Long Time Approximation of Ergodic SDEs and Multi-Level Monte Carlo

K. C. Zygalakis

School of Mathematics, University of Edinburgh

Numerical schemes for SDEs and SPDEs
Université Lille 1
01 June 2016

THE UNIVERSITY of EDINBURGH

Collaborators

Mike Giles (Oxford)

Lukasz Szpruch (Edinburgh)

Sebastian Vollmer (Oxford)

Preprint can be found on ArXiv (arXiv:1605.01384)

THE UNIVERSITY *of* EDINBURGH

Overview

Overview

THE UNIVERSITY of EDINBURGH

Ergodic SDEs

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

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \phi(X_s)ds = \int_{\mathbb{R}^d} \phi(x)\pi(x)dx.$$

$\pi(x)$ 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

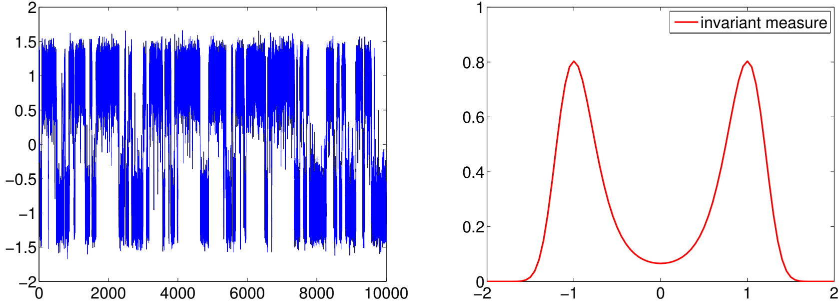

Figure: Long trajectory and invariant measure

Example problem

Standard Bayesian Model with i.i.d. data

1. prior distribution $p_0(x)$ on $x \in \mathbb{R}^{\text{dim}}$
2. $d_i \overset{\text{i.i.d.}}{\sim} p(d|x_0)$ for $i := 1, \ldots, N$
3. posterior $\pi(x) := p(x|\{d_i\}_{i=1}^N) \propto p_0(x) \prod_{i=1}^N p(d_i|x).$

How to sample from $\pi(x)$?

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

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

2. Use a Metropolis-Hastings type of algorithm

Use (1) as proposal within MCMC framework ($y_n = x_n$)

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

where

$$\alpha(y_n, x_{n+1}) = \min\left(1, \frac{\pi(x_{n+1})q(x_{n+1}, y_n)}{\pi(y_n)q(y_n, x_{n+1})}\right)$$

Things to consider when making a choice

1. The first approach (numerical analysis) introduces bias in the calculation of $\pi(x)$
2. The second approach (computational statistics) removes the bias from the calculation of $\pi(x)$

However

- Computational Statistics approach might be expensive in the presence of big data.

Computational Complexity

The quantity of interest is

$$\pi(g) = \int_{\mathbb{R}^d} g(x)\pi(x)dx = \lim_{T\to\infty} \mathbb{E}(g(X_T))$$

We are measuring accuracy in terms of the mean square error

$$\text{MSE} := \mathbb{E}(\pi(g) - \hat{g})^2$$

where $\hat{g}$ is the approximation obtain from the algorithm of choice (numerical analysis, computational statistics).

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

- numerical analysis approach: $\mathcal{O}(\epsilon^{-3})$
- computational statistics approach: $\mathcal{O}(\epsilon^{-2})$

THE UNIVERSITY of EDINBURGH

Multi-level Monte Carlo

THE UNIVERSITY of EDINBURGH

Generic approach for finite time

$$dX_t = \nabla \log \pi(X_t)dt + \sqrt{2}dW_t, \quad P = g(X_T).$$

The simplest estimator for $\mathbb{E}(P)$ is an average of N path simulations

$$\hat{Y} = N^{-1} \sum_{i=1}^{N} \widehat{P}^{(i)}$$

with $\widehat{P}^{(i)} = g(X_T^h)$.

Decomposition of MSE

$$\mathbb{E}\left[\left(\hat{Y}-\mathbb{E}(P)\right)^2\right] = \mathbb{E}\left[\left(\hat{Y}-\mathbb{E}(\widehat{P})+\mathbb{E}(\widehat{P})-\mathbb{E}(P)\right)^2\right]$$
$$= N^{-1}\mathbb{V}[\widehat{P}]+\left(\mathbb{E}(\widehat{P})-\mathbb{E}(P)\right)^2$$

- first term is due to the variance of the estimator
- second term is due to bias (finite time steps)-weak convergence

For the Euler method the combined MSE error is $\mathcal{O}(N^{-1}+h^2)$. To make this equal to $\mathcal{O}(\epsilon^2)$ we need

$$N=\mathcal{O}(\epsilon^{-2}), \quad h=\mathcal{O}(\epsilon) \Longrightarrow \quad \text{cost}=\mathcal{O}(N^{-1}h)=\mathcal{O}(\epsilon^{-3})$$

THE UNIVERSITY of EDINBURGH

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

THE UNIVERSITY of EDINBURGH

Complexity theorem

$$\left|\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]\right| \le c_1 h_\ell^\alpha, \quad \mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] \le c_2 h_\ell^\beta,$$

for some $\alpha \ge 1/2$, $\beta > 0$, $c_1 > 0$ and $c_2 > 0$, the computational complexity of the resulting multilevel estimate needed to achieve the accuracy ε is proportional to

$$\mathcal{C} = \begin{cases} \varepsilon^{-2}, & \beta > \gamma, \\ \varepsilon^{-2}\log^2(\varepsilon), & \beta = \gamma, \\ \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < \gamma \end{cases}$$

where the cost of the algorithm is of order $h^{-\gamma}$.

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.

THE UNIVERSITY of EDINBURGH

Overview

THE UNIVERSITY of EDINBURGH

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

MLMC in time

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

Now consider $(0 = T_0 < T_1 < T_2, \dots, T_i <, \dots,)$ 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

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

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 $T_i := \frac{\log 2}{2m}\beta(i+1)$

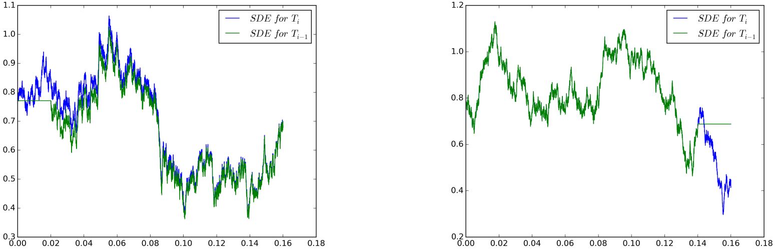

Figure: Shifted Couplings

Overview

THE UNIVERSITY *of* EDINBURGH

Coupling Numerical solutions

Consider

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

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

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

General Numerical Analysis

Definition of L^2-regularity

We will say a process $(x_k^h)_{k\in\mathbb{N}}$ is L^2-regular **uniformly in time** if it can be written as

$$S_{h,\xi_k}(x_k) - S_{h,\xi_k}(y_k) = x^h_{k+1,x_k} - x^h_{k+1,y_k} = x_k - y_k + Z_{k+1}, \quad \forall k \geq 0,$$

and for all $k \geq 0$ there exists constants $K > 0$ and random variables $\mathcal{R}_k$ and $\mathcal{H}_{k+1}$ such that

$$\begin{aligned}\mathbb{E}_k[|x^h_{k+1,x_k} - x^h_{k+1,y_k}|^2] \leq& (1-Kh)|x_k - y_k|^2 + \mathcal{R}_k \\ \mathbb{E}_k[|Z_{k+1}|^2] \leq& \mathbb{E}_k(\mathcal{H}_{k+1})|x_k - y_k|^2 h.\end{aligned}$$

Moreover we assume that there exists constants $C_{\mathcal{R}}, C_{\mathcal{H}}$ and β s.t

$$\sup_{n\geq 1} \mathbb{E}\left[\sum_{i=1}^{n} \mathcal{R}_i\right] \leq C_{\mathcal{R}} h^{2\beta}$$

$$\sup_{n\geq 1} \mathbb{E}\left[\mathcal{H}_n^2\right] \leq C_{\mathcal{H}}.$$

Assumptions for the numerical method

Consider two process $(x_k^f)_{2k\in\mathbb{N}}$ and $(x_k^c)_{k\in\mathbb{N}}$ and assume that

H0 For $q > 1$ there exist functions $H_k^f := H(k, f, q)$ and $H_k^c := H(k, c, q)$

$$\mathbb{E}|x_k^f|^q \leq H_k^f \quad \text{and} \quad \sup_{k\geq 0} \mathbb{E}|x_k^c|^q \leq H_k^f,$$

with $\sup_{k\geq 0} H_k^f < \infty$ and $\sup_{k\geq 0} H_k^c < \infty$

H1

$$\mathcal{L}\left(S^f_{h,\xi}(x)\right) = \mathcal{L}\left(S^c_{h,\tilde{\xi}}(x)\right)$$

H2 The process $(x_k^f)_{2k\in\mathbb{N}}$ is L^2 regular uniformly in time.

THE UNIVERSITY of EDINBURGH

Exponential contraction of numerical trajectories

Theorem

Take $(x_n^f)_{2n\in\mathbb{N}}$ and $(x_n^c)_{n\in\mathbb{N}}$ with $h \in (0,1]$ and assume that **H0-H2** *hold. Moreover, assume that there exists constants $c_s > 0$, $c_w > 0$ and $\alpha \geq \frac{1}{2}$, $\beta \geq 0$ with $\alpha \geq \frac{\beta}{2} + 1$ such that for all $n \geq 1$*

$$|\mathbb{E}_{n-1}(x^c_{n,x^c_{n-1}} - x^f_{n,x^c_{n-1}})| \leq c_w(1 + |x^c_{n-1}|^p)h^\alpha,$$

and

$$\mathbb{E}_{n-1}[|x^c_{n,x^c_{n-1}} - x^f_{n,x^c_{n-1}}|^2] \leq c_s(1 + |x^c_{n-1}|^{2p})h^{\beta+1},$$

where $p \geq 1$. Fix $\zeta \in (0,1)$. Then the global error is given by

$$\mathbb{E}[(x^c_{T/h,y_0} - x^f_{T/h,x_0})^2] \leq |x_0 - y_0|^2 e^{-K\zeta T} + \sum_{j=1}^{n} e^{(j-(n-1))K\zeta h}\mathbb{E}(\mathcal{R}_{n-1})$$
$$+ h^{\beta+1}\sum_{j=1}^{n} C_j(\zeta)e^{(j-(n-1))K\zeta h},$$

where $C_n(\zeta) := \left(c_s H_{1,n} + \frac{(4c_w+2)[4c_w H_{1,n} + c_s(H_{3,n}+2H_{2,n})]}{4(1-\zeta)K}\right)$ with $H_{1,n} := (1 + \mathbb{E}[|x^c_{n-1}|^{2p}])$, $H_{2,n} := (1 + \mathbb{E}[|x^c_{n-1}|^{4p}])$, $H_{3,n} := \mathbb{E}[\mathcal{H}_{n-1}]$.

Decay of MLMC variance

Let $g(\cdot)$ be a Lipschitz function. Define

$$h_i = 2^{-i}, \quad T_i \sim -\frac{\beta}{K\zeta}\left(\log h_0 + i \log 2\right), \quad \forall i \geq 0.$$

Then the resulting MLMC variance is given by.

$$\mathbb{V}[\Delta_i] \leq C 2^{-\beta i}, \quad \Delta_i = g\left(x^{(f,i)}_{\frac{t_{l-1}}{h_{i-1}}}\right) - g\left(x^{(c,i)}_{\frac{t_{i-1}}{h_{i-1}}}\right)$$

Hence we can apply the complexity theorem

Examples of Numerical Methods

Euler-Maryama method

When the Euler method is used both for the fine and the coarse paths, if we assume that $U(x)$ is Lipchitz we have

$$\begin{aligned}\mathbb{E}_{n-1}[|x_{n,x_{n-1}} - x_{n,y_{n-1}}|^2] \leq & (1-(2m-L^2h)h)|x_{n-1}-y_{n-1}|^2 \\ \mathbb{E}_k[|Z_n|^2] \leq & h^2L^2|x_{n-1}-y_{n-1}|^2\end{aligned}$$

while in addition we have the following one-step error estimates

$$\left\|\mathbb{E}[x^f_{1,x} - x^c_{1,x}]\right\| \leq \frac{h^{3/2}}{2}L\left(\mathbb{E}\left[\frac{\sqrt{h}}{2}\left(L\|x\| + \|\nabla U(0)\|\right)\right] + \frac{2}{\pi}\right). \tag{1}$$

$$\mathbb{E}\left\|x^f_{1,x} - x^c_{1,x}\right\|^2 \leq h^3\frac{L^2}{4}\left(\frac{h}{2}(\|x\|^2 + \|\nabla U(0)\|^2) + d\right) \tag{2}$$

Implying that $\alpha = 3/2$ (with additional regularity assumptions $\alpha = 2$), $\beta = 2$ hence giving computational complexity of $\mathcal{O}(\epsilon^{-2})$

THE UNIVERSITY of EDINBURGH

Beyond Lipchitz $U(x)$

One can alleviate the assumption about $U(x)$ being Lipchitz by using for example the implicit-Euler method, which can be verified that it satisfies the necessary conditions that yield computational complexity of $\mathcal{O}(\epsilon^{-2})$.

Overview

1. Introduction
   - Ergodic SDEs
   - Multi-level Monte Carlo
2. Coupling of SDEs
3. New MLMC framework
   - General Numerical Analysis
   - Examples of numerical methods
4. Stochastic Gradient Langevin Algorithm
5. Numerical Investigations

THE UNIVERSITY of EDINBURGH

Revisiting Big Data problem

We are interested in sampling with respect to

$$\pi(x) = p(\theta| \{d_i\}_{i=1}^{N}) \propto p_0(x) \prod_{i=1}^{N} p(d_i|x)$$

Here

$$\nabla \log \pi(x) = \nabla \log p_0(x) + \sum_{i=1}^{N} \nabla \log p(d_i|x)$$

Numerical discretization

Problem: When $N \gg 1$ standard approaches might be computationally infeasible

Stochastic Gradient Langevin Dynamics

$$x_{k+1} = S_{h,\xi_k,\tau_k}(x_k).$$

where

$$S_{h,\xi,\tau}(x) = x + h\widehat{f}(x) + \sqrt{2h}\xi, \quad \widehat{f}(x) = \left(\nabla \log p_0(x) + \frac{N}{n}\sum_{i=1}^{n} \nabla \log p(d_{\tau_i}|x)\right)$$

and

$$\mathbb{E}_\tau \widehat{f}(x) = \nabla \log p(x), \quad \forall \quad n \leq N.$$

Numerical Algorithm

1. Set $x_0^{(f,i)} = x_0$, then simulate according to $S_{h_i,\xi,\tau}(x)$ for $\frac{t_i - t_{i-1}}{h_i}$ steps with independent random input;
2. set $x_0^{(c,i)} = x_0$ and $x_0^{(f,i)} = x^{h_i}_{\frac{t_i - t_{i-1}}{h_i}}$, then simulate $(x_\cdot^{(f,i)}, x_\cdot^{(c,i)})$ jointly according to

$$\left(x_{k+1}^{(f,i)}, x_{k+1}^{(c,i)}\right) = \left(S_{h_i,\xi_{k,2},\tau_k^{\mathcal{F},2}} \circ S_{h_i,\xi_{k,1},\tau_k^{\mathcal{F},1}}(x_k^{(f,i)}), S_{h_{i-1},\frac{1}{\sqrt{2}}(\xi_{k,1}+\xi_{k,2}),\tau_k^{\mathcal{C}}}(x_k^{(c,i)})\right).$$

3. set

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

Possible choices of coupling

We need

$$\mathcal{L}\left(\tau^{\mathcal{F},1}\right) = \mathcal{L}\left(\tau^{\mathcal{F},2}\right) = \mathcal{L}\left(\tau^{\mathcal{C}}\right).$$

This is satisfied when

- an independent sample of $\{1, \ldots, N\}$ without replacement denoted as $\tau^{\mathcal{C}}_{\text{ind}}$ called independent coupling;
- a draw of s samples without replacement from $\left(\tau^{\mathcal{F},1}, \tau^{\mathcal{F},2}\right)$ denoted as $\tau^{\mathcal{C}}_{\text{union}}$ called union coupling;
- the concatination of a draw of $\frac{n}{2}$ samples without replacement from $\tau^{\mathcal{F},1}$ and a draw of $\frac{n}{2}$ samples without replacement from $\tau^{\mathcal{F},2}$ (provided that n is even) $\tau^{\mathcal{C}}_{\text{strat}}$ called stratified coupling.

One-step error

Nothing changes in terms of the exponent α however for all the different data coupling choices we have

$$\mathbb{E}\left\| S_{\frac{h}{2},\xi,2,\tau^{\mathcal{F},2}} \circ S_{\frac{h}{2},\xi_1,\tau^{\mathcal{F},1}}(x) - S_{h,\frac{1}{\sqrt{2}}(\xi_1+\xi_2),\tau^{\mathcal{C}}}(x)\right\|^2 \leq Ch^2$$

implying that $\beta = 1$, and hence the computational complexity is of $\mathcal{O}(\epsilon^{-2}\log\epsilon^3)$.

Overview

Ornstein Uhlenbeck process

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

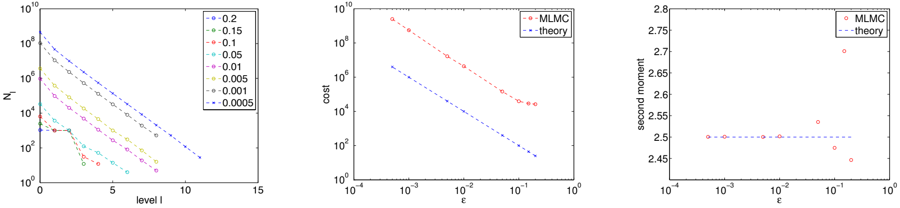

Figure: MLMC results for the OU process for $g(x) = x^2$ and $k = 0.4$

Logistic Regression

Data $d_i \in \{-1, 1\}$ are modelled by

$$p(d_i|\iota_i, x) = f(d_i x^t \iota_i)$$

where $f(z) = \frac{1}{1+\exp(-z)} \in [0,1]$ and $\iota_i \in \mathbb{R}^d$ are fixed covariates. We put a Gaussian prior $\mathcal{N}(0, C_0)$ on x, for simplicity we use $C_0 = I$ subsequently. By Bayes' rule the posterior π satisfies

$$\pi(x) \propto \exp\left(-\frac{1}{2}\|x\|^2_{C_0}\right)\prod_{i=1}^{N} f(y_i x^T \iota_i).$$

We consider $d = 3$ and $N = 100$ data points and choose the covariate to be

$$\iota = \begin{pmatrix} \iota_{1,1} & \iota_{1,2} & 1 \\ \iota_{2,1} & \iota_{2,2} & 1 \\ \vdots & \vdots & \vdots \\ \iota_{100,1} & \iota_{100,2} & 1 \end{pmatrix}$$

for a fixed sample of $\iota_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $i = 1, \ldots 100$ and we take $n = 10$.

Numerical experiment

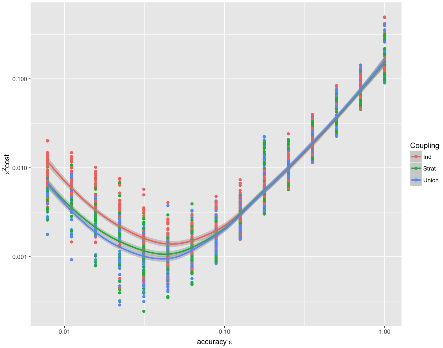

Figure: Cost of MLMC (sequential CPU time) SGLD for Bayesian Logistic Regression for decreasing accuracy parameter ϵ and different couplings

Conclusions and future work

- We have extended the MLMC framework to infinite time for a class of ergodic SDEs
- We have achieved the computational complexity as for MCMC methods without the need of accept reject (we can make our samples unbiased)

1. Study tamed methods
2. Investigate different couplings for SGLD to bring computational complexity down to $\mathcal{O}(\epsilon^{-2})$

Thank you for your attention!