998]

$$m_k(\boldsymbol{\mu}_k) = F(\boldsymbol{\mu}_k) \qquad \nabla m_k(\boldsymbol{\mu}_k) = \nabla F(\boldsymbol{\mu}_k)$$

- The Carter condition [Carter, 1989, Carter, 1991]

$$||\nabla F(\boldsymbol{\mu}_k) - \nabla m_k(\boldsymbol{\mu}_k)|| \leq \eta \, ||\nabla m_k(\boldsymbol{\mu}_k)|| \qquad \eta \in (0,\, 1)$$

- Asymptotic gradient bound [Heinkenschloss and Vicente, 2002]

$$||\nabla F(\boldsymbol{\mu}_k) - \nabla m_k(\boldsymbol{\mu}_k)|| \leq \xi \min\{||\nabla m_k(\boldsymbol{\mu}_k)|| \,,\, \Delta_k\} \qquad \xi > 0$$

Asymptotic gradient bound permits the use of an ***error indicator****:* φ_k

$$||\nabla F(\boldsymbol{\mu}) - \nabla m_k(\boldsymbol{\mu})|| \leq \xi \varphi_k(\boldsymbol{\mu}) \qquad \xi > 0$$
$$\varphi_k(\boldsymbol{\mu}_k) \leq \kappa_\varphi \min\{||\nabla m_k(\boldsymbol{\mu}_k)|| \,,\, \Delta_k\}$$

Trust region method with inexact gradients [Kouri et al., 2013]

1: **Model update**: Choose model m_k and error indicator φ_k

$$\varphi_k(\boldsymbol{\mu}_k) \leq \kappa_\varphi \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\}$$

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

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

3: **Step acceptance**: Compute actual-to-predicted reduction

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

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** $\Delta_{k+1} \in (0, \gamma\,||\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k||]$ **end if**

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

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

Trust region method with inexact gradients [Kouri et al., 2013]

1: **Model update**: Choose model m_k and error indicator φ_k

$$\varphi_k(\boldsymbol{\mu}_k) \leq \kappa_\varphi \min\{||\nabla m_k(\boldsymbol{\mu}_k)|| \, , \, \Delta_k\}$$

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

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

3: **Step acceptance**: Compute actual-to-predicted reduction

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

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** $\Delta_{k+1} \in (0, \gamma\, ||\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k||]$ **end if**

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

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

Trust region method with inexact gradients and objective

1: **Model update**: Choose model m_k and error indicator φ_k

$$\vartheta_k(\boldsymbol{\mu}_k) \leq \kappa_\vartheta \Delta_k \qquad \varphi_k(\boldsymbol{\mu}_k) \leq \kappa_\varphi \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\}$$

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

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

3: **Step acceptance**: Compute approximation of actual-to-predicted reduction

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

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**

Asymptotic accuracy requirements on approximation model [Zahr, 2016]

$$\begin{aligned} |F(\boldsymbol{\mu}_k) - F(\boldsymbol{\mu}) + m_k(\boldsymbol{\mu}) - m_k(\boldsymbol{\mu}_k)| &\leq \zeta \vartheta_k(\boldsymbol{\mu}) & \zeta &> 0 \\ \vartheta_k(\boldsymbol{\mu}_k) &\leq \kappa_\vartheta \Delta_k & \kappa_\vartheta &\in (0,\, 1) \end{aligned}$$

Asymptotic accuracy requirements on inexact objective evaluations [Kouri et al., 2014]

$$\begin{aligned} |F(\boldsymbol{\mu}_k) - F(\boldsymbol{\mu}) + \psi_k(\boldsymbol{\mu}) - \psi_k(\boldsymbol{\mu}_k)| &\leq \sigma \theta_k(\boldsymbol{\mu}) \qquad \sigma > 0 \\ \theta_k(\hat{\boldsymbol{\mu}}_k)^\omega &\leq \eta \min\{m_k(\boldsymbol{\mu}_k) - m_k(\hat{\boldsymbol{\mu}}_k),\, r_k\} \end{aligned}$$

$$\omega,\, \eta \in (0,\, 1),\, r_k \to 0$$

Trust region ingredients for **global convergence**

Approximation models

$$m_k(\boldsymbol{\mu}),\ \psi_k(\boldsymbol{\mu})$$

Error indicators

$$\begin{aligned}
|F(\boldsymbol{\mu}_k) - F(\boldsymbol{\mu}) + m_k(\boldsymbol{\mu}) - m_k(\boldsymbol{\mu}_k)| &\leq \zeta \vartheta_k(\boldsymbol{\mu}) & \zeta > 0 \\
||\nabla F(\boldsymbol{\mu}) - \nabla m_k(\boldsymbol{\mu})|| &\leq \xi \varphi_k(\boldsymbol{\mu}) & \xi > 0 \\
|F(\boldsymbol{\mu}_k) - F(\boldsymbol{\mu}) + \psi_k(\boldsymbol{\mu}) - \psi_k(\boldsymbol{\mu}_k)| &\leq \sigma \theta_k(\boldsymbol{\mu}) & \sigma > 0
\end{aligned}$$

Adaptivity

$$\begin{aligned}
\vartheta_k(\boldsymbol{\mu}_k) &\leq \kappa_\vartheta \Delta_k \\
\varphi_k(\boldsymbol{\mu}_k) &\leq \kappa_\varphi \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\ \Delta_k\} \\
\theta_k(\hat{\boldsymbol{\mu}}_k)^\omega &\leq \eta \min\{m_k(\boldsymbol{\mu}_k) - m_k(\hat{\boldsymbol{\mu}}_k),\ r_k\}
\end{aligned}$$

Global convergence

$$\liminf_{k\to\infty} \ ||\nabla F(\boldsymbol{\mu}_k)|| = 0$$

Proposed approach: managed inexactness

Replace expensive PDE with inexpensive approximation model

- **Reduced-order models** used for *inexact PDE evaluations*
- **Partially converged solutions** used for *inexact PDE evaluations*
- **Anisotropic sparse grids** used for *inexact integration* of risk measures

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

Manage inexactness with trust region method

- Embedded in globally convergent **trust region** method
- **Error indicators** to account for *all* sources of inexactness
- **Refinement** of approximation model using *greedy algorithms*

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad F(\boldsymbol{\mu}) \qquad \longrightarrow \qquad \begin{aligned} &\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} && m_k(\boldsymbol{\mu}) \\ &\text{subject to} && ||\boldsymbol{\mu} - \boldsymbol{\mu}_k|| \leq \Delta_k \end{aligned}$$

Source of inexactness: **projection-based model reduction**

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

$$\boldsymbol{u} \approx \boldsymbol{\Phi} \boldsymbol{u}_r$$

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

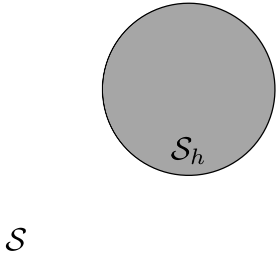
- Substitute into $\boldsymbol{r}(\boldsymbol{u},\, \boldsymbol{\mu}) = 0$ and project onto columnspace of a test basis $\boldsymbol{\Psi} \in \mathbb{R}^{n_{\boldsymbol{u}} \times k_{\boldsymbol{u}}}$ to obtain a square system

$$\boldsymbol{\Psi}^T \boldsymbol{r}(\boldsymbol{\Phi} \boldsymbol{u}_r,\, \boldsymbol{\mu}) = 0$$

Connection to finite element method: hierarchical subspaces

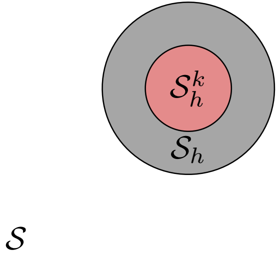
$$\mathcal{S}$$

- $\mathcal{S}$ - infinite-dimensional trial space



- $\mathcal{S}$ - infinite-dimensional trial space
- $\mathcal{S}_h$ - (large) finite-dimensional trial space

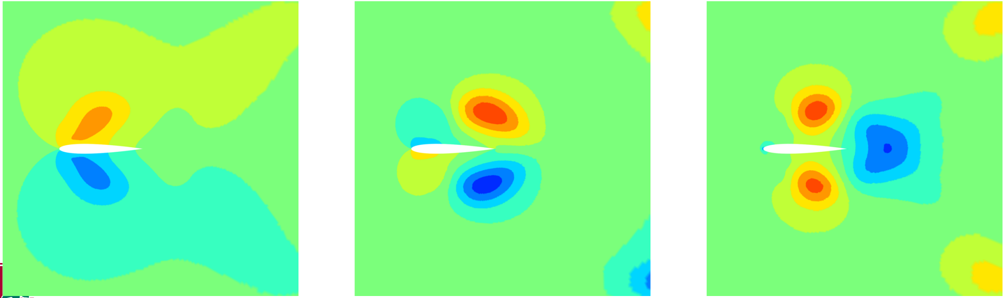
Connection to finite element method: hierarchical subspaces



- $\mathcal{S}$ - infinite-dimensional trial space
- $\mathcal{S}_h$ - (large) finite-dimensional trial space
- $\mathcal{S}_h^k$ - (small) finite-dimensional trial space
- $\mathcal{S}_h^k \subset \mathcal{S}_h \subset \mathcal{S}$

Few global, data-driven basis functions v. many local ones

- Instead of using traditional *local* shape functions, use **global shape functions**
- Instead of a-priori, analytical shape functions, leverage data-rich computing environment by using **data-driven modes**



Definition of $\boldsymbol{\Psi}$: minimum-residual reduced-order models

A ROM possesses the **minimum-residual property** if $\boldsymbol{\Psi}^T \boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\ \boldsymbol{\mu}) = 0$ is equivalent to the optimality condition of

$$\underset{\boldsymbol{u}_r \in \mathbb{R}^{k_{\boldsymbol{u}}}}{\text{minimize}} \quad ||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\ \boldsymbol{\mu})||_{\boldsymbol{\Theta}} \qquad \boldsymbol{\Theta} \succ 0$$

which requires

$$\boldsymbol{\Psi}(\boldsymbol{u},\ \boldsymbol{\mu}) = \boldsymbol{\Theta} \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}}(\boldsymbol{u},\ \boldsymbol{\mu}) \boldsymbol{\Phi}$$

Definition of $\boldsymbol{\Psi}$: minimum-residual reduced-order models

A ROM possesses the **minimum-residual property** if $\boldsymbol{\Psi}^T\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\ \boldsymbol{\mu}) = 0$ is equivalent to the optimality condition of

$$\underset{\boldsymbol{u}_r \in \mathbb{R}^{k_{\boldsymbol{u}}}}{\text{minimize}} \quad ||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} \qquad \boldsymbol{\Theta} \succ 0$$

which requires

$$\boldsymbol{\Psi}(\boldsymbol{u},\, \boldsymbol{\mu}) = \boldsymbol{\Theta}\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}}(\boldsymbol{u},\, \boldsymbol{\mu})\boldsymbol{\Phi}$$

Implications of the minimum-residual property

- ("Optimality") For any $\boldsymbol{u} \in \text{col}(\boldsymbol{\Phi})$,

$$||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} \leq ||\boldsymbol{r}(\boldsymbol{u},\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}}$$

- (Monotonicity) For any $\text{col}(\boldsymbol{\Phi}') \subseteq \text{col}(\boldsymbol{\Phi})$,

$$||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} \leq ||\boldsymbol{r}(\boldsymbol{\Phi}'\boldsymbol{u}_r',\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}}$$

- (Interpolation) If $\boldsymbol{u}(\boldsymbol{\mu}) \in \text{col}(\boldsymbol{\Phi})$, then

$$\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\, \boldsymbol{\mu}) = 0 \qquad \text{and} \qquad \boldsymbol{u}(\boldsymbol{\mu}) = \boldsymbol{\Phi}\boldsymbol{u}_r$$

Definition of $\frac{\partial \boldsymbol{u}_r}{\partial \boldsymbol{\mu}}$: minimum-residual reduced sensitivities

Traditional sensitivity analysis ($\boldsymbol{\Theta} = \boldsymbol{I}$)

$$\frac{\partial \boldsymbol{u}_r}{\partial \boldsymbol{\mu}} = -\left[\sum_{j=1}^{n_{\boldsymbol{u}}} \boldsymbol{r}_j \boldsymbol{\Phi}^T \frac{\partial^2 \boldsymbol{r}_j}{\partial \boldsymbol{u} \partial \boldsymbol{u}} \boldsymbol{\Phi} + \left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}} \boldsymbol{\Phi}\right)^T \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}} \boldsymbol{\Phi}\right]^{-1} \left(\sum_{j=1}^{n_{\boldsymbol{u}}} \boldsymbol{r}_j \boldsymbol{\Phi}^T \frac{\partial^2 \boldsymbol{r}_j}{\partial \boldsymbol{u} \partial \boldsymbol{\mu}} + \left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}} \boldsymbol{\Phi}\right)^T \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\mu}}\right)$$

+ Guaranteed to produce *exact* derivatives of ROM quantities of interest
– Requires 2nd derivatives of $\boldsymbol{r}$
– $\boldsymbol{\Phi} \frac{\partial \boldsymbol{u}_r}{\partial \boldsymbol{\mu}}$ not guaranteed to be good approximate of $\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\mu}}$

Definition of $\frac{\partial \boldsymbol{u}_r}{\partial \boldsymbol{\mu}}$: minimum-residual reduced sensitivities

Minimum-residual sensitivity analysis

$$\widehat{\frac{\partial \boldsymbol{u}_r}{\partial \boldsymbol{\mu}}} = \arg\min_{\boldsymbol{a}} \left\|\boldsymbol{\Phi}^{\partial}\boldsymbol{a} - \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\mu}}\right\|_{\boldsymbol{\Theta}^{\partial}} = -\left[\left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}}\boldsymbol{\Phi}^{\partial}\right)^T \boldsymbol{\Theta}^{\partial} \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}}\boldsymbol{\Phi}^{\partial}\right]^{-1} \left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}}\boldsymbol{\Phi}^{\partial}\right)^T \boldsymbol{\Theta}^{\partial} \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\mu}}$$

- \+ **Minimum-residual property** – optimality, monotonicity, interpolation
- \+ Does not require 2nd derivatives of $\boldsymbol{r}$
- – $\widehat{\frac{\partial \boldsymbol{u}_r}{\partial \boldsymbol{\mu}}} \neq \frac{\partial \boldsymbol{u}_r}{\partial \boldsymbol{\mu}}$, i.e., it is not the exact sensitivity[2]

[2] These quantities agree if $\boldsymbol{\Phi}^{\partial} = \boldsymbol{\Phi}$ and either $\boldsymbol{\Psi}$ is constant or the primal ROM is exact [Zahr, 2016]

Hyperreduction to reduce complexity of nonlinear terms

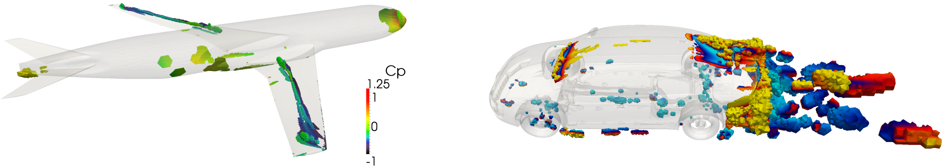
Despite **reduced dimensionality**, $\mathcal{O}(n_{\boldsymbol{u}})$ operations are required to evaluate

$$\boldsymbol{\Psi}^T \boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\, \boldsymbol{\mu}) \qquad \boldsymbol{\Psi}^T \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{u}}(\boldsymbol{\Phi}\boldsymbol{u}_r,\, \boldsymbol{\mu})\boldsymbol{\Phi}$$

Solution: only perform minimization over a *subset* of the spatial domain

$$\underset{\boldsymbol{u}_r \in \mathbb{R}^{k_{\boldsymbol{u}}}}{\text{minimize}} \quad ||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} \implies \underset{\boldsymbol{u}_r \in \mathbb{R}^{k_{\boldsymbol{u}}}}{\text{minimize}} \quad \left|\left|\boldsymbol{P}^T \boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r,\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}}$$

and **hyperreduced** model[3] is independent of $n_{\boldsymbol{u}}$



Sample mesh for CRM (left) and Passat (right) [Washabaugh, 2016]

[3] Masked minimum-residual property and weaker definitions of optimality, monotonicity, and interpolation hold

Trust region ingredients for global convergence

Approximation models

$$m_k(\boldsymbol{\mu}),\ \psi_k(\boldsymbol{\mu})$$

Error indicators

$$\begin{aligned}
|F(\boldsymbol{\mu}_k) - F(\boldsymbol{\mu}) + m_k(\boldsymbol{\mu}) - m_k(\boldsymbol{\mu}_k)| &\leq \zeta \vartheta_k(\boldsymbol{\mu}) & \zeta > 0 \\
||\nabla F(\boldsymbol{\mu}) - \nabla m_k(\boldsymbol{\mu})|| &\leq \xi \varphi_k(\boldsymbol{\mu}) & \xi > 0 \\
|F(\boldsymbol{\mu}_k) - F(\boldsymbol{\mu}) + \psi_k(\boldsymbol{\mu}) - \psi_k(\boldsymbol{\mu}_k)| &\leq \sigma \theta_k(\boldsymbol{\mu}) & \sigma > 0
\end{aligned}$$

Adaptivity

$$\begin{aligned}
\vartheta_k(\boldsymbol{\mu}_k) &\leq \kappa_\vartheta \Delta_k \\
\varphi_k(\boldsymbol{\mu}_k) &\leq \kappa_\varphi \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\ \Delta_k\} \\
\theta_k(\hat{\boldsymbol{\mu}}_k)^\omega &\leq \eta \min\{m_k(\boldsymbol{\mu}_k) - m_k(\hat{\boldsymbol{\mu}}_k),\ r_k\}
\end{aligned}$$

Global convergence

$$\liminf_{k \to \infty}\ ||\nabla F(\boldsymbol{\mu}_k)|| = 0$$

Trust region method: ROM approximation model

Approximation models based on reduced-order models

$$m_k(\boldsymbol{\mu}) = \mathcal{J}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu}) \qquad \psi_k(\boldsymbol{\mu}) = \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu}),\, \boldsymbol{\mu})$$

Error indicators from residual-based error bounds

$$\begin{aligned}
\vartheta_k(\boldsymbol{\mu}) &= ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}_k),\, \boldsymbol{\mu}_k)||_{\boldsymbol{\Theta}} + ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} \\
\varphi_k(\boldsymbol{\mu}) &= ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} + \left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\Psi}_k \boldsymbol{\lambda}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}^{\boldsymbol{\lambda}}} \\
\theta_k(\boldsymbol{\mu}) &= 0
\end{aligned}$$

Adaptivity to refine basis at trust region center

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \boldsymbol{u}(\boldsymbol{\mu}_k) & \boldsymbol{\lambda}(\boldsymbol{\mu}_k) & \texttt{POD}(\boldsymbol{U}_k) & \texttt{POD}(\boldsymbol{V}_k) \end{bmatrix}$$

$$\boldsymbol{U}_k = \begin{bmatrix} \boldsymbol{u}(\boldsymbol{\mu}_0) & \cdots & \boldsymbol{u}(\boldsymbol{\mu}_{k-1}) \end{bmatrix} \qquad \boldsymbol{V}_k = \begin{bmatrix} \boldsymbol{\lambda}(\boldsymbol{\mu}_0) & \cdots & \boldsymbol{\lambda}(\boldsymbol{\mu}_{k-1}) \end{bmatrix}$$

$$\text{Interpolation property} \implies \vartheta_k(\boldsymbol{\mu}_k) = \varphi_k(\boldsymbol{\mu}_k) = 0$$

Approximation models based on reduced-order models

$$m_k(\boldsymbol{\mu}) = \mathcal{J}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu}) \qquad \psi_k(\boldsymbol{\mu}) = \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu}),\, \boldsymbol{\mu})$$

Error indicators from residual-based error bounds

$$\begin{aligned}
\vartheta_k(\boldsymbol{\mu}) &= ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}_k),\, \boldsymbol{\mu}_k)||_{\boldsymbol{\Theta}} + ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} \\
\varphi_k(\boldsymbol{\mu}) &= ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} + \left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\Psi}_k \boldsymbol{\lambda}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}^{\boldsymbol{\lambda}}} \\
\theta_k(\boldsymbol{\mu}) &= 0
\end{aligned}$$

Adaptivity to refine basis at trust region center

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \boldsymbol{u}(\boldsymbol{\mu}_k) & \boldsymbol{\lambda}(\boldsymbol{\mu}_k) & \mathtt{POD}(\boldsymbol{U}_k) & \mathtt{POD}(\boldsymbol{V}_k) \end{bmatrix}$$

$$\boldsymbol{U}_k = \begin{bmatrix} \boldsymbol{u}(\boldsymbol{\mu}_0) & \cdots & \boldsymbol{u}(\boldsymbol{\mu}_{k-1}) \end{bmatrix} \qquad \boldsymbol{V}_k = \begin{bmatrix} \boldsymbol{\lambda}(\boldsymbol{\mu}_0) & \cdots & \boldsymbol{\lambda}(\boldsymbol{\mu}_{k-1}) \end{bmatrix}$$

$$\text{Interpolation property} \implies \vartheta_k(\boldsymbol{\mu}_k) = \varphi_k(\boldsymbol{\mu}_k) = 0$$

$$\liminf_{k\to\infty} \; ||\nabla \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu}_k),\, \boldsymbol{\mu}_k)|| = 0$$

Proposed approach: managed inexactness

Replace expensive PDE with inexpensive approximation model

- **Reduced-order models** used for *inexact PDE evaluations*
- **Partially converged solutions** used for *inexact PDE evaluations*
- **Anisotropic sparse grids** used for *inexact integration* of risk measures

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

Manage inexactness with trust region method

- Embedded in globally convergent **trust region** method
- **Error indicators** to account for *all* sources of inexactness
- **Refinement** of approximation model using *greedy algorithms*

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad F(\boldsymbol{\mu}) \qquad \longrightarrow \qquad \begin{aligned} &\underset{\boldsymbol{\mu}\in\mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} && m_k(\boldsymbol{\mu}) \\ &\text{subject to} && ||\boldsymbol{\mu}-\boldsymbol{\mu}_k|| \leq \Delta_k \end{aligned}$$

Source of inexactness: **Partially Converged Solution (PCS)**

A τ-partially converged primal solution $\boldsymbol{u}^{\tau}(\boldsymbol{\mu})$ is any $\boldsymbol{u}$ satisfying

$$||\boldsymbol{r}(\boldsymbol{u},\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} \leq \tau$$

A τ_1-τ_2-partially converged adjoint solution $\boldsymbol{\lambda}^{\tau_1,\,\tau_2}(\boldsymbol{\mu})$ is any $\boldsymbol{\lambda}$ satisfying

$$\left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{u}^{\tau_1}(\boldsymbol{\mu}),\, \boldsymbol{\lambda},\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}^{\boldsymbol{\lambda}}} \leq \tau_2$$

Trust region method: ROM/PCS approximation model

Approximation models based on ROMs and partially converged solutions

$$m_k(\boldsymbol{\mu}) = \mathcal{J}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu}) \qquad \psi_k(\boldsymbol{\mu}) = \mathcal{J}(\boldsymbol{u}^{\tau_k}(\boldsymbol{\mu}),\, \boldsymbol{\mu})$$

Error indicators from residual-based error bounds

$$\vartheta_k(\boldsymbol{\mu}) = \left|\left|\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}_k),\, \boldsymbol{\mu}_k)\right|\right|_{\boldsymbol{\Theta}} + \left|\left|\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}}$$
$$\varphi_k(\boldsymbol{\mu}) = \left|\left|\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}} + \left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\Psi}_k \boldsymbol{\lambda}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}^{\boldsymbol{\lambda}}}$$
$$\theta_k(\boldsymbol{\mu}) = \left|\left|\boldsymbol{r}(\boldsymbol{u}^{\tau_k}(\boldsymbol{\mu}_k),\, \boldsymbol{\mu}_k)\right|\right|_{\boldsymbol{\Theta}} + \left|\left|\boldsymbol{r}(\boldsymbol{u}^{\tau_k}(\boldsymbol{\mu}),\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}}$$

Adaptivity to refine basis at trust region center

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \boldsymbol{u}^{\alpha_k}(\boldsymbol{\mu}_k) & \boldsymbol{\lambda}^{\alpha_k,\,\beta_k}(\boldsymbol{\mu}_k) & \texttt{POD}(\boldsymbol{U}_k) & \texttt{POD}(\boldsymbol{V}_k) \end{bmatrix}$$
$$\boldsymbol{U}_k = \begin{bmatrix} \boldsymbol{u}^{\alpha_0}(\boldsymbol{\mu}_0) & \cdots & \boldsymbol{u}^{\alpha_{k-1}}(\boldsymbol{\mu}_{k-1}) \end{bmatrix} \qquad \boldsymbol{V}_k = \begin{bmatrix} \boldsymbol{\lambda}^{\alpha_0,\,\beta_0}(\boldsymbol{\mu}_0) & \cdots & \boldsymbol{\lambda}^{\alpha_{k-1},\,\beta_{k-1}}(\boldsymbol{\mu}_{k-1}) \end{bmatrix}$$

and $\alpha_k,\, \beta_k,\, \tau_k$ selected such that

$$\vartheta_k(\boldsymbol{\mu}_k) \leq \kappa_\vartheta \Delta_k \qquad \varphi_k(\boldsymbol{\mu}_k) \leq \kappa_\varphi \min\{\left|\left|\nabla m_k(\boldsymbol{\mu}_k)\right|\right|,\, \Delta_k\}$$
$$\theta_k^\omega(\hat{\boldsymbol{\mu}}_k) \leq \eta \min\{m_k(\boldsymbol{\mu}_k) - m_k(\hat{\boldsymbol{\mu}}_k),\, r_k\}$$

Trust region method: ROM/PCS approximation model

Approximation models based on ROMs and partially converged solutions

$$m_k(\boldsymbol{\mu}) = \mathcal{J}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu}) \qquad \psi_k(\boldsymbol{\mu}) = \mathcal{J}(\boldsymbol{u}^{\tau_k}(\boldsymbol{\mu}),\, \boldsymbol{\mu})$$

Error indicators from residual-based error bounds

$$\vartheta_k(\boldsymbol{\mu}) = ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}_k),\, \boldsymbol{\mu}_k)||_{\boldsymbol{\Theta}} + ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}}$$
$$\varphi_k(\boldsymbol{\mu}) = ||\boldsymbol{r}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}} + \left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu}),\, \boldsymbol{\Psi}_k \boldsymbol{\lambda}_r(\boldsymbol{\mu}),\, \boldsymbol{\mu})\right|\right|_{\boldsymbol{\Theta}^{\boldsymbol{\lambda}}}$$
$$\theta_k(\boldsymbol{\mu}) = ||\boldsymbol{r}(\boldsymbol{u}^{\tau_k}(\boldsymbol{\mu}_k),\, \boldsymbol{\mu}_k)||_{\boldsymbol{\Theta}} + ||\boldsymbol{r}(\boldsymbol{u}^{\tau_k}(\boldsymbol{\mu}),\, \boldsymbol{\mu})||_{\boldsymbol{\Theta}}$$

Adaptivity to refine basis at trust region center

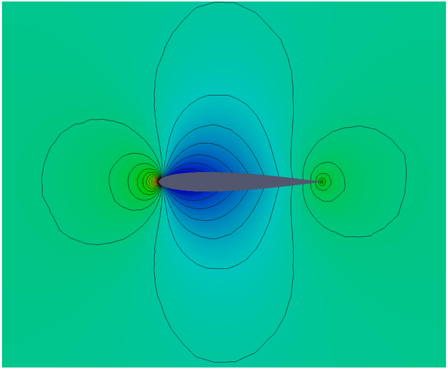
$$\boldsymbol{\Phi}_k = \begin{bmatrix} \boldsymbol{u}^{\alpha_k}(\boldsymbol{\mu}_k) & \boldsymbol{\lambda}^{\alpha_k,\,\beta_k}(\boldsymbol{\mu}_k) & \texttt{POD}(\boldsymbol{U}_k) & \texttt{POD}(\boldsymbol{V}_k) \end{bmatrix}$$
$$\boldsymbol{U}_k = \begin{bmatrix} \boldsymbol{u}^{\alpha_0}(\boldsymbol{\mu}_0) & \cdots & \boldsymbol{u}^{\alpha_{k-1}}(\boldsymbol{\mu}_{k-1}) \end{bmatrix} \qquad \boldsymbol{V}_k = \begin{bmatrix} \boldsymbol{\lambda}^{\alpha_0,\,\beta_0}(\boldsymbol{\mu}_0) & \cdots & \boldsymbol{\lambda}^{\alpha_{k-1},\,\beta_{k-1}}(\boldsymbol{\mu}_{k-1}) \end{bmatrix}$$

and $\alpha_k,\ \beta_k,\ \tau_k$ selected such that

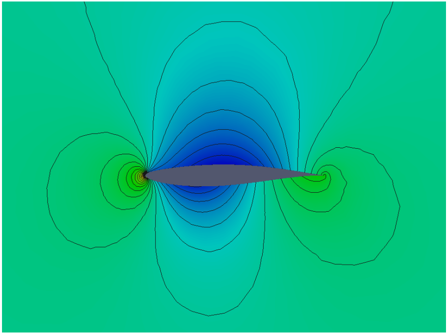
$$\vartheta_k(\boldsymbol{\mu}_k) \leq \kappa_\vartheta \Delta_k \qquad \varphi_k(\boldsymbol{\mu}_k) \leq \kappa_\varphi \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\}$$
$$\theta_k^\omega(\hat{\boldsymbol{\mu}}_k) \leq \eta \min\{m_k(\boldsymbol{\mu}_k) - m_k(\hat{\boldsymbol{\mu}}_k),\, r_k\}$$

$$\liminf_{k\to\infty} \; ||\nabla \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu}_k),\, \boldsymbol{\mu}_k)|| = 0$$

Pressure discrepancy minimization (Euler equations)



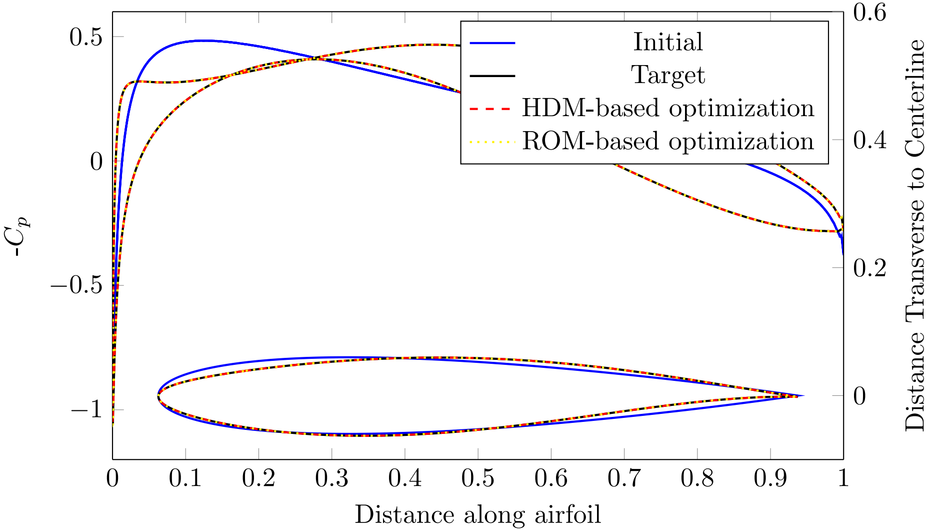
NACA0012: Initial



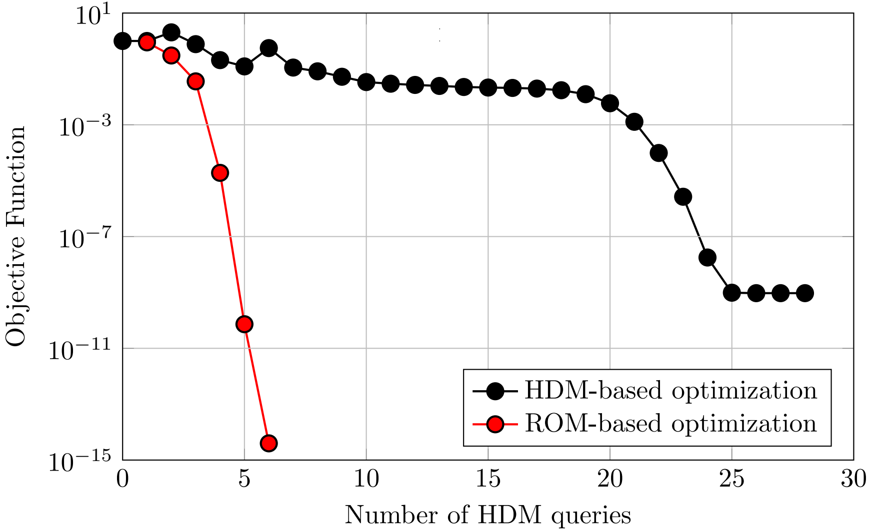
RAE2822: Target

Pressure field for airfoil configurations at $M_\infty = 0.5$, $\alpha = 0.0^\circ$

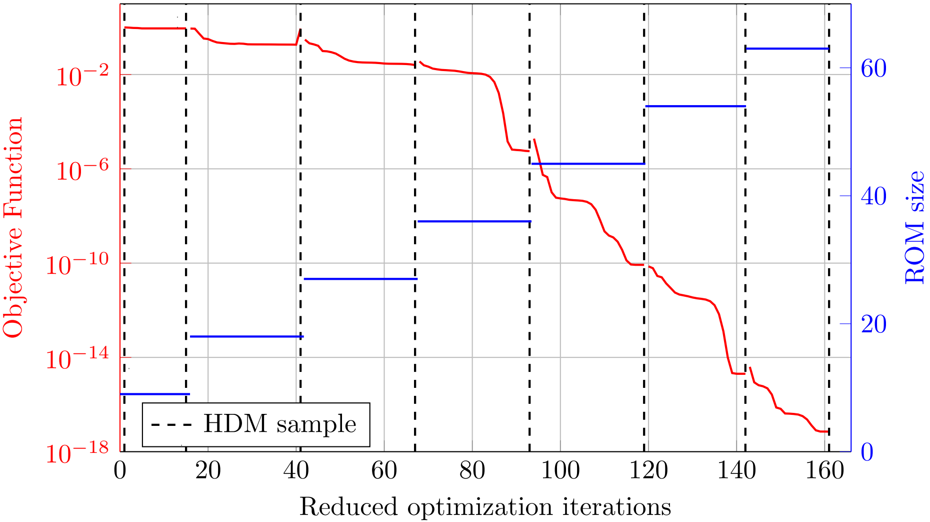
Proposed method: recovers target airfoil



Proposed method: 4× fewer HDM queries



At the cost of ROM queries

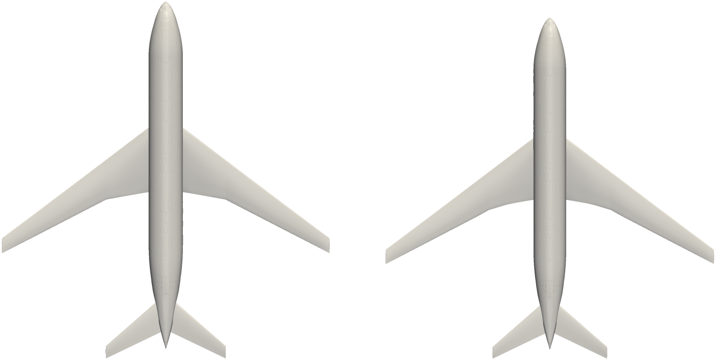


Shape optimization of aircraft in turbulent flow

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^4}{\text{minimize}} \quad -L_z(\boldsymbol{\mu})/L_x(\boldsymbol{\mu})$$
$$\text{subject to} \quad L_z(\boldsymbol{\mu}) = \bar{L}_z$$

- **Flow**: $M = 0.85 \quad \alpha = 2.32^\circ \quad Re = 5 \times 10^6$
- **Equations**: RANS with Spalart-Allmaras
- **Solver**: Vertex-centered finite volume method
- **Mesh**: **11.5M** nodes, **68M** tetra, **69M** DOF

$$\boldsymbol{\mu} = [\mathbf{L} \quad r_x \quad \phi \quad r_z]$$



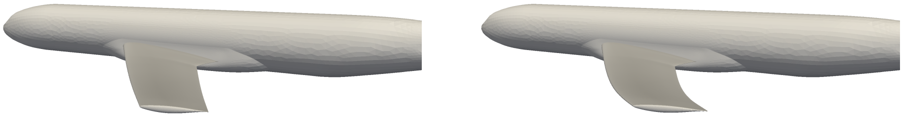
Wingspan

Shape optimization of aircraft in turbulent flow

$$\begin{aligned} \underset{\boldsymbol{\mu}\in\mathbb{R}^4}{\text{minimize}} \quad & -L_z(\boldsymbol{\mu})/L_x(\boldsymbol{\mu}) \\ \text{subject to} \quad & L_z(\boldsymbol{\mu}) = \bar{L}_z \end{aligned}$$

- **Flow**: $M = 0.85 \quad \alpha = 2.32^\circ \quad Re = 5 \times 10^6$
- **Equations**: RANS with Spalart-Allmaras
- **Solver**: Vertex-centered finite volume method
- **Mesh**: **11.5M** nodes, **68M** tetra, **69M** DOF

$$\boldsymbol{\mu} = [L \quad \mathbf{r_x} \quad \phi \quad r_z]$$



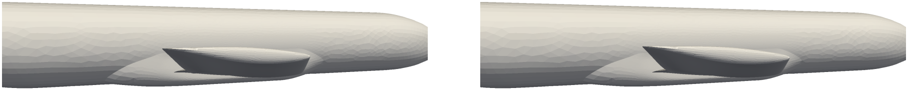
Localized sweep

Shape optimization of aircraft in turbulent flow

$$\begin{aligned} \underset{\boldsymbol{\mu}\in\mathbb{R}^4}{\text{minimize}} \quad & -L_z(\boldsymbol{\mu})/L_x(\boldsymbol{\mu}) \\ \text{subject to} \quad & L_z(\boldsymbol{\mu}) = \bar{L}_z \end{aligned}$$

- **Flow**: $M = 0.85 \quad \alpha = 2.32^\circ \quad Re = 5\times 10^6$
- **Equations**: RANS with Spalart-Allmaras
- **Solver**: Vertex-centered finite volume method
- **Mesh**: **11.5M** nodes, **68M** tetra, **69M** DOF

$$\boldsymbol{\mu} = [L \quad r_x \quad \phi \quad r_z]$$



Twist

Shape optimization of aircraft in turbulent flow

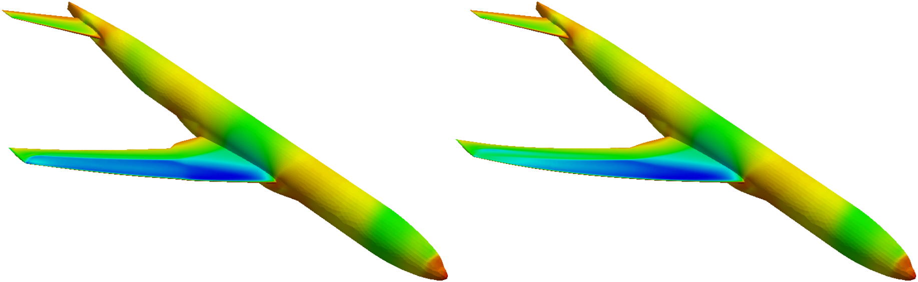
$$\begin{aligned} \underset{\boldsymbol{\mu} \in \mathbb{R}^4}{\text{minimize}} \quad & -L_z(\boldsymbol{\mu})/L_x(\boldsymbol{\mu}) \\ \text{subject to} \quad & L_z(\boldsymbol{\mu}) = \bar{L}_z \end{aligned}$$

- **Flow**: $M = 0.85 \quad \alpha = 2.32^\circ \quad Re = 5 \times 10^6$
- **Equations**: RANS with Spalart-Allmaras
- **Solver**: Vertex-centered finite volume method
- **Mesh**: **11.5M** nodes, **68M** tetra, **69M** DOF

$$\boldsymbol{\mu} = [L \quad r_x \quad \phi \quad \mathbf{r_z}]$$

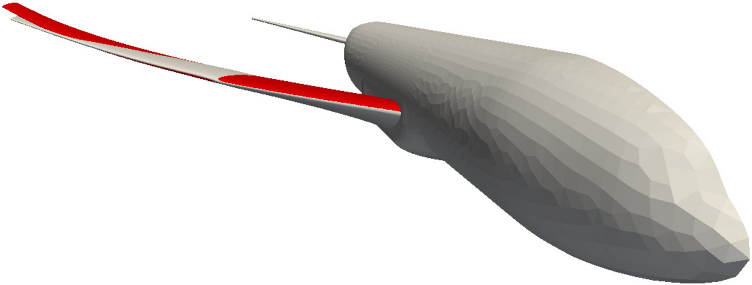
Localized dihedral

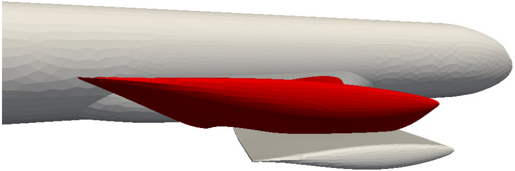
Optimized shape: reduction in **2.2** drag counts



Baseline (left) and optimized (right) shape – colored by C_p

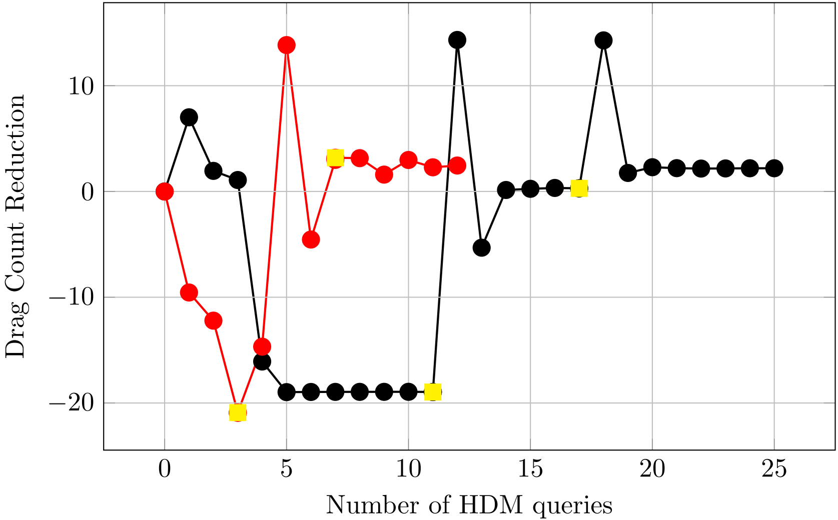
Optimized shape: reduction in **2.2** drag counts





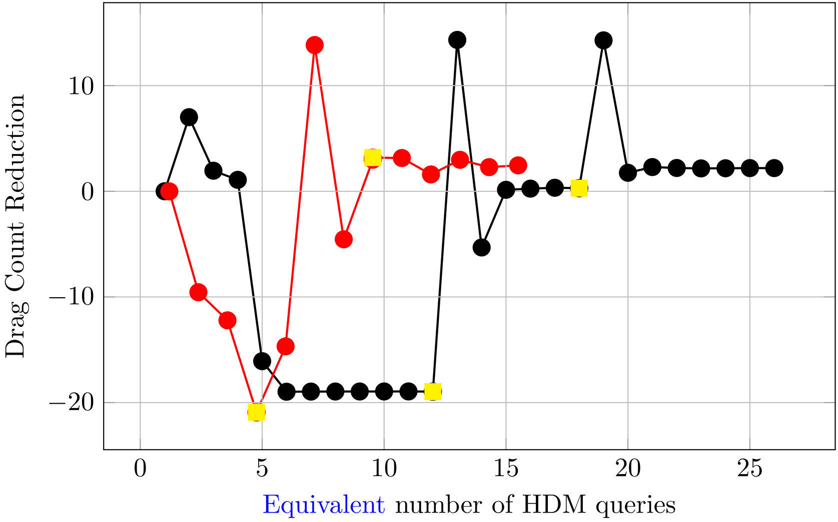
Baseline (gray) and optimized shape (red) – 2× magnification

Proposed method: **2x** reduction in number of HDM queries



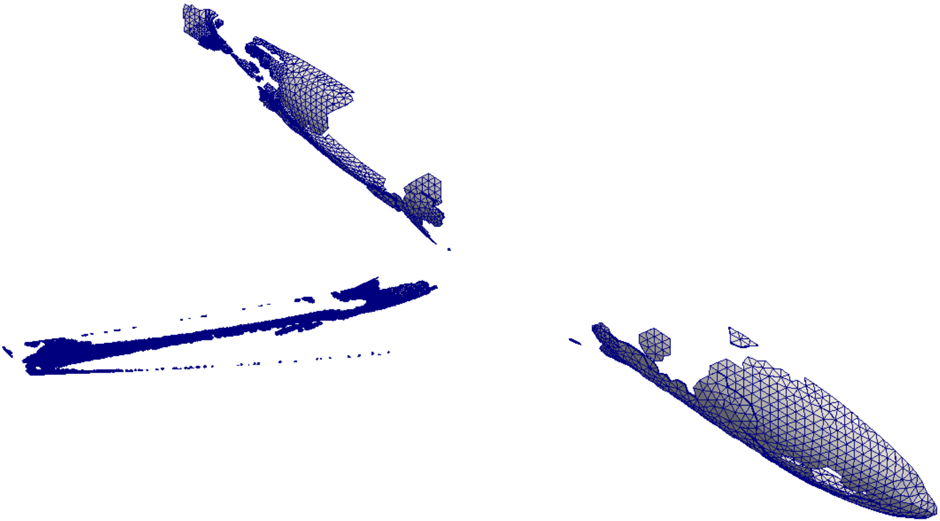
(—●—) HDM optimization, (—●—) ROM optimization,
(■) Augmented Lagrangian update

Proposed method: **1.6x** reduction in overall cost



(—●—) HDM optimization, (—●—) ROM optimization,
(■) Augmented Lagrangian update

Sample mesh has **0.6%** the nodes of the full mesh



The sample mesh at an intermediate iteration with **72k nodes**
(vs. the full mesh with **11.5M nodes**)

Stochastic PDE-constrained optimization formulation

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

- $\boldsymbol{r} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \to \mathbb{R}^{n_{\boldsymbol{u}}}$ — 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
- $\mathbb{E}[\mathcal{F}] \equiv \int_{\boldsymbol{\Xi}} \mathcal{F}(\boldsymbol{\xi})\rho(\boldsymbol{\xi})\, d\boldsymbol{\xi}$

First source of inexactness: anisotropic sparse grids

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

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

$$\Downarrow$$

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

[Kouri et al., 2013, Kouri et al., 2014]

Second source of inexactness: reduced-order models

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

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

$$\Downarrow$$

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

$$\Downarrow$$

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

Proposed approach: managed inexactness

Replace expensive PDE with inexpensive approximation model

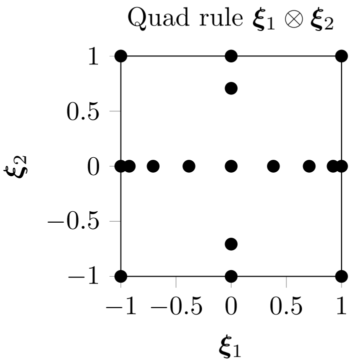
- **Reduced-order models** used for *inexact PDE evaluations*
- **Partially converged solutions** used for *inexact PDE evaluations*
- **Anisotropic sparse grids** used for *inexact integration* of risk measures

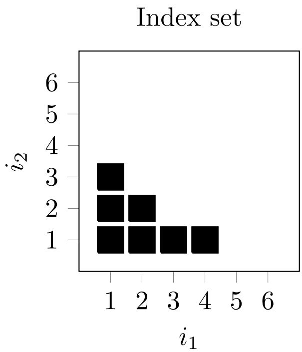
$$\underset{\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad F(\boldsymbol{\mu}) \qquad \longrightarrow \qquad \underset{\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad m_k(\boldsymbol{\mu})$$

Manage inexactness with trust region method

- Embedded in globally convergent **trust region** method
- **Error indicators** to account for *all* sources of inexactness
- **Refinement** of approximation model using *greedy algorithms*

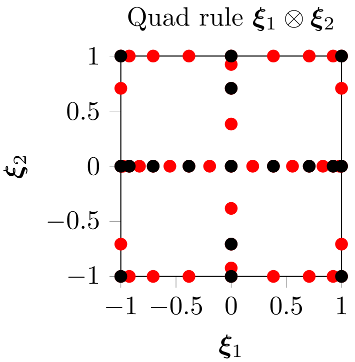
$$\underset{\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} \quad F(\boldsymbol{\mu}) \qquad \longrightarrow \qquad \begin{aligned} &\underset{\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\mu}}}}{\text{minimize}} && m_k(\boldsymbol{\mu}) \\ &\text{subject to} && ||\boldsymbol{\mu} - \boldsymbol{\mu}_k|| \leq \Delta_k \end{aligned}$$

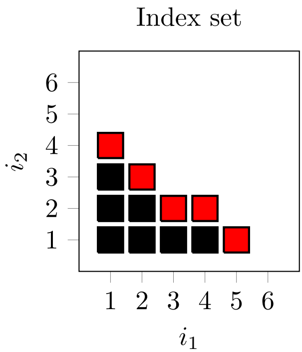




Index set ($\mathcal{I}$) – ● Neighbors ($\mathcal{N}(\mathcal{I})$) – ●

Source of inexactness: **anisotropic sparse grids**





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

Trust region method: ROM/SG approximation model

Approximation models built on two sources of inexactness

$$m_k(\boldsymbol{\mu}) = \mathbb{E}_{\mathcal{I}_k}\left[\mathcal{J}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu},\,\cdot),\,\boldsymbol{\mu},\,\cdot)\right]$$
$$\psi_k(\boldsymbol{\mu}) = \mathbb{E}_{\mathcal{I}'_k}\left[\mathcal{J}(\boldsymbol{\Phi}'_k \boldsymbol{u}_r(\boldsymbol{\mu},\,\cdot),\,\boldsymbol{\mu},\,\cdot)\right]$$

Error indicators that account for both sources of error

$$\vartheta_k(\boldsymbol{\mu}) = ||\boldsymbol{\mu} - \boldsymbol{\mu}_k||$$
$$\varphi_k(\boldsymbol{\mu}) = \alpha_1 \mathcal{E}_1(\boldsymbol{\mu};\,\mathcal{I}_k,\,\boldsymbol{\Phi}_k) + \alpha_2 \mathcal{E}_2(\boldsymbol{\mu};\,\mathcal{I}_k,\,\boldsymbol{\Phi}_k) + \alpha_3 \mathcal{E}_4(\boldsymbol{\mu};\,\mathcal{I}_k,\,\boldsymbol{\Phi}_k)$$
$$\theta_k(\boldsymbol{\mu}) = \beta_1(\mathcal{E}_1(\boldsymbol{\mu};\,\mathcal{I}'_k,\,\boldsymbol{\Phi}'_k) + \mathcal{E}_1(\boldsymbol{\mu}_k;\,\mathcal{I}'_k,\,\boldsymbol{\Phi}'_k)) + \beta_2(\mathcal{E}_3(\boldsymbol{\mu};\,\mathcal{I}'_k,\,\boldsymbol{\Phi}'_k) + \mathcal{E}_3(\boldsymbol{\mu}_k;\,\mathcal{I}'_k,\,\boldsymbol{\Phi}'_k))$$

Reduced-order model errors

$$\mathcal{E}_1(\boldsymbol{\mu};\,\mathcal{I},\,\boldsymbol{\Phi}) = \mathbb{E}_{\mathcal{I}\cup\mathcal{N}(\mathcal{I})}\left[||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\,\cdot),\,\boldsymbol{\mu},\,\cdot)||\right]$$
$$\mathcal{E}_2(\boldsymbol{\mu};\,\mathcal{I},\,\boldsymbol{\Phi}) = \mathbb{E}_{\mathcal{I}\cup\mathcal{N}(\mathcal{I})}\left[\left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\,\cdot),\,\boldsymbol{\Psi}\boldsymbol{\lambda}_r(\boldsymbol{\mu},\,\cdot),\,\boldsymbol{\mu},\,\cdot)\right|\right|\right]$$

Sparse grid truncation errors

$$\mathcal{E}_3(\boldsymbol{\mu};\,\mathcal{I},\,\boldsymbol{\Phi}) = \mathbb{E}_{\mathcal{N}(\mathcal{I})}\left[|\mathcal{J}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\,\cdot),\,\boldsymbol{\mu},\,\cdot)|\right]$$
$$\mathcal{E}_4(\boldsymbol{\mu};\,\mathcal{I},\,\boldsymbol{\Phi}) = \mathbb{E}_{\mathcal{N}(\mathcal{I})}\left[||\nabla\mathcal{J}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\,\cdot),\,\boldsymbol{\mu},\,\cdot)||\right]$$

Final requirement for convergence: **Adaptivity**

With the approximation model, $m_k(\boldsymbol{\mu})$, and gradient error indicator, $\varphi_k(\boldsymbol{\mu})$

$$m_k(\boldsymbol{\mu}) = \mathbb{E}_{\mathcal{I}_k}\left[\mathcal{J}(\boldsymbol{\Phi}_k \boldsymbol{u}_r(\boldsymbol{\mu},\,\cdot\,),\, \boldsymbol{\mu},\,\cdot\,)\right]$$
$$\varphi_k(\boldsymbol{\mu}) = \alpha_1 \mathcal{E}_1(\boldsymbol{\mu};\, \mathcal{I}_k,\, \boldsymbol{\Phi}_k) + \alpha_2 \mathcal{E}_2(\boldsymbol{\mu};\, \mathcal{I}_k,\, \boldsymbol{\Phi}_k) + \alpha_3 \mathcal{E}_4(\boldsymbol{\mu};\, \mathcal{I}_k,\, \boldsymbol{\Phi}_k)$$

the sparse grid $\mathcal{I}_k$ and reduced-order basis $\boldsymbol{\Phi}_k$ must be constructed such that the gradient condition holds

$$\varphi_k(\boldsymbol{\mu}_k) \leq \kappa_\varphi \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\}$$

Define dimension-adaptive greedy method to target each source of error such that the stronger conditions hold

$$\mathcal{E}_1(\boldsymbol{\mu}_k;\, \mathcal{I},\, \boldsymbol{\Phi}) \leq \frac{\kappa_\varphi}{3\alpha_1} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\}$$
$$\mathcal{E}_2(\boldsymbol{\mu}_k;\, \mathcal{I},\, \boldsymbol{\Phi}) \leq \frac{\kappa_\varphi}{3\alpha_2} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\}$$
$$\mathcal{E}_4(\boldsymbol{\mu}_k;\, \mathcal{I},\, \boldsymbol{\Phi}) \leq \frac{\kappa_\varphi}{3\alpha_3} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\}$$

while $\mathcal{E}_4(\mathbf{\Phi}, \mathcal{I}, \boldsymbol{\mu}_k) > \frac{\kappa_\varphi}{3\alpha_3} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,, \Delta_k\}$ **do**

Refine index set: Dimension-adaptive sparse grids

$$\mathcal{I}_k \leftarrow \mathcal{I}_k \cup \{\mathbf{j}^*\} \qquad \text{where} \qquad \mathbf{j}^* = \underset{\mathbf{j} \in \mathcal{N}(\mathcal{I}_k)}{\arg\max}\, \mathbb{E}_{\mathbf{j}}\left[||\nabla \mathcal{J}(\mathbf{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu}, \cdot), \boldsymbol{\mu}, \cdot)||\right]$$

Adaptivity: Dimension-adaptive greedy method

while $\mathcal{E}_4(\mathbf{\Phi}, \mathcal{I}, \boldsymbol{\mu}_k) > \frac{\kappa_\varphi}{3\alpha_3} \min\{\|\nabla m_k(\boldsymbol{\mu}_k)\|, \Delta_k\}$ **do**

Refine index set: Dimension-adaptive sparse grids

$$\mathcal{I}_k \leftarrow \mathcal{I}_k \cup \{\mathbf{j}^*\} \qquad \text{where} \qquad \mathbf{j}^* = \underset{\mathbf{j}\in\mathcal{N}(\mathcal{I}_k)}{\arg\max}\ \mathbb{E}_{\mathbf{j}}\left[\|\nabla\mathcal{J}(\mathbf{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu}, \cdot), \boldsymbol{\mu}, \cdot)\|\right]$$

Refine reduced-order basis: Greedy sampling

while $\mathcal{E}_1(\mathbf{\Phi}, \mathcal{I}, \boldsymbol{\mu}_k) > \frac{\kappa_\varphi}{3\alpha_1} \min\{\|\nabla m_k(\boldsymbol{\mu}_k)\|, \Delta_k\}$ **do**

$$\mathbf{\Phi}_k \leftarrow \begin{bmatrix}\mathbf{\Phi}_k & \boldsymbol{u}(\boldsymbol{\mu}_k, \boldsymbol{\xi}^*) & \boldsymbol{\lambda}(\boldsymbol{\mu}_k, \boldsymbol{\xi}^*)\end{bmatrix}$$

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}\in\mathbf{\Xi}_{\mathbf{j}^*}}{\arg\max}\ \rho(\boldsymbol{\xi})\,\|\boldsymbol{r}(\mathbf{\Phi}_k\boldsymbol{u}_r(\boldsymbol{\mu}_k, \boldsymbol{\xi}), \boldsymbol{\mu}_k, \boldsymbol{\xi})\|$$

end while

while $\mathcal{E}_4(\mathbf{\Phi}, \mathcal{I}, \boldsymbol{\mu}_k) > \frac{\kappa_\varphi}{3\alpha_3} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||, \Delta_k\}$ **do**

Refine index set: Dimension-adaptive sparse grids

$$\mathcal{I}_k \leftarrow \mathcal{I}_k \cup \{\mathbf{j}^*\} \qquad \text{where} \qquad \mathbf{j}^* = \underset{\mathbf{j}\in\mathcal{N}(\mathcal{I}_k)}{\arg\max}\ \mathbb{E}_{\mathbf{j}}\left[||\nabla\mathcal{J}(\mathbf{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu}, \cdot), \boldsymbol{\mu}, \cdot)||\right]$$

Refine reduced-order basis: Greedy sampling

while $\mathcal{E}_1(\mathbf{\Phi}, \mathcal{I}, \boldsymbol{\mu}_k) > \frac{\kappa_\varphi}{3\alpha_1} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||, \Delta_k\}$ **do**

$$\mathbf{\Phi}_k \leftarrow \begin{bmatrix}\mathbf{\Phi}_k & \boldsymbol{u}(\boldsymbol{\mu}_k, \boldsymbol{\xi}^*) & \boldsymbol{\lambda}(\boldsymbol{\mu}_k, \boldsymbol{\xi}^*)\end{bmatrix}$$

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}\in\mathbf{\Xi}_{\mathbf{j}^*}}{\arg\max}\ \rho(\boldsymbol{\xi})\,||\boldsymbol{r}(\mathbf{\Phi}_k\boldsymbol{u}_r(\boldsymbol{\mu}_k, \boldsymbol{\xi}), \boldsymbol{\mu}_k, \boldsymbol{\xi})||$$

end while

while $\mathcal{E}_2(\mathbf{\Phi}, \mathcal{I}, \boldsymbol{\mu}_k) > \frac{\kappa_\varphi}{3\alpha_2} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||, \Delta_k\}$ **do**

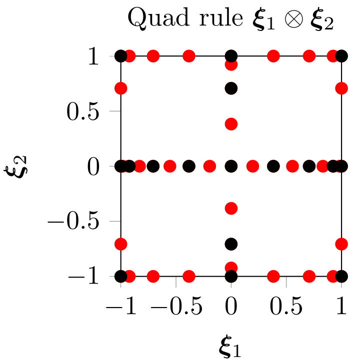
$$\mathbf{\Phi}_k \leftarrow \begin{bmatrix}\mathbf{\Phi}_k & \boldsymbol{u}(\boldsymbol{\mu}_k, \boldsymbol{\xi}^*) & \boldsymbol{\lambda}(\boldsymbol{\mu}_k, \boldsymbol{\xi}^*)\end{bmatrix}$$

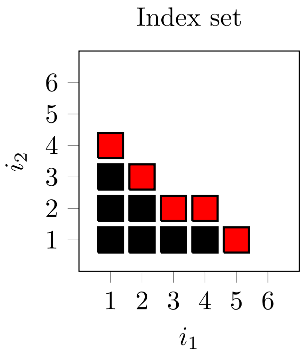
$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}\in\mathbf{\Xi}_{\mathbf{j}^*}}{\arg\max}\ \rho(\boldsymbol{\xi})\left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}(\mathbf{\Phi}_k\boldsymbol{u}_r(\boldsymbol{\mu}_k, \boldsymbol{\xi}), \mathbf{\Psi}_k\boldsymbol{\lambda}_r(\boldsymbol{\mu}_k, \boldsymbol{\xi}), \boldsymbol{\mu}_k, \boldsymbol{\xi})\right|\right|$$

end while

end while

Anisotropic sparse grid quadrature: neighbors





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

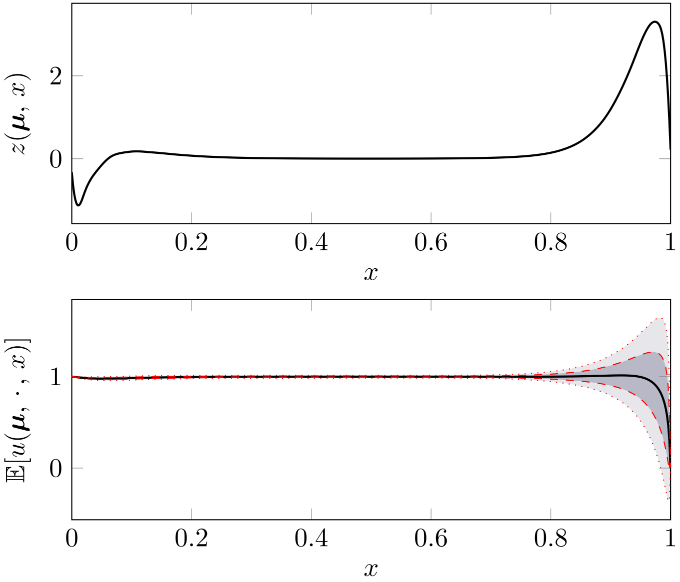
$$-\nu(\boldsymbol{\xi})\partial_{xx}u(\boldsymbol{\mu}, \boldsymbol{\xi}, x) + u(\boldsymbol{\mu}, \boldsymbol{\xi}, x)\partial_x u(\boldsymbol{\mu}, \boldsymbol{\xi}, x) = z(\boldsymbol{\mu}, x) \quad x \in (0, 1), \quad \boldsymbol{\xi} \in \boldsymbol{\Xi}$$
$$u(\boldsymbol{\mu}, \boldsymbol{\xi},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

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

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

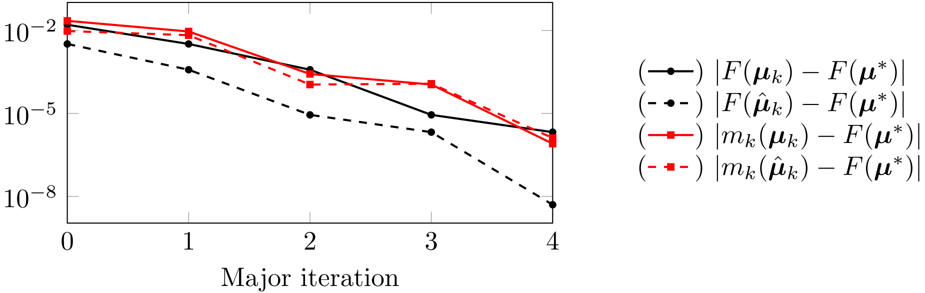
- **Parametrization:** $z(\boldsymbol{\mu}, x)$ – cubic splines with 51 knots, $n_{\boldsymbol{\mu}} = 53$

Optimal control and statistics



Optimal control and corresponding mean state (——) ± one (- - -) and two (······) standard deviations

Global convergence without pointwise agreement

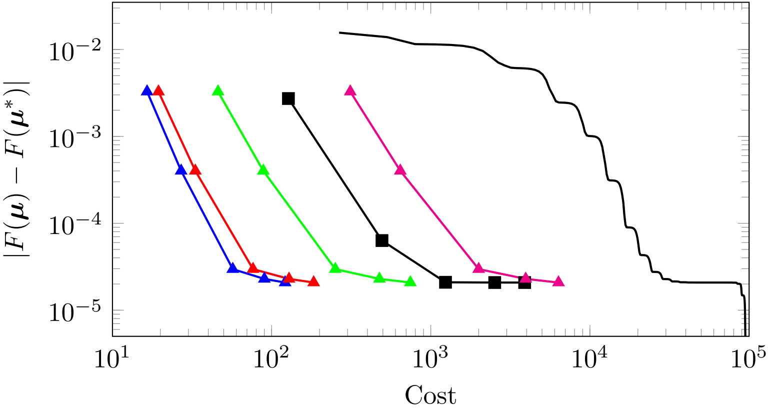


$F(\boldsymbol{\mu}_k)$	$m_k(\boldsymbol{\mu}_k)$	$F(\hat{\boldsymbol{\mu}}_k)$	$m_k(\hat{\boldsymbol{\mu}}_k)$	$\|\|\nabla F(\boldsymbol{\mu}_k)\|\|$	ρ_k	Success?
6.6506e-02	7.2694e-02	5.3655e-02	5.9922e-02	2.2959e-02	1.0257e+00	1.0000e+00
5.3655e-02	5.9593e-02	5.0783e-02	5.7152e-02	2.3424e-03	9.7512e-01	1.0000e+00
5.0783e-02	5.0670e-02	5.0412e-02	5.0292e-02	1.9724e-03	9.8351e-01	1.0000e+00
5.0412e-02	5.0292e-02	5.0405e-02	5.0284e-02	9.2654e-05	8.7479e-01	1.0000e+00
5.0405e-02	5.0404e-02	5.0403e-02	5.0401e-02	8.3139e-05	9.9946e-01	1.0000e+00
5.0403e-02	5.0401e-02	-	-	2.2846e-06	-	-

Convergence history of trust region method built on two-level approximation

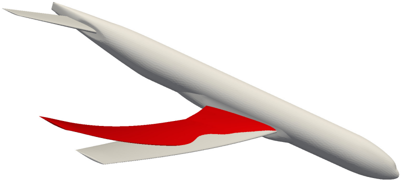
Significant reduction in cost, even if (largest) ROM only 10× faster than HDM

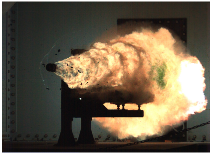
$$\text{Cost} = \text{nHdmPrim} + 0.5 \times \text{nHdmAdj} + \tau^{-1} \times (\text{nRomPrim} + 0.5 \times \text{nRomAdj})$$

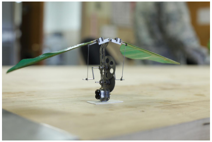


5-level isotropic SG (—), dimension-adaptive SG [Kouri et al., 2014] (-■-), and proposed ROM/SG for $\tau = 1$ (-▲-), $\tau = 10$ (-▲-), $\tau = 100$ (-▲-), $\tau = \infty$ (-▲-)

- Framework introduced for accelerating **deterministic** and **stochastic** PDE-constrained optimization problems
  - Adaptive *model reduction*
  - *Partially converged* primal and adjoint solutions
  - Dimension-adaptive *sparse grids*
- Inexactness **managed** with flexible **trust region** method
- Applied to variety of problems in computational mechanics and outperforms state-of-the-art methods
  - **1.6**× speedup on (deterministic) shape design of aircraft
  - **100**× speedup on (stochastic) optimal control of 1D flow

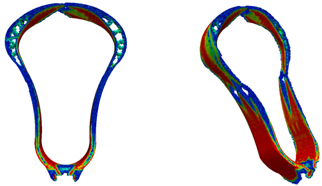


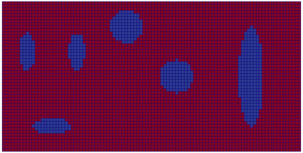


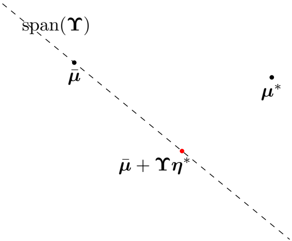


Extension to problems with **many parameters**

- Topology optimization[4] and inverse problems
- **Nested reduction** of state and parameter
- Multifidelity trust region method to globalize **state** reduction
- Linesearch/subspace method to globalize **parameter** reduction

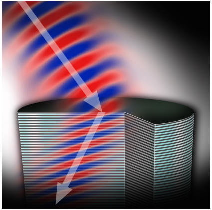


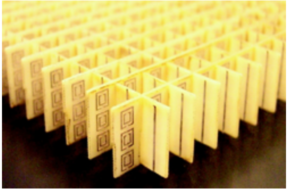




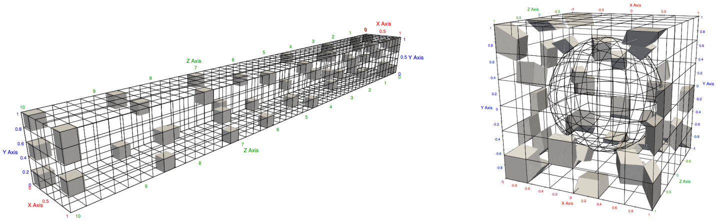
[4]Increasingly relevant due to emergence of Additive Manufacturing – ***MIT Technology Review*, Top 10 Technological Breakthrough 2013**

Extension to **multiscale** problems





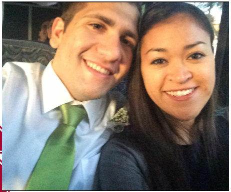
- **Existing multiscale methods** are extremely expensive
  - Single simulation: 203 hours ($\approx$ 8.5 days), 41760 cores [Knap et. al., 2016]
  - Not amenable to optimization (many-query)
- **Hyperreduced models** at each scale [Zahr et al., 2016a] – embedded in trust region optimization framework to *design microstructure* to achieve *macroscale objectives*

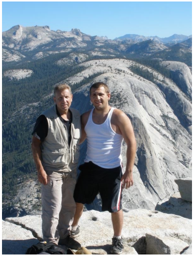


Hyperreduced model for macroscale (left) and microstructure (right)

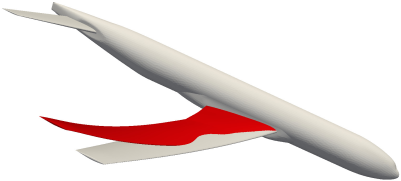
Acknowledgments

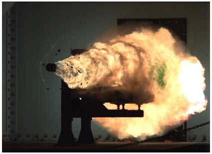
- **Funding**: DOE CSGF, Boeing
- **Advisor**: Charbel Farhat
- **Practicum mentor**: Per-Olof Persson
- **Committee**: Michael Saunder, Walter Murray, Louis Durlofsky
- **FRG**: Kyle Washabaugh, Alex Main, Todd Chapman, Raunak Borker, Kevin Carlberg, Philip Avery, Grace Fontanilla, Tatiana Wilson
- **Family**: Theresa Yates, Mom and Dad, Grandma and Grandpa, Bob and Emily, Uncle Jack, Aunt Nini and Allie, Yates Family

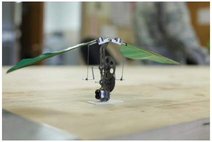




- Framework introduced for accelerating **deterministic** and **stochastic** PDE-constrained optimization problems
  - Adaptive *model reduction*
  - *Partially converged* primal and adjoint solutions
  - Dimension-adaptive *sparse grids*
- Inexactness **managed** with flexible **trust region** method
- Applied to variety of problems in computational mechanics and outperforms state-of-the-art methods
  - **1.6**× speedup on (deterministic) shape design of aircraft
  - **100**× speedup on (stochastic) optimal control of 1D flow







References I

Alexandrov, N. M., Dennis Jr, J. E., Lewis, R. M., and Torczon, V. (1998).
A trust-region framework for managing the use of approximation models in optimization.
Structural Optimization, 15(1):16–23.

Carter, R. G. (1989).
Numerical optimization in hilbert space using inexact function and gradient evaluations.

Carter, R. G. (1991).
On the global convergence of trust region algorithms using inexact gradient information.
SIAM Journal on Numerical Analysis, 28(1):251–265.

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

Heinkenschloss, M. and Vicente, L. N. (2002).
Analysis of inexact trust-region sqp algorithms.
SIAM Journal on Optimization, 12(2):283–302.

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.

Moré, J. J. (1983).
Recent developments in algorithms and software for trust region methods.
Springer.

Washabaugh, K. (2016).
Faster Fidelity For Better Design: A Scalable Model Order Reduction Framework For Steady Aerodynamic Design Applications.
PhD thesis, Stanford University.

Zahr, M. J. (2016).
Adaptive model reduction to accelerate optimization problems governed by partial differential equations.
PhD thesis, Stanford University.

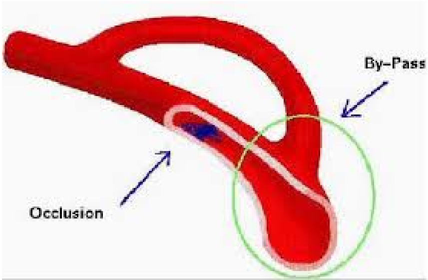
Zahr, M. J., Avery, P., and Farhat, C. (2016a).
A multilevel projection-based model reduction framework for efficient multiscale modeling.
International Journal for Numerical Methods in Engineering.

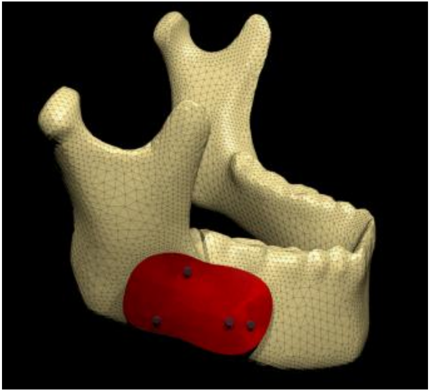
Zahr, M. J. and Persson, P.-O. (In review, 2016).
An adjoint method for a high-order discretization of deforming domain conservation laws for optimization of flow problems.
Journal of Computational Physics.

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

PDE optimization is **ubiquitous** in science and engineering

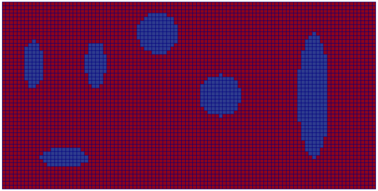
Design: Find system that optimizes performance metric, satisfies constraints

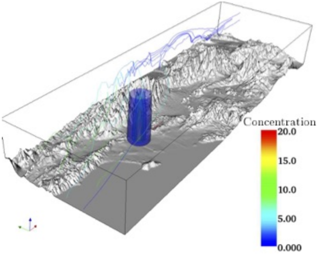




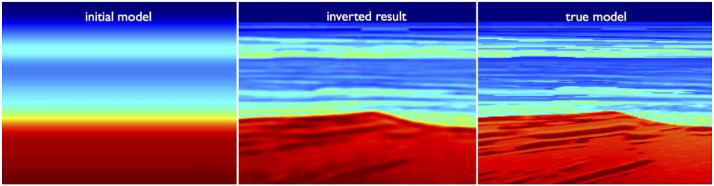
Shape design of arterial bypass (left) and shape/topology design of patient-specific implant (right)

Inverse problems: Infer the problem setup given solution observations





Left: Material inversion – find inclusions from acoustic, structural measurements
Right: Source inversion – find source of airborne contaminant from downstream measurements



Full waveform inversion – estimate subsurface of Earth's crust from acoustic measurements

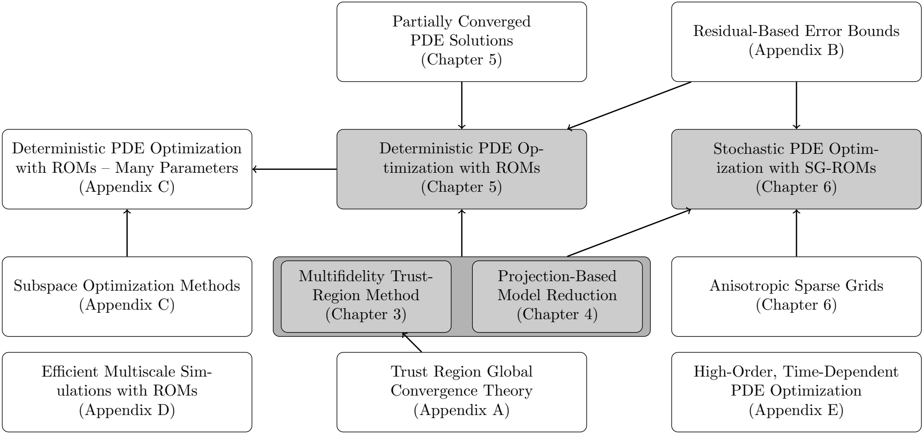
Applications in computational mechanics: dynamic

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., 2016b]

Thesis organization



Overview of global convergence theory[5]

Let $\{\boldsymbol{\mu}_k\}$ be a sequence of iterates produced by the algorithm and suppose there exists $\epsilon > 0$ such that $||\nabla m_k(\boldsymbol{\mu}_k)|| > 0$

Lemma 1: $\Delta_k \to 0$

- Fraction of Cauchy decrease
- $|F(\boldsymbol{\mu}_k) - F(\hat{\boldsymbol{\mu}}_k) + \psi_k(\hat{\boldsymbol{\mu}}_k) - \psi_k(\boldsymbol{\mu}_k)| \leq \sigma \left[\eta \min\{m_k(\boldsymbol{\mu}_k) - m_k(\hat{\boldsymbol{\mu}}_k),\, r_k\}\right]^{1/\omega}$

Lemma 2: $\rho_k \to 1$

- Fraction of Cauchy decrease
- $|F(\boldsymbol{\mu}_k) - F(\hat{\boldsymbol{\mu}}_k) + m_k(\hat{\boldsymbol{\mu}}_k) - m_k(\boldsymbol{\mu}_k)| \leq \zeta \Delta_k$

Theorem 1: $\liminf ||\nabla F(\boldsymbol{\mu}_k)|| = 0$

- Contradiction from Lemma 1 and 2 $\implies \liminf ||\nabla m_k(\boldsymbol{\mu}_k)|| = 0$
- $||\nabla F(\boldsymbol{\mu}_k) - \nabla m_k(\boldsymbol{\mu}_k)|| \leq \xi \min\{||\nabla m_k(\boldsymbol{\mu})||\,,\, \Delta_k\}$

[5] Closely parallels convergence theory in [Moré, 1983, Kouri et al., 2014]

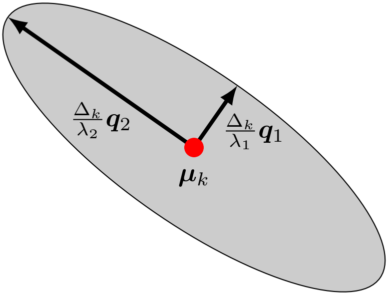
An interpretation of **error-aware** trust regions

Let $\boldsymbol{\vartheta}_k(\boldsymbol{\mu})$ be a vector-valued error indicator such that $\vartheta_k(\boldsymbol{\mu}) = ||\boldsymbol{\vartheta}_k(\boldsymbol{\mu})||_2$ and

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

Then, to first order[6],

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



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

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

Optimization of the Rosenbrock function

$$\underset{\boldsymbol{\mu}\in\mathbb{R}^2}{\text{minimize}} \quad F(\boldsymbol{\mu}) \equiv 100(\mu_2 - \mu_1^2)^2 + (1-\mu_1)^2.$$

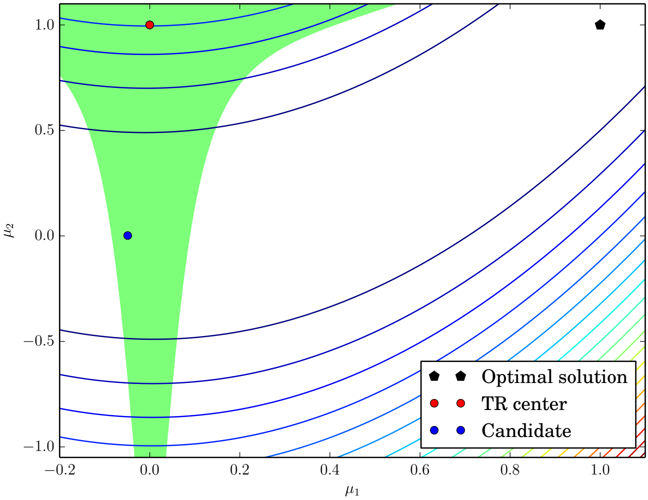
using the approximation models and error indicators

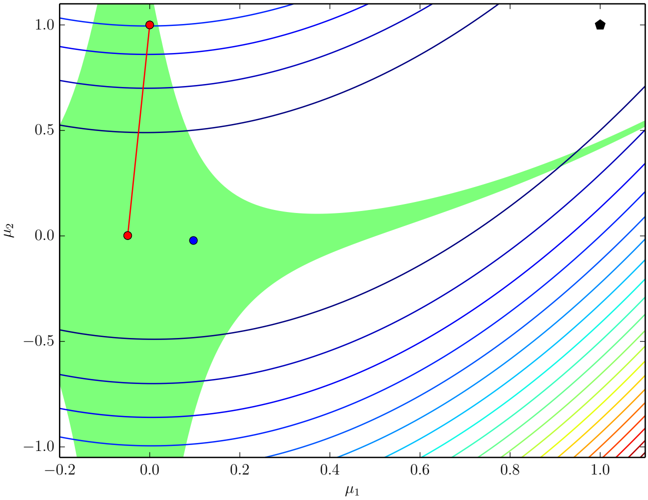
$$\begin{aligned}
m_k(\boldsymbol{\mu}) &\equiv G_k(\boldsymbol{\mu};\ \epsilon_k,\ \delta_k)\\
\psi_k(\boldsymbol{\mu}) &\equiv F(\boldsymbol{\mu})\\
\vartheta_k(\boldsymbol{\mu}) &\equiv |F(\boldsymbol{\mu}) - G_k(\boldsymbol{\mu};\, \epsilon_k,\, \delta_k)| + |F(\boldsymbol{\mu}_k) - G_k(\boldsymbol{\mu}_k;\, \epsilon_k,\, \delta_k)|\\
\varphi_k(\boldsymbol{\mu}) &\equiv ||\nabla F(\boldsymbol{\mu}) - \nabla G_k(\boldsymbol{\mu};\, \epsilon_k,\, \delta_k)||\\
\theta_k(\boldsymbol{\mu}) &\equiv 0
\end{aligned}$$

where $G_k(\boldsymbol{\mu};\, \epsilon_k,\, \delta_k)$ is the inexact quadratic approximation of F at $\boldsymbol{\mu}_k$

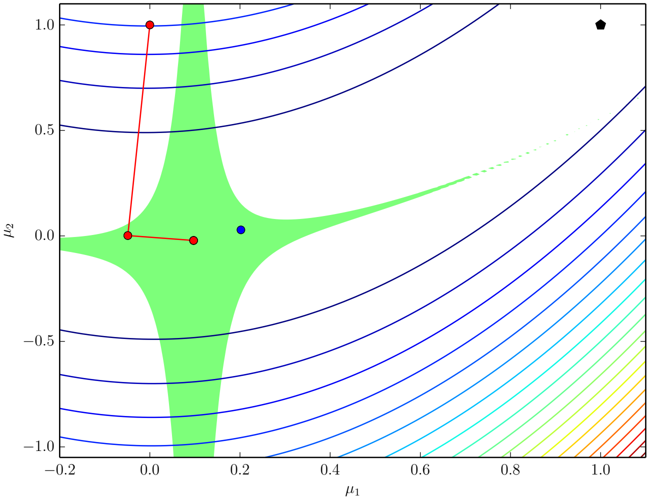
$$G_k(\boldsymbol{\mu};\, \epsilon,\, \delta) \equiv F(\boldsymbol{\mu}_k) + \epsilon + (\nabla F(\boldsymbol{\mu}_k) + \delta\mathbf{1})^T(\boldsymbol{\mu} - \boldsymbol{\mu}_k) + \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_k)^T \nabla^2 F(\boldsymbol{\mu}_k)(\boldsymbol{\mu} - \boldsymbol{\mu}_k)$$

A look at **error-aware** trust regions

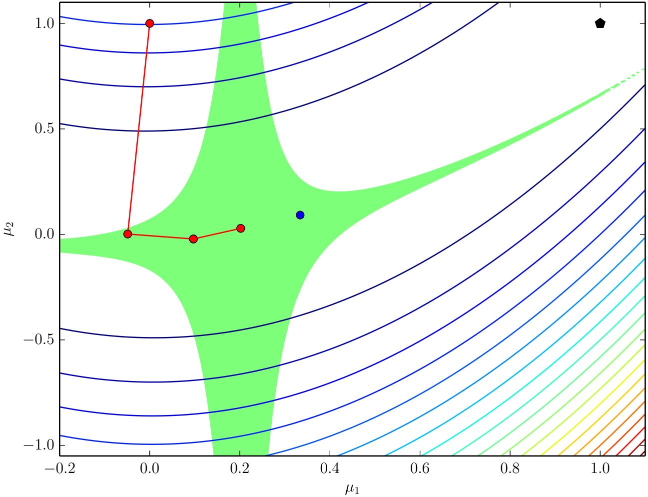



1.0
0.5
0.0
−0.5
−1.0
μ_2
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
μ_1

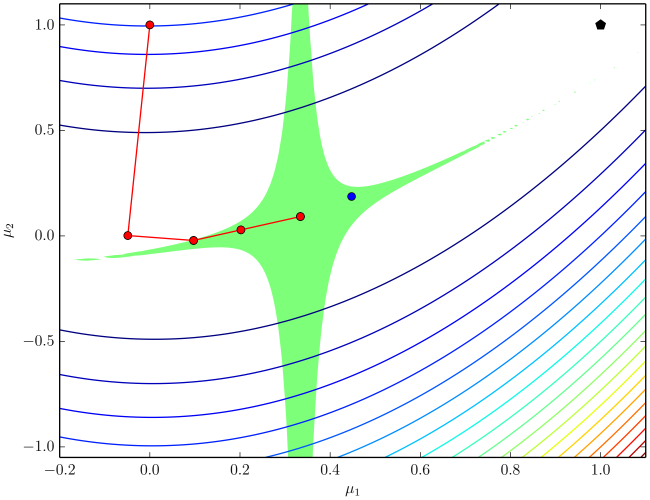
A look at **error-aware** trust regions



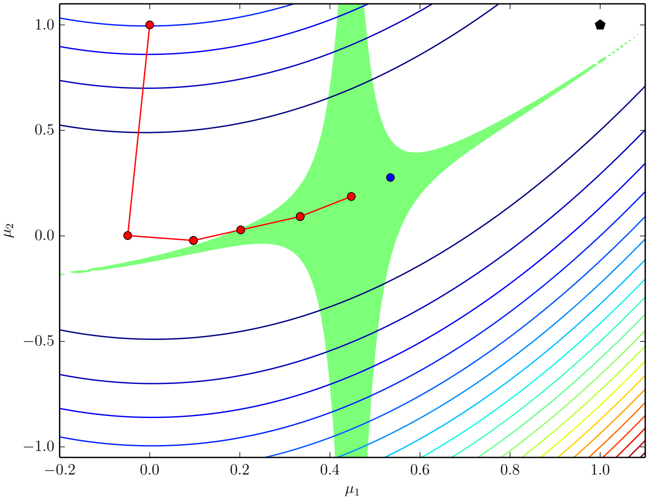
A look at **error-aware** trust regions



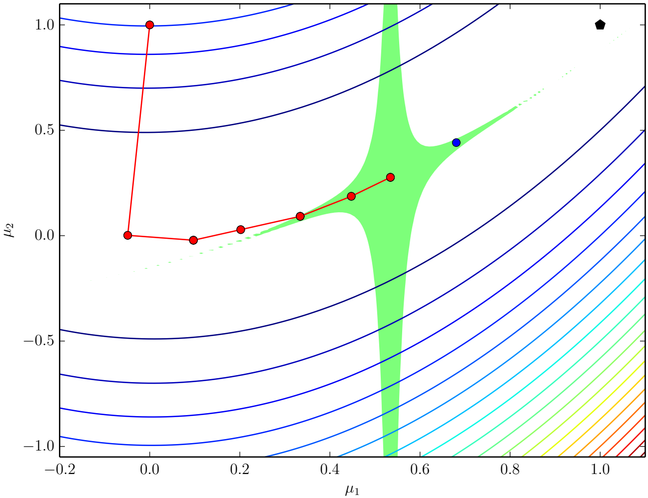
A look at **error-aware** trust regions



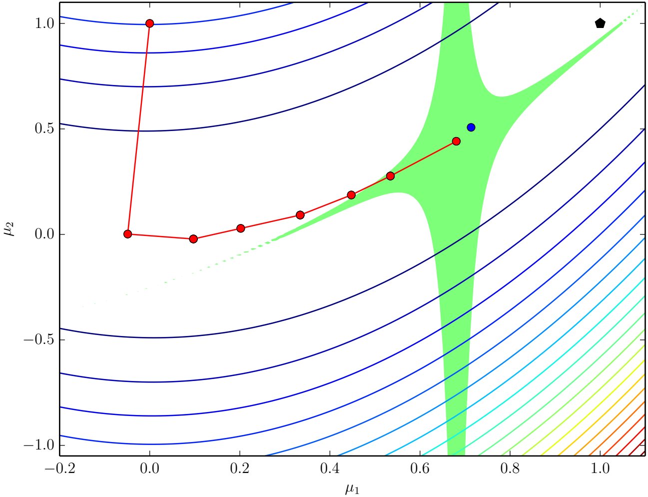
A look at **error-aware** trust regions



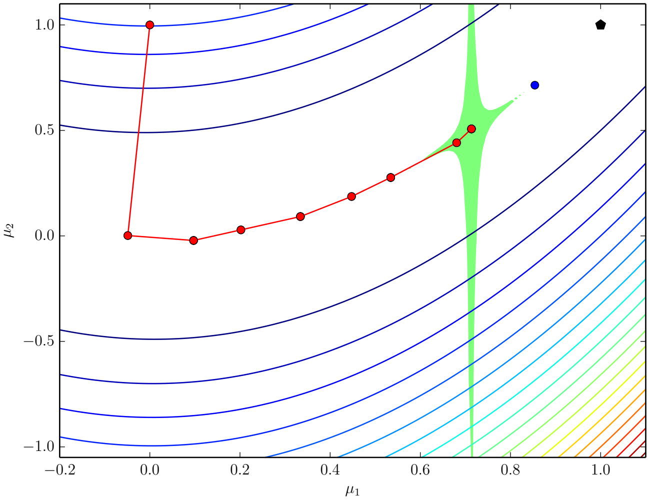
A look at **error-aware** trust regions



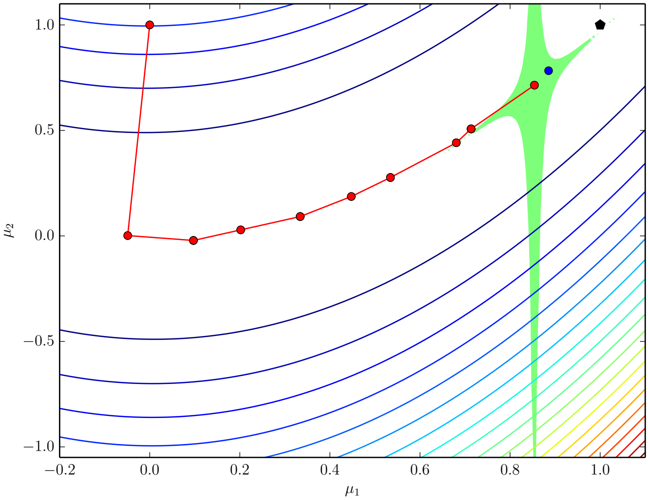
A look at **error-aware** trust regions



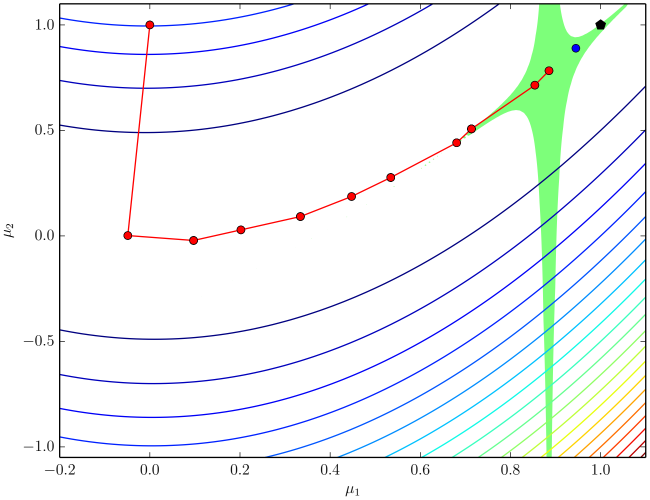
A look at **error-aware** trust regions



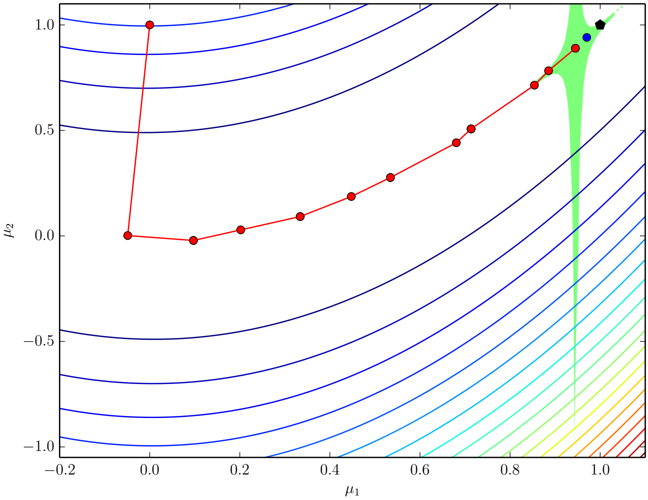
A look at **error-aware** trust regions



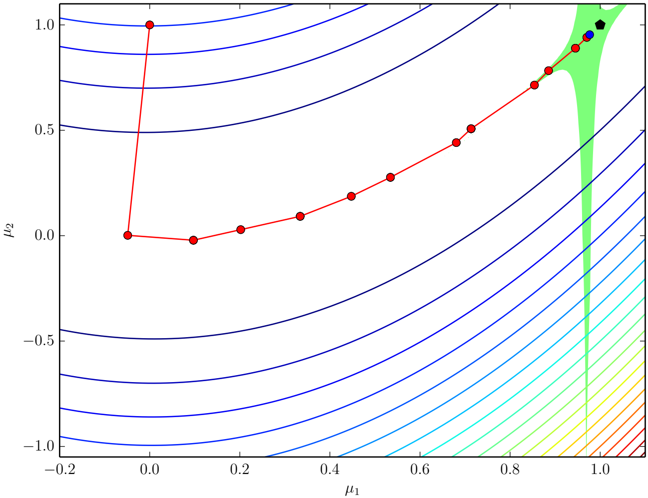
A look at **error-aware** trust regions



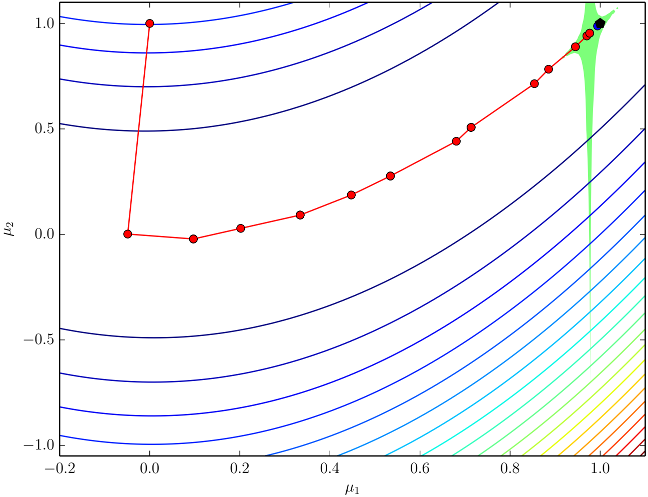
A look at **error-aware** trust regions



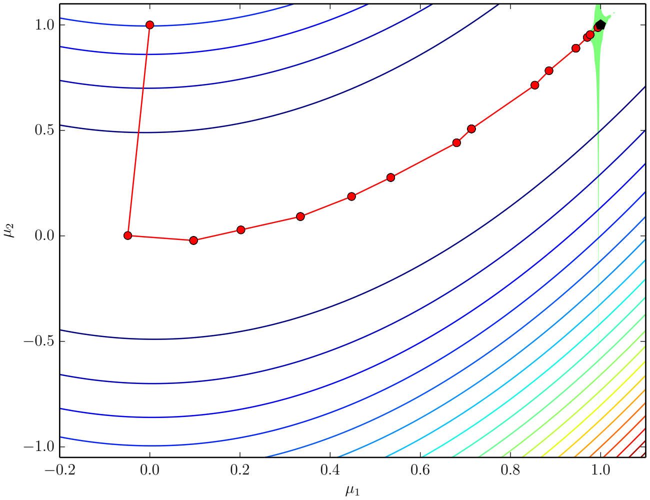
A look at **error-aware** trust regions



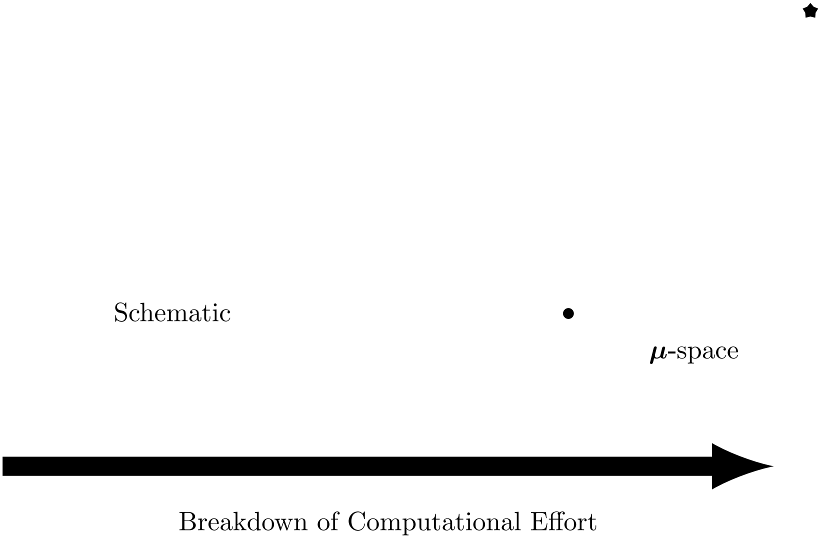
A look at **error-aware** trust regions

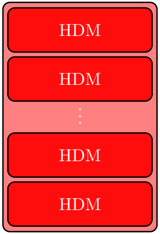


A look at **error-aware** trust regions

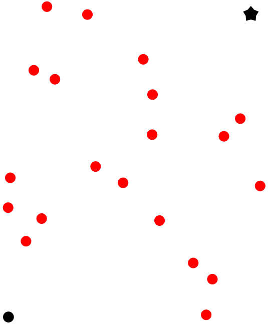


Offline-online approach to optimization with ROMs





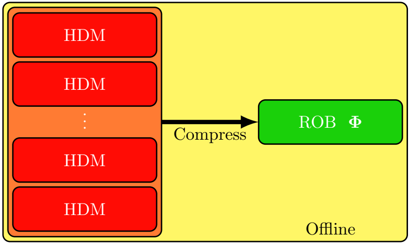
Schematic



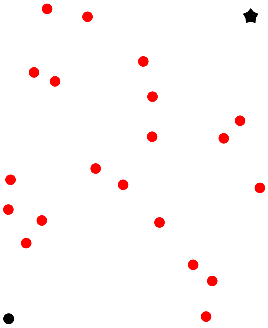
μ-space

Breakdown of Computational Effort

Offline-online approach to optimization with ROMs



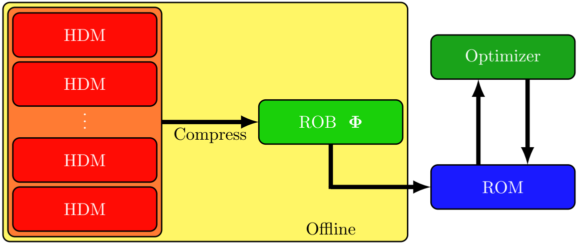
Schematic



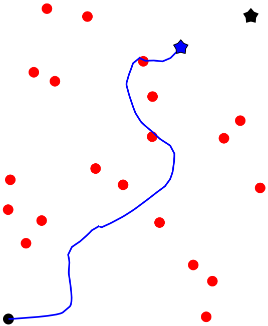
μ-space

Breakdown of Computational Effort

Offline-online approach to optimization with ROMs

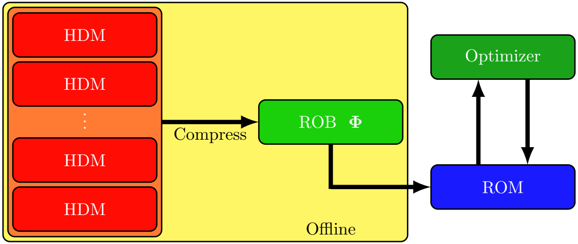


Schematic

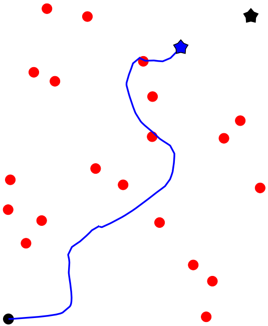


μ-space

Breakdown of Computational Effort



Schematic

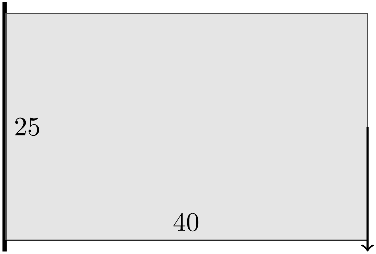


μ-space

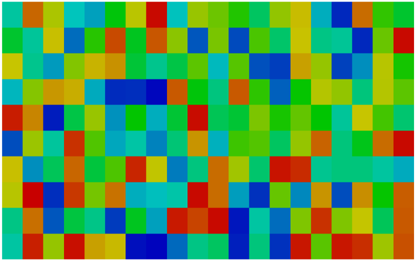
Breakdown of Computational Effort

No convergence **Scales exponentially with N_μ**

- *Greedy* Training
  - 5000 candidate points (LHS)
  - 50 snapshots
  - Error indicator: $||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r, \boldsymbol{\mu})||$
- State reduction ($\boldsymbol{\Phi}$)
  - POD
  - $k_{\boldsymbol{u}} = 25$
  - Polynomialization acceleration

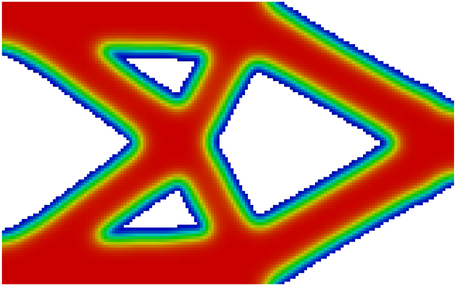


Stiffness maximization, volume constraint

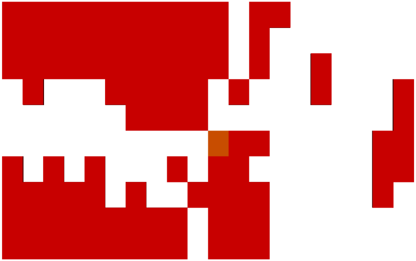


Parametrization with $n_{\boldsymbol{\mu}} = 200$

Numerical demonstration: offline-online breakdown



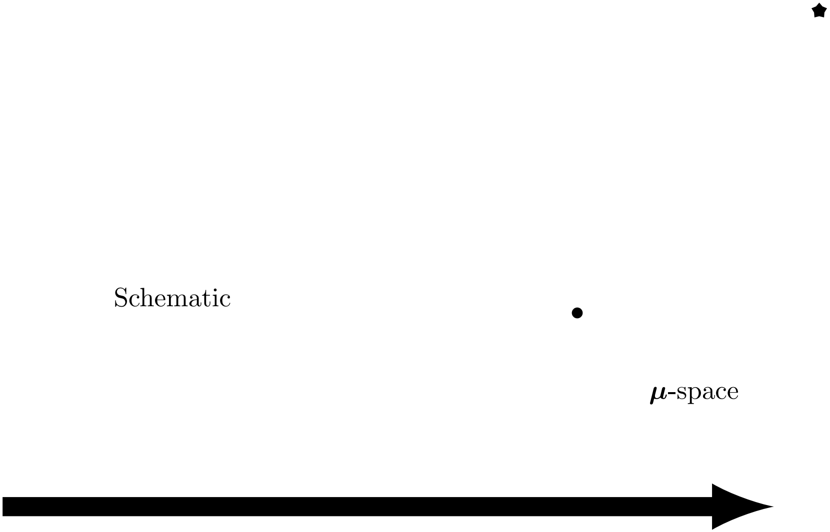
Optimal Solution
(1.97×10^4 s)



ROM Solution

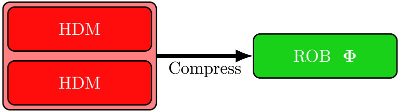
HDM Solution	**ROB** Construction	**Greedy** Algorithm	**ROM** Optimization
2.84×10^3 s	5.48×10^4 s	1.67×10^5 s	30 s
1.26%	24.36%	74.37%	0.01%

Schematic

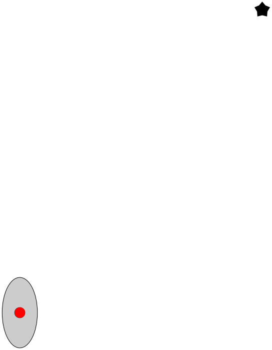


Breakdown of Computational Effort

Trust region framework for optimization with ROMs

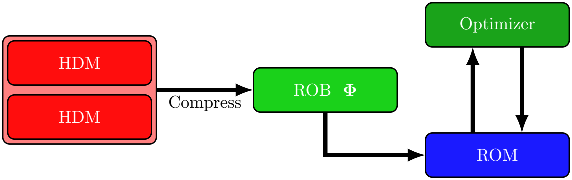


Schematic

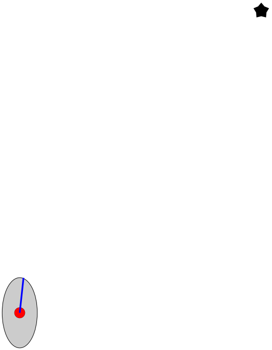


μ-space

Breakdown of Computational Effort



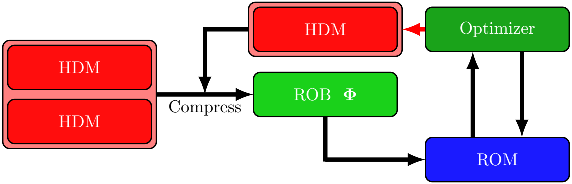
Schematic



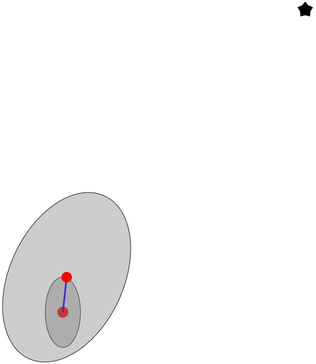
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust region framework for optimization with ROMs

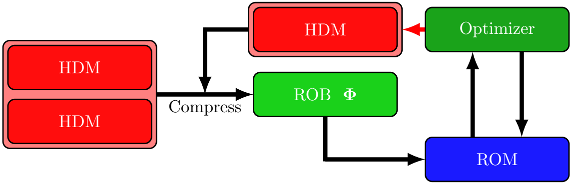


Schematic

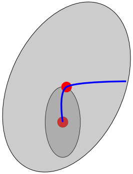


$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

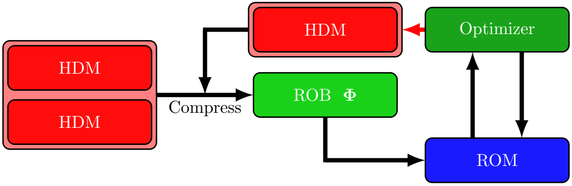


Schematic

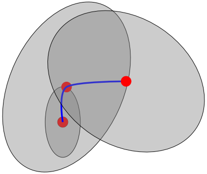


μ-space

Breakdown of Computational Effort



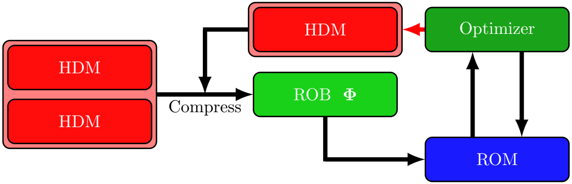
Schematic



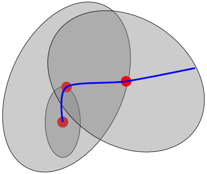
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust region framework for optimization with ROMs

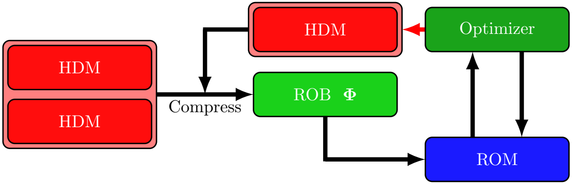


Schematic

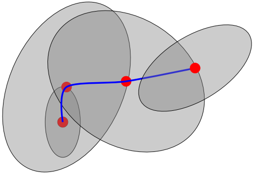


μ-space

Breakdown of Computational Effort



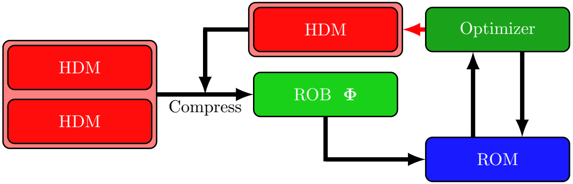
Schematic



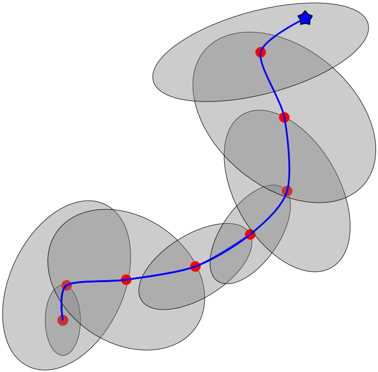
$\boldsymbol{\mu}$-space

Breakdown of Computational Effort

Trust region framework for optimization with ROMs

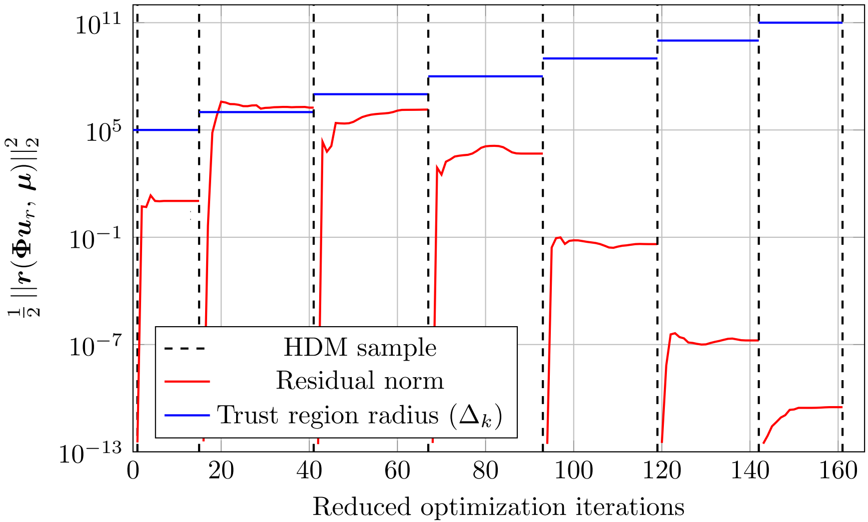


Schematic



$\boldsymbol{\mu}$-space

Breakdown of Computational Effort



10^{11}
10^{5}
10^{-1}
10^{-7}
10^{-13}
0
20
40
60
80
100
120
140
160
Reduced optimization iterations
$\frac{1}{2}||\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r, \boldsymbol{\mu})||_2^2$
HDM sample
Residual norm
Trust region radius (Δ_k)

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

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

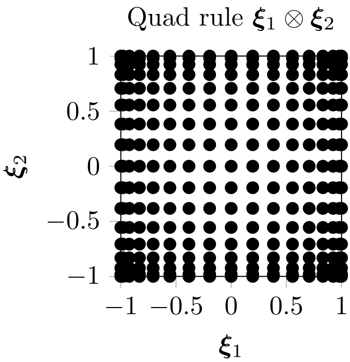
Neighbors: Let $\mathcal{I}^c = \mathbb{N}^{n_{\boldsymbol{\xi}}} \setminus \mathcal{I}$

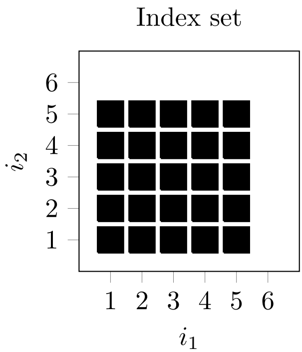
$$\mathcal{N}(\mathcal{I}) = \{\boldsymbol{i} \in \mathcal{I}^c \mid \boldsymbol{i} - \boldsymbol{e}_j \in \mathcal{I},\, j = 1,\, \ldots,\, n_{\boldsymbol{\xi}}\}$$

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

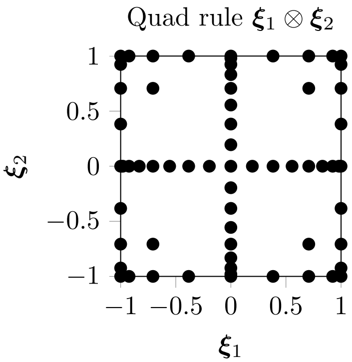
$$\mathbb{E} - \mathbb{E}_{\mathcal{I}} = \sum_{\mathbf{i}\in\mathcal{I}^c} \Delta_1^{i_1} \otimes \cdots \otimes \Delta_{n_{\boldsymbol{\xi}}}^{i_{n_{\boldsymbol{\xi}}}} \approx \sum_{\mathbf{i}\in\mathcal{N}(\mathcal{I})} \Delta_1^{i_1} \otimes \cdots \otimes \Delta_{n_{\boldsymbol{\xi}}}^{i_{n_{\boldsymbol{\xi}}}}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

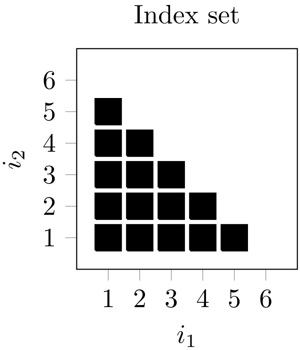
Tensor product quadrature





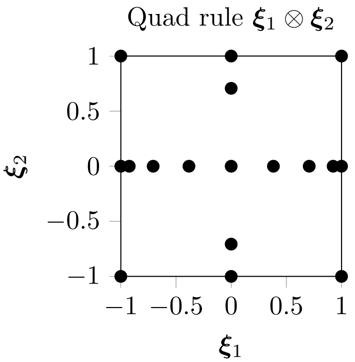
Index set ($\mathcal{I}$) – ● Neighbors ($\mathcal{N}(\mathcal{I})$) – ●

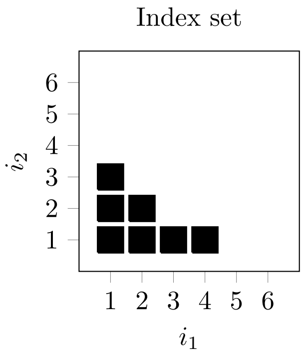




Index set $(\mathcal{I})$ – ● Neighbors $(\mathcal{N}(\mathcal{I}))$ – ●

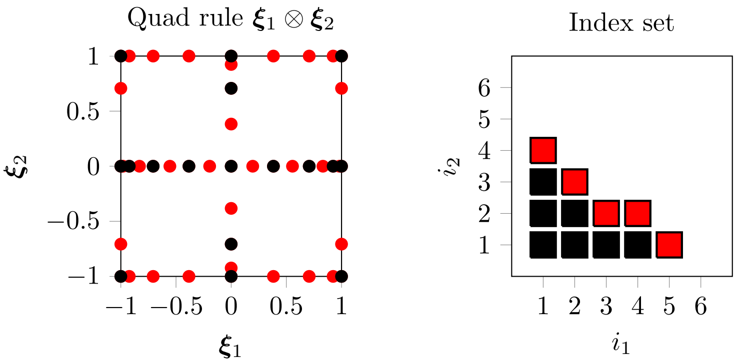
Anisotropic sparse grid quadrature





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

Anisotropic sparse grid quadrature: neighbors



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

Derivation of gradient error indicator

For brevity, let

$$
\begin{aligned}
\mathcal{J}(\boldsymbol{\xi}) &\leftarrow \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\nabla\mathcal{J}(\boldsymbol{\xi}) &\leftarrow \nabla\mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\mathcal{J}_r(\boldsymbol{\xi}) &= \mathcal{J}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\nabla\mathcal{J}_r(\boldsymbol{\xi}) &= \nabla\mathcal{J}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\boldsymbol{r}_r(\boldsymbol{\xi}) &= \boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\boldsymbol{r}_r^{\boldsymbol{\lambda}}(\boldsymbol{\xi}) &= \boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\Psi}\boldsymbol{\lambda}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi})
\end{aligned}
$$

Separate total error into contributions from ROM inexactness and SG truncation

$$
||\mathbb{E}[\nabla\mathcal{J}] - \mathbb{E}_{\mathcal{I}}[\nabla\mathcal{J}_r]|| \leq \mathbb{E}\left[||\nabla\mathcal{J} - \nabla\mathcal{J}_r||\right] + ||\mathbb{E}\left[\nabla\mathcal{J}_r\right] - \mathbb{E}_{\mathcal{I}}\left[\nabla\mathcal{J}_r\right]||
$$

Derivation of gradient error indicator

For brevity, let

$$
\begin{aligned}
\mathcal{J}(\boldsymbol{\xi}) &\leftarrow \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\nabla \mathcal{J}(\boldsymbol{\xi}) &\leftarrow \nabla \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\mathcal{J}_r(\boldsymbol{\xi}) &= \mathcal{J}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\nabla \mathcal{J}_r(\boldsymbol{\xi}) &= \nabla \mathcal{J}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\boldsymbol{r}_r(\boldsymbol{\xi}) &= \boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi}) \\
\boldsymbol{r}_r^{\boldsymbol{\lambda}}(\boldsymbol{\xi}) &= \boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\Psi}\boldsymbol{\lambda}_r(\boldsymbol{\mu},\, \boldsymbol{\xi}),\, \boldsymbol{\mu},\, \boldsymbol{\xi})
\end{aligned}
$$

Separate total error into contributions from ROM inexactness and SG truncation

$$
\begin{aligned}
||\mathbb{E}[\nabla \mathcal{J}] - \mathbb{E}_{\mathcal{I}}[\nabla \mathcal{J}_r]|| &\leq \mathbb{E}\left[||\nabla \mathcal{J} - \nabla \mathcal{J}_r||\right] + ||\mathbb{E}\left[\nabla \mathcal{J}_r\right] - \mathbb{E}_{\mathcal{I}}\left[\nabla \mathcal{J}_r\right]|| \\
&\leq \zeta' \mathbb{E}\left[\alpha_1 \,||\boldsymbol{r}|| + \alpha_2 \left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}\right|\right|\right] + \mathbb{E}_{\mathcal{I}^c}\left[||\nabla \mathcal{J}_r||\right]
\end{aligned}
$$

Derivation of gradient error indicator

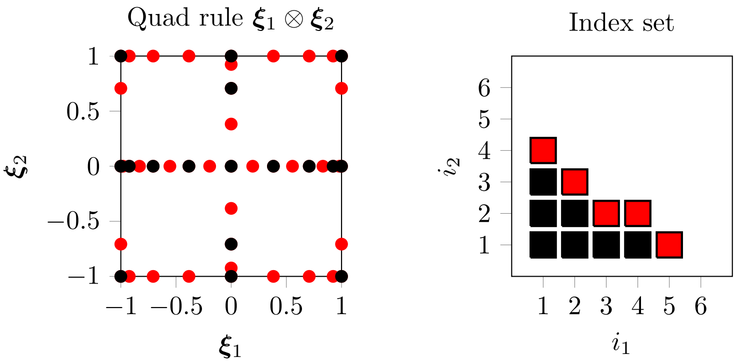
For brevity, let

$$\begin{aligned}
\mathcal{J}(\boldsymbol{\xi}) &\leftarrow \mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu},\,\boldsymbol{\xi}),\,\boldsymbol{\mu},\,\boldsymbol{\xi})\\
\nabla\mathcal{J}(\boldsymbol{\xi}) &\leftarrow \nabla\mathcal{J}(\boldsymbol{u}(\boldsymbol{\mu},\,\boldsymbol{\xi}),\,\boldsymbol{\mu},\,\boldsymbol{\xi})\\
\mathcal{J}_r(\boldsymbol{\xi}) &= \mathcal{J}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\,\boldsymbol{\xi}),\,\boldsymbol{\mu},\,\boldsymbol{\xi})\\
\nabla\mathcal{J}_r(\boldsymbol{\xi}) &= \nabla\mathcal{J}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\,\boldsymbol{\xi}),\,\boldsymbol{\mu},\,\boldsymbol{\xi})\\
\boldsymbol{r}_r(\boldsymbol{\xi}) &= \boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\,\boldsymbol{\xi}),\,\boldsymbol{\mu},\,\boldsymbol{\xi})\\
\boldsymbol{r}_r^{\boldsymbol{\lambda}}(\boldsymbol{\xi}) &= \boldsymbol{r}^{\boldsymbol{\lambda}}(\boldsymbol{\Phi}\boldsymbol{u}_r(\boldsymbol{\mu},\,\boldsymbol{\xi}),\,\boldsymbol{\Psi}\boldsymbol{\lambda}_r(\boldsymbol{\mu},\,\boldsymbol{\xi}),\,\boldsymbol{\mu},\,\boldsymbol{\xi})
\end{aligned}$$

Separate total error into contributions from ROM inexactness and SG truncation

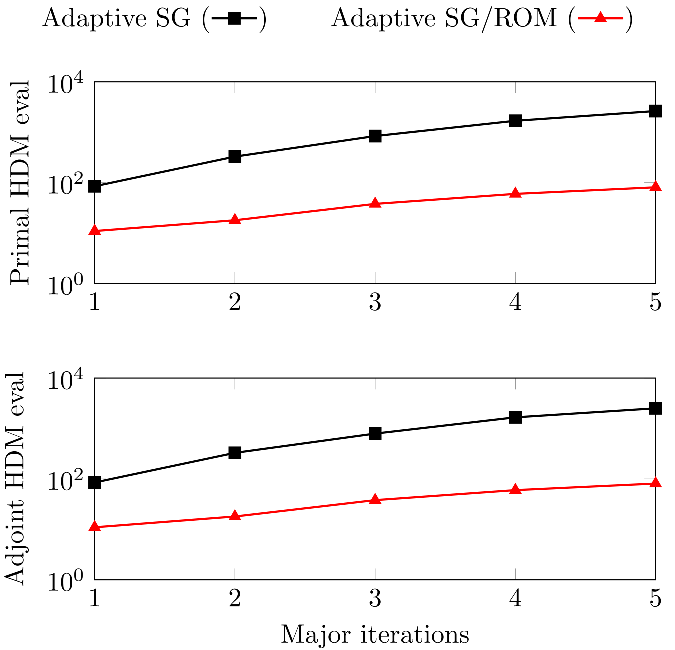
$$\begin{aligned}
||\mathbb{E}[\nabla\mathcal{J}] - \mathbb{E}_{\mathcal{I}}[\nabla\mathcal{J}_r]|| &\leq \mathbb{E}\left[||\nabla\mathcal{J} - \nabla\mathcal{J}_r||\right] + ||\mathbb{E}\left[\nabla\mathcal{J}_r\right] - \mathbb{E}_{\mathcal{I}}\left[\nabla\mathcal{J}_r\right]||\\
&\leq \zeta'\mathbb{E}\left[\alpha_1\,||\boldsymbol{r}|| + \alpha_2\,\left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}\right|\right|\right] + \mathbb{E}_{\mathcal{I}^c}\left[||\nabla\mathcal{J}_r||\right]\\
&\lesssim \zeta\left(\mathbb{E}_{\mathcal{I}\cup\mathcal{N}(\mathcal{I})}\left[\alpha_1\,||\boldsymbol{r}|| + \alpha_2\,\left|\left|\boldsymbol{r}^{\boldsymbol{\lambda}}\right|\right|\right] + \alpha_3\mathbb{E}_{\mathcal{N}(\mathcal{I})}\left[||\nabla\mathcal{J}_r||\right]\right)
\end{aligned}$$

Adaptivity: Dimension-adaptive greedy method

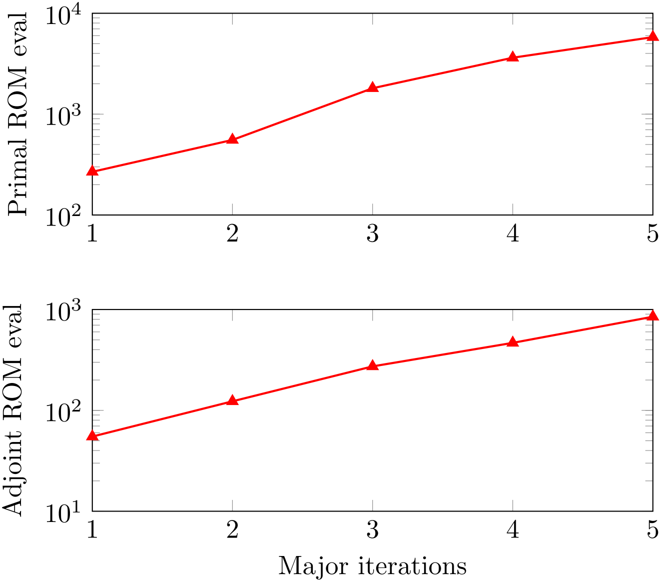


Index set ($\mathcal{I}$) – ●   Neighbors ($\mathcal{N}(\mathcal{I})$) – ●

Significant reduction in number of queries to HDM in comparison to state-of-the-art [Kouri et al., 2014]



At a price ... a large number of ROM evaluations



Extension to **time-dependent** problems

- **Applications**: inverse problems, optimal flapping flight and swimming[7] and design of helicopter blades, wind turbines, and turbomachinery
- Monolithic **space-time** formulation of reduced-order model
  - Increased speed due to natural **parallelism** in *space and time*
  - Treat as **steady state** problem in $n_{sd} + 1$ dimensions
- **Error indicators and adaptivity** algorithms in space-time setting to solve with multifidelity trust region method

Un-optimized flapping motion (left), optimal control (center), and optimal control and time-morphed geometry (right)

[7] insight into bio-locomotion, design of micro-aerial vehicles