Uncertainty Quantification in Inverse Problems

Aretha Teckentrup

School of Mathematics, University of Edinburgh
Alan Turing Institute, London

Newton Institute, Cambridge - June 15, 2018

Outline

1. Inverse Problems
2. Surrogate models
3. Model discrepancy
4. Further challenges

Inverse Problems

Definition and Applications

- An inverse problem is concerned with determining causal factors from observed results.
- In mathematical terms, we want to determine model inputs based on data comprised of observable model outputs.

Inverse Problems

Definition and Applications

- An inverse problem is concerned with determining causal factors from observed results.
- In mathematical terms, we want to determine model inputs based on data comprised of observable model outputs.
- Inverse problems appear in many application areas, including
  - the determination of an earthquake's epicentre using observations of seismic waves on the earth's surface,
  - the detection of flaws or cracks within a concrete structure from acoustic or electromagnetic measurements at its surface.
- The model is frequently given by a differential equation, where initial conditions, boundary conditions and/or coefficients are viewed as inputs, and observables of the solution are the outputs.

Inverse Problems

Mathematical Formulation

- Suppose we are given a mathematical model of a process, represented by a (typically non-linear) map $\mathcal{G}$.

Inverse Problems

Mathematical Formulation

- Suppose we are given a mathematical model of a process, represented by a (typically non-linear) map $\mathcal{G}$.
- We are interested in the following inverse problem: given observational data $y \in \mathbb{R}^{d_y}$, determine model parameters $u \in U$ such that

$$y = \mathcal{G}(u) + \eta,$$

  where
  - η represents observational noise, due to for example measurement error, and
  - $U \subseteq \mathbb{R}^{d_u}$ is compact. $\Rightarrow$ separable Banach space U in general

Inverse Problems

Mathematical Formulation

- Suppose we are given a mathematical model of a process, represented by a (typically non-linear) map $\mathcal{G}$.
- We are interested in the following inverse problem: given observational data $y \in \mathbb{R}^{d_y}$, determine model parameters $u \in U$ such that

$$y = \mathcal{G}(u) + \eta,$$

  where
  - η represents observational noise, due to for example measurement error, and
  - $U \subseteq \mathbb{R}^{d_u}$ is compact. $\Rightarrow$ separable Banach space U in general
- Simply "inverting $\mathcal{G}$" is not possible, since
  - we do not know the value of η, and
  - the problem is typically ill-posed/ill-conditioned.

Inverse Problems

Bayesian Formulation [Kaipio, Somersalo '04], [Stuart '10]

- The Bayesian approach treats u, y and η as random variables.

Inverse Problems

Bayesian Formulation [Kaipio, Somersalo '04], [Stuart '10]

- The Bayesian approach treats u, y and η as random variables.
- We choose a prior density (pdf) π_0 on u.

Inverse Problems

Bayesian Formulation [Kaipio, Somersalo '04], [Stuart '10]

- The Bayesian approach treats u, y and η as random variables.
- We choose a prior density (pdf) π_0 on u.
- Under the measurement model $y = \mathcal{G}(u) + \eta$ with $\eta \sim N(0, \Gamma)$, we have $y|u \sim N(\mathcal{G}(u), \Gamma)$, and the likelihood of the data y is

$$\mathcal{P}(y|u) \eqsim \exp\Big(- \frac{1}{2} \|y - \mathcal{G}(u)\|^2_{\Gamma^{-1}} \Big) =: \exp\big(- \Phi(u) \big).$$

Inverse Problems

Bayesian Formulation [Kaipio, Somersalo '04], [Stuart '10]

- The Bayesian approach treats u, y and η as random variables.
- We choose a prior density (pdf) π_0 on u.
- Under the measurement model $y = \mathcal{G}(u) + \eta$ with $\eta \sim N(0, \Gamma)$, we have $y|u \sim N(\mathcal{G}(u), \Gamma)$, and the likelihood of the data y is

$$\mathcal{P}(y|u) \eqsim \exp\Big(- \frac{1}{2} \|y - \mathcal{G}(u)\|^2_{\Gamma^{-1}} \Big) =: \exp\big(- \Phi(u)\big).$$

- Using Bayes' Theorem, we obtain the posterior pdf π^y on $u|y$, given by

$$\pi^y(u) = \frac{1}{Z} \exp\big(- \Phi(u)\big) \pi_0(u),$$

  where the normalising constant Z ensures that we have a probability distribution.

Inverse Problems

Computational Challenges

- The posterior pdf π^y allows for uncertainty quantification in $u|y$.

Inverse Problems

Computational Challenges

- The posterior pdf π^y allows for uncertainty quantification in $u|y$.

- The goal of simulations is usually
  - to sample from the posterior π^y, e.g. using Markov chain Monte Carlo methods, or
  - to compute the maximum a-posteriori (MAP) estimate

$$u^{\mathrm{MAP}} := \operatorname{argmax} \pi^y(u).$$

Inverse Problems

Computational Challenges

- The posterior pdf π^y allows for uncertainty quantification in $u|y$.
- The goal of simulations is usually
  - to sample from the posterior π^y, e.g. using Markov chain Monte Carlo methods, or
  - to compute the maximum a-posteriori (MAP) estimate

$$u^{\mathrm{MAP}} := \operatorname{argmax} \pi^y(u).$$

- In the case of models $\mathcal{G}$ given by differential equations:
  - the computation of $\Phi(u)$ is typically very costly, and surrogate models (emulators, reduced order models, ...) are frequently used to reduce computational cost $\Rightarrow$ how does this influence the posterior pdf?
  - there is usually a discrepancy between the model $\mathcal{G}$ and the real world $\Rightarrow$ how can we incorporate this?

Surrogate models

Approximation of the Posterior

- Recall: $\pi^y(u) = \frac{1}{Z} \exp\big(-\Phi(u)\big)\pi_0(u)$
- Approximating the negative log-likelihood Φ by a surrogate model Φ_N, we obtain $\pi_N^y(u) = \frac{1}{Z_N} \exp\big(-\Phi_N(u)\big)\pi_0(u)$.

Surrogate models

Approximation of the Posterior

- Recall: $\pi^y(u) = \frac{1}{Z} \exp\big(-\Phi(u)\big)\pi_0(u)$
- Approximating the negative log-likelihood Φ by a surrogate model Φ_N, we obtain $\pi_N^y(u) = \frac{1}{Z_N} \exp\big(-\Phi_N(u)\big)\pi_0(u)$.
- There is large variety of surrogate models to choose from, which are either
  - deterministic, including
    - interpolation methods (e.g. Lagrange interpolation, radial basis function (RBF) interpolation, . . .)
    - regression methods (e.g. (penalised) least squares, . . .)
    - reduced order models (e.g. proper orthogonal decomposition, . . .)
  - or stochastic, including
    - randomised projection methods (e.g. projecting the observation or parameter space onto a random low-dimensional subspace, . . .)
    - statistical interpolation/regression methods (e.g. Gaussian process emulators, RBF interpolation at random points, . . .)

Surrogate models

Approximation of the Posterior

- When the surrogate model Φ_N is stochastic, replacing Φ by Φ_N gives a random approximation to π^y:

$$\pi^{y,N}_{\text{rand}}(u, \omega) = \frac{1}{Z_N^{\text{rand}}(\omega)} \exp\big(-\Phi_N(u, \omega)\big)\pi_0(u).$$

- A deterministic approximation of π^y is obtained by either fixing ω, or by taking the marginal approximation

$$\pi^{y,N}_{\text{marg}}(u) = \frac{1}{\mathbb{E}(Z_N^{\text{rand}})} \mathbb{E}(\exp\big(-\Phi_N(u, \cdot)\big))\pi_0(u).$$

Surrogate models

Error bounds

- Under certain assumptions, we can prove that the approximate posterior densities converge to the true posterior density, as the approximate log-likelihood Φ_N converges to the true log-likelihood Φ.

Surrogate models

Error bounds

- Under certain assumptions, we can prove that the approximate posterior densities converge to the true posterior density, as the approximate log-likelihood Φ_N converges to the true log-likelihood Φ.
- This justifies the use of surrogate models in Bayesian inverse problems, and shows that the error in the distribution of $u|y$ is small when Φ_N is close to Φ.
- For more details, see [Marzouk, Xiu '09], [Stuart, T, '18], [Lie, Sullivan, T '17].

Surrogate models

Choice of surrogate model

There's a wide choice of surrogate models Φ_N:

- $\Rightarrow$ Max Gunzburger's talk earlier today,
- Gaussian process emulators [Bliznyuk et al, '08]
- dimension reduction in parameter space [Constantine et al '16]
- dimension reduction in observation space ("big data") [Le et al, '17]
- projection-based reduced order models [Arridge et al '06],
- generalised Polynomial Chaos [Marzouk, Najm, Rahn '07],
- sparse grid collocation [Marzouk, Xiu '09]

Surrogate models

Choice of surrogate model

There's a wide choice of surrogate models Φ_N:

- $\Rightarrow$ Max Gunzburger's talk earlier today,
- Gaussian process emulators [Bliznyuk et al, '08]
- dimension reduction in parameter space [Constantine et al '16]
- dimension reduction in observation space ("big data") [Le et al, '17]
- projection-based reduced order models [Arridge et al '06],
- generalised Polynomial Chaos [Marzouk, Najm, Rahn '07],
- sparse grid collocation [Marzouk, Xiu '09]

The choice of surrogate model depends on whether:

- the model $\mathcal{G}$ is given as a black box, or
- the equations for $\mathcal{G}$ are known explicitly.

Model discrepancy

Formulation

- The observed data y was assumed to be of the form $y = \mathcal{G}(u) + \eta$, with $\mathcal{G}$ our mathematical model of the process generating the data.
- The error term η incorporates
  - measurement error η_1, due for example instruments having only finite accuracy,
  - model discrepancy η_2, due to the difference between our model and the real world.

Model discrepancy

Formulation

- The observed data y was assumed to be of the form $y = \mathcal{G}(u) + \eta$, with $\mathcal{G}$ our mathematical model of the process generating the data.
- The error term η incorporates
  - measurement error η_1, due for example instruments having only finite accuracy,
  - model discrepancy η_2, due to the difference between our model and the real world.
- A typical assumption is $\eta \sim N(m, \Gamma)$, in which case the set-up of the Bayesian inverse problem is well-defined [Kaipio, Somersalo '04], [Stuart '10].
- How do we choose m and Γ?

Model discrepancy
Estimation

- In some cases, we have some structural information about model discrepancy (e.g. simplified physics).

Model discrepancy

Estimation

- In some cases, we have some structural information about model discrepancy (e.g. simplified physics).
- The model discrepancy typically has to be estimated, which is a hard problem!

Model discrepancy

Estimation

- In some cases, we have some structural information about model discrepancy (e.g. simplified physics).
- The model discrepancy typically has to be estimated, which is a hard problem!
- Some recent progress has been made in [Plumlee '17] to overcome identifiability issues. The key ideas are:
  - to define the model discrepancy as the difference between the real world data y and the "best model" $\mathcal{G}(u^*)$, where u^* is defined by minimising a loss function.
  - to explicitly use the definition of u^* to derive a distribution on η_2 (cf [Kennedy, O'Hagan '01]).
  - the resulting distribution on η_2 involves the gradient (first derivative) of the computer model, and the model discrepancy is orthogonal to the gradient of the model.

Further challenges

- How should we collect the data y?
  - optimal experimental design [Alexanderian et al, '16]
  - $\Rightarrow$ Dave Wood's talk earlier today

Further challenges

- How should we collect the data y?
  - optimal experimental design [Alexanderian et al, '16]
  - $\Rightarrow$ Dave Wood's talk earlier today
- What if standard sampling methods such as Markov chain Monte Carlo are too expensive?
  - e.g. ensemble Kalman Filter [Iglesias et al, '13]
  - e.g. Laplace approximation [Lu et al '17]

Further challenges

- How should we collect the data y?
  - optimal experimental design [Alexanderian et al, '16]
  - $\Rightarrow$ Dave Wood's talk earlier today
- What if standard sampling methods such as Markov chain Monte Carlo are too expensive?
  - e.g. ensemble Kalman Filter [Iglesias et al, '13]
  - e.g. Laplace approximation [Lu et al '17]
- How can I efficiently use a hierarchy of models?
  - multilevel and multi-fidelity methods [Multilevel Community Webpage]
  - $\Rightarrow$ Richard Wilkinson's talk later today

References I

Mulilevel Community Webpage.
https://people.maths.ox.ac.uk/gilesm/mlmc_community.html.

A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *A fast and scalable method for a-optimal design of experiments for infinite-dimensional bayesian nonlinear inverse problems*, SIAM Journal on Scientific Computing, 38 (2016), pp. A243–A272.

S. Arridge, J. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and M. Vauhkonen, *Approximation Errors and Model Reduction with an Application in Optical Diffusion Tomography*, Inverse Problems, 22 (2006), p. 175.

N. Bliznyuk, D. Ruppert, C. Shoemaker, R. Regis, S. Wild, and P. Muguntan, *Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation*, Journal of Computational and Graphical Statistics, 17 (2008).

P. G. Constantine, C. Kent, and T. Bui-Thanh, *Accelerating markov chain monte carlo with active subspaces*, SIAM Journal on Scientific Computing, 38 (2016), pp. A2779–A2805.

References II

M. A. Iglesias, K. J. Law, and A. M. Stuart, *Ensemble kalman methods for inverse problems*, Inverse Problems, 29 (2013), p. 045001.

J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, Springer, 2004.

M. C. Kennedy and A. O'Hagan, *Bayesian calibration of computer models*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63 (2001), pp. 425–464.

E. B. Le, A. Myers, T. Bui-Thanh, and Q. P. Nguyen, *A Data-Scalable Randomized Misfit Approach for Solving Large-Scale PDE-Constrained Inverse Problems*, Inv. Probl., 33 (2017), p. 065003.

H. C. Lie, T. J. Sullivan, and A. L. Teckentrup, *Random forward models and log-likelihoods in Bayesian inverse problems*.
Available as arXiv preprint 1712.05717, 2017.

Y. Lu, A. Stuart, and H. Weber, *Gaussian approximations for probability measures on r^d*, SIAM/ASA Journal on Uncertainty Quantification, 5 (2017), pp. 1136–1165.

References III

Y. Marzouk and D. Xiu, *A Stochastic Collocation Approach to Bayesian Inference in Inverse Problems*, Communications in Computational Physics, 6 (2009), pp. 826–847.

Y. M. Marzouk, H. N. Najm, and L. A. Rahn, *Stochastic Spectral Methods for Efficient Bayesian Solution of Inverse Problems*, Journal of Computational Physics, 224 (2007), pp. 560–586.

M. Plumlee, *Bayesian calibration of inexact computer models*, Journal of the American Statistical Association, 112 (2017), pp. 1274–1285.

A. Stuart, *Inverse Problems: A Bayesian Perspective*, Acta Numerica, 19 (2010), pp. 451–559.

A. M. Stuart and A. L. Teckentrup, *Posterior Consistency for Gaussian Process Approximations of Bayesian Posterior Distributions*, Mathematics of Computation, 87 (2018), pp. 721–753.