Ergodic Stochastic Differential Equations and Sampling: A numerical analysis prespective

K. C. Zygalakis

School of Mathematics, University of Edinburgh

The Alan Turing Institute

Workshop on Multiscale methods for stochastic dynamics
University of Geneva
31 January 2017

Collaborators

Assyr Abdulle (EPFL)

Gilles Vilmart (Geneva)

Grigorios Pavliotis (Imperial)

Andrew Duncan (Sussex)

Mike Giles (Oxford)

Lukasz Szpruch (Edinburgh)

Sebastian Vollmer (Warwick)

Overview

Overview

Ergodic SDEs I

Consider the stochastic differential equation

$$dX_t = \nabla \log \pi(X_t)dt + \sqrt{2}dW_t.$$

Under appropriate assumptions on $\nabla \log \pi(x)$ one can show that its dynamics are ergodic with respect to $\pi(x) : \mathbb{R}^d \mapsto \mathbb{R}$ i.e

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \phi(X_s)ds = \mathbb{E}_\pi[f] := \int_{\mathbb{R}^d} f(x)\pi(x)dx$$

$\pi(x)$ also satisfies the equation

$$\mathcal{L}^*\pi(x) = 0$$

$\mathcal{L}^*$ is the adjoint of

$$\mathcal{L} := \nabla \log \pi(x) \cdot \nabla_x + \Delta_x$$

Ergodic SDEs II

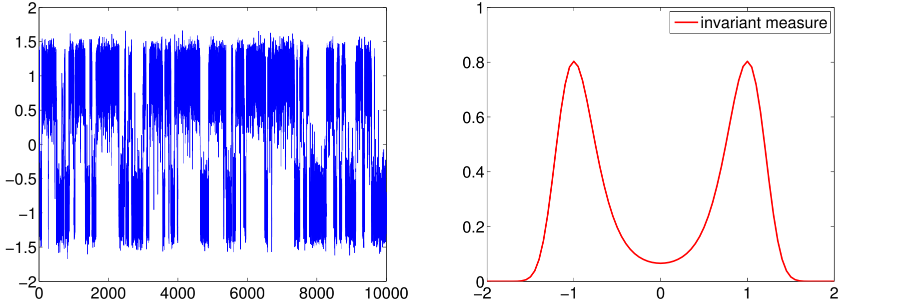


Figure: Long trajectory and invariant measure

Ergodic SDEs III

One can use

$$\pi_T(f) := \frac{1}{T}\int_0^T f(X_s)\,ds$$

as an estimator for $\mathbb{E}_\pi[f]$, for $T \gg 1$. A natural way to measure the efficiency of such estimator is the mean square error (MSE) given by

$$\begin{aligned}\mathrm{MSE}(T) :&= \mathbb{E}|\pi_T(f) - \pi(f)|^2 = (\mathbb{E}\pi_T(f) - \pi(f))^2 + \mathbb{E}(\pi_T(f) - \mathbb{E}\pi_T(f))^2 \\ &= \mu_T^2 + \sigma_T^2 \simeq \mu_T^2 + \frac{\sigma_f^2}{T} \quad \text{if the central limit theorem holds}\end{aligned}$$

Here

$$\sigma_f^2 = 2\left\langle f - \mathbb{E}_\pi[f], (-\mathcal{L})^{-1}(f - \mathbb{E}_\pi[f])\right\rangle_\pi$$

We note here that for geometrically ergodic SDEs $\mu_T \to 0$ exponentially fast.

THE UNIVERSITY of EDINBURGH THE ALAN TURING INSTITUTE

Numerical ergodic averages

In practice we cannot solve our SDE exactly and we have to approximate in some way. We denote its approximation by $X_n^{\Delta t}$

$$\hat{\pi}_T(f) = \frac{1}{N}\sum_{n=0}^{N} f(X_n^{\Delta t}), \quad n\Delta t = T.$$

If the approximation is ergodic one has

$$\lim_{T\to\infty} \hat{\pi}_T(f) = \mathbb{E}_{\hat{\pi}}[f] := \int_{\mathbb{R}^d} f(x)\hat{\pi}(x)dx$$

THE UNIVERSITY of EDINBURGH
THE ALAN TURING INSTITUTE

Two different approaches

1. Solve the underlying SDE with a numerical method for large times

$$x_{n+1}^{\Delta t} = x_n^{\Delta t} - \Delta t \nabla \log \pi(x_n^{\Delta t}) + \sqrt{2\Delta t}\xi_n$$

2. Use a Metropolis-Hastings type of algorithm. For example MALA

   Use (1) as proposal within MCMC framework ($y_n^{\Delta t} = x_n^{\Delta t}$).

$$y_{n+1}^{\Delta t} = \begin{cases} x_{n+1}^{\Delta t} & \text{with probability} \quad \alpha(y_n^{\Delta t}, x_{n+1}^{\Delta t}) \\ y_n^{\Delta t} & \text{with probability} \quad 1 - \alpha(y_n^{\Delta t}, x_{n+1}^{\Delta t}) \end{cases}$$

   where

$$\alpha(y_n^{\Delta t}, x_{n+1}^{\Delta t}) = \min\left(1, \frac{\pi(x_{n+1}^{\Delta t}) q_{\Delta t}(x_{n+1}^{\Delta t}, y_n^{\Delta t})}{\pi(y_n^{\Delta t}) q_{\Delta t}(y_n^{\Delta t}, x_{n+1}^{\Delta t})}\right)$$

Things to consider when making a choice

$$\widehat{\text{MSE}}(T) := \mathbb{E}|\hat{\pi}_T^{\Delta t}(f) - \pi(f)|^2 = (\mathbb{E}\hat{\pi}_T^{\Delta t}(f) - \pi(f))^2 + \mathbb{E}(\hat{\pi}_T^{\Delta t}(f) - \mathbb{E}\hat{\pi}_T^{\Delta t}(f))^2$$
$$= \hat{\mu}^2_{T,\Delta t} + \hat{\sigma}^2_{T,\Delta t} \simeq \hat{\mu}^2_{T,\Delta t} + \frac{\hat{\sigma}^2_{f,\Delta t}}{T} \quad \text{if the central limit theorem holds}$$

---

1. The first approach (numerical analysis) introduces (asymptotic) bias in the calculation of $\pi(x)$, since $\hat{\pi}^{\Delta t} \neq \pi$
2. The second approach (computational statistics) removes the (asymptotic) bias from the calculation of $\pi(x)$, since $\hat{\pi}^{\Delta t} = \pi$

However

- Computational Statistics approach might be expensive in the presence of big data.
- Numerical Analysis approach permits for the construction of non-reversible algorithms with much smaller asymptotic variance than the computational statistics one.

Computational Complexity

For $\widehat{\text{MSE}}$ to be of $\mathcal{O}(\epsilon^2)$ we have the following computational complexity

- numerical analysis approach: $\mathcal{O}(\epsilon^{-3})$.
  - Asymptotic bias is of $\mathcal{O}(\Delta t)$ for a first order method.
  - Finite time error decays like $\frac{\hat{\sigma}^2_{f,\Delta t}}{T}$.
- computational statistics approach: $\mathcal{O}(\epsilon^{-2})$.
  - No asymptotic bias.
  - Finite time error decays like $\frac{\hat{\sigma}^2_{f,\Delta t}}{T}$.

Three desirable features

A desirable approach should have three different features

1. Small asymptotic bias (this can already be zero with MCMC methods).
2. Small asymptotic variance.
3. Optimal computational Complexity of $\mathcal{O}(\epsilon^{-2})$ (already optimal for MCMC methods).

We will try and address these by

1. Characterise the asymptotic bias of numerical integrators (without Metropolizing them).
2. Introduce non reversible samples with small asymptotic variance.
3. Make the Computational Complexity of numerical approach optimal with the use of Multi-Level Monte Carlo.

Overview

THE UNIVERSITY of EDINBURGH

THE ALAN TURING INSTITUTE

Backward Error Analysis

Given a numerical method for

$$dX = h(X)dt + g(X)dW_t,$$

can I find an equation

$$d\widetilde{X} = \hat{h}_{\Delta t}(\widetilde{X})dt + \hat{g}_{\Delta t}(\widetilde{X})dW_t,$$

that my integrator approximates better? The type of approximation one is interested is the weak approximation

$$\mathbb{E}(f(X_T)) - \mathbb{E}(f(X_n^{\Delta t}))$$

This approach will allows us to obtain the following expansion

$$\hat{\pi}^{\Delta t}(x) = \pi(x) + \Delta t \pi_1(x) + \Delta t^2 \pi_2(x) + \cdots$$

Weak Taylor Expansions

Consider the (deterministic) PDE

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad u(x,0) = f(x),$$

Then,

$$u(f,x,t) = \mathbb{E}\left(f(X_t)|X_0 = x\right),$$

which can be (formally) expanded in Taylor series $u = \sum_{j\geq 0} \frac{t^j \mathcal{L}^j}{j!}\phi(x)$.

Similarly consider $U(f,x,t) = \mathbb{E}\left(f(X_t^{\Delta t})|X_0^{\Delta t} = x\right)$ Here no such PDE exists but one can try. In particular, we will assume

Assumption

The numerical solution has the following expansion

$$U(f,x,\Delta t) = f(x) + \Delta t \mathcal{A}_0 f(x) + \Delta t^2 \mathcal{A}_1 f(x) + \dots,$$

Modified generator

$$U(\phi, x, h) - \phi(x) = \sum_{j \geq 1} \frac{h^j}{j!} \tilde{\mathcal{L}}^j \phi$$

with

$$\tilde{\mathcal{L}} = \mathcal{L} + \sum_{j \geq 1} L_j$$

L_j can be computed recursively

$$L_n = A_n - \frac{1}{2} (\mathcal{L} L_{n-1} + L_{n-1} \mathcal{L}) - \cdots - \frac{1}{(n+1)!} \mathcal{L}^{n+1}$$

Numerical expansion

Lemma [DF12]

Under appropriate assumptions (restrict ourself on $\mathbb{T}^d$) the numerical invariant measure $\hat{\pi}^{\Delta t}$ admits the following expansion

$$\hat{\pi}^{\Delta t}(x) = \pi(x) + \sum_{n=1}^{M} \Delta t^n \pi_n(x), \quad \int_{\mathbb{T}^d} \pi_n(x) dx = 0, \quad n \geq 1$$

where

$$\mathcal{L}^* \pi_n = -\sum_{l=1}^{n} (L_l)^* \pi_{n-l},$$

Order conditions

Theorem [AVZ13]

Consider

$$dX_t = f(X_t)dt + g(X_t)dW_t$$

on $\mathbb{T}^d$ solved by an ergodic numerical method. If

$$A_j^* \rho_\infty = 0, \quad j = 1, \cdots r - 1$$

then

$$\mathbb{E}_{\hat{\pi}^{\Delta t}}[f] - \mathbb{E}_\pi[f] = \mathcal{O}(\Delta t^r)$$

Thus the order of ergodic convergence for the numerical integrator is r.

Ways to satisfy the order conditions

An obvious way to satisfy the order conditions is to use a method of weak order r since in this case

$$A_j = \frac{\mathcal{L}^{j+1}}{(j+1)!}, \quad \text{for all} \quad j < r$$

and thus

$$A_j^* \pi = 0, \quad j = 1, \cdots, r-1$$

However if the structure of π is known then one can take advantage of it and satisfy the order conditions without necessarily using a method of high weak order.

Reducing asymptotic bias with additive noise

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + \Delta t \tilde{g}_{\Delta t}(X_n^{\Delta t}) + \sqrt{2\Delta t}\xi_n$$

with

$$\tilde{g}_{\Delta t} = g + \Delta t g_1, \quad g_1 = -\left[\frac{1}{2}g'g + \frac{1}{2}\Delta g\right], \quad g(x) = \nabla \log \pi(x)$$

has weak order of convergence $\beta = 1$ for finite time but the error for the asymptotic bias is of order $\alpha = 2$ (a calculation shows that for this method $A_1^*\pi = 0$). In principle able to construct further perturbations (idea similar to modified integrators [ACV12])

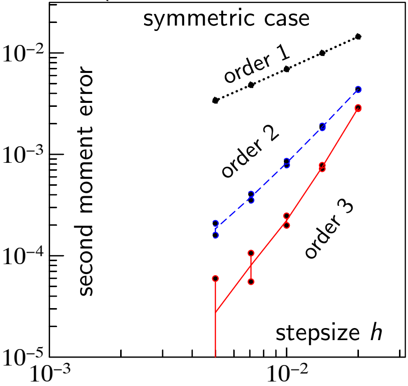
$$\tilde{g}_{\Delta t} = f + \sum_{i=1}^{k} \Delta t^i g_i, \quad \text{such that} \quad \alpha = k+1, \beta = 1$$

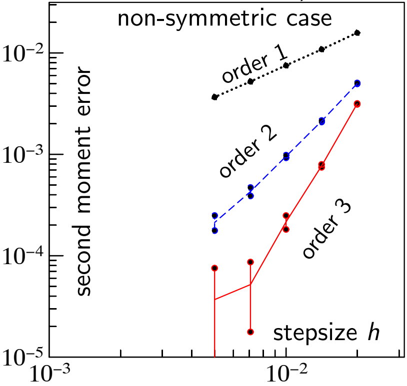
THE UNIVERSITY of EDINBURGH

THE ALAN TURING INSTITUTE

Numerical Investigation

(double-well potential, long trajectories of length $T = 10^8$).





$V(x) = (1 - x^2)^2$ $\qquad$ $V(x) = (1 - x^2)^2 - x/2.$

Overview

Non reversible Langevin equation

We consider the SDE

$$dX_t = (\nabla \log \pi(X_t) + \gamma(X_t))dt + \sqrt{2}dW_t, \quad \nabla \cdot (\pi(x)\gamma(x)) = 0.$$

This SDE remains ergodic with respect to π, but it has non-reversible dynamics. The choice of γ is not unique but one we study here is

$$\gamma(x) = \beta J \nabla \log \pi(x), \quad J^T = -J$$

One can show

$$\sigma_f^2(\beta) \leq \sigma_f^2(0)$$

hence the reversible dynamics $\beta = 0$ is the worse choice in terms of achieving small asymptotic variance.

Lie Trotter splitting

$$X_{n+1}^{\Delta t} = \Theta_{\Delta t} \circ \Phi_{\Delta t}(X_n^{\Delta t}),$$

where $\Phi_{\Delta t}(x)$ is a integrator that approximates the flow map corresponding to the deterministic dynamics

$$\frac{dx_t}{dt} = \gamma(x_t),$$

and $\Theta_{\Delta t}(x)$ which approximates the reversible dynamics

$$dx_t = \nabla \log \pi(x_t) dt + \sqrt{2} dW_t$$

Examples of reversible integrators

- Given $X_n^{\Delta t} \in \mathbb{R}^d$, sample Y from the proposal density $q_{\Delta t}(\cdot \,|\, X_n^{\Delta t})$.
- With probability

$$\alpha(X_n^{\Delta t}, Y) = 1 \wedge \frac{\pi(Y) q_{\Delta t}(X_n^{\Delta t} \,|\, Y)}{\pi(X_n^{\Delta t}) q_{\Delta t}(Y \,|\, X_n^{\Delta t})},$$

  set $\Theta_{\Delta t}(X_n^{\Delta t}) = Y$
- Otherwise, set $\Theta_{\Delta t}(X_n^{\Delta t}) = X_n^{\Delta t}$.

One can choose various options for proposal density $q_{\Delta t}$, in particular we consider

Random Walk Metropolis Hastings (RWMH): In this case the proposal is given by

$$q_{\Delta t}(\cdot \,|\, x) \sim \mathcal{N}(x, \Delta t),$$

Metropolis Adjusted Langevin Algorithm (MALA):

$$q_{\Delta t}(\cdot \,|\, x) \sim \mathcal{N}(x + \nabla \log \pi(x) \Delta t, 2\Delta t).$$

We will also consider the *Barker rule* sampler, which uses a RWMH proposal and with acceptance probability given by

$$\alpha(X, Y) = \frac{\pi(Y)}{\pi(X) + \pi(Y)}.$$

THE UNIVERSITY of EDINBURGH THE ALAN TURING INSTITUTE

Asymptotic Bias of the splitting methods

The fact that the integrator used for the reversible dynamics preserves the invariant measure is important as it implies that order of convergence of the integrator used for the non-reversible dynamics is a lower bound for the asymptotic bias of the corresponding splitting method.

Asymptotic variance of the splitting methods

Two sources of error

- One associated with the discreteness of the approach.

$$\sigma^2_{f,\Delta t} - \sigma^2_f = -2\Delta t \mathrm{Var}_\pi(f) + o(\Delta t)$$

- One associated with the choice of the numerical discretizations for the reversible and non-reversible part

$$\widehat{\sigma}^2_{f,\Delta t} - \sigma^2_{f,\Delta t} = \mathcal{O}(\Delta t)$$

If we combine both we see that the numerical asymptotic variance is an $\mathcal{O}(\Delta t)$ perturbation of the true one hence the numerical scheme inherits the good properties of the corresponding nonreversible SDE.

Gaussian Case

$$dX_t = AX_t dt + \Sigma_t dW_t$$

We consider

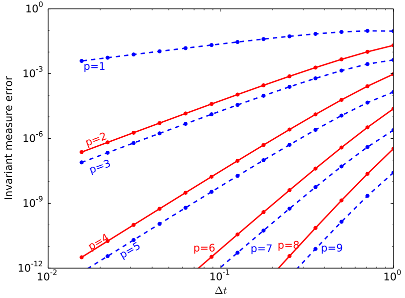
$$\Theta_{\Delta t}(x) = e^{A\Delta t}x + Q\xi_n, \quad QQ^T = \int_0^{\Delta t} e^{A(\Delta t - s)}\Sigma\Sigma^T e^{A^T(\Delta t - s)}ds.$$

Here it is possible to calculate the asymptotic bias and variance analytically since the theory of modified equations allows us to find a new SDE

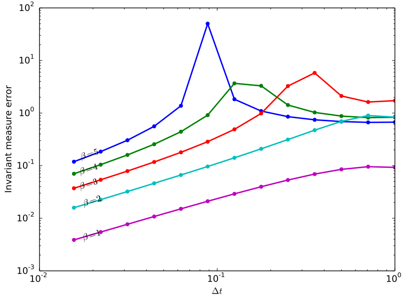
$$d\tilde{X}_t = \hat{A}\tilde{X}_t dt + \hat{\Sigma}_t dW_t$$

that the numerical method is solving exactly.

Gaussian Case: Asymptotic Bias

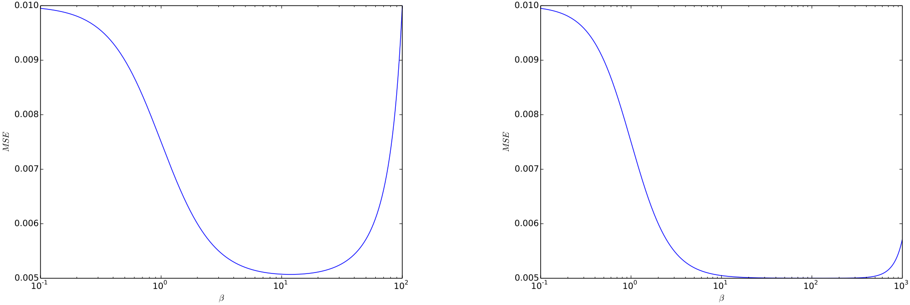


(a) Asymptotic Bias vs Δt for different order methods for the non-reversible dynamics



(b) Asymptotic Bias vs Δt for different values of β

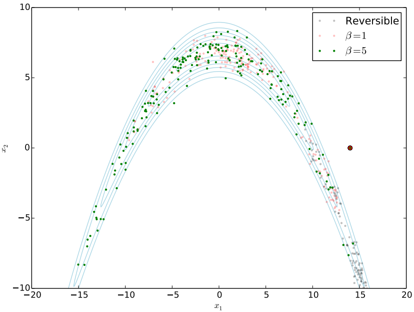
Gaussian Case: MSE



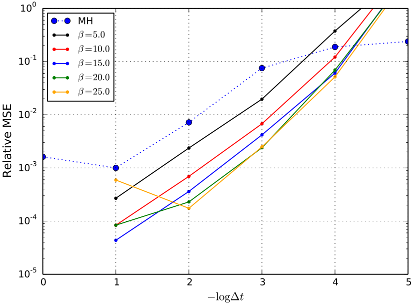
(a) First order method (b) Second order method

Figure: MSE for two different methods applied to the non-reversible part.
$T = 10^2, \quad h = 10^{-4}$

Banana shape distribution



(a) Samples for different values of β



(b) MSE when MALA is used for the reversible part (fixed budget of 6×10^6 gradient evaluations)

Figure: Dynamics and MSE error for the banana shaped distribution

Bayesian Logistic Regression

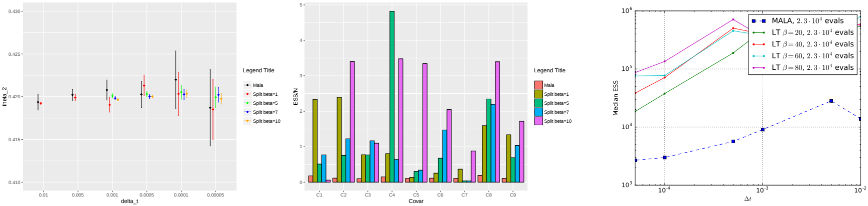
Given Data set $(X_i, Y_i)_{1\leq i\leq N}$ posterior distribution over regression coefficients θ

$$\pi(\theta|(X_i, Y_i)_{1\leq i\leq N}) \propto \exp\left(\sum_{i=1}^{N} Y_i\theta^T X_i - \log\left(1 + e^{\theta^T X_i}\right) - \frac{1}{2}\theta^T\Sigma^{-1}\theta\right)$$

Gaussian prior

$$\pi_0(\theta) = \mathcal{N}(0, \Sigma), \quad \Sigma = 100 I_{d\times d}$$

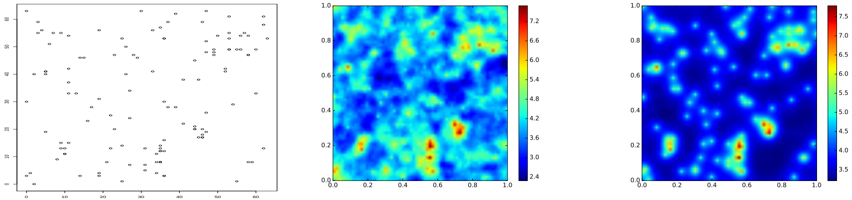
Diabetes data set $d = 8, N = 532$



(a) (b) (c)

Figure: Logistic regression for the Pima indian data set. The computational budget is set to $N = 10^6$ gradient evaluations

Spatial Model I



(a) Observed position data (b) Average Inferred Poisson intensity for MALA (c) Average Inferred Poisson intensity for splitting scheme

$$\Lambda_{i,j} = \exp(Y_{i,j}), \quad Y = (Y_{i,j}, i,j = 1, \cdots 64) \sim \mathcal{N}(\mu\mathbf{1}, \Sigma)$$

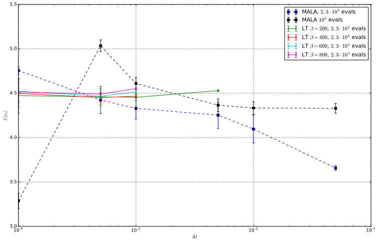
$$\Sigma_{i,j,i',j'} = \sigma^2 \left[-\frac{-\{(i-i')^2 + (j-j')^2\}^{1/2}}{64\beta} \right], \quad i,j,i',j' = 1 \cdots, 64.$$

$$f(y|x) \propto \prod_{i,j=1}^{64} \exp\{(x_{i,j}y_{i,j}) - m\exp(y_{i,j})\} \exp\{-0.5(y-\mu\mathbf{1})^T\Sigma^{-1}(y-\mu\mathbf{1})\}$$

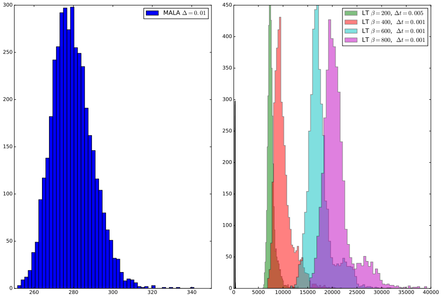
THE UNIVERSITY of EDINBURGH

THE ALAN TURING INSTITUTE

Spatial Model II



(a) Estimator of first covariate



(b) Histogram of effective sample sizes

Figure: Results for the inference of the log-Gaussian cox process. The computational budget is set to $N = 10^4$ gradient evaluations.

Overview

THE UNIVERSITY of EDINBURGH

THE ALAN TURING INSTITUTE

Multi-level Monte Carlo

To estimate $\mathbb{E}[P]$ where P can be approximated by $\widehat{P}_l$ using $h_\ell = 2^{-\ell}T$ uniform time steps, we use

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$$

$\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ is estimated using N_ℓ simulations with the same Brownian path $W(t)$ for both $\widehat{P}_\ell$ and $\widehat{P}_{\ell-1}$,

$$\hat{Y}_\ell = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(\widehat{P}_\ell^{(i)} - \widehat{P}_{\ell-1}^{(i)} \right)$$

Because of the strong convergence, on finer levels $\mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ is small and so few paths are required.

Same Brownian path $W(t) \Longrightarrow$ strong convergence $\Longrightarrow$ small variance

Modified Multilevel approach

- Note that $\widehat{P}_\ell$ appears twice, in $\mathbb{E}[\widehat{P}_{\ell+1} - \widehat{P}^c_\ell]$ and $\mathbb{E}[\widehat{P}^f_\ell - \widehat{P}_{\ell-1}]$, and $\widehat{P}^f_\ell = \widehat{P}^c_\ell$ naturally leads to cancellation and the telescoping sum.
- It may be better to use a different approximation for $\widehat{P}^f_\ell$ and $\widehat{P}^c_{\ell-1}$ in $\mathbb{E}[\widehat{P}^f_\ell - \widehat{P}^c_{\ell-1}]$, provided $\mathbb{E}[\widehat{P}^f_\ell] = \mathbb{E}[\widehat{P}^c_\ell]$.

A new MLMC:

$$\mathbb{E}[\widehat{P}^f_L] = \mathbb{E}[\widehat{P}^f_0] + \sum_{l=1}^{L} \mathbb{E}[\widehat{P}^f_\ell - \widehat{P}^c_{\ell-1}]$$

The complexity theorem is still valid.

Main Challenge

We want to extend the MLMC framework for $T \to \infty$. However for a typical SDE the constants c_1, c_2 will grow exponential with time T.

Approach:

- Restrict ourselves to a certain class of ergodic SDEs with log-concave invariant densities.
- These SDEs have exponentially contracting properties when driven by the same Brownian motion
- Exploit the exponentially contracting property of the SDE on the level of the numerical discretization by appropriately coupling of the fine and the coarse level. These will yield uniform in time estimates for the appropriate differences between the fine and the coarse paths.

Contracting properties

For the simplicity of notation take $U(x) = \log \pi(x)$ and assume that there exists $m \geq 0$ such that

$$U(y) \leq U(x) + \langle \nabla U(x), y - x \rangle - \frac{m}{2} \|x - y\|^2$$

We define

$$\psi_{s,t,W}(x) := x + \int_s^t \nabla U(X_r) dr + \int_s^t \sqrt{2} dW_r, \quad x \in \mathbb{R}^d.$$

and $X_T = \psi_{0,T,W}(X_0)$ and $Y_T = \psi_{0,T,W}(Y_0)$. Then

$$\mathbb{E}||X_T - Y_T||^2 \leq \mathbb{E}||X_0 - Y_0||^2 e^{-2mT}$$

THE UNIVERSITY of EDINBURGH
THE ALAN TURING INSTITUTE

MLMC in time

$$\lim_{T\to\infty} \mathbb{E}(g(X_T)) = \pi(g),$$

Now consider $(0 = T_0 < T_1 < T_2, \ldots, T_i <, \ldots,)$ and a sequence of random variables $(\Delta_i)_{i\geq 0}$ satisfying

$$\mathbb{E}\Delta_i = \begin{cases} \mathbb{E}g(X_{T_i}) - \mathbb{E}g(X_{T_{i-1}}) & i \geq 1 \\ \mathbb{E}g(X_{T_i}) & i = 0 \end{cases}$$

$$\pi(g) = \sum_{i=1}^{\infty} \mathbb{E}(\Delta_i).$$

THE UNIVERSITY of EDINBURGH

THE ALAN TURING INSTITUTE

Properties of the paths at different levels

We will construct the fine $X^{(f,i)}$ and the coarse $X^{(c,i)}$ paths in a way to satisfy

$$\mathcal{L}(X^{(f,i)}) = \mathcal{L}(X_{T_i}), \quad \mathcal{L}(X^{(c,i)}) = \mathcal{L}(X_{T_{i-1}}), \quad \forall i \geq 0,$$

and

$$\mathbb{E}\|X^{(f,i)} - X^{(c,i)}\|^2 \leq \mathbb{E}\|X_{T_i} - X_{T_{i-1}}\|^2.$$

Construction:

- Take $X^{(f,i)}(0) = \psi_{0,(T_i - T_{i-1}),\tilde{W}}(X(0))$
- Set $X^{(f,i)}(T_{i-1}) = \psi_{0,T_{i-1},W}(X^{(f,i)}(0)), \quad X^{(c,i)}(T_{i-1}) = \psi_{0,T_{i-1},W}(X(0))$.

THE UNIVERSITY of EDINBURGH

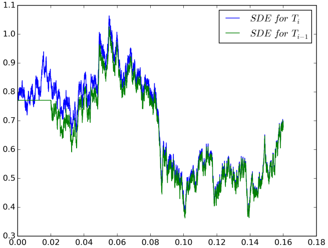
THE ALAN TURING INSTITUTE

Illustrations of couplings

We have

$$\mathbb{E}||X^{(f,i)}(T_{i-1}) - X^{(c,i)}(T_{i-1})||^2 \leq \mathbb{E}||X^{(f,i)}(0) - X(0)||^2 e^{-2mT_{i-1}}.$$

which leads to small variance for the choice of $T_i := \frac{\log 2}{2m}\beta(i+1)$



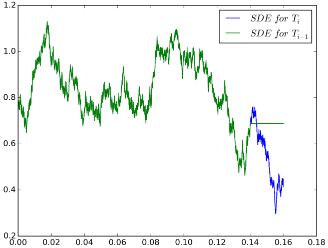


Figure: Shifted Couplings

THE UNIVERSITY of EDINBURGH

THE ALAN TURING INSTITUTE

Coupling Numerical solutions

Consider

$$x_{k+1}^h = S^f_{h,\xi_k}(x_k^h), \quad y_{k+1}^h = S^c_{h,\tilde{\xi}_k}(y_k^h), \quad P_h(x,\cdot) = \mathcal{L}\left(S^f_{h,\xi}(x)\right)$$

The coupling arises by evolving both fine and course paths jointly, over a time interval of length $T_i - T_{i-1}$, by doing two steps for the finer level (with the time step h_i) and one on the coarser level (with the time step h_{i-1}) using the discretization of the same Brownian path.

Numerical Algorithm

1. Set $x_0^{(f,i)} = x_0$, then simulate according to P_{h_i} up to $x^{(f,i)}_{\frac{t_i - t_{i-1}}{h_i}}$;
2. set $x_0^{(c,i)} = x_0$ and $x_0^{(f,i)} = x^{(f,i)}_{\frac{t_i - t_{i-1}}{h_i}}$, then simulate $(x_\cdot^{(f,i)}, x_\cdot^{(c,i)})$ jointly according to

$$\left(x_{k+1}^{(f,i)}, x_{k+1}^{(c,i)}\right) = \left(S^f_{h_i,\xi_{k,2}} \circ S^f_{h_i,\xi_{k,1}}(x_k^{(f,i)}), S^c_{h_{i-1},\frac{1}{\sqrt{2}}(\xi_{k,1}+\xi_{k,2})}(x_k^{(c,i)})\right).$$

3. set

$$\Delta_i := g\left(x^{(f,i)}_{\frac{t_{i-1}}{h_{i-1}}}\right) - g\left(x^{(c,i)}_{\frac{t_{i-1}}{h_{i-1}}}\right)$$

THE UNIVERSITY of EDINBURGH

THE ALAN TURING INSTITUTE

Ornstein Uhlenbeck process

$$dX_t = -\kappa X_t dt + \sqrt{2} dW_t,$$

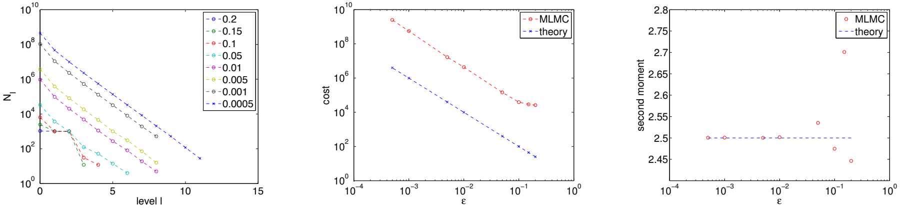


Figure: MLMC results for the OU process for $g(x) = x^2$ and $k = 0.4$

Non Lipschitz example

$$dX_t = -\left(X_t^3 + X_t\right) dt + \sqrt{2} dW_t,$$

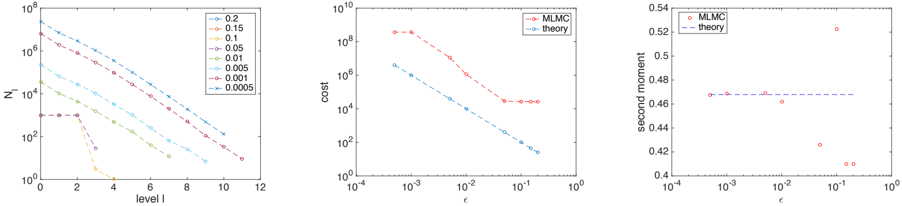


Figure: MLMC results process for $g(x) = x^2$

Bayesian Logistic Regression

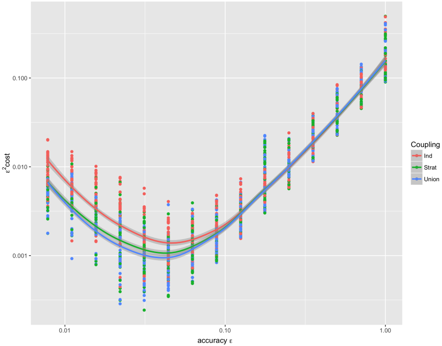


Figure: Cost of MLMC (sequential CPU time) SGLD for Bayesian Logistic Regression for decreasing accuracy parameter ϵ and different couplings

Overview

THE UNIVERSITY of EDINBURGH

THE ALAN TURING INSTITUTE

Conclusions

- Understanding the long time properties of a numerical integration has important implications for applications
- Obtained an expansion that approximates the numerical invariant measure, used then to obtain order conditions.
- Used the order conditions to study the properties of non-reversible Langevin samplers
- Illustrated applications of non-reversible samplers to "real" data sets
- Discussed about the extension of Multi-Level Monte Carlo to infinite time, which allows for optimal computational complexity

References

T. Shardlow. Modified equations for stochastic differential equations. *BIT*, 46(1): 111–125, (2006).

K. C. Zygalakis. On the existence and the applications of modified equations for stochastic differential equations. *SIAM J. Sci. Comput.* 33, 102-130, (2011).

A. Abdulle, D. Cohen, G. Villmart, K. C. Zygalakis. High order weak methods for stochastic differential equations based on modified equations. *SIAM J. Sci. Comput.* 34(3):A1800-A1823, (2012).

A. Debussche, E. Faou. Weak backward error analysis for SDEs. *SIAM J. Num. Anal.* 50(3):1735-1752, (2012).

T. Lelievre, F. Nier, G. Pavliotis. Optimal nonreversible linear drift for the convergence to equilibrium of a diffusion. J. Stat. Phys. 152(2) 237-274 (2013).

A. Abdulle, G. Villmart, K. C. Zygalakis. High order numerical approximation of the invariant measure of ergodic SDEs. *SIAM J. Num. Anal.* 52(4):1600-1622, (2014).

A. Abdulle, G. Villmart, K. C. Zygalakis. Long-run accuracy of Lie-Trotter splitting methods for second order stochastic dynamics. *SIAM J. Num. Anal.* 53(1),1-16, (2015).

A. Duncan, T. Lelievre, G. Pavliotis, Variance Reduction using Nonreversible Langevin Samplers, J. Stat. Phys. 163(3) 457-491, (2016).

M. B. Giles, L. Szpruch, S. Vollmer, K. C. Zygalakis, Multilevel Monte Carlo methods for the approximation of invariant distribution of Stochastic Differential Equations, arXiv:1605.01384, (2016).

A. Duncan, G. Pavliotis, K. C. Zygalakis. Nonreversible Langevin Samplers: Splitting Schemes, Analysis and Implementation, arXiv:1701.04247 (2017)

Thank you for your attention!