A Scalable Variational Approach for Solving Data-Consistent Stochastic Inverse Problems

Tim Wildey
Sandia National Laboratories
Center for Computing Research
Computational Mathematics Department

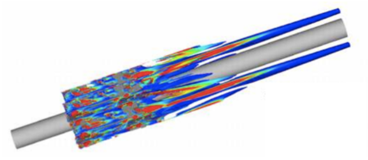
Collaborators: Troy Butler (CU-Denver), Tan Bui-Thanh (UT-Austin), Brad Marvin (UT-Austin)

SIAM Optimization Conference
May 31 - June 3, 2023

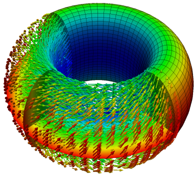
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of
Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear
-NA

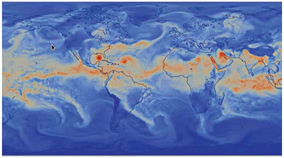
Sandia National Laboratories is a multimission laboratory managed and opera
owned subsidiary of Honeywell International Inc., for the U.S. Department of E
DE-NA0003525.

Flow in Nuclear Reactor (Turbulent CFD)

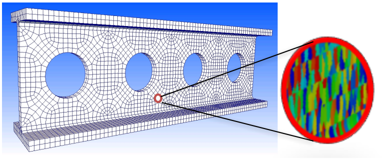


Tokamak Equilibrium (MHD)





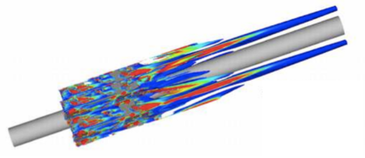
Climate Modeling



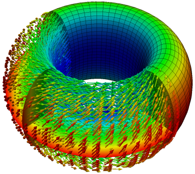
Multi-scale Materials Modeling

Motivation

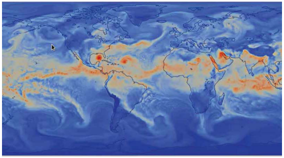
Flow in Nuclear Reactor (Turbulent CFD)



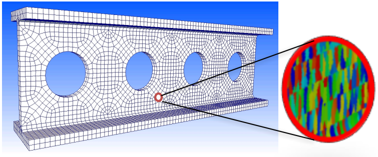
Tokamak Equilibrium (MHD)



We are looking to move beyond forward simulation ...

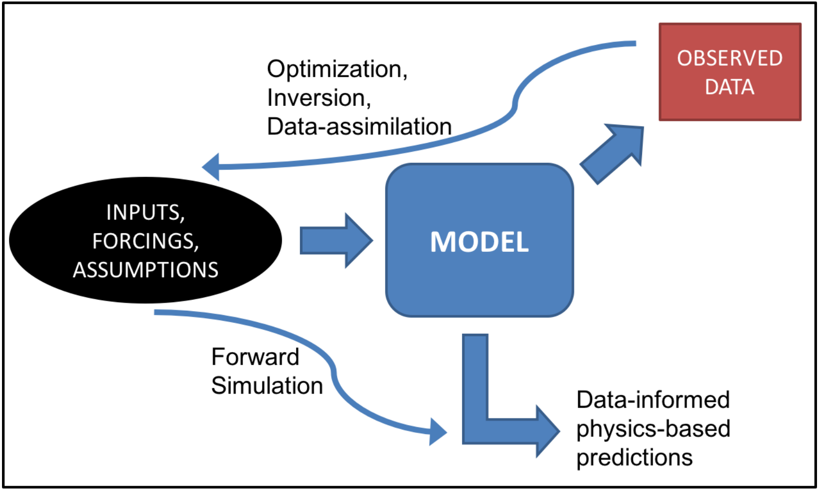


Climate Modeling

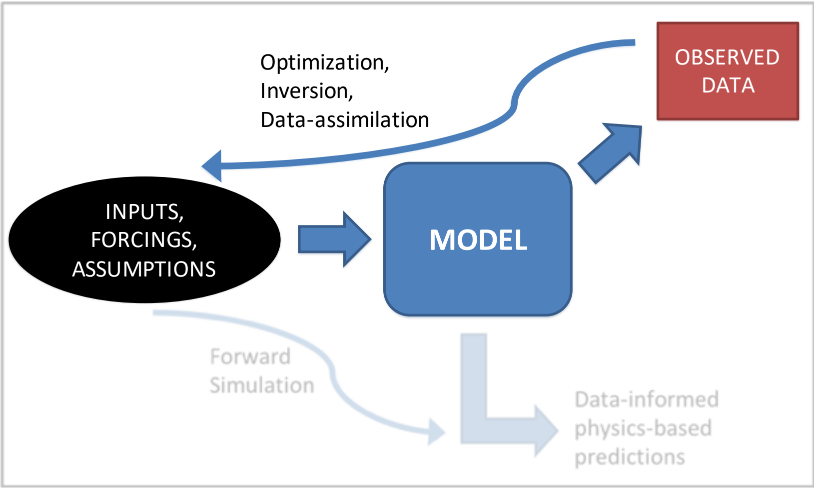


Multi-scale Materials Modeling

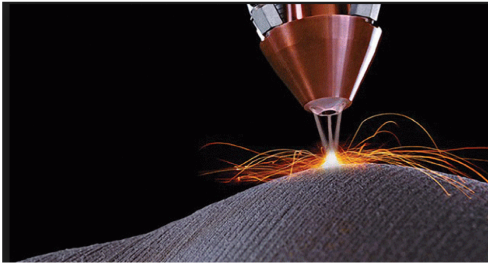
Data-informed Physics-Based Predictions



Data-informed Physics-Based Predictions



Motivation



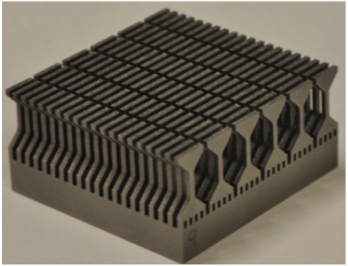
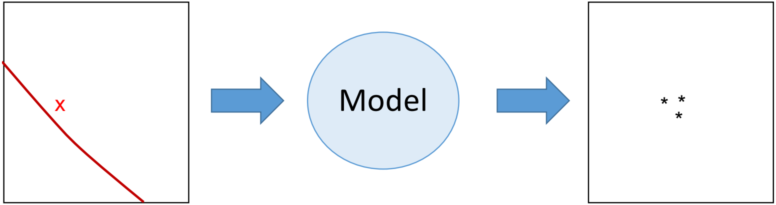


Figure: Additive manufacturing and high-throughput testing provides new data science challenges.

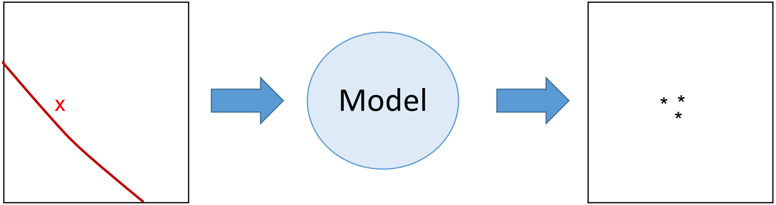
A Deterministic Inverse Problem



Problem

Given some observed data, find $\lambda \in \Lambda$ that best predicts the data.

A Deterministic Inverse Problem

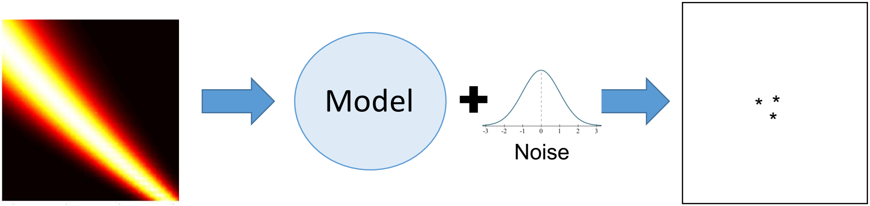


Problem

Given some observed data, find $\lambda \in \Lambda$ that best predicts the data.

- Solutions may not be unique without additional assumptions.
- Requires solving several deterministic forward problems.

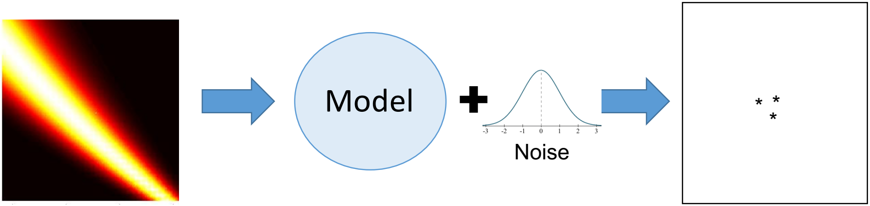
A Stochastic Inverse Problem



Problem

Given some observed data and an assumed noise model, find the parameters that are most likely to have produced the data.

A Stochastic Inverse Problem



Problem

Given some observed data and an assumed noise model, find the parameters that are most likely to have produced the data.

- Solutions may not be unique without additional assumptions.
- Requires solving several deterministic forward problems.

Motivation

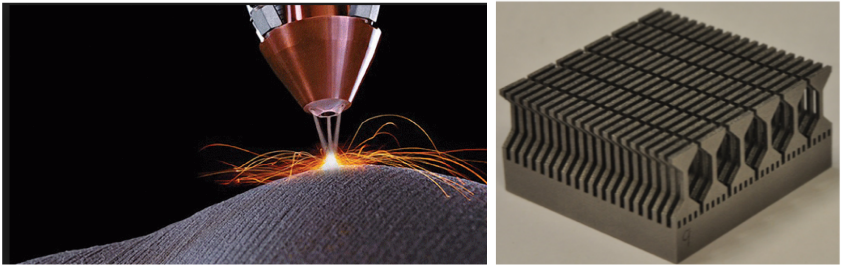
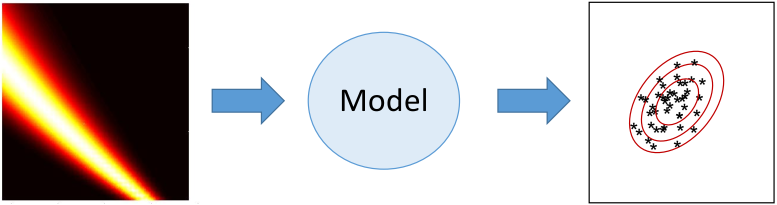


Figure: Additive manufacturing and high-throughput testing provides new data science challenges.

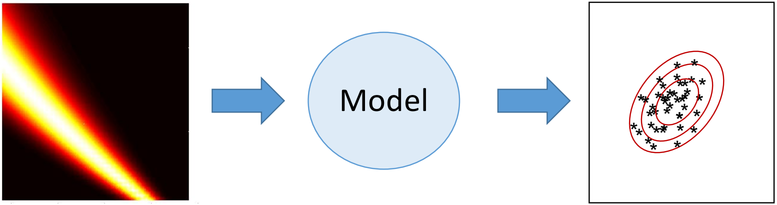
A Different Stochastic Inverse Problem



Problem

Given a probability density on observations, find a probability density on Λ such that the push-forward matches the given density on the observed data.

A Different Stochastic Inverse Problem



Problem

Given a probability density on observations, find a probability density on Λ such that the push-forward matches the given density on the observed data.

- Solutions may not be unique without additional assumptions.
- **We only need to solve a single stochastic forward problem.**

Notation

We assume we are given:

1. A finite-dimensional **parameter space**, Λ.
2. A **parameter-to-observation/data map**, $Q : \Lambda \to \mathcal{D} = Q(\Lambda)$
3. A **observed/target probability measure** on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, denoted $\mathbb{P}_{\mathcal{D}}^{\text{obs}}$, with density $\pi_{\mathcal{D}}^{\text{obs}}$ (typically from experimental data)
4. An **initial probability measure** on $(\Lambda, \mathcal{B}_{\Lambda})$, denoted $\mathbb{P}_{\Lambda}^{\text{init}}$, with density $\pi_{\Lambda}^{\text{init}}$ (typically from prior beliefs or expert knowledge)

Notation

We assume we are given:

1. A finite-dimensional **parameter space**, Λ.
2. A **parameter-to-observation/data map**, $Q : \Lambda \to \mathcal{D} = Q(\Lambda)$
3. A **observed/target probability measure** on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, denoted $\mathbb{P}_{\mathcal{D}}^{\text{obs}}$, with density $\pi_{\mathcal{D}}^{\text{obs}}$ (typically from experimental data)
4. An **initial probability measure** on $(\Lambda, \mathcal{B}_{\Lambda})$, denoted $\mathbb{P}_{\Lambda}^{\text{init}}$, with density $\pi_{\Lambda}^{\text{init}}$ (typically from prior beliefs or expert knowledge)

We need to compute:

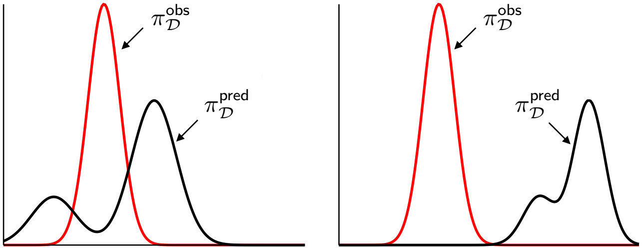
1. The **push-forward of the initial density** through the model.

- In other words, **we need to solve a forward UQ problem using the initial.**
- We use $\pi_{\mathcal{D}}^{\text{pred}}$ to denote this push-forward density.

A Key Assumption

Predictability Assumption

We assume that the observed probability measure, $\mathbb{P}_{\mathcal{D}}^{\text{obs}}$, is absolutely continuous with respect to the push-forward of the initial, $\mathbb{P}_{\mathcal{D}}^{\text{pred}}$.



Good Initial

Bad Initial
(Cannot predict all observations)

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \; \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

Corollary

The updated measure of Λ is 1.

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

Corollary

The updated measure of Λ is 1.

Theorem

$\mathbb{P}_\Lambda^{up}$ is stable with respect to perturbations in $\mathbb{P}_\mathcal{D}^{obs}$ and in $\mathbb{P}_\Lambda^{init}$.

For details: [Combining Push-forward Measures and Bayes' Rule to Construct Consistent Solutions to Stochastic Inverse Problems, BJW. SISC 40 (2), 2018.]

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

The updated density is:

$$\pi_\Lambda^{\text{up}}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda) \frac{\pi_\mathcal{D}^{\text{obs}}(Q(\lambda))}{\pi_\mathcal{D}^{\text{pred}}(Q(\lambda))}.$$

- Both $\pi_\Lambda^{\text{init}}$ and $\pi_\mathcal{D}^{\text{obs}}$ are given.
- Computing $\pi_\mathcal{D}^{\text{pred}}$ requires a forward propagation of the initial density.

A Parameterized Nonlinear System

Example

Consider a parameterized nonlinear system of equations:

$$\begin{aligned} \lambda_1 u_1^2 + u_2^2 &= 1, \\ u_1^2 - \lambda_2 u_2^2 &= 1 \end{aligned}$$

- Quantity of interest is the second component: $Q(\lambda) = u_2$.
- Two inputs, one output
- Given $\pi_{\mathcal{D}}^{\text{obs}} \sim N(0.3, 0.025^2)$.
- Given a uniform initial density.
- Use 10,000 samples from the initial and a standard KDE to approximate the push-forward.
- Use standard rejection sampling to generate samples from $\pi_{\Lambda}^{\text{up}}$.

A Parameterized Nonlinear System

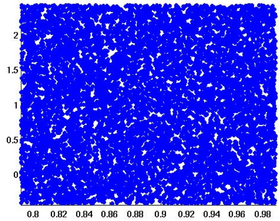
Example

Consider a parameterized nonlinear system of equations:

$$\begin{aligned} \lambda_1 u_1^2 + u_2^2 &= 1, \\ u_1^2 - \lambda_2 u_2^2 &= 1 \end{aligned}$$

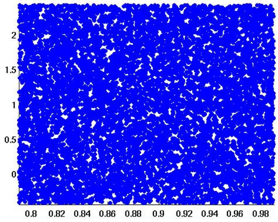
- Quantity of interest is the second component: $Q(\lambda) = u_2$.
- Two inputs, one output
- Given $\pi_{\mathcal{D}}^{\text{obs}} \sim N(0.3, 0.025^2)$.
- Given a uniform initial density.
- Use 10,000 samples from the initial and a standard KDE to approximate the push-forward.
- Use standard rejection sampling to generate samples from $\pi_{\Lambda}^{\text{up}}$.

A Parameterized Nonlinear System

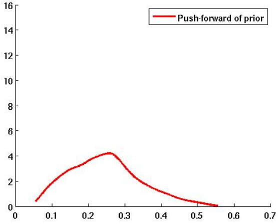


Initial

A Parameterized Nonlinear System

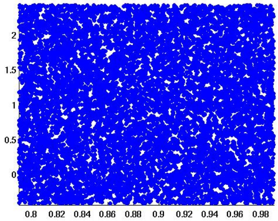


Initial

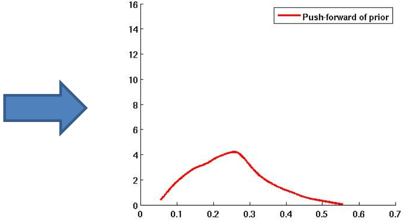


Push-forward of Initial

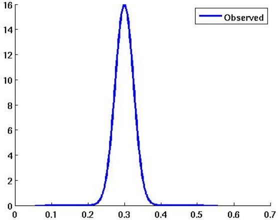
A Parameterized Nonlinear System



Initial

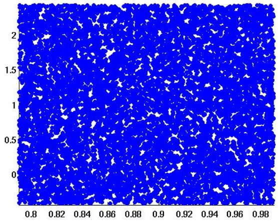


Push-forward of Initial

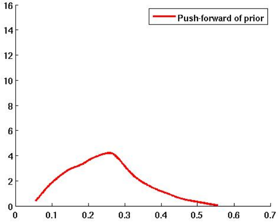


Observed density

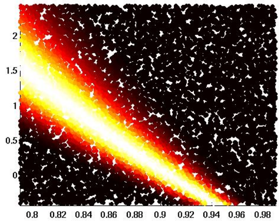
A Parameterized Nonlinear System



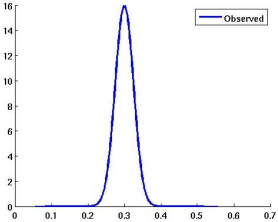
Initial



Push-forward of Initial

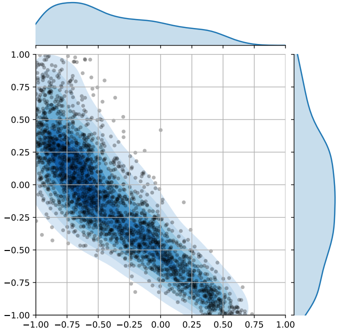


Updated



Observed density

A Parameterized Nonlinear System



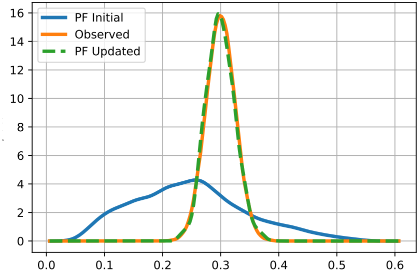


Figure: Samples from the updated density (left) and a comparison of $\pi_{\mathcal{D}}^{\text{obs}}$, $\pi_{\mathcal{D}}^{\text{pred}}$ and push-forward of the updated density (right).

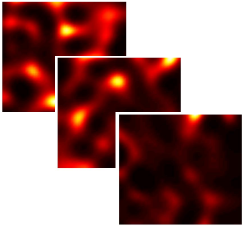
A Higher-Dimensional PDE-based Example

Example

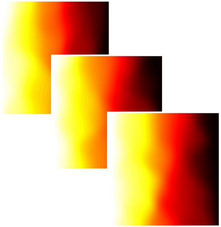
Consider the classical model for single-phase incompressible flow in porous media,

$$-\nabla \cdot (K(\lambda, x)\nabla p(\lambda)) = 0, \quad x \in \Omega = [0,1]^2,$$

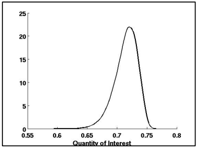
with $p = 1$ along the left boundary and $p = 0$ along the right boundary.
We assume $\log(K(\lambda, x))$ is described by a Karhunen-Loeve (KL) expansion.
We use correlation length of 0.01 in each direction and retain 100 terms.



Realizations of permeability



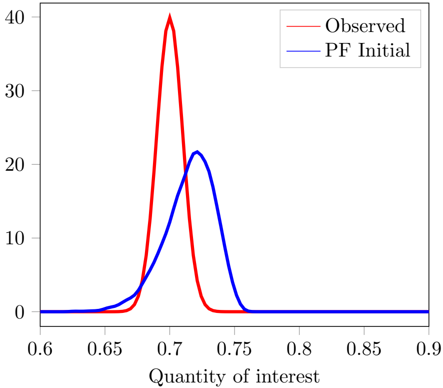
Realizations of pressure



PDF for QoI

A Higher-Dimensional PDE-based Example

- The QoI is the pressure at $(0.0540, 0.5487)$.
- Use 10,000 samples and a KDE to approximate the push-forward of the initial.
- Assume $\pi_{\mathcal{D}}^{\text{obs}} \sim N(0.7, 1.0\text{E-4})$.



The 1D Marginals of Updated Density Tell Us Nothing

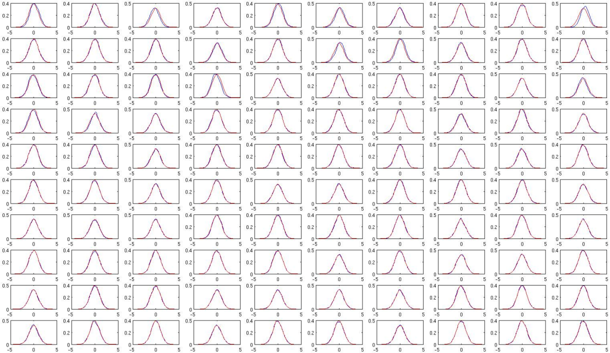
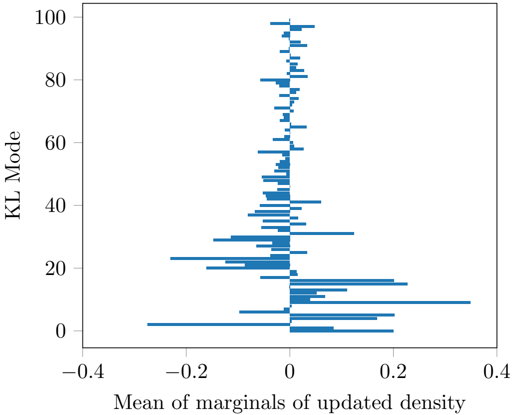
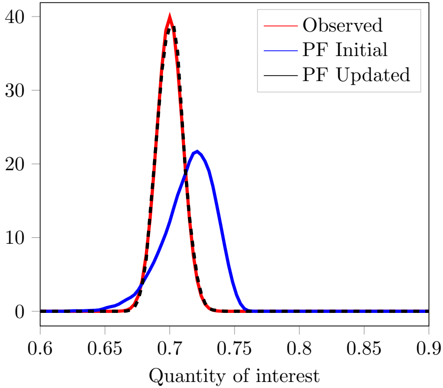


Figure: The 1D marginals from the initial (red) and the updated (blue) for the first KL-mode (upper-left) through the 100th KL-mode (lower-right).

The Means Are Slightly More Interesting



Push-forward of the Updated Matches the Observations



	Mean	Variance
Observed	0.7	1.0E-4
PF-Updated	0.7001	1.0170E-4

Relationship with Statistical Bayesian Inference

Using Bayes theorem we can define a posterior density [Stuart 2010; Gelman et al 2013; Jaynes 1998, ...]:

$$\tilde{\pi}_{\Lambda}^{\text{post}}(\lambda|q) = \pi_{\Lambda}^{\text{init}}(\lambda)\frac{\pi(q|\lambda)}{C}.$$

Relationship with Statistical Bayesian Inference

Using Bayes theorem we can define a posterior density [Stuart 2010; Gelman et al 2013; Jaynes 1998, ...]:

$$\tilde{\pi}_{\Lambda}^{\text{post}}(\lambda|q) = \pi_{\Lambda}^{\text{init}}(\lambda)\frac{\pi(q|\lambda)}{C}.$$

Example

Let $\Lambda = [-1, 1]$ and consider the simple nonlinear map

$$q(\lambda) = \lambda^p, \quad p = 1, 3, 5, \ldots.$$

Here, p is not uncertain and are used to vary the nonlinearity.

Relationship with Statistical Bayesian Inference

Using Bayes theorem we can define a posterior density [Stuart 2010; Gelman et al 2013; Jaynes 1998, ...]:

$$\tilde{\pi}_{\Lambda}^{\text{post}}(\lambda|q) = \pi_{\Lambda}^{\text{init}}(\lambda)\frac{\pi(q|\lambda)}{C}.$$

Example

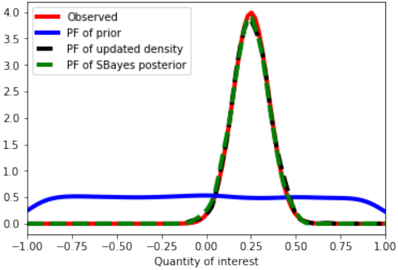
Let $\Lambda = [-1, 1]$ and consider the simple nonlinear map

$$q(\lambda) = \lambda^p, \quad p = 1, 3, 5, \ldots.$$

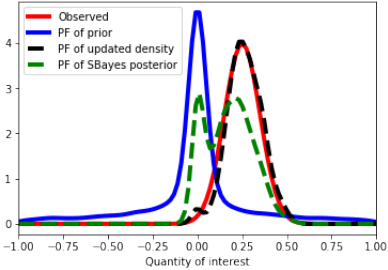
Here, p is not uncertain and are used to vary the nonlinearity.

- Assume a uniform initial/prior and $\pi_{\mathcal{D}}^{\text{obs}} \sim N(0.25, 0.1^2)$.
- For the statistical Bayesian approach, we use an observed value of $\hat{q} = 0.25$ and assume a Gaussian noise model $\eta \sim N(0, 0.1^2)$.

Comparing Push-forwards for Linear and Nonlinear Maps

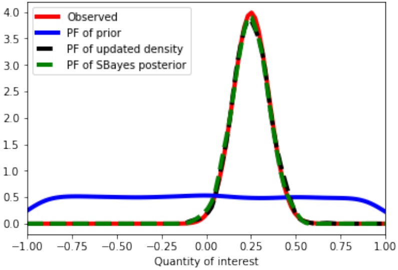


Linear map ($p = 1$)

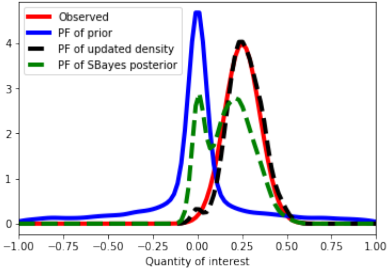


Nonlinear map ($p = 5$)

Comparing Push-forwards for Linear and Nonlinear Maps



Linear map ($p = 1$)



Nonlinear map ($p = 5$)

The Bayesian and measure-theoretic formulations **solve different problems**, **give different densities** and **make different predictions**.

Connections with Deterministic Optimization

Consider the linear map $q = A\lambda$ and $\pi_{\Lambda}^{\text{init}} \sim N(\overline{\lambda}, \Gamma_{\text{init}})$, and $\pi_{\mathcal{D}}^{\text{obs}} \sim N(\overline{q}, \Gamma_{\text{obs}})$. The updated density is given by,

$$\pi_{\Lambda}^{\text{up}}(\lambda) \sim \exp\Big(-\big(\underbrace{\underbrace{\frac{1}{2}\|\Gamma_{\text{obs}}^{-1/2}(A\lambda - \overline{q})\|_{\mathbb{R}^m}^2}_{\text{Data mismatch}} + \underbrace{\frac{1}{2}\|\Gamma_{\text{obs}}^{-1/2}(\lambda - \overline{\lambda})\|_{\mathbb{R}^n}^2}_{\text{Tikhonov regularization}}}_{\text{Same as statistical Bayesian}} - \underbrace{\frac{1}{2}\|\Gamma_{A}^{-1/2}A(\lambda - \overline{\lambda})\|_{\mathbb{R}^m}^2}_{\text{Un-regularization}}\big)\Big)$$

where $\Gamma_A = A\Gamma_{\text{init}}A^*$.

Connections with Deterministic Optimization

Consider the linear map $q = A\lambda$ and $\pi_\Lambda^{\text{init}} \sim N(\overline{\lambda}, \Gamma_{\text{init}})$, and $\pi_\mathcal{D}^{\text{obs}} \sim N(\overline{q}, \Gamma_{\text{obs}})$. The updated density is given by,

$$\pi_\Lambda^{\text{up}}(\lambda) \sim \exp\Big(-\big(\underbrace{\underbrace{\frac{1}{2}\|\Gamma_{\text{obs}}^{-1/2}(A\lambda - \overline{q})\|_{\mathbb{R}^m}^2}_{\text{Data mismatch}} + \underbrace{\frac{1}{2}\|\Gamma_{\text{obs}}^{-1/2}(\lambda - \overline{\lambda})\|_{\mathbb{R}^n}^2}_{\text{Tikhonov regularization}}}_{\text{Same as statistical Bayesian}} - \underbrace{\frac{1}{2}\|\Gamma_A^{-1/2}A(\lambda - \overline{\lambda})\|_{\mathbb{R}^m}^2}_{\text{Un-regularization}}\big)\Big)$$

where $\Gamma_A = A\Gamma_{\text{init}}A^*$.

The point that maximizes the updated density, denoted λ_{MUD}, is defined by

$$\lambda_{\text{MUD}} = \underset{\lambda \in \Lambda}{\operatorname{argmin}} \quad J(\lambda),$$

where

$$J(\lambda) = \frac{1}{2}\|\Gamma_{\text{obs}}^{-1/2}(A\lambda - \overline{q})\|_{\mathbb{R}^m}^2 + \frac{1}{2}\|\Gamma_{\text{init}}^{-1/2}(\lambda - \overline{\lambda})\|_{\mathbb{R}^n}^2 - \frac{1}{2}\|\Gamma_A^{-1/2}A(\lambda - \overline{\lambda})\|_{\mathbb{R}^m}^2.$$

Connections with Deterministic Optimization

We can rewrite the regularization terms as,

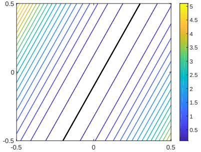
$$\frac{1}{2}\|\Gamma_{\text{init}}^{-1/2}(\lambda - \overline{\lambda})\|_{\mathbb{R}^n}^2 - \frac{1}{2}\|\Gamma_A^{-1/2}A(\lambda - \overline{\lambda})\|_{\mathbb{R}^m}^2 = \frac{1}{2}(\lambda - \overline{\lambda})^T \underbrace{(\Gamma_{\text{init}}^{-1} - A^*(A\Gamma_{\text{init}}A^*)^{-1}A)}_{R}(\lambda - \overline{\lambda})$$

Regularization matrix: $R = \Gamma_{\text{init}}^{-1} - A^*(A\Gamma_{\text{init}}A^*)^{-1}A$.

If A is invertible, then $R = 0$.
$\Rightarrow$ Regularization is "turned off" if there is a unique solution.

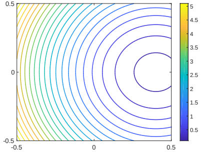
More generally, regularization is only applied in directions not informed by the map/data.

Bayesian MAP Point:



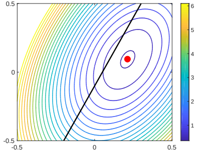
Data mismatch

+



Tikhonov regularization

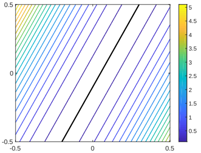
=



Objective function

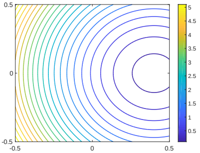
Connections with Deterministic Optimization

Bayesian MAP Point:



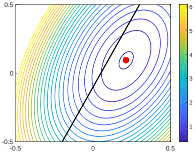
Data mismatch

+



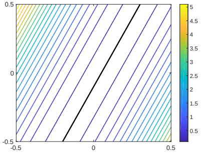
Tikhonov regularization

=



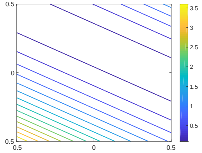
Objective function

MUD Point:



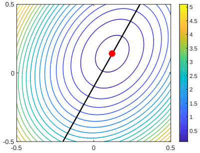
Data mismatch

+



Modified regularization

=



Modified objective function

Comparison of Computational Effort (Naive Approach)

MAP Point

Preliminary computations:

None

Objective function evaluation:

Inputs: A, x, $\Gamma_{\text{init}}^{-1}$, $\overline{x}$, $\overline{q}$, Γ_{obs}^{-1}

Compute: $q = Ax$ (1 forward solve)

Set: $\Delta x = x - \overline{x}$ and $\Delta q = q - \overline{q}$

Return: $J(x) = \frac{1}{2}\Delta q^T \Gamma_{\text{obs}}^{-1} \Delta q + \frac{1}{2}\Delta x^T \Gamma_{\text{init}}^{-1} \Delta x$

Gradient evaluation:

Inputs: A^*, Δx, $\Gamma_{\text{init}}^{-1}$, Δq, Γ_{obs}^{-1}

Compute: $g = A^* \Gamma_{\text{obs}}^{-1} \Delta q$ (1 adjoint solve)

Return: $\nabla J(x) = g + \Gamma_{\text{init}}^{-1} \Delta x$

Hessian evaluation on vector v:

Inputs: A, A^*, $\Gamma_{\text{init}}^{-1}$, Γ_{obs}^{-1}, v

Compute: $z = Av$ (1 forward solve)

Compute: $w = A^* \Gamma_{\text{obs}}^{-1} z$ (1 adjoint solve)

Return: $Hv = w + \Gamma_{\text{init}}^{-1} v$

Comparison of Computational Effort (Naive Approach)

MAP Point

Preliminary computations:

None

Objective function evaluation:

Inputs: A, x, $\Gamma_{\text{init}}^{-1}$, $\overline{x}$, $\overline{q}$, Γ_{obs}^{-1}

Compute: $q = Ax$ (1 forward solve)

Set: $\Delta x = x - \overline{x}$ and $\Delta q = q - \overline{q}$

Return: $J(x) = \frac{1}{2}\Delta q^T \Gamma_{\text{obs}}^{-1} \Delta q + \frac{1}{2}\Delta x^T \Gamma_{\text{init}}^{-1} \Delta x$

Gradient evaluation:

Inputs: A^*, Δx, $\Gamma_{\text{init}}^{-1}$, Δq, Γ_{obs}^{-1}

Compute: $g = A^* \Gamma_{\text{obs}}^{-1} \Delta q$ (1 adjoint solve)

Return: $\nabla J(x) = g + \Gamma_{\text{init}}^{-1} \Delta x$

Hessian evaluation on vector v:

Inputs: A, A^*, $\Gamma_{\text{init}}^{-1}$, Γ_{obs}^{-1}, v

Compute: $z = Av$ (1 forward solve)

Compute: $w = A^* \Gamma_{\text{obs}}^{-1} z$ (1 adjoint solve)

Return: $Hv = w + \Gamma_{\text{init}}^{-1} v$

MUD Point

Preliminary computations:

Compute: $\Gamma_A = (A\Gamma_{\text{init}} A^*)^{-1}$ (M forward and adjoint solves)

Compute: $\hat{x} = A^* \Gamma_{\text{obs}}^{-1} \overline{q}$ (1 adjoint solve)

Compute: $\tilde{q} = A\overline{x}$ (1 forward solve)

Compute: $\tilde{z} = A^* \Gamma_A \tilde{q}$ (1 adjoint solve)

Objective function evaluation:

Inputs: A, A^*, x, $\Gamma_{\text{init}}^{-1}$, $\overline{x}$, $\overline{q}$, $\tilde{q}$, Γ_{obs}^{-1}, Γ_A

Compute: $q = Ax$ (1 forward solve)

Compute: $w = A^* \Gamma_A (q - \tilde{q})$ (1 adjoint solve)

Set: $\Delta x = x - \overline{x}$ and $\Delta q = q - \overline{q}$

Return: $J(x) = \frac{1}{2}\Delta q^T \Gamma_{\text{obs}}^{-1} \Delta q + \frac{1}{2}\Delta x^T \Gamma_{\text{init}}^{-1} \Delta x - \frac{1}{2}\Delta x^T w$

Gradient evaluation:

Inputs: A^*, Δx, $\Gamma_{\text{init}}^{-1}$, q, Γ_{obs}^{-1}, Γ_A, $\tilde{z}$, $\hat{x}$

Compute: $g = A^* (\Gamma_{\text{obs}}^{-1} - \Gamma_A) q$ (1 adjoint solve)

Return: $\nabla J(x) = g + \tilde{z} - \hat{x} + \Gamma_{\text{init}}^{-1} \Delta x$

Hessian evaluation on vector v:

Inputs: A, A^*, $\Gamma_{\text{init}}^{-1}$, Γ_{obs}^{-1}, Γ_A, v

Compute: $z = Av$ (1 forward solve)

Compute: $w = A^* (\Gamma_{\text{obs}}^{-1} - \Gamma_A) z$ (1 adjoint solve)

Return: $Hv = w + \Gamma_{\text{init}}^{-1} v$

Example: Random Linear Maps

Example

Consider the following random linear map,

$$A : \mathbb{R}^n \to \mathbb{R}^m,$$

where $n > m$ (more parameters than observations) and each entry is U(0,1).

The covariance of the prior and observed densities are given:

$$\Gamma_{\text{init}} = 1.0\mathbb{I}_n, \quad \Gamma_{\text{obs}} = 0.25\mathbb{I}_m.$$

The mean of the initial and the mean of the observations are also U(0,1).

We assess the scalability with increasing n and m.

Trust-region Newton-CG method and reduce gradient to 1e-12.

For each case, repeat 25 times and report the average performance.

Scaling With Dimension of Input Space (n)

Dimension of Output Space is Fixed: $m = 10$

n	MAP Point					MUD Point				
	Iters	CG	Misfit	F+A	Time	Iters	CG	Misfit	F+A	Time
100	9.1	48.2	12.2	114.7	0.012	9.2	48.5	2.84e-13	147.6	0.012
500	9.2	55.1	60.7	128.6	0.026	9.2	55.6	3.96e-13	161.8	0.027
1000	10.6	61.7	121.9	144.6	0.109	10.7	61.9	3.06e-13	178.9	0.110
2000	11.0	62.9	242.8	147.8	0.549	11.0	63.9	2.76e-13	183.9	0.553
4000	11.2	65.3	484.1	153.0	2.88	11.3	66.8	3.75e-13	190.6	2.90
8000	12.4	71.4	965.6	167.7	19.7	12.6	73.0	8.59e-13	206.7	20.0

Table: Comparison of the number of optimization iterations, the number of CG iterations, the value of the final data misfit, the average number of combined forward and adjoint solves and the average run time in seconds.

Scaling With Dimension of Output Space (m)

Dimension of Input Space is Fixed: $n = 1000$

m	MAP Point					MUD Point				
	Iters	CG	Misfit	F+A	Time	Iters	CG	Misfit	F+A	Time
5	10.8	54.8	115.7	131.3	0.108	10.9	55.4	3.47e-13	156.6	0.109
10	10.6	61.7	121.9	144.6	0.109	10.7	61.9	3.06e-13	178.9	0.110
20	10.6	60.8	122.9	142.8	0.118	10.6	61.3	5.07e-13	197.5	0.118
40	10.9	57.2	125.4	136.2	0.128	10.9	57.8	2.79e-13	231.4	0.129
80	10.9	65.4	126.4	152.6	0.138	11.0	65.9	2.21e-13	327.7	0.139
160	11.8	82.4	131.8	188.4	0.202	11.6	79.7	1.18e-12	517.4	0.212

Table: Comparison of the number of optimization iterations, the number of CG iterations, the value of the final data misfit, the average number of combined forward and adjoint solves and the average run time in seconds.

The Initial Measure in the Infinite-Dimensional Case

Similar to the Bayesian case [Stuart 2010], we have a Radon-Nikodym derivative

$$\frac{d\mathbb{P}_\Lambda^{\text{up}}}{d\mathbb{P}_\Lambda^{\text{init}}}(\lambda) = \frac{\pi_\mathcal{D}^{\text{obs}}(Q(\lambda))}{\pi_\mathcal{D}^{\text{pred}}(Q(\lambda))},$$

$\implies$ we can consider random fields if $\mathbb{P}_\Lambda^{\text{init}}$ is chosen appropriately.

The Initial Measure in the Infinite-Dimensional Case

Similar to the Bayesian case [Stuart 2010], we have a Radon-Nikodym derivative

$$\frac{d\mathbb{P}_\Lambda^{\text{up}}}{d\mathbb{P}_\Lambda^{\text{init}}}(\lambda) = \frac{\pi_\mathcal{D}^{\text{obs}}(Q(\lambda))}{\pi_\mathcal{D}^{\text{pred}}(Q(\lambda))},$$

$\implies$ we can consider random fields if $\mathbb{P}_\Lambda^{\text{init}}$ is chosen appropriately.

Let Γ_{init} be defined by the biharmonic operator:

$$\Gamma_{\text{init}} = \mathcal{A}^{-2} = (\delta I + \gamma\nabla\cdot(\Theta\nabla))^{-2},$$

such that half powers, $\Gamma_{\text{init}}^{1/2}$ and $\Gamma_{\text{init}}^{-1/2}$, are easy to compute.

Linear-Gaussian Case: SVD and Low-Rank Approximations

- Assume $Q(\lambda) = A\lambda$, $\pi_{\Lambda}^{\text{init}} \sim \mathrm{N}(\overline{\lambda}, \Gamma_{\text{init}})$, and $\pi_{\mathcal{D}}^{\text{obs}} \sim \mathrm{N}(\overline{q}, \Gamma_{\text{obs}})$.
- Set $A^* = M^{-1}A^\top$ and $U^* = U^\top M$ where M is the mass matrix associated with the discretized field.
- Similar to the Bayesian case [Bui-Thanh et al 2013], define $\tilde{A} \in \mathbb{R}^{n\times m}$ by

$$\tilde{A} = \Gamma_{\text{init}}^{1/2} A^* \Gamma_{\text{obs}}^{-1/2}$$

and denote the SVD of $\tilde{A}$ by

$$\tilde{A} = U\Sigma V^\top, \quad U \in \mathbb{R}^{n\times m}, \quad \Sigma \in \mathbb{R}^{m\times m}, \quad V \in \mathbb{R}^{m\times m},$$

where we impose $U^*U = \mathbb{I}_m$, $V^\top V = \mathbb{I}_m$ and $VV^\top = \mathbb{I}_m$.

Linear-Gaussian Case: SVD and Low-Rank Approximations

- Assume $Q(\lambda) = A\lambda$, $\pi_\Lambda^{\text{init}} \sim N(\overline{\lambda}, \Gamma_{\text{init}})$, and $\pi_{\mathcal{D}}^{\text{obs}} \sim N(\overline{q}, \Gamma_{\text{obs}})$.
- Set $A^* = M^{-1}A^\top$ and $U^* = U^\top M$ where M is the mass matrix associated with the discretized field.
- Similar to the Bayesian case [Bui-Thanh et al 2013], define $\tilde{A} \in \mathbb{R}^{n\times m}$ by

$$\tilde{A} = \Gamma_{\text{init}}^{1/2} A^* \Gamma_{\text{obs}}^{-1/2}$$

  and denote the SVD of $\tilde{A}$ by

$$\tilde{A} = U\Sigma V^\top, \quad U \in \mathbb{R}^{n\times m}, \quad \Sigma \in \mathbb{R}^{m\times m}, \quad V \in \mathbb{R}^{m\times m},$$

  where we impose $U^*U = \mathbb{I}_m$, $V^\top V = \mathbb{I}_m$ and $VV^\top = \mathbb{I}_m$.
- Solve efficiently using matrix-free randomized algorithms [Halko et al 2011].
- Number of PDE solves scales linearly with the desired rank.
- Effective rank depends on information content of data [Bui-Thanh et al 2013]

Linear Gaussian Case: Hessians and Covariances

The Hessian and covariance of the updated density can be expressed as,

$$
\begin{aligned}
H_{\text{up}} &= \Gamma_{\text{init}}^{-1/2}(\mathbb{I}_n + U\left(\Sigma^2 - \mathbb{I}_m\right)U^*)\Gamma_{\text{init}}^{-1/2}, \\
\Gamma_{\text{up}} &= \Gamma_{\text{init}}^{1/2}\left(\mathbb{I}_n + UGU^*\right)\Gamma_{\text{init}}^{1/2},
\end{aligned}
$$

where $G \in \mathbb{R}^{m\times m}$ is a diagonal matrix with entries

$$
G_{ii} = \frac{\Sigma_{ii}^2 - 1}{\Sigma_{ii}^2}.
$$

Linear Gaussian Case: Hessians and Covariances

The Hessian and covariance of the updated density can be expressed as,

$$H_{\text{up}} = \Gamma_{\text{init}}^{-1/2}(\mathbb{I}_n + U\left(\Sigma^2 - \mathbb{I}_m\right)U^*)\Gamma_{\text{init}}^{-1/2},$$
$$\Gamma_{\text{up}} = \Gamma_{\text{init}}^{1/2}\left(\mathbb{I}_n + UGU^*\right)\Gamma_{\text{init}}^{1/2},$$

where $G \in \mathbb{R}^{m\times m}$ is a diagonal matrix with entries

$$G_{ii} = \frac{\Sigma_{ii}^2 - 1}{\Sigma_{ii}^2}.$$

The Hessian and covariance of the posterior in Bayesian approach are similar:

$$H_{\text{post}} = \Gamma_{\text{prior}}^{-1/2}(\mathbb{I}_n + U\Sigma^2 U^*)\Gamma_{\text{prior}}^{-1/2},$$
$$\Gamma_{\text{post}} = \Gamma_{\text{prior}}^{1/2}\left(\mathbb{I}_n + UDU^*\right)\Gamma_{\text{prior}}^{1/2},$$

where

$$D_{ii} = \frac{\Sigma_{ii}^2}{\Sigma_{ii}^2 + 1}.$$

Linear Gaussian Case: Hessians

There exists a unitary matrix $\tilde{U}$ such that the Hessian can be expressed as,

$$H_{\text{up}} = M\Gamma_{\text{init}}^{-1/2}\tilde{U}\left\{\left[\begin{array}{c|c}\Sigma^2 & 0\\ \hline 0 & 0\end{array}\right] + \left[\begin{array}{c|c}0 & 0\\ \hline 0 & \mathbb{I}\end{array}\right]\right\}\tilde{U}^*\Gamma_{\text{init}}^{-1/2},$$

with regularization only applied in the data-uninformed directions

Linear Gaussian Case: Hessians

There exists a unitary matrix $\tilde{U}$ such that the Hessian can be expressed as,

$$H_{\text{up}} = M\Gamma_{\text{init}}^{-1/2}\tilde{U}\left\{\left[\begin{array}{c|c}\Sigma^2 & 0 \\ \hline 0 & 0\end{array}\right] + \left[\begin{array}{c|c}0 & 0 \\ \hline 0 & \mathbb{I}\end{array}\right]\right\}\tilde{U}^*\Gamma_{\text{init}}^{-1/2},$$

with regularization only applied in the data-uninformed directions

The Bayesian approach gives a similar expression:

$$H_{\text{post}} = \Gamma_{\text{prior}}^{-1/2}(U\Sigma^2U^* + I)\Gamma_{\text{prior}}^{-1/2} = \Gamma_{\text{prior}}^{-1/2}\tilde{U}\left\{\left[\begin{array}{c|c}\Sigma^2 & 0 \\ \hline 0 & 0\end{array}\right] + \left[\begin{array}{c|c}\mathbb{I} & 0 \\ \hline 0 & \mathbb{I}\end{array}\right]\right\}\tilde{U}^*\Gamma_{\text{prior}}^{-1/2},$$

with regularization applied in all directions.

Example: Linear Advection Diffusion

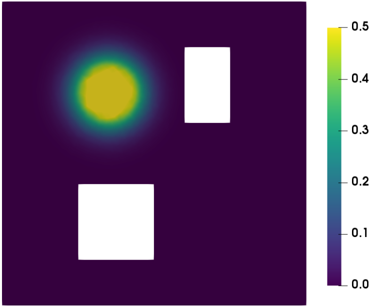
Consider the advection-diffusion equation

$$\begin{aligned} \frac{\partial u}{\partial t} - \kappa \Delta u + \mathrm{v} \cdot \nabla u = 0 \quad & \forall (x,t) \in \Omega \times (0,T), \\ \nabla u \cdot \mathrm{n} = 0 \quad & \forall (x,t) \in \partial\Omega \times (0,T), \\ u(\cdot, 0) = \lambda \quad & \forall x \in \Omega, \end{aligned}$$

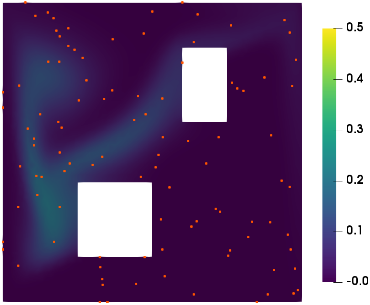
where κ is a diffusion coefficient and v is a non time-varying velocity field.

- Use FEniCS for forward model and hIPPYlib for optimization.
- Observations are pointwise observations of u at $t = 3$
- Generate $\pi_{\mathcal{D}}^{\text{obs}}$ using data-generating point in Λ and Gaussian noise.
- Compare the data-generating point and the MUD point.
- Dimension of both the discretized parameter and state spaces is 7863.
- Optimization with inexact-Newton globalized with a backtracking line search.

Example: Linear Advection Diffusion



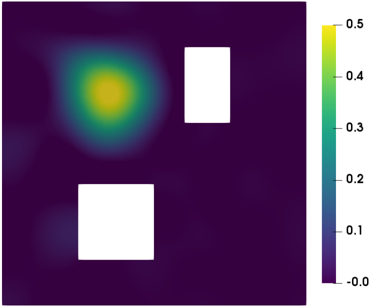
(a) $t = 0$



(b) $t = 3$

Figure: Maximum of data-generating distribution (left) and corresponding observations at $t = 3$ (right).

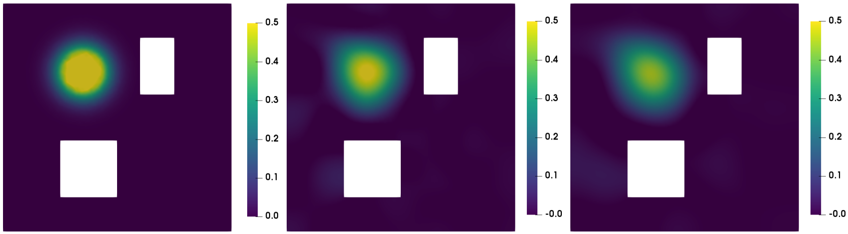
Example: Linear Advection Diffusion



(a) λ_{MUD}

(b) MUD Predictions

Figure: Maximum of the updated density (left) and the corresponding predictions at $t = 3$ (right).



(a) $t = 0$ (b) λ_{MUD} (c) Bayesian λ_{MAP}

Figure: Data-generating point (left), MUD point (middle), and MAP point (right).

Example: Linear Advection Diffusion

Parameters		MAP Point			MUD Point		
N_d	rank	N_s	J_{mis}	$\|\|\nabla J\|\|_2$	N_s	J_{mis}	$\|\|\nabla J\|\|_2$
10	10	62	3.032e-01	1.156e-09	54	4.784e-17	1.061e-06
20	20	78	1.584e+00	2.124e-08	100	3.710e-13	3.318e-07
50	29	128	1.262e+01	4.014e-07	142	3.689e+00	2.979e-07
100	36	150	1.595e+01	6.327e-08	148	4.153e+00	5.128e-08
200	43	170	1.850e+01	5.962e-07	172	1.045e+01	2.137e-07
500	49	182	1.786e+01	4.324e-07	192	8.802e+00	6.837e-07

Table: Scaling study for coarse mesh ($n = 2023$). N_d is the dimension of the data space, rank is the number of eigenvectors used in the low rank decomposition, N_s is the total number of model evaluations, J_{mis} is final value of the misfit portion of the cost function, and $||\nabla J||_2$ is the final value of the norm of the gradient.

Note what happens to J_{mis} when the rank is less than N_d.

Example: Linear Advection Diffusion

Parameters		MAP Point			MUD Point		
N_d	rank	N_s	J_{mis}	$\|\|\nabla J\|\|_2$	N_s	J_{mis}	$\|\|\nabla J\|\|_2$
10	10	70	1.234e-01	4.200e-08	60	2.535e-15	2.046e-06
20	20	92	3.525e+00	9.896e-09	86	1.937e-11	2.845e-07
50	30	130	1.141e+01	3.238e-07	130	8.924e+00	2.361e-07
100	38	162	1.519e+01	3.837e-07	172	9.352e+00	5.971e-08
200	40	166	1.631e+01	1.796e-07	168	1.185e+01	6.281e-07
500	48	194	1.755e+01	8.434e-08	194	4.699e+00	4.714e-08

Table: Scaling study for refined mesh ($n = 7863$). N_d is the dimension of the data space, rank is the number of eigenvectors used in the low rank decomposition, N_s is the total number of model evaluations, J_{mis} is final value of the misfit portion of the cost function, and $||\nabla J||_2$ is the final value of the norm of the gradient.

Note what happens to J_{mis} when the rank is less than N_d.

Partial Linearization for Nonlinear Maps

- For nonlinear maps, the push-forward is typically non-Gaussian
- "De-regularization" term is difficult to characterize.
- Often find MAP point, λ_{MAP}, to obtain initial guess for λ_{MUD}.
- We can linearize around λ_{MAP}:

$$f(\lambda) \approx f(\lambda_{\text{MAP}}) + A(\lambda - \lambda_{\text{MAP}}).$$

- Use this linearization to approximate the de-regularization term:

$$J_{PL}(\lambda) = \frac{1}{2}||\Gamma_{\text{obs}}^{-1/2}(f(\lambda) - \overline{q})||_{\mathbb{R}^m} + \frac{1}{2}||R^{1/2}(\lambda - \overline{\lambda})||_{L^2(\Omega)},$$
$$R = \Gamma_{\text{init}}^{-1} - A^*(A\Gamma_{\text{init}}A^*)^{-1}A,$$

which allows us to employ the scalable approximation procedures.

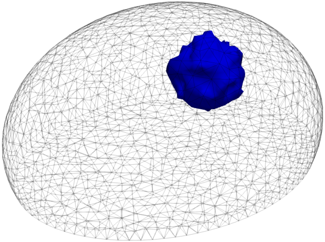
Example: Hyper-elasticity

Consider the equations for the displacement of a compressible Mooney-Rivlin solid:

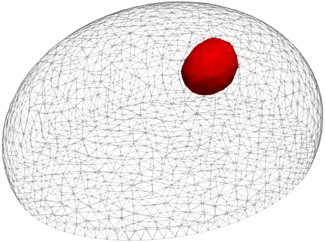
$$\min_{u \in V} \quad \Pi(u) = \int_\Omega \psi(u)\ dx - \int_\Omega B \cdot u\ dx - \int_{\partial\Omega} T \cdot u\, dS,$$
$$\psi(u) = C_1(\bar{I}_1 - 3) + C_2(\bar{I}_2 - 3) + C_3(J-1)^2,$$

- Given a data-generating distribution on the first coefficient in the strain-energy density C_1 ($\lambda = \log(C_1)$).
- Pointwise observations of the boundary displacement.
- Maximum of data-generating distribution is piecewise constant with value of 0.3 inside a small inclusion and 0.1 everywhere else in the domain.
- Dimension of the parameter and state spaces is 36905 and 14487 respectively.
- Optimization used inexact-Newton CG.

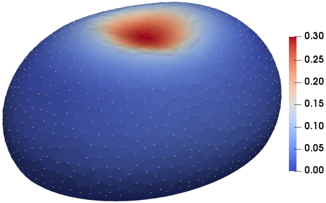
Example: Hyper-elasticity



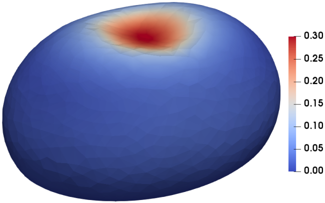
(a) Data-generating contour



(b) λ_{MUD} contour



(c) Data-generating predictions



(d) λ_{MUD} predictions

Conclusions and Future Work

- Some inverse problems are best posed as constructive pullback probability measures
- Solution may not be unique, i.e., multiple probability measures can have the same pushforward
- Data-consistent inversion borrows ideas from Bayesian inference to obtain **existence, uniqueness and stability**
- Standard approach relies on density estimation on observations
- Shown we can leverage connections with deterministic optimization to construct scalable approximations
- Future work will connect this with [Pilosov et al 2023] for parameter estimation.

Thanks! Questions?

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, ASCR, Early Career Research Program and the CHaRMNET MMICC Center.

Thank you for your attention!

Questions?