Probability Measures on Numerical Solutions of ODEs and PDEs for Uncertainty Quantification and Inference

Patrick R. Conrad (Warwick Statistics)
Mark Girolami (Warwick Statistics)
Andrew Stuart (Warwick Mathematics)
Simo Särkkä (Aalto BECS)
Konstantinos Zygalakis (Southampton Mathematics)

June 8, 2015

This work is partly supported by the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom (grants EP/K034154/1 and EP/J016934/2)

Statistics with deterministic numerical simulations

If we have a physical model $u(\theta)$ requiring the solution of differential equations, we form an approximation,

$$\|u(\theta) - U_h(\theta)\| \leq \psi(h) \to 0 \text{ as } h \to 0.$$

Applied in forward UQ:

$$\mathbb{E}_\theta[\Phi(u(\theta))] \approx \mathbb{E}_\theta[\Phi(U_h(\theta))],$$

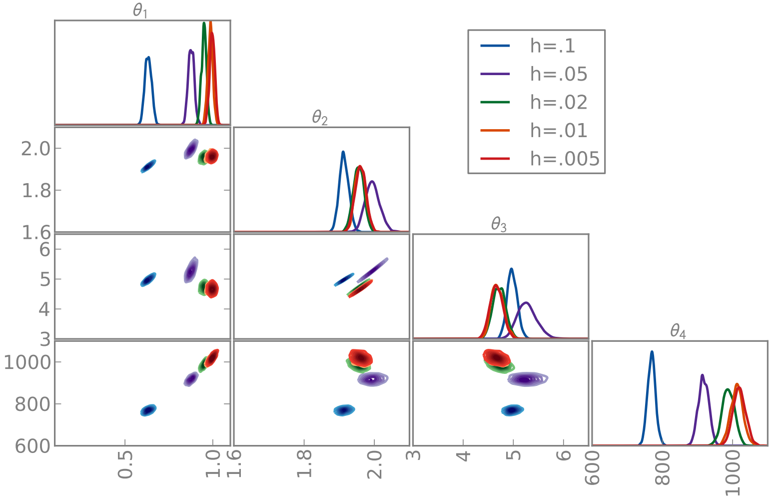
or a Bayesian inverse problem:

$$\mu(\theta) \propto \mathcal{L}(u(\theta)|d)p(\theta) \approx \mathcal{L}(U_h(\theta)|d)p(\theta).$$

This neglects uncertainty in the solver $U_h(\theta)$! Our analysis will be **over-confident** about the solution u.

Bayesian posterior with deterministic solver

Posterior is over-confident at finite h values



Statistics with random numerical simulations

Instead, form a randomized solver,

$$\mathbb{E}_{\omega}\|u(\theta) - U_{h,\omega}(\theta)\| \leq \psi(h) \to 0 \text{ as } h \to 0.$$

Applied in forward uncertainty quantification:

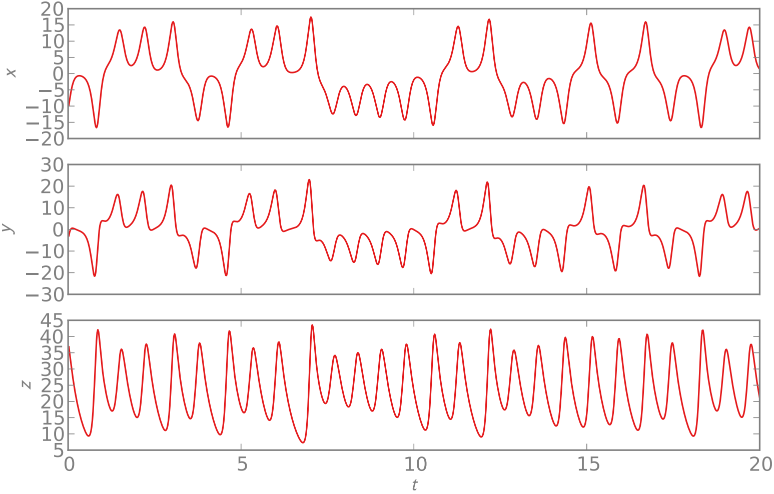
$$\mathbb{E}_{\theta}[\Phi(u(\theta))] \approx \mathbb{E}_{\theta,\omega}[\Phi(U_{h,\omega}(\theta))],$$

or a Bayesian inverse problem:

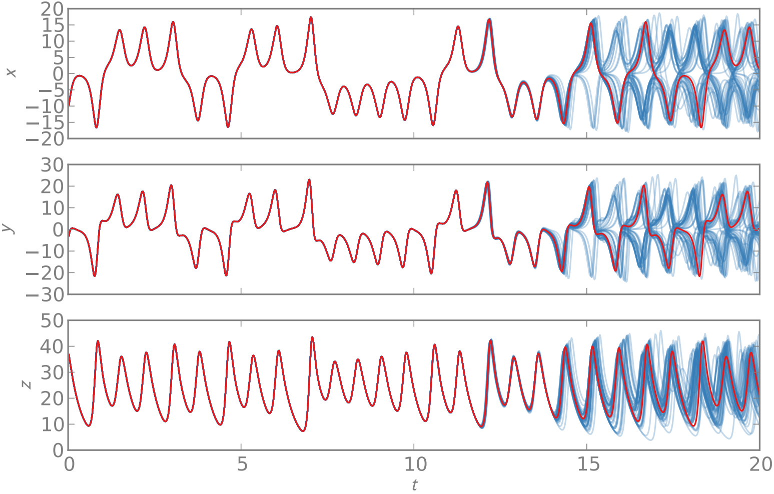
$$\mu(\theta) \propto \mathcal{L}(u(\theta)|d)p(\theta) \approx \int \mathcal{L}(U_{h,\omega}(\theta)|d)p(\theta)d\omega.$$

Allow **consistent** statistical analysis across multiple resolutions

Lorenz with Classical 4th order Runge-Kutta



Lorenz with Randomized 4th order Runge-Kutta



Lorenz movie

Overview

- Our aim is to **quantify uncertainty** in existing solvers for combination with statistical methods
- Describe uncertainty as a **measure** over solutions that contracts to the true solution
- Construct Monte Carlo samples by **perturbing** the **discretization** with random Gaussian fields
- Developed for both ODE and PDE solvers

Disclaimers:

- Not a route to faster converging solvers or eliminating bias
- Assume that analytic solution $u(\theta)$ is our objective

Existing approaches to statistical error models

Deterministic error indicators are well developed, e.g., an h refinement indicator

$$e(t) = U^h(t) - U^{h/2}(t)$$

suggest measure

$$\mu(t) = \mathcal{N}(U^h(t), e(t)^2)$$

Pointwise i.i.d. Gaussian error is too simplistic; correlations impact later analysis

Related work on statistical treatment of discretization error

- O'Hagan (1992); Skilling (1991); Diaconis (1988)
- Chkrebtii, Campbell, Girolami, Calderhead (2014)
- Schober, Duvenaud, Hennig (2014)

Randomized ODE solvers

Consider the ODE:

$$\frac{du}{dt} = f(u), \quad u(0) = u_0.$$

Integral equation

Choose a fixed step size h. For $u_k = u(kh)$. For $t \in [t_k, t_{k+1}]$:

$$u(t) = u_k + \int_{t_k}^{t} f(u(s))\, ds$$

Some approximation is required to create a numeric method.

Randomized ODE solvers (cont)

One-step numerical method

For $U_k \approx u(kh)$:

$$U_{k+1} = \Psi_h(U_k), \quad U_0 = u_0.$$

Continuous approximation:

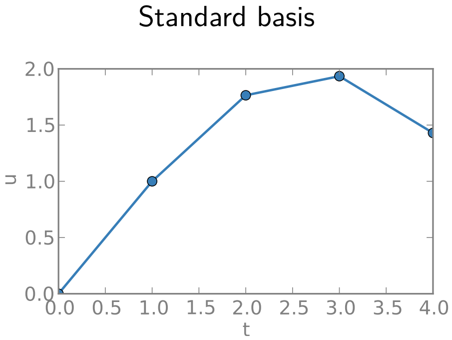
$$U(t) \approx \Psi_{t-t_k}(U_k)$$

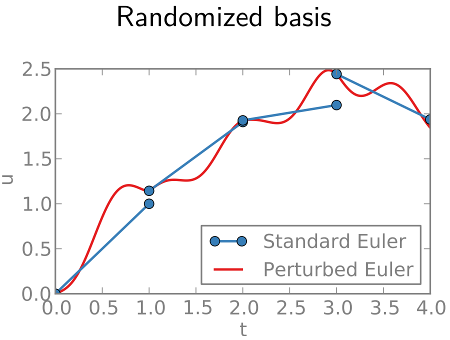
Randomized numerical method

Assume the flow map is perturbed by a Gaussian process, $\xi_k(\cdot)$ defined on $[0, h]$, where $\xi_k(\cdot) = 0$. This gives approximation $U(t)$ for $t \in [t_k, t_{k+1}]$:

$$\begin{aligned} U(t) &= \Psi_{t-t_k}(U_k) + \xi_k(t - t_k), \\ U_{k+1} &= \Psi_h(U_k) + \xi_k(h). \end{aligned}$$

Illustration of randomized ODE step





Randomized solver is **locally** Gaussian, but **globally** non-Gaussian

Assumptions

Assumption 1

Let there exist $K > 0, p \geq 1$ such that, for all $t \in [0, h]$,

$$\mathbb{E}\left|\xi(t)\xi(t)^T\right|_{\mathrm{F}}^2 \leq Kt^{2p+1}.$$

Furthermore, assume there is a constant σ, independent of h, such that

$$\mathbb{E}[\xi(h)\xi(h)^T] = \sigma h^{2p+1} I.$$

Assumption 2

The function f and a sufficient number of its derivatives are bounded uniformly in $\mathbb{R}^n$ in order to ensure that f is globally Lipschitz and that the numerical flow-map Ψ_h has uniform local truncation error of order $q+1$ with respect to the true flow-map Φ_h:

$$\sup_{u \in \mathbb{R}^n} |\Psi_t(u) - \Phi_t(u)| \leq Kt^{q+1}.$$

Convergence result

Theorem

Under Assumptions 1 and 2 it follows that there is $K > 0$ such that

$$\sup_{0 \le kh \le T} \mathbb{E}|u_k - U_k|^2 \le K h^{2\min\{p,q\}}.$$

Furthermore

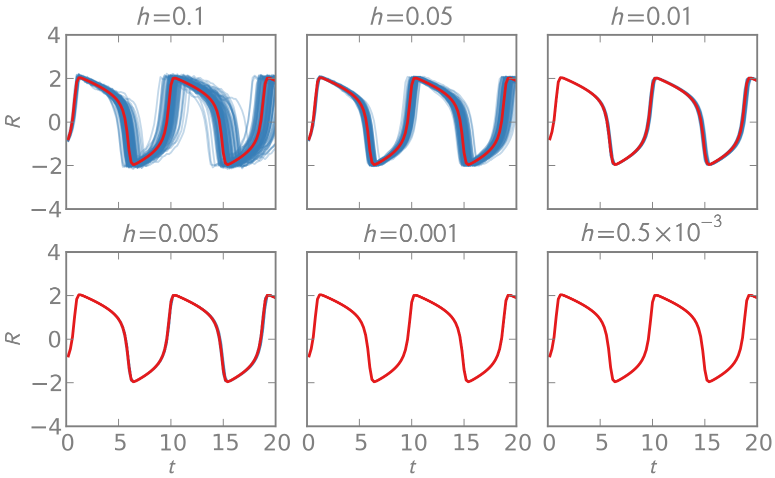
$$\sup_{0 \le t \le T} \mathbb{E}|u(t) - U(t)| \le K h^{\min\{p,q\}}.$$

Scaling of Noise

- Optimal scaling of noise is $p = q$.
- Then deterministic rate of convergence is unaffected.
- But maximal noise is added to the system.
- Fit constant σ to an error estimator.

Convergence of random solutions

Draws from the random solver for fixed σ



Backward error analysis

Modified (Stochastic Differential) Equation

$$\frac{du^h}{dt} = f(u^h) + h^q \sum_{l=0}^{q} h^\ell f_\ell(u^h) + \sqrt{\sigma h^{2q}} \frac{dW}{dt}, \quad u^h(0) = u_0$$

Theorem

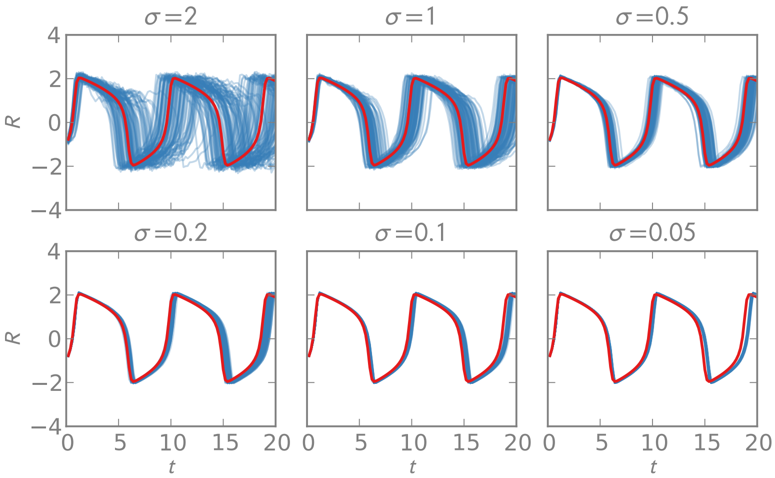
Under Assumptions 1 and 2, for Φ a C^∞ function with all derivatives bounded uniformly on $\mathbb{R}^n$, there is a choice of $\{f_\ell\}_{\ell=0}^q$ such that

$$\Big|\Phi(u(T)) - \mathbb{E}\Phi(U_k)\Big| \le Kh^q, \quad kh = T.$$

and

$$\Big|\mathbb{E}\Phi(u^h(T)) - \mathbb{E}\Phi(U_k)\Big| \le Kh^{2q+1}, \quad kh = T.$$

Draws from the random solver for fixed $h = 0.1$



Choosing scale of random perturbations

The scale σ is problem dependent, choose it to match the classical error indicator, by sampling

$$p(\sigma) \propto \exp\left[-d\left(\mathcal{N}(\mathbb{E}[U^h_\sigma], \mathbb{V}[U^h_\sigma]), \mathcal{N}(U^h(t), e(t)^2)\right)\right],$$

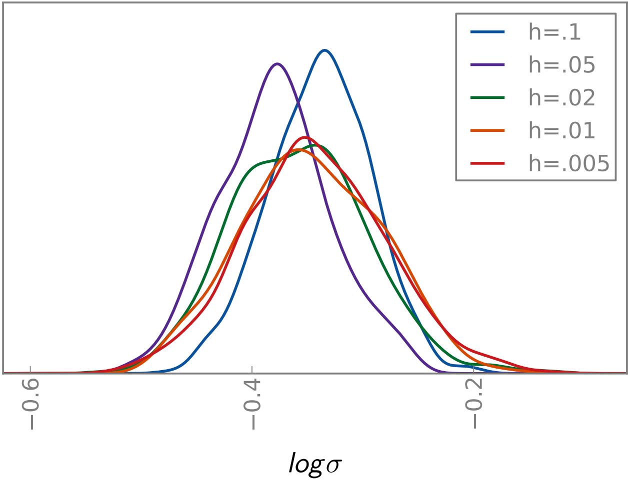
or optimizing

$$\min_\sigma d\left(\mathcal{N}(\mathbb{E}[U^h_\sigma], \mathbb{V}[U^h_\sigma]), \mathcal{N}(U^h(t), e(t)^2)\right).$$

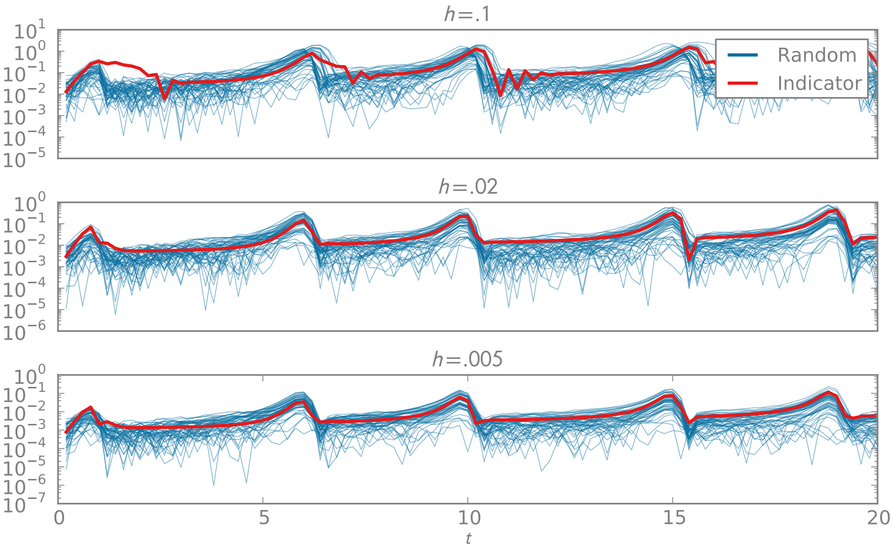
The Bhattacharyya distance works well

$$d\left(\mathcal{N}(\mu_p, \sigma_p^2), \mathcal{N}(\mu_q, \sigma_q^2)\right) = \frac{1}{4}\left(\ln\frac{1}{4}\left(\frac{\sigma_p^2}{\sigma_q^2} + \frac{\sigma_q^2}{\sigma_p^2} + 2\right)\right) + \frac{1}{4}\left(\frac{(\mu_p - \mu_q)^2}{\sigma_p^2 + \sigma_q^2}\right)$$

Density of scale parameter in FitzHugh-Nagumo



Calibrated difference from deterministic solution



Bayesian posterior

For a true ODE problem,

$$\frac{du}{dt} = f(u, \theta), \qquad u(0) = u_0,$$

construct the true posterior,

$$\mathbb{P}(\theta|\mathbf{d}) \propto \pi(\theta)\mathcal{L}(u(t, \theta)|\mathbf{d}).$$

Given a numerical approximation, $U^h(t)$, construct approximate posterior,

$$\approx \mathbb{P}_h(\theta|\mathbf{d}) \propto \pi(\theta)\mathcal{L}(U^h(t, \theta)|\mathbf{d}).$$

Typically convergent, in the sense that [Cotter, Dashti, Stuart]

$$d_{\text{Hell}}\left(\mathbb{P}_h(\theta|\mathbf{d}), \mathbb{P}(\theta|\mathbf{d})\right) \to 0 \text{ as } h \to 0$$

Modifying the inference problem

Deterministic solver

$$\mathbb{P}_h(\theta|\mathbf{d}) \propto \pi(\theta)\mathcal{L}(U^h(t,\theta)|\mathbf{d}).$$

Deterministic error indicator

$$\mathbb{P}_h(\theta|\mathbf{d}) \propto \pi(\theta)\int \mathcal{L}(U^h(t)+\xi(t)|\mathbf{d})d\xi(t)$$

$$\xi(t) \sim \mathcal{GP}(0, e(t)^2),\ \text{either i.i.d. or AR(1)}$$

Randomized solver

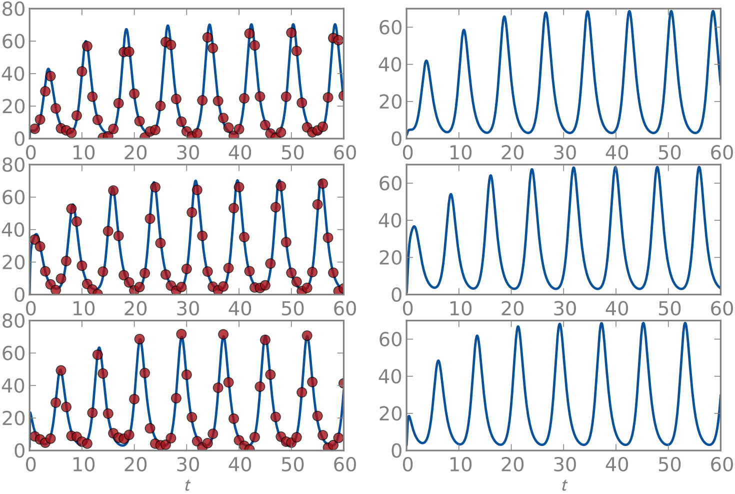
$$\mathbb{P}_h(\theta|\mathbf{d}) \propto \pi(\theta)\int \mathcal{L}(U^h_\sigma(t|\xi)|\mathbf{d})d\xi$$

Apply noisy pseudomarginal MCMC to sample integrals

Repressilator inference

Protein concentration

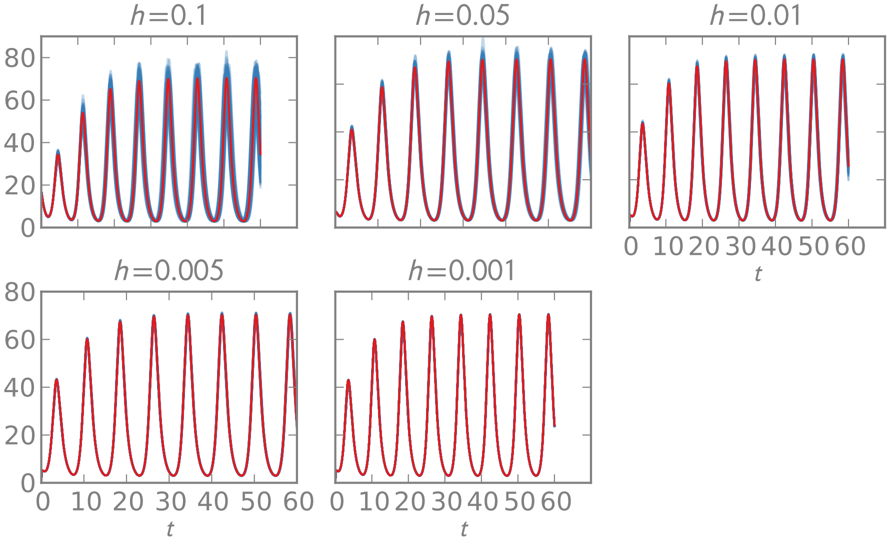
mRNA



Elowitz and Leibler, 2000

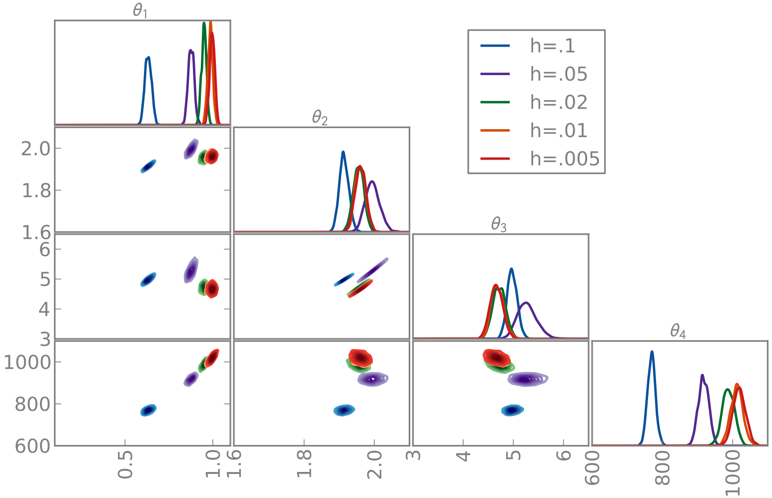
Repressilator random integrals

Inference uses 2nd order Runge-Kutta



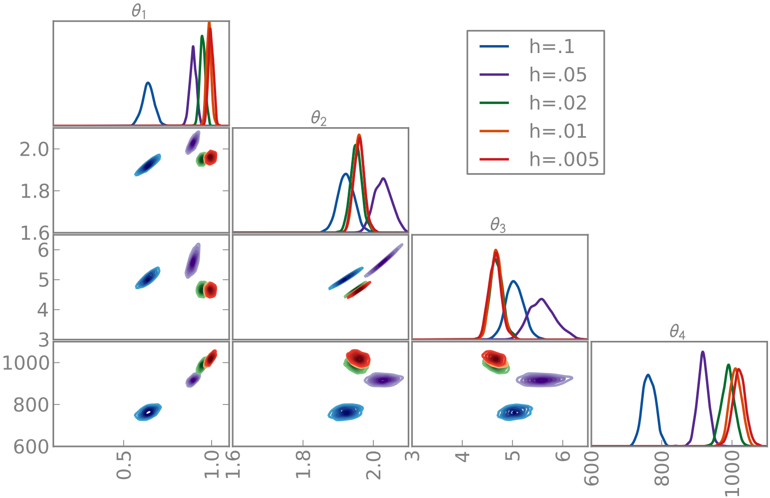
Repressilator posterior with deterministic solver

Posterior is over-confident at finite h values



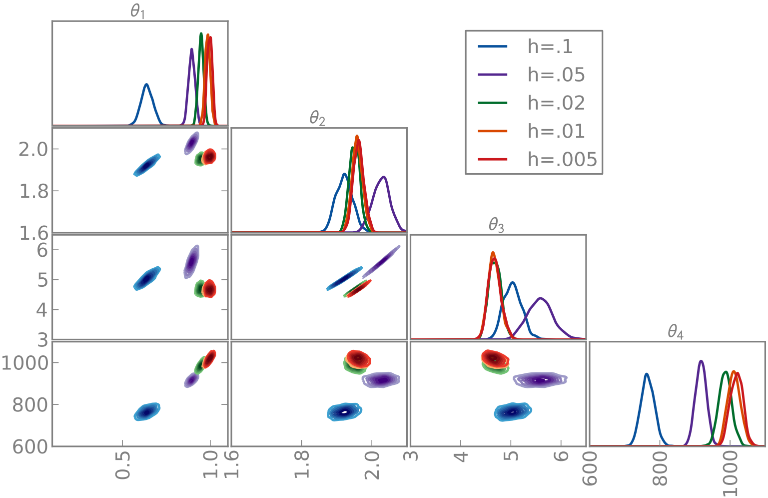
Repressilator posterior with error indicator and i.i.d.

Uncorrelated model error has little impact



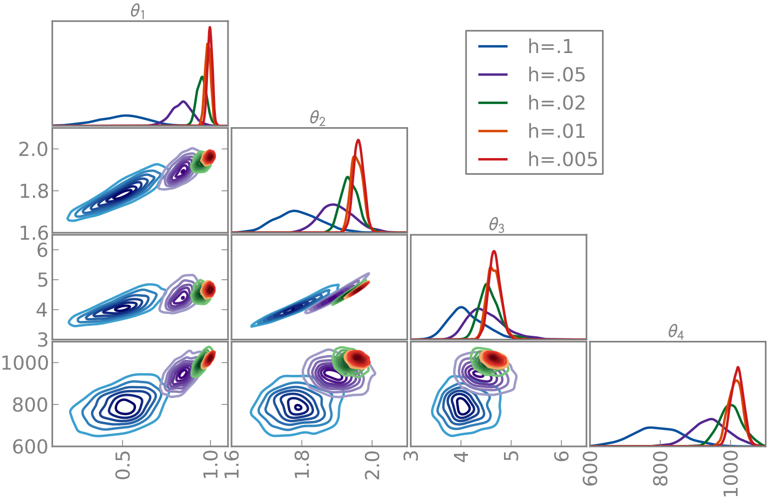
Repressilator posterior with error indicator and AR(1)

Simple correlation model has little impact



Repressilator posterior with random solver

Posterior still contains bias, but posterior width reflects error



Summary of random ODE solvers

1. Insert uncertainty into discretisation with local Gaussian processes
2. Prove convergence of random solver and backwards error analysis
3. Scale noise in practice by matching error indicators
4. Demonstrate improved results on inference

Randomizing standard PDE solvers

Weak Form

$$u \in \mathcal{V} : \; a(u, v) = r(v), \quad \forall v \in \mathcal{V}.$$

Galerkin Method

$$u^h \in \mathcal{V}^h : a(u^h, v) = r(v), \quad \forall v \in \mathcal{V}^h.$$

Then

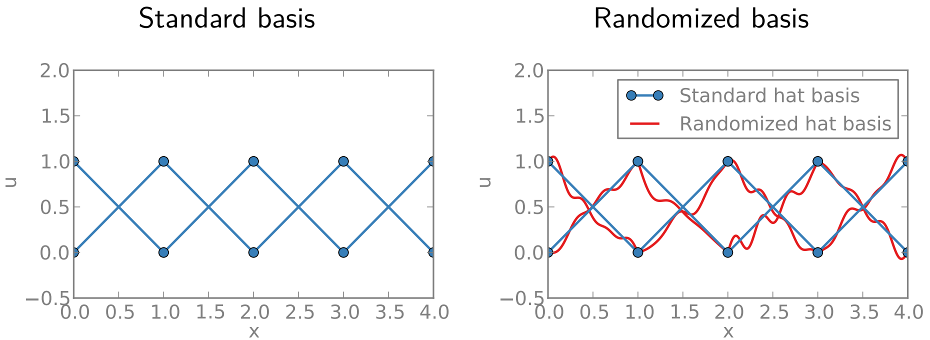
$$\mathcal{V}^h = \mathrm{span}\{\Phi_j = \Phi_j^{\mathrm{s}}\}_{j=1}^{J}.$$

Randomized Galerkin Method

$\mathcal{V}^h$ comprises small randomized perturbations of the standard Galerkin method:

$$\mathcal{V}^h = \mathrm{span}\{\Phi_j = \Phi_j^{\mathrm{s}} + \Phi_j^{\mathrm{r}}\}_{j=1}^{J}.$$

Illustration of randomized PDE basis



- For nodal points x_k, $\Phi_j^r(x_k) = 0$.
- Choose $\operatorname{supp} \Phi_j^s = \operatorname{supp} \Phi_j^r$ to maintain sparsity
- Greater flexibility in choosing properties of random field
- Randomness generated in advance, solver step is unaffected

Assumptions

Assumption 1

The $\{\Phi^{\mathrm{r}}_j\}_{j=1}^J$ are independent, mean zero, Gaussian random fields, with the same support as the $\{\Phi^{\mathrm{s}}_j\}$, and satisfying

$$\Phi^{\mathrm{r}}_j(x_k) = 0, \quad \sum_{j=1}^{J} \mathbb{E}\|\Phi^{\mathrm{r}}_j\|_a^2 \le Ch^{2q}.$$

Assumption 2

The true solution u of problem (30) is in $L^\infty(D)$. Furthermore the standard deterministic interpolant of the true solution, defined by

$$v^{\mathrm{s}} := \sum_{j=1}^{J} u(x_j)\Phi^{\mathrm{s}}_j,$$

satisfies $\|u - v^{\mathrm{s}}\|_a \le Ch^p$.

Convergence result

Theorem

Under Assumptions 1 and 2 it follows that the random approximation U^h satisfies

$$\mathbb{E}\|u - U^h\|_a^2 \leq Ch^{2\min\{p,q\}}.$$

Corollary

Consider the Poisson equation with Dirichlet boundary conditions and a random perturbation of the piecewise linear FEM approximation, with $p = q = 1$. Under Assumptions 1 and 2 it follows that the random approximation U^h satisfies

$$\mathbb{E}\|u - U^h\|_{L^2} \leq Ch^2.$$

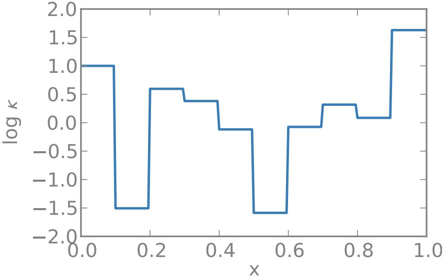
Elliptic PDE inverse problem

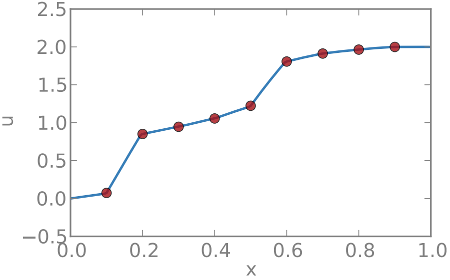
Standard elliptic inversion problem:

$$\nabla \cdot (\kappa(x) \nabla u(x)) = 4x$$

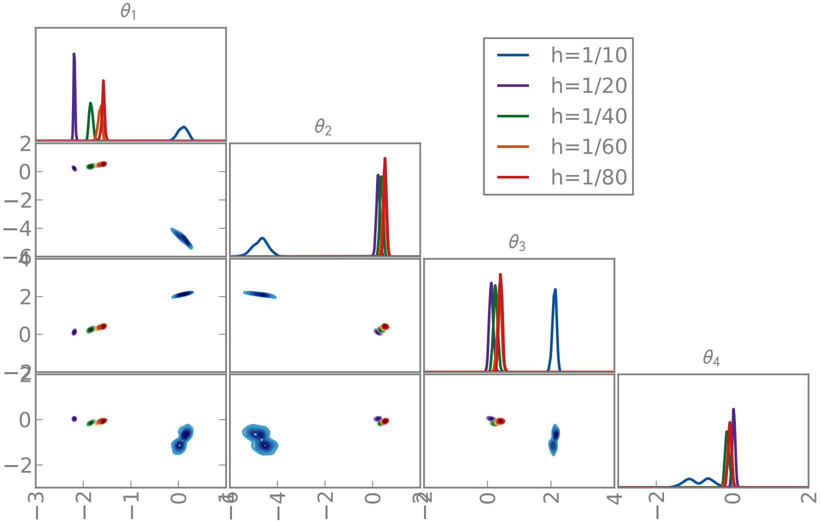
$$u(0) = 0, u(1) = 2$$

Data with small i.i.d. Gaussian error

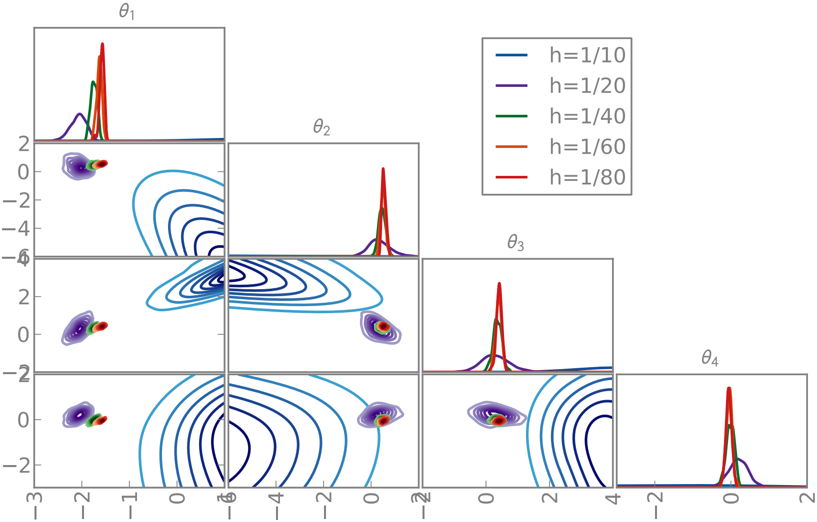




Elliptic inference with standard solver



Elliptic inference with random solver

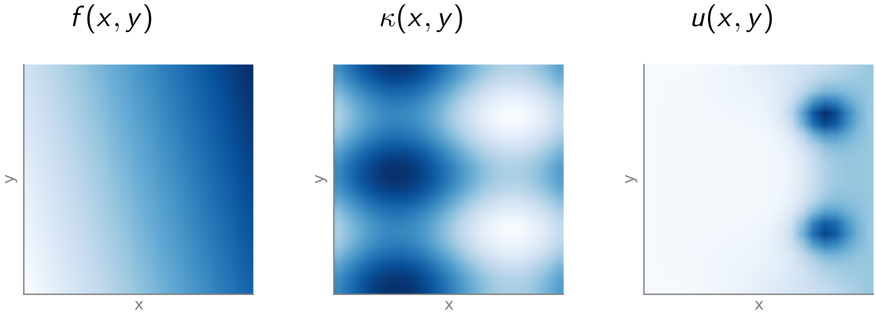


2D Elliptic problem

Standard 2D elliptic equation:

$$\nabla \cdot (\kappa(x, y)\nabla u(x, y)) = f(x, y)$$

Solved on a 30×30 grid. Perturbation fields are $\mathcal{N}(0, (-\triangle)^{-2})$

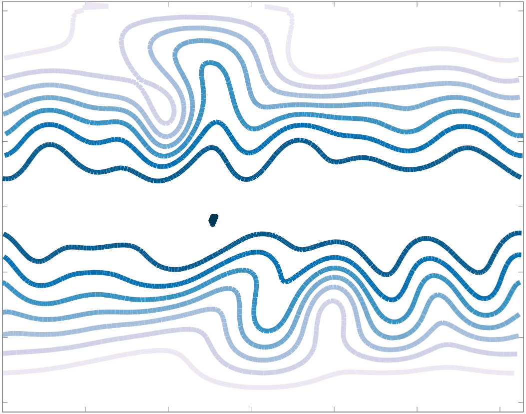


Perturbations in random solutions

Top PCA modes of perturbation are shown

Shallow water equation solver

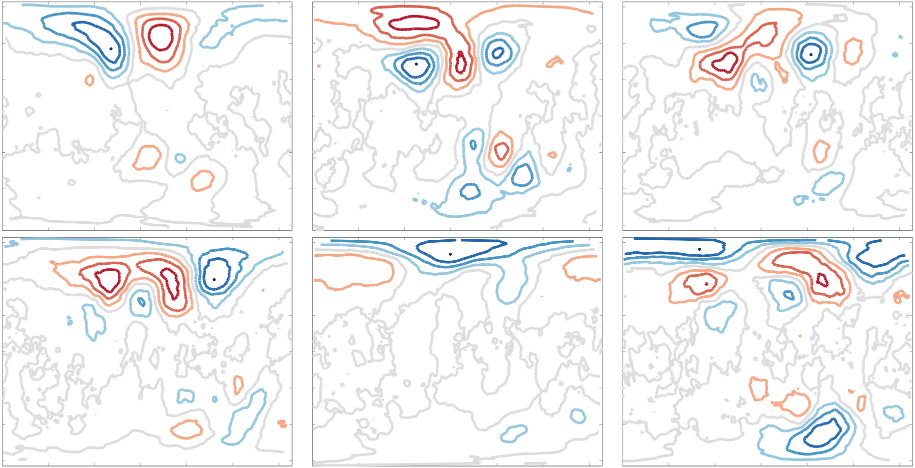
- Vortex shedding on a global shallow water model
- Advanced mimetic finite element scheme $\approx$ 10,000 DOF
- Introduced perturbations to ODE step and roughly calibrated scale



Solver due to Thuburn, Cotter (2014)

Perturbations to shallow water solver

Top PCA modes of perturbation are shown



Contributions

- Our aim is to **quantify uncertainty** in existing solvers for combination with statistical methods
- Describe uncertainty as a **measure** over solutions that contracts to the true solution
- Construct Monte Carlo samples by **perturbing** the **discretization** with random Gaussian fields
- Developed for both ODE and PDE solvers
- Simple construction can be adapted to many useful solvers
- Demonstrated more consistent statistical analysis using randomized solvers

Future work

- Study other classes of PDEs and backwards error analysis
- Extend to other types of model error (e.g., dimension reduction)
- Apply to other problem classes, e.g., Stochastic Differential Equations

- Study rates of convergence of forward UQ and Bayesian posteriors
- Computational issues surrounding pseudomarginal posteriors
- Leverage efficient statistical methods, e.g., multilevel sampling
- Extensions to Bayesian inference on infinite dimensional spaces

- Apply to large, real-world applications, such as shallow water equations