Probabilistic Numerics for Differential Equations

Tim Sullivan[1]

with J. Cockayne[2], M. Girolami[2], H. C. Lie[1], C. J. Oates[3], and A. M. Stuart[4]

[1]Free University of Berlin / Zuse Institute Berlin, Germany
[2]The University of Warwick, UK
[3]University of Technology Sydney, Australia
[4]California Institute of Technology, USA

CSC Seminar

MPI for Dynamics of Complex Technical Systems, Magdeburg, DE

22 November 2016

Overview

What is "Probabilistic Numerics"?

- Beginning with a seminal papers[1] of Kadane (1985), Diaconis (1988), O'Hagan (1992), and Skilling (1992) there has been interest in giving probabilistic answers to ostensibly deterministic problems, e.g. quadrature, optimisation, solution of differential equations.
- In some sense, this is a Bayesian statistician's natural approach to numerical analysis, phrasing computational tasks as inference problems using finite/incomplete/imperfect information.
- Disadvantage: costs more, and gives 'fuzzier' answers, so why bother?
- Advantage: fold uncertainty arising from numerical error into inferences, and propagate this uncertainty through later computations.
  - Replicability of results $\neq$ accuracy.
  - Good data + bad/overconfident model $\implies$ faulty inferences.
  - In many practical examples from physical, social, and data sciences, we *know* that our numerical solutions are coarse approximations of models that are themselves approximate.

[1] But actually going all the way back to Poincaré (1896).

Motivating Example I: Bayesian Quadrature

$$\int_0^1 \exp\left(\cosh\left(\frac{x + 2x^2 + \cos x}{3 + \sin x^3}\right)\right) \mathrm{d}x = ???$$

Diaconis (1988) offers a Bayesian approach to quadrature:

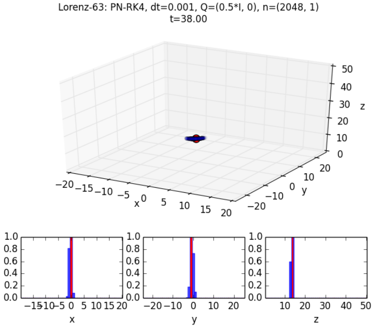
- Put a prior μ on the space of integrands $f : [a, b] \to \mathbb{R}$.
- Evaluate the integrand at nodes $\{x_i\}_{i=1}^n \subset [a, b]$ to get $f(x_i) = y_i$ (possibly with errors).
- The posterior distribution

$$\mathbb{P}_{f\sim\mu}\left(\textstyle\int_a^b f(x)\,\mathrm{d}x \,\middle|\, f(x_i) = y_i \text{ for } i = 1, \dots, n\right)$$

  is the Bayesian statistician's estimate of the integral.
- Brownian motion prior $\longleftrightarrow$ posterior mean is linear interpolation; Sul′din (1959, 1960): associated trapezoidal rule is optimal (Bayes) for quadratic loss.
- Integrated BM prior $\longleftrightarrow$ cubic spline interpolation.
- How are the posterior variance and classical numerical analysis error bounds related? Does the posterior concentrate on the truth?

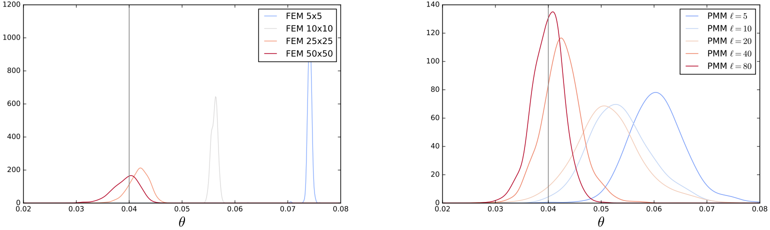
$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \sigma(y-x) \\ x(\rho - z) - y \\ xy - \beta z \end{bmatrix} \qquad \begin{aligned} \sigma &= 10 \\ \rho &= 28 \\ \beta &= 8/3 \end{aligned}$$

- Integrated with a time step $\tau = 10^{-3}$ using classical RK4.
- Global error is of order τ^4... right?



- Determistic trajectory in red;
- 2048 PN trajectories with Gaussian perturbations with variance $\frac{1}{2}\tau^9$, in blue.
- Why? Morally, this models accumulated local errors of the same scale as the s.d. of the perturbations.

$$
\begin{aligned}
-\theta \Delta u + \theta^{-1}(u^3 - u) &= 0 && \text{in } D = (0,1)^2, \\
u(x_1, x_2) &= +1 && \text{for } x_1 = 0 \text{ or } 1; \\
u(x_1, x_2) &= -1 && \text{for } x_2 = 0 \text{ or } 1.
\end{aligned}
$$



(a) FE (deterministic) (b) PMM (probabilistic)

Figure: Comparison of posteriors for θ with various forward models (likelihoods). Ground truth $\theta^\dagger = 0.04$, with prior $\theta \sim \mathrm{Unif}(0.02, 0.15)$. More details later.

Overview

1. Introduction

2. PN for ODEs
   - Sampling-Based PN for ODEs (Forward Problems)
   - PN for ODE Inverse Problems
   - Filtering-Based Approaches

3. Probabilistic Meshless Methods for PDEs
   - PMM for PDEs (Forward Problems)
   - PMM in PDE Inverse Problems

4. Closing Remarks

Overview

1. Introduction

2. PN for ODEs
   - Sampling-Based PN for ODEs (Forward Problems)
   - PN for ODE Inverse Problems
   - Filtering-Based Approaches

3. Probabilistic Meshless Methods for PDEs
   - PMM for PDEs (Forward Problems)
   - PMM in PDE Inverse Problems

4. Closing Remarks

Setting

- We consider the following autonomous ODE for $u\colon [0,T] \to \mathcal{S}$, $\mathcal{S} = \mathbb{R}^n$ or a separable Hilbert space, with vector field $f\colon \mathcal{S} \to \mathcal{S}$:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} u(t) &= f(u(t)), && \text{for } t \in [0,T], \\ u(0) &= u_0, \end{aligned} \tag{1}$$

  where the initial state $u_0 \in \mathcal{S}$ is given.
- Write $\Phi^t\colon \mathcal{S} \to \mathcal{S}$ for the flow: $u(t) = \Phi^t(u_0)$.
- Aim: Build a probabilistic numerical approximation to solutions of (1), with convergence guarantees.
- Motivation: The ODE solver is often a forward model (likelihood) in a Bayesian inverse problem, so a statistical description of the discretisation error is essential for valid inferences.
- Following Conrad et al. (2016), our will be an ensemble-based approach, cf. the global Gaussian process approach of Schober et al. (2014).

We need assumptions on (1) the underlying exact flow, (2) the deterministic numerical method, and (3) the random perturbation.

Assumption 1 (re: exact flow)

Suppose that f is smooth enough that, for $|t|$ small enough, its flow map Φ^t is globally Lipschitz with Lipschitz constant $1 + L|t|$:

$$\|\Phi^t(u) - \Phi^t(v)\| \leq (1 + L|t|)\|u - v\|.$$

Assumption 1 holds if, e.g., f is one-sided Lipschitz:

$$\langle f(u) - f(v), u - v \rangle \leq \mu \|u - v\|^2 \tag{2}$$

for all $u, v \in \mathcal{S}$, for some constant $\mu \in \mathbb{R}$; in this case, Φ^t has Lipschitz constant $\exp(\mu|t|)$, which for small enough $|t|$ is dominated by $1 + 2|\mu||t|$.

PN for ODEs

- Fix a time step $\tau > 0$. Set $t_k := k\tau$ and $u_k := u(t_k)$.
- Let $\Psi^\tau \colon \mathcal{S} \to \mathcal{S}$ be a one-step numerical integrator for the ODE (1) with time step τ.
- This class includes all Runge–Kutta methods and Taylor methods.
- That is, Ψ^τ is a numerical flow map, an approximation to the exact flow Φ^τ. This numerical flow produces a sequence of deterministic approximations to the solution of the ODE (1),

$$U_{k+1} := \Psi^\tau(U_k) \approx u_{k+1} = \Phi^\tau(u_k).$$

Assumption 2 (re: numerical flow)

Suppose that the numerical flow-map Ψ^τ has uniform local truncation error of order $q+1$: for some constant $C \geq 0$,

$$\sup_{u \in \mathcal{S}} \|\Psi^\tau(u) - \Phi^\tau(u)\| \leq C\tau^{q+1}.$$

PN for ODEs (Euler)

- Integral formulation for $u(t)$, $t_k \leq t \leq t_{k+1}$:

$$u(t) = u_k + \int_{t_k}^{t} f(u(s))\,\mathrm{d}s = \int_{t_k}^{t} g(s)\,\mathrm{d}s.$$

- Numerical integrators amount to a choice of $g(\,\cdot\,)$, subject to the common-sense criterion that $g(t_k) = f(U_k)$.
- If we posit e.g. a Gaussian random field for g, then we get a sequence of random approximations to u:

$$\begin{aligned} U(t) &= U_k + (t - t_k) f(U_k) + \xi_k(t - t_k) \\ U_{k+1} &= U_k + \tau f(U_k) + \xi_k(\tau), \end{aligned}$$

where $\mathbb{E}g = f(U_k)$ and $\xi_k = g - \mathbb{E}g$.
- Extend this idea to more general means (deterministic integrators).

PN for ODEs (General)

- Let $(\Theta, \mathcal{F}, \mathbb{P})$ be an abstract probability space, assumed to be rich enough for all the following arguments.
- We now define a new randomised one-step integrator

$$U_{k+1} := \boxed{\Psi^\tau}(U_k) + \xi_k(\tau)$$

with $\xi_k(t) := \int_0^t \chi_k(s)\,\mathrm{d}s$, where $\chi_k \sim \mathcal{N}(0, C^\tau)$ are Gaussian. The covariance structure should reflect the smoothness of f and the accuracy of Ψ^τ in terms of numbers of derivatives.

- This definition not only provides for forward propagation of the the numerical state U_k, but also a continuous output via

$$U(t) = \boxed{\Psi^{t-t_k}}(U_k) + \xi_k(t - t_k) \quad \text{for } t \in [t_k, t_{k+1}].$$

- Note: this approach is intended to model sub-grid effects rather than sub-floating point effects, though a comprehensive analysis would include the latter (Hairer et al., 2008; Mosbach and Turner, 2009).

Assumption 3 (re: random perturbation)

Suppose that $\xi_k(t) := \int_0^t \chi_k(s)\,\mathrm{d}s$, where $\chi_k \sim \mathcal{N}(0, C^\tau)$ are i.i.d., and that there are constants $C \geq 0$ and $p \geq 1$ such that, for all $t \in [0, \tau]$, $\mathbb{E}\|\xi_k(t) \otimes \xi_k(t)\| \leq Ct^{2p+1}$, and, in particular, $\mathbb{E}\|\xi_k(t)\|^2 \leq Ct^{2p+1}$.

- Prime Gaussian example is a scaled integrated Brownian motion:

$$\xi_k(t) = \tau^{p-1} \int_0^t B(s)\,\mathrm{d}s.$$

- Our $\tau \to 0$ convergence results only use the highlighted bound on the second moment, $\mathbb{E}\xi_k = 0$, and independence of the ξ_k, so the Gaussian structure is not essential to the construction.
- We can and should enlist the help of numerical analysis to inform other priors for the truncation error for finer analysis.

Strong Convergence Result

In the absence of noise:

$$\max_{0 \le k \le T/\tau} \| u_k - U_k \| \le C \tau^q, \qquad \sup_{0 \le t \le T} \| u(t) - U(t) \| \le C \tau^q.$$

Strong Convergence Result

In the absence of noise:

$$\max_{0 \leq k \leq T/\tau} \| u_k - U_k \|^2 \leq C\tau^{2q}, \qquad \sup_{0 \leq t \leq T} \| u(t) - U(t) \|^2 \leq C\tau^{2q}.$$

Strong Convergence Result

In the absence of noise:

$$\max_{0 \leq k \leq T/\tau} \|u_k - U_k\|^2 \leq C\tau^{2q}, \qquad \sup_{0 \leq t \leq T} \|u(t) - U(t)\|^2 \leq C\tau^{2q}.$$

Theorem 4 (Conrad et al., 2016)

Under Assumptions 1–3, with Gaussian ξ, there exists $C \geq 0$ such that

$$\max_{0 \leq k \leq T/\tau} \mathbb{E}\left[\|u_k - U_k\|^2\right] \leq C\tau^{2p \wedge 2q}, \tag{3}$$

$$\sup_{0 \leq t \leq T} \mathbb{E}\left[\|u(t) - U(t)\|^2\right] \leq C\tau^{2p \wedge 2q}. \tag{4}$$

- A natural choice of scaling for the noise is $p = q$, for maximal uncertainty consistent with the deterministic convergence rate.
- But there is still plenty of scope for numerical analysis and domain expertise to inform the fine structure of ξ (covariance structure, non-Gaussian structure, ...).

Stronger Convergence Result

The previous result, asserting convergence in $L^\infty\big([0,T];L^2(\Theta,\mathbb{P};\mathcal{S})\big)$, can be strengthened to convergence in $L^2\big(\Theta,\mathbb{P};L^\infty([0,T];\mathcal{S})\big)$:

Theorem 5 (Lie, Stuart & S., in prep.)

Under Assumptions 1–3, with non-Gaussian ξ, there exists $C \geq 0$ such that

$$\mathbb{E}\left[\max_{0\leq k\leq T/\tau} \|u_k - U_k\|^2\right] \leq C\tau^{2p\wedge 2q}. \tag{5}$$

If Assumption 3 is strengthened to $\mathbb{E}\left[\sup_{0\leq t\leq\tau} \|\xi_0(t)\|^2\right] \leq C\tau^{2p+1}$, then

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \|u(t) - U(t)\|^2\right] \leq C\tau^{2p\wedge 2q}. \tag{6}$$

Other Convergence Results

If Assumptions 1 (the flow being globally Lipschitz) and 2 (the numerical integrator having globally bounded truncation error) are relaxed to local bounds, then we get convergence without a rate:

Theorem 6 (Lie, Stuart & S., in prep.)

If, on each closed and bounded ball $\overline{\mathbb{B}_R(0)} \subset \mathcal{S}$*, the flow* Φ^t *has Lipschitz constant* $1 + L_R|t|$ *and* $\|\Phi^\tau - \Psi^\tau\|_\infty \leq C_R \tau^{1+q}$*, then*

$$\mathbb{E}\left[\max_{0 \leq k \leq T/\tau} \|u_k - U_k\|^2\right] \to 0 \text{ as } \tau \to 0.$$

We expect control of the growth rates of the L_R and C_R to give a convergence rate for the PN integrator, as in the case of integrators for Itô SDEs (Higham et al., 2002).

Overview

1 Introduction

2 PN for ODEs
- Sampling-Based PN for ODEs (Forward Problems)
- PN for ODE Inverse Problems
- Filtering-Based Approaches

3 Probabilistic Meshless Methods for PDEs
- PMM for PDEs (Forward Problems)
- PMM in PDE Inverse Problems

4 Closing Remarks

Example: FitzHugh–Nagumo

FitzHugh–Nagumo Oscillator

Nonlinear oscillator $u\colon [0, T] \to \mathbb{R}^2$:

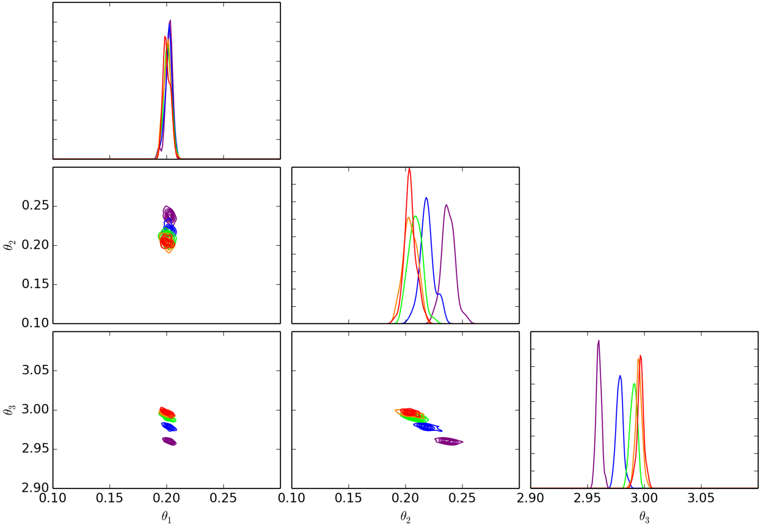
$$\frac{\mathrm{d}u}{\mathrm{d}t} = f(u) := \begin{bmatrix} u_1 - \frac{u_1^3}{3} + u_2 \\ -\frac{1}{\theta_3}(u_1 - \theta_1 + \theta_2 u_2) \end{bmatrix}$$

Note that f is not globally Lipschitz, but is one-sided Lipschitz!

- Aim: infer $\theta \in \mathbb{R}^3_{>0}$ from observations $y_i = u(t_i^{\mathrm{obs}}) + \eta_i$ at some discrete times $t_i^{\mathrm{obs}} = 0, 1, \dots, 40$, $\eta_i \sim \mathcal{N}(0, 10^{-3} I)$ i.i.d.
- Take ground truth $u(0) = (-1, 1)$ and $\theta = (0.2, 0.2, 3)$; generate data from a reference trajectory using RK4 with time step $\tau = 10^{-3}$.
- Infer θ using PN explicit Euler solvers with noise

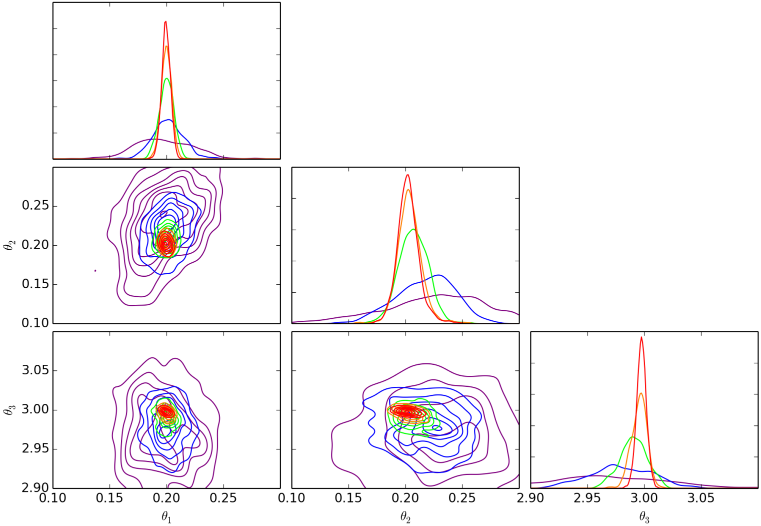
$$\mathbb{E}\big[\xi_k(\tau) \otimes \xi_k(\tau)\big] = \sigma I \tau^{2p+1} \qquad p = q = 1.$$

- Take log-normal prior for θ and compute the Bayesian posterior $\mathbb{E}_\xi \mathbb{P}[\theta | y, \tau, \xi]$ for various $\tau > 0$ and $\sigma \geq 0$.



The deterministic posteriors are over-confident at all values of the time step $\tau = 0.1, 0.05, 0.02, 0.01, 0.005$, do not overlap, and are biased.

Example: FitzHugh–Nagumo



The PN-Euler posteriors for $\tau = 0.1, 0.05, 0.02, 0.01, 0.005$ are less confident and overlap more, though are still biased.

Next Steps

- Relaxed regularity assumptions: convergence rates for locally Lipschitz flows and locally accurate deterministic solvers. Connections to numerical analysis for random dynamical systems.
- Construct structure-preserving PN integrators, e.g. for Hamiltonian dynamics the PN perturbation should not push the trajectory off the energy contours. Connections to stochastic analysis on manifolds, thermostats in molecular dynamics.
- 'On the fly' calibration of the noise covariance, and non-Gaussian structure. Connection to local error estimation and adaptivity, and to hierarchical Bayesian inversion.

Overview

1. Introduction
2. PN for ODEs
   - Sampling-Based PN for ODEs (Forward Problems)
   - PN for ODE Inverse Problems
   - Filtering-Based Approaches
3. Probabilistic Meshless Methods for PDEs
   - PMM for PDEs (Forward Problems)
   - PMM in PDE Inverse Problems
4. Closing Remarks

Filtering Approach in Brief

- For $\mathbb{R}$-valued u, Schober et al. (2014) propose an alternative approach, based upon a linear Gaussian Kálmán filter.
- Prior model for u: q-times integrated Brownian motion. Hence, $X = (u, u', u'', \dots, u^{(q)})$ solves the linear Itô SDE

$$
\mathrm{d}X(t) = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 \end{bmatrix} X(t)\,\mathrm{d}t + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma \end{bmatrix} \mathrm{d}B(t).
$$

- Data: pointwise observations of the vector field f (i.e. indirect observation of u').
- For $1 \leq q \leq 4$, the minimum posterior variance is achieved by evaluating at the Runge–Kutta points, and then classical RKq is the posterior mean.
- Ongoing work: runtime calibration of σ, matching it to classical error indicators.

Overview

1. Introduction

2. PN for ODEs
   - Sampling-Based PN for ODEs (Forward Problems)
   - PN for ODE Inverse Problems
   - Filtering-Based Approaches

3. Probabilistic Meshless Methods for PDEs
   - PMM for PDEs (Forward Problems)
   - PMM in PDE Inverse Problems

4. Closing Remarks

PN for (Elliptic) PDEs

- Huge zoo of numerical methods for PDEs $\implies$ many possible PN constructions.
- PN perturbation of FE bases for elliptic PDEs treated by Conrad et al. (2016):
  - qualitatively similar examples to ODE case — some correction of over-confident posteriors;
  - but the construction of the PN-FEM basis is tricky.
- Owhadi (2016) offers a game-theoretic approach: elementary gambles ("gamblets") on the value of PDE solution u given observations of action of test functions on the RHS; near-linear complexity if given a hierarchical structure.
- Here we build a meshless construction using Gaussian processes.

Overview

1. Introduction

2. PN for ODEs
   - Sampling-Based PN for ODEs (Forward Problems)
   - PN for ODE Inverse Problems
   - Filtering-Based Approaches

3. Probabilistic Meshless Methods for PDEs
   - PMM for PDEs (Forward Problems)
   - PMM in PDE Inverse Problems

4. Closing Remarks

Strong Formulation of Elliptic PDE

- General elliptic PDE on a bounded Lipschitz domain $D \subset \mathbb{R}^d$:

$$
\begin{aligned}
\mathcal{A}u(x) &= g(x) && \text{in } D, \\
\mathcal{B}u(x) &= 0 && \text{on } \partial D.
\end{aligned}
$$

- Assume we have a Green's function available, defined by

$$
\begin{aligned}
\mathcal{A}G(x; x') &= \delta(x - x') && \text{in } D, \\
\mathcal{B}G(x; x') &= 0 && \text{on } \partial D.
\end{aligned}
$$

Example: Poisson's equation with Dirichlet BCs

$$
\begin{aligned}
\mathcal{A}u &:= -\nabla \cdot (\kappa \nabla u) \\
\mathcal{B}u &:= \operatorname{trace} u
\end{aligned}
$$

- We aim to infer a GP emulator for u given observations of the right-hand side g.
- First construct a prior measure for u, before evaluation of $g(x)$, which incorporates information from the known form of the linear system.
- Assume $g \in \mathcal{H}(\Lambda)$, where
  - $\Lambda \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a positive definite function, and
  - $\mathcal{H}(\Lambda)$ is the reproducing kernel Hilbert space induced by Λ.
- Consider

$$k(x, x') := \iint_D G(x; z)\Lambda(z, z')G(x'; z') \, \mathrm{d}z\mathrm{d}z',$$

which is the natural kernel for the problem in the sense that its native RKHS $\mathcal{H}(k)$ consists exactly of functions u that satisfy the boundary conditions and such that $\mathcal{A}u \in \mathcal{H}(\Lambda)$.

'Natural' Posterior Measure for u

- Now construct the posterior for u given observations $\boldsymbol{g} = g(X^\circ)$ of the RHS g at locations $X^\circ = \{x_i^\circ\}_{i=1}^n$.
- For subsets $A = \{a_i\}$, $B = \{b_j\}$ of D:

$$K(A, B) = [k(a_i, b_j)]_{ij}$$
$$\mathcal{A}K(A, B) = [\mathcal{A}k(a_i, b_j)]_{ij} \qquad \text{etc.}$$

- $\bar{\mathcal{A}}K(A, B)$ denotes $\mathcal{A}$ applied to the second argument of k.

Theorem 7 (Posterior Measure for u)

$u|(g(X^\circ) = \boldsymbol{g})$ is Gaussian $\mathcal{N}(\mu, C)$, where, for finite $X \subset D$,

$$\mu(X) := \mathcal{A}K(X^\circ, X)\big(\mathcal{A}\bar{\mathcal{A}}K(X^\circ, X^\circ)\big)^{-1}\boldsymbol{g}$$
$$C(X) := K(X, X) - \bar{\mathcal{A}}K(X, X^\circ)\big(\mathcal{A}\bar{\mathcal{A}}K(X^\circ, X^\circ)\big)^{-1}\mathcal{A}K(X^\circ, X)$$

'Practical' Prior and Posterior

- The 'natural' kernel k encodes the boundary conditions, but is difficult to access in practice.
- Instead, following Cialenco et al. (2012, Lemma 2.2) we work with a 'practical' kernel $\hat{k}$ of the form

$$\hat{k}(x, x') := \int_D \tilde{k}(x, z)\tilde{k}(z, x') \, \mathrm{d}z,$$

  where $\tilde{k}$ is a 'generic' kernel such as Whittle–Matérn or Wendland.
- Boundary conditions are now enforced by adding additional observations $\boldsymbol{b}$ on ∂D.

Theorem 8 (Posterior Measure for u)

$u|\boldsymbol{g}, \boldsymbol{b} \sim \mathcal{N}(\mu, C)$, where, for $\mathcal{L} := \begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix}^\top$ and $X \subset D$,

$$\mu(X) := \mathcal{L}\hat{K}(X^\circ, X)\big(\mathcal{L}\bar{\mathcal{L}}\hat{K}(X^\circ, X^\circ)\big)^{-1}[\boldsymbol{g}^\top, \boldsymbol{b}^\top]^\top$$

$$C(X) := \hat{K}(X, X) - \bar{\mathcal{L}}\hat{K}(X, X^\circ)\big(\mathcal{L}\bar{\mathcal{L}}\hat{K}(X^\circ, X^\circ)\big)^{-1}\mathcal{L}\hat{K}(X^\circ, X)$$

Accuracy and Contraction of the Posterior for u

Theorem 9 (Cockayne et al., 2016, Prop. 6–Theorem 8)

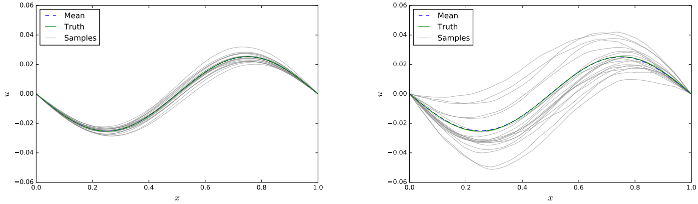
Let $u^\dagger \in \mathcal{H}(\hat{k})$ be the exact solution to the PDE, and $u^{\mathrm{PN}} := u|\boldsymbol{g}, \boldsymbol{b}$ the PN solution (posterior GP). Then $\mu = \mathbb{E}[u^{\mathrm{PN}}]$ coincides with the symmetric collocation solution of Fasshauer (1999) and

$$
\begin{aligned}
\text{local accuracy:} \quad & |\mathbb{E}[u^{\mathrm{PN}}(x)] - u^\dagger(x)| \leq \sigma(x) \|u^\dagger\|_{\hat{k}}, \\
\text{minimax rate:} \quad & \sigma(x) = O\big(h^{\beta - d/2 - \rho}\big), \\
\text{contraction:} \quad & \mathbb{P}\big[\|u^{\mathrm{PN}} - u^\dagger\|_{L^2(D)} > \varepsilon\big] = O\left(\frac{h^{2\beta - d - 2\rho}}{\varepsilon}\right),
\end{aligned}
$$

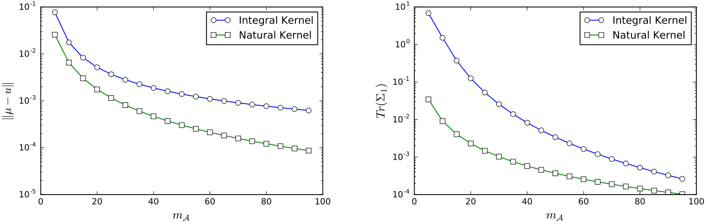
where

- *h denotes the fill distance of the observation sites $X^\circ \subset D$;*
- *$\mathcal{H}(\hat{k})$ is norm-equivalent to Sobolev space $H^\beta(D)$;*
- *the PDE is of order $\rho < \beta - d/2$, $d = \dim D$.*

$$-u''(x) = \sin(2\pi x) \quad \text{with Dirichlet BCs on } [0,1]$$



(a) 'Natural' kernel, $\#X^\circ = 39$ (b) Integral kernel, $\#X^\circ = 39$



(c) Posterior error (d) Posterior covariance

Overview

Bayesian Parameter Inference with PN: Simple Example

- We now consider PDE-constrained parameter inference problems.
- Continue with the previous 1d example:

$$-\theta u''(x) = \sin(2\pi x) \quad \text{with Dirichlet BCs on } [0,1]$$

- Infer θ given observations of u with $\theta = \theta^\dagger = 1$ at $x = 0.25$ and 0.75, corrupted by additive noise $\mathcal{N}(0, 10^{-6} I)$.
- Compare PN to its deterministic counterpart, symmetric collocation.

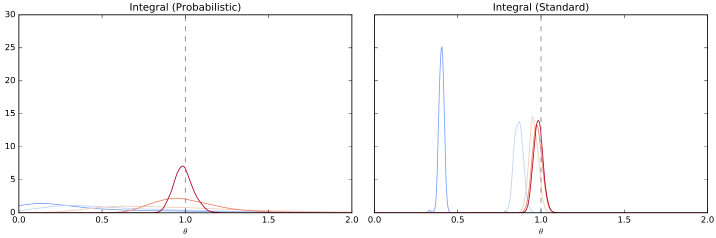


Figure: Posterior distributions for θ using integral Wendland kernel.

Bayesian Parameter Inference with PN: Simple Example

- We now consider PDE-constrained parameter inference problems.
- Continue with the previous 1d example:

$$-\theta u''(x) = \sin(2\pi x) \quad \text{with Dirichlet BCs on } [0,1]$$

- Infer θ given observations of u with $\theta = \theta^\dagger = 1$ at $x = 0.25$ and 0.75, corrupted by additive noise $\mathcal{N}(0, 10^{-6} I)$.
- Compare PN to its deterministic counterpart, symmetric collocation.

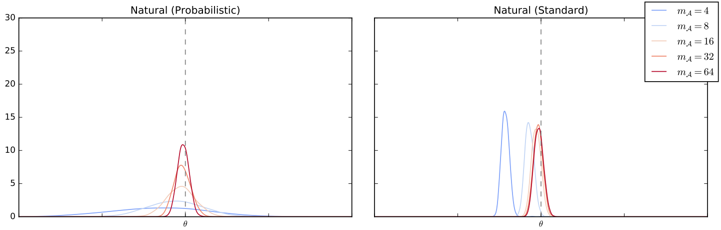


Figure: Posterior distributions for θ using optimal 'natural' kernel.

- We now consider PDE-constrained parameter inference problems.
- Continue with the previous 1d example:

$$-\theta u''(x) = \sin(2\pi x) \quad \text{with Dirichlet BCs on } [0,1]$$

- Infer θ given observations of u with $\theta = \theta^{\dagger} = 1$ at $x = 0.25$ and 0.75, corrupted by additive noise $\mathcal{N}(0, 10^{-6} I)$.
- Compare PN to its deterministic counterpart, symmetric collocation.

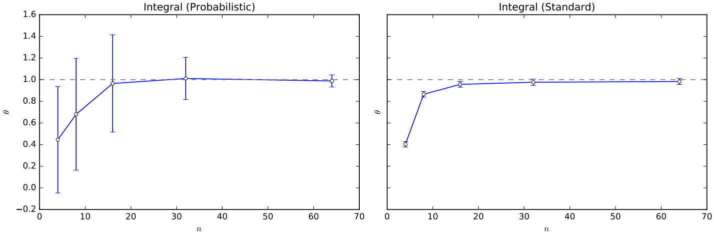


Figure: Posterior credible (1 s.d.) intervals for θ using integral Wendland kernel, as a function of $n = \#X^{\circ}$.

Bayesian Parameter Inference with PN: Simple Example

- We now consider PDE-constrained parameter inference problems.
- Continue with the previous 1d example:

$$-\theta u''(x) = \sin(2\pi x) \quad \text{with Dirichlet BCs on } [0,1]$$

- Infer θ given observations of u with $\theta = \theta^\dagger = 1$ at $x = 0.25$ and 0.75, corrupted by additive noise $\mathcal{N}(0, 10^{-6} I)$.
- Compare PN to its deterministic counterpart, symmetric collocation.

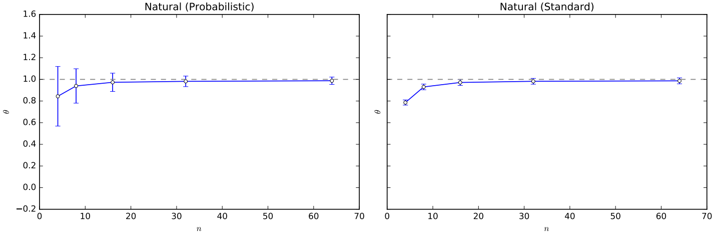


Figure: Posterior credible (1 s.d.) intervals for θ using optimal 'natural' kernel, as a function of $n = \#X^\circ$.

More generally, for the Poisson problem

$$
\begin{aligned}
-\nabla \cdot (\theta \nabla u) &= g && \text{in } D \subset \mathbb{R}^d, \\
u &= b && \text{on } \partial D,
\end{aligned}
$$

with data $y_i = u^\dagger(x_i; \kappa^\dagger) + \xi_i$, $\xi \sim \mathcal{N}(0, \Gamma)$, we obtain contraction of the posterior for θ using the PN forward solver provided the idealised problem is also contractive and the observation set X° is 'dense enough':

Theorem 10 (Cockayne et al., 2016, Theorem 11)

If the posterior for θ under the idealised exact solution $u^\dagger$ contracts in probability to $\delta_{\theta^\dagger}$, then so too does the posterior for θ under the PN solution $u^{\mathrm{PN}} := u | g(X^\circ)$, provided the fill distance h of X° and the number of data points n scale as

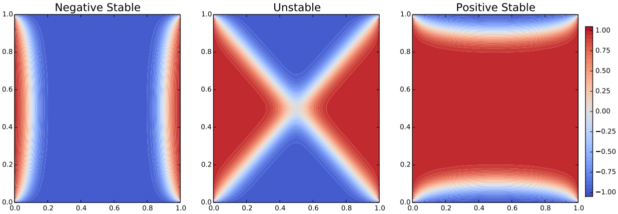
$$
h = o\left(\frac{1}{n^{1/(\beta - d/2 - \rho)}} \right).
$$

$$
\begin{aligned}
-\theta \Delta u + \theta^{-1}(u^3 - u) &= g && \text{in } D = (0,1)^2, \\
u(x_1, x_2) &= +1 && \text{for } x_1 = 0 \text{ or } 1; \\
u(x_1, x_2) &= -1 && \text{for } x_2 = 0 \text{ or } 1.
\end{aligned}
$$

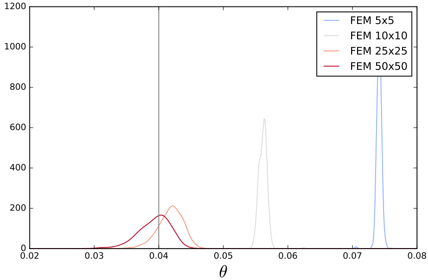
- 'Mild' nonlinearity: linear + monotone.
- Extend PMM to handle this nonlinearity by introducing a latent variable z, which is later marginalised out:

$$
-\theta \Delta u - \theta^{-1} u = z, \qquad \theta^{-1} u^3 = g - z.
$$

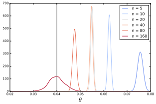
- Exhibits multiple solutions for $\theta \approx 0.04$, $g = 0$:



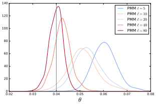
We try to recover θ from 16 observations of u on a 4×4 regular interior grid, corrupted by $\mathcal{N}(0, \frac{1}{10} I)$ noise.



(a) FE



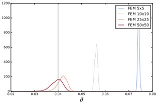
(b) SCM



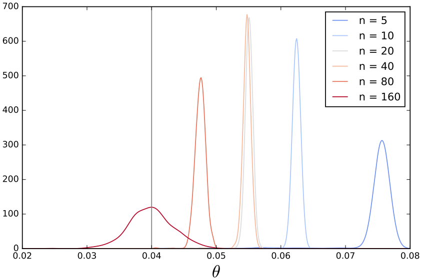
(c) PMM

Figure: Comparison of posteriors for θ with various forward models (likelihoods). Ground truth $\theta^\dagger = 0.04$, with prior $\theta \sim \mathrm{Unif}(0.02, 0.15)$. PN forward model uses squared exponential kernel, marginalising over a half-range Cauchy length scale parameter. Details of pseudomarginal MCMC etc. in Cockayne et al. (2016, Section 7).

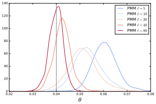
We try to recover θ from 16 observations of u on a 4×4 regular interior grid, corrupted by $\mathcal{N}(0, \frac{1}{10}I)$ noise.



(a) FE



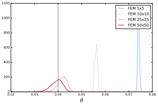
(b) SCM



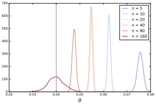
(c) PMM

Figure: Comparison of posteriors for θ with various forward models (likelihoods). Ground truth $\theta^{\dagger} = 0.04$, with prior $\theta \sim \mathrm{Unif}(0.02, 0.15)$. PN forward model uses squared exponential kernel, marginalising over a half-range Cauchy length scale parameter. Details of pseudomarginal MCMC etc. in Cockayne et al. (2016, Section 7).

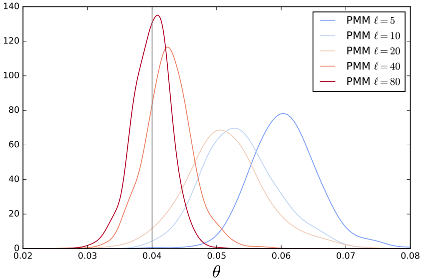
We try to recover θ from 16 observations of u on a 4×4 regular interior grid, corrupted by $\mathcal{N}(0, \frac{1}{10} I)$ noise.



(a) FE



(b) SCM



(c) PMM

Figure: Comparison of posteriors for θ with various forward models (likelihoods). Ground truth $\theta^{\dagger} = 0.04$, with prior $\theta \sim \mathrm{Unif}(0.02, 0.15)$. PN forward model uses squared exponential kernel, marginalising over a half-range Cauchy length scale parameter. Details of pseudomarginal MCMC etc. in Cockayne et al. (2016, Section 7).

Allen–Cahn: A-Optimal Experimental Design

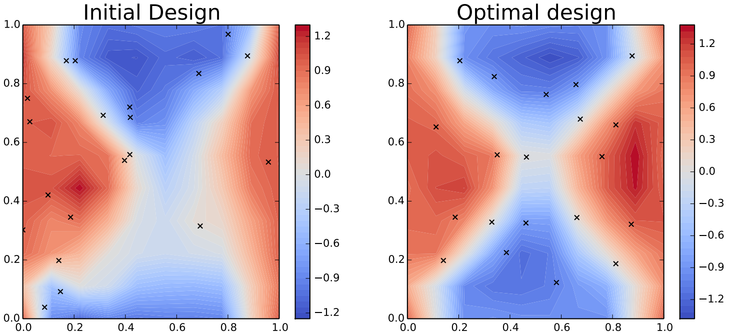


Figure: A-optimal (minimal trace of the posterior covariance) locations of 20 RHS evaluations for the PMM forward solution u. In practice, the optimal designs for $\theta = \theta^\dagger = 0.04$ and other values of θ were indistinguishable in the 'eyeball norm'.

Next Steps

- Statistical methods for the approximation of the Green's function or the 'natural' optimal kernel (White and Stuart, 2009; Fasshauer, 2012) — if the cost-accuracy tradeoff relative to the 'practical' integral kernel makes it worthwhile.
- Connections to Bayesian numerical homogenisation (Owhadi, 2015) and the gamblet method (Owhadi, 2016).
- Extensions to evolutionary problems and more profound nonlinearities.
- Escape from the Gaussian Alcatraz!

Overview

General Comments

- PN offers ways to fold uncertainty arising from numerical error into inferences, and propagate this uncertainty to later computations.
  - Thus, we don't confuse the replicability of deterministic simulations with their accuracy.
  - Good data + appropriately skeptical model $\implies$ sound inferences.
- For both ODEs and PDEs, we have a good idea of how to proceed with Gaussian priors. In some cases, we can see how Gaussian and non-Gaussian priors have the same high-precision limits, but we can't expect this Gaussian universality to always hold true.
- Numerical analysis expertise is needed to build more realistic priors.
- Statistical expertise is needed to explore their posteriors.

`http://probabilistic-numerics.org`
`https://github.com/jcockayne/bayesian_pdes`

General Comments

- PN offers ways to fold uncertainty arising from numerical error into inferences, and propagate this uncertainty to later computations.
  - Thus, we don't confuse the replicability of deterministic simulations with their accuracy.
  - Good data + appropriately skeptical model $\implies$ sound inferences.
- For both ODEs and PDEs, we have a good idea of how to proceed with Gaussian priors. In some cases, we can see how Gaussian and non-Gaussian priors have the same high-precision limits, but we can't expect this Gaussian universality to always hold true.
- Numerical analysis expertise is needed to build more realistic priors.
- Statistical expertise is needed to explore their posteriors.

http://probabilistic-numerics.org
https://github.com/jcockayne/bayesian_pdes

Thank You

Bibliography I

I. Cialenco, G. E. Fasshauer, and Q. Ye. Approximation of stochastic partial differential equations by a kernel-based collocation method. *Int. J. Comput. Math.*, 89(18):2543–2561, 2012. URL http://dx.doi.org/10.1080/00207160.2012.688111.

J. Cockayne, C. J Oates, T. J. Sullivan, and M. Girolami. Probabilistic meshless methods for partial differential equations and Bayesian inverse problems, 2016. URL http://arxiv.org/pdf/1605.07811.

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*, 2016. URL http://dx.doi.org/10.1007/s11222-016-9671-0.

P. Diaconis. Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics, IV, Vol. 1 (West Lafayette, Ind., 1986)*, pages 163–175. Springer, New York, 1988.

G. E. Fasshauer. Solving differential equations with radial basis functions: multilevel methods and smoothing. *Adv. Comput. Math.*, 11(2-3):139–159, 1999. URL http://dx.doi.org/10.1023/A:1018919824891. Radial basis functions and their applications.

G. E. Fasshauer. Green's functions: taking another look at kernel approximation, radial basis functions, and splines. In *Approximation theory XIII: San Antonio 2010*, volume 13 of *Springer Proc. Math.*, pages 37–63. Springer, New York, 2012. URL http://dx.doi.org/10.1007/978-1-4614-0772-0_4.

Bibliography II

E. Hairer, R. I. McLachlan, and A. Razakarivony. Achieving Brouwer's law with implicit Runge–Kutta methods. *BIT*, 48(2):231–243, 2008. URL http://dx.doi.org/10.1007/s10543-008-0170-3.

D. J. Higham, X. Mao, and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063 (electronic), 2002. URL http://dx.doi.org/10.1137/S0036142901389530.

J. B. Kadane. Parallel and sequential computation: a statistician's view. *J. Complexity*, 1(2): 256–263, 1985. URL http://dx.doi.org/10.1016/0885-064X(85)90014-7. Complexity of approximately solved problems (Morningside Heights, N.Y., 1985).

S. Mosbach and A. G. Turner. A quantitative probabilistic investigation into the accumulation of rounding errors in numerical ODE solution. *Comput. Math. Appl.*, 57(7):1157–1167, 2009. URL http://dx.doi.org/10.1016/j.camwa.2009.01.020.

A. O'Hagan. Some Bayesian numerical analysis. In *Bayesian statistics, 4 (Peñíscola, 1991)*, pages 345–363. Oxford Univ. Press, New York, 1992.

H. Owhadi. Bayesian numerical homogenization. *Multiscale Model. Simul.*, 13(3):812–828, 2015. URL http://dx.doi.org/10.1137/140974596.

H. Owhadi. Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games, 2016. URL http://arxiv.org/pdf/1503.03467v4.

H. Poincaré. *Calcul des Probabilites*. Georges Carré, Paris, 1896.

M. Schober, D. Duvenaud, and P. Hennig. Probabilistic ODE solvers with Runge–Kutta means. In *Advances in Neural Information Processing Systems 25*, 2014.

Bibliography III

J. Skilling. Bayesian solution of ordinary differential equations. In *Maximum Entropy and Bayesian Methods*, pages 23–37. Springer, 1992.

A. V. Sul′din. Wiener measure and its applications to approximation methods. I. *Izv. Vysš. Učebn. Zaved. Matematika*, 1959(6 (13)):145–158, 1959.

A. V. Sul′din. Wiener measure and its applications to approximation methods. II. *Izv. Vysš. Učebn. Zaved. Matematika*, 1960(5 (18)):165–179, 1960.

D. White and A. Stuart. Green's functions by Monte Carlo. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pages 627–636. Springer, Berlin, 2009. URL `http://dx.doi.org/10.1007/978-3-642-04107-5_41`.