Gaussian process regression in inverse problems and Markov chain Monte Carlo

Aretha Teckentrup

School of Mathematics, University of Edinburgh

Joint work with:

Tianming Bai and Konstantinos Zygalakis (University of Edinburgh)

Workshop on Probabilistic Numerics - September 15, 2022

THE UNIVERSITY *of* EDINBURGH
School of Mathematics

Outline

Bayesian Inverse Problems

Mathematical formulation

- An inverse problem is concerned with determining causal factors from observed results.
- In mathematical terms, we want to determine system inputs based on (partial and noisy) observations of system outputs.

Bayesian Inverse Problems

Mathematical formulation

- An inverse problem is concerned with determining causal factors from observed results.
- In mathematical terms, we want to determine system inputs based on (partial and noisy) observations of system outputs.
- We are interested in the following inverse problem: given observational data $y \in \mathbb{R}^{d_y}$, determine model parameters $u \in U \subseteq \mathbb{R}^{d_u}$ such that

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

where η represents observational noise, due to for example measurement error.

Bayesian Inverse Problems

Mathematical formulation

- An inverse problem is concerned with determining causal factors from observed results.
- In mathematical terms, we want to determine system inputs based on (partial and noisy) observations of system outputs.
- We are interested in the following inverse problem: given observational data $y \in \mathbb{R}^{d_y}$, determine model parameters $u \in U \subseteq \mathbb{R}^{d_u}$ such that

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

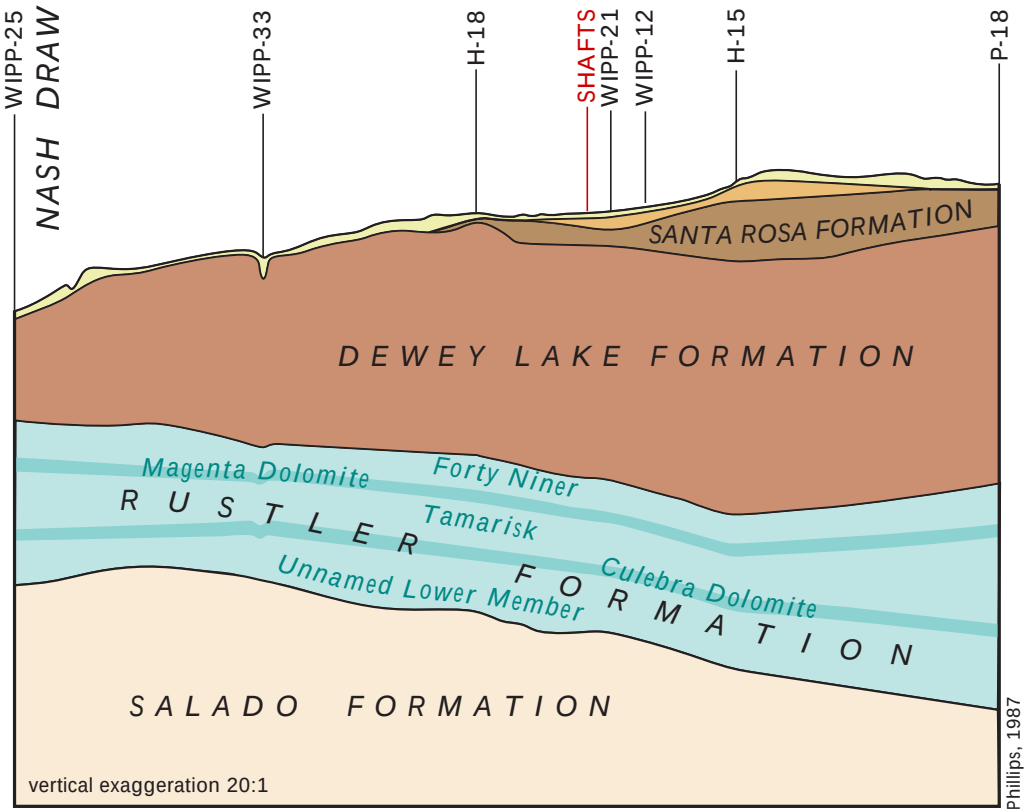
where η represents observational noise, due to for example measurement error.

- Inverse problems appear in many applications areas including medical and astronomical imaging, geophysics, climate and weather . . .

Bayesian inverse problems

Example problem in geophysics

- **Goal**: reconstruct the hydraulic conductivity k of the subsurface given noisy measurements of the water pressure $\{p(x_i)\}$.



Cross-section at WIPP

Bayesian inverse problems

Example problem in geophysics

- **Goal**: reconstruct the hydraulic conductivity k of the subsurface given noisy measurements of the water pressure $\{p(x_i)\}$.
- Unknown $u \in \mathbb{R}^{d_u}$: coefficients in a basis expansion

$$k(x; u) = \phi_0(x) + \sum_{j=1}^{d_u} u_j \phi_j(x),$$

  where $\{\phi_j\}_{j=1}^{d_u}$ are linearly independent and ϕ_0 is s.t. k is positive.

- Map $\mathcal{G}$: implicitly defined by $u \mapsto \{p(x_i; u)\}_{i=1}^{d_y}$, where p is the solution of

$$-\nabla \cdot (k(x; u) \nabla p(x; u)) = g(x).$$

  (This equation comes from Darcy's law plus conservation of mass: conductivity k, pressure head p, sources/sinks g.)

- Observations: $y = \{p(x_i; u) + \eta_i\}_{i=1}^{d_y} \in \mathbb{R}^{d_y}$.

Bayesian Inverse Problems

Mathematical Formulation

- Recall: We are interested in the following inverse problem: given observational data $y \in \mathbb{R}^{d_y}$, determine model parameters $u \in U \subseteq \mathbb{R}^{d_u}$ such that

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

where η represents observational noise, due to for example measurement error.

Bayesian Inverse Problems

Mathematical Formulation

- Recall: We are interested in the following inverse problem: given observational data $y \in \mathbb{R}^{d_y}$, determine model parameters $u \in U \subseteq \mathbb{R}^{d_u}$ such that

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

  where η represents observational noise, due to for example measurement error.

- Simply "inverting $\mathcal{G}$" is not possible, since
  - we do not know the value of η, and
  - the problem is typically ill-posed and/or ill-conditioned.

Bayesian Inverse Problems

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

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

Bayesian Inverse Problems

Mathematical Formulation [Kaipio, Somersalo '04] [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 [Kaipio, Somersalo '04] [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

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

Bayesian Inverse Problems

Mathematical Formulation [Kaipio, Somersalo '04] [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

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

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

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 (MCMC) 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 (MCMC) methods, or
  - to compute the maximum a-posteriori (MAP) estimate

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

- MCMC requires repeated evaluation of the likelihood $\exp\big(-\Phi(u)\big)$, often $10^4 - 10^6$ evaluations in practical applications.
- The computation of $\exp\big(-\Phi(u)\big)$ is typically very costly, for example if $\mathcal{G}$ involves the solution of a partial 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 (MCMC) methods, or
  - to compute the maximum a-posteriori (MAP) estimate

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

- MCMC requires repeated evaluation of the likelihood $\exp\big(-\Phi(u)\big)$, often $10^4 - 10^6$ evaluations in practical applications.
- The computation of $\exp\big(-\Phi(u)\big)$ is typically very costly, for example if $\mathcal{G}$ involves the solution of a partial differential equation.
- To make computations feasible, we approximate Φ (or $\mathcal{G}$) by a surrogate model (emulator, reduced order model, ...).

Gaussian process regression

Surrogate models

- There are many choices of surrogate models, including
  - projection-based reduced order models [Arridge et al '06],
  - generalised Polynomial Chaos [Marzouk, Najm, Rahn '07],
  - sparse grid collocation [Marzouk, Xiu '09],
  - ...

Gaussian process regression

Surrogate models

- There are many choices of surrogate models, including
  - projection-based reduced order models [Arridge et al '06],
  - generalised Polynomial Chaos [Marzouk, Najm, Rahn '07],
  - sparse grid collocation [Marzouk, Xiu '09],
  - ...
- We will use Gaussian process regression [Sacks et al '89].
  - This can be used as a black-box methodology, based on N runs of $\mathcal{G}$.
  - But aspects of the physical model (e.g. the PDE) can also be incorporated.

Gaussian process regression

Surrogate models

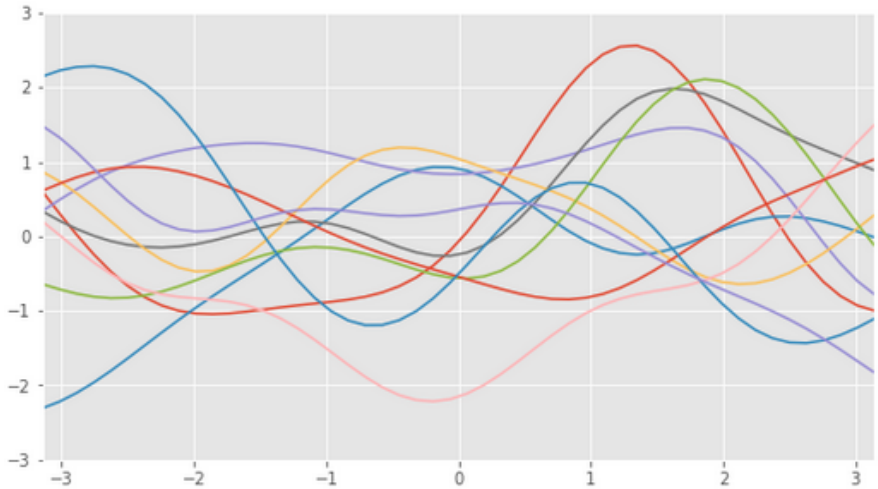
- There are many choices of surrogate models, including
  - projection-based reduced order models [Arridge et al '06],
  - generalised Polynomial Chaos [Marzouk, Najm, Rahn '07],
  - sparse grid collocation [Marzouk, Xiu '09],
  - ...
- We will use Gaussian process regression [Sacks et al '89].
  - This can be used as a black-box methodology, based on N runs of $\mathcal{G}$.
  - But aspects of the physical model (e.g. the PDE) can also be incorporated.
  - Gaussian process regression is a Bayesian procedure, which allows for uncertainty quantification. $\leftarrow$ important later on!

Gaussian process regression

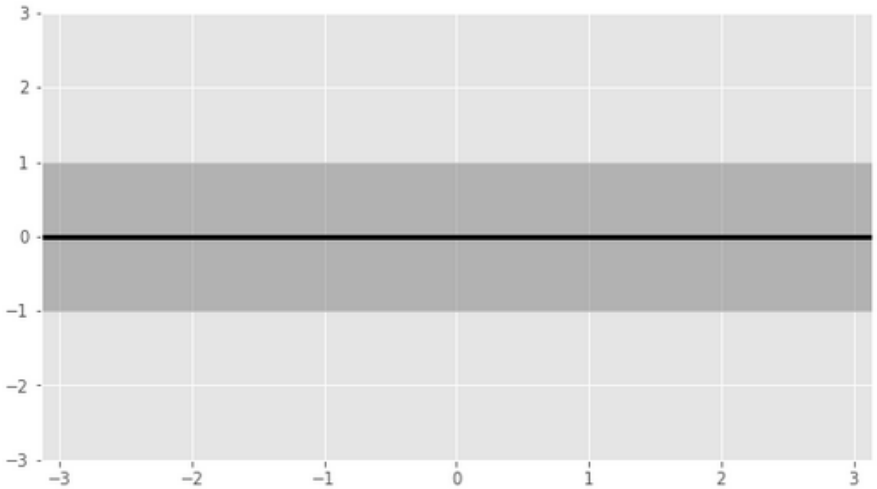
Set-up [Rasmussen, Williams '06]

- Gaussian process regression is a Bayesian methodology to emulate a function $f : U \to \mathbb{R}$, e.g. $f = \Phi$ or $f = \mathcal{G}_j$, $j = 1, \ldots, d_y$.
- We put a Gaussian process prior $\mathrm{GP}(0, k)$ on f, where k is chosen to reflect properties of f.

  For $\{u_i\}_{i=1}^m \subseteq U$, the random variables $\{f(u_i)\}_{i=1}^m$ follow a joint Gaussian distribution with $\mathbb{E}[f(u_i)] = 0$ and $\mathbb{C}[f(u_i), f(u_j)] = k(u_i, u_j)$.



Sample paths



Mean and standard deviation

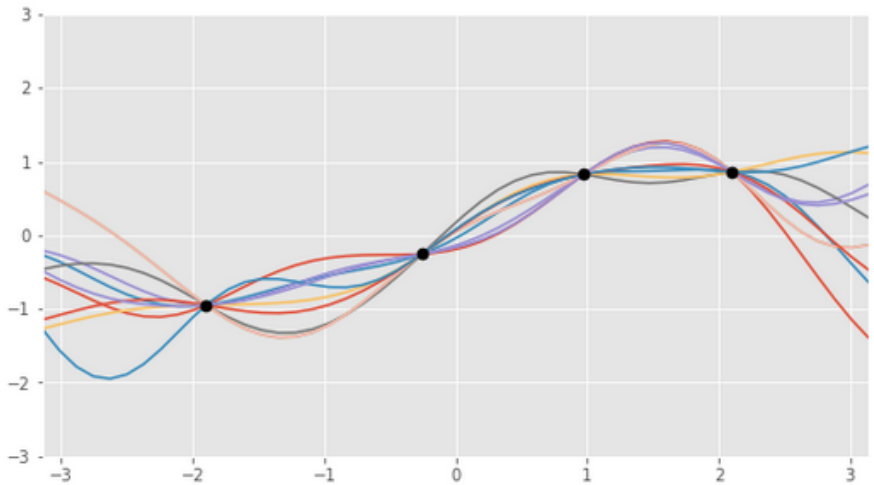
Gaussian process regression

Predictive distribution [Rasmussen, Williams '06]

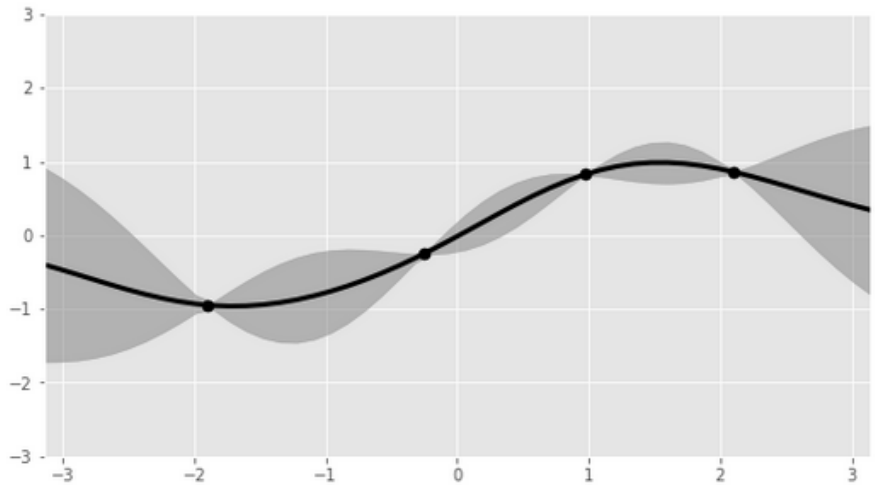
- The Gaussian process posterior $f_N \sim \mathrm{GP}(m_N^f, k_N)$ on $f|d$ is obtained by conditioning the prior on function values $d = \{u^n, f(u^n)\}_{n=1}^N$:

$$m_N^f(u) = k(u, D_N)^T K(D_N, D_N)^{-1} f(D_N),$$
$$k_N(u, u') = k(u, u') - k(u, D_N)^T K(D_N, D_N)^{-1} k(u', D_N),$$

where $D_N = \{u^n\}_{n=1}^N$, $k(u, D_N) = [k(u, u^1), \ldots, k(u, u^N)] \in \mathbb{R}^N$ and $K(D_N, D_N) \in \mathbb{R}^{N \times N}$ is the matrix with ij^{th} entry equal to $k(u^i, u^j)$.



Sample paths



Mean and standard deviation

Gaussian process regression

Predictive distribution [Rasmussen, Williams '06]

- The predictive mean m_N^f is a linear combination of kernel functions:

$$m_N^f(u) = \sum_{n=1}^{N} \alpha_n k(u, u^n), \qquad \text{for known } \alpha \in \mathbb{R}^N,$$

  and can be evaluated cheaply for a given $u \in U$.

- Computing α has cost $\mathcal{O}(N^3)$, but N is typically small!

- Example: 1d diffusion equation with $d_u = 2$ and $N = 20$

PDE solve in Firedrake	Predictive mean
$0.3s$	$0.00007s$

Gaussian process regression

Predictive distribution [Rasmussen, Williams '06]

- The predictive mean m_N^f is a linear combination of kernel functions:

$$m_N^f(u) = \sum_{n=1}^{N} \alpha_n k(u, u^n), \qquad \text{for known } \alpha \in \mathbb{R}^N,$$

  and can be evaluated cheaply for a given $u \in U$.

- Computing α has cost $\mathcal{O}(N^3)$, but N is typically small!

- Example: 1d diffusion equation with $d_u = 2$ and $N = 20$

PDE solve in Firedrake	Predictive mean
$0.3s$	$0.00007s$

- The predictive mean m_N^f is an interpolant of f, and for isotropic kernels $k(u, u') = k(\|u - u'\|_2)$, a radial basis function interpolant.

- The predictive process f_N is a random interpolant of f, reflecting the uncertainty in $f(u)$ away from the design points $D_N = \{u^n\}_{n=1}^N$.

Approximations of the Posterior

Bayesian posterior with Gaussian process regression

- Recall: $\pi^y(u) = \frac{1}{Z} \exp\left(-\frac{1}{2} \|y - \mathcal{G}(u)\|^2_{\Gamma^{-1}} \right) \pi_0(u)$
- For the remainder of the talk, assume that we approximate $\mathcal{G}$ by Gaussian process regression. Similar results hold for Φ.

Approximations of the Posterior

Bayesian posterior with Gaussian process regression

- Recall: $\pi^y(u) = \frac{1}{Z} \exp\big(-\frac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Gamma^{-1}}\big)\pi_0(u)$
- For the remainder of the talk, assume that we approximate $\mathcal{G}$ by Gaussian process regression. Similar results hold for Φ.
- Since the surrogate model $\mathcal{G}_N$ is stochastic, a deterministic approximation of π^y is obtained:
  - by taking the mean-based approximation, or

$$\pi^y_{N,\mathrm{mean}}(u) = \frac{1}{Z_N^{\mathrm{mean}}} \exp\big(-\frac{1}{2}\|y - m_N^{\mathcal{G}}(u)\|^2_{\Gamma^{-1}}\big)\pi_0(u),$$

  - by taking the marginal approximation

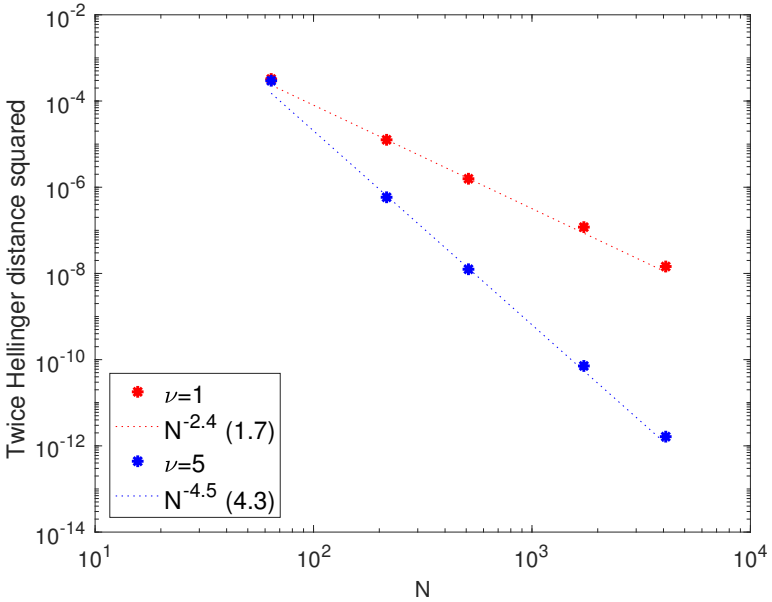
$$\pi^y_{N,\mathrm{marg}}(u) = \frac{1}{\mathbb{E}(Z_N^{\mathrm{rand}})}\mathbb{E}\left(\exp\big(-\frac{1}{2}\|y - \mathcal{G}_N(u)\|^2_{\Gamma^{-1}}\big)\right)\pi_0(u).$$

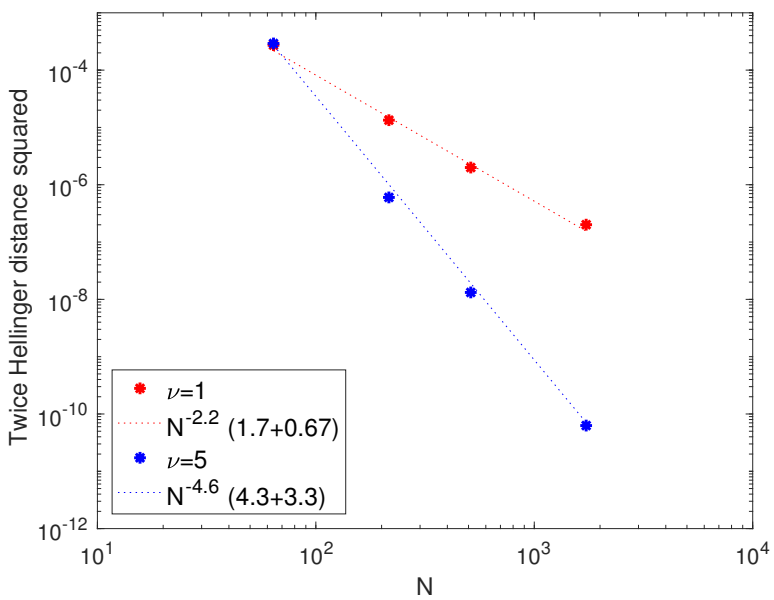
Approximations of the Posterior

Convergence as $N \to \infty$ [Stuart, T '18], [Lie, Sullivan, T '18], [T 20]

- Both approximations converge to the true posterior π^y as $N \to \infty$.

  Example: 1d diffusion equation with $d_u = 3$



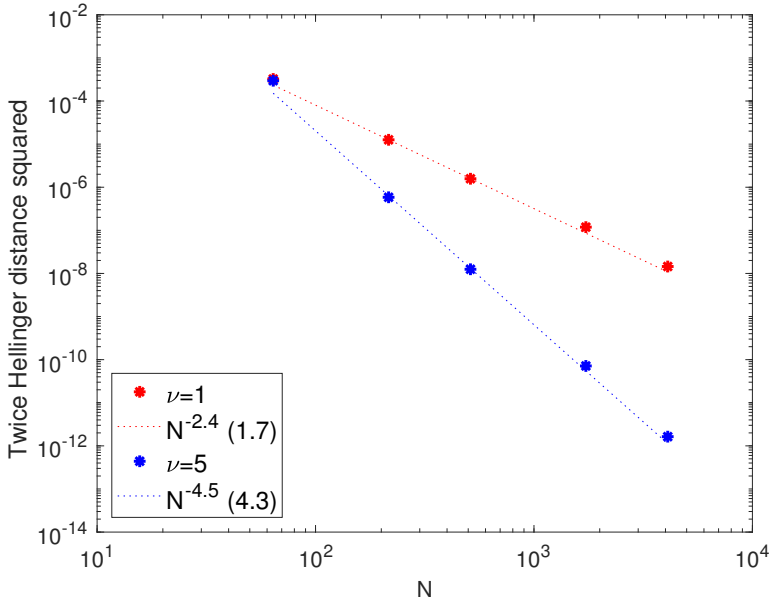


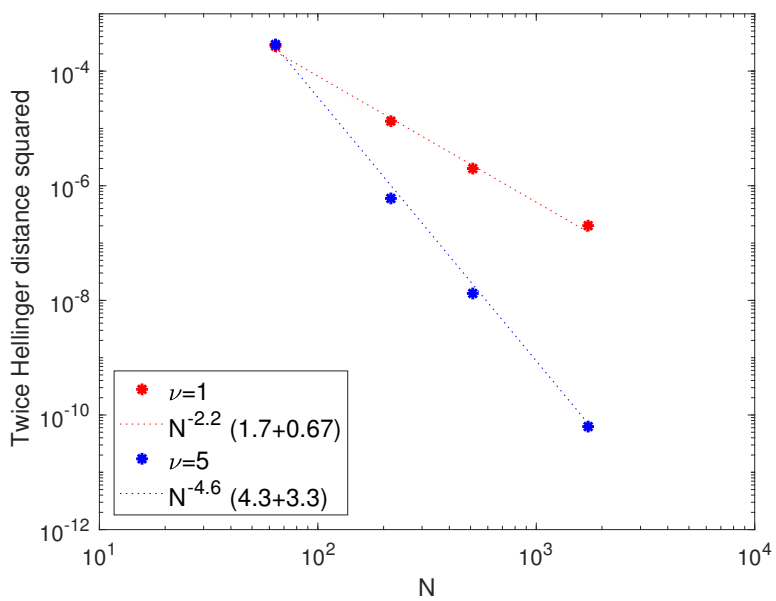
Left: Mean-based approximation. Right: Marginal approximation.

Approximations of the Posterior

Convergence as $N \to \infty$ [Stuart, T '18], [Lie, Sullivan, T '18], [T 20]

- Both approximations converge to the true posterior π^y as $N \to \infty$.
  Example: 1d diffusion equation with $d_u = 3$





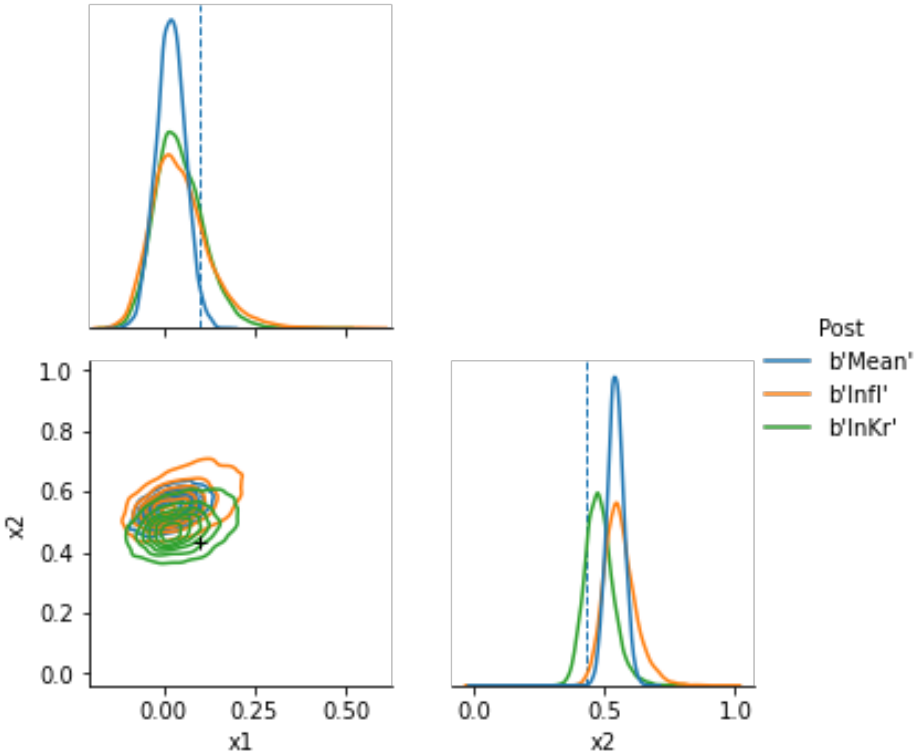
Left: Mean-based approximation. Right: Marginal approximation.

- Error in Hellinger distance depends on $\left\|\mathcal{G} - m_N^{\mathcal{G}}\right\|_{L^2(U;\mathbb{R}^{d_y})}$ and $\left\|\mathbb{E}\left(\|\mathcal{G} - m_N^{\mathcal{G}}\|^{1+\delta}\right)^{\frac{1}{1+\delta}}\right\|_{L^2(U)}$ for any $\delta > 0$, respectively.

Approximations of the Posterior

Mean-based and marginal approximations [Bai, T, Zygalakis in prep]

- For small N, the difference between $\pi^y_{N,\text{mean}}$ and $\pi^y_{N,\text{marg}}$ can be significant.

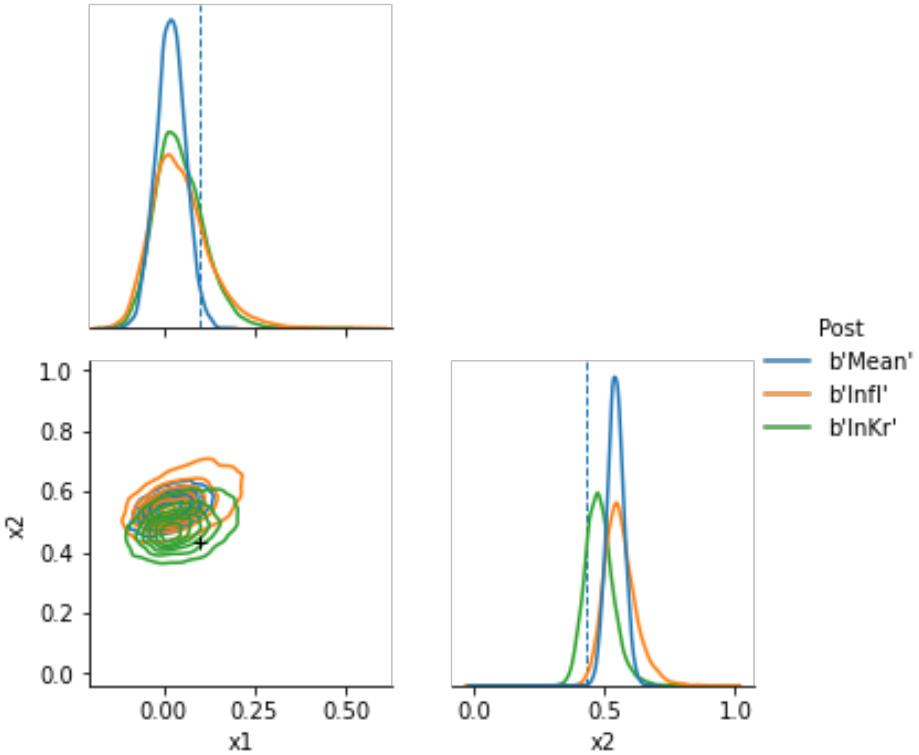


Example: 1d diffusion equation with $d_u = 2$ and $N = 20$

Approximations of the Posterior

Mean-based and marginal approximations [Bai, T, Zygalakis in prep]

- For small N, the difference between $\pi^y_{N,\text{mean}}$ and $\pi^y_{N,\text{marg}}$ can be significant.
- Only $\pi^y_{N,\text{marg}}$ uses the uncertainty in $\mathcal{G}_N$, modelling the error in the surrogate model.
- Using $\pi^y_{N,\text{mean}}$ can lead to biased predictions with high confidence.



Example: 1d diffusion equation with $d_u = 2$ and $N = 20$

Approximations of the Posterior

Marginal likelihood [Bai, T, Zygalakis in prep]

- With $\mathcal{G}_N \sim \mathrm{GP}(m_N^{\mathcal{G}}, k_N)$, we can analytically compute the marginal likelihood $\mathbb{E}\left(\exp\left(-\frac{1}{2}\|y - \mathcal{G}_N(u)\|^2_{\Gamma^{-1}}\right)\right)$.

Approximations of the Posterior

Marginal likelihood [Bai, T, Zygalakis in prep]

- With $\mathcal{G}_N \sim \mathrm{GP}(m_N^{\mathcal{G}}, k_N)$, we can analytically compute the marginal likelihood $\mathbb{E}\left(\exp\left(-\frac{1}{2}\|y - \mathcal{G}_N(u)\|^2_{\Gamma^{-1}}\right)\right)$.
- We have $\mathcal{G}_N(u) = m_N^{\mathcal{G}}(u) + \xi$, with $\xi \sim \mathrm{N}(0, k_N(u,u))$. Hence

$$\mathbb{E}\left(\exp\left(-\frac{1}{2}\|y - \mathcal{G}_N(u)\|^2_{\Gamma^{-1}}\right)\right)$$

$$= \frac{1}{\sqrt{(2\pi)^{d_y}\det\left(\Sigma(u)\right)}} \int_{\mathbb{R}^{d_y}} \exp\left(-\frac{||y - m_N^{\mathcal{G}}(u) - \xi||^2_{\Gamma^{-1}}}{2}\right) \exp\left(-\frac{||\xi||^2_{\Sigma^{-1}(u)}}{2}\right) d\xi$$

$$\propto \frac{1}{\sqrt{\det\left(\Gamma + \Sigma(u)\right)}} \exp\left(-\frac{||y - m_N^{\mathcal{G}}(u)||^2_{(\Gamma+\Sigma(u))^{-1}}}{2}\right),$$

where $\Sigma(u) = k_N(u,u)$.

Approximations of the Posterior

Variance inflation

- Compared to the mean-based likelihood

$$\frac{1}{\sqrt{\det(\Gamma)}} \exp\left(-\frac{||y - m_N^{\mathcal{G}}(u)||^2_{\Gamma^{-1}}}{2}\right),$$

the marginal likelihood

$$\frac{1}{\sqrt{\det(\Gamma + \Sigma(u))}} \exp\left(-\frac{||y - m_N^{\mathcal{G}}(u)||^2_{(\Gamma+\Sigma(u))^{-1}}}{2}\right),$$

is a form of variance inflation.

Approximations of the Posterior

Variance inflation

- Variance inflation is an emerging tool to improve Bayesian inference in complex models, see e.g. [Conrad et al '17], [Calvetti et al '18], [Cui, Fox, Neumayer '20].
- It is closely related to the well-established inclusion of modelling error [Kennedy, O'Hagan '01]:

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

with $\tilde{\eta} \sim \mathrm{N}(m, C)$.

Approximations of the Posterior

Variance inflation

- Variance inflation is an emerging tool to improve Bayesian inference in complex models, see e.g. [Conrad et al '17], [Calvetti et al '18], [Cui, Fox, Neumayer '20].
- It is closely related to the well-established inclusion of modelling error [Kennedy, O'Hagan '01]:

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

with $\tilde{\eta} \sim \mathrm{N}(m, C)$.

- Using Gaussian process regression, we have
  - a parameter-dependent variance inflation $\Sigma(u)$, rather than assuming that the error in $m_N^{\mathcal{G}}(u)$ is independent of u.
  - the variance $\Sigma(u)$ easily computed.
  - a model for $\Sigma(u)$ that is readily tuned.

Markov chain Monte Carlo methods

- In practice, we need to use sampling methods such as MCMC to sample from target density $\pi = \pi^y_{N,\text{mean}}$ or $\pi = \pi^y_{N,\text{marg}}$.

Markov chain Monte Carlo methods

- In practice, we need to use sampling methods such as MCMC to sample from target density $\pi = \pi^y_{N,\text{mean}}$ or $\pi = \pi^y_{N,\text{marg}}$.

ALGORITHM 1. (Metropolis Hastings)

- Choose $u^{(1)}$ with $\pi(u^{(1)}) > 0$.
- At state $u^{(i)}$, sample a proposal u' from density $q(u' \mid u^{(i)})$.
- Accept sample u' with probability

$$\alpha(u' \mid u^{(i)}) = \min\left(1, \frac{\pi(u')\, q(u^{(i)} \mid u')}{\pi(u^{(i)})\, q(u' \mid u^{(i)})}\right),$$

i.e. $u^{(i+1)} = u'$ with probability $\alpha(u' \mid u^{(i)})$; otherwise stay at $u^{(i+1)} = u^{(i)}$.

- The proposal density q is chosen to be easy to sample from.
- The accept/reject step is added in order to obtain samples from π.
- Knowledge of the normalising constant Z of π is not required.

Approximations of the Posterior

MALA proposals [Roberts, Tweedie '96], [Bai, T, Zygalakis in prep]

- In inverse problems, we often have high dimensional parameters u, and we require a an efficient choice of proposals such as MALA:

$$u' = u^{(i)} + \beta \nabla \log \pi(u^{(i)}) + \sqrt{2\beta}\xi_i, \qquad \text{where} \qquad \xi_i \sim \mathcal{N}(0, \mathrm{I})$$

Approximations of the Posterior

MALA proposals [Roberts, Tweedie '96], [Bai, T, Zygalakis in prep]

- In inverse problems, we often have high dimensional parameters u, and we require a an efficient choice of proposals such as MALA:

$$u' = u^{(i)} + \beta \nabla \log \pi(u^{(i)}) + \sqrt{2\beta}\xi_i, \qquad \text{where} \qquad \xi_i \sim \mathcal{N}(0, \mathrm{I})$$

- For $\pi = \pi^y_{N,\mathrm{mean}}$ and $\pi = \pi^y_{N,\mathrm{marg}}$, the gradient of the log-likelihood exists provided $k(\cdot, u^n)$ is differentiable. For common choices of k, e.g. $k(u, u') = \sigma^2 \exp(-\frac{\|u-u'\|_2^2}{\lambda})$, the gradient can be computed explicitly.

Approximations of the Posterior

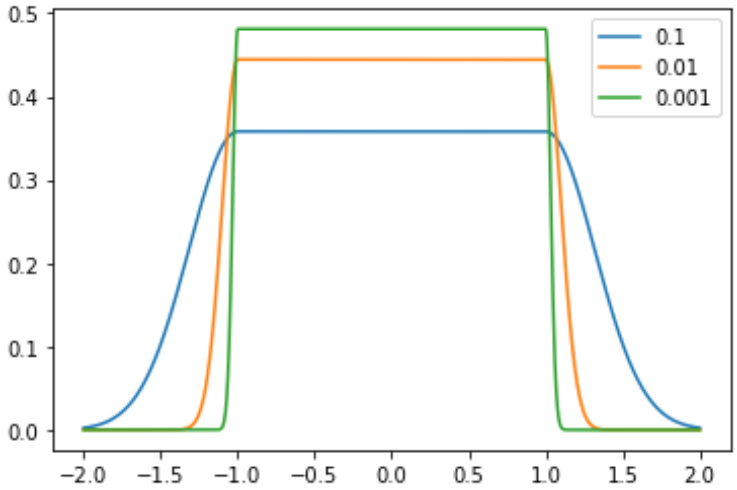
MALA proposals [Roberts, Tweedie '96], [Bai, T, Zygalakis in prep]

- In inverse problems, we often have high dimensional parameters u, and we require a an efficient choice of proposals such as MALA:

$$u' = u^{(i)} + \beta \nabla \log \pi(u^{(i)}) + \sqrt{2\beta}\xi_i, \qquad \text{where} \qquad \xi_i \sim \mathcal{N}(0, \mathrm{I})$$

- For $\pi = \pi^y_{N,\mathrm{mean}}$ and $\pi = \pi^y_{N,\mathrm{marg}}$, the gradient of the log-likelihood exists provided $k(\cdot, u^n)$ is differentiable. For common choices of k, e.g. $k(u, u') = \sigma^2 \exp(-\frac{\|u-u'\|_2^2}{\lambda})$, the gradient can be computed explicitly.

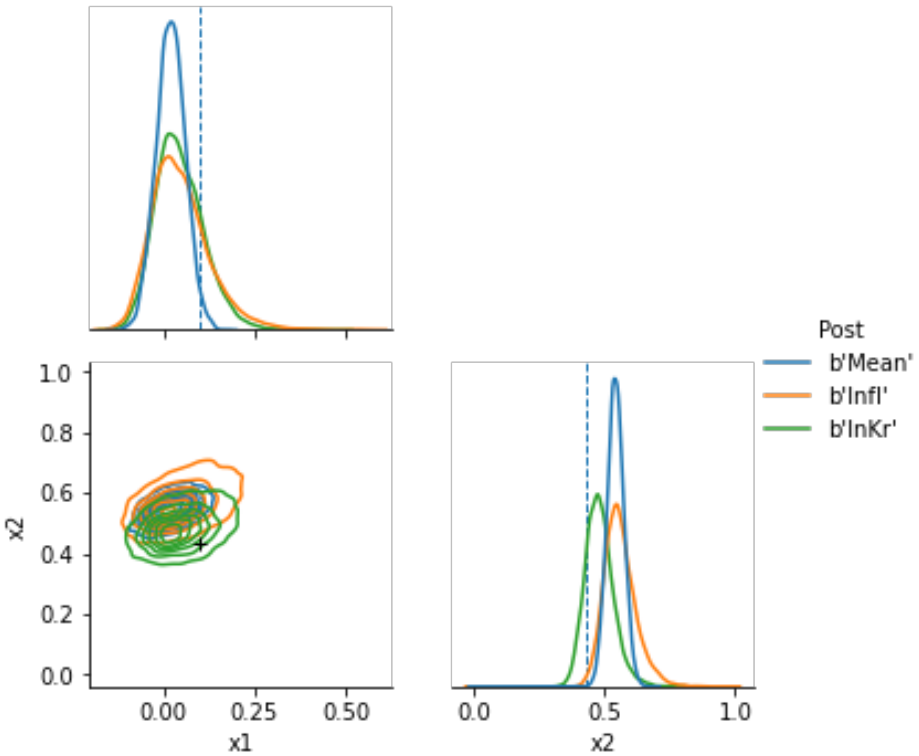
- Some priors, such as the uniform prior, require smoothing using MoreauYoshida regularisation [Pereyra '16].



Incorporating PDE constraints

Spatial correlation [Bai, T, Zygalakis in prep]

- So far, we have general $\mathcal{G}$.
- Since, $\mathcal{G} : U \to \mathbb{R}^{d_y}$, we have to choose a prior $\mathrm{GP}(0, k)$, with $k : U \times U \to \mathbb{R}^{d_y \times d_y}$.

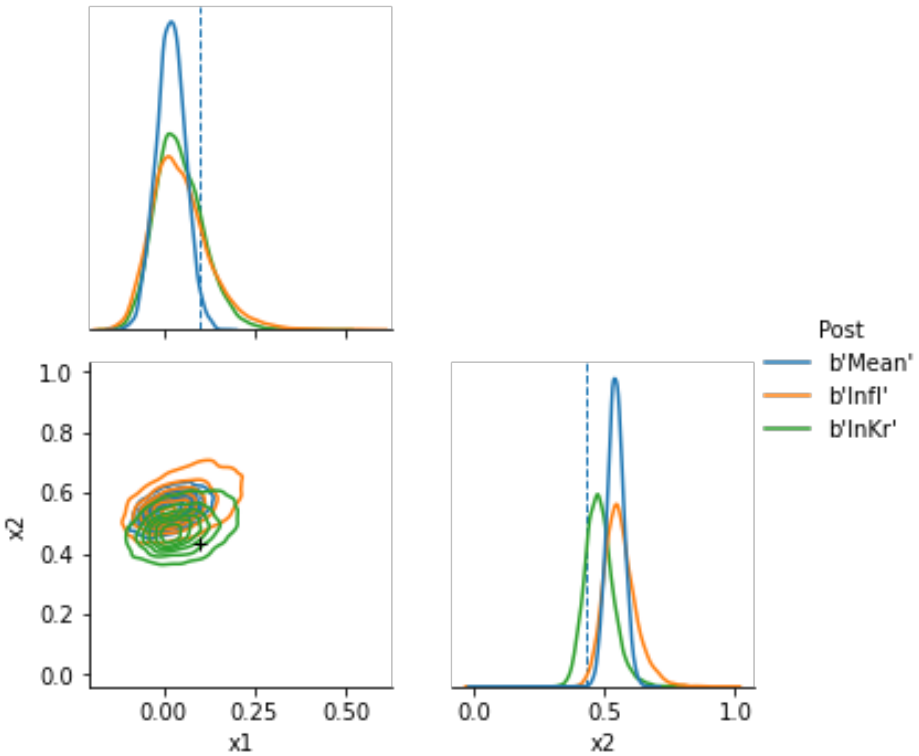


Example: 1d diffusion equation with $d_u = 2$ and $N = 20$

Incorporating PDE constraints

Spatial correlation [Bai, T, Zygalakis in prep]

- So far, we have general $\mathcal{G}$.
- Since, $\mathcal{G} : U \to \mathbb{R}^{d_y}$, we have to choose a prior $\mathrm{GP}(0, k)$, with $k : U \times U \to \mathbb{R}^{d_y \times d_y}$.
- The easiest choice is $k(u, u) = \tilde{k}(u, u')\mathrm{I}$, with $\tilde{k} : U \times U \to \mathbb{R}$, which corresponds to approximating each entry independently.
  $\to$ orange line
- But in the PDE application, $\mathcal{G}(u) = \{p(x_i; u)\}_{i=1}^{d_y}$, so we expect entries to be correlated.



Example: 1d diffusion equation with $d_u = 2$ and $N = 20$

Incorporating PDE constraints

Spatial correlation [Bai, T, Zygalakis in prep]

- To include spatial correlation, we first define a GP prior on the PDE solution:

$$p(x, u) \sim \mathrm{GP}(0, k_1(u, u')k_2(x, x')).$$

- A prior in observation space is then obtained:

$$\{p(x_i; u)\}_{i=1}^{d_y} = \mathcal{G}(u) \sim \mathrm{GP}(0, k_1(u, u')K_2),$$

where $K_2 \in \mathbb{R}^{d_y \times d_y}$ has entries $(K_2)_{ij} = k_2(x_i, x_j)$. (This works for any linear observation operator.)

Incorporating PDE constraints

Spatial correlation [Bai, T, Zygalakis in prep]

- To include spatial correlation, we first define a GP prior on the PDE solution:

$$p(x, u) \sim \mathrm{GP}(0, k_1(u, u')k_2(x, x')).$$

- A prior in observation space is then obtained:

$$\{p(x_i; u)\}_{i=1}^{d_y} = \mathcal{G}(u) \sim \mathrm{GP}(0, k_1(u, u')K_2),$$

  where $K_2 \in \mathbb{R}^{d_y \times d_y}$ has entries $(K_2)_{ij} = k_2(x_i, x_j)$. (This works for any linear observation operator.)

- We can use a standard kernel such as $k_2(x, x') = \sigma^2 \exp(-\frac{\|x-x'\|_2^2}{\lambda})$. $\rightarrow$ green line

- We can instead also use kernels from physics-informed Gaussian process regression, see e.g. the survey [Swiler et al '20].

Incorporating PDE constraints

Spatial correlation [Bai, T, Zygalakis in prep]

- To include spatial correlation, we first define a GP prior on the PDE solution:

$$p(x, u) \sim \mathrm{GP}(0, k_1(u, u')k_2(x, x')).$$

- A prior in observation space is then obtained:

$$\{p(x_i; u)\}_{i=1}^{d_y} = \mathcal{G}(u) \sim \mathrm{GP}(0, k_1(u, u')K_2),$$

where $K_2 \in \mathbb{R}^{d_y \times d_y}$ has entries $(K_2)_{ij} = k_2(x_i, x_j)$. (This works for any linear observation operator.)

- We can use a standard kernel such as $k_2(x, x') = \sigma^2 \exp(-\frac{\|x-x'\|_2^2}{\lambda})$. $\rightarrow$ green line
- We can instead also use kernels from physics-informed Gaussian process regression, see e.g. the survey [Swiler et al '20].
- Introducing spatial correlation changes k_N, but not $m_N^{\mathcal{G}}$!

Incorporating PDE constraints

Further constraints [Bai, T, Zygalakis in prep]

- Physics-informed Gaussian process regression can incorporate a variety of constraints, e.g. boundary conditions, monotonicity, divergence-free fields, ...
- In 1d, boundary conditions can be incorporated easily by adding additional entries to $\mathcal{G}$. (Extrapolation vs interpolation!)

Incorporating PDE constraints

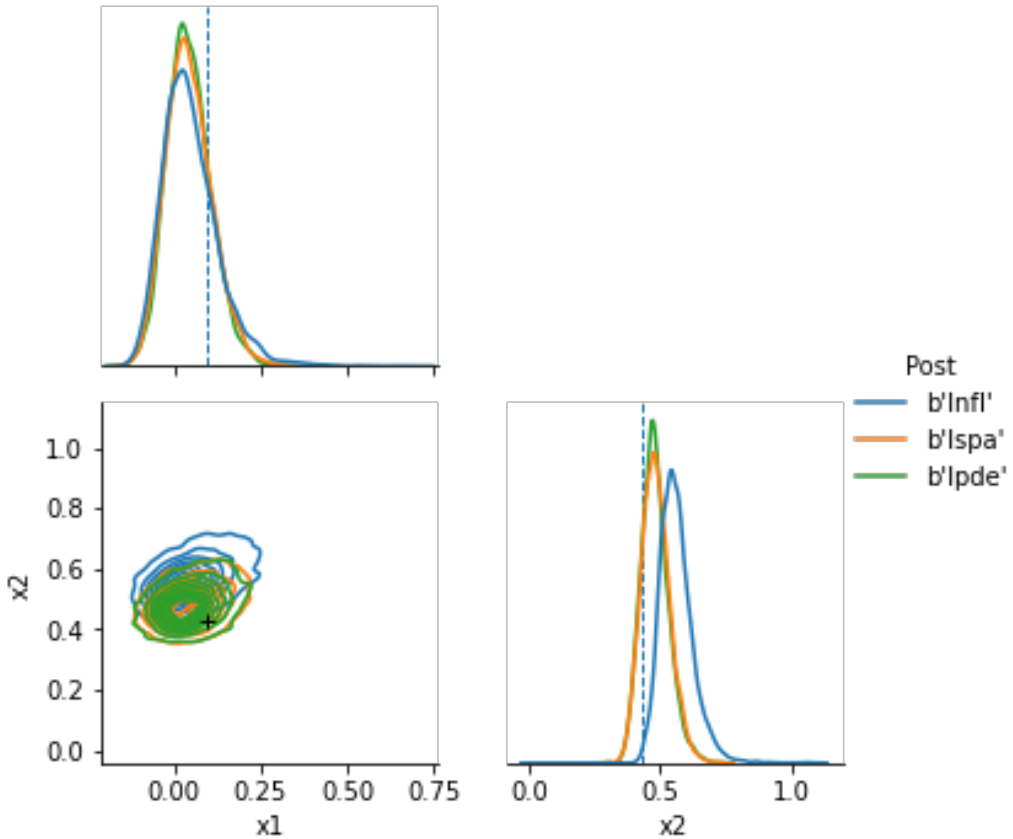
Further constraints [Bai, T, Zygalakis in prep]

- Physics-informed Gaussian process regression can incorporate a variety of constraints, e.g. boundary conditions, monotonicity, divergence-free fields, ...
- In 1d, boundary conditions can be incorporated easily by adding additional entries to $\mathcal{G}$. (Extrapolation vs interpolation!)
- For linear PDEs $\mathcal{L}_u\, p(x;u) = g(x)$, such as the diffusion equation with $\mathcal{L}_u\, p(x;u) = -\nabla \cdot (k(x;u)\nabla p(x;u))$, we can also incorporate $\mathcal{L}_u$ [Raissi, Perdikaris, Karniadakis '17]:
  - Suppose we also have available data $\tilde{y} = \{g(\tilde{x}_j) + \tilde{\eta}_j\}_{j=1}^{d_{\tilde{y}}}$.
  - We put a joint Gaussian process prior on p and g by first putting a prior on p. The corresponding prior on $g = \mathcal{L}_u\, p$ then follows.
  - Then we follow the same procedure as before to obtain a prior on $\{p(x_i;u), g(\tilde{x}_j)\}$.

Incorporating PDE constraints

Further constraints [Bai, T, Zygalakis in prep]

- Marginal approximations with and without spatial correlation.
  Example: 1d diffusion equation with $d_u = 2$ and $N = 20$



Conclusions

- We discussed how surrogate models can be used to obtain computationally cheaper approximations to Bayesian posterior distributions.
- Gaussian process regression is a popular choice of surrogate model. It can be used as a black-box model, or we can incorporate information about the model (e.g. PDE constraints).
- The predictive variance in Gaussian process regression can be used to define the marginal approximation, which is better suited for Bayesian inference in complex models.

References I


D. Calvetti, M. Dunlop, E. Somersalo, and A. Stuart, *Iterative updating of model error for Bayesian inversion*, Inverse Problems, 34 (2018), p. 025008.


P. R. Conrad, M. Girolami, S. Särkkä, A. Stuart, and K. Zygalakis, *Statistical analysis of differential equations: introducing probability measures on numerical solutions*, Statistics and Computing, 27 (2017), pp. 1065–1082.


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.


H. C. Lie, T. J. Sullivan, and A. L. Teckentrup, *Random forward models and log-likelihoods in Bayesian inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 1600–1629.


M. Pereyra, *Proximal Markov chain Monte Carlo algorithms*, Statistics and Computing, 26 (2016), pp. 745–760.

References II

M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Machine learning of linear differential equations using gaussian processes*, Journal of Computational Physics, 348 (2017), pp. 683–693.

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

G. O. Roberts and R. L. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernoulli, (1996), pp. 341–363.

J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, *Design and analysis of computer experiments*, Statistical science, 4 (1989), pp. 409–423.

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.

References III

L. P. Swiler, M. Gulian, A. L. Frankel, C. Safta, and J. D. Jakeman, *A survey of constrained Gaussian process regression: Approaches and implementation challenges*, Journal of Machine Learning for Modeling and Computing, 1 (2020).

A. L. Teckentrup, *Convergence of Gaussian process regression with estimated hyper-parameters and applications in Bayesian inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, 8 (2020), pp. 1310–1337.

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