Surrogate Models in Bayesian Inverse Problems

Aretha Teckentrup

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

Joint work with:

Andrew Stuart (Caltech)

Han Cheng Lie and Tim Sullivan (Free University Berlin)

Newton Institute, Cambridge - February 8, 2018

THE UNIVERSITY of EDINBURGH
School of Mathematics

The Alan Turing Institute

Outline

1. Bayesian Inverse Problems
2. Approximations of the Posterior
3. Example 1: Gaussian Process Emulators
4. Example 2: Randomised Misfit Models
5. Conclusions

Bayesian 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.

Bayesian 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.

Bayesian Inverse Problems

Mathematical Formulation

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

Bayesian 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 is a separable Banach space (finite or infinite dimensional), for this presentation let $U \subseteq \mathbb{R}^{d_u}$ be compact.

Bayesian 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 is a separable Banach space (finite or infinite dimensional), for this presentation let $U \subseteq \mathbb{R}^{d_u}$ be compact.
- Simply "inverting $\mathcal{G}$" is not possible, since
  - we do not know the value of η, and
  - the problem is typically ill-posed.

Bayesian Inverse Problems

Mathematical Formulation [Stuart '10]

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

Bayesian Inverse Problems

Mathematical Formulation [Stuart '10]

- The Bayesian approach treats u, y and η as random variables.
- We choose a prior measure μ_0 on u with density π_0.

Bayesian Inverse Problems

Mathematical Formulation [Stuart '10]

- The Bayesian approach treats u, y and η as random variables.
- We choose a prior measure μ_0 on u with density π_0.
- 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

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

Bayesian Inverse Problems

Mathematical Formulation [Stuart '10]

- The Bayesian approach treats u, y and η as random variables.
- We choose a prior measure μ_0 on u with density π_0.
- 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 measure μ^y on $u|y$ with density π^y, given by

$$\pi^y(u) = \frac{1}{Z} \exp\big(- \Phi(u) \big) \pi_0(u), \qquad \Big(\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z} \exp\big(- \Phi(u) \big) \Big)$$

where $Z = \mathbb{E}_{\pi_0}\Big(\exp\big(- \Phi(u) \big) \Big)$.

Bayesian Inverse Problems

Computational Challenges

- 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

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

Bayesian Inverse Problems

Computational Challenges

- 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

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

- The computation of Φ is typically very costly, since the evaluation of the map $\mathcal{G}$ involves the solution of a differential equation.

Bayesian Inverse Problems

Computational Challenges

- 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

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

- The computation of Φ is typically very costly, since the evaluation of the map $\mathcal{G}$ involves the solution of a differential equation.

- To make computations feasible, we approximate Φ by a surrogate model (emulator, reduced order model, ...) in order to
  - reduce the cost of evaluating Φ, and/or
  - reduce the number of times we need to evaluate Φ.

Approximations of the Posterior

Surrogate models

- 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 $\pi_N^y(u) = \frac{1}{Z_N} \exp\big(-\Phi_N(u)\big)\pi_0(u)$.

Approximations of the Posterior

Surrogate models

- 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 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 space $\mathbb{R}^{\mathrm{d}_y}$ onto a random low-dimensional subspace, . . .)
    - statistical interpolation/regression methods (e.g. Gaussian process emulators, Lagrange interpolation at random points, . . .)

Approximations of the Posterior

Measuring accuracy

- To measure the error between π_N^y and π^y, we use the Hellinger distance

$$d_{\text{hell}}(\pi^y, \pi_N^y) = \left(\frac{1}{2} \int_U \left(\sqrt{\frac{\pi^y(u)}{\pi_0(u)}} - \sqrt{\frac{\pi_N^y(u)}{\pi_0(u)}} \right)^2 \pi_0(u) \mathrm{d}u \right)^{1/2}.$$

Approximations of the Posterior

Measuring accuracy

- To measure the error between π_N^y and π^y, we use the Hellinger distance

$$d_{\text{hell}}(\pi^y, \pi_N^y) = \left(\frac{1}{2} \int_U \left(\sqrt{\frac{\pi^y(u)}{\pi_0(u)}} - \sqrt{\frac{\pi_N^y(u)}{\pi_0(u)}} \right)^2 \pi_0(u) \mathrm{d}u \right)^{1/2}.$$

- The Hellinger distance is an upper bound on the Total Variation distance.
- For $f \in L^2_{\pi^y} \cap L^2_{\pi_N^y}$, we have the bound

$$\left| \mathbb{E}_{\pi^y}(f) - \mathbb{E}_{\pi_N^y}(f) \right| \leq C(f)\, d_{\text{hell}}(\pi^y, \pi_N^y).$$

Approximations of the Posterior

Deterministic surrogate models

Theorem [Stuart, ALT '18] (related earlier work in [Stuart '10])

Suppose $\sup_{u \in U} |\Phi(u) - \Phi_N(u)|$ can be bounded uniformly in N. Then there exists a constant C, independent of N, such that

$$d_{\text{hell}}(\pi^y, \pi_N^y) \leq C \left\| \Phi - \Phi_N \right\|_{L^2_{\pi_0}(U)} .$$

Proof: The assumption is sufficient to prove that the normalising constant Z_N satisfies

$$C_1 \leq Z_N \leq C_2,$$

for some positive constants C_1, C_2 independent of N.

The claim then follows from the local Lipschitz continuity of the exponential function.

Approximations of the Posterior

Stochastic surrogate models

- 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

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

Approximations of the Posterior

Stochastic surrogate models

Theorem ([Stuart, ALT '18], [Lie, Sullivan, ALT '17])

Suppose $\sup_{u \in U} |\Phi(u) - \mathbb{E}[\Phi_N(u)]|$ can be bounded uniformly in N. Under some regularity assumptions on the distribution of Φ_N, there exist constants $C_1, C_2 \geq 0$ and $r \geq 1$, independent of N, such that

$$d_{\text{hell}}(\pi^y, \pi^{y,N}_{\text{marg}}) \leq C_1 \left\| \mathbb{E}\left(|\Phi - \Phi_N|^r\right)^{\frac{1}{r}} \right\|_{L^2_{\pi_0}(U)},$$

and

$$\mathbb{E}\left(d_{\text{hell}}(\pi^y, \pi^{y,N}_{\text{rand}})^2\right)^{1/2} \leq C_2 \left\| \mathbb{E}\Big(|\Phi - \Phi_N|^{2r}\Big)^{\frac{1}{2r}} \right\|_{L^2_{\pi_0}(U)}.$$

The regularity assumptions need to be sufficient to prove moment bounds on Z_N^{rand}: $\mathbb{E}[(Z_N^{\text{rand}})^{\pm p}]$ is bounded independently of N for some $p > 2$.

The claim then follows from local Lipschitz continuity of exponential function.

Example 1: Gaussian Process Emulators

Simple Derivation [Rasmussen, Williams '06]

- Gaussian process emulators (also known as *kriging*) are a statistical version of interpolation.

Example 1: Gaussian Process Emulators

Simple Derivation [Rasmussen, Williams '06]

- Gaussian process emulators (also known as *kriging*) are a statistical version of interpolation.
- We view Φ as a sample of a Gaussian process $\Phi_0 \sim \mathrm{GP}(m(\cdot), k(\cdot, \cdot))$, where $m(\cdot)$ and $k(\cdot, \cdot)$ are chosen to reflect the smoothness and typical length scales of Φ.

Example 1: Gaussian Process Emulators

Simple Derivation [Rasmussen, Williams '06]

- Gaussian process emulators (also known as *kriging*) are a statistical version of interpolation.
- We view Φ as a sample of a Gaussian process $\Phi_0 \sim \mathrm{GP}(m(\cdot), k(\cdot,\cdot))$, where $m(\cdot)$ and $k(\cdot,\cdot)$ are chosen to reflect the smoothness and typical length scales of Φ.
- We evaluate Φ at design points $D = \{u^n\}_{n=1}^N \subseteq U$, obtaining function values $\{\Phi(u^n)\}_{n=1}^N$.

Example 1: Gaussian Process Emulators

Simple Derivation [Rasmussen, Williams '06]

- Gaussian process emulators (also known as *kriging*) are a statistical version of interpolation.
- We view Φ as a sample of a Gaussian process $\Phi_0 \sim \mathrm{GP}(m(\cdot), k(\cdot, \cdot))$, where $m(\cdot)$ and $k(\cdot, \cdot)$ are chosen to reflect the smoothness and typical length scales of Φ.
- We evaluate Φ at design points $D = \{u^n\}_{n=1}^N \subseteq U$, obtaining function values $\{\Phi(u^n)\}_{n=1}^N$.
- We condition Φ_0 on the observed values $\{\Phi(u^n)\}_{n=1}^N$, leading to the Gaussian process emulator Φ_N.

Example 1: Gaussian Process Emulators

Approximations of the Posterior

Theorem [Stuart, ALT, '18]

For the deterministic surrogate model $\mathbb{E}(\Phi_N)$, we have

$$d_{\text{hell}}(\pi^y, \pi^y_N) \leq C \, \|\Phi - \mathbb{E}[\Phi_N]\|_{L^2_{\pi_0}(U)} .$$

For the random surrogate model Φ_N, we have for any $\delta > 0$

$$d_{\text{hell}}(\pi^y, \pi^{y,N}_{\text{marg}}) \leq C \left\| \mathbb{E}\left(|\Phi - \Phi_N|^{1+\delta}\right)^{\frac{1}{1+\delta}} \right\|_{L^2_{\pi_0}(U)}$$

$$\mathbb{E}\left(d_{\text{hell}}(\pi^y, \pi^{y,N}_{\text{rand}})^2\right)^{1/2} \leq C \left\| \mathbb{E}\left(|\Phi - \Phi_N|^{2+\delta}\right)^{\frac{1}{2+\delta}} \right\|_{L^2_{\pi_0}(U)} .$$

Example 1: Gaussian Process Emulators

Approximations of the Posterior - Technical details

Lemma [Stuart, ALT '18]

There exist positive constants C_1, C_2, C_3, independent of N, s.t.

$$C_1 \leq \mathbb{E}\big((Z_N^{\mathrm{rand}})^p\big) \leq C_2, \quad \text{and} \quad C_2 \leq \mathbb{E}\big((Z_N^{\mathrm{rand}})^{-p}\big) \leq C_3,$$

for all $1 \leq p < \infty$.

Proof: Uses convergence of $\mathbb{E}[\Phi_N]$ and $\mathbb{V}[\Phi_N]$, Fernique's Theorem, Borell-TIS inequality and Sudakov-Fernique inequality.

Example 1: Gaussian Process Emulators

Convergence Analysis [Stuart, ALT, '18], [Wendland '04]

- Gaussian process emulators are closely related to scattered data approximation, in particular radial basis function interpolation if a stationary covariance kernel is used:

$$\Phi_0 \sim \mathrm{GP}(0, k(\| \cdot - \cdot \|))$$

$$\Rightarrow \qquad \mathbb{E}[\Phi_N(u)] = \sum_{n=1}^{N} \alpha_n k(\|u - u^n\|) \quad \text{and} \quad \mathbb{E}[\Phi_N(u^n)] = \Phi(u^n).$$

Example 1: Gaussian Process Emulators

Convergence Analysis [Stuart, ALT, '18], [Wendland '04]

- Gaussian process emulators are closely related to scattered data approximation, in particular radial basis function interpolation if a stationary covariance kernel is used:

$$\Phi_0 \sim \mathrm{GP}(0, k(\| \cdot - \cdot \|))$$

$$\Rightarrow \qquad \mathbb{E}[\Phi_N(u)] = \sum_{n=1}^{N} \alpha_n k(\|u - u^n\|) \quad \text{and} \quad \mathbb{E}[\Phi_N(u^n)] = \Phi(u^n).$$

- Under certain regularity assumptions on the design points D and the functions Φ and Φ_N, we have

$$\|\Phi - \mathbb{E}[\Phi_N]\|_{L^2(U)} \to 0, \qquad \text{and} \quad \|\mathbb{V}[\Phi_N]^{\frac{1}{2}}\|_{L^2(U)} \to 0,$$

as $N \to \infty$.

Example 1: Gaussian Process Emulators

Convergence Analysis [Stuart, ALT, '18], [Wendland '04]

- Gaussian process emulators are closely related to scattered data approximation, in particular radial basis function interpolation if a stationary covariance kernel is used:

$$\Phi_0 \sim \mathrm{GP}(0, k(\| \cdot - \cdot \|))$$

$$\Rightarrow \qquad \mathbb{E}[\Phi_N(u)] = \sum_{n=1}^{N} \alpha_n k(\|u - u^n\|) \quad \text{and} \quad \mathbb{E}[\Phi_N(u^n)] = \Phi(u^n).$$

- Under certain regularity assumptions on the design points D and the functions Φ and Φ_N, we have

$$\|\Phi - \mathbb{E}[\Phi_N]\|_{L^2(U)} \to 0, \qquad \text{and} \quad \|\mathbb{V}[\Phi_N]^{\frac{1}{2}}\|_{L^2(U)} \to 0,$$

as $N \to \infty$.

- The rate of convergence depends on the regularity assumptions, and error estimates exist for a wide range of scenarios.

Example 1: Gaussian Process Emulators

Scattered Data Approximation [Wendland '04]

- Suppose we use the family of Matèrn covariances

$$k_{\mathrm{Mat}}(u,u') = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\|u-u'\|}{\lambda} \right)^{\nu} B_\nu \left(\frac{\|u-u'\|}{\lambda} \right).$$

Example 1: Gaussian Process Emulators

Scattered Data Approximation [Wendland '04]

- Suppose we use the family of Matèrn covariances

$$k_{\mathrm{Mat}}(u,u') = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\|u-u'\|}{\lambda}\right)^{\nu} B_{\nu}\left(\frac{\|u-u'\|}{\lambda}\right).$$

- With design points $D = \{u^n\}_{n=1}^N$, define the fill distance

$$h_D = \sup_{u \in U} \inf_{u^n \in D} \|u - u^n\|. \qquad h_D \sim N^{-1/d_u}$$

Example 1: Gaussian Process Emulators

Scattered Data Approximation [Wendland '04]

- Suppose we use the family of Matèrn covariances

$$k_{\mathrm{Mat}}(u,u') = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\|u-u'\|}{\lambda}\right)^{\nu} B_\nu \left(\frac{\|u-u'\|}{\lambda}\right).$$

- With design points $D = \{u^n\}_{n=1}^N$, define the fill distance

$$h_D = \sup_{u \in U} \inf_{u^n \in D} \|u - u^n\|. \qquad h_D \sim N^{-1/d_u}$$

Theorem (see e.g. [Wendland '04], [Stuart, ALT '18])

Suppose U satisfies an interior cone condition. We have, for h_D sufficiently small,

$$\|\Phi - \mathbb{E}[\Phi_N]\|_{L^2(U)} \leq C\, h_D^{\nu+d_u/2}\, \|\Phi\|_{H^{\nu+d_u/2}(U)},$$

with C independent of N and Φ. Furthermore, $\|\mathbb{V}[\Phi_N]^{\frac{1}{2}}\|_{L^2(U)} \leq C\, h_D^{\nu}$.

Example 1: Gaussian Process Emulators

Scattered Data Approximation [Wendland '04]

- Suppose we use the family of Matèrn covariances

$$k_{\mathrm{Mat}}(u,u') = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\|u-u'\|}{\lambda} \right)^{\nu} B_{\nu}\left(\frac{\|u-u'\|}{\lambda} \right).$$

- With design points $D = \{u^n\}_{n=1}^N$, define the fill distance

$$h_D = \sup_{u \in U} \inf_{u^n \in D} \|u - u^n\|. \qquad h_D \sim N^{-1/d_u}$$

Theorem (see e.g. [Wendland '04], [Stuart, ALT '18])

Suppose U satisfies an interior cone condition. We have, for h_D sufficiently small,

$$\|\Phi - \mathbb{E}[\Phi_N]\|_{L^2(U)} \le C\, h_D^{\nu+d_u/2}\, \|\Phi\|_{H^{\nu+d_u/2}(U)},$$

with C independent of N and Φ. Furthermore, $\|\mathbb{V}[\Phi_N]^{\frac{1}{2}}\|_{L^2(U)} \le C\, h_D^{\nu}$.

- Convergence rates are available for $\Phi \in H^{\tau}(U)$, for $\tau > d_u/2$.

Example 1: Gaussian Process Emulators

Scattered Data Approximation [Rieger, Wendland '17]

- Suppose we use the family of separable Matèrn covariances

$$k_{\mathrm{sepMat}}(u, u') = \prod_{i=1}^{d_u} k_{\mathrm{Mat}}(u_i, u_i'), \qquad U = [-1, 1]^{d_u}.$$

Example 1: Gaussian Process Emulators

Scattered Data Approximation [Rieger, Wendland '17]

- Suppose we use the family of separable Matèrn covariances

$$k_{\text{sepMat}}(u, u') = \prod_{i=1}^{d_u} k_{\text{Mat}}(u_i, u_i'), \qquad U = [-1, 1]^{d_u}.$$

- Suppose D is a Smolyak sparse grid based on Clenshaw-Curtis points.

Example 1: Gaussian Process Emulators

Scattered Data Approximation [Rieger, Wendland '17]

- Suppose we use the family of separable Matèrn covariances

$$k_{\text{sepMat}}(u, u') = \prod_{i=1}^{d_u} k_{\text{Mat}}(u_i, u_i'), \qquad U = [-1, 1]^{d_u}.$$

- Suppose D is a Smolyak sparse grid based on Clenshaw-Curtis points.

Theorem [Rieger, Wendland '17]

We have

$$\|\Phi - \mathbb{E}[\Phi_N]\|_{L^\infty(U)} \leq C N^{-\nu} \, (\log N)^{(\nu+2)(d_u-1)+d_u+1} \, \|\Phi\|_{H^{\nu+1/2}_{\otimes^{d_u}}(U)},$$

with C independent of N and Φ. Furthermore,

$$\|\mathbb{V}[\Phi_N]^{\frac{1}{2}}\|_{L^2(U)} \leq C N^{-\nu} \, (\log N)^{(\nu+2)(d_u-1)+d_u+1}.$$

Example 2: Randomised Misfit Models

Simple Derivation [Le et al, '17]

Main idea: project high-dimensional data y onto low-dimensional space.

Example 2: Randomised Misfit Models

Simple Derivation [Le et al, '17]

Main idea: project high-dimensional data y onto low-dimensional space.

We introduce a random vector $\sigma \in \mathbb{R}^{d_y}$, and approximate

$$\begin{aligned}
\Phi(u) &= \frac{1}{2\sigma_\eta^2}\|y - \mathcal{G}(u)\|_2^2 \\
&= \frac{1}{2\sigma_\eta^2}\left(y - \mathcal{G}(u)\right)^T \mathbb{E}[\sigma\sigma^T]\left(y - \mathcal{G}(u)\right) \quad \text{with} \quad \mathbb{E}[\sigma\sigma^T] = \mathrm{I} \\
&= \frac{1}{2\sigma_\eta^2}\mathbb{E}\left[\left(y - \mathcal{G}(u)\right)^T \sigma\sigma^T \left(y - \mathcal{G}(u)\right)\right] \\
&= \frac{1}{2\sigma_\eta^2}\mathbb{E}\left[\left|\sigma^T \left(y - \mathcal{G}(u)\right)\right|^2\right] \\
&\approx \frac{1}{2\sigma_\eta^2}\frac{1}{N}\sum_{i=1}^N \left|{\sigma^{(i)}}^T \left(y - \mathcal{G}(u)\right)\right|^2 \quad \text{with i.i.d.} \quad \sigma^{(1)}, \dots, \sigma^{(N)} \\
&=: \Phi_N(u)
\end{aligned}$$

Example 2: Randomised Misfit Models

Properties

- A suitable distribution for σ is the ℓ-sparse distribution: for $\ell \in [0,1)$, let $s = \frac{1}{1-\ell} \geq 1$ and set, for $j = 1, \dots, d_y$,

$$\sigma_j = \sqrt{s} \begin{cases} 1, & \text{with probability } \frac{1}{2s}, \\ 0, & \text{with probability } \ell = 1 - \frac{1}{s}, \\ -1, & \text{with probability } \frac{1}{2s}. \end{cases}$$

- In [Le et al, '17], the authors are interested in computing

$$u^{\text{MAP}} := \text{argmax } \pi^{y,N}_{\text{rand}}(u)$$

and show that the required number of evaluations of $\mathcal{G}$ and it's adjoint are drastically reduced in inexact Newton methods. The cost savings are roughly $\frac{N}{d_y}$.

Example 2: Randomised Misfit Models

Convergence Analysis

Lemma [Lie, Sullivan, ALT, '17]

Suppose the entries of σ are i.i.d. ℓ-sparse, for some $\ell \in [0,1)$. Then there exist constants $\tilde{C}, C$, independent of N, such that

$$\left(\mathbb{E}_\sigma \left[d_{\text{hell}}(\pi^y, \pi^{y,N}_{\textsf{rand}})^2\right]\right)^{1/2} \leq \tilde{C} \left\| \mathbb{E}\Big(|\Phi - \Phi_N|^2\Big)^{\frac{1}{2}} \right\|_{L^2_{\pi_0}(U)} \leq \frac{C}{\sqrt{N}}.$$

Proof has two main ingredients:

1. $\mathbb{E}[(Z_N^{\text{rand}})^{\pm p}]$ is bounded independently of N for any $p \leq \infty$, since $0 \leq \Phi_N(u) \leq \sqrt{s}\,\Phi(u)$.
2. Standard estimates for Monte Carlo estimators.

Conclusions

- We discussed how surrogate models can be used to obtain computationally cheaper approximations to Bayesian posterior distributions.
- The Hellinger distance between the true and approximate posterior can be bounded directly in terms of the error in the surrogate model, measured in a suitable norm.
- The theory is generally applicable, both in terms of the mathematical model $\mathcal{G}$ involved and the choice of surrogate model.

References

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.

C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*, (2006).

C. Rieger and H. Wendland, *Sampling inequalities for sparse grids*, Numerische Mathematik, 136 (2017), pp. 439–466.

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.

H. Wendland, *Scattered Data Approximation*, vol. 17, Cambridge University Press, 2004.