Computational Advances in Data-Consistent Inversion: Measure-Theoretic Methods for Improving Predictions

Michael Pilosov

Advisor: Troy Butler

University of Colorado Denver

November 6, 2020

Department of Mathematical & Statistical Sciences

UNIVERSITY OF COLORADO DENVER

The one where we describe why any of this matters.

Broad Goals of Uncertainty Quantification

- Make inferences and predictions
- Quantify and reduce uncertainties (aleotoric, epistemic)
- Be *accurate* and *precise*
- Design “efficient” experiments
- Collect and use data “intelligently”

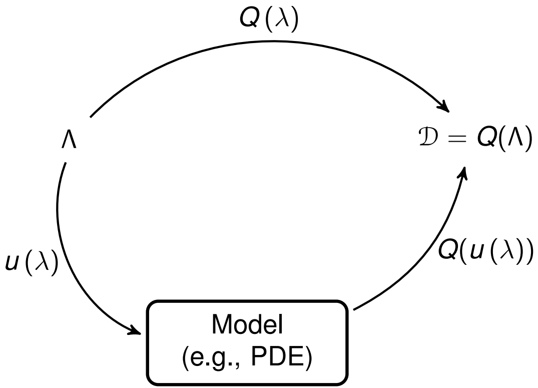
The one where we define the letters we use and what they mean.

- State variable: u (e.g. heat, energy, pressure, deflection)
- Parameters: λ (e.g. source term, diffusion, boundary data)
- Model: $\mathcal{M}(u, \lambda) = 0$, so $u(\lambda)$
- Quantity of Interest (QoI) map, (piecewise smooth):

$$Q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_D \end{bmatrix}, \quad \text{where} \quad q_i : u(\lambda) \to \mathbb{R}$$

- We write $Q(\lambda) := Q(u(\lambda))$ to make the dependence on λ explicit.

The one where we illustrate how a QoI map relates to outputs.



Definition (Stochastic Forward Problem (SFP))

Given a probability measure $\mathbb{P}_\Lambda$ on $(\Lambda, \mathcal{B}_\Lambda)$, and QoI map Q, the *stochastic forward problem* is to determine a measure, $\mathbb{P}_\mathcal{D}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$ that satisfies

$$\mathbb{P}_\mathcal{D}(E) = \mathbb{P}_\Lambda\left(Q^{-1}(E)\right), \; \forall \, E \in \mathcal{B}_\mathcal{D}. \tag{1.1}$$

Definition (Stochastic Forward Problem (SFP))

Given a probability measure $\mathbb{P}_\Lambda$ on $(\Lambda, \mathcal{B}_\Lambda)$, and QoI map Q, the *stochastic forward problem* is to determine a measure, $\mathbb{P}_\mathcal{D}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$ that satisfies

$$\mathbb{P}_\mathcal{D}(E) = \mathbb{P}_\Lambda\left(Q^{-1}(E)\right), \ \forall\, E \in \mathcal{B}_\mathcal{D}. \tag{1.2}$$

Definition (Stochastic Inverse Problem (SIP))

Given a probability measure, $\mathbb{P}_\mathcal{D}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$ the *stochastic inverse problem* is to determine a probability measure, $\mathbb{P}_\Lambda$, on $(\Lambda, \mathcal{B}_\Lambda)$ satisfying

$$\mathbb{P}_\Lambda(Q^{-1}(E)) = \mathbb{P}_\mathcal{D}(E), \ \forall\, E \in \mathcal{B}_\mathcal{D}. \tag{1.3}$$

Equation (1.3) is referred to as the *consistency condition*.

Definition (Consistent Solution and Density)

If $\mathbb{P}_\Lambda$ or $\mathbb{P}_\mathcal{D}$ absolutely continuous w.r.t μ_Λ or $\mu_\mathcal{D}$, resp, then we write

$$\pi_\Lambda := \frac{d\mathbb{P}_\Lambda}{d\mu_\Lambda} \text{ or } \pi_\mathcal{D} := \frac{d\mathbb{P}_\mathcal{D}}{d\mu_\mathcal{D}}$$

to denote the Radon-Nikodym derivatives of $\mathbb{P}_\Lambda$ and $\mathbb{P}_\mathcal{D}$, resp.

In such a case, we can rewrite (1.2) and (1.3) using these pdfs:

$$\mathbb{P}_\Lambda(Q^{-1}(E)) = \int_{Q^{-1}(E)} \pi_\Lambda(\lambda)\, d\mu_\Lambda = \int_E \pi_\mathcal{D}(Q(\lambda))\, d\mu_\mathcal{D} = \mathbb{P}_\mathcal{D}(E)$$

Definition (Initial Distribution)

When $\mathbb{P}_\Lambda$ in (1.2) quantifies the characterization of uncertainty in parameter variability before observations on QoI are taken into account, it is referred to as the *initial measure* $\mathbb{P}_{\text{in}}$.

Given a dominating μ_Λ on $(\Lambda, \mathcal{B}_\Lambda)$, the Radon-Nikodym derivative π_{in} w.r.t μ_Λ is referred to as the *initial distribution*.

Definition (Predicted Distribution)

The *predicted distribution* is the push-forward of π_{in} under the map Q, and is denoted as π_{pr}.

Given as Radon-Nikodym derivative (w.r.t $\mu_{\mathcal{D}}$) of pushforward measure:

$$\mathbb{P}_{\text{pr}}(E) = \mathbb{P}_{\text{in}}\left(Q^{-1}(E)\right), \; \forall\, E \in \mathcal{B}_{\mathcal{D}}. \tag{1.4}$$

Definition (Observed Distribution)

When $\mathbb{P}_{\mathcal{D}}$ in (1.3) quantifies the characterization of uncertainty in the QoI data, it is referred to as the *observed measure*, $\mathbb{P}_{\text{ob}}$.

Given a dominating $\mu_{\mathcal{D}}$ on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, the Radon-Nikodym derivative $\mathbb{P}_{\text{ob}}$ w.r.t. $\mu_{\mathcal{D}}$ is referred to as the *observed distribution* π_{ob}.

The one where we define the solution to the Stochastic Inverse Problem.

We now have all of the definitions required for the *updated distribution* which solves the SIP:

$$\pi_{\text{up}}(\lambda) := \pi_{\text{in}}(\lambda)\frac{\pi_{\text{ob}}(Q(\lambda))}{\pi_{\text{pr}}(Q(\lambda))}. \tag{1.5}$$

The one with some practical considerations.

- May evaluate π_{up} directly for any sample of Λ (one model solve).
- Stable w.r.t. perturbations in the Total Variation metric.
- Accuracy proportional approximation error of the predicted density.
- Approximate π_{pr} with density estimation using samples from π_{in}.
- We (currently) use Gaussian KDE:
  - Let D be the dimension of $\mathcal{D}$
  - Let N be the number of samples from π_{in} propagated through Q.
  - Converges at a rate of $\mathcal{O}(N^{-4/(4+D)})$ in mean-squared error.
  - Converges at a rate of $\mathcal{O}(N^{-2/(4+D)})$ in L^1-error.

Bayesian approach:

- Modeling epistemic uncertainties in data.
- Data obtained from a true, but unknown, parameter value, $\lambda^\dagger$.
- Fundamentally solving a different problem.

Definition (Deterministic Forward Problem (DFP))

Given a space Λ, and QoI map Q, the *deterministic forward problem* is to determine the values, $q \in \mathcal{D}$ that satisfy

$$q = Q(\lambda), \; \forall \, \lambda \in \Lambda. \tag{1.6}$$

Definition (Deterministic Inverse Problem (DIP) Under Uncertainty)

Given a noisy datum (or data-vector) $d = q + \xi$, $q \in \mathcal{D}$, the *deterministic inverse problem* is to determine the parameter $\lambda \in \Lambda$ which minimizes

$$\|Q(\lambda) - d\| \tag{1.7}$$

where ξ is a random variable (or vector) drawn from a distribution characterizing the uncertainty in observations due to measurement errors.

- ξ is some unobservable perturbation to the true output.
- ξ arises from epistemic uncertainty (e.g. the precision of available measurement equipment).

The one where we distinguish ourselves from the Bayesian Inverse Problem.

- The *posterior* is a conditional density:

$$\pi_{\text{post}}(\lambda \,|\, d)$$

- π_{post} proportional to the product of π_{prior} and $L_{\mathfrak{D}}$ [3, 2, 1, 4]:

$$\pi_{\text{post}}(\lambda \,|\, d) := \pi_{\text{prior}}(\lambda) \frac{L_{\mathfrak{D}}(q|\lambda)}{C} \tag{1.8}$$

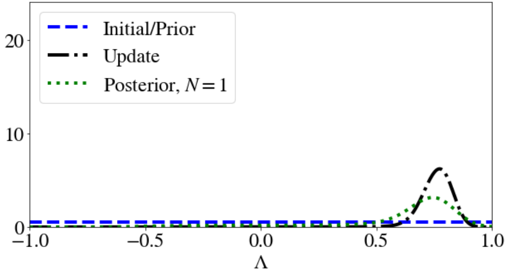
- The *evidence* term C ensures integration to unity. Given by:

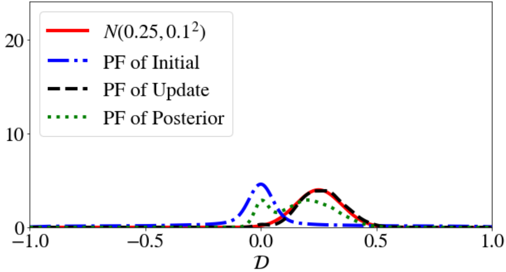
$$C = \int_{\Lambda} \pi_{\text{prior}}(\lambda) \, L_{\mathfrak{D}}(q|\lambda) \, d\lambda$$

The one where we provide an illustrative example.

- Suppose $\Lambda = [-1, 1] \subset \mathbb{R}$ and $Q(\lambda) = \lambda^5$ so that $\mathcal{D} = [-1, 1]$
- $\pi_{\text{in}} \sim \mathcal{U}([-1, 1])$
- $\pi_{\text{ob}} \sim N(0.25, 0.1^2)$
- $d \in \mathcal{D}$ with $d = Q(\lambda^\dagger) + \xi$ where $\xi \sim N(0, 0.1^2)$
- $\pi_{\text{prior}} = \pi_{\text{in}}$ and $d = 0.25$ so $L_{\mathcal{D}} = \pi_{\text{ob}}$

The one where we provide an illustrative example.





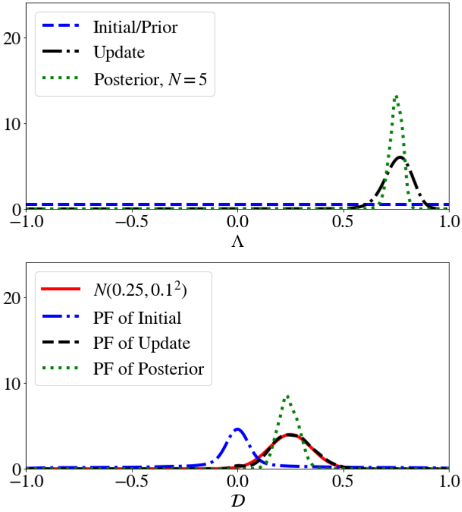
The one where we provide an illustrative example.

What happens as we collect more data?

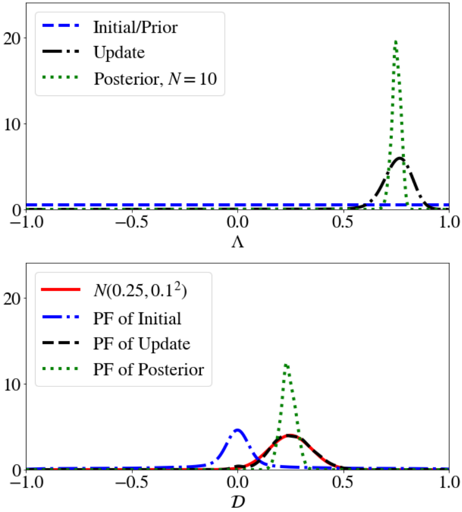
SIP: Use N to estimate mean of observed.

DIP: Likelihood function incorporates more terms.

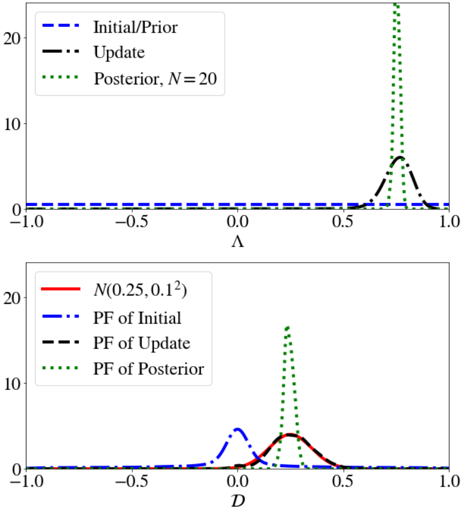
The one where we provide an illustrative example.



The one where we provide an illustrative example.



The one where we provide an illustrative example.



The one where we define the Maximal Updated Density (MUD) point.

$$\lambda^{\mathsf{MUD}} := \arg\max \pi_{\mathsf{up}}(\lambda) \tag{2.1}$$

The one where we create a unifying framework.

- Recall $\|\mathbf{x}\|_C^2 := (\mathbf{x}, \mathbf{x})_C = \mathbf{x}^T C\mathbf{x}$.
- Non-degenerative $\Sigma_{\text{pred}}^{-1}$, Σ_{obs}^{-1}, $\Sigma_{\text{init}}^{-1}$ play the role of C.
- Suppose that $\pi_{\text{in}} = \pi_{\text{prior}} \sim \mathcal{N}(\lambda_0, \Sigma_{\text{init}})$.
- Suppose Q is linear and that $\pi_{\text{ob}} = \pi_{\text{like}} \sim \mathcal{N}(\mathbf{y}, \Sigma_{\text{obs}})$.
- Linearity of Q implies that $Q(\lambda) = A\lambda$ for some $A \in \mathbb{R}^{d\times p}$, and that $\pi_{\text{pr}} \sim \mathcal{N}(Q(\lambda_0), \Sigma_{\text{pred}})$, where

$$\Sigma_{\text{pred}} := A\Sigma_{\text{init}}A^\top. \tag{2.2}$$

The one with the regularization equations.

$$\pi_{\text{post}}(\lambda \mid d) = \frac{\pi_{\text{prior}}(\lambda)\pi_{\text{like}}(d \mid \lambda)}{\int_\Lambda \pi_{\text{like}}(d \mid \lambda)\, \pi_{\text{prior}}(\lambda) d\mu_\Lambda}$$

$$\pi_{\text{up}}(\lambda) = \pi_{\text{in}}(\lambda) \frac{\pi_{\text{ob}}(Q(\lambda))}{\pi_{\text{pr}}(Q(\lambda))}$$

Tikhonov	$T(\lambda) := \lVert Q(\lambda) - \mathbf{y} \rVert^2_{\Sigma_{\text{obs}}^{-1}} + \lVert \lambda - \lambda_0 \rVert^2_{\Sigma_{\text{init}}^{-1}}$
Data-Consistent	$J(\lambda) := T(\lambda) - \lVert Q(\lambda) - Q(\lambda_0) \rVert^2_{\Sigma_{\text{pred}}^{-1}}$

The one where an example highlights a key difference.

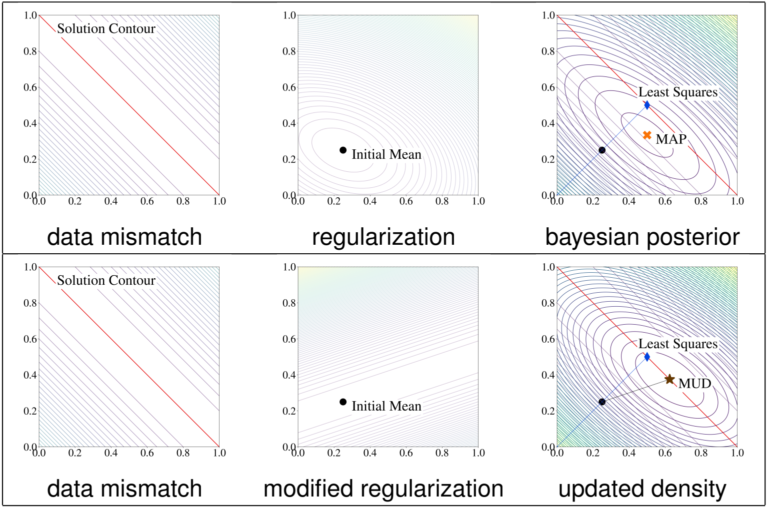
- $A = \begin{bmatrix} 1 & 1 \end{bmatrix}$
- 2-D input, 1-D output $\implies$ rank-deficient
- Details:

$$\lambda_0 = \begin{bmatrix} 0.25 & 0.25 \end{bmatrix}^\top$$

$$\Sigma_{\text{init}} = \begin{bmatrix} 1 & -0.25 \\ -0.25 & 0.5 \end{bmatrix}$$

$$\mathbf{y} = 1, \text{ and } \Sigma_{\text{obs}} = \begin{bmatrix} 0.25 \end{bmatrix}$$

The one that kind of says it all.



- Posterior covariance:

$$\Sigma_{\text{post}} := (A^\top \Sigma_{\text{obs}}^{-1} A + \Sigma_{\text{init}}^{-1})^{-1} \tag{2.3}$$

- Using Woodbury identity and (2.2):

$$\Sigma_{\text{post}} = \Sigma_{\text{init}} - \Sigma_{\text{init}} A^\top \left[\Sigma_{\text{pred}} + \Sigma_{\text{obs}}\right]^{-1} A \Sigma_{\text{init}} \tag{2.4}$$

- Interpretation: Σ_{post} is a rank d correction (or update) of Σ_{init}.
- $\Sigma_{\text{pred}} + \Sigma_{\text{obs}}$ is invertible because it is the sum of two s.p.d matrices.
- Rewrite using analytical expression for the MAP point:

$$\lambda^{\text{MAP}} = \lambda_0 + \Sigma_{\text{post}} A^\top \Sigma_{\text{obs}}^{-1} (\mathbf{y} - b - A\lambda_0). \tag{2.5}$$

The one where we make some convenient manipulations.

- Let

$$R := \Sigma_{\text{init}}^{-1} - A^\top \Sigma_{\text{pred}}^{-1} A. \tag{2.6}$$

- Using this R, rewrite $J(\lambda)$ as

$$J(\lambda) := \|\mathbf{y} - Q(\lambda)\|^2_{\Sigma_{\text{obs}}^{-1}} + \|\lambda - \lambda_0\|^2_R. \tag{2.7}$$

- R is the *effective regularization* in $J(\lambda)$ in the DCI framework:

$$\Sigma_{\text{up}} := \left(A^\top \Sigma_{\text{obs}}^{-1} A + R\right)^{-1} \tag{2.8}$$

- Since R is not invertible, Woodbury's identity cannot be applied (yet).

The one where we make some convenient manipulations.

- *Using linear algebra ...*

$$\Sigma_{\text{up}} = \Sigma_{\text{init}} - \Sigma_{\text{init}} A^\top \Sigma_{\text{pred}}^{-1} \left[\Sigma_{\text{pred}} - \Sigma_{\text{obs}}\right] \Sigma_{\text{pred}}^{-1} A \Sigma_{\text{init}}. \tag{2.9}$$

- Substitute Σ_{up} for Σ_{post} in (2.5):

$$\lambda^{\text{MUD}} = \lambda_0 + \Sigma_{\text{up}} A^\top \Sigma_{\text{obs}}^{-1} (\mathbf{y} - b - A\lambda_0). \tag{2.10}$$

- Substituting (2.9) into (2.10) and simplifying, we have

$$\lambda^{\text{MUD}} = \lambda_0 + \Sigma_{\text{init}} A^\top \Sigma_{\text{pred}}^{-1} (\mathbf{y} - b - A\lambda_0). \tag{2.11}$$

Theorem

Suppose $Q(\lambda) = A\lambda + b$ *for some full rank* $A \in \mathbb{R}^{d\times p}$ *with* $d \leq p$ *and* $b \in \mathbb{R}^d$.

If $\pi_{\text{in}} \sim N(\lambda_0, \Sigma_{init})$, $\pi_{\text{ob}} \sim N(\mathbf{y}, \Sigma_{obs})$, *and the predictability assumption holds, then*

(a) *There exists a unique* λ^{MUD}.

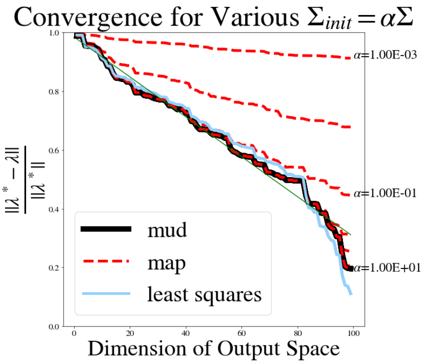
(b) $Q(\lambda^{MUD}) = \mathbf{y}$.

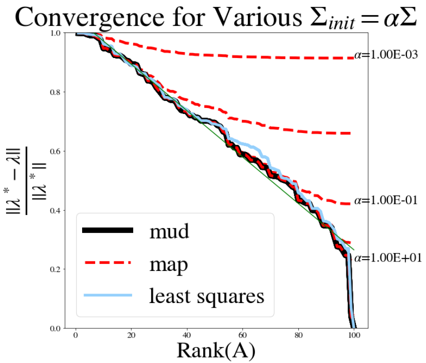
(c) *If* $d = p$, λ^{MUD} *is given by* A^{-1}. *If* $d < p$, λ^{MUD} *is given by* (2.11) *and the covariance associated with this point is given by* (2.9).

The one where we address a key assumption

- Predictability Assumption: π_{pr} is a dominating measure for π_{ob}
- Linear case: involves eigenvalues of covariances:
  - min eigenvalue Σ_{pred} > max eigenvalue Σ_{obs}

The one where we show how rank and dimension impact our solutions.





Example: scaling random diagonal initial covariances

The one where we leverage this framework for general streams of data.

- Measurement devices M_j generating repeated noisy data, $1 \leq j \leq d$.
- $d_{j,i}$ is the ith noisy datum for the jth measurement, where $1 \leq i \leq N_j$.
- Unbiased additive error model for the measurement noise:

$$d_{j,i} = M_j(\lambda^\dagger) + \xi_i, \ \xi_i \sim N(0, \sigma_j^2), \quad 1 \leq i \leq N_j. \tag{2.12}$$

We now construct a d-dimensional vector-valued map from data obtained on the d measurement devices.

The one with the Weighted Mean Error (WME) map $Q_{WME}(\lambda)$.

$$Q_{\mathrm{WME},j}(\lambda) := \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \frac{M_j(\lambda) - d_{j,i}}{\sigma_j}. \tag{2.13}$$

- $Q_{\mathrm{WME},j}(\lambda^\dagger)$ is the sample avg of N_j draws from an i.i.d. $N(0, N_j)$.
- Observed data are generated according to fixed (truth) $\lambda^\dagger$ in (2.12).
- For each component, $Q_{\mathrm{WME},j}(\lambda^\dagger) \sim N(0, 1)$.
- π_{ob} is a $N(\mathbf{0}_{d\times 1}, \mathbf{I}_{d\times d})$ due to the structure of $Q_{\mathrm{WME}}(\lambda)$.

The one where measurements impact the predictability assumption.

- The jth diagonal component of Σ_{pred} is given by the predicted variance associated with using the scalar-valued $Q_{\text{WME},j}$.
- The associated predicted variance for the jth component is given by:

$$\frac{N_j}{\sigma_j^2} M_j \Sigma_{\text{init}} M_j^\top . \tag{2.14}$$

- Σ_{init} non-degenerative and M_j non-trivial row vector, which implies that the **predicted variance grows linearly** with N_j.

The following result is now an immediate consequence of Theorem 2.1.

Corollary

If $\pi_{\text{in}} \sim N(\lambda_0, \Sigma_{init})$ *and data are obtained for* d *linearly independent measurements on* Λ *with an additive noise model with i.i.d. Gaussian noise for each measurement, then* **there exists a minimum number of data points obtained for each of the measurements** *such that there exists a unique* λ^{MUD} *and* $Q_{WME}(\lambda^{MUD}) = 0$.

The one where we violate some assumptions (and see what happens).

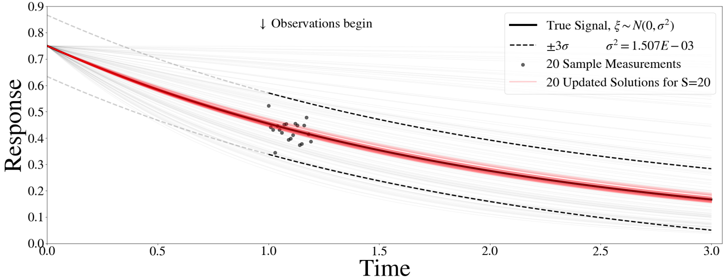
Consider the exponential decay problem with uncertain decay rate λ:

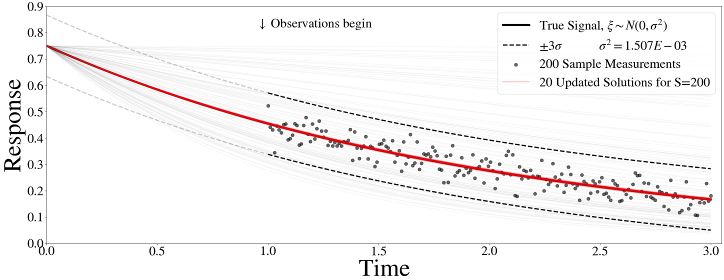
$$\begin{cases} \frac{\partial u}{\partial t} & = \lambda u(t),\ 0 < t \leq 3, \\ u(0) & = 0.75, \end{cases}$$

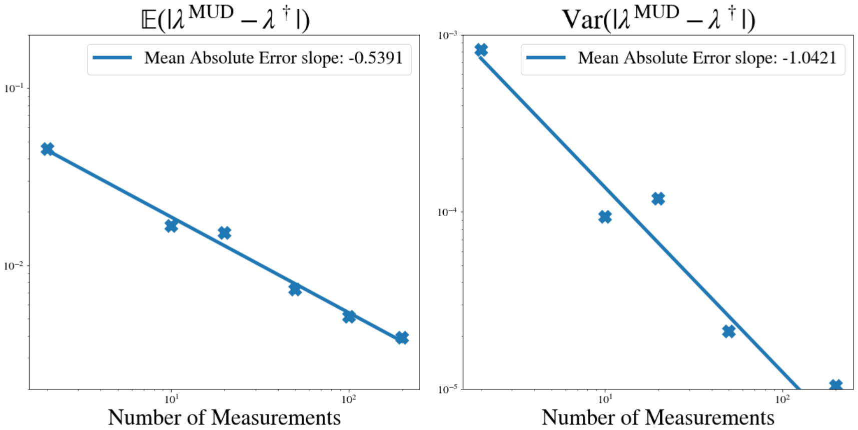
with solution

$$u(t;\lambda) = u_0 \exp(-\lambda t),\ u_0 = 0.75, \tag{2.15}$$

and measurements occur from $t = 1$ until $t = 3$ at rate of 100Hz.


↓ Observations begin
True Signal, $\xi \sim N(0, \sigma^2)$
$\pm 3\sigma$ $\sigma^2 = 1.507E-03$
20 Sample Measurements
20 Updated Solutions for S=20
Response
Time


↓ Observations begin
True Signal, $\xi \sim N(0, \sigma^2)$
$\pm 3\sigma$ $\sigma^2 = 1.507E-03$
200 Sample Measurements
20 Updated Solutions for S=200
Response
Time


$\mathbb{E}(|\lambda^{MUD} - \lambda^{\dagger}|)$
Mean Absolute Error slope: -0.5391
10^{-1}
10^{-2}
10^{1}
10^{2}
Number of Measurements
$\mathrm{Var}(|\lambda^{MUD} - \lambda^{\dagger}|)$
Mean Absolute Error slope: -1.0421
10^{-3}
10^{-4}
10^{-5}
10^{1}
10^{2}
Number of Measurements

The one where we violate some assumptions (and see what happens).

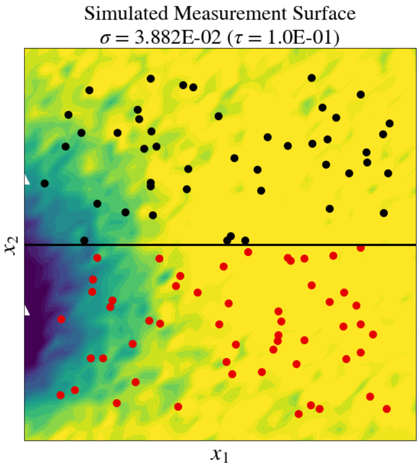
Consider the Poisson problem:

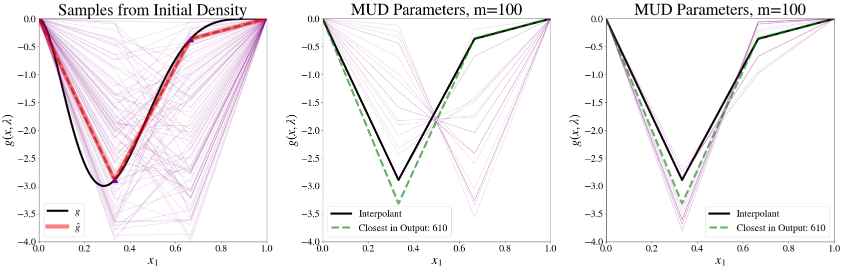
$$\begin{cases} -\nabla \cdot \nabla u &= f(x), \quad \text{on } x \in \Omega, \\ u &= 0, \quad \text{on } \Gamma_T \cup \Gamma_B, \\ \frac{\partial u}{\partial \mathbf{n}} &= g(x_2), \quad \text{on } \Gamma_L, \\ \frac{\partial u}{\partial \mathbf{n}} &= 0, \quad \text{on } \Gamma_R, \end{cases} \tag{2.16}$$

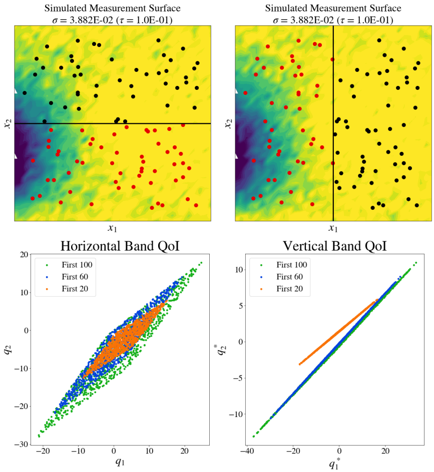
where $x = (x_1, x_2) \in \Omega = (0,1)^2$ is the spatial domain.

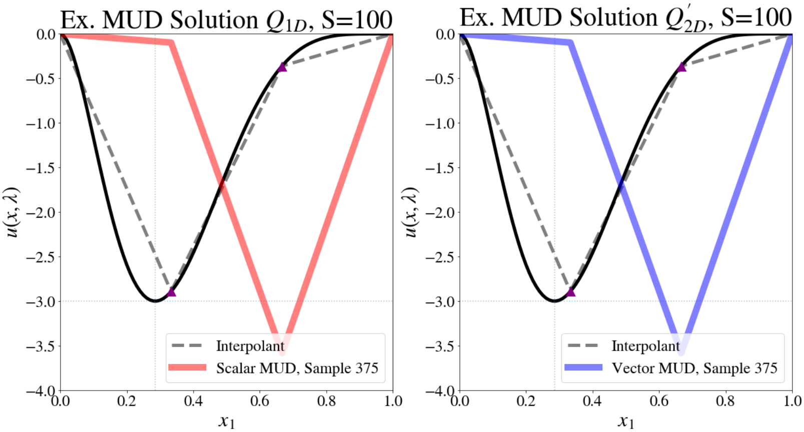
- Γ_T, Γ_B, Γ_L, and Γ_R, denote the top, bottom, left, and right boundaries.
- The outward normal derivative is denoted by $\frac{\partial u}{\partial \mathbf{n}}$.
- The forcing function is $f = 10 \exp\left(\|x - 0.5\|^2 / 0.02\right)$.

- $g(x_2)$ is uncertain parameter, i.e., λ defines an uncertain function.
- To generate the noisy data, we use $g(x_2) \propto x_2^2(x_2 - 1)^5$.
- Constant of proportionality chosen so $\min g = -3$ at $x_2 = \frac{2}{7}$.
- Piecewise-linear finite elements on a triangulation of a 36×36 mesh.
- 100 randomly placed sensors in the subdomain $(0.05, 0.95)^2 \subset \Omega$.
- Repeated 20 times to study variation due to realizations of noisy data.
- Limited to $N = 1000$ samples from initial density.

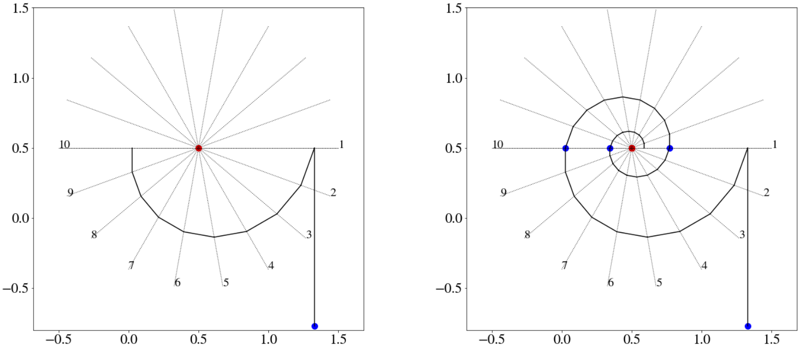

Simulated Measurement Surface
σ = 3.882E-02 (τ = 1.0E-01)
x2
x1


Samples from Initial Density
g
$\hat{g}$
MUD Parameters, m=100
Interpolant
Closest in Output: 610
MUD Parameters, m=100
Interpolant
Closest in Output: 610
$g(x, \lambda)$
x_1
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−4.0
0.2
0.4
0.6
0.8
1.0

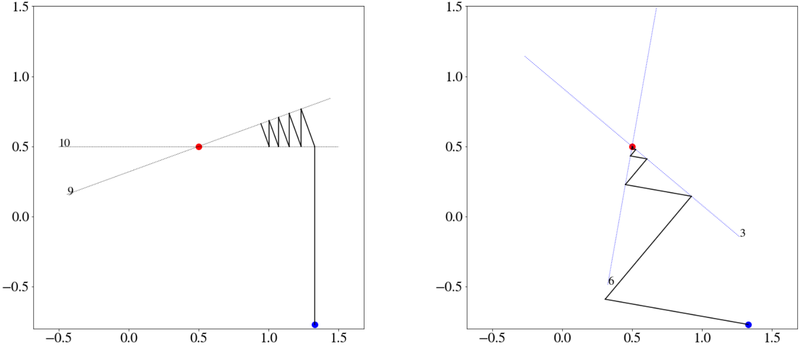

Simulated Measurement Surface
$\sigma = 3.882E\text{-}02$ ($\tau = 1.0E\text{-}01$)
x_1
x_2
Horizontal Band QoI
Vertical Band QoI
First 100
First 60
First 20
q_1
q_2
q_1^*
q_2^*

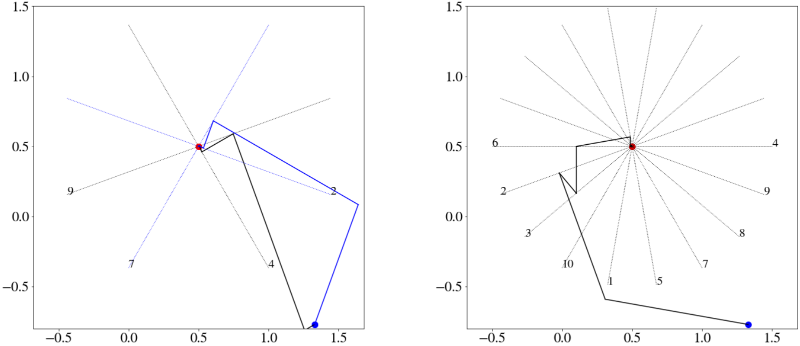

Ex. MUD Solution Q_{1D}, S=100
Ex. MUD Solution Q'_{2D}, S=100
$u(x, \lambda)$
x_1
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−4.0
0.0
0.2
0.4
0.6
0.8
1.0
Interpolant
Scalar MUD, Sample 375
Vector MUD, Sample 375

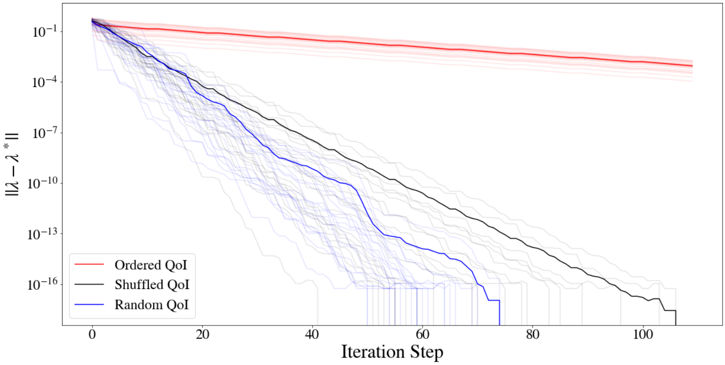
The one with the small problems in many batches.



QoI defined by 10 equispaced rotations of the unit vector $[0, 1]$ through the first two Euclidean quadrants.


1.5
1.0
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
1.5
10
9
3
6


1.5
1.0
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
1.5
9
2
7
4
1.5
1.0
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
1.5
6
4
2
9
3
8
10
7
1
5



10^{-1}
10^{-4}
10^{-7}
10^{-10}
10^{-13}
10^{-16}
$\|\lambda - \lambda^*\|$
Ordered QoI
Shuffled QoI
Random QoI
0
20
40
60
80
100
Iteration Step

The one where we convince you to trust our numerics.

- Public repository hosted on Github.com (`github.com/mathematicalmichael/thesis`)
- Github Actions implements Continuous Integration / Deployment
- Each change is validated for reproducibility
- `makefile` for convenience (`make <filename>`)
  - dissertation + presentation (LaTeX, themes, style files)
  - every example, convergence result (Python)
  - every image in every figure
- PyPi published implementation of main methods: `pip install mud`
- Unit tests aid in ensuring integrity of functions
- Docker guarantees software runtime (ran on `x86` and `arm`) `docker pull mathematicalmichael/python:thesis(latex:thesis)`

M. Allmaras, W. Bangareth, J.M. Linhart, J. Polanco, F. Wang, K. Wang, J. Webster, and S. Zedler.
Estimating parameters in physical models through Bayesian inversion: A complete example.
2013.

J.O. Berger.
Statistical Decision Theory and Bayesian Analysis.
Springer-Verlag, 1985.

S. Myers R. Walpole, R. Myers and K. Ye.
Probability & Statistics for Engineers & Scientists.
Pearson Education, 2007.

Ralph C. Smith.
Uncertainty Quantification: Theory, Implementation, and Applications.
Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2013.