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

Matthew J. Zahr

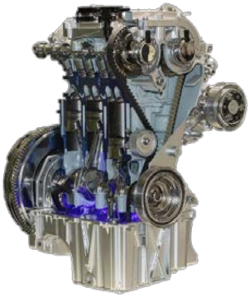
Advisor: Charbel Farhat
Computational and Mathematical Engineering
Stanford University

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

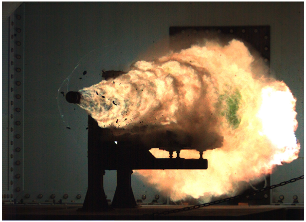
SIAM Annual Meeting
MS 137: Model Reduction of Parametrized PDEs
Boston, Massachusetts, USA
July 15, 2016

Multiphysics optimization – a key player in next-gen problems

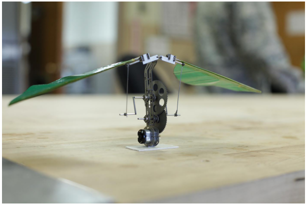
*Current interest in computational physics reaches far beyond analysis of a single configuration of a physical system into **design** (shape and topology) and **control** in an **uncertain** setting*



Engine System



EM Launcher



Micro-Aerial Vehicle

PDE-constrained optimization under uncertainty

Goal: Efficiently solve stochastic PDE-constrained optimization problems

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

- $\boldsymbol{r} : \mathbb{R}^{n_{\boldsymbol{u}}} \times \mathbb{R}^{n_{\boldsymbol{\mu}}} \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \to \mathbb{R}^{n_{\boldsymbol{u}}}$ — 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***

Proposed approach: managed two-level inexactness

Two levels of inexactness to obtain an inexpensive, approximation model

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

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

Proposed approach: managed two-level inexactness

Two levels of inexactness to obtain an inexpensive, approximation model

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

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

Manage inexactness with trust-region method

- Embedded in globally convergent **trust-region** method
- **Error indicators** to account for *both* sources of inexactness
- **Refinement** of integral approximation and reduced-order model via *dimension-adaptive* sparse grids and a *greedy method* over collocation nodes

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

The connection between the objective function and model

- 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 error indicator: φ_k

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

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

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

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

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

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

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

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

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

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

Inexact objective function evaluations

- Asymptotic objective decrease bound [Kouri et al., 2014]

$$|F(\boldsymbol{\mu}_k) - F(\hat{\boldsymbol{\mu}}_k) + \psi_k(\hat{\boldsymbol{\mu}}_k) - \psi_k(\boldsymbol{\mu}_k)| \leq \sigma \min\{m_k(\boldsymbol{\mu}_k) - m_k(\hat{\boldsymbol{\mu}}_k),\, r_k\}^{1/\omega}$$

  where $\omega \in (0,\, 1)$, $r_k \to 0$, $\sigma > 0$

Asymptotic objective decrease bound permits the use of error indicator: θ_k

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

Trust region method ingredients for global convergence

Approximation models

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

Error indicators

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

Adaptivity

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

Global convergence

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

First layer 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 layer of inexactness: reduced-order models

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

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

$$\Downarrow$$

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

$$\Downarrow$$

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

First two ingredients for global convergence

Approximation models built on two levels of inexactness

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

Error indicators that account for both sources of error

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

Reduced-order model errors

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

Sparse grid truncation errors

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

Derivation of gradient error indicator

For brevity, let

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

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

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

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

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

Separate total error into contributions from ROM inexactness and SG truncation

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

Final requirement for convergence: Adaptivity

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

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

The final requirement for convergence is to construct the sparse grid $\mathcal{I}_k$ and reduced-order basis $\boldsymbol{\Phi}_k$ such that the gradient condition hold

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

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

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

Adaptivity: Dimension-adaptive greedy method

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

Refine index set: Dimension-adaptive sparse grids

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

Adaptivity: Dimension-adaptive greedy method

$$\textbf{while} \quad \mathcal{E}_4(\boldsymbol{\Phi}, \mathcal{I}, \boldsymbol{\mu}_k) > \frac{\kappa}{3\alpha_3} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\} \ \textbf{do}$$

Refine index set: Dimension-adaptive sparse grids

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

Refine reduced-order basis: Greedy sampling

$$\textbf{while} \quad \mathcal{E}_1(\boldsymbol{\Phi}, \mathcal{I}, \boldsymbol{\mu}_k) > \frac{\kappa}{3\alpha_1} \min\{||\nabla m_k(\boldsymbol{\mu}_k)||\,,\, \Delta_k\} \ \textbf{do}$$

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

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

end while

Adaptivity: Dimension-adaptive greedy method

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

Refine index set: Dimension-adaptive sparse grids

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

Refine reduced-order basis: Greedy sampling

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

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

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

end while

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

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

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

end while

end while

Optimal control of steady Burgers' equation

- **Optimization problem:**

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

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

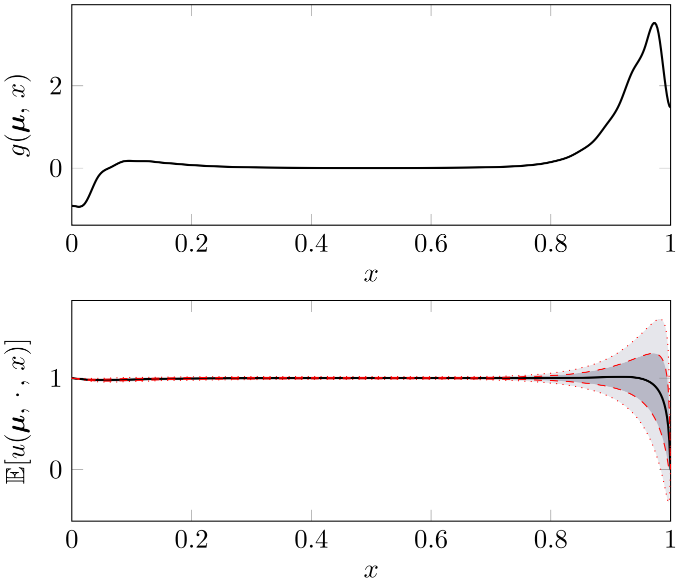
$$-\nu(\boldsymbol{\xi})\partial_{xx}u(\boldsymbol{\mu},\, \boldsymbol{\xi},\, x) + u(\boldsymbol{\mu},\, \boldsymbol{\xi},\, x)\partial_x u(\boldsymbol{\mu},\, \boldsymbol{\xi},\, x) = z(\boldsymbol{\mu},\, x) \quad x \in (0,\, 1), \quad \boldsymbol{\xi} \in \boldsymbol{\Xi}$$
$$u(\boldsymbol{\mu},\,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

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

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

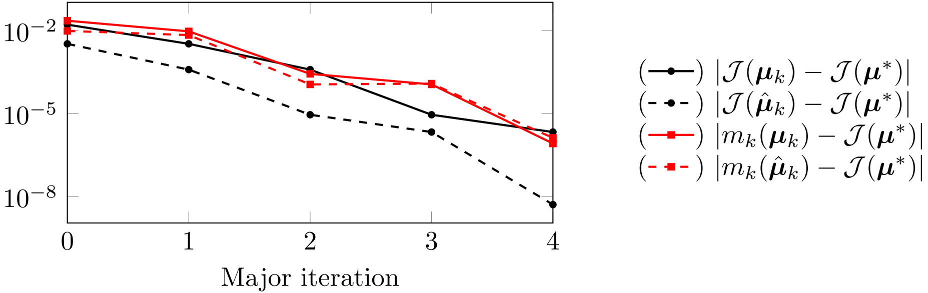
- **Parametrization:** $z(\boldsymbol{\mu},\, x)$ – cubic splines with 51 knots, $n_{\boldsymbol{\mu}} = 53$

Optimal control and statistics



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

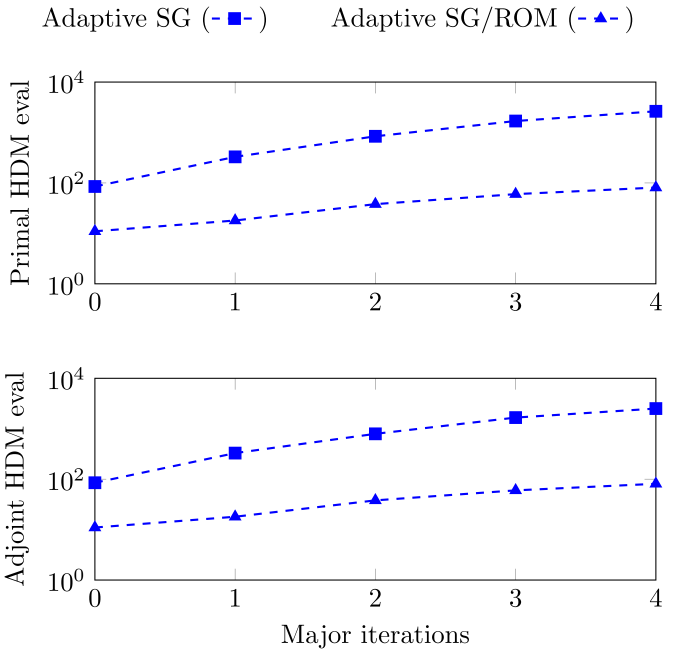
Global convergence without pointwise agreement



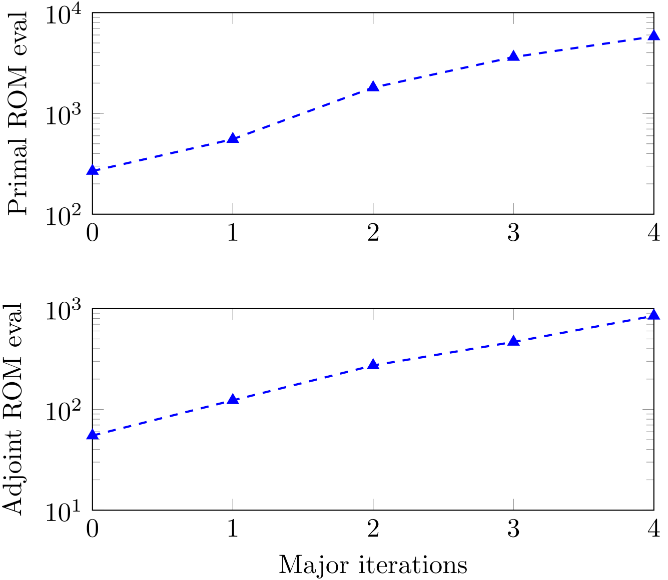
$\mathcal{J}(\boldsymbol{\mu}_k)$	$m_k(\boldsymbol{\mu}_k)$	$\mathcal{J}(\hat{\boldsymbol{\mu}}_k)$	$m_k(\hat{\boldsymbol{\mu}}_k)$	$\lvert\lvert\nabla\mathcal{J}(\boldsymbol{\mu}_k)\rvert\rvert$	ρ_k	Success?
6.6506e-02	7.2694e-02	5.3655e-02	5.9922e-02	2.2959e-02	1.0257e+00	1.0000e+00
5.3655e-02	5.9593e-02	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 number of queries to HDM in comparison to state-of-the-art [Kouri et al., 2014]

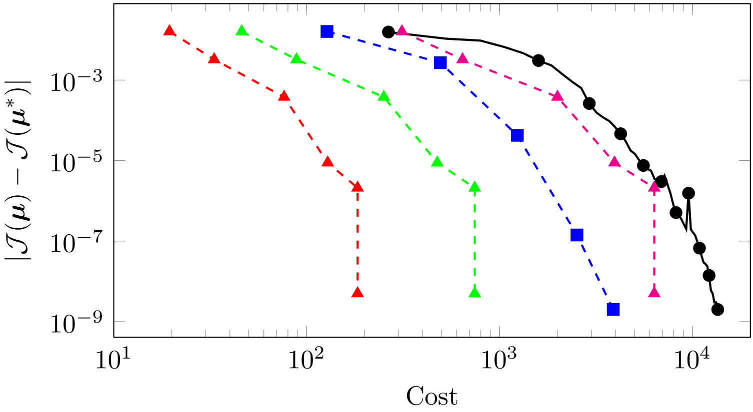


At a price ... a large number of ROM evaluations



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 method of [Kouri et al., 2014] (- ■ -), and proposed ROM/SG for $\tau = 1$ (- ▲ -), $\tau = 10$ (- ▲ -), $\tau = 100$ (- ▲ -)

Leveraging and managing two-levels of inexactness for efficient stochastic PDE-constrained optimization

Summary

- Two-level approximation of moments of quantities of interest of SPDE
  - *Anisotropic sparse grids* - inexact integration
  - *Reduced-order models* - inexact evaluations
- Two-level inexactness managed through trust-region method
- Significant decrease in number of HDM queries vs. state-of-the-art

Future work

- Incorporate nonlinear constraints
- Local reduced-order models for improved efficiency

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.

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.

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.