SAND2022-12342C

Sandia National Laboratories

Model Discrepancy Sensitivities to Enable Decision-Making

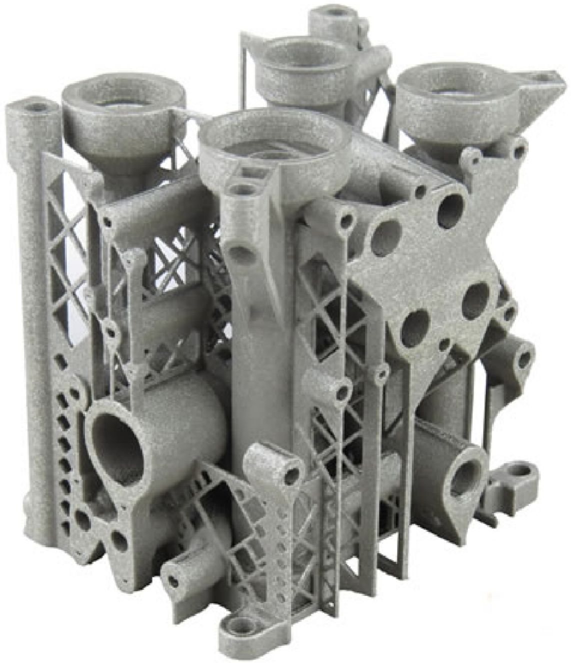
Joseph Hart and Bart van Bloemen Waanders

Sandia National Laboratories

SIAM Mathematics of Data Science, September 27, 2022

U.S. DEPARTMENT OF ENERGY NNSA

All models are wrong...



...but some are useful (George Box 76)

- Computational models support decision making when:
  - ✓ the models are computationally efficient enough
  - ✓ uncertainty may be accounted for and propagated through the analysis
- Many models of complex systems do not meet these criteria
- Data science supports the development of computationally efficient models, but may introduce additional errors and uncertainty.

Optimization of Approximate Models

$$\min_{z \in \mathcal{Z}} J(\tilde{S}(z), z)$$

- J is the objective
- z is a design, control, or inversion parameter
- $\tilde{S}(z)$ is an approximate model

Our goals are:

- Use the limited high-fidelity evaluations to improve the solution
- Characterize uncertainty in the optimal solution due to $S - \tilde{S}$

Learning Optimal Solution Updates

Approximate Optimal Solution

$$\tilde{z}(\theta) = \underset{z}{\text{argmin}} \quad J(\tilde{S}(z), z)$$

High-fidelity Data

$$S(z_l) - \tilde{S}(z_l)$$

Learning Optimal Solution Updates

Discrepancy Parameterization

$$\delta(z,\theta) \approx S(z) - \tilde{S}(z)$$

Approximate Optimal Solution

$$\tilde{z}(\theta) = \underset{z}{\text{argmin}}\ J(\tilde{S}(z) + \delta(z,\theta), z)$$

Post-Optimality Sensitivity

$$\frac{\partial \tilde{z}}{\partial \theta}\Delta\theta = -H^{-1}B\Delta\theta$$

Optimal Solution Posterior

$$\Theta \sim \pi_{post} \Rightarrow z^* \approx \frac{\partial \tilde{z}}{\partial \theta}\Theta$$

High-fidelity Data

$$S(z_l) - \tilde{S}(z_l)$$

Bayesian Inversion

$$\pi_{post} \propto \pi_{likelihood}\,\pi_{prior}$$

Prior Discrepancy

π_{prior} = Laws, length scales, etc.

Illustrative Example

$$\min_z \frac{1}{2}\int_0^1 (\tilde{S}(z) - T(x))^2 dx + \frac{\beta}{2}\int_0^1 z\mathcal{E}z$$

where $\tilde{S}(z)$ is the solution operator for

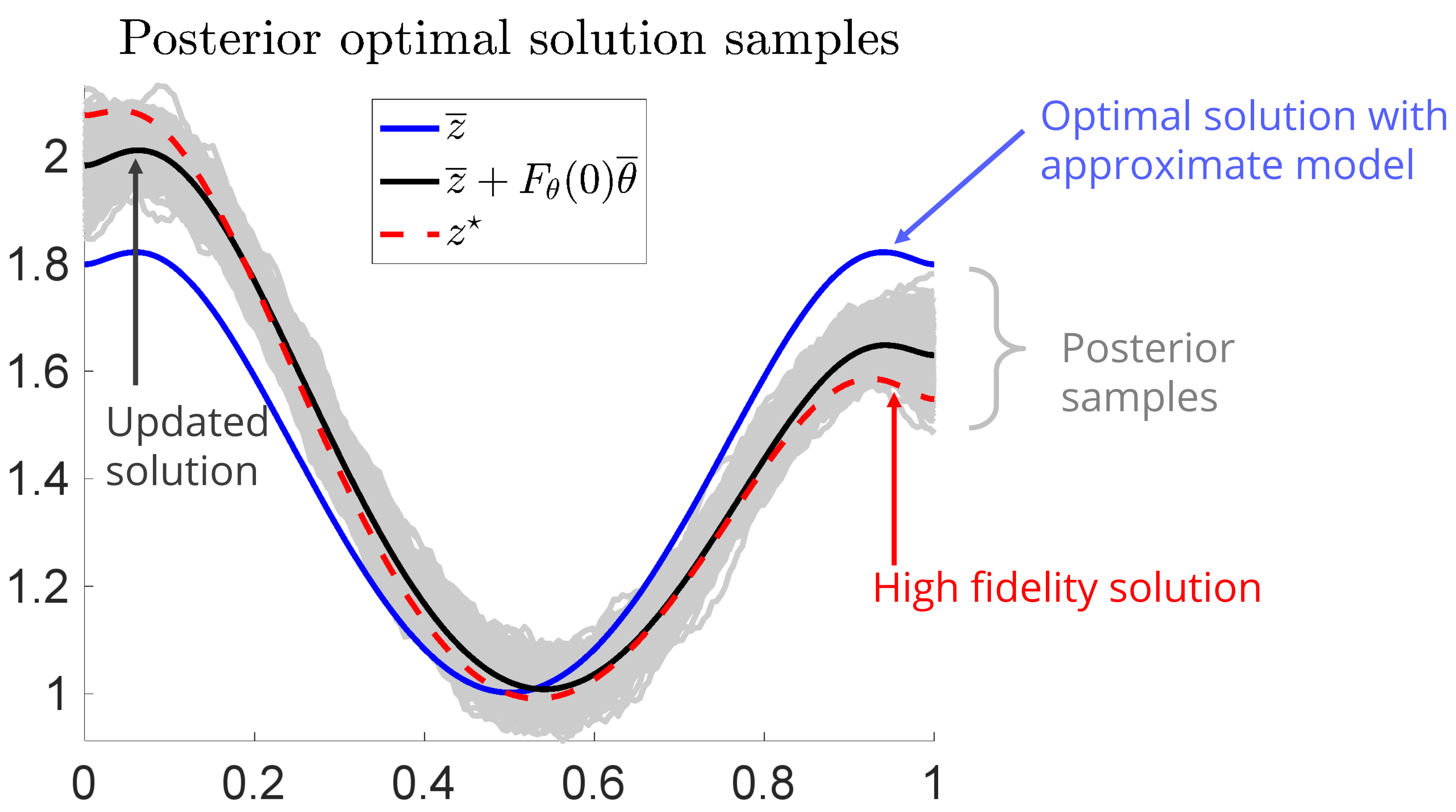
$$\begin{aligned} -\kappa u'' &= z & &\text{on } (0,1) \\ \kappa u' &= hu & &\text{on } \{0,1\} \end{aligned}$$

The high-fidelity model S solves

$$\begin{aligned} -\kappa u'' + vu' &= z & &\text{on } (0,1) \\ \kappa u' &= hu & &\text{on } \{0,1\} \end{aligned}$$

Given the high-fidelity solution $S(z)$ for 2 different source terms, improve and characterize uncertainty in the low-fidelity optimal source.

Optimal Solution Posterior



Learning Optimal Solution Updates

Discrepancy Parameterization

$$\delta(z,\theta) \approx S(z) - \tilde{S}(z)$$

Approximate Optimal Solution

$$\tilde{z}(\theta) = \underset{z}{\text{argmin}}\, J(\tilde{S}(z) + \delta(z,\theta), z)$$

Post-Optimality Sensitivity

$$\frac{\partial \tilde{z}}{\partial \theta}\Delta\theta = -H^{-1}B\Delta\theta$$

Optimal Solution Posterior

$$\Theta \sim \pi_{post} \Longrightarrow z^{*} \approx \frac{\partial \tilde{z}}{\partial \theta}\Theta$$

High-fidelity Data

$$S(z_l) - \tilde{S}(z_l)$$

Bayesian Inversion

$$\pi_{post} \propto \pi_{likelihood}\,\pi_{prior}$$

Prior Discrepancy

π_{prior} = Laws, length scales, etc.

Model Discrepancy Representation

- General form for a (discretized) operator

$$\sum_{i=1}^{m} f_i(z)\phi_i$$

- Since post-optimality analysis only depends on the mixed (z,θ) derivative, assume f_i's are linear, Reisz representation yields

$$\sum_{i=1}^{m} \left(\theta_{i,0} + \sum_{j=1}^{n} \theta_{i,j} (z, \psi_j)_{\mathcal{Z}} \right) \phi_i$$

$$\downarrow$$

$$\delta(\mathbf{z}, \theta) = \left(\mathbf{I}_m \quad \mathbf{I}_m \otimes \mathbf{z}^T \mathbf{M}_z \right) \theta$$

- Discretized $\delta : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$ is parameterized by $\theta \in \mathbb{R}^p$
- $p = m(n+1)$ so the dimension of θ may be $\mathcal{O}(\text{mesh size}^2)$
- Evaluate $\delta(z,\theta)$ efficiently using Kronecker product
- $(M_z)_{i,j} = (\psi_i, \psi_j)_Z$ - mass matrix that defines the inner product on $\mathcal{Z}_h$

Learning Optimal Solution Updates

Discrepancy Parameterization

$\delta(z,\theta) \approx S(z) - \tilde{S}(z)$

Approximate Optimal Solution

$\tilde{z}(\theta) = \underset{z}{\text{argmin}}\, J(\tilde{S}(z) + \delta(z,\theta), z)$

High-fidelity Data

$S(z_l) - \tilde{S}(z_l)$

Post-Optimality Sensitivity

$\frac{\partial \tilde{z}}{\partial \theta}\Delta\theta = -H^{-1}B\Delta\theta$

Bayesian Inversion

$\pi_{post} \propto \pi_{likelihood}\, \pi_{prior}$

Prior Discrepancy

π_{prior} = Laws, length scales, etc.

Optimal Solution Posterior

$\Theta \sim \pi_{post} \Rightarrow z^* \approx \frac{\partial \tilde{z}}{\partial \theta}\Theta$

Post-optimality Sensitivities

$$\min_{\boldsymbol{z}} \mathbf{J}(\tilde{\mathbf{S}}(\boldsymbol{z}) + \boldsymbol{\delta}(\boldsymbol{z}, \theta), \boldsymbol{z}) \quad (1)$$

- $\tilde{\mathbf{z}}^\star$ solves (1) when $\delta(\mathbf{z}, \theta_0) = \mathbf{0}$, the problem solved in practice
- Under mild assumptions, applying the Implicit Function Theorem to

$$\nabla J(\tilde{\mathbf{z}}^\star, \theta_0) = \mathbf{0}$$

gives

$$\mathcal{F} : \mathcal{N}(\theta_0) \to \mathcal{N}(\tilde{\mathbf{z}}^\star)$$

such that $\mathcal{F}(\theta_0)$ solves (1) when $\theta = \theta_0$ and

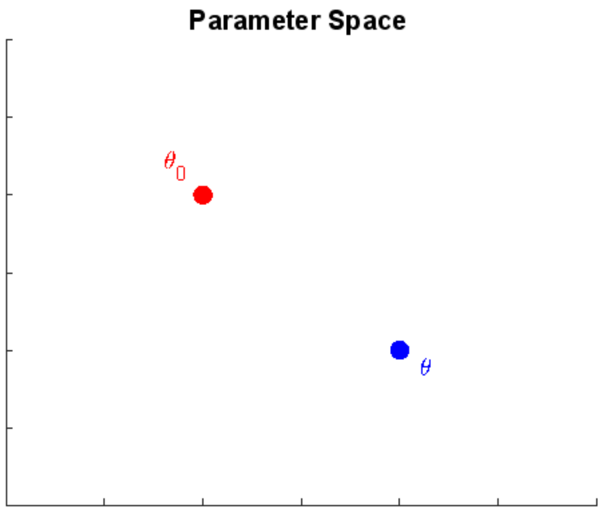
$$\mathcal{F}'_\theta(\theta_0) = -\mathcal{H}^{-1}\mathcal{B}$$

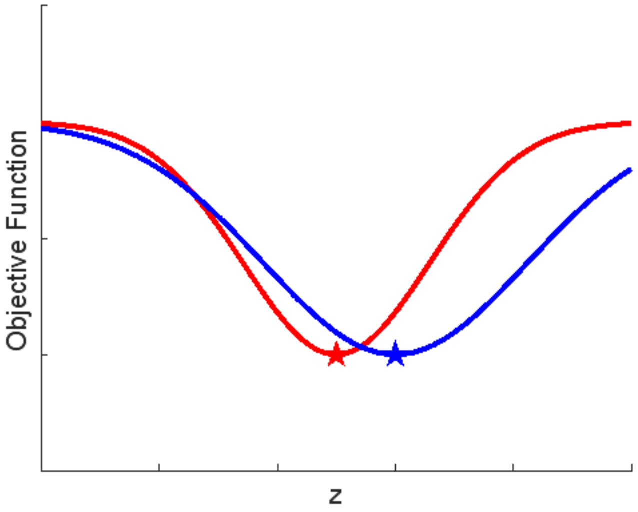
is the sensitivity of the optimal solution with respect to model discrepancy

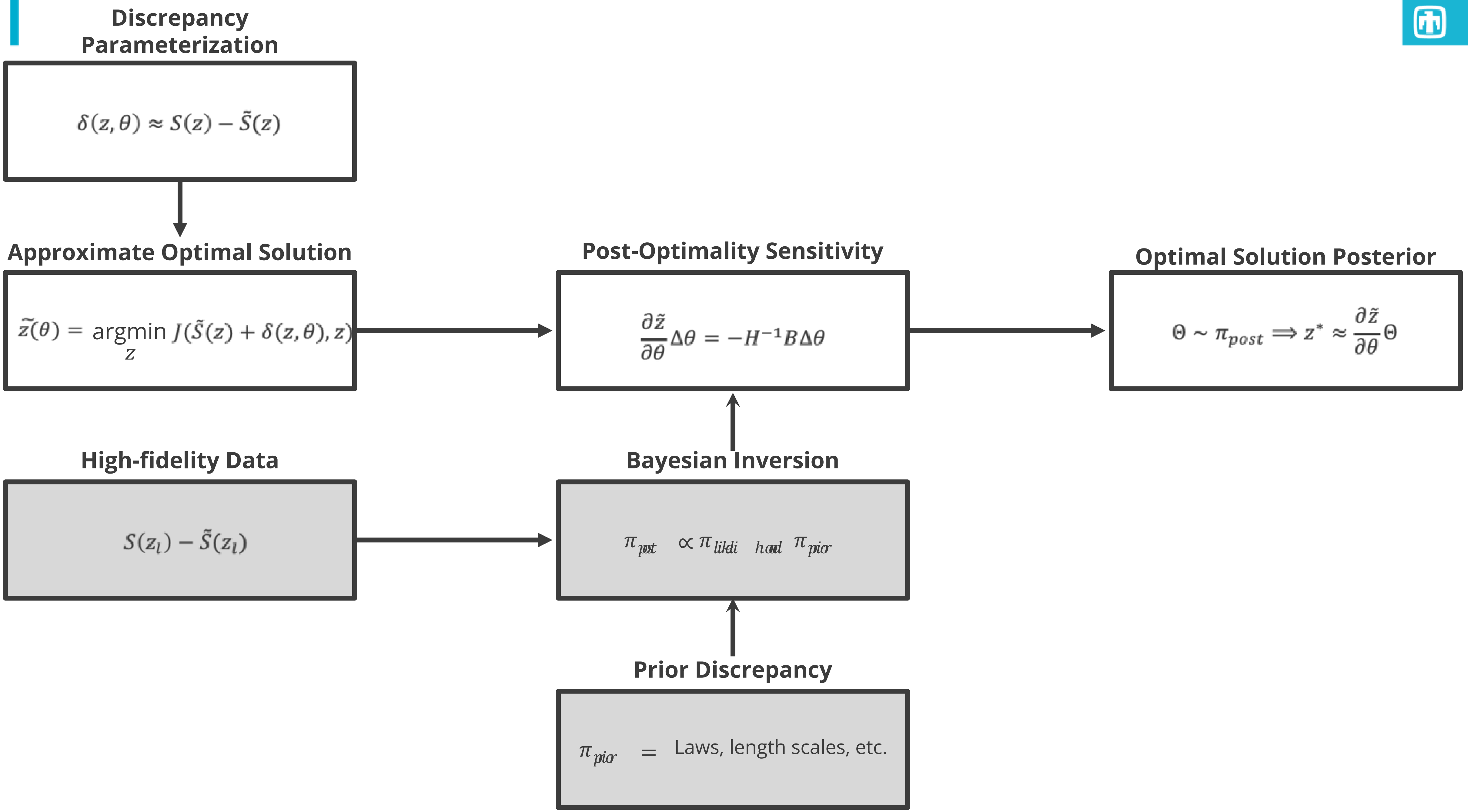
Post-optimality Sensitivities

$$\mathcal{F}'_\theta(\theta_0) = -\mathcal{H}^{-1}\mathcal{B}$$

- $\mathcal{H}$ is the Hessian of the objective function with respect to z
- $\mathcal{B}$ is the mixed second derivative of the objective with respect to z and θ
- Acts like a Newton step to update the optimal solution after a perturbation of the model discrepancy






Discrepancy Parameterization
$\delta(z,\theta) \approx S(z) - \tilde{S}(z)$
Approximate Optimal Solution
$\tilde{z}(\theta) = \underset{z}{\text{argmin}}\, J(\tilde{S}(z) + \delta(z,\theta), z)$
Post-Optimality Sensitivity
$\frac{\partial \tilde{z}}{\partial \theta}\Delta\theta = -H^{-1}B\Delta\theta$
Optimal Solution Posterior
$\Theta \sim \pi_{post} \Rightarrow z^* \approx \frac{\partial \tilde{z}}{\partial \theta}\Theta$
High-fidelity Data
$S(z_l) - \tilde{S}(z_l)$
Bayesian Inversion
$\pi_{post} \propto \pi_{likelihood}\, \pi_{prior}$
Prior Discrepancy
π_{prior} = Laws, length scales, etc.

Bayesian Inverse Problem - Prior Discrepancy

- Measure size of $\boldsymbol{\delta}$:

$$||\boldsymbol{\delta}(\boldsymbol{z},\boldsymbol{\theta})||_{\boldsymbol{L}}^2 = \boldsymbol{\theta}^T \begin{pmatrix} \boldsymbol{L} & \boldsymbol{L} \otimes \boldsymbol{z}^T \boldsymbol{M}_z \\ \boldsymbol{L} \otimes \boldsymbol{M}_z \boldsymbol{z} & \boldsymbol{L} \otimes \boldsymbol{M}_z \boldsymbol{z}\boldsymbol{z}^T \boldsymbol{M}_z \end{pmatrix} \boldsymbol{\theta}$$

- Marginalize out $\boldsymbol{z}$:

$$\mathbb{E}_{\boldsymbol{z}}[||\boldsymbol{\delta}(\boldsymbol{z},\boldsymbol{\theta})||_{\boldsymbol{L}}^2] = \boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta}$$

where

$$\boldsymbol{M}_{\boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{L} & \boldsymbol{L} \otimes \overline{\boldsymbol{z}}^T \boldsymbol{M}_z \\ \boldsymbol{L} \otimes \boldsymbol{M}_z \overline{\boldsymbol{z}} & \boldsymbol{L} \otimes \boldsymbol{E} \end{pmatrix}$$

- $\boldsymbol{L}$ encodes known physics of the discrepancy - in our case a Laplacian like operator and $\boldsymbol{L}^{-1}$ represents the prior covariance
- $\boldsymbol{\Gamma}$ is a covariance matrix on the control space $\mathcal{Z}$
- Hence $\boldsymbol{M}_{\boldsymbol{\theta}}$ defines an inner product for $\boldsymbol{\theta}$ to measure the size of the model discrepancy $\boldsymbol{\delta}(\boldsymbol{z},\boldsymbol{\theta})$ according to our prior knowledge imposed in $\boldsymbol{L}$ and $\boldsymbol{\Gamma}$

Bayesian Inverse Problem - notation

- for notational simplicity, we define

$$\mathbf{A}_\ell = \left(\mathbf{I}_m \quad \mathbf{I}_m \otimes \mathbf{z}_\ell^T \mathbf{M}_z\right) \in \mathbb{R}^{m \times p}, \qquad \ell = 1, 2, \ldots, N,$$

  so that $\delta(\mathbf{z}_\ell, \theta) = \mathbf{A}_\ell \theta$, and the concatenation of these matrices

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_N \end{pmatrix} \in \mathbb{R}^{mN \times p}$$

  so that $\mathbf{A}\theta \in \mathbb{R}^{mN}$ corresponds to evaluating $\delta(\mathbf{z}, \theta)$ for the inputs $\mathbf{z}_\ell$.

- let $\mathbf{b} \in \mathbb{R}^{mN}$ be defined by stacking $\mathbf{y}_\ell = \mathbf{S}(\mathbf{z}_\ell) - \tilde{\mathbf{S}}(\mathbf{z}_\ell)$, $\ell = 1, 2, \ldots, N$, into a vector so that we seek

$$\mathbf{A}\theta \approx \mathbf{b}$$

- infinite number of θ directions because the problem is underdetermined

Bayesian Inversion Problem

- Arrange data and discrepancy representation so that we seek θ such that

$$\mathbf{A}\theta \approx \mathbf{b}$$

- Given Gaussian prior and noise models, linearity of $\delta(\mathbf{z}, \theta)$ in θ, the posterior is Gaussian with a negative log probability density function

$$\frac{1}{2\alpha}\left(\mathbf{A}\theta - \mathbf{b}\right)^T\left(\mathbf{A}\theta - \mathbf{b}\right) + \frac{1}{2}\theta^T\mathbf{M}_\theta\theta.$$

- α balances the dependence of prior and data misfit
- The posterior mean is

$$\bar{\theta} = \frac{1}{\alpha}\mathbf{\Sigma}\mathbf{A}^T\mathbf{b}$$

  and the posterior covariance is

$$\mathbf{\Sigma} = \left(\mathbf{M}_\theta + \frac{1}{\alpha}\mathbf{A}^T\mathbf{A}\right)^{-1}.$$

Bayesian Inversion Problem – Enabling Sampling

- The goal is to sample from a Gaussian distribution which may be generated by multiplying a factor of the covariance matrix with a standard normal random vector and adding the mean

- But how do we invert the sum?

$$\mathbf{\Sigma} = \left(\mathbf{M}_\theta + \frac{1}{\alpha}\mathbf{A}^T\mathbf{A}\right)^{-1}$$

1. Factorize $\mathbf{A}$ to rewrite $\mathbf{M}_\theta + \frac{1}{\alpha}\mathbf{A}^T\mathbf{A}$
2. Invert $\mathbf{M}_\theta + \frac{1}{\alpha}\mathbf{A}^T\mathbf{A}$
3. Factorize $\mathbf{\Sigma}$
4. Compute matrix-vector products for posterior samples

Bayesian Inversion Problem – 4 step process

1. Factorize $\mathbf{A}$ to rewrite $\mathbf{M}_\theta + \frac{1}{\alpha}\mathbf{A}^T\mathbf{A}$

- compute GSVD of $\mathbf{A}$ in the $\mathbf{M}_\theta = \mathbf{C}\mathbf{C}^T$ inner product
- exploit Kronecker structure
- factor out $\mathbf{M}_\theta$ to enable inversion: $\boldsymbol{\Sigma}^{-1} = \frac{1}{\alpha}\mathbf{C}\mathbf{X}\mathbf{C}^T$
- where

$$\mathbf{X} = \alpha\mathbf{I} + \mathbf{C}^T\boldsymbol{\Psi}\boldsymbol{\Phi}^2\boldsymbol{\Psi}^T\mathbf{C}.$$

conveniently consists of identity, orthogonal, and diagonal matrices

Bayesian Inversion Problem – 4 step process

2. Invert $\boldsymbol{\Sigma}$

- applying Sherman-Morrison-Woodbury to $\mathbf{X}$ yields

$$\boldsymbol{\Sigma} = \mathbf{M}_{\theta}^{-1} - \boldsymbol{\Psi}\mathbf{P}\boldsymbol{\Psi}^T.$$

- where $\mathbf{P} \in \mathbb{R}^{mN \times mN}$ is a diagonal matrix whose entries are given by $\frac{\lambda_i}{\lambda_i + \alpha\rho_j}$, a combination of the eigenvalues from $\mathbf{L}$ and $\mathbf{G}$ (object from computing GSVD of $\mathbf{A}$), along with the noise covariance α

- enables calculation of $\overline{\theta} = \frac{1}{\alpha}\boldsymbol{\Sigma}\mathbf{A}^T\mathbf{b}$

Bayesian Inversion Problem – 4 step process

3. Factorize $\boldsymbol{\Sigma}$

- factorization is needed to sample Gaussian
- compute eigendecomposition of $\mathbf{X} = \alpha\mathbf{I} + \mathbf{C}^T\boldsymbol{\Psi}\boldsymbol{\Phi}^2\boldsymbol{\Psi}^T\mathbf{C}$

$$\mathbf{X} = \mathbf{Q}\boldsymbol{\Upsilon}\mathbf{Q}^T.$$

- $\mathbf{Q}$ is an orthonormal matrix with collected eigenvectors

$$\boldsymbol{\Sigma} = \alpha\mathbf{C}^{-T}\mathbf{X}^{-1}\mathbf{C}^{-1} = \mathbf{T}\mathbf{T}^T$$
$$\mathbf{T} = \sqrt{\alpha}\mathbf{C}^{-T}\mathbf{Q}\boldsymbol{\Upsilon}^{-\frac{1}{2}}.$$

- this facilitates efficient sampling

Bayesian Inversion Problem – 4 step process

4. Compute matrix-vector products for posterior samples

- computing matrix-vector products with Σ and $\mathbf{T}$ gives a sum over eigen-pairs of an operator which scales with the mesh resolution
- rewrite the sum in terms of linear solves in the direction of the incoming vectors
- this gives expressions for posterior samples which scale with the number of high-fidelity data compute posterior samples as

$$\mathbf{T}\omega = \mathbf{T}\left(\mathbf{I}_{n+1} \otimes \mathbf{V}^{\mathrm{L}}\right)^T \nu = \hat{\theta} + \tilde{\theta}$$

where

$$\hat{\theta} = \sqrt{\alpha} \sum_{i=1}^{N} \frac{1}{\sqrt{\lambda_i}} \begin{pmatrix} s_i \hat{\mathbf{u}}_i \\ \hat{\mathbf{u}}_i \otimes \mathbf{M}_{\mathbf{z}}^{-1} \mathbf{\Gamma}^{-1} \mathbf{w}_i \end{pmatrix} \quad \text{and} \quad \tilde{\theta} = \sum_{k=1}^{n-N+1} \begin{pmatrix} \tilde{s}_k \tilde{\mathbf{u}}_k \\ \tilde{\mathbf{u}}_k \otimes \tilde{\mathbf{w}}_k \end{pmatrix}$$

with

$$\hat{\mathbf{u}}_i = (\alpha \mathbf{L} + \lambda_i \mathbf{I})^{-\frac{1}{2}} \nu_i \qquad \tilde{s}_k = -\overline{\mathbf{z}}^T \mathbf{\Gamma}^{-\frac{1}{2}} \tilde{\mathbf{z}}_k$$

$$\tilde{\mathbf{u}}_k = \mathbf{L}^{-\frac{1}{2}} \nu_{N+k} \qquad \tilde{\mathbf{w}}_k = \mathbf{M}_{\mathbf{z}}^{-1} \mathbf{\Gamma}^{-\frac{1}{2}} \tilde{\mathbf{z}}_k$$

Posterior Samples for Discrepancy

- Posterior samples take the form

$$\overline{\theta} + \hat{\theta} + \tilde{\theta}$$

where the mean is

$$\overline{\theta} = \frac{1}{\alpha} \sum_{\ell=1}^{N} \left[\begin{pmatrix} a_\ell u_\ell \\ u_\ell \otimes \mathbf{M}_z^{-1} \mathbf{\Gamma}^{-1} (z_\ell - \overline{z}) \end{pmatrix} - \sum_{i=1}^{N} b_{i,\ell} \begin{pmatrix} s_i \boldsymbol{u}_{i,\ell} \\ \boldsymbol{u}_{i,\ell} \otimes \mathbf{M}_z^{-1} \mathbf{\Gamma}^{-1} \boldsymbol{w}_i \end{pmatrix} \right]$$

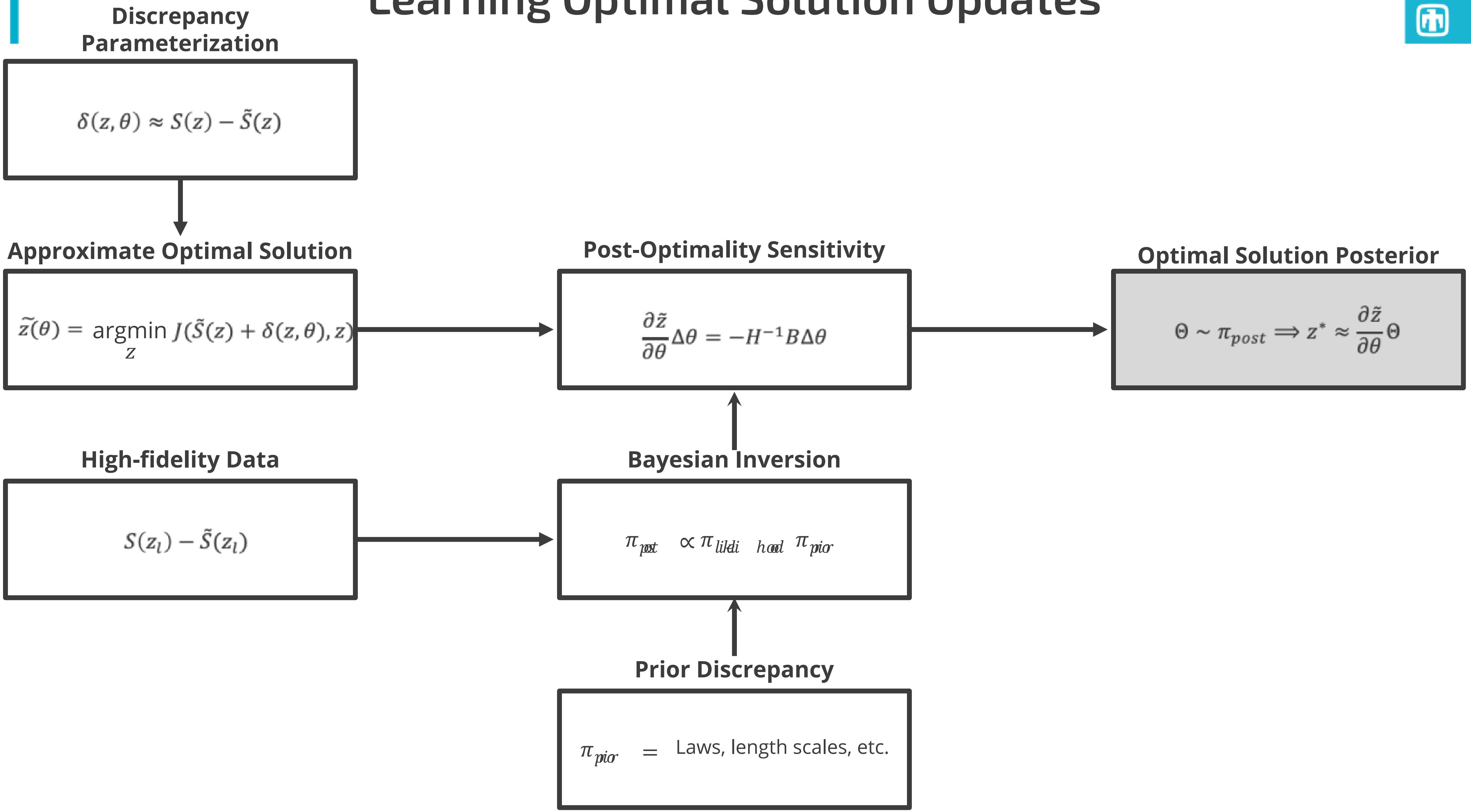
uncertainty in the data informed directions is

$$\hat{\theta} = \sqrt{\alpha} \sum_{i=1}^{N} \frac{1}{\sqrt{\lambda_i}} \begin{pmatrix} s_i \hat{u}_i \\ \hat{u}_i \otimes \mathbf{M}_z^{-1} \mathbf{\Gamma}^{-1} \boldsymbol{w}_i \end{pmatrix}$$

and uncertainty in the data uninformed directions is

$$\tilde{\theta} = \sum_{k=1}^{n-N+1} \begin{pmatrix} \tilde{s}_k \tilde{\boldsymbol{u}}_k \\ \tilde{u}_k \otimes \tilde{\boldsymbol{w}}_k \end{pmatrix}$$

Learning Optimal Solution Updates



Propagating Samples Through Post-optimality Sensitivities

Sample

$$\mathcal{F}'_\theta(\theta_0)(\overline{\theta} + \hat{\theta} + \tilde{\theta}) = -\mathcal{H}^{-1}(\mathcal{B}\overline{\theta} + \mathcal{B}\hat{\theta} + \mathcal{B}\tilde{\theta})$$

$$\mathbf{B}\overline{\theta} = \frac{1}{\alpha}\tilde{\mathbf{S}}_z^T \nabla_{u,u}\mathbf{J}\left[\sum_{\ell=1}^{N}\left(u_\ell - \sum_{i=1}^{N} b_{i,\ell}(e^T \boldsymbol{g}_i)\boldsymbol{u}_{i,\ell}\right)\right] + \frac{1}{\alpha}\sum_{\ell=1}^{N}(\nabla_u J u_\ell)\mathbf{\Gamma}^{-1}(z_\ell - \overline{z})$$
$$- \frac{1}{\alpha}\sum_{\ell=1}^{N}\sum_{i=1}^{N} b_{i,\ell}(\nabla_u \mathbf{J}\boldsymbol{u}_{i,\ell})\mathbf{\Gamma}^{-1}\boldsymbol{w}_i$$

$$\mathbf{B}\hat{\theta} = \sqrt{\alpha}\tilde{\mathbf{S}}_z^T \nabla_{u,u}\mathbf{J}\left(\sum_{i=1}^{N}\frac{e^T \boldsymbol{g}_i}{\sqrt{\lambda_i}}\hat{u}_i\right) + \sqrt{\alpha}\sum_{i=1}^{N}\frac{\nabla_u \mathbf{J}\hat{u}_i}{\sqrt{\lambda_i}}\mathbf{\Gamma}^{-1}\boldsymbol{w}_i$$

and

$$\mathbf{B}\tilde{\theta} = \sum_{k=1}^{n-N+1}(\nabla_u \mathbf{J}\tilde{u}_k)\,\mathbf{\Gamma}^{-\frac{1}{2}}\tilde{z}_k.$$

A Fluid Flow Example

Optimal design of a flow controller

$$\min_z \frac{1}{2} \int_\chi \mathbf{v}_y(z)^2 + \frac{\beta}{2} \int_\Omega ||\mathbf{z}||^2$$

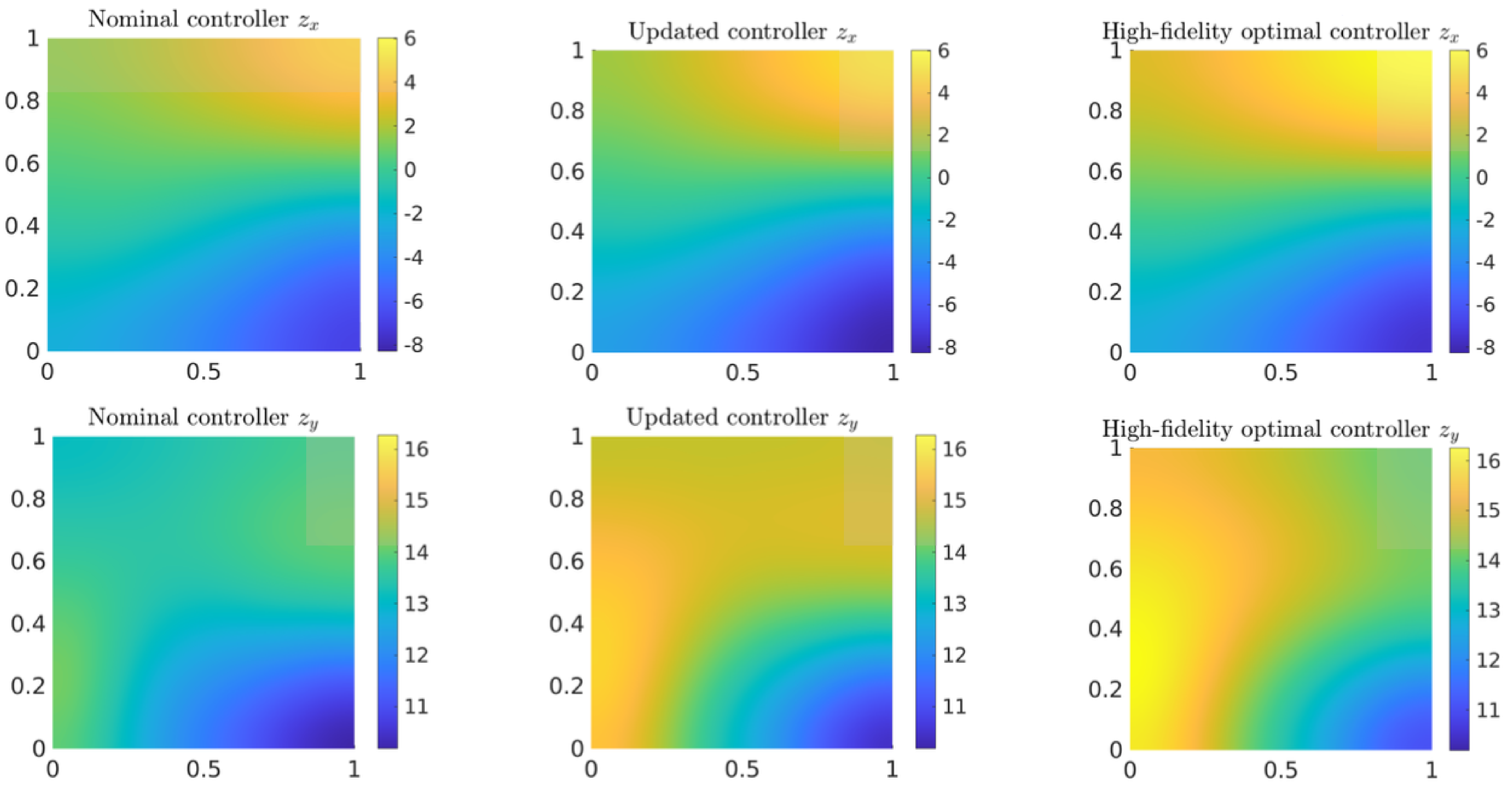
constrained by the Stokes equations

$$\begin{aligned} -\mu \nabla \mathbf{v} + \nabla p &= \mathbf{g} + \mathbf{z} && \text{on } \Omega \\ \nabla \cdot \mathbf{v} &= 0 && \text{on } \Omega \end{aligned}$$

as a simplification of the Navier-Stokes equations

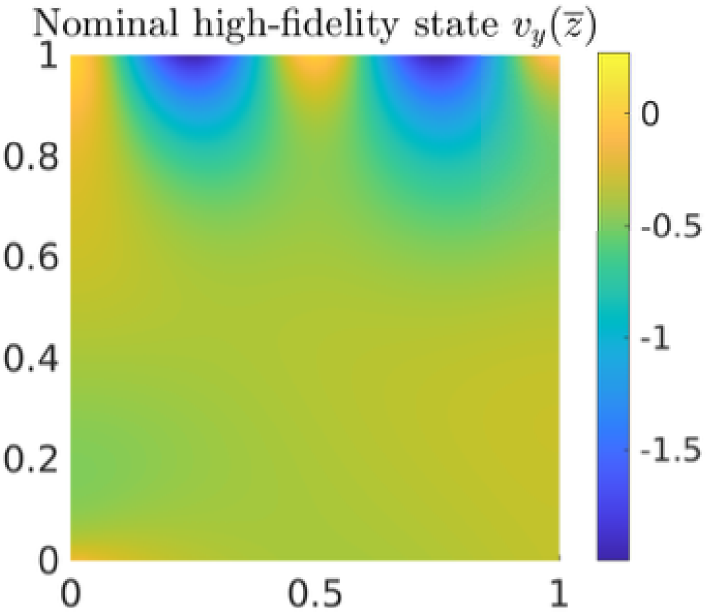
$$\begin{aligned} -\mu \nabla \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} + \nabla p &= \mathbf{g} + \mathbf{z} && \text{on } \Omega \\ \nabla \cdot \mathbf{v} &= 0 && \text{on } \Omega \end{aligned}$$

Comparison of Controllers

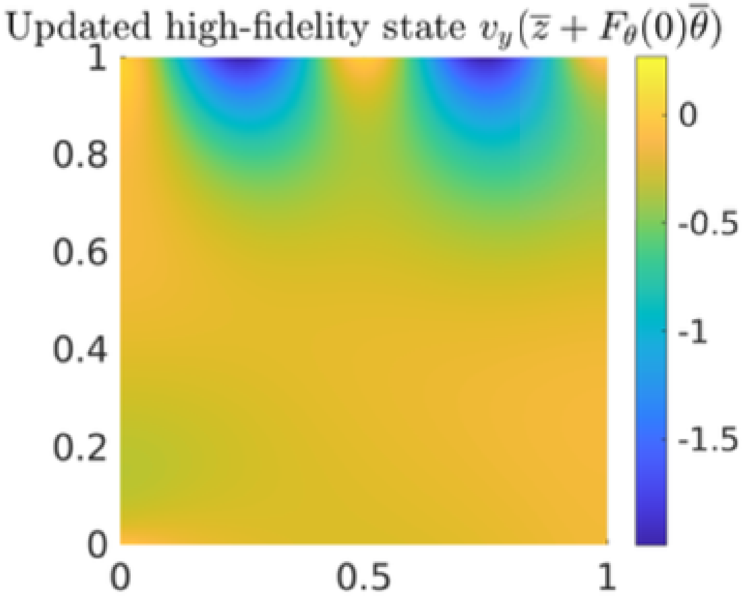


Comparison of States

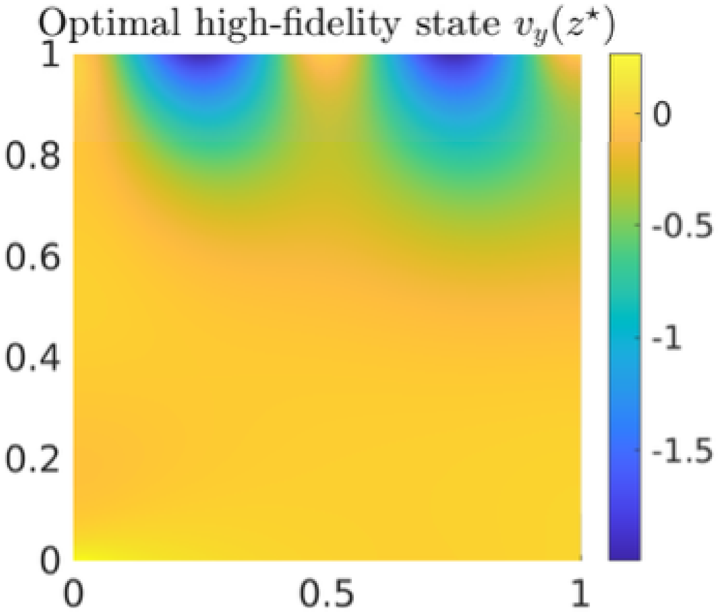
Navier-Stokes solve
with nominal control



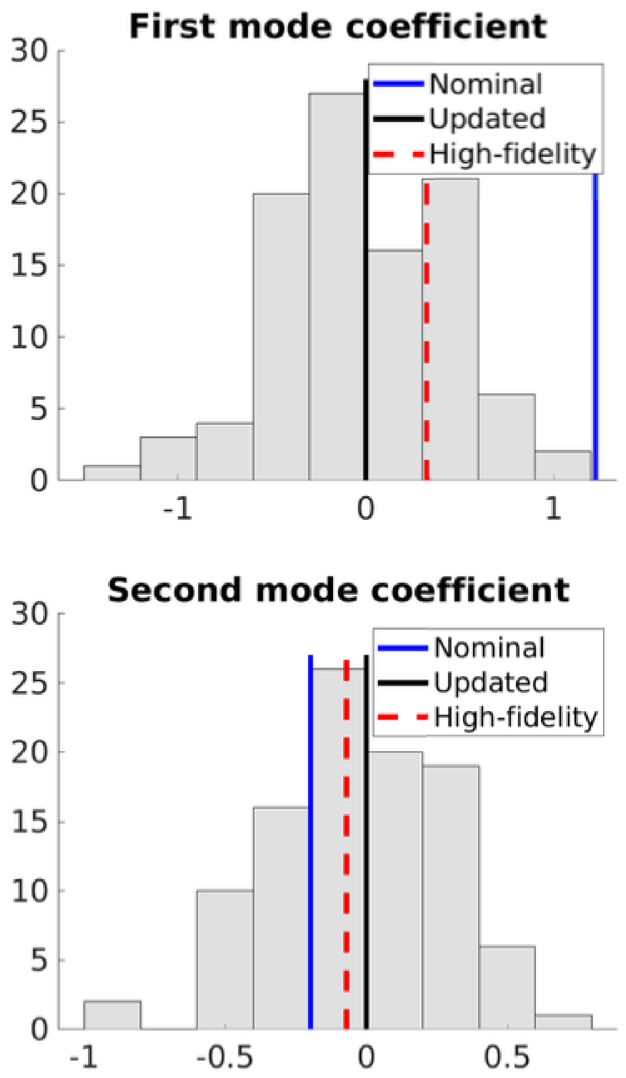
Navier-Stokes solve
with updated control



Navier-Stokes solve
with optimal control



Posterior Controller Uncertainty



- KL representation
- Histogram of posterior
- Goal is for updated to be as close as possible to high-fidelity

Conclusions

- Developed a framework to learn updates of low-fidelity optimal solutions using limited high-fidelity data

- Builds on linear approximation in post-optimality sensitivity analysis

- The discrepancy representation, inverse problem formulation, and judicious linear algebra manipulations enables closed form solution for posterior samples

- Kronecker product representation of the discrepancy facilitates computation which scales with $\dim(\mathcal{U})$ and $\dim(\mathcal{Z})$, not $\dim(\mathcal{U} \otimes \mathcal{Z})$

- Approach is non-intrusive to the high-fidelity data and hence applicable to wide range of applications

✓ Joseph Hart and Bart van Bloemen Waanders, "Hyper-Differential Sensitivity Analysis With Respect to Model Discrepancy (in preparation)