Probabilistic methods for uncertainty quantification of the error in numerical solvers of differential equations

Assyr Abdulle, Giacomo Garegnani

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Candidacy Exam, 20 March 2018

Outline

Outline

Probabilistic methods – why?

Consider Lorenz equation (atmospheric convection)

$$\begin{aligned} x' &= \sigma(y - x), & x(0) &= -10, \\ y' &= x(\rho - z) - y, & y(0) &= -1, \\ z' &= xy - \beta z, & z(0) &= 40. \end{aligned}$$

For $\rho = 28$, $\sigma = 10$, $\beta = 8/3$ chaotic behaviour.

$\implies$ Numerical integration gives unreliable solutions.

Probabilistic methods – why?

Consider Lorenz equation (atmospheric convection)

$$\begin{aligned} x' &= \sigma(y - x), & x(0) &= -10, \\ y' &= x(\rho - z) - y, & y(0) &= -1, \\ z' &= xy - \beta z, & z(0) &= 40. \end{aligned}$$

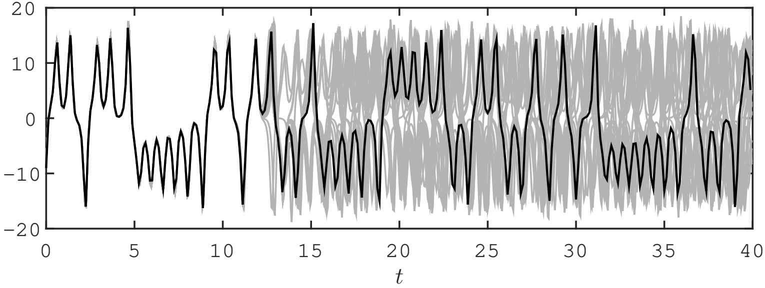
For $\rho = 28$, $\sigma = 10$, $\beta = 8/3$ chaotic behaviour.

$\implies$ Numerical integration gives unreliable solutions.

Goal

Establish a probability measure over the numerical solution given by classical methods.

Probabilistic methods – why?

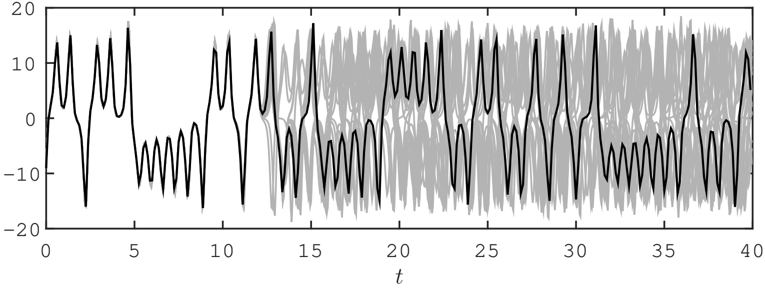


Time evolution of the first component of Lorenz equation

Black line $\rightarrow$ deterministic solution

Gray lines $\rightarrow$ probabilistic solutions.

Probabilistic methods – why?



Time evolution of the first component of Lorenz equation

Black line $\rightarrow$ deterministic solution

Gray lines $\rightarrow$ probabilistic solutions.

Chaotic behaviour appears frequently in nonlinear differential equations.

Outline

Notation

Autonomous dynamical system, function $f\colon \mathbb{R}^d \to \mathbb{R}^d$ and the ODE

$$y' = f(y), \quad y(0) = y_0.$$

Flow of the equation $\varphi_t\colon \mathbb{R}^d \to \mathbb{R}^d$ such that

$$y(t) = \varphi_t(y_0).$$

One-step method: numerical flow Ψ_h such that

$$y_{n+1} = \Psi_h(y_n).$$

Runge-Kutta methods: flow implicitly defined by

$$K_i = y_n + h\sum_{j=1}^{s} a_{ij} f(K_j),$$

$$\Psi_h(y_n) = y_n + h\sum_{i=1}^{s} b_i f(K_i).$$

Probabilistic methods for ODEs

Filtering methods for ODEs: fix a prior on $y(t)$ (Gaussian process), update with evaluations of $f(y)$ [Kersting and Hennig, 2016]

Randomised methods for ODEs: random perturbation of deterministic numerical solutions $\rightarrow$ sampling [Conrad et al., 2016]

Probabilistic methods for ODEs

Filtering methods for ODEs: fix a prior on $y(t)$ (Gaussian process), update with evaluations of $f(y)$ [Kersting and Hennig, 2016]

Randomised methods for ODEs: random perturbation of deterministic numerical solutions $\rightarrow$ sampling [Conrad et al., 2016]

Additive noise method [Conrad et al., 2016]

Stochastic process $\{Y_n\}_{n=1,2,\dots}$ with recurrence

$$Y_{n+1} = \underbrace{\Psi_h(Y_n)}_{\text{deterministic}} + \underbrace{\xi_n(h)}_{\text{random}}.$$

Main assumption: $\{\xi_n\}_{n=0,1,\dots}$ iid such that for $p > 1$ and $Q \in \mathbb{R}^{d \times d}$

$$\mathbb{E}\,\xi_n(h) = 0, \quad \mathbb{E}\,\xi_n(h)\xi_n(h)^T = Qh^{2p+1}.$$

Additive noise method [Conrad et al., 2016]

Stochastic process $\{Y_n\}_{n=1,2,\dots}$ with recurrence

$$Y_{n+1} = \underbrace{\Psi_h(Y_n)}_{\text{deterministic}} + \underbrace{\xi_n(h)}_{\text{random}}.$$

Main assumption: $\{\xi_n\}_{n=0,1,\dots}$ iid such that for $p > 1$ and $Q \in \mathbb{R}^{d\times d}$

$$\mathbb{E}\,\xi_n(h) = 0, \quad \mathbb{E}\,\xi_n(h)\xi_n(h)^T = Qh^{2p+1}.$$

Properties

If Ψ_h is of order q and for $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ smooth

- Strong convergence: $\mathbb{E}\|y(hn) - Y_n\| \leq Ch^{\min\{p,q\}}$,
- Weak convergence: $|\Phi\big(y(hn)\big) - \mathbb{E}\,\Phi(Y_n)| \leq Ch^{\min\{2p,q\}}$,
- Good qualitative behavior in Bayesian inverse problems.

Additive noise method [Conrad et al., 2016]

Stochastic process $\{Y_n\}_{n=1,2,\dots}$ with recurrence

$$Y_{n+1} = \underbrace{\Psi_h(Y_n)}_{\text{deterministic}} + \underbrace{\xi_n(h)}_{\text{random}}.$$

Main assumption: $\{\xi_n\}_{n=0,1,\dots}$ iid such that for $p > 1$ and $Q \in \mathbb{R}^{d\times d}$

$$\mathbb{E}\,\xi_n(h) = 0, \quad \mathbb{E}\,\xi_n(h)\xi_n(h)^T = Qh^{2p+1}.$$

Issues

- Robustness: $\Psi_h(Y_{n-1}) > 0 \;\not\Rightarrow\; \mathbb{P}(Y_n < 0) = 0$,
- Geometric properties are not conserved from Ψ_h. For example if $I(y) = y^T S y$ and $I(\Psi_h(y_0)) = I(y_0)$

$$I(Y_1) = I(y_0) + 2\xi_0(h)^T S \Psi_h(y_0) + \xi_0(h)^T S \xi_0(h).$$

Random time steps

Intrinsic noise: Random time-stepping Runge-Kutta (RTS-RK)

$$Y_{n+1} = \Psi_{H_n}(Y_n),$$

Main assumption: $\{H_n\}_{n=0,1,\dots}$ iid such that for $h, C > 0$ and $p > 1$

$$H_n > 0 \text{ a.s.}, \quad \mathbb{E}\, H_n = h, \quad \operatorname{Var} H_n = Ch^{2p}.$$

Example: $H_n \overset{\text{iid}}{\sim} \mathcal{U}(h - h^p, h + h^p)$.

Random time steps

Theorem (Weak convergence)

There exists $C > 0$ independent of h such that for all $\Phi \colon \mathbb{R}^d \to \mathbb{R}$ smooth

$$|\mathbb{E}\,\Phi(Y_k) - \Phi(y(kh)))| \leq Ch^{\min\{2p-1,q\}}.$$

Random time steps

Theorem (Weak convergence)

There exists $C > 0$ independent of h such that for all $\Phi \colon \mathbb{R}^d \to \mathbb{R}$ smooth

$$|\mathbb{E}\,\Phi(Y_k) - \Phi(y(kh)))| \leq Ch^{\min\{2p-1,q\}}.$$

Theorem (Mean square convergence)

There exists $C > 0$ independent of h such that

$$\left(\mathbb{E}\|Y_k - y(t_k)\|^2\right)^{1/2} \leq Ch^{\min\{p-1/2,q\}}.$$

Random time steps

Theorem (Weak convergence)

There exists $C > 0$ independent of h such that for all $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ smooth

$$|\mathbb{E}\,\Phi(Y_k) - \Phi(y(kh)))| \leq Ch^{\min\{2p-1,q\}}.$$

Theorem (Mean square convergence)

There exists $C > 0$ independent of h such that

$$\left(\mathbb{E}\|Y_k - y(t_k)\|^2\right)^{1/2} \leq Ch^{\min\{p-1/2,q\}}.$$

Consequences

- Reasonable choice $p = q + 1/2$
- $\mathbb{E}\|Y_k - y(t_k)\| \leq Ch^{\min\{p-1/2,q\}}$ (strong order)

Random time steps

Theorem (Monte Carlo estimators)

For $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ smooth, Monte Carlo estimators $\hat{Z} = M^{-1} \sum_{i=1}^{M} \Phi(Y_N^{(i)})$ of $Z = \Phi(Y_N)$ satisfy

$$\mathrm{MSE}(\hat{Z}) \leq C\Big(h^{2\min\{2p-1,q\}} + \frac{h^{2\min\{p-1/2,q\}}}{M}\Big),$$

where C is a positive constant independent of h and M and

$$\mathrm{MSE}(\hat{Z}) = \mathbb{E}\left(\hat{Z} - \Phi(y(t_N))\right)^2.$$

Random time steps

Theorem (Monte Carlo estimators)

For $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ smooth, Monte Carlo estimators $\hat{Z} = M^{-1}\sum_{i=1}^{M}\Phi(Y_N^{(i)})$ of $Z = \Phi(Y_N)$ satisfy

$$\mathrm{MSE}(\hat{Z}) \leq C\left(h^{2\min\{2p-1,q\}} + \frac{h^{2\min\{p-1/2,q\}}}{M}\right),$$

where C is a positive constant independent of h and M and

$$\mathrm{MSE}(\hat{Z}) = \mathbb{E}\left(\hat{Z} - \Phi(y(t_N))\right)^2.$$

Consequence

For reasonable choice $p = q + 1/2$, $\mathrm{MSE}(\hat{Z})$ converges independently of M with h (quality of the estimation independent of the number of paths)

Outline

Conservation of first integrals – Additive noise

Recall: $Y_{n+1} = \Psi_h(Y_n) + \xi_n(h)$, with $\mathbb{E}\,\xi_n(h)\xi_n(h)^\top = h^{2p+1}Q$

Linear first integrals: $I(y) = v^\top y$ such that $I(\Psi_h(Y_1)) = I(y_0)$. Then

$$I(Y_1) = v^\top(y_0 + \xi_0(h)) \implies \mathbb{E}\,I(Y_1) = I(y_0) \text{ iff } \mathbb{E}\,\xi_0(h) = 0.$$

Quadratic first integrals: $I(y) = y^\top Sy$ such that $I(\Psi_h(Y_1)) = I(y_0)$. Then

$$\begin{aligned} I(Y_1) &= I(y_0) + 2\xi_0(h)^T S\Psi_h(y_0) + \xi_0(h)^T S\xi_0(h), \\ \implies \mathbb{E}\,I(Y_1) &= I(y_0) + Q : Sh^{2p+1}, \quad (\text{with } \mathbb{E}\,\xi_0(h) = 0) \end{aligned}$$

Quadratic first integrals are not conserved on average!

Conservation of first integrals – Random time steps

Theorem (Conservation of invariants)

If the Runge-Kutta scheme defined by Ψ_h conserves an invariant $I(y)$ for an ODE, then the RTS-RK method conserves $I(y)$ for the same ODE.

Proof

If $I(\Psi_h(y)) = I(y)$ for any h, then $I(\Psi_{H_0}(y)) = I(y)$ for any value that H_0 can assume.

Symplecticity – Random time steps

Theorem (Symplecticity of the flow map)

If the flow Ψ_h of the deterministic integrator is symplectic, then the flow of the RTS-RK method is symplectic.

Idea of the proof

Adaptive time steps ruin symplectic properties if not carefully selected. Nonetheless, if the time steps are chosen independently of the solution, the flow is symplectic.

Symplecticity – Random time steps

Theorem (Symplecticity of the flow map)

If the flow Ψ_h of the deterministic integrator is symplectic, then the flow of the RTS-RK method is symplectic.

Idea of the proof

Adaptive time steps ruin symplectic properties if not carefully selected. Nonetheless, if the time steps are chosen independently of the solution, the flow is symplectic.

Remark

The symplecticity of the flow is not enough to guarantee good approximation of the Hamiltonian for long time spans.

Symplecticity – Random time steps

Energy $Q\colon \mathbb{R}^{2d} \to \mathbb{R}$ and Hamiltonian system

$$y' = J^{-1}\nabla Q(y), \quad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Symplectic integrator Ψ_h of order q.

Theorem (Conservation of the Hamiltonian)

There exist positive constants κ, C_1, C_2, C_3, independent of h such that

$$\mathbb{E}|Q(Y_n) - Q(y_0)| \leq C_1 e^{-\kappa/2h}(1 + h^{2p-1}) + C_3 h^q + C_4 n^{1/2} h^{p+q}.$$

Symplecticity – Random time steps

Energy $Q\colon \mathbb{R}^{2d} \to \mathbb{R}$ and Hamiltonian system

$$y' = J^{-1}\nabla Q(y), \quad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Symplectic integrator Ψ_h of order q.

Theorem (Conservation of the Hamiltonian)

There exist positive constants κ, C_1, C_2, C_3, independent of h such that

$$\mathbb{E}|Q(Y_n) - Q(y_0)| \leq C_1 e^{-\kappa/2h}(1 + h^{2p-1}) + C_3 h^q + C_4 n^{1/2} h^{p+q}.$$

Consequence

Up to times $n = \mathcal{O}(h^{-2p})$ (balance between h^q and h^{p+q} terms) same conservation as deterministic symplectic method.

Outline

Bayesian inverse problems

Goal

Given $\vartheta \in \mathbb{R}^n$, $f_\vartheta \colon \mathbb{R}^d \to \mathbb{R}^d$ and the ODE

$$y' = f_\vartheta(y), \quad y(0) = y_{0,\vartheta} \in \mathbb{R}^d,$$

retrieve the true value ϑ^* from observations of $y(t)$, $t > 0$.

Bayesian inverse problems

Goal

Given $\vartheta \in \mathbb{R}^n$, $f_\vartheta \colon \mathbb{R}^d \to \mathbb{R}^d$ and the ODE

$$y' = f_\vartheta(y), \quad y(0) = y_{0,\vartheta} \in \mathbb{R}^d,$$

retrieve the true value ϑ^* from observations of $y(t)$, $t > 0$.

Bayesian setting: fix prior $\pi_{\text{prior}}(\vartheta)$, consider $\mathcal{G} \colon \mathbb{R}^n \to \mathbb{R}^m$ and the observation model

$$\mathcal{Y} = \underbrace{\mathcal{G}(\vartheta^*)}_{\text{forward}} + \underbrace{\eta}_{\text{noise}}, \quad \varepsilon \sim \pi_{\text{noise}},$$

then the posterior distribution (density) is

$$\pi(\vartheta \mid \mathcal{Y}) \propto \pi_{\text{prior}}(\vartheta)\pi_{\text{noise}}(\mathcal{Y} - \mathcal{G}(\vartheta)).$$

Bayesian inverse problems

Obtaining a sample $\{\vartheta^{(i)}\}_{i=0}^{N}$ from $\pi(\vartheta \mid \mathcal{Y})$.

Algorithm: Metropolis-Hastings.

Given $\vartheta^{(0)} \in \mathbb{R}^n$, proposal $q \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $N \in \mathbb{N}$;

Compute $\pi(\vartheta^{(0)} \mid \mathcal{Y})$;

for $i = 0, \ldots, N$ **do**

Draw $\bar{\vartheta}$ from $q(\vartheta^{(i)}, \cdot)$;

Set $\vartheta^{(i+1)} = \bar{\vartheta}$ with probability

$$\alpha(\vartheta^{(i)}, \bar{\vartheta}) = \min\left\{1, \frac{\pi(\bar{\vartheta} \mid \mathcal{Y}) q(\vartheta^{(i)}, \bar{\vartheta})}{\pi(\vartheta^{(i)} \mid \mathcal{Y}) q(\bar{\vartheta}, \vartheta^{(i)})}\right\}$$

otherwise set $\vartheta^{(i+1)} = \vartheta^{(i)}$;

end

Bayesian inverse problems

Obtaining a sample $\{\vartheta^{(i)}\}_{i=0}^{N}$ from $\pi(\vartheta \mid \mathcal{Y})$.

Algorithm: Metropolis-Hastings.

Given $\vartheta^{(0)} \in \mathbb{R}^n$, proposal $q\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $N \in \mathbb{N}$;

Compute $\pi(\vartheta^{(0)} \mid \mathcal{Y})$;

for $i = 0, \ldots, N$ **do**

Draw $\bar{\vartheta}$ from $q(\vartheta^{(i)}, \cdot)$;

Set $\vartheta^{(i+1)} = \bar{\vartheta}$ with probability

$$\alpha(\vartheta^{(i)}, \bar{\vartheta}) = \min\left\{1, \frac{\pi(\bar{\vartheta} \mid \mathcal{Y})q(\vartheta^{(i)}, \bar{\vartheta})}{\pi(\vartheta^{(i)} \mid \mathcal{Y})q(\bar{\vartheta}, \vartheta^{(i)})}\right\}$$

otherwise set $\vartheta^{(i+1)} = \vartheta^{(i)}$;

end

Bayesian inverse problems

The posterior $\pi(\vartheta \mid \mathcal{Y})$ is not computable, approximate with

$$\pi^{h}(\vartheta \mid \mathcal{Y}) \propto \pi_{\text{prior}}(\vartheta)\pi_{\text{noise}}(\mathcal{Y} - \mathcal{G}^{h}(\vartheta)).$$

Bayesian inverse problems

The posterior $\pi(\vartheta \mid \mathcal{Y})$ is not computable, approximate with

$$\pi^h(\vartheta \mid \mathcal{Y}) \propto \pi_{\mathrm{prior}}(\vartheta)\pi_{\mathrm{noise}}(\mathcal{Y} - \mathcal{G}^h(\vartheta)).$$

Properties

If Ψ_h is of order q

- $d_{\mathrm{Hell}}(\pi^h, \pi) \to 0$ for $h \to 0$ with rate q
- fast MH iterations for explicit Ψ_h (and h coarse)
- explores complex posterior distributions

Bayesian inverse problems

The posterior $\pi(\vartheta \mid \mathcal{Y})$ is not computable, approximate with

$$\pi^h(\vartheta \mid \mathcal{Y}) \propto \pi_{\text{prior}}(\vartheta)\pi_{\text{noise}}(\mathcal{Y} - \mathcal{G}^h(\vartheta)).$$

Issue

- π^h concentrated around values “far” from $\vartheta^* \to$ non-predictive posterior

Bayesian inverse problems

The posterior $\pi(\vartheta \mid \mathcal{Y})$ is not computable, approximate with

$$\pi^{\boldsymbol{h},\mathrm{RTS}}(\vartheta \mid \mathcal{Y}) \propto \pi_{\mathrm{prior}}(\vartheta)\mathbb{E}^{\mathbf{H}}\pi_{\mathrm{noise}}(\mathcal{Y} - \mathcal{G}^{\mathbf{H}}(\vartheta)),$$

where $\mathbf{H} = (H_0, H_1, \ldots)$.

Bayesian inverse problems

The posterior $\pi(\vartheta \mid \mathcal{Y})$ is not computable, approximate with

$$\pi^{h,\mathrm{RTS}}(\vartheta \mid \mathcal{Y}) \propto \pi_{\mathrm{prior}}(\vartheta)\mathbb{E}^{\mathbf{H}}\pi_{\mathrm{noise}}(\mathcal{Y} - \mathcal{G}^{\mathbf{H}}(\vartheta)),$$

where $\mathbf{H} = (H_0, H_1, \ldots)$.

Properties

If $\Psi_h \to \varphi_h$ for $h \to 0$

- $d_{\mathrm{Hell}}(\pi^{h,\mathrm{RTS}}, \pi) \to 0$ for $h \to 0$ [Lie et al., 2017]
- "correct" the non-predictive behaviour of deterministic approximations
- explores complex posterior distributions

Bayesian inverse problems

The posterior $\pi(\vartheta \mid \mathcal{Y})$ is not computable, approximate with

$$\pi^{\boldsymbol{h},\mathrm{RTS}}(\vartheta \mid \mathcal{Y}) \propto \pi_{\mathrm{prior}}(\vartheta)\mathbb{E}^{\mathbf{H}}\pi_{\mathrm{noise}}(\mathcal{Y} - \mathcal{G}^{\mathbf{H}}(\vartheta)),$$

where $\mathbf{H} = (H_0, H_1, \dots)$.

Issues

- Approximation of $\mathbb{E}^{\mathbf{H}}\,\pi_{\mathrm{noise}}(\mathcal{Y} - \mathcal{G}^{\mathbf{H}}(\vartheta))$ is required
- Employ pseudo-marginal MH $\rightarrow$ slow mixing for small noise
- Employ noisy pseudo-marginal MH $\rightarrow$ inexact posterior distributions

Outline

Numerical experiments – Geometric properties

Consider the perturbed Kepler equation (model for two-body problem)

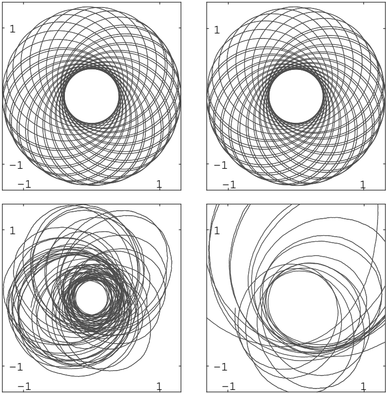
$$q_1' = p_1, \quad p_1' = -\frac{q_1}{\|q\|^3} - \frac{\delta q_1}{\|q\|^5},$$
$$q_2' = p_2, \quad p_2' = -\frac{q_2}{\|q\|^3} - \frac{\delta q_2}{\|q\|^5}.$$

The angular momentum is conserved (quadratic first integral)

$$I(p, q) = q_1 p_2 - q_2 p_1$$

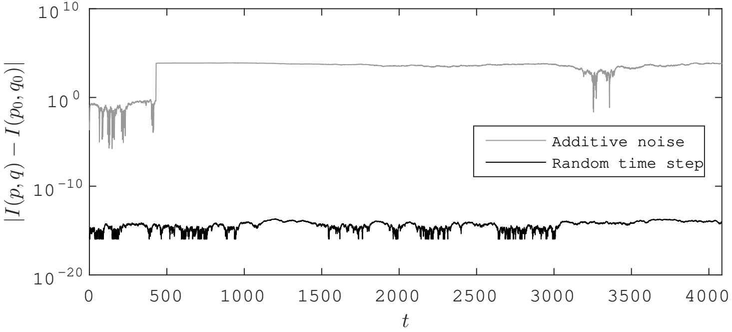
$\rightarrow$ employ a Gauss method (implicit midpoint rule).

Numerical experiments – Geometric properties



RTS-RK (first row), Additive noise (second row). Time $0 \leq t \leq 200$ and $200 \leq t \leq 400$ (left and right)

Numerical experiments – Geometric properties



Conservation of the angular momentum (quadratic first integral)

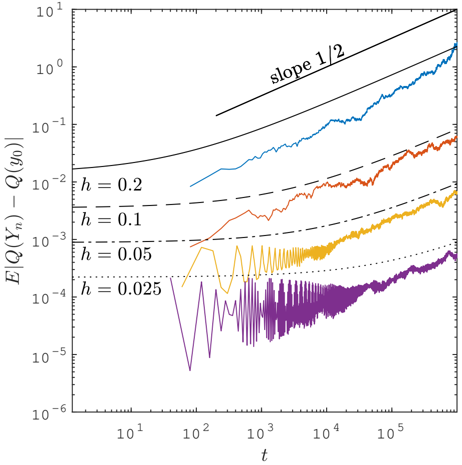
Numerical experiments – Geometric properties

Consider the pendulum system, Hamiltonian with energy

$$Q(p, q) = \frac{1}{2}p^2 - \cos(q).$$

Energy is separable $\rightarrow$ employ Störmer-Verlet (or symplectic Euler).

Numerical experiments – Geometric properties



Mean error on the Hamiltonian for different values of the time step h.

Numerical experiments – Bayesian inverse problems

Consider the Hénon-Heiles system (motion of a star around a galactic center), Hamiltonian with energy

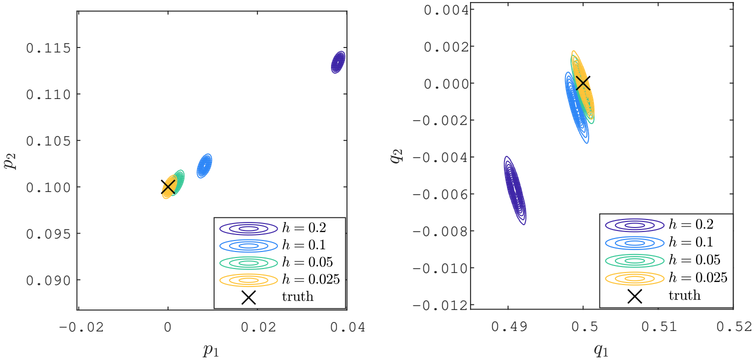
$$E(p, q) = \frac{1}{2}\|p\|^2 + \frac{1}{2}\|q\|^2 + q_1^2 q_2 - \frac{1}{3} q_2^3.$$

Chaotic problem for certain levels of energy.

Goal

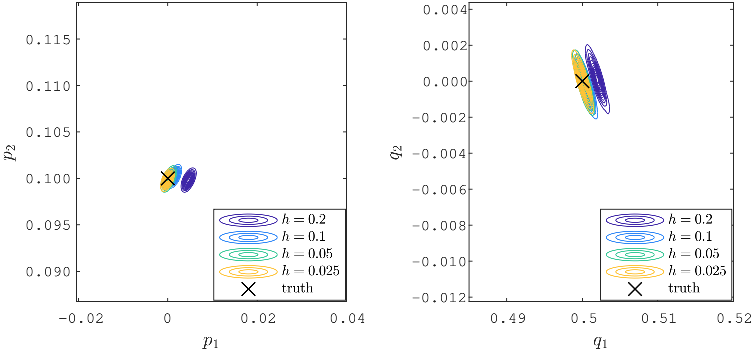
Find posterior $\pi((p_0, q_0) \mid \mathcal{Y})$ over the initial condition from a single observation of $(p(10), q(10))$

Numerical experiments – Bayesian inverse problems



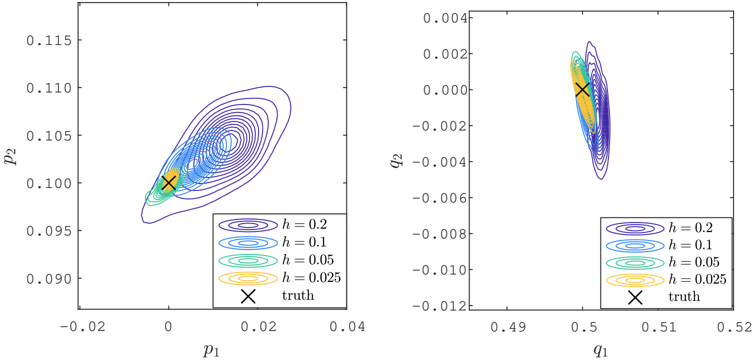
Posterior distributions given by deterministic Heun method.

Numerical experiments – Bayesian inverse problems



Posterior distributions given by deterministic Störmer-Verlet method.

Numerical experiments – Bayesian inverse problems



Posterior distributions given by RTS-RK Störmer-Verlet method.

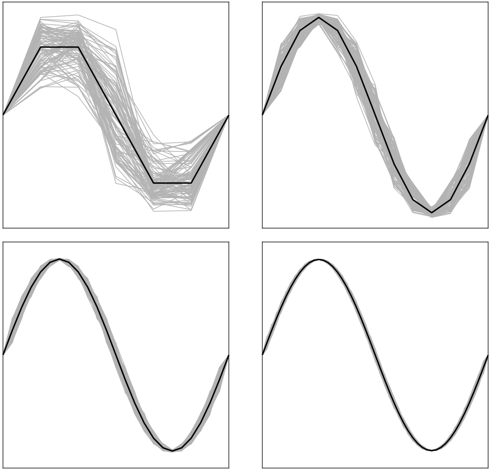
Outline

Research plan

Future research will cover the following topics

- Analysis of modelling errors in Bayesian inverse problems
- Probabilistic methods for PDEs, extension of the RTS-RK method?
- Adaptive time stepping probabilistic algorithms for ODEs
- Particle filter approach to sampling methods – a bridge between sampling and filtering probabilistic methods

Research plan – preliminary results



Probabilistic solutions of $-\Delta u = sin(2\pi x)$ with random meshes.

References I

[Conrad et al., 2016] Conrad, P. R., Girolami, M., Särkkä, S., Stuart, A., and Zygalakis, K. (2016).
Statistical analysis of differential equations: introducing probability measures on numerical solutions.
Stat. Comput.

[Kersting and Hennig, 2016] Kersting, H. and Hennig, P. (2016).
Active uncertainty calibration in Bayesian ODE solvers.
In *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI 2016)*, pages 309–318. AUAI Press.

[Lie et al., 2017] Lie, H. C., Sullivan, T. J., and Teckentrup, A. L. (2017).
Random forward models and log-likelihoods in bayesian inverse problems.