Grand challenges for uncertainty quantification in inverse problems

Youssef Marzouk

Department of Aeronautics and Astronautics
Center for Computational Science and Engineering
Statistics and Data Science Center
Earth Resources Laboratory

Massachusetts Institute of Technology
http://uqgroup.mit.edu

Support from AFOSR, DOE, NSF, ONR

17 June 2021

Inverse problems

Inverse problems formalize certain **fundamental questions:**

- How to integrate models with data?
- How to learn "cause" from "effect?"

Inverse problems

Inverse problems formalize certain **fundamental questions:**

- How to integrate models with data?
- How to learn "cause" from "effect?"

These questions cut across a **broad spectrum of science and engineering fields:**

- Seismic inversion, structural health monitoring, geophysical remote sensing, imaging, weather prediction, chemical kinetics

Inverse problems formalize certain **fundamental questions:**

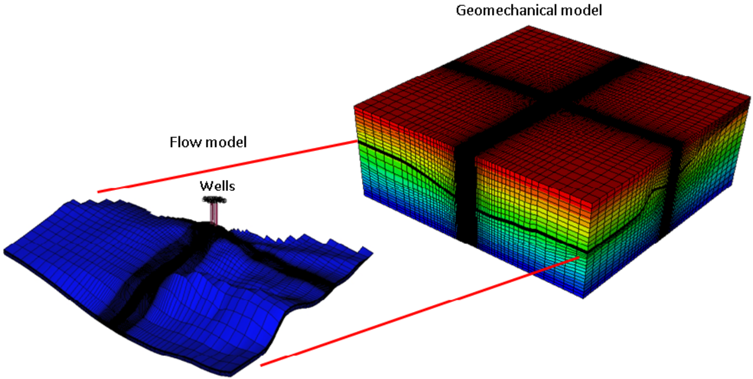
- How to integrate models with data?
- How to learn "cause" from "effect?"

These questions cut across a **broad spectrum of science and engineering fields:**

- Seismic inversion, structural health monitoring, geophysical remote sensing, imaging, weather prediction, chemical kinetics

Related notions: *data assimilation, history matching, model calibration, parameter estimation, signal recovery, ...*

Coupled flow and geomechanics



Learning aquifer and fault properties from observations of injection-induced seismicity [Jagalur, Jha, Wang, Juanes, M GRL 2018]

Common characteristics of inverse problems:

- Observations are **indirectly** related to parameters
- Observations are limited in number *(or maybe not!)*
- Observations are noisy
- Parameters are often high dimensional

Statistical inference for inverse problems

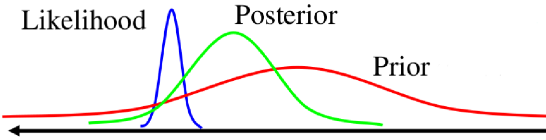
Why is a statistical perspective useful?

- To characterize *uncertainty* in the inverse solution, and understand how it depends on the number/quality of observations, features of the forward model, prior information, etc.
- To choose useful observations *(optimal experimental design)*
- To make probabilistic *predictions*
- To guide decision-making and **optimal design** under uncertainty
- To address questions of model misspecification and model validity

We take a *Bayesian* statistical perpective...

Bayesian inference

$$\pi(x|y) = \frac{\pi(y|x)\pi(x)}{\pi(y)}$$



Bayesian notion: model parameters X are treated as random variables

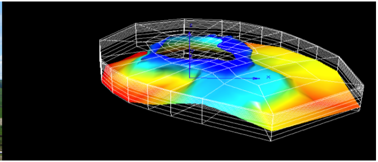
Notation

- X are model parameters; Y are the data; here, assume both to be *finite-dimensional* and to have probability densities
- $\pi(x)$ is the *prior* probability density
- $\mathcal{L}_y(x) := \pi(y|x)$ is the likelihood function
- $\pi(x|y)$ is the *posterior* probability density
- Normalizing constant $\pi(y)$ is the *evidence* or *marginal likelihood*

Bayesian inference in inverse problems

Observations $\color{red}{y}$

Parameters $\color{blue}{x}$



$$\pi_{\text{pos}}(\color{blue}{x}) := \underbrace{\pi(\color{blue}{x}|\color{red}{y}) \propto \mathcal{L}_{\color{red}{y}}(\color{blue}{x})\ \pi_{\text{pr}}(\color{blue}{x})}_{\textbf{Bayes' rule}}$$

- Need to characterize the posterior distribution (density π_{pos})
- This is a challenging task since:
  - $x \in \mathbb{R}^d$ is typically **high-dimensional**
  - Evaluations of π_{pos} may be computationally **expensive**
    - For instance: $\mathcal{L}_y$ may follow from $y \sim \mathcal{N}(\mathbf{f}(x), \Sigma)$, with expensive forward operator $\mathbf{f}$
- In what sense does π_{pos} truly capture the key uncertainties?

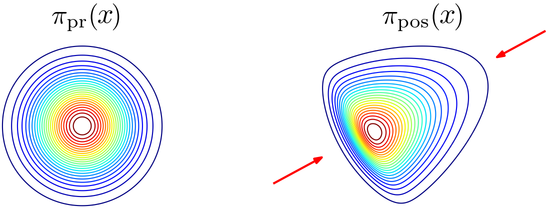
Topics for this talk

A glimpse of **grand challenges** + efforts to address them:

1. High dimensionality of parameters and data
2. Computationally intensive forward models
3. Achieving “honest” uncertainty quantification

An empirical observation

In many inverse problems, the data inform only a low-dimensional subspace of the parameters



$$\mathbb{R}^d = \underbrace{X_r}_{\pi_{\text{pos}} \neq \pi_{\text{pr}}} + \underbrace{X_\perp}_{\pi_{\text{pos}} \approx \pi_{\text{pr}}}$$

- This structure is now well understood in the linear–Gaussian case [Spantini *et al.* SISC 2015]
- We have new *certified* approximations in the nonlinear case [Zahm *et al.* arXiv:1807.03712]

Bayesian models, low rank

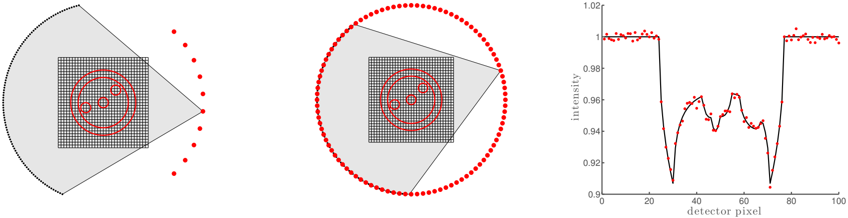
In an ill-posed inverse problem, the "data misfit Hessian" may have a compact or rapidly-decaying spectrum.

- Hessian of negative log-likelihood is $H \equiv -\nabla_x^2 \log \mathcal{L}_y(x) = G^\top \Gamma_{\mathrm{obs}}^{-1} G$
- Limited number of observations
- *Smoothing* character of the forward operator

How does **low rank** of H interact with the prior, and affect the structure of the posterior?

Answer: look at *generalized eigenvalues* of $(H, \Gamma_{\mathrm{pr}}^{-1})$

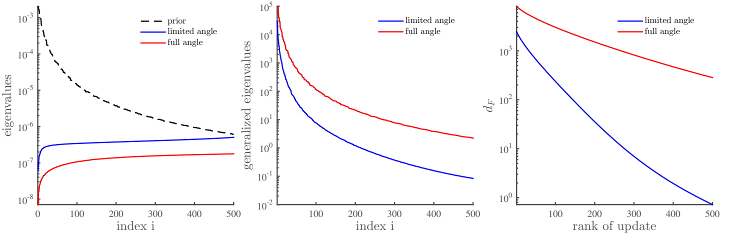
Example: computerized tomography



Example motivated by **real-time X-ray imaging** of logs entering a sawmill, for automatic quality control (see `http://finnos.fi`)

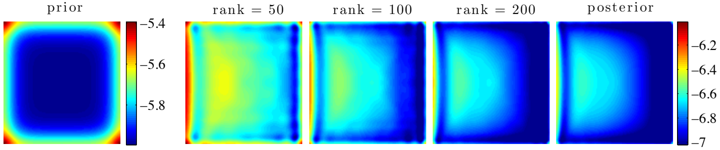
Example: computerized tomography

"Weaker data" $\Rightarrow$ faster decay of generalized eigenvalues $\Rightarrow$ better low-rank approximations



Limited angle case: $r \approx 200$ sufficient. Full angle: $r \approx 800$ needed.

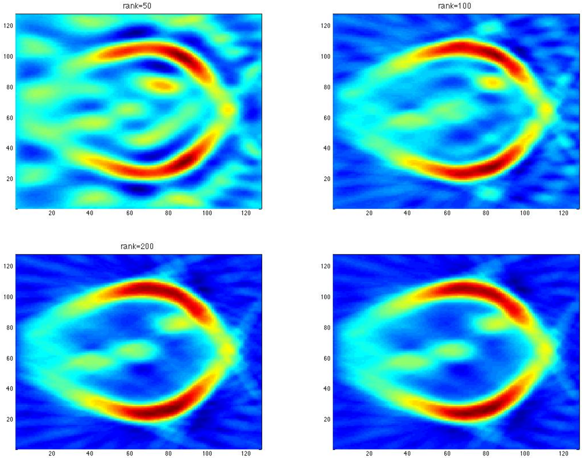
Variance fields:



Example: computerized tomography

Posterior mean is a **low-rank** function of the data *(data compression)*:

$$\mu_{\text{pos}}(y) = \Gamma_{\text{pos}}\, G^\top \Gamma_{\text{obs}}^{-1}\, y \approx \sum_{i=1}^{r} \delta_i (1+\delta_i^2)^{-1} w_i v_i^\top\, y =: A_r\, y$$



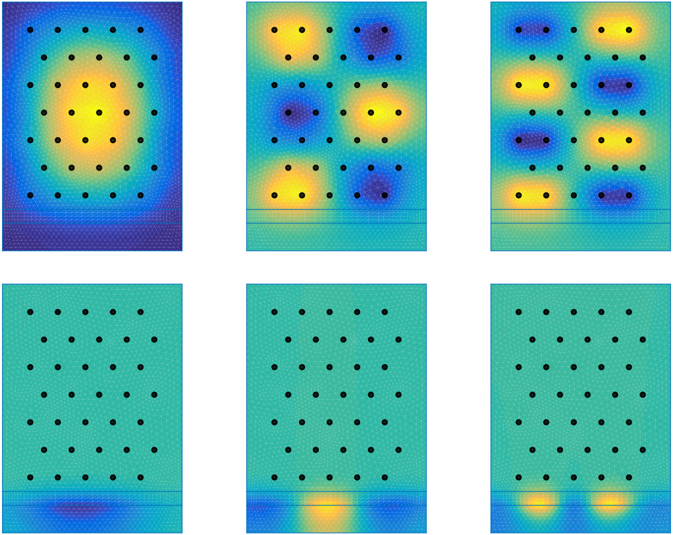
Pre-computing A_r *offline* enables fast real-time reconstructions!

Goal-oriented dimension reduction

Suppose we care about some ultimate prediction goal, e.g,. $z = \mathcal{O}x$

- There might be data-informed directions unimportant to the QoI z
- Further dimension reduction is possible! [Spantini *et al.* SISC 2017]

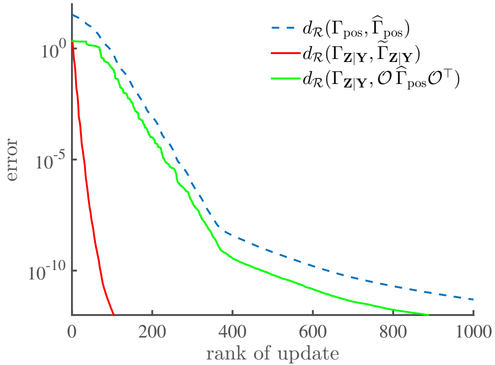
Intuition behind the goal-oriented approximation



- (*top*) Naïve approximation (*bottom*) Optimal approximation.

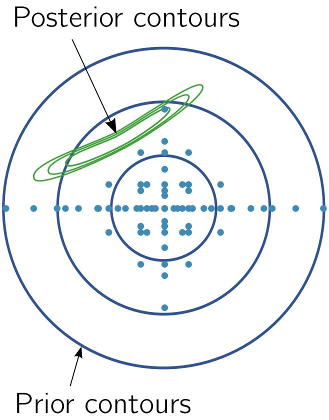
Goal-oriented dimension reduction

- Error plot for an optimal approximation $\widehat{\Gamma}_{\mathbf{z}|\mathbf{y}} = \Gamma_{\mathbf{z}} - KK^\top$



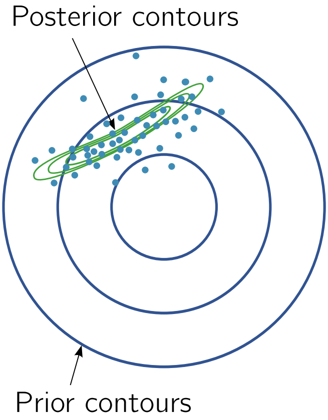
Surrogate models and approximations

- **Surrogates** for the *forward model* or *likelihood* are very useful for Bayesian inference
- Simple approach: construct an approximation (of either) over the *prior* distribution
  - Convergence of this approximation (e.g., in L_p^2) yields convergence to the true posterior [M & Xiu 2009]



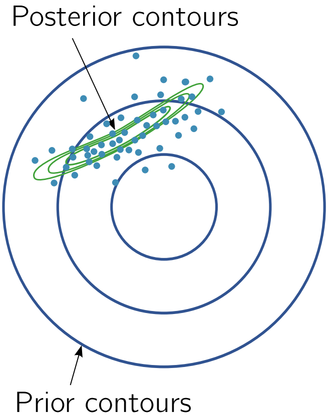
Inference with computationally intensive models

- *Posterior-focused* surrogates can improve efficiency
  - Polynomial approximations
  - Data-driven model reduction
  - Radial basis functions
- Machine learning (e.g., nonlinear regression via *neural networks*) offers new ways of building accurate surrogates



Inference with computationally intensive models

- With any surrogate, samples are then drawn from an **approximate** posterior
- Approximation cost is borne *a priori;* must balance with sampling error



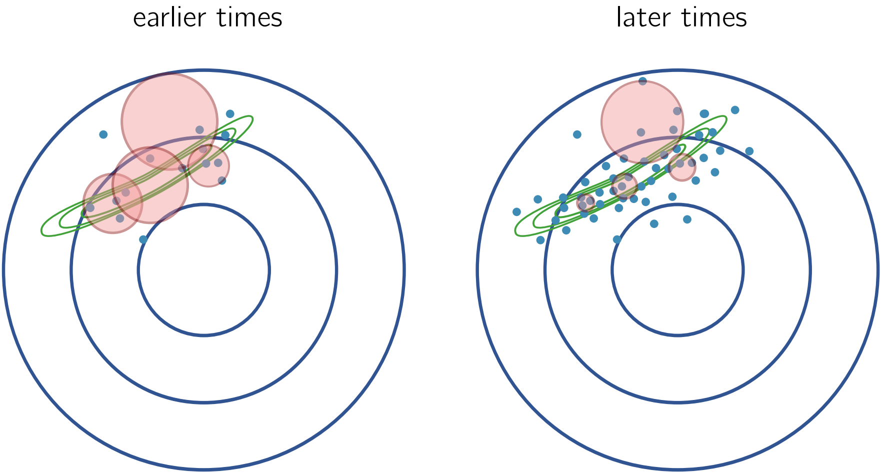
Asymptotically exact MCMC

Sampling from the exact posterior:

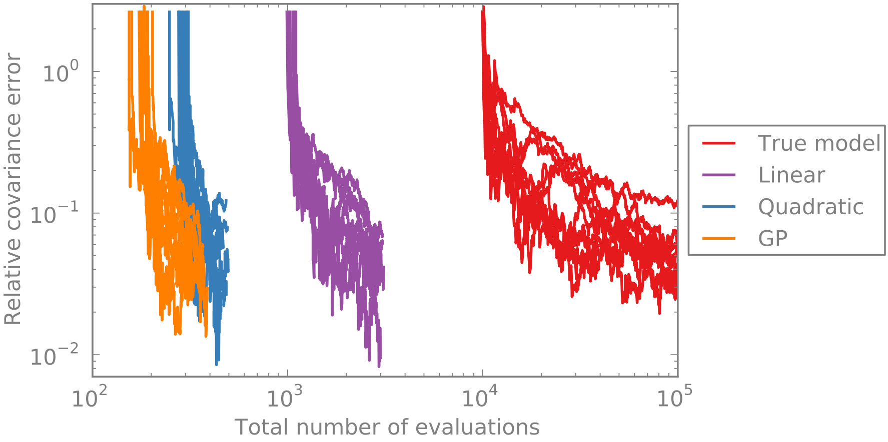
- Delayed-acceptance schemes [Christen & Fox 2005]: at least one full model evaluation per accepted sample
- Consider a different approach: *asymptotically exact* Markov chain Monte Carlo (MCMC), via incremental and infinite refinement of surrogates
  - Posterior exploration and surrogate construction occur *simultaneously*
  - Asymptotic exactness: convergence of surrogate tied to stationarity of the MCMC chain

[Conrad, M, Pillai, Smith JASA 2016; Conrad, Davis, M, Pillai, Smith SIAM JUQ 2018; Davis et al. arXiv:2006.00032]

Local approximation illustration



Local approximation: accuracy versus cost



Groundwater tracer transport model

- Nonlinear PDE for hydraulic head

$$\nabla \cdot (h\kappa \nabla h) = -f_h$$

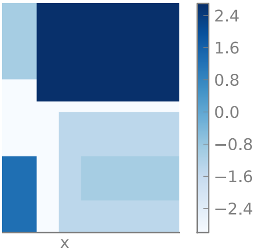
- Darcy velocity $(u, v) = -h\kappa\nabla h$ then enters tracer transport equation:

$$\frac{\partial c}{\partial t} + \nabla \cdot \left(\left(d_m \mathbf{I} + d_l \begin{bmatrix} u^2 & uv \\ uv & v^2 \end{bmatrix} \right) \nabla c \right) - \begin{bmatrix} u \\ v \end{bmatrix} \cdot \nabla c = -f_t,$$

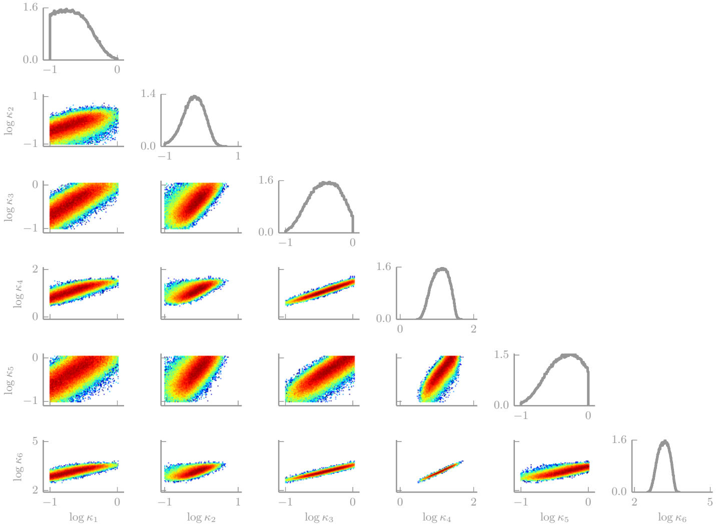
- Tracer advects according to velocity and well forcing

- Observe tracer concentration at well locations, at several times, with Gaussian error
- Infer for piecewise constant conductivities; log-normal priors
- Forward model takes about 13 seconds to evaluate

Log-conductivity field ($\log \kappa$)



Tracer transport problem: posterior distribution



Takeaways on surrogate modeling for Bayesian inference

- It can be *essential* to tractability
- **When does it work well?**
  - *Smooth* parameter-to-observable mappings
  - Low parameter dimension, or high dimensions with suitable *structure*
  - Sufficiently large and relevant training sets
- **When is it difficult?**
  - Very irregular parameter-to-observable maps
  - Small training sets
  - High dimensions without understanding/exploiting structure
  - Combinations of the above!

Topic #3: have we truly quantified uncertainty?

- Complete/honest UQ requires careful *joint* physical and statistical modeling
- Key issue: **model misspecification**
  - Does our likelihood truly model the data-generating process?
  - Are our prior assumptions realistic, or sufficiently unrestrictive?
- What happens to the posterior under misspecification?
  - Worst case: asymptotically in the size of the data, one might become arbitrarily confident about the wrong conclusions!
- Can we devise methods that are somehow *robust* to these issues?

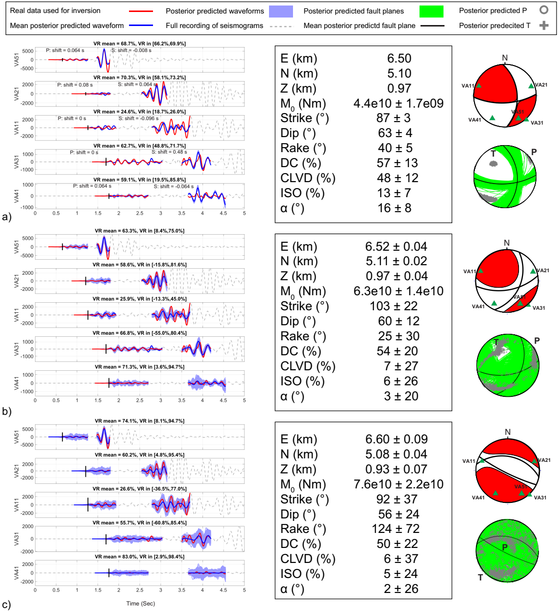
Example: moment tensor inversion

$$p(\mathbf{m}|\mathcal{D}, \mathbf{x}^*, \mathbf{v}^*)$$

$$p(\mathbf{m}|\mathcal{D}, \mathbf{v}^*) = \int p(\mathbf{m}, \mathbf{x}|\mathcal{D}, \mathbf{v}^*)\, \mathrm{d}\mathbf{x}$$

$$p(\mathbf{m}|\mathcal{D}) = \int p(\mathbf{m}, \mathbf{x}, \mathbf{v}|\mathcal{D})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{v}$$

[Chen, M, Toksöz GJI 2018]



Thanks for your attention!