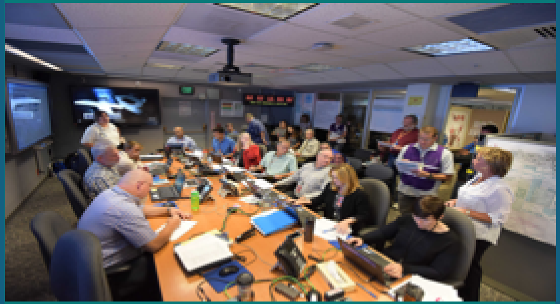
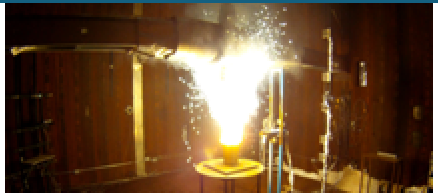
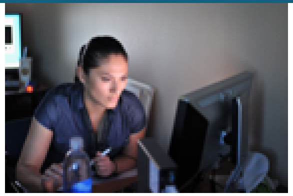
SAND2022-6603C

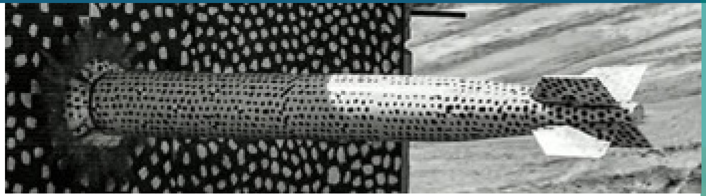
Sandia National Laboratories

Hyper-differential sensitivity analysis with respect to model discrepancy









June 9, 2022

PRESENTED BY

Joseph Hart and Bart van Bloemen Waanders Sandia National Laboratories

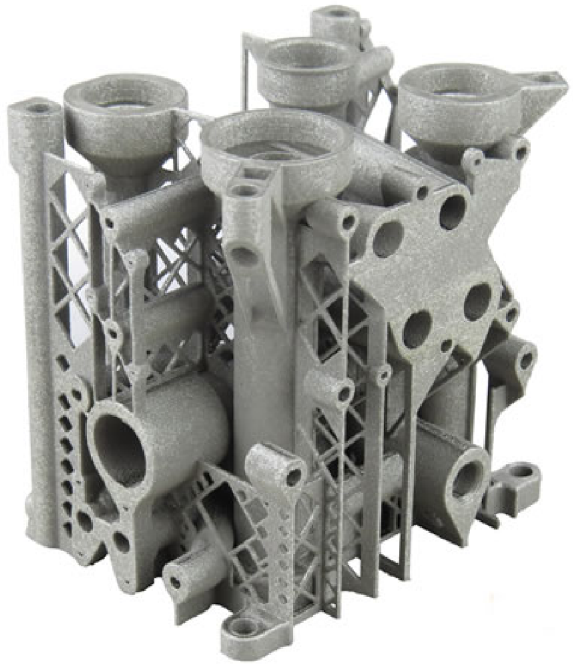
ECCOMAS Congress 2022

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2022-XXXX

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

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

Optimization of approximate models

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

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

Our goals are:

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

Assume that we can

- solve the optimization problem constrained by $\tilde{S}$
- evaluate a high-fidelity model $S(z)$ at a small number of inputs z

Algorithmic Pipeline

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

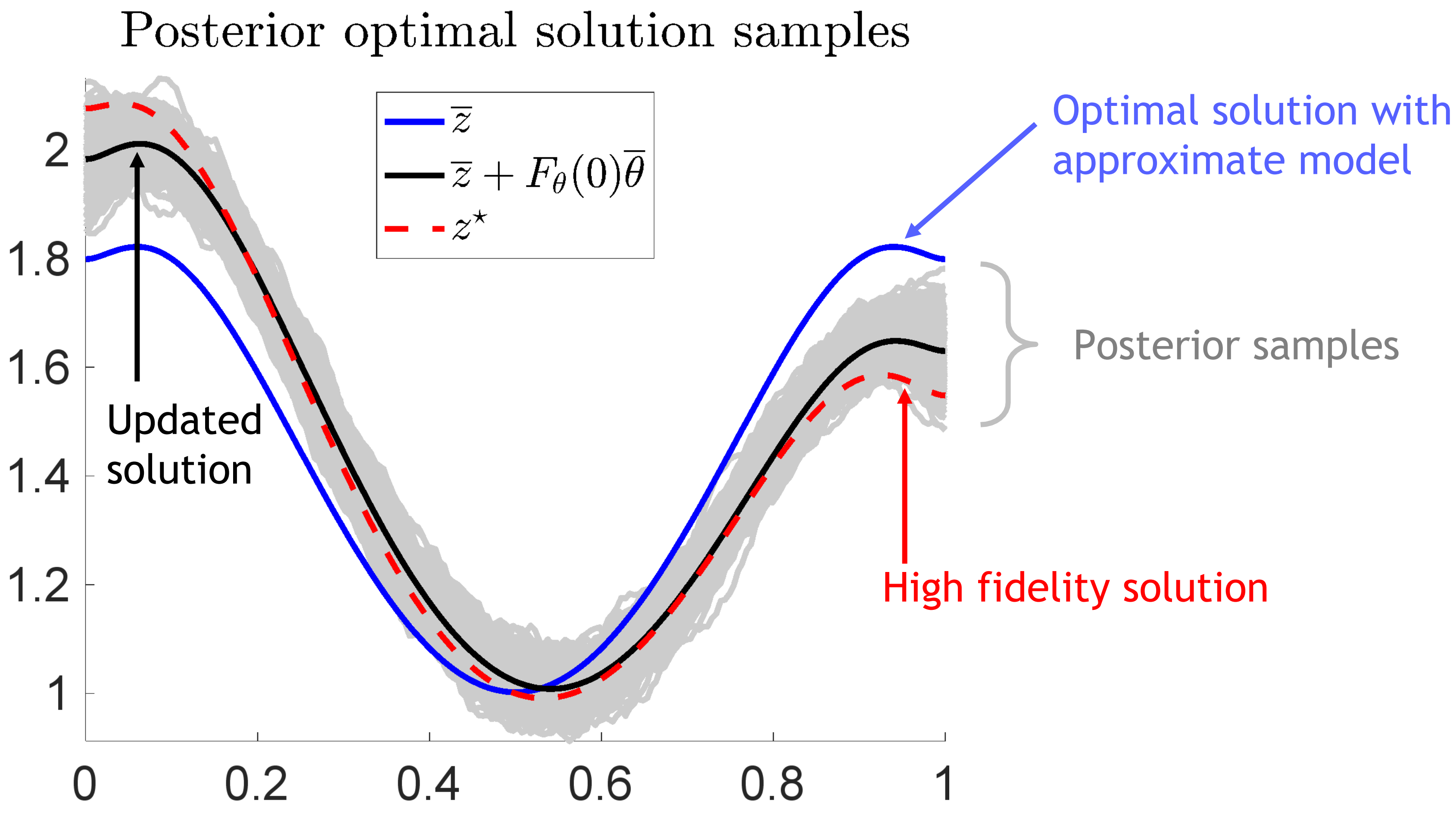
$$\begin{aligned} -\kappa u'' &= z && \text{on } (0,1) \\ \kappa u' &= hu && \text{on } \{0,1\} \end{aligned}$$

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 samples



Algorithmic Pipeline

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

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

Parameterizing model discrepancy

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

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

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

$$\delta(z,\theta) = 0$$

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

- $\delta(z,\theta)$ interpolates between the optimization problems
- parameterize $\delta(z,\theta)$ in a basis expansion with coefficients θ

Algorithmic Pipeline

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

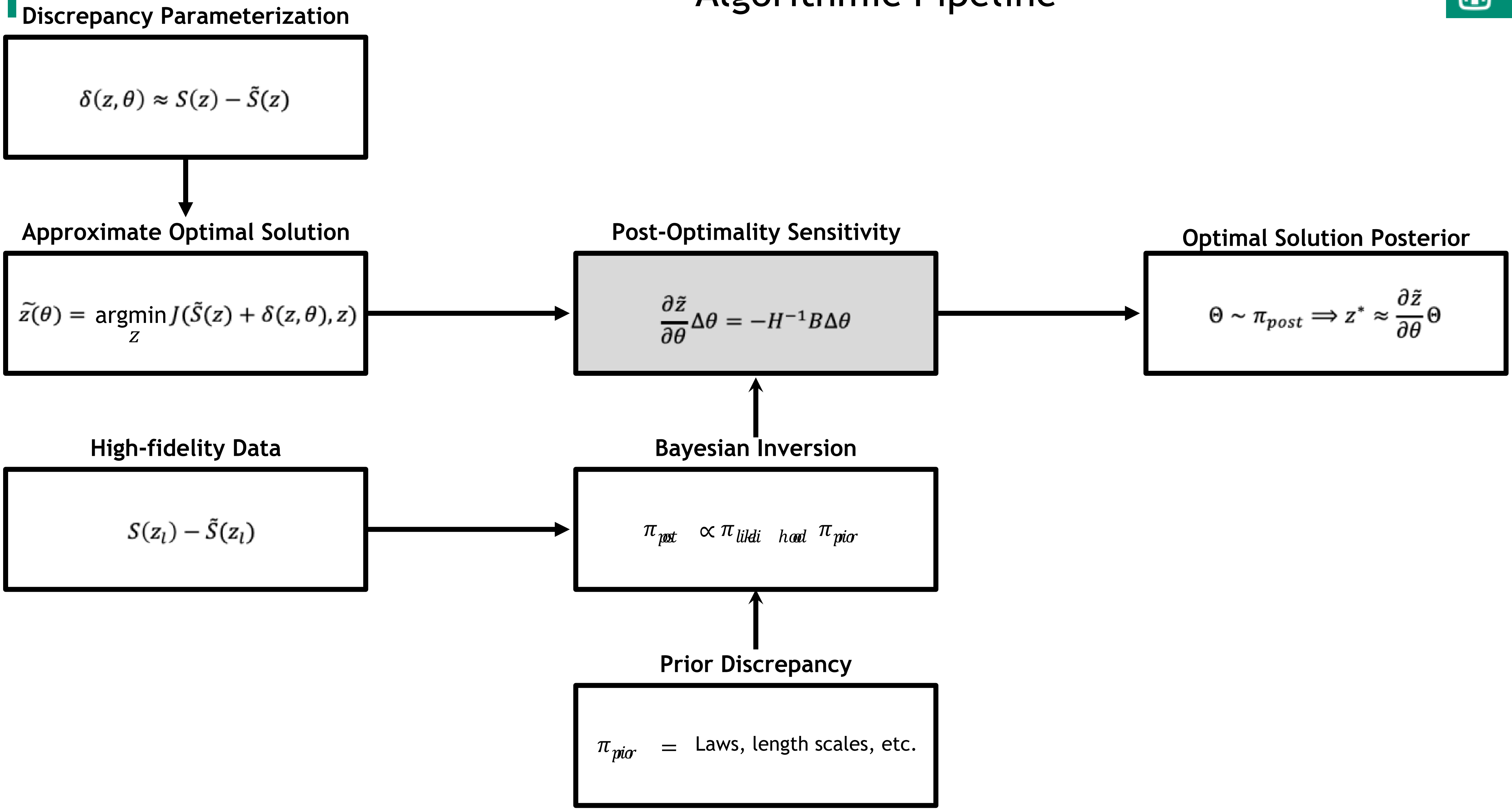
$$\sum_{i=1}^{m}\left(\theta_{i,0}+\sum_{j=1}^{n}\theta_{i,j}(z,\psi_j)_{\mathcal{Z}}\right)\phi_i$$

$$\downarrow$$

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

- discretized $\delta:\mathbb{R}^n\times\mathbb{R}^p\rightarrow\mathbb{R}^m$ 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 $\mathcal{Z}_h$

Algorithmic Pipeline



Post-optimality sensitivities (Hyper-Differential Sensitivity Analysis)

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

- $\tilde{\mathbf{z}}^\star$ solves (1) when $\delta(\mathbf{z},\theta_0) = \mathbf{0}$, the problem solved in practice
- Under mild assumptions, the implicit function theorem 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

Algorithmic Pipeline

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 $\boldsymbol{\delta}$:

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

- marginalize out $\boldsymbol{z}$:

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

- 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

$$\mathbf{A}_\ell = \begin{pmatrix}\mathbf{I}_m & \mathbf{I}_m \otimes \mathbf{z}_\ell^T \mathbf{M}_z\end{pmatrix} \in \mathbb{R}^{m \times p}, \qquad \ell = 1, 2, \ldots, N,$$

- 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 the evaluation of $\delta(\mathbf{z}, \theta)$ for the inputs $\mathbf{z}_\ell$, $\ell = 1, 2, \ldots, N$.
- 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

- 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

- 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 a 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

Posterior samples for discrepancy

- Posterior samples take the form

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

where the mean is

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

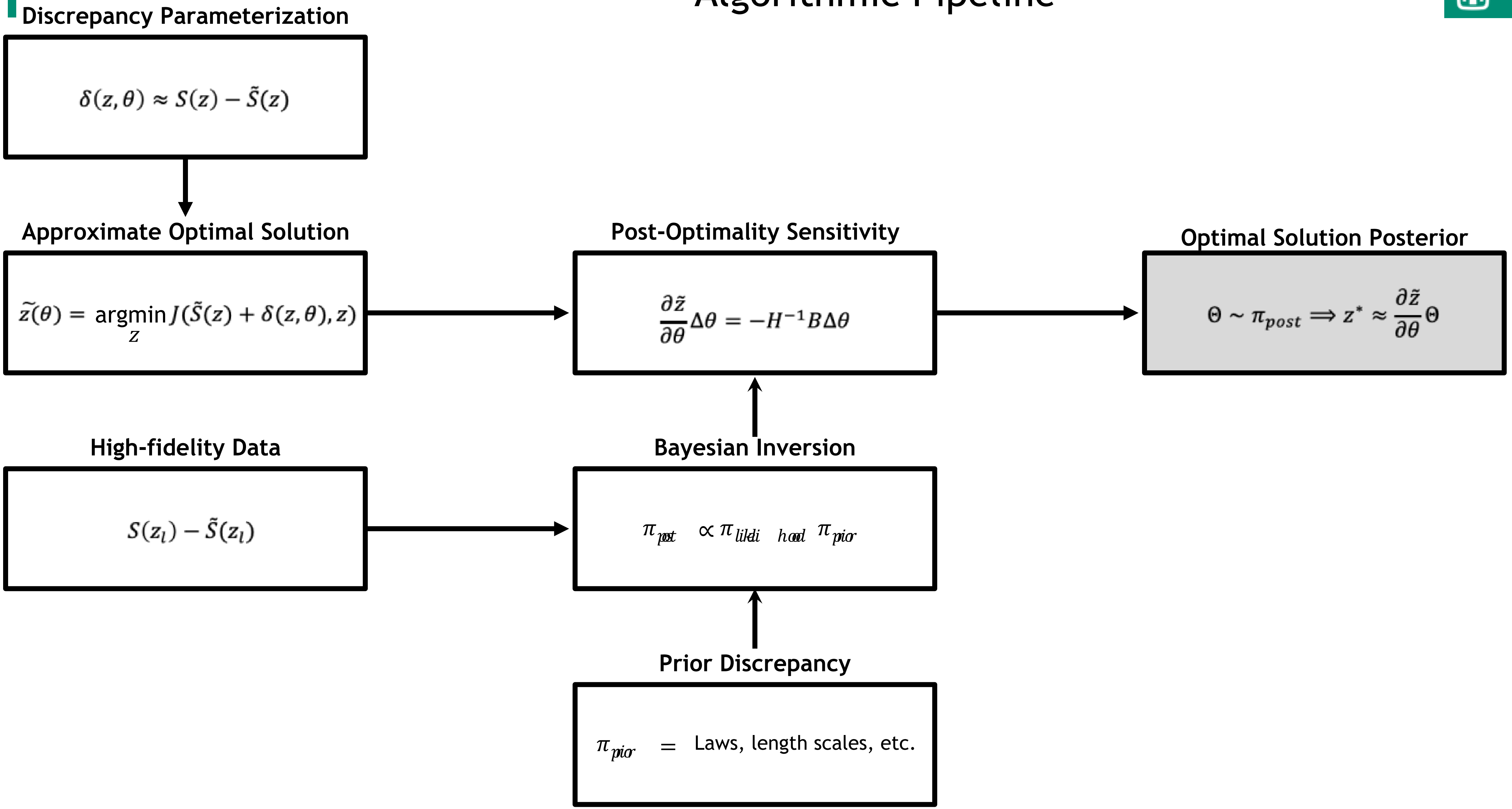
uncertainty in the data informed directions is

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

and uncertainty in the data uninformed directions is

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

Algorithmic Pipeline



Propagating samples through post-optimality sensitivities

Sample

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

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

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

and

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

A fluid flow example to illustrate

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

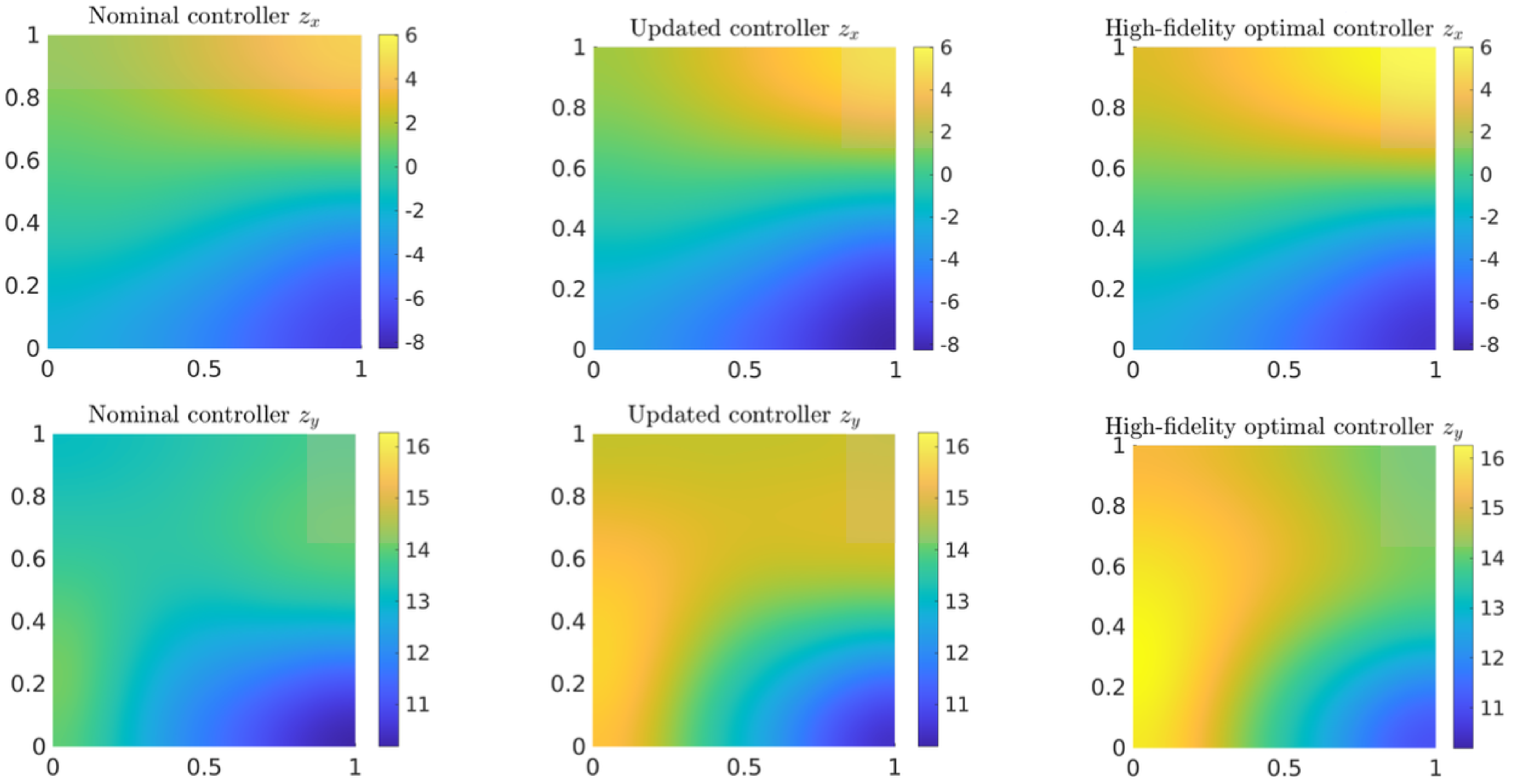
$$\begin{aligned} -\mu\nabla\mathbf{v} + \nabla p &= \mathbf{g} + \mathbf{z} && \text{on } \Omega \\ \nabla \cdot \mathbf{v} &= 0 && \text{on } \Omega \end{aligned}$$

as a simplification of the Navier-Stokes equations

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

How does modeling error effect the flow controller design?

Comparison of controllers



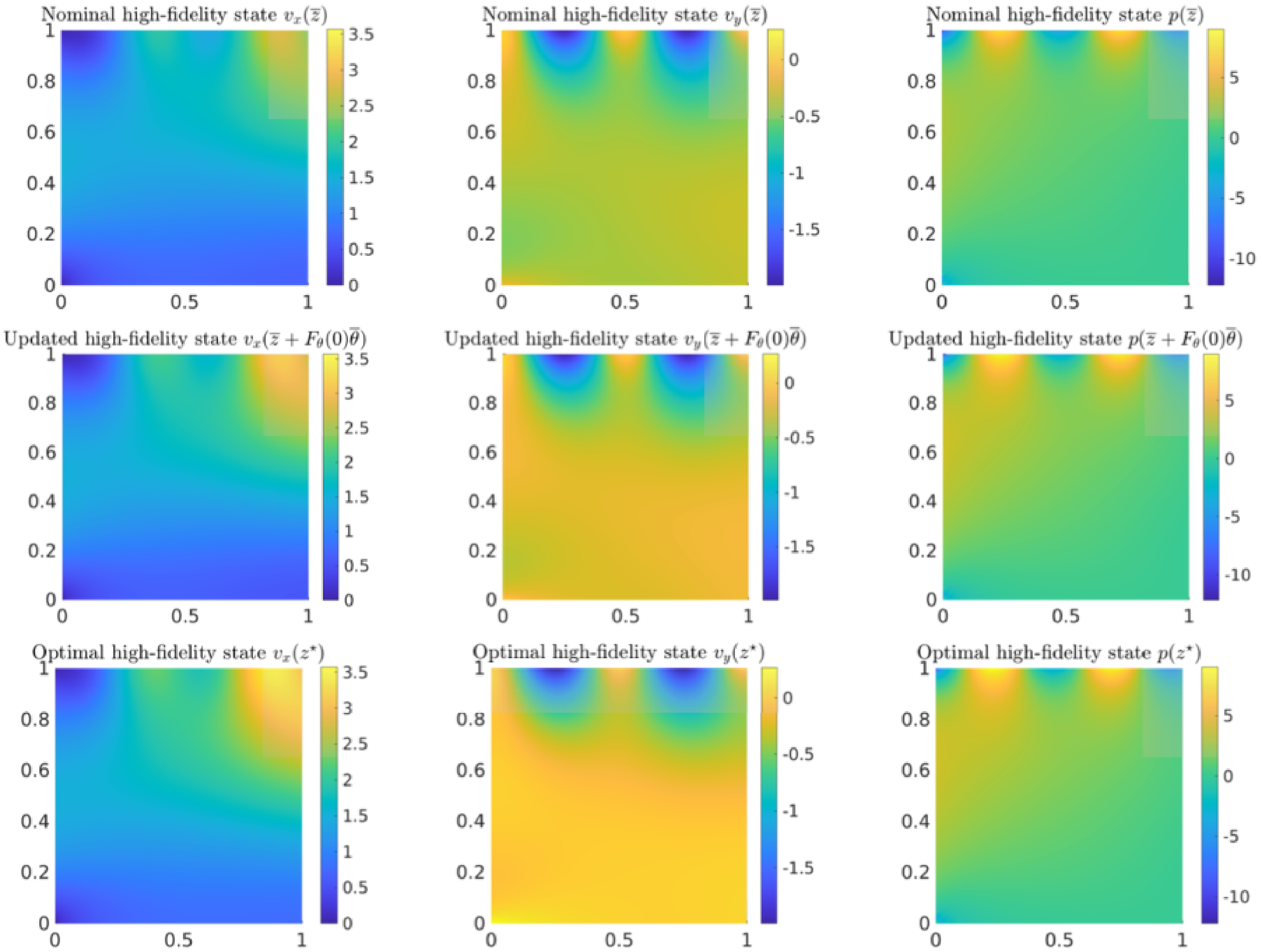
Comparison of Stokes and Navier-Stokes

v_x v_y P

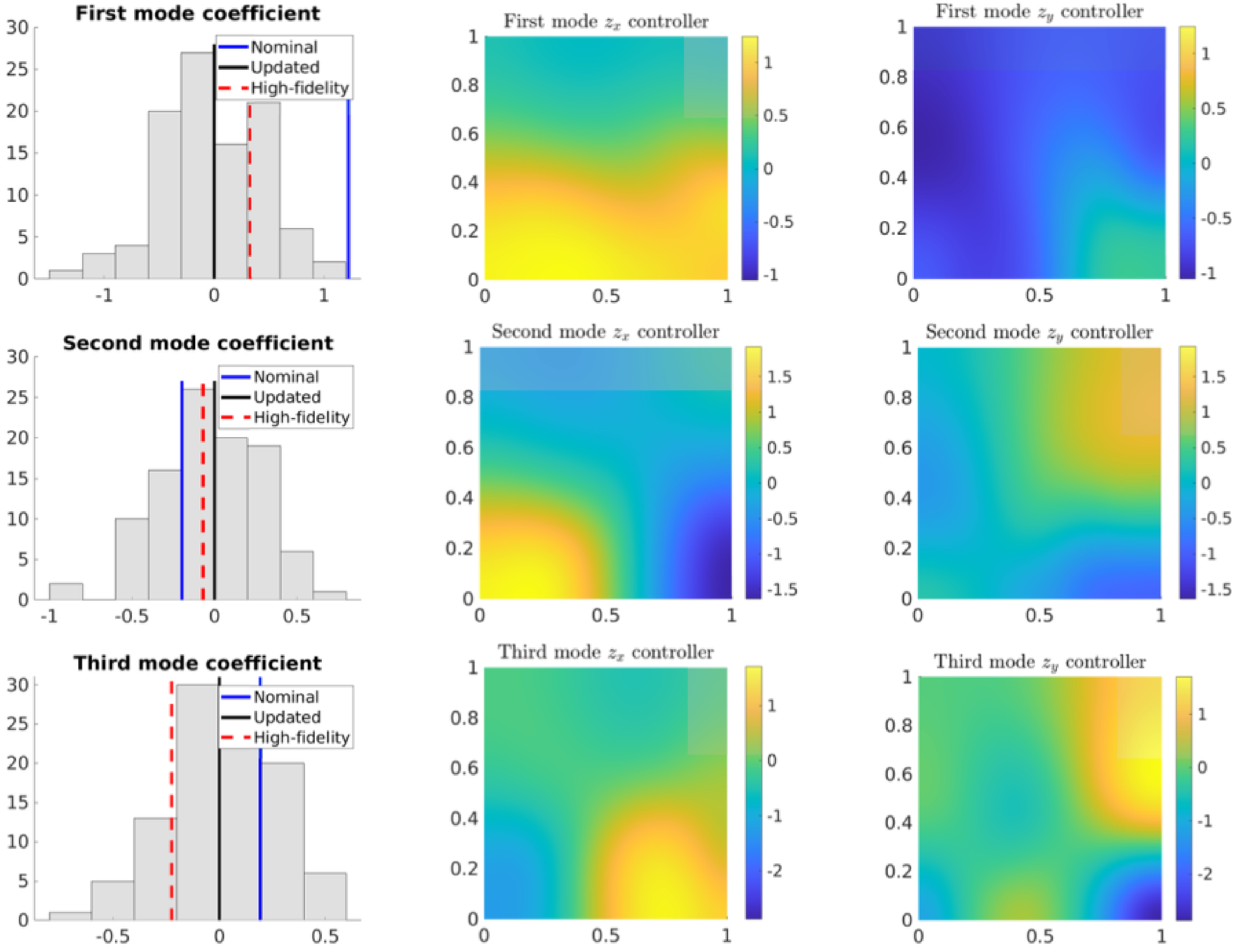
Original Stokes with nominal controller

Navier-Stokes solve with nominal control

Navier Stokes with updated controller



Posterior controller uncertainty



Conclusions

- Developed framework to leverage high-fidelity data to improve low fidelity optimization and characterize uncertainty
- 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