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

Michael Pilosov

Advisor: Troy Butler

University of Colorado Denver

October 29, 2020

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”

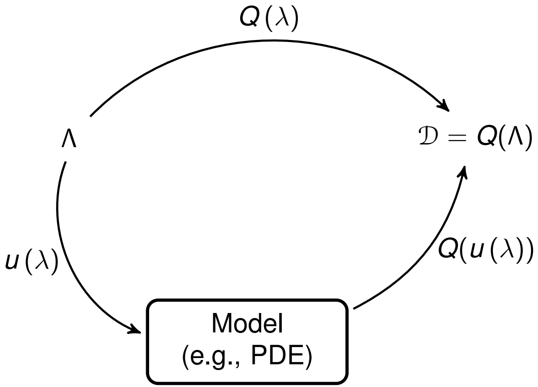
The one where we define the letters we use and what they mean.

- 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 inputs 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* $\mathbb{P}_{\text{in}}$.
If a dominating measure μ_Λ exists on $(\Lambda, \mathcal{B}_\Lambda)$, the *initial distribution* π_{in} is given by the Radon-Nikodym derivative of $\mathbb{P}_{\text{in}}$ w.r.t the measure μ_Λ.

Definition (Predicted Distribution)

The *predicted distribution* (or density) is the push-forward density of π_{in} under the map Q, and is denoted as π_{pr}.
Given as the Radon-Nikodym derivative (w.r.t $\mu_{\mathcal{D}}$) of the 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 density* π_{ob}.

The one where we define the solution to the SIP.

We now have all of the definitions required to summarize the density-based solution to the SIP, known as the *updated density* as:

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

Practical Considerations

- We approximate π_{pr} using density estimation on forward propagation of samples from π_{in}
- May evaluate π_{up} directly for any sample of Λ (one model solve)
- Accuracy of the computed updated density is proportional to accuracy of approximation of the predicted density
- 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
- Stable w.r.t. perturbations in the Total Variation metric

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

Bayesian approach: modeling epistemic uncertainties in data on a QoI obtained from a true, but unknown, parameter value, $\lambda^\dagger$.

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

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

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

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.

In the above definition, ξ is some unobservable perturbation to the true output, arising 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, denoted by $\pi_{\text{post}}(\lambda \mid d)$, proportional to the product of the prior and data-likelihood function [3, 2, 1, 4]:

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

where we emphasize the use of π_{post} to distinguish the *posterior* from the updated density π_{up} in (1.5).

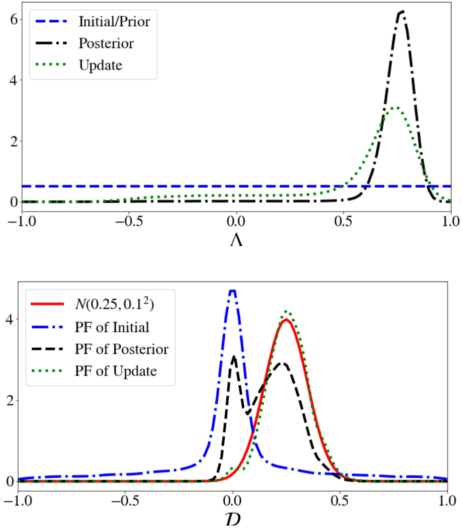
The *evidence* term C ensures the posterior density integrates to one:

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

The one where we provide an illustrative example.

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

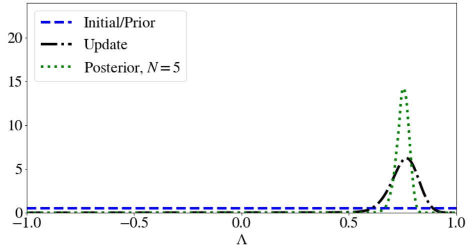


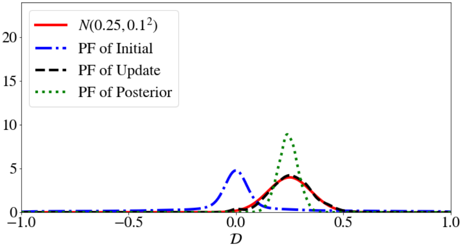
The one where we provide an illustrative example.

What happens as we collect more data?

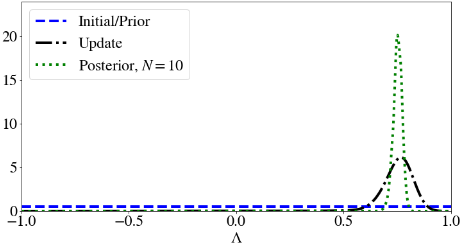
SIP: Use N to estimate mean of observed
DIP: likelihood function incorporates more data

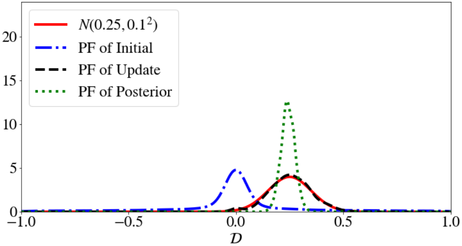
The one where we provide an illustrative example.



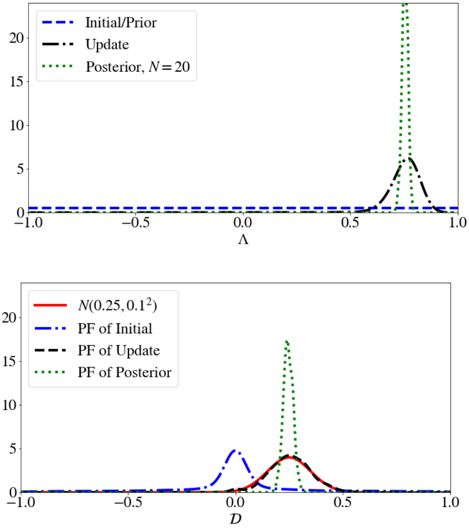


The one where we provide an illustrative example.





The one where we provide an illustrative example.



We formally define the maximal updated density (MUD) point as

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

The one where we create a unifying framework.

Let $\|\mathbf{x}\|_C^2 := (\mathbf{x}, \mathbf{x})_C = \mathbf{x}^T C \mathbf{x}$.
Inverse covariances associated with non-degenerative multivariate Gaussian distributions will play the role of C.
Suppose that the initial and prior densities are both given by the same $\mathcal{N}(\lambda_0, \Sigma_{\text{init}})$ distribution.

Additionally, suppose the map Q is linear and that the data-likelihood and observed densities are both given by the same $\mathcal{N}(\mathbf{y}, \Sigma_{\text{obs}})$ distribution.

The linearity of Q implies that $Q(\lambda) = A\lambda$ for some $A \in \mathbb{R}^{d \times p}$, and that the predicted density follows a $\mathcal{N}(Q(\lambda_0), \Sigma_{\text{pred}})$ distribution where

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

The one with the regularization equations.

$\pi_{up}(\lambda) = \pi_{in}(\lambda)\dfrac{\pi_{ob}(Q(\lambda))}{\pi_{pr}(Q(\lambda))}$	$\pi_{post}(\lambda \mid d) = \dfrac{\pi_{prior}(\lambda)\pi_{like}(d \mid \lambda)}{\int_\Lambda \pi_{like}(d \mid \lambda)\pi_{prior}(\lambda) d\mu_\Lambda}$
Tikhonov	$T(\lambda) := \Vert Q(\lambda) - \mathbf{y} \Vert^2_{\Sigma^{-1}_{obs}} + \Vert \lambda - \lambda_0 \Vert^2_{\Sigma^{-1}_{init}}$
Data-Consistent	$J(\lambda) := T(\lambda) - \Vert Q(\lambda) - Q(\lambda_0) \Vert^2_{\Sigma^{-1}_{pred}}$

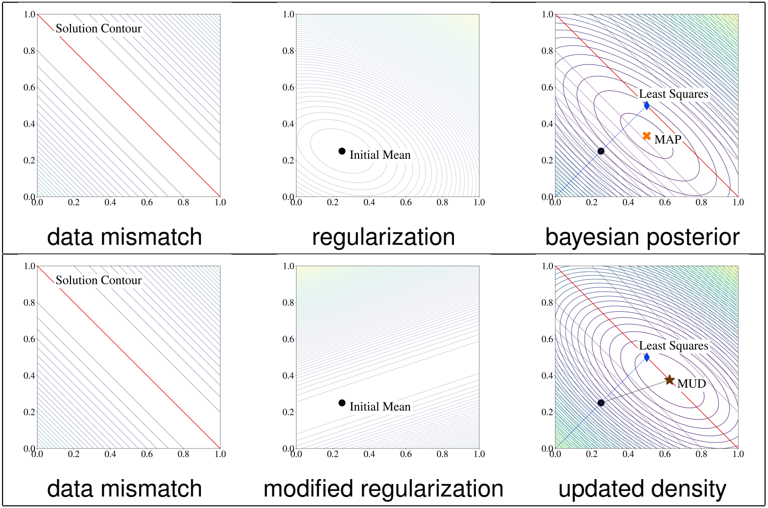
The one where an example highlights a key difference.

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

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

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

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


Solution Contour
Initial Mean
Least Squares
MAP
data mismatch
regularization
bayesian posterior
Solution Contour
Initial Mean
Least Squares
MUD
data mismatch
modified regularization
updated density

The posterior covariance is formally given by

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

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} as 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.
We rewrite the closed form expression for the MAP poing given in [5] as

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

We define

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

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

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

In this form, we identify R as the *effective regularization* in $J(\lambda)$ due to the formulation in the data-consistent 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).

We derive using several identities

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

Substitute Σ_{up} for Σ_{post} in (2.5) to write the point that minimizes J as:

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

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

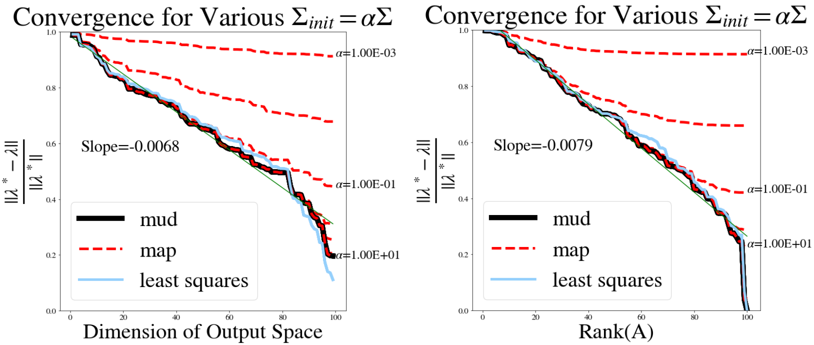
(c) *If* $d = p$, λ^{MUD} *is given by* A^{-1}. *If* $d < p$, λ^{MUD} *is given by* (2.11) *and the covariance associated with this point is given by* (2.9).

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.

Suppose $\exists$ d measurement devices generating repeated noisy data.

For each $1 \leq j \leq d$, denote by $\mathcal{M}_j(\lambda^\dagger)$ the jth measurement device.
N_j is number of noisy data obtained for $\mathcal{M}_j(\lambda^\dagger)$.

$d_{j,i}$ is the ith noisy datum for the jth measurement, where $1 \leq i \leq N_j$.

Assume an unbiased additive error model for the measurement noise, with independent identically distributed (i.i.d.) Gaussian errors so that

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

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.

The weighted mean error (WME) map, denoted by $Q_{\text{WME}}(\lambda)$ has jth component, denoted by $Q_{\text{WME},j}(\lambda)$, given by

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

$Q_{\text{WME},j}(\lambda^\dagger)$ is the sample avg of N_j random draws from an i.i.d. $N(0, N_j)$. By assumption, the observed data are generated according to the fixed true physical parameter vector given by $\lambda^\dagger$ in (2.12).

Subsequently, each component of $Q_{\text{WME}}(\lambda^\dagger)$ is a random draw from an $N(0, 1)$ distribution.
Therefore, with this choice of data-defined QoI map, we specify π_{ob} as a $N(\mathbf{0}_{d\times 1}, \mathbf{I}_{d\times d})$ distribution.

The one where measurements impact the predictability assumption.

The jth diagonal component of the predicted covariance matrix is given by the predicted variance associated with using the scalar-valued $Q_{\text{WME},j}$.
Then, the associated predicted variance is given by

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

Since Σ_{init} is assumed to be non-degenerative and M_j is a non-trivial row vector, this predicted variance grows linearly with N_j.

In other words, the jth diagonal component of the predicted covariance has the form $\beta_j N_j$ for some $\beta_j > 0$.

Let $N_{\text{min},j}$ denote the minimum N_j for $1 \leq j \leq N$ necessary to make the jth diagonal components sufficiently large so that the smallest eigenvalue of the predicted covariance is larger than 1.

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.

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 begin at $t = 1$.

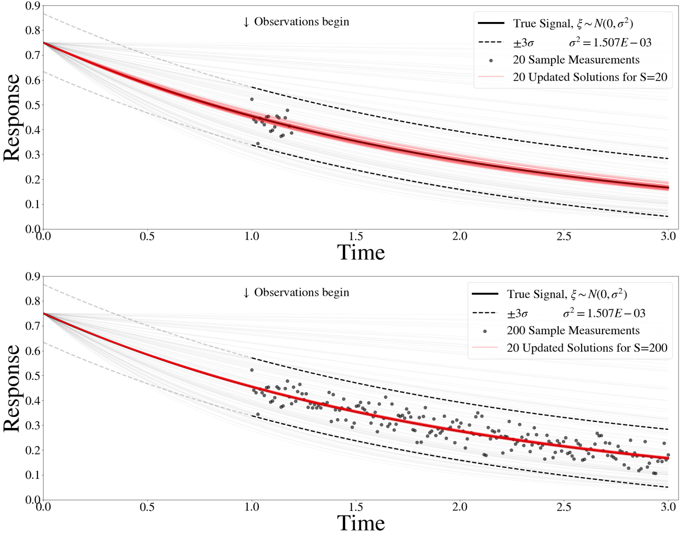


Figure: Curves for exponential decay model with various decay coefficients. Dashed curves denote 99% probability intervals for noise. The true signal is

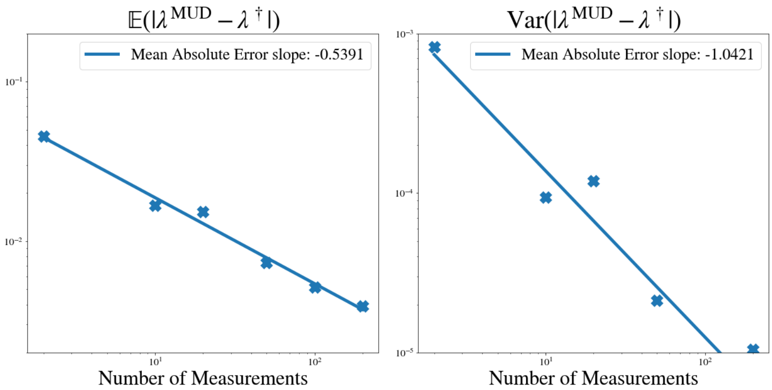


Figure: The mean (left) and variance (right) of absolute errors in MUD estimates as a function of the number of data points used. These statistics are computed over 20 trials.

The one where we violate some assumptions.

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 of this domain
- $\frac{\partial u}{\partial \mathbf{n}}$ denotes the usual outward normal derivative.
- The forcing function f is taken to be $10 \exp\left(\|x - 0.5\|^2 / 0.02\right)$.

- $g(x_2)$ is uncertain parameter, i.e., the parameter is 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 the subsequent variation in MUD points due to different realizations of noisy data.

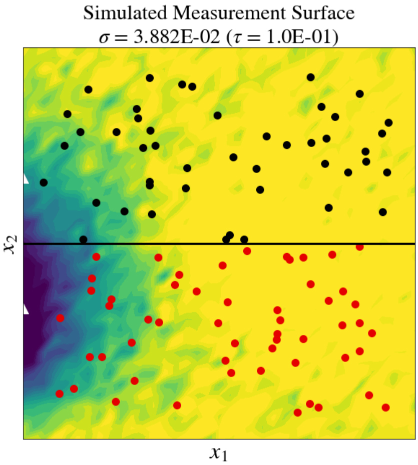


Figure: A representative noisy perturbation of the reference response surface. Locations of the randomly chosen spatial data used to construct both Q_{1D} and Q_{2D} are shown as black and red dots.

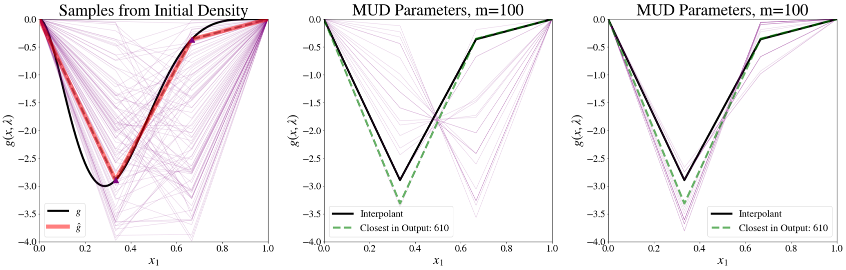


Figure: In the left plot, the reference $g(x_2)$ is shown as the solid black curve with its interpolant onto the spline basis shown as a dotted red curve. The dashed blue line represents the sample from parameter space which most closely predicts noiseless data, which we refer to as the projection of g. The purple curves in the center and right plots show the variability in MUD estimates of $g(x_2)$ for the 20 different realizations of noisy data. The center plot uses Q_{1D} and the right plot uses Q_{2D} to construct the MUD estimates.

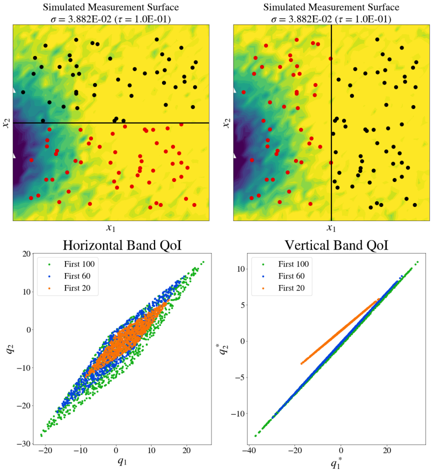


Figure: $N = 1000$ parameter evaluations for both methods of partitioning Ω.

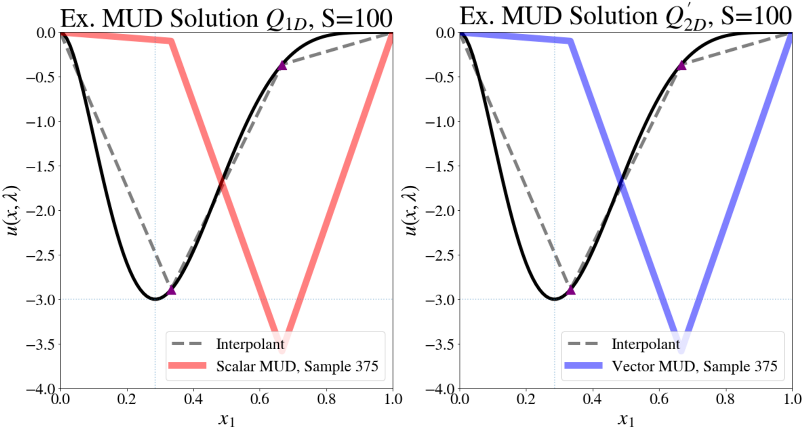


Figure: Side-by-side comparison of an example solution to the SIP using Q_{1D} (left) compared to using Q'_{2D}, juxtaposed against a plot of g.

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

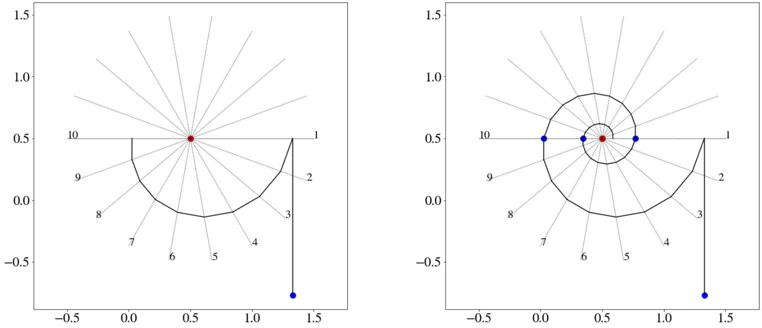
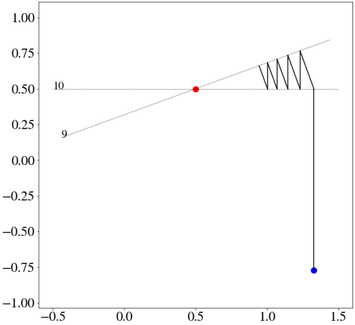


Figure: Dotted lines show the solution contours for each row of the operator A. (Left): First epoch for iterating through 10 QoI. (Right): Three more epochs allows our estimate to get much closer to the true value.



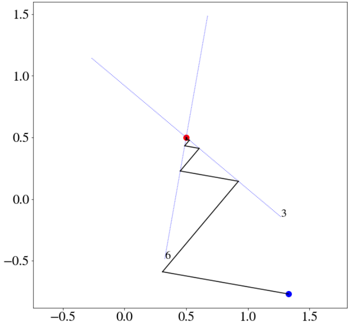
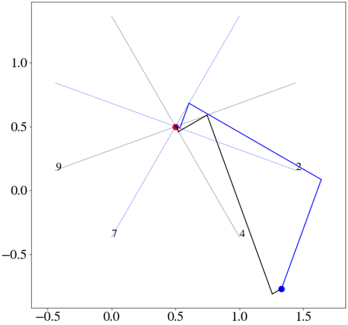


Figure: Iterating through five epochs of two QoI, each formed by picking two of the ten available rows of A at random. The random directions chosen on the left exhibit more redundancy than those on the right, so the same amount of iteration results in less accuracy.



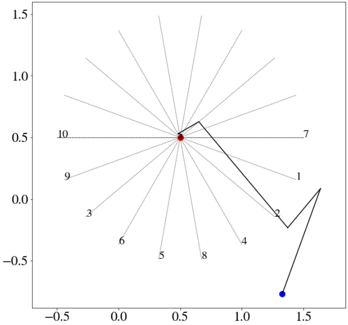


Figure: (Left): Subsets of available QoI components can be chosen to exhibit minimal redudancy and lead to expedited convergence. (Right): Random components of the QoI map used for each iterative step. This leads to an overall similar level of precision in this example, without the need to use gradients.

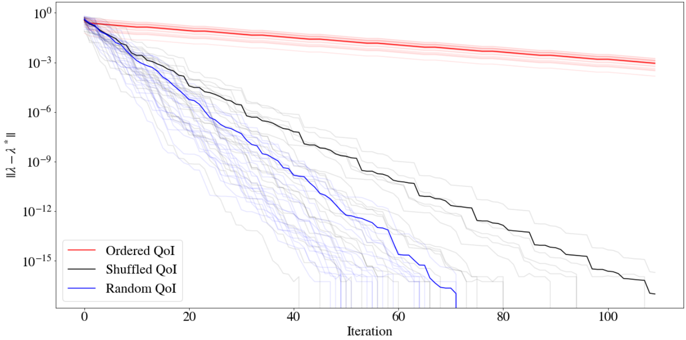


Figure: 20 initial means are chosen and iterated on for three approaches for ordering QoI. Individual experiments are transparent and the mean error is shown as a solid line for each approach.

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.
- Albert Tarantola.

Inverse Problem Theory and Methods for Model Parameter Estimation.
siam, 2005.