

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

Michael Pilosov

Advisor: Troy Butler

University of Colorado Denver

October 31, 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 (aleatoric, 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.*

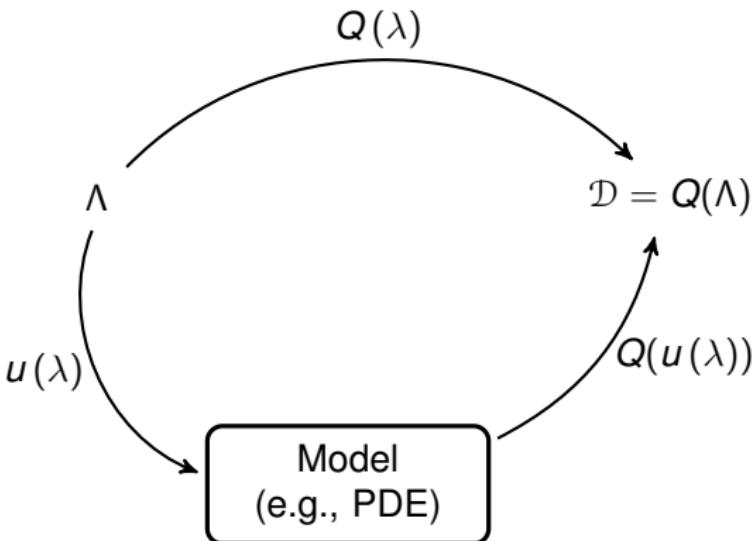
- State variable:  $u$  (e.g. heat, energy, pressure, deflection)
- Parameters:  $\lambda$  (e.g. source term, diffusion, boundary data)
- Model:  $\mathcal{M}(u, \lambda) = 0$ , so  $u(\lambda)$
- Quantity of Interest (QoI) map, (piecewise smooth):

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

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

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



## 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

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

## 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

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

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  $\mu_\Lambda$  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*  $\mathbb{P}_{in}$ .

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



## Definition (Predicted Distribution)

The *predicted distribution* (or density) is the push-forward density of  $\pi_{\text{in}}$  under the map  $Q$ , and is denoted as  $\pi_{\text{pr}}$ .

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

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



## 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*  $\pi_{\text{ob}}$ .



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

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

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



*The one with some practical considerations.*

- Approximate  $\pi_{\text{pr}}$  using density estimation using samples from  $\pi_{\text{in}}$ .
- May evaluate  $\pi_{\text{up}}$  directly for any sample of  $\Lambda$  (one model solve).
- Accuracy proportional to that of approximation to the predicted density.
- We (currently) use Gaussian KDE:
  - » Let  $D$  be the dimension of  $\mathcal{D}$
  - » Let  $N$  be the number of samples from  $\pi_{\text{in}}$  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.
- Data obtained from a true, but unknown, parameter value,  $\lambda^\dagger$ .
- Fundamentally solving a different problem.

### Definition (Deterministic Forward Problem (DFP))

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

$$q = Q(\lambda), \forall \lambda \in \Lambda. \quad (1.6)$$



*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  $\xi$  is a random variable (or vector) drawn from a distribution characterizing the uncertainty in observations due to measurement errors.

- $\xi$  is some unobservable perturbation to the true output.
- $\xi$  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, denoted by  $\pi_{\text{post}}(\lambda | d)$ .
- $\pi_{\text{post}}$  proportional to the product of  $\pi_{\text{prior}}$  and  $L_{\mathcal{D}}$  [3, 2, 1, 4]:

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

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

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

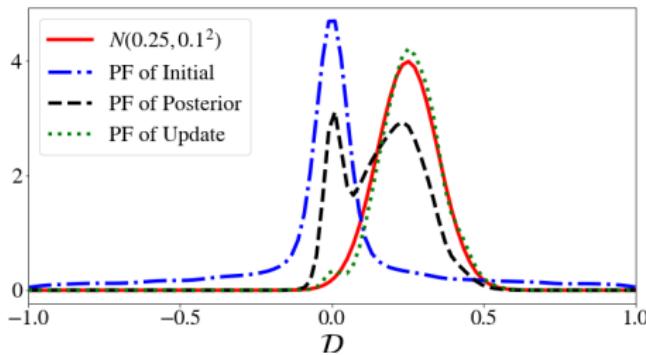
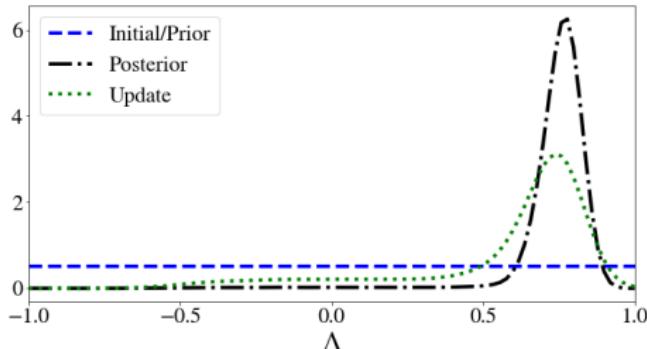


*The one where we provide an illustrative example.*

- 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.*

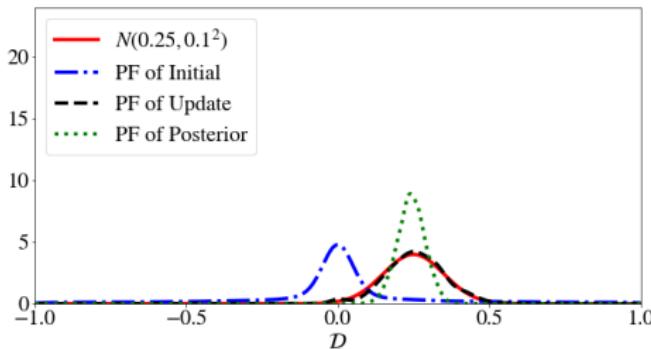
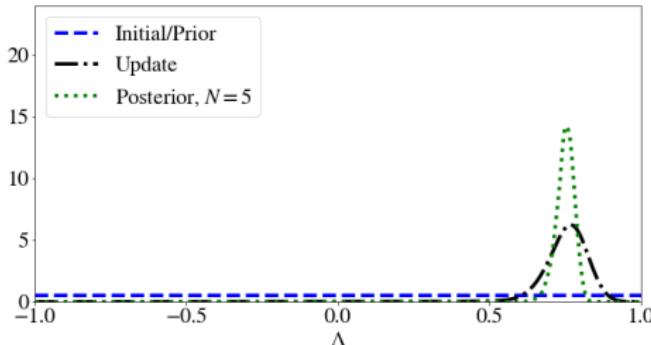
*What happens as we collect more data?*

SIP: Use  $N$  to estimate mean of observed

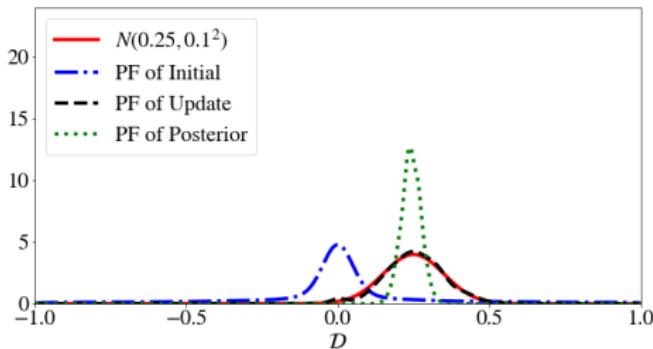
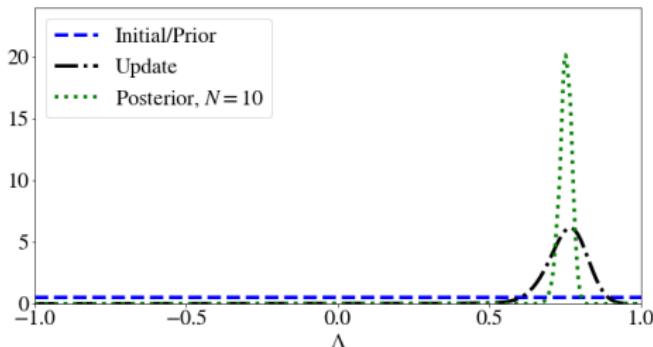
DIP: likelihood function incorporates more data



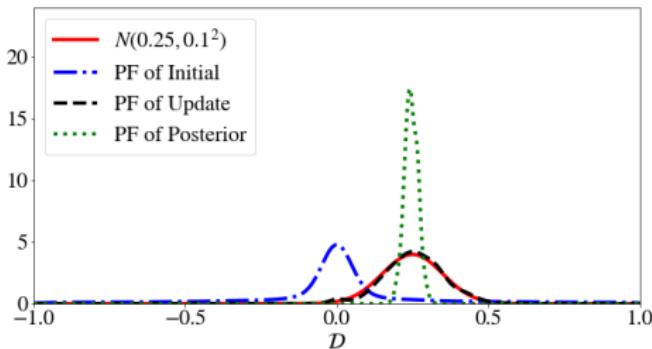
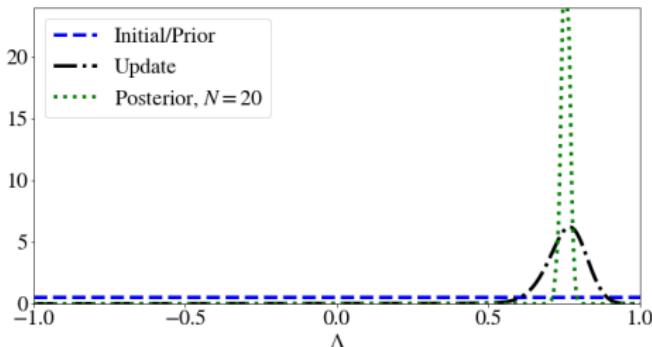
*The one where we provide an illustrative example.*



*The one where we provide an illustrative example.*



*The one where we provide an illustrative example.*



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

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



*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  $\Sigma_{\text{pred}}^{-1}$ ,  $\Sigma_{\text{obs}}^{-1}$ ,  $\Sigma_{\text{init}}^{-1}$  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

$$\Sigma_{\text{pred}} := A \Sigma_{\text{init}} A^\top. \quad (2.2)$$



*The one with the regularization equations.*

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

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

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



*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:

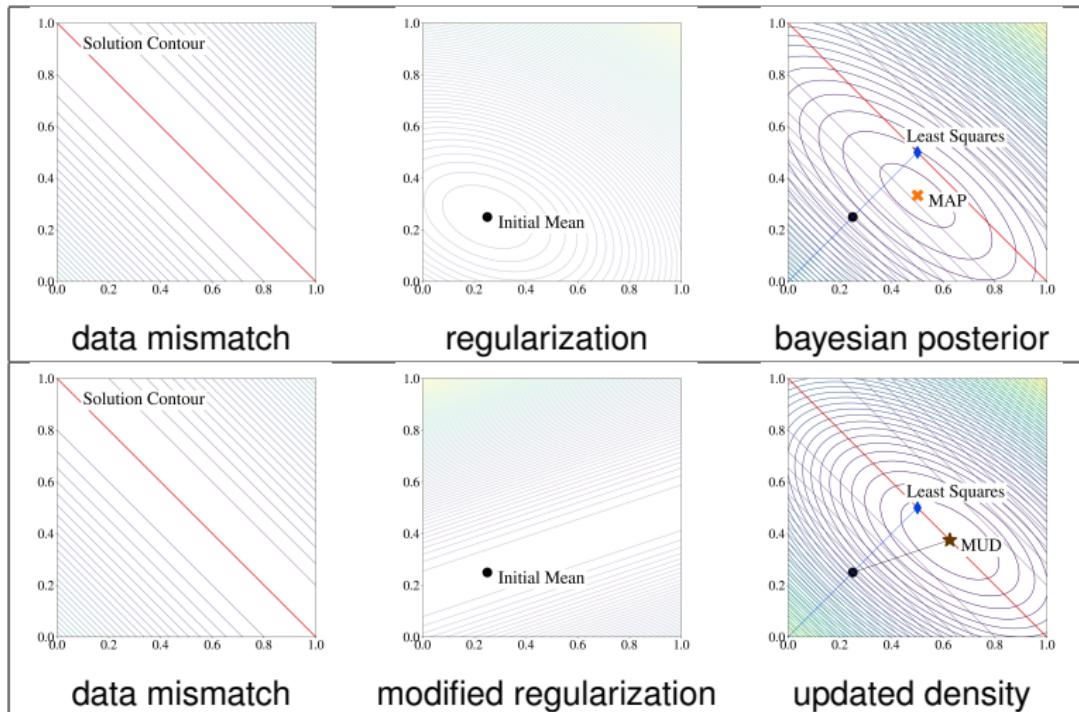
$$\lambda_0 = \begin{bmatrix} 0.25 & 0.25 \end{bmatrix}^\top$$

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

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



*The one that kind of says it all.*



- Posterior covariance:

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

- Using Woodbury identity and (2.2):

$$\Sigma_{\text{post}} = \Sigma_{\text{init}} - \Sigma_{\text{init}} \mathbf{A}^{\top} [\Sigma_{\text{pred}} + \Sigma_{\text{obs}}]^{-1} \mathbf{A} \Sigma_{\text{init}} \quad (2.4)$$

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

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



*The one where we make some convenient manipulations.*

- Let

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

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

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

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

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

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

*The one where we make some convenient manipulations.*

- *Using linear algebra ...*

$$\Sigma_{\text{up}} = \Sigma_{\text{init}} - \Sigma_{\text{init}} A^\top \Sigma_{\text{pred}}^{-1} [\Sigma_{\text{pred}} - \Sigma_{\text{obs}}] \Sigma_{\text{pred}}^{-1} A \Sigma_{\text{init}}. \quad (2.9)$$

- Substitute  $\Sigma_{\text{up}}$  for  $\Sigma_{\text{post}}$  in (2.5):

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

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

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



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

- (a) There exists a unique  $\lambda^{\text{MUD}}$ .
- (b)  $Q(\lambda^{\text{MUD}}) = \mathbf{y}$ .
- (c) If  $d = p$ ,  $\lambda^{\text{MUD}}$  is given by  $A^{-1}$ . If  $d < p$ ,  $\lambda^{\text{MUD}}$  is given by (2.11) and the covariance associated with this point is given by (2.9).



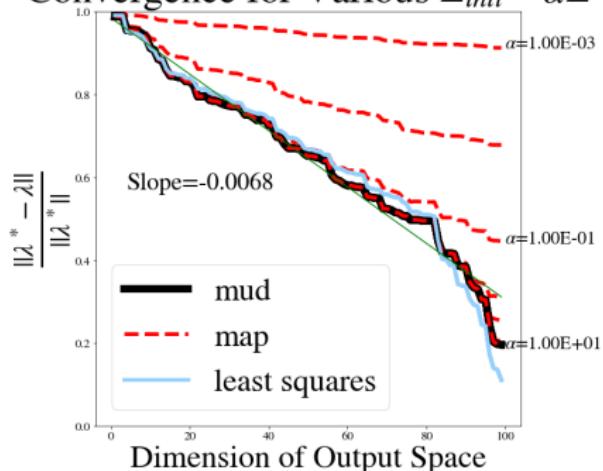
*The one where we address a key assumption*

- Predictability Assumption:  $\pi_{\text{pr}}$  is a dominating measure for  $\pi_{\text{ob}}$
- Linear case: involves eigenvalues of covariances:
  - » min eigenvalue  $\Sigma_{\text{pred}} >$  max eigenvalue  $\Sigma_{\text{obs}}$

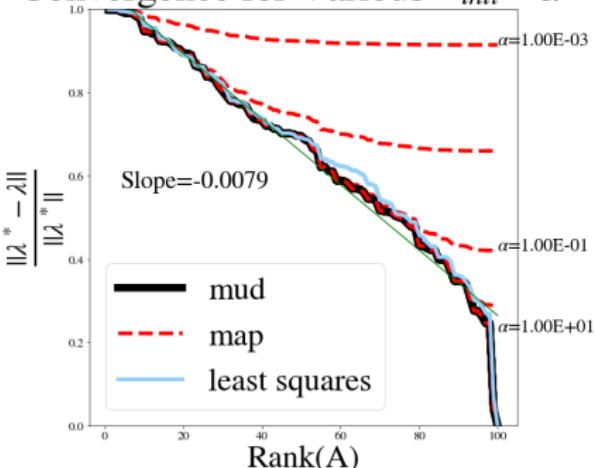


The one where we show how rank and dimension impact our solutions.

Convergence for Various  $\Sigma_{init} = \alpha \Sigma$



Convergence for Various  $\Sigma_{init} = \alpha \Sigma$



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  $j$ th measurement device.
- $N_j$  is number of noisy data obtained for  $\mathcal{M}_j(\lambda^\dagger)$ .
- $d_{j,i}$  is the  $i$ th noisy datum for the  $j$ th measurement, where  $1 \leq i \leq N_j$ .
- Unbiased additive error model for the measurement noise.
- Noise is independent identically distributed (i.i.d.) Gaussian errors, so

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

*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  $j$ th component, denoted by  $Q_{\text{WME},j}(\lambda)$ , given by

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

- $Q_{\text{WME},j}(\lambda^\dagger)$  is the sample avg of  $N_j$  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  $\pi_{\text{ob}}$  as a  $N(\mathbf{0}_{d \times 1}, \mathbf{I}_{d \times d})$  distribution.

*The one where measurements impact the predictability assumption.*

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

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

- $\Sigma_{\text{init}}$  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  $\pi_{\text{in}} \sim N(\lambda_0, \Sigma_{\text{init}})$  and data are obtained for  $d$  linearly independent measurements on  $\Lambda$  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  $\lambda^{\text{MUD}}$  and  $Q_{\text{WME}}(\lambda^{\text{MUD}}) = 0$ .**



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

Consider the exponential decay problem with uncertain decay rate  $\lambda$ :

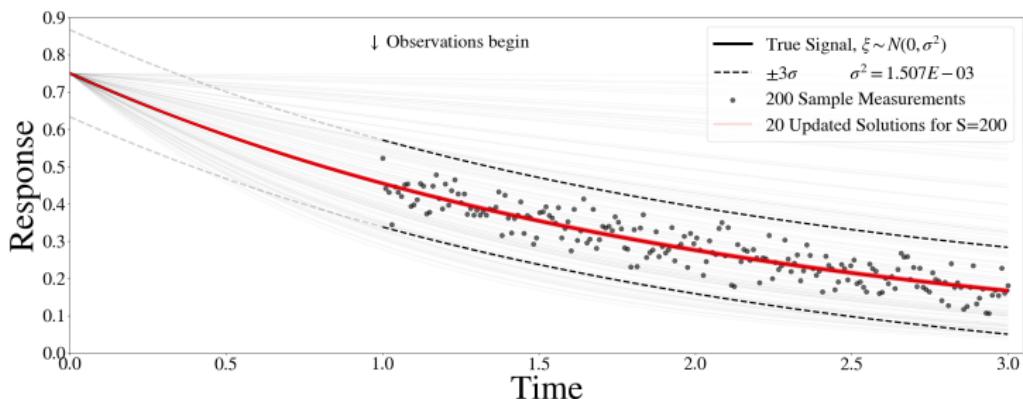
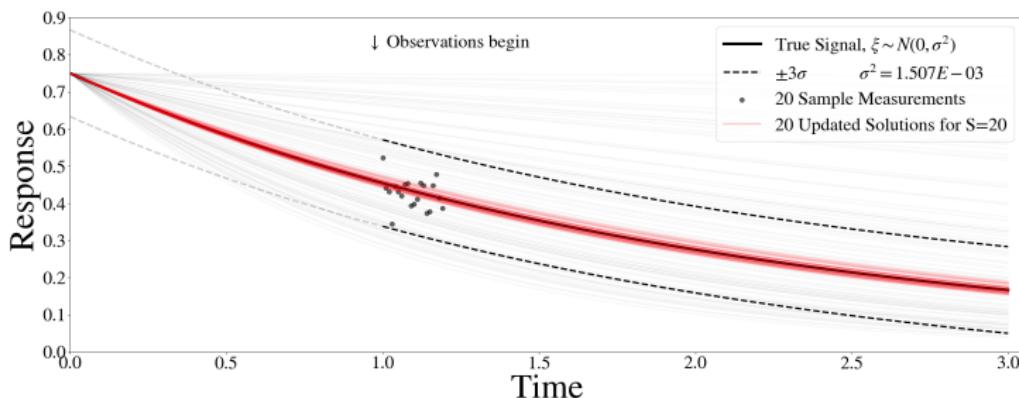
$$\begin{cases} \frac{\partial u}{\partial t} &= \lambda u(t), \quad 0 < t \leq 3, \\ u(0) &= 0.75, \end{cases}$$

with solution

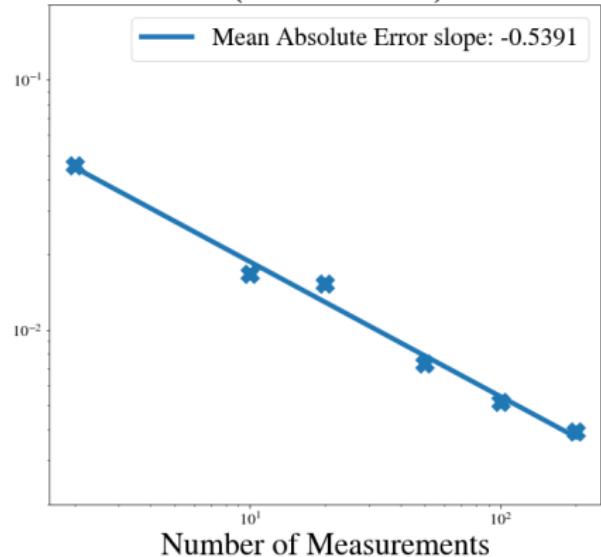
$$u(t; \lambda) = u_0 \exp(-\lambda t), \quad u_0 = 0.75, \tag{2.15}$$

and measurements begin at  $t = 1$ , continuing until  $t = 3$ .

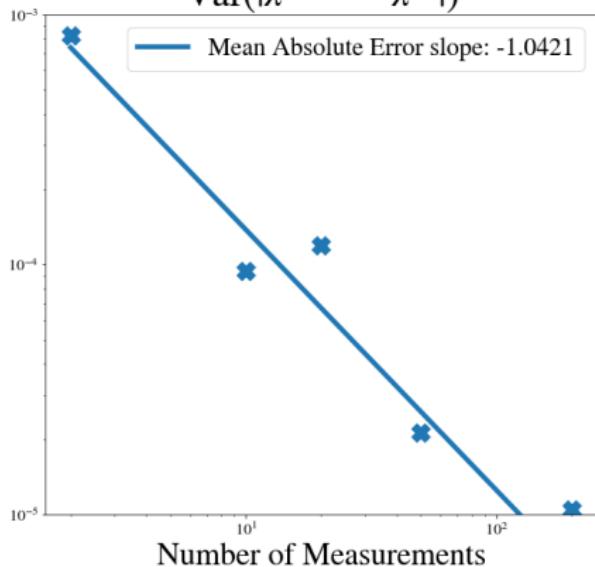




$$\mathbb{E}(|\lambda^{\text{MUD}} - \lambda^\dagger|)$$



$$\text{Var}(|\lambda^{\text{MUD}} - \lambda^\dagger|)$$



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

Consider the Poisson problem:

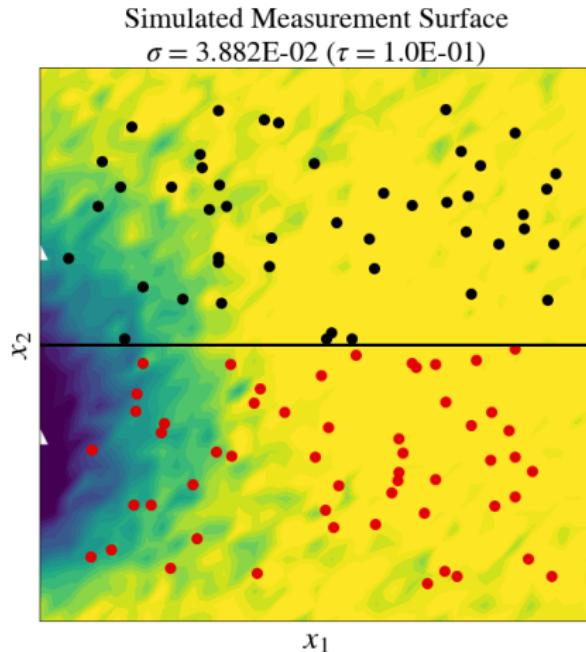
$$\begin{cases} -\nabla \cdot \nabla u = f(x), & \text{on } x \in \Omega, \\ u = 0, & \text{on } \Gamma_T \cup \Gamma_B, \\ \frac{\partial u}{\partial \mathbf{n}} = g(x_2), & \text{on } \Gamma_L, \\ \frac{\partial u}{\partial \mathbf{n}} = 0, & \text{on } \Gamma_R, \end{cases} \quad (2.16)$$

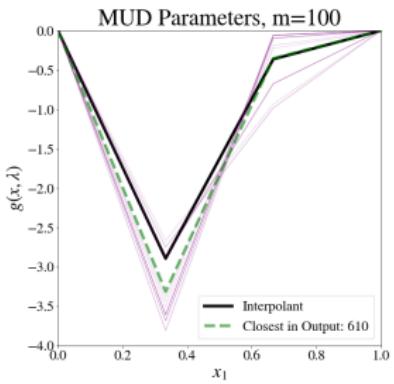
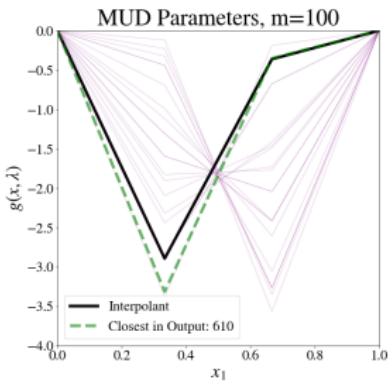
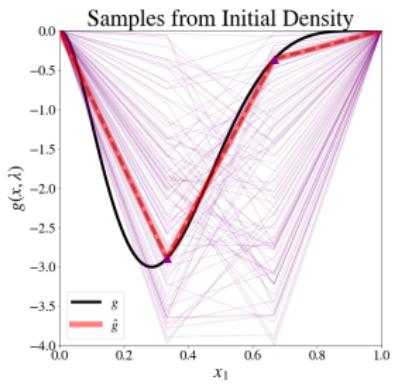
where  $x = (x_1, x_2) \in \Omega = (0, 1)^2$  is the spatial domain.

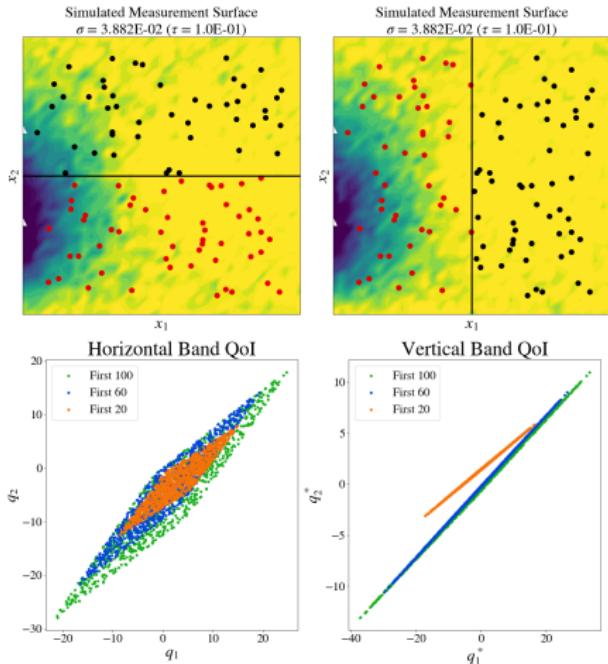
- $\Gamma_T$ ,  $\Gamma_B$ ,  $\Gamma_L$ , and  $\Gamma_R$ , denote the top, bottom, left, and right boundaries.
- The outward normal derivative is denoted by  $\frac{\partial u}{\partial \mathbf{n}}$ .
- The forcing function is  $f = 10 \exp \left( \|x - 0.5\|^2 / 0.02 \right)$ .

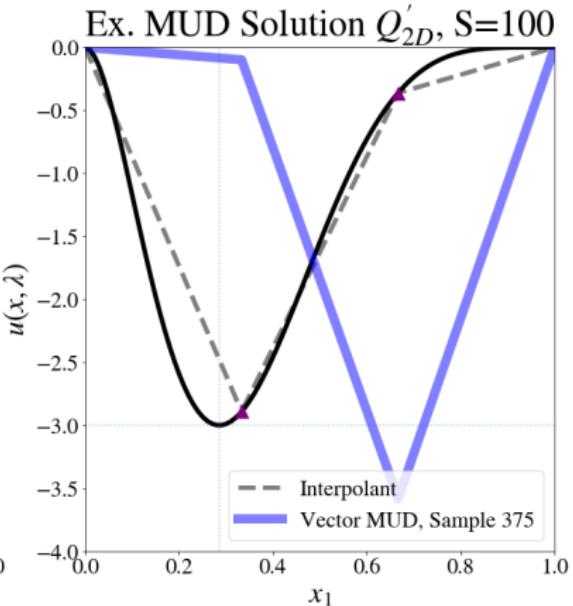
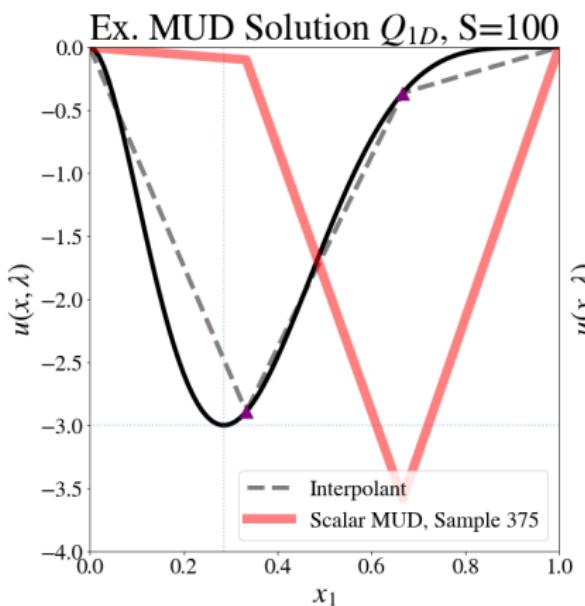


- $g(x_2)$  is uncertain parameter, i.e.,  $\lambda$  defines an uncertain function.
- To generate the noisy data, we use  $g(x_2) \propto x_2^2(x_2 - 1)^5$ .
- Constant of proportionality chosen so  $\min g = -3$  at  $x_2 = \frac{2}{7}$ .
- Piecewise-linear finite elements on a triangulation of a  $36 \times 36$  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.

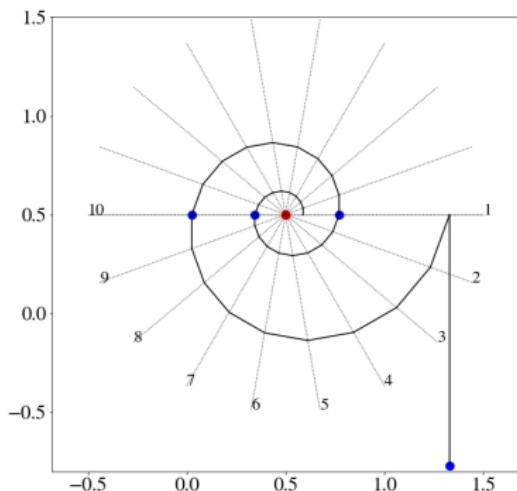
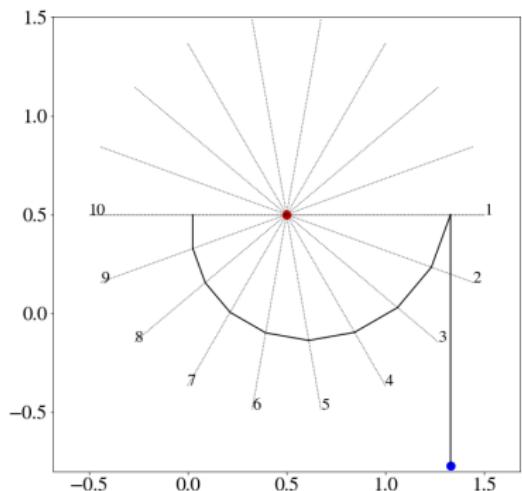




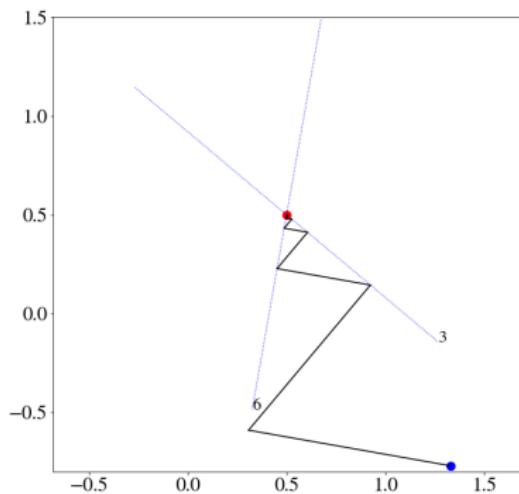
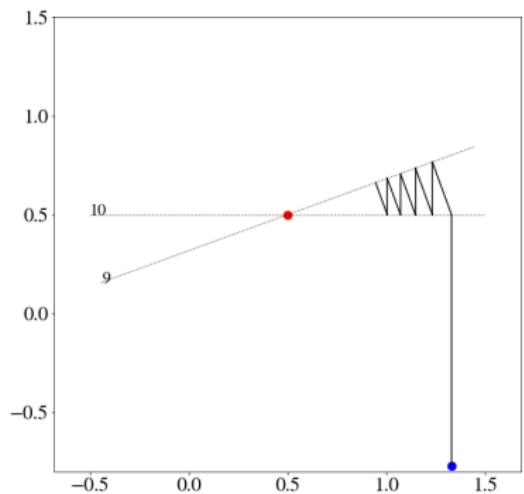


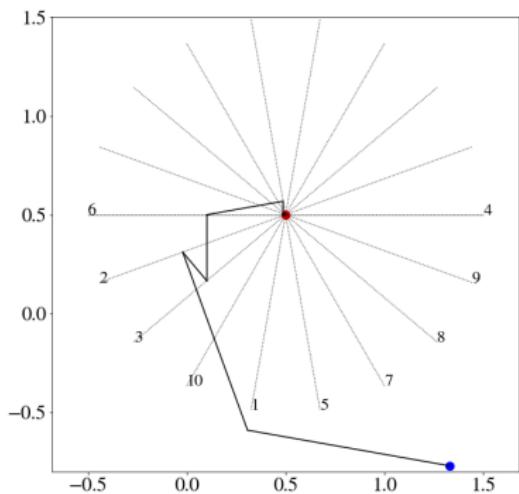
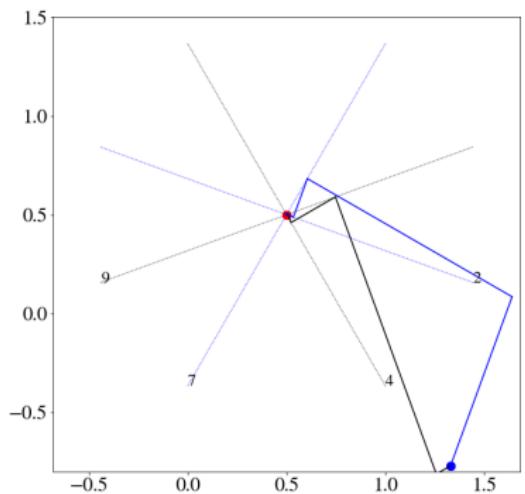


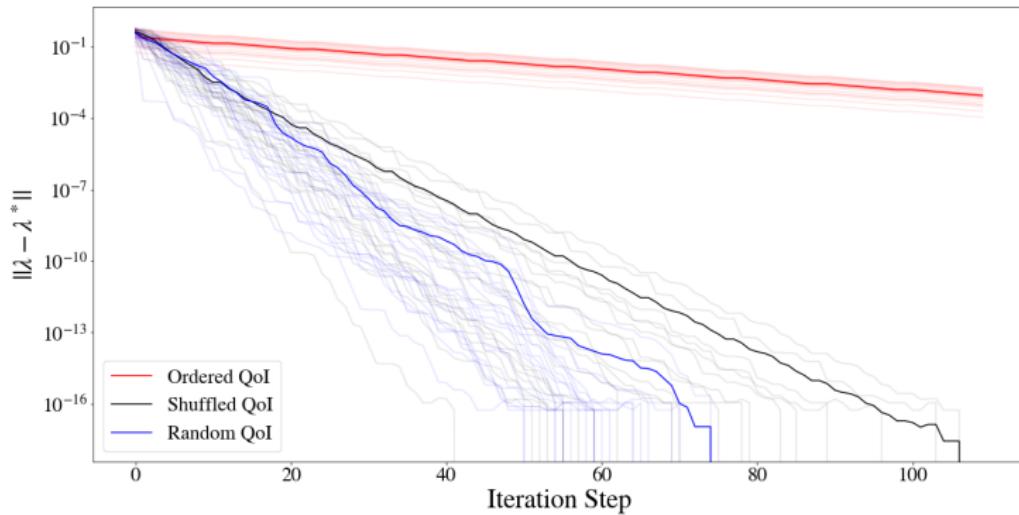
*The one with the small problems in many batches.*



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







*The one where we convince you to trust our numerics.*

- Public repository hosted on Github.com  
([github.com/mathematicalmichael/thesis](https://github.com/mathematicalmichael/thesis))
- Github Actions implements Continuous Integration / Deployment
- Each change is validated for reproducibility
- makefile for convenience (`make <filename>`)
  - » dissertation + presentation (L<sup>A</sup>T<sub>E</sub>X, 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.