How Geometric Features of Quantity of Interest Maps Impact Errors and Convergence of Sample-Based Solutions to Stochastic Inverse Problems

Michael Pilosov

Advisor: Troy Butler

University of Colorado Denver

October 22, 2020

Broad Goals of Uncertainty Quantification

Why does any of this matter?

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

Notation and Terminology

- State variable: u (e.g. heat, energy, pressure, deflection)
- Parameters: λ (e.g. source term, diffusion, boundary data)
- Deterministic model: $\mathcal{M}(u, \lambda) = 0$,

$$\mathcal{M} : \lambda \to u(\lambda)$$

- Quantity of Interest map (QoI) - at least pcw differentiable
  - Functional of the solution

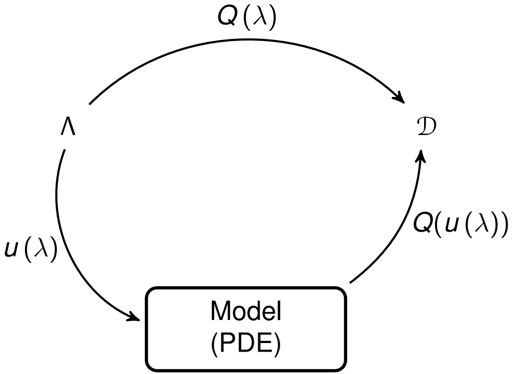
$$q : u(\lambda) \to \mathbb{R}$$

  - Can be vector valued

$$Q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_d \end{bmatrix}$$

  - $Q(\lambda) := Q(u(\lambda))$

A QoI Map relates Inputs to Outputs



Defining the Quantity of Interest Map

Definition (Inverse Problem)

Given $\mathbb{P}_{\text{ob}}$ on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ the **inverse problem** is to determine a probability measure $\mathbb{P}_{\text{up}}$ on $(\Lambda, \mathcal{B}_{\Lambda})$ such that

$$\mathbb{P}_{\text{up}}(Q^{-1}(E)) = \mathbb{P}_{\text{ob}}(E), \tag{1.1}$$

for all events $E \in \mathcal{B}_{\mathcal{D}}$, where

$$\pi_{\text{up}} = \frac{d\,\mathbb{P}_{\text{up}}}{d\mu_{\Lambda}} \quad \text{and} \quad \pi_{\text{ob}} = \frac{d\,\mathbb{P}_{\text{ob}}}{d\mu_{\mathcal{D}}}.$$

(1.1) defines a "Consistent Solution," yields a Consistency Condition.

Note: We use the notation $\mathbb{P}$ and π throughout this work to relate measures to their associated densities (i.e., Radon-Nikodym derivatives), which exist under the assumption of a dominating (Lebesgue) volume measure μ.

Perspectives

- We seek the $\mathbb{P}_{\text{up}}$ whose push-forward measure matches $\mathbb{P}_{\text{ob}}$
- In measure-theoretic terms, $\mathbb{P}_{\text{up}}$ is a pull-back measure of $\mathbb{P}_{\text{ob}}$

Definition (Observed Density)

The density π_{ob} in (1.1) represents the uncertainty in QoI data.

Definition (Initial Density)

π_{in} encodes prior beliefs about λ's (before evidence is accounted for)

Summarizing

- "Push-forward" **initial** beliefs using Q to **compare** to **observed** (data)
- Solve forward problem to construct solution to inverse problem
- The push-forward density of π_{in} under the map Q is denoted by π_{pr}

Definition (Predicted Density)

π_{pr} is given as the Radon-Nikodym derivative (with respect to $\mu_{\mathcal{D}}$) of the push-forward probability measure defined by:

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

The Updated Density solves the SIP

These definitions are combined to form the **updated density**:

$$\pi_{\text{up}}(\lambda) = \pi_{\text{in}}(\lambda) \frac{\pi_{\text{ob}} Q(\lambda)}{\pi_{\text{pr}} Q(\lambda)}, \;\; \lambda, \in \Lambda. \tag{1.3}$$

- π_{ob} and π_{pr} defined on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ are evaluated at $Q(\lambda)$
- The map Q impacts the structure of the update
- $\mathcal{D}$ itself depends on Q
- Primary effort in solving for π_{up} (in (1.3)) requires constructing π_{pr}
- This is because π_{in} and π_{ob} are given *a priori* (often parametric)
- Updated derived through use of Disintegration Theorem in [**?**]
- Existence and Uniqueness given a predictability assumption

Assumption (Predictability Assumption)

The measure associated with π_{ob} is absolutely continuous with respect to the measure associated with π_{pr}.

The requirement is guaranteed if the following is satisfied:

$$\exists\, C > 0 \text{ s.t. } \pi_{\text{ob}}(d) \leq C\, \pi_{\text{pr}}(d) \text{ for a.e. } d \in \mathcal{D}, \tag{1.4}$$

where $d = Q(\lambda)$ for some $\lambda \in \Lambda$. By [**?**], if (1) holds, we have:

Theorem (Existence and Uniqueness)

For any set $A \in \mathcal{B}_\Lambda$, the solution $\mathbb{P}_{\text{up}}$ given defined by

$$\mathbb{P}_{\text{up}}(A) = \int_{\mathcal{D}} \left(\int_{\Lambda \in Q^{-1}(d)} \pi_{\text{in}}\, \lambda \frac{\pi_{\text{ob}}(d)}{\pi_{\text{pr}}(d)}\, d\mu_{\Lambda,d}\lambda \right) d\mu_{\mathcal{D}}(d),\ \forall\, A \in \mathcal{B}_\Lambda \tag{1.5}$$

is a consistent solution, and is unique up to choice of $\mathbb{P}_{\text{in}}$ on $(\Lambda, \mathcal{B}_\Lambda)$.

Stability

All the stability and convergence results presented are with respect to:

Definition

Total Variation / Statistical Distance

$$d_{\mathrm{TV}}(\mathbb{P}_f, \mathbb{P}_g) := \int |f - g| \, d\mu, \tag{1.6}$$

where f, g are the densities (Radon-Nikodym derivatives with respect to μ) associated with measures $\mathbb{P}_f, \mathbb{P}_g$, respectively.

Definition (Stability of Updates I)

We say that $\mathbb{P}_{\text{up}}$ is *stable* with respect to perturbations in $\mathbb{P}_{\text{ob}}$ if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$d_{\text{TV}}(\mathbb{P}_{\text{ob}}, \widehat{\mathbb{P}_{\text{ob}}}) < \delta \implies d_{\text{TV}}(\mathbb{P}_{\text{up}}, \widehat{\mathbb{P}_{\text{up}}}) < \varepsilon. \tag{1.7}$$

In [**?**], it is shown that $d_{\text{TV}}(\widehat{\mathbb{P}_{\text{up}}}, \mathbb{P}_{\text{up}}) = d_{\text{TV}}(\widehat{\mathbb{P}_{\text{ob}}}, \mathbb{P}_{\text{ob}})$, implying that:

Theorem

$\mathbb{P}_{\text{up}}$ *is stable with respect to perturbations in* $\mathbb{P}_{\text{ob}}$*.*

Definition (Stability of Updates II)

Let $\{\mathbb{P}_{\Lambda,d}\}_{d\in\mathcal{D}}$ and $\left\{\widehat{\mathbb{P}_{\Lambda,d}}\right\}_{d\in\mathcal{D}}$ be the conditional probabilities defined by the disintegration of $\mathbb{P}_{\text{in}}$ and $\widehat{\mathbb{P}_{\text{in}}}$, respectively.
We say that $\mathbb{P}_{\text{up}}$ is *stable* with respect to perturbations in $\mathbb{P}_{\text{in}}$ if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that for almost every $d \in \text{supp}(\pi_{\text{ob}})$,

$$d_{\text{TV}}(\mathbb{P}_{\Lambda,d}, \widehat{\mathbb{P}_{\Lambda,d}}) < \delta \implies d_{\text{TV}}(\mathbb{P}_{\text{up}}, \widehat{\mathbb{P}_{\text{up}}}) < \varepsilon. \tag{1.8}$$

Theorem

$\mathbb{P}_{\text{up}}$ *is stable with respect to perturbations in the initial.*

Properties of the Updated Density

- Taken together, these stability results provide assurances that the updated we obtain is accurate up to the level of experimental error polluting π_{ob} and error in incorrectly specifying initial assumptions.
- Given that specifying the definition of a “true” initial is somewhat nebulous, we are less interested in the consequences of the latter conclusion.
- Generating samples from π_{pr} requires a numerical approximation to π_{pr}, which introduces **additional errors** in π_{pr}.

- Let $\widehat{\pi_{\text{pr}}}$ be a computational approximation to π_{pr} and $\widehat{\pi_{\text{up}}}$ the associated approximate updated π_{up}
- The conditional densities from the Disintegration theorem are

$$\frac{\widehat{d\mathbb{P}_{\Lambda,d}}}{d\mu_{\Lambda,d}(\lambda)} = \frac{\pi_{\text{in}}(\lambda)}{\widehat{\pi_{\text{pr}}(d)}}$$

- To approximate the push-forward of the initial density, we require:

Assumption

There exists some $C > 0$ such that

$$\pi_{\text{ob}}(d) \leq C\widehat{\pi_{\text{pr}}(d)} \text{ for a.e. } d \in \mathcal{D}.$$

Assumption

There exists some $C > 0$ such that

$$\pi_{\text{ob}}(d) \leq C\widehat{\pi_{\text{pr}}(d)} \text{ for a.e. } d \in \mathcal{D}.$$

If this assumption is satisfied, we can prove the following citeBJW18:

Theorem

The error in the approximate updated is:

$$d_{TV}(\mathbb{P}_{\text{up}}, \widehat{\mathbb{P}_{\text{up}}}) \leq C d_{TV}(\mathbb{P}_{\text{pr}}, \widehat{\mathbb{P}_{\text{pr}}}), \tag{1.9}$$

where the C is the constant taken from (3)*.*

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 is 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

The one where we distinguish ourselves from the Bayesians.

The one with the regularization equations.

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

Table: The λ which minimizes these functionals also maximizes the updated PDF (left) and the Bayesian posterior PDF (right).
$T(\lambda)$ is the typical functional often associated with Tikhonov regularization.
The $J(\lambda)$ has an additional term subtracted from $T(\lambda)$ coming from the predicted density that serves as “unregularization” in data–informed directions.

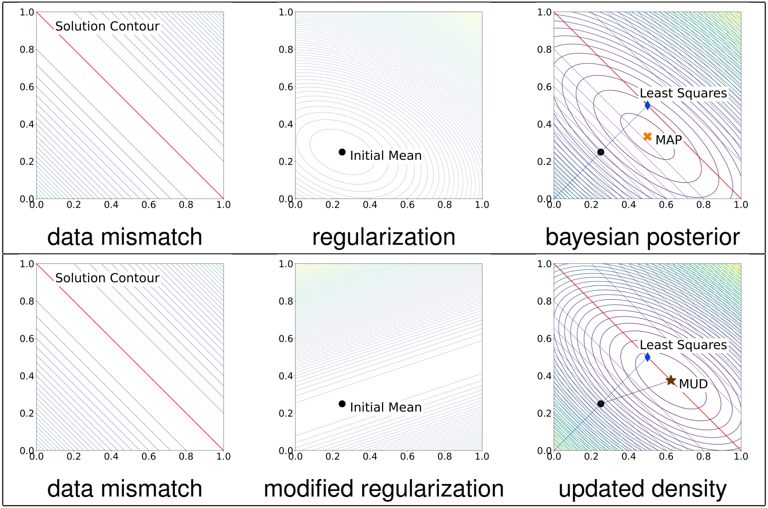


Figure: Gaussian data mismatch for a 2-to-1 linear map (left plots). Gaussian initial/prior induce different regularization terms (middle plots), which leads to different optimization functions (right plots) and parameter estimates.

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

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

The one where we show that our approach works even when some assumptions are violated.

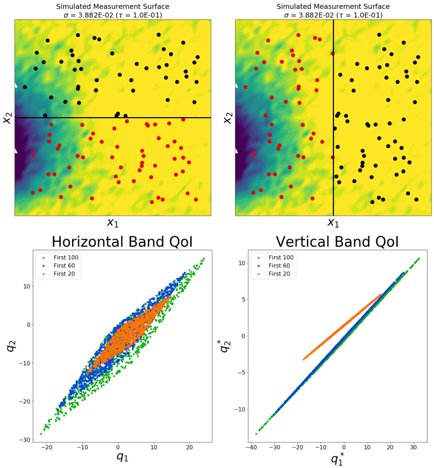


Figure: $N = 1000$ parameter evaluations for both methods of partitioning Ω.

The one with the small problems in large batches.

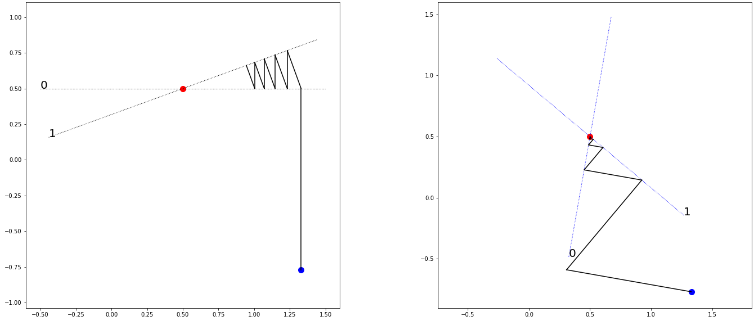
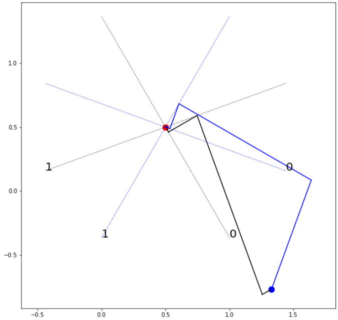


Figure: Iterating through five epochs of two QoI, each formed by picking two of the ten available rows of A at random. The random directions chosen on the left exhibit more redundancy than those on the right, so the same amount of iteration results in less accuracy.

The one with the small problems in large batches.



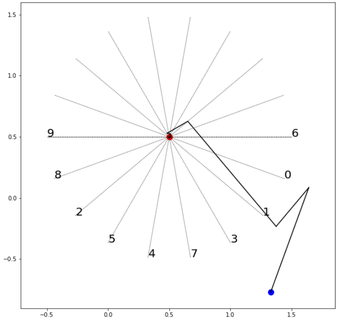


Figure: If we are careful with how we construct our maps or choose our iteration strategy, we can achieve considerably more accurate solutions with the same computational cost.

The one with the small problems in large batches.

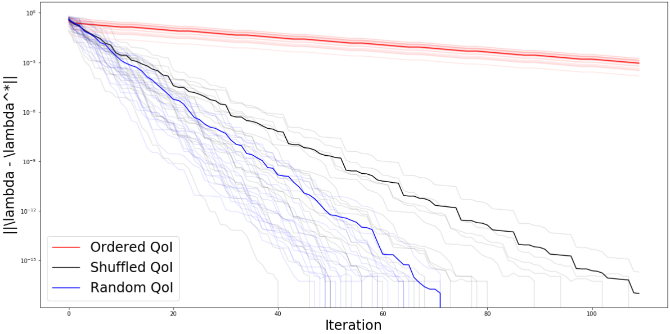


Figure: 20 initial means are chosen and iterated on for three approaches for ordering QoI. Individual experiments are transparent and the mean error is shown as solid lines.

How do I know I can trust you?

You don't. But I enabled you to check for yourself.

- 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 (LaTeX)
  - presentation (beamer)
  - every example, convergence result (Python)
  - every image in every figure
- Unit tests aid in ensuring integrity of functions
- PyPi published implementation of main methods: `pip install mud`
- Docker guarantees software runtime (on `x86`, `armv8` forthcoming): `docker pull mathematicalmichael/python:thesis`(`latex:thesis`)