Probabilistic solvers for ODE's and Bayesian inference of parametrized models

Master Project - Master in CSE

Giacomo Garegnani
Supervisor: Prof. Assyr Abdulle
Experts: Prof. Fabio Nobile, Dr. Kostas Zygalakis

EPFL

02/02/2017

Outline of the presentation

1. Introduction on Bayesian inference and MCMC
2. Probabilistic solvers for ODE's
3. Bayesian inference inverse problems with differential equations

Consider Ω event space, $\mathcal{A}$ σ-algebra, P probability measure and $(\Omega, \mathcal{A}, P)$. Given A, B in Ω, Bayes' formula reads

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \propto P(B \mid A)P(A).$$

Normalization constant $P(B)$ can be replaced as

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{\int_\Omega P(B \mid A)P(A)},$$

as $P(A \mid B)$ is a probability distribution.

Problem. Consider two events A, B in Ω and the probability space $(\Omega, \mathcal{A}, P)$. We want to infer the probability distribution of A given B as

$$\underbrace{\pi(A \mid B)}_{\text{posterior}} \propto \overbrace{\mathcal{Q}(A)}^{\text{prior}} \underbrace{\mathcal{L}(B \mid A)}_{\text{likelihood}}$$

In models parametrized by a parameter θ, we deduce the distribution of θ through observations $\mathcal{Y}_n = \{y_1, y_2, \dots, y_n\}$ as

$$\pi(\theta \mid \mathcal{Y}_n) \propto \mathcal{Q}(\theta)\mathcal{L}(\mathcal{Y}_n \mid \theta).$$

Goal. Approximate the expectation under the distribution $\pi(\theta \mid \mathcal{Y})$ of a functional of the parameter $\theta \in \mathbb{R}^{N_p}$ with a Monte Carlo sum, i.e.,

$$\mathbb{E}^{\pi}\left[g(\theta)\right] = \int_{\mathbb{R}^{N_p}} g(\theta)\pi(\mathrm{d}\theta \mid \mathcal{Y}) \approx \frac{1}{N}\sum_{k=1}^{N} g(\theta^{(k)}),$$

where $\theta^{(k)}$ are realizations of θ.

Problem. How do we generate samples $\theta^{(k)}$, with $k = 1, \ldots, N$ so that the approximation above holds?
$\rightsquigarrow$ MCMC [Gilks, 2005, e.g.]

Idea. Generate samples $\theta^{(k)}$, with from a Markov chain with kernel P until the chain reaches its stationary distribution. Different choices of the Markov kernel lead to different MCMC algorithms.

Metropolis-Hastings (MH). Choose a *proposal distribution* $q(x, y)$ such that

$$\int_{\mathbb{R}^{N_p}} q(x, y)\mathrm{d}y = 1,$$

Then, given current sample $\theta^{(i)}$, the transition kernel P_{MH} giving $\theta^{(i+1)}$ is defined as

$$P_{\mathrm{MH}}(\theta^{(i)}, \theta^{(i+1)}) := \alpha(\theta^{(i)}, \theta^{(i+1)})q(\theta^{(i)}, \theta^{(i+1)}) + \delta_{\theta^{(i)}}(\theta^{(i+1)})\rho(\theta^{(i)}),$$

where $\alpha(\theta^{(i)}, \vartheta)$ is the *acceptance probability* and is given by

$$\alpha(\theta^{(i)}, \vartheta) = \min\left\{\frac{\pi(\vartheta)q(\vartheta, \theta^{(i)})}{\pi(\theta^{(i)})q(\theta^{(i)}, \vartheta)}, 1\right\},$$

and ρ is defined as

$$\rho(\theta^{(i)}) := 1 - \int_{\mathbb{R}^{N_p}} \alpha(\theta^{(i)}, x)q(\theta^{(i)}, x)\mathrm{d}x.$$

Algorithm 1 Metropolis-Hastings.

Given $\theta^{(0)} \in \mathbb{R}^{N_p}$, $N \in \mathbb{N}_0$.
for $i = 0, \ldots, N$ **do**
  Draw ϑ from $q(\theta^{(i)}, \cdot)$;
  Compute the acceptance probability $\alpha(\theta^{(i)}, \vartheta)$ as

$$\alpha(\theta^{(i)}, \vartheta) = \min\left\{ \frac{\pi(\vartheta) q(\vartheta, \theta^{(i)})}{\pi(\theta^{(i)}) q(\theta^{(i)}, \vartheta)}, 1 \right\};$$

  Draw u from $\mathcal{U}(0, 1)$;
  if $\alpha > u$ **then**
    Accept ϑ, set $\theta_{i+1} = \vartheta$;
  else
    set $\theta_{i+1} = \theta^{(i)}$
  end if
end for

Remark. If the proposal distribution is symmetric, i.e., $q(x,y) = q(y,x)$, then

$$\alpha(\theta^{(i)}, \vartheta) = \min\left\{\frac{\pi(\vartheta)}{\pi(\theta^{(i)})}, 1\right\}.$$

For example, Gaussian proposal [Kaipio and Somersalo, 2005]

$$q(x,y) \propto \exp\left(-\frac{1}{2}(x-y)^T\Sigma^{-1}(x-y)\right).$$

Problems. Two main problems

1. How to choose an efficient proposal distribution?
2. How to modify MH if it is not possible to evaluate the posterior distribution?

Problem. Bad proposal distribution $q(x, y) \implies$ inefficient algorithms. Measure efficiency with *acceptance ratio*.

Idea. Adapt $q(x, y)$ to obtain a chosen acceptance ratio α^*. Choose $q(x, y)$ Gaussian, the new guess ϑ is

$$\vartheta = \theta_k + S_n z_n, \quad Z_n \sim \mathcal{N}(0, I),$$

where $S_n \in \mathbb{R}^{N_p \times N_p}$ is lower triangular definite positive. Then,

$$S_{n+1} S_{n+1}^T = S_n \left(I + \eta_n \left(\alpha(\theta^{(n)}, \vartheta) - \alpha^* \right) \frac{z_n z_n^T}{z_n^T z_n} \right) S_n^T,$$

where $\eta_n \xrightarrow{n \to \infty} 0$.

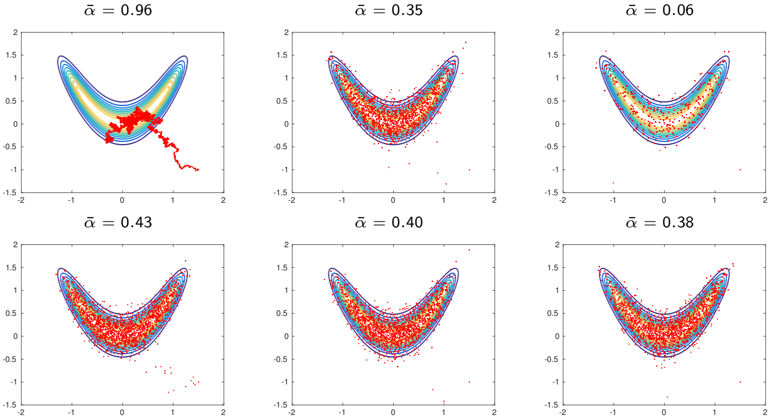
Two-dimensional distribution π with density [Kaipio and Somersalo, 2005]

$$\pi(X) \propto \exp\left(-10(X_1^2 - X_2)^2 - (X_1 - 0.25)^4\right),$$

Setting of the experiment. Given $\sigma = \{0.01, 0.5, 2.0\}$, compare

- standard MH with Gaussian proposal with covariance σI,
- RAM with $S_0 = \sigma I$ and $\alpha^* = 0.4$.

Draw $N = 5000$ samples and compute final acceptance ratio $\bar{\alpha}$.



Samples produced by MH and RAM for the distribution with standard MH (first row) and RAM (second row).

[Andrieu et al., 2010, Doucet et al., 2015, Medina-Aguayo et al., 2016]

Problem. Impossible to evaluate $\pi(\theta)$ (no closed form available).

Idea. Find evaluable $\pi(\theta, \xi)$ that admits $\pi(\theta)$ as marginal distribution, then compute

$$\hat{\pi}_M(\theta) = \frac{1}{M} \sum_{i=1}^{M} \pi(\theta, \xi^{(i)}),$$

with $\xi^{(i)}$, $i = 1, \ldots, M$ realizations of ξ. Use $\hat{\pi}_M$ for $\alpha(\theta^{(i)}, \vartheta)$.

Remark. The rest of MH is unchanged.

Outline of the presentation

1. Introduction on Bayesian inference and MCMC
2. Probabilistic solvers for ODE's
3. Bayesian inference inverse problems with differential equations

Problem. Given $f\colon \mathbb{R}^d \to \mathbb{R}^d$ and the autonomous ODE

$$u'(t) = f(u), \quad u(0) = u_0,$$

build a probabilistic numerical solution. There exists flow map $\Phi_t(y)$ such that

$$u(t) = \Phi_t(u_0).$$

Idea. Given $h > 0$, the flow map of a Runge-Kutta method $\Psi_h(y)$ is

$$u_{k+1} = \Psi_h(u_k), \quad k = 0, 1, \dots,$$

consider $\xi_k(h)$ i.i.d. random variables in $\mathbb{R}^d$ and compute

$$U_{k+1} = \underbrace{\Psi_h(U_k)}_{\text{deterministic}} + \overbrace{\xi_k(h)}^{\text{random}}, \quad k = 0, 1, \dots,$$

Deterministic component $\rightsquigarrow$ Runge-Kutta methods.

Definition (Runge-Kutta methods)

Given $s \in \mathbb{N}^*$, $(b_i)_{i=1}^s$, $(a_{ij})_{i,j=1}^s$, $h > 0$, one step of an s-stage Runge-Kutta method is

$$\begin{aligned} K_i &= f(U_0 + h\textstyle\sum_{j=1}^s a_{ij}K_j), \quad i = 1, \dots, s, \\ U_1 &= U_0 + h\textstyle\sum_{i=1}^s b_i K_i. \end{aligned}$$

The method is of order q if $\exists C > 0$ independent of h such that

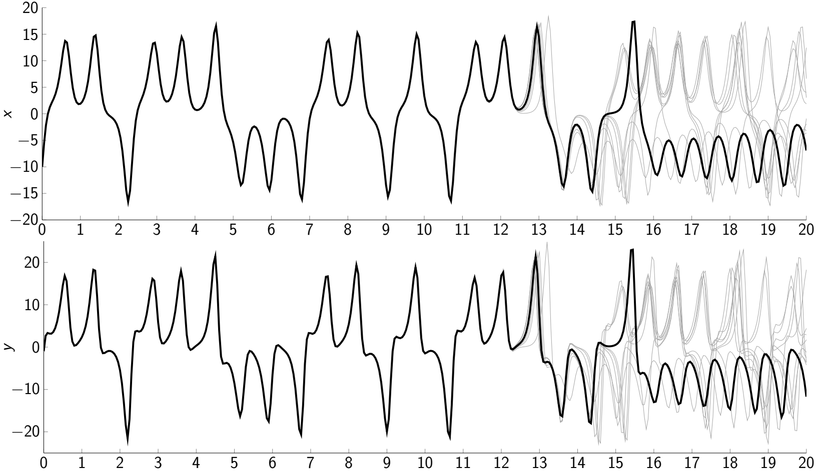
$$\|u(h) - U_1\| \leq Ch^{q+1}.$$

Monte Carlo computations $\rightsquigarrow$ Explicit methods privileged.

Consider chaotic differential equation, e.g., Lorenz system

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

Small perturbation $\implies$ uncontrollable deviation of the solution.
Deterministic solvers not reliable for any time step $h > 0$.
$\rightsquigarrow$ Family of M probabilistic numerical solutions.



Components x and y of Lorenz system with deterministic solver (thick black) and probabilistic solver (light gray). Gauss method on 2 stages as $\Psi_h(y)$.

Relevant properties analyzed

- strong order of convergence,
- weak order of convergence,
- behavior of Monte Carlo approximations.

Recall. The probabilistic method is defined as

$$U_{k+1} = \Psi_h(U_k) + \xi_k(h), \quad k = 0, 1, \dots,$$

for suitable random variables $\xi_k(h)$.

It is possible to prove a result of strong convergence.

Definition (Strong convergence)

The probabilistic method has strong order r if $\exists C > 0$ independent of h such that for h small enough

$$\sup_{t_k = kh} \mathbb{E}|U_k - u(t_k)| \leq Ch^r.$$

Assumption (Variance of random variables)

The variables $\xi_k(t)$ satisfy for $p \geq 1$

$$\mathbb{E}|\xi_k(t)\xi_k(t)^T|_F^2 \leq Kt^{2p+1}.$$

Furthermore, there exists a matrix Q independent of h such that

$$\mathbb{E}[\xi_k(h)\xi_h(h)^T] = Qh^{2p+1},$$

Assumption (Order of the deterministic component)

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 $q+1$, i.e.,

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

Theorem (Strong Convergence)

Under the assumptions above, there exists $K > 0$ such that

$$\sup_{0<kh<T} \mathbb{E}|u_k - U_K|^2 \leq Kh^{2\min\{p,q\}}.$$

Idea of the proof. Compute truncation error between exact and numerical solutions, divide deterministic and probabilistic contribution and derive a recurrence on the error. Apply then discrete Gronwall's lemma to bound the error.

Definition (Weak convergence)

The probabilistic method has weak order r if $\exists C > 0$ independent of h such that for any function φ sufficiently smooth

$$\sup_{t_k = kh} |\mathbb{E}[\varphi(U_k)] - \varphi(u(t_k))| \leq Ch^r,$$

for h small enough.

Idea. Introduce a modified SDE

$$\mathrm{d}\tilde{u} = f^h \tilde{u} \, \mathrm{d}t + \sqrt{h^{2p} Q} \, \mathrm{d}W ,$$

and study the convergence of U_k and $\tilde{u}$ to u for $h \to 0$.

Theorem (Weak convergence)

For any function φ sufficiently smooth

$$|\varphi(u(T)) - \mathbb{E}[\varphi(U_k)]| \leq K h^{\min\{2p,q\}}, \quad kh = T,$$

and

$$|\mathbb{E}[\varphi(\tilde{u}(T))] - \mathbb{E}[\varphi(U_k)]| \leq K h^{2p+1}, \quad kh = T.$$

Idea of the proof. Use techniques of backwards error analysis, finding a modified ODE and SDE such that the numerical error is of higher order.

Problem. Study convergence of Monte Carlo approximations.

Numerical solution $\rightsquigarrow \mathcal{Q}_h$ (inaccessible),
M samples of numerical solution $\rightsquigarrow Q_h^M$ (accessible),
Exact solution $\rightsquigarrow \delta_u$.

Convergence scheme, we expect

$$\mathcal{Q}_h^M \xrightarrow{M\to\infty} \mathcal{Q}_h \xrightarrow{h\to 0} \delta_u.$$

Second convergence already treated, first unclear [Kersting and Hennig, 2016].

Consider φ a regular function, M trajectories of the numerical solution and the estimator

$$\hat{Z} = \frac{1}{M} \sum_{i=1}^{M} \varphi\left(U_N^{(i)}\right).$$

Goal. Estimate the convergence of the MSE of $\hat{Z}$.

Remark. Thanks to weak convergence result,

$$\mathrm{MSE}(\hat{Z}) \leq \mathrm{Var}(\hat{Z}) + Ch^{2\min\{2p,q\}}.$$

$\rightsquigarrow$ bound the variance of $\hat{Z}$ with a function of h and M.

Consider only one-dimensional problems and bound the variance of the numerical solution. Recall that

$$\mathbb{E}[\xi_k(h)^2] = Qh^{2p+1}.$$

Lemma (Variance of the numerical solution)

Consider a one-dimensional ODE and the probabilistic method with Ψ any Runge-Kutta scheme on s stages. Then, if h is small enough, $\exists C_1, C_2 > 0$ such that

$$\mathrm{Var}(U_k) \leq C_1 \,\mathrm{Var}(U_0) + C_2 Q h^{2p}, \quad k = 1, \dots, N.$$

Remark. If the initial condition is deterministic, the variance is bounded by

$$\mathrm{Var}(U_k) \leq C_2 Q h^{2p}.$$

Assume φ is Lipschitz with constant C_L, then

$$\mathrm{Var}(\varphi(U_k)) \leq C_L^2 \,\mathrm{Var}(U_k).$$

Therefore, the following result is trivially proved.

Theorem (Bound of the MSE)

The following bound for the MSE of $\hat{Z}$ is valid

$$\mathrm{MSE}(\hat{Z}) \leq C_1 h^{2\min\{2p,q\}} + \frac{C_2}{M}(\mathrm{Var}(U_0) + h^{2p}).$$

Remark. If the initial condition is deterministic

$$\mathrm{MSE}(\hat{Z}) = \mathcal{O}(h^{2\min\{2p,q\}}) + \mathcal{O}(M^{-1}h^{2p}),$$

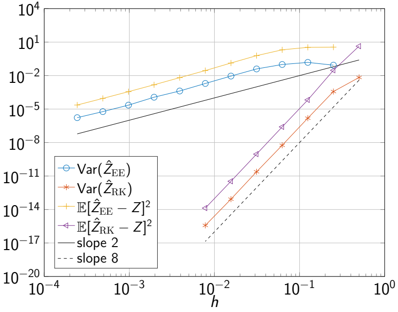
hence depending on p and q one can choose $M = \mathcal{O}(1)$.

Goal. Verify the properties of Monte Carlo estimators derived above. Test problem is the FitzHug-Nagumo equation

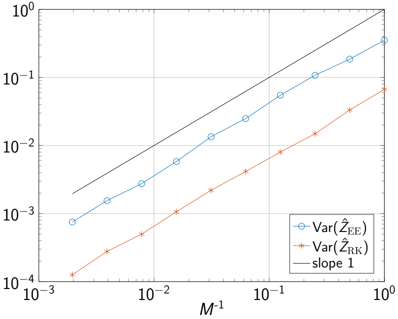
$$
\begin{aligned}
x' &= c\left(x - \frac{x^3}{3} + y\right), \quad x(0) = -1,\\
y' &= -\frac{1}{c}(x - a + by), \quad y(0) = 1,
\end{aligned}
$$

Use order one and order four methods, Euler Explicit and fourth-order Runge-Kutta.

$$
\text{(EE)} \ \begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array} \qquad \text{(RK4)} \ \begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array}
$$



(a) Variation of h.



(b) Variation of M.

Numerical experiment. Deterministic methods Explicit Euler and 4th-order Runge-Kutta, deterministic U_0 and $p = q$. Compute variance and bias of $\hat{Z}$ varying h and M.

Recall. Bound on the MSE

$$\mathrm{MSE}(\hat{Z}) \leq C_1 h^{2\min\{2p,q\}} + C_2 M^{-1}(\mathrm{Var}(U_0) + h^{2p}).$$

Problem. If $\mathrm{Var}(U_0) > 0$, convergence with h is not valid anymore. In this case

$$\mathrm{MSE}(\hat{Z}) \leq C_1 h^{2\min\{p,q\}} + C_2 \mathrm{Var}(U_0) M^{-1}.$$

If the two contributions are balanced and the error is measured as $\mathrm{MSE}(\hat{Z})^{1/2}$ we obtain

$$\mathrm{cost} = M\frac{T}{h} = \mathcal{O}\left(\varepsilon^{-2-1/\min\{p,q\}}\right),$$

where ε is the desired accuracy.

Idea. Apply multi-level techniques to reduce computational cost. Applying standard MLMC [Giles, 2008] we get

$$\mathrm{cost} = \begin{cases} \mathcal{O}\left(\left|\log_2 \varepsilon^{1/q}\right|\varepsilon^{-2}\right), & \text{if } q \leq p, \\ \mathcal{O}\left(\left|\log_2 \varepsilon^{1/2p}\right|\varepsilon^{-2}\right), & \text{if } q \geq 2p, \\ \mathcal{O}\left(\left|\log_2 \varepsilon^{1/q}\right|\varepsilon^{-2}\right), & \text{if } p < q \leq 2p. \end{cases}$$

Remark. Computational cost is remarkably lower than in standard Monte Carlo. **Not necessary if deterministic initial condition**.

We considered the probabilistic method for ODE's

$$U_{k+1} = \Psi_h(U_k) + \xi_k(h), \quad k = 0, 1, \dots,$$

and studied

1. weak and strong convergence [Conrad et al., 2016],
2. convergence of Monte Carlo estimators,
3. multi-level techniques for random initial conditions.

In particular, we found that the MSE of the Monte Carlo estimator converges **independently of the number of samples** for deterministic U_0.

Outline of the presentation

1. Introduction on Bayesian inference and MCMC
2. Probabilistic solvers for ODE's
3. Bayesian inference inverse problems with differential equations

Consider θ unknown parameter in $\mathbb{R}^{N_p}$ (true value $\bar{\theta}$) and

$$u_\theta'(t) = f_\theta(u_\theta(t)), \quad u_\theta(0) = u_0.$$

Consider observations

$$\mathcal{Y}_N = \{y_1, y_2, \dots, y_N\}, \quad y_i = u_{\bar{\theta}}(t_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \Gamma).$$

<u>Goal.</u> Apply Bayesian inference techniques to get distribution of θ with

$$\pi(\theta \mid \mathcal{Y}_N) \propto \overbrace{\mathcal{Q}(\theta)}^{\text{prior (known)}} \quad \underbrace{\mathcal{L}(\mathcal{Y}_N \mid \theta)}_{\text{likelihood (unknown)}},$$

where the likelihood is not known since the ODE has no closed form solution $u_\theta(t)$.

Problem. Approximate $\mathcal{L}(\mathcal{Y}_N \mid u_\theta(t))$. Probabilistic integrator with time step h gives

$$\mathcal{L}(\mathcal{Y} \mid \theta) \approx \mathcal{L}_h(\mathcal{Y} \mid \theta),$$

but $\mathcal{L}_h$ is not accessible, repeated sampling with M trajectories

$$\mathcal{L}(\mathcal{Y} \mid \theta) \approx \mathcal{L}_h^M(\mathcal{Y} \mid \theta).$$

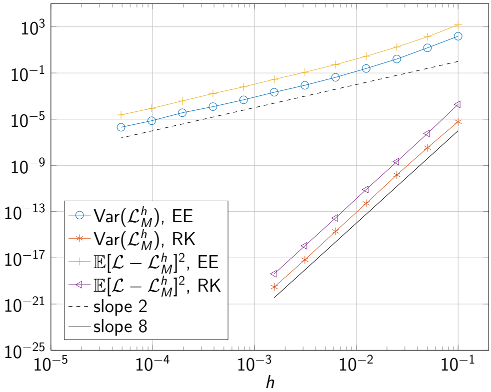
Hence, for each value of θ we have

$$\pi_h^M(\theta \mid \mathcal{Y}_N) \xrightarrow{M\to\infty} \pi_h(\theta \mid \mathcal{Y}_N) \xrightarrow{h\to 0} \pi(\theta \mid \mathcal{Y}_N).$$

Result on MSE of probabilistic method implies

$$\mathrm{MSE}(\mathcal{L}_h^M(\mathcal{Y} \mid \theta)) \leq C_1 h^{2\min\{2p,q\}} + \frac{C_2}{M} h^{2p}.$$

$\rightsquigarrow$ **Run pseudo-marginal MCMC.**



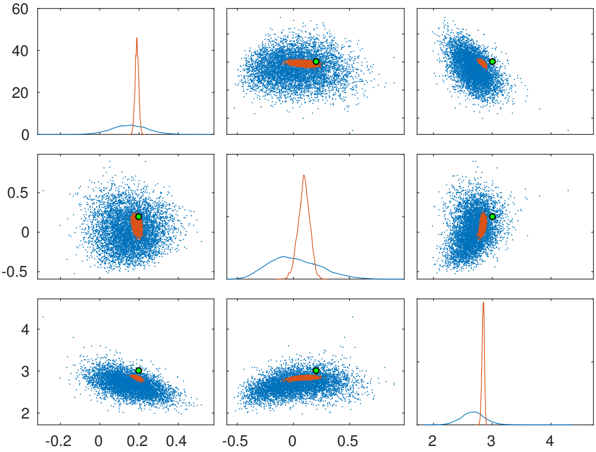
Approximation of the likelihood for FitzHug-Nagumo problem. Reference solution obtained with RK4 and fine time-step, ten observations $\mathcal{Y}_{10}$ at equispaced times between $t_1 = 1$ and $t_{10} = 10$. Probabilistic method with $p = q$.

Why the probabilistic integrator and not a deterministic Runge-Kutta method? Consider the FitzHug-Nagumo problem

$$x' = c\left(x - \frac{x^3}{3} + y\right), \quad x(0) = -1,$$
$$y' = -\frac{1}{c}(x - a + by), \quad y(0) = 1,$$

with parameter $\theta = (a, b, c)^T$.
Empirically [Conrad et al., 2016] the posterior accounts better for the numerical error with the probabilistic method.

Experiment. Run 50000 iterations of RAM with deterministic method and RAM applied to pseudo-marginal MCMC with probabilistic method.



Posterior distribution for θ. Probabilistic method (blue dots), deterministic method (red dots), true value (thick green).

Problem. Determine how the posterior π_h^M converges to the true posterior π with respect to h and M.

Assumptions. Consider the following assumptions

1. probabilistic integrator with $p = q$,
2. π and π_h^M admit densities $\pi(x)$ and $\pi_h^M(x)$,
3. denote as $\mathbb{E}^\xi[\cdot]$ and $\mathbb{E}^\pi[\cdot]$ the expectation with respect to the r.v. ξ and the distribution π respectively.

Notion of distance. We consider the Hellinger distance, i.e., given distributions μ and ν with densities f and g

$$d^2_{\mathrm{Hell}}(\mu, \nu) := \frac{1}{2} \int_{\mathbb{R}^n} \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 \mathrm{d}x$$

Recall. With $p = q$ we have, thanks to the result on Monte Carlo estimators we have, for any fixed $\theta \in \mathbb{R}^{N_p}$

$$\begin{aligned}\mathrm{MSE}(\mathcal{L}_h^M(\mathcal{Y} \mid \theta)) &= \mathbb{E}^\xi[(\mathcal{L}(\mathcal{Y} \mid \theta) - \mathcal{L}_h^M(\mathcal{Y} \mid \theta)^2] \\ &\leq C(\theta) h^{2q}.\end{aligned}$$

We can then derive the convergence result

$$\mathbb{E}^\xi[d_{\mathrm{Hell}}(\pi(\theta \mid \mathcal{Y}), \pi_h^M(\theta \mid \mathcal{Y}))] \leq \left(\sqrt{\frac{1}{2} \sup_{\theta \in \mathbb{R}^{N_p}} C(\theta)}\right) h^q.$$

Hence, convergence is independent of M.

Problem. Consider $g\colon \mathbb{R}^{N_p} \to \mathbb{R}$ and the Monte Carlo approximation

$$\mathbb{E}^{\pi}\left[g(\theta)\right] \approx \frac{1}{N}\sum_{k=1}^{N} g(\theta^{(k)}).$$

Running MCMC with approximated likelihood $\mathcal{L}_h^M$, the Monte Carlo estimation approximates

$$\mathbb{E}^{\pi_h^M}\left[g(\theta)\right] \approx \frac{1}{N}\sum_{k=1}^{N} g(\theta^{(k)}).$$

Goal. Estimate in terms of h and M the quantity

$$\mathrm{MSE}[\mathbb{E}^{\pi_h^M}\left[g(\theta)\right]] = \mathbb{E}^{\xi}[(\mathbb{E}^{\pi_h^M}\left[g(\theta)\right] - \mathbb{E}^{\pi}\left[g(\theta)\right])^2].$$

It is possible to show that

$$\mathbb{E}^{\xi}\left|\mathbb{E}^{\pi}\left[g(\theta)\right]-\mathbb{E}^{\pi_h^M}\left[g(\theta)\right]\right|\leq\|g\|_{\infty}Ch^{q},$$
$$\mathrm{Var}^{\xi}(\mathbb{E}^{\pi_h^M}\left[g(\theta)\right])\leq\left\|g^2\right\|_{\infty}Ch^{2q}.$$

Hence, we have

$$\begin{aligned}\mathrm{MSE}[\mathbb{E}^{\pi_h^M}\left[g(\theta)\right]]&=(\mathbb{E}^{\xi}[\mathbb{E}^{\pi}\left[g(\theta)\right]-\mathbb{E}^{\pi_h^M}\left[g(\theta)\right]])^2+\mathrm{Var}^{\xi}(\mathbb{E}^{\pi_h^M}\left[g(\theta)\right])\\&\leq C\|g\|_{\infty}^{2}h^{2q}.\end{aligned}$$

We consider the parabolic PDE (Brusselator problem)

$$\begin{aligned}
\frac{\partial u}{\partial t} &= 1 + u^2 v + \alpha \frac{\partial^2 u}{\partial x^2}, \quad u = u(x,t) \\
\frac{\partial v}{\partial t} &= 3u - u^2 v + \alpha \frac{\partial^2 v}{\partial x^2}, \quad v = v(x,t), \quad x \in \Omega = (0,1), \quad t \geq 0 \\
u(0,t) &= u(1,t) = 1, \\
v(0,t) &= v(1,t) = 3, \\
u(x,0) &= 1 + \sin(2\pi x), \\
v(x,0) &= 3.
\end{aligned}$$

Discretization with the method of lines on $N+2$ points leads to a stiff system of ODE's with stiffness index

$$\lambda = 4\alpha(N+1)^2.$$

Goal. Estimate from observations the parameter α. The stiffness index

$$\lambda = 4\alpha(N+1)^2.$$

varies with respect to α. How to obtain stable approximations of $\mathcal{L}(\mathcal{Y} \mid \alpha)$?

Idea. Consider α to be bounded in $I_\alpha = [0, \alpha_{\mathrm{max}}]$ and use MCMC with truncated Gaussian proposal $q(x, y)$.

Problem. Using explicit Euler as deterministic component we obtain the time step restriction

$$h < \frac{1}{8\alpha(N+1)^2} \leq \frac{1}{8\alpha_{\mathrm{max}}(N+1)^2} =: \bar{h}$$

$\rightsquigarrow$ high computational cost

Idea. Use stabilized explicit method Runge-Kutta-Chebyshev of first order on s stages as deterministic component. It is known that if

$$s \geq \max \left\{ 2, \left\lceil \sqrt{\frac{1}{2} h\lambda} \right\rceil \right\},$$

the method is stable. Therefore, the maximum number of stages is

$$\bar{s} = \max \left\{ 2, \left\lceil \sqrt{2h\alpha_{\max}(N+1)^2} \right\rceil \right\}.$$

Furthermore, **we can adapt s during MCMC** to be the minimum number of step required for stability at each guess $\alpha^{(i)}$.

h	0.2	0.1	0.05	0.025	0.0125
$\hat{\alpha}$ $(\cdot 10^{-2})$	2.509	2.0448	1.996	2.055	1.990
mean s	11.89	7.87	6.00	4.92	4.00
mean cost $(\cdot 10^{3})$	0.59	0.79	1.20	1.97	3.20
$\bar{s}$	64	46	32	23	16
max cost $(\cdot 10^{3})$	3.2	4.6	6.4	9.2	12.8

Results obtained with observations given by true value $\bar{\alpha} = 0.02$ and $I_\alpha = [0, 1]$. Computational cost = n. of function evaluations.

Remark. With explicit Euler the minimum value of time step is $\bar{h} = 4.98 \cdot 10^{-7}$ with computational cost per iteration $2 \cdot 10^{7}$.
$\rightsquigarrow$ remarkable computational gain.

We considered Bayesian inference inverse problems involving ODE's and studied

1. convergence of the likelihood estimator $\mathcal{L}_h^M(\mathcal{Y} \mid \theta)$,
2. qualitative behavior of deterministic vs. probabilistic approximation,
3. convergence of the posterior distribution in the Hellinger distance,
4. convergence of the expected value under approximated posterior,
5. application of stabilized explicit methods in stiff problems.

Conclusion and future work

In this project, we

1. presented a survey on existing MCMC techniques to perform Bayesian inference,
2. analyzed a probabilistic solver for ODE's studying weak and strong convergence [Conrad et al., 2016], as well as the behavior of Monte Carlo estimates (novel contribution),
3. performed Bayesian inference of parametric ODE's and study the convergence of posterior distributions and Monte Carlo estimates.

In future work, we will be concerned with a classification of the problems for which a probabilistic interpretation is needed. In particular, we believe chaotic ODE's (e.g., Lorenz) have to be regarded with utmost attention.

[Andrieu et al., 2010] Andrieu, C., Doucet, A., and Holenstein, R. (2010).
Particle Markov chain Monte Carlo methods.
J. R. Stat. Soc. Ser. B. Stat. Methodol., pages 269 – 342.

[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.

[Doucet et al., 2015] Doucet, A., Pitt, M. K., Deligiannidis, G., and Kohn, R. (2015).
Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator.
Biometrika, pages 1 – 19.

References II

[Giles, 2008] Giles, M. B. (2008).
Multilevel Monte Carlo path simulation.
Operations Research, 56(3):607–617.

[Gilks, 2005] Gilks, W. R. (2005).
Markov chain Monte Carlo.
Encyclopedia of Biostatistics, 4.

[Hairer et al., 2002] Hairer, E., Lubich, C., and Wanner, G. (2002).
Geometric numerical integration: structure-preserving algorithms for ordinary differential equations.
Springer-Verlag, Berlin and New York.

References III

[Kaipio and Somersalo, 2005] Kaipio, J. and Somersalo, E. (2005).
Statistical and Computational Inverse Problems.
Applied Mathematical Sciences, 160. Springer.

[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.

[Medina-Aguayo et al., 2016] Medina-Aguayo, F. J., Lee, A., and Roberts, G. O. (2016).
Stability of noisy Metropolis–Hastings.
Stat. Comput., 26(6):1187–1211.

References IV

[Vihola, 2012] Vihola, M. (2012).
Robust adaptive Metropolis algorithm with coerced acceptance rate.
Stat. Comput., 22(5):997–1008.