Multilevel Monte Carlo for Stochastic McKean-Vlasov Equations

Lukasz Szpruch

School of Mathemtics
University of Edinburgh

joint work with Shuren Tan (Edinburgh)

MV-SDEs

d-dimensional McKean-Vlasov SDEs:

$$dX_t = b(X_t, \mathbb{P}_t)dt + \sigma(X_t, \mathbb{P}_t)dW_t, \quad \mathbb{P}_t = \mathbb{P} \circ X_t^{-1} = Law(X_t), \quad t \in [0, T],$$

where $\{W_t\}_{t \geq 0}$ is k-dimensional Brownian motion and $\mathbb{P}$ is a probability measure on $C([0, T], \mathbb{R}^d)$.
Example:

$$dY_t = \mathbb{E}[b(y, Y_t)]|_{y=Y_t} dt + dW_t = \int_{\mathcal{C}} b(Y_t, y)\mathbb{P}_t(dy)dt + dW_t.$$

Goal:

$$\mathbb{E}[G(X_T)] \quad \text{or} \quad \mathbb{E}[P((X_t)_{0 \leq t \leq T}]$$

for $G : \mathbb{R}^d \to \mathbb{R}$ or $P : C([0, T], \mathbb{R}^d) \to \mathbb{R}$

Classical Framework

Existence and uniqueness of the solution holds if $\forall x, y \in \mathbb{R}^d \quad \forall \mathbb{P}, \mathbb{Q} \in \mathcal{P}_2(\mathbb{R}^d)$

$$|b(x,\mathbb{P}) - b(y,\mathbb{Q})| + |\sigma(x,\mathbb{P}) - \sigma(y,\mathbb{Q})| \leq L(|x-y| + W_2(\mathbb{P},\mathbb{Q})),$$

where 2-Wasserstein distance, $W_2(\cdot,\cdot)$, is defined as

$$W_2(\mu,\nu) = \inf_{\gamma} \left[\int_{\mathbb{R}^{2d}} |u-v|^2 \gamma(du,dv);\ \gamma(\cdot \times \mathbb{R}^d) = \mu, \gamma(\mathbb{R}^d \times \cdot) = \nu \right].$$

(see Sznitman 1991)

Motivation

- MV-SDEs gives probabilistic interpretation of nonlinear McKean-Vlasov PDEs which weak formulation with $f(\cdot) \in C_K^\infty(\mathbb{R}^d)$ is given

$$\begin{cases} \frac{\partial}{\partial t}\langle \mathbb{P}_t, f \rangle & = \langle \mathbb{P}_t, \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x, \mathbb{P}_t)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^d b_i(x, \mathbb{P}_t)\frac{\partial f}{\partial x_i}(x)\rangle, \\ \mathbb{P}_0 & = \mathbb{P} \circ X_0^{-1} = Law(X_0), \end{cases}$$

  where $a(x, \mathbb{P}_t) = \sigma(X_t, \mathbb{P}_t)^T \sigma(X_t, \mathbb{P}_t)$.
- Applications:
  - Lagrangian models (Bossy,Jabir,Talay, 2011)
  - Navier-Stokes equation for the vorticity of a two-dimensional incompressible fluid flow and many more (Bossy,Jourdain,Meleard,Reygner,Talay...)
  - Mean-Field Games (Lasry, Lions, 2007, Chassagneux, Crisan, Delarue, 2015)
  - Stochastic Local Volatility Models (Gyongy, 1996, Guyon, Henry-Labordere 2011, Jourdain ,Zhou 2016)

Propagation of chaos

- *stochastic interacting particles* $(X_t^{i,N})$ are solutions to $(\mathbb{R}^d)^N$ dimensional SDEs

$$\begin{cases} dX_t^{i,N} & = b(X_t^{i,N}, \mathbb{P}_t^N)dt + \sigma(X_t^{i,N}, \mathbb{P}_t^N)dW_t^i, \quad i = 1, \dots, N, \\ \mathbb{P}_t^N & := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}, \quad t \geq 0, \end{cases}$$

  where $\{X_0^{i,N}\}_{i=1,\dots,N}$ are i.i.d samples with the law $\mathbb{P}_0$ and $\{W_t^i\}_{i=1,\dots,N}$ are independent Brownian motions.
- Very rich modelling framework:
  - interaction agents in economics (Lasry, Lions, 2007)
  - neuronal networks (Delarue, Inglis, Rubenthaler, Tanre, 2015)
  - systemic risk (Carmona, Fouque, Sun, 2013)
- $X_t^{i,N}$ converge weakly to X_t^i when $N \to \infty$.

Euler Scheme

Consider MV-SDEs

$$dX_t = \int_{\mathbb{R}^d} b(X_t, y)\mathbb{P}_t(dy)dt + \int_{\mathbb{R}^d} \sigma(X_t, y)\mathbb{P}_t(dy)dW_t,$$

Euler scheme with time-step $h = T/M$, i=1,...,N,

$$Y_{k+1}^{i,N} = Y_k^{i,N} + \frac{1}{N}\sum_{j=1}^{N} b(Y_k^{i,N}, Y_k^{j,N})h + \frac{1}{N}\sum_{j=1}^{N} \sigma(Y_k^{i,N}, Y_k^{j,N})\Delta W_{k+1}^i.$$

- Due to the particle interactions, its implementation requires N^2 arithmetic operations at each step.
- Euler scheme converges with weak rate of order $((\sqrt{N})^{-1} + h)$ Bossy, Talay (1997) Antonelli, Kohatsu-Higa (2002), Bossy, Jourdain (2002).
- Notice that the same "sample" is used to approximate MV-SDEs and to evaluate the $\mathbb{E}[G(X_T)]$.

A cost of the propagation of chaos

- Consider mean-square-error

$$\mathbb{E}\left[\left(\mathbb{E}[f(X_T)] - \frac{1}{N}\sum_{i=1}^{n} f(Y_T^{i,N})\right)^2\right]$$

- bias and statistical error are in a nonlinear relationship
- Consider iid samples

$$\bar{X}_{k+1} = \bar{X}_k + b(\bar{X}_k, \mathbb{P}_{kh})h + \sigma(\bar{X}_k, \mathbb{P}_{kh})\Delta W_{k+1}, \quad \mathbb{P}_{kh} = \mathbb{P} \circ (\bar{X}_k)^{-1}.$$

- Error decomposition

$$\mathbb{E}[f(X_T)] - \frac{1}{N}\sum_{i=1}^{N} f(Y_T^{i,N}) = \left(\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]\right)$$

$$+ \left(\mathbb{E}[f(\bar{X}_T)] - \frac{1}{N}\sum_{i=1}^{N} f((\bar{X}_T^i)\right)$$

$$+ \frac{1}{N}\sum_{i=1}^{N}\left(f((\bar{X}_T^i) - f(Y_T^{i,N})\right).$$

Cost

- Typical mean-square error

$$\mathbb{E}\left[\left(\mathbb{E}[f(X_T)] - \frac{1}{N}\sum_{i=1}^{N} f(Y_T^{i,N})\right)^2\right] \le C(h^2 + \frac{1}{N} + \frac{1}{N}),$$

- Cost $\mathcal{C}_\gamma = N^\gamma h^{-1}$, $\gamma = 1$ no-interacting Kernel, $\gamma = 2$ interacting Kernel.
- For the root-mean-square-error ϵ the cost is $\mathcal{C}_1 = \epsilon^{-3}$ or $\mathcal{C}_2 = \epsilon^{-5}$
- Example of the non-interacting Kernel particle system:

$$Y_{k+1}^{i,N} = Y_k^{i,N} + b(Y_k^{i,N}, \frac{1}{N}\sum_{j=1}^{N} f(Y_k^{j,N}))h + \sigma \Delta W_{k+1}^i.$$

MLMC for standard SDEs

Idea of Giles (2006), Heinrich (2001) was to explore the identity

$$\mathbb{E}[P_L] = \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell - P_{\ell-1}],$$

where $P_\ell := P(Y^{M_\ell})$ with $P : C([0,T], \mathbb{R}^d) \to \mathbb{R}$ and $\{Y^{M_\ell}\}$, $\ell = 0 \dots L$, being discrete time approximation of process X with M_ℓ number of time steps.

This identity leads to an unbiased estimator of $\mathbb{E}[P(Y^{M_L})]$,

$$\frac{1}{N_0} \sum_{i=1}^{N_0} P_0^{(i,0)} + \sum_{\ell=1}^{L} \left\{ \frac{1}{N_\ell} \sum_{i=1}^{N_l} (P_\ell^{(i,\ell)} - P_{\ell-1}^{(i,\ell)}) \right\},$$

where $P_\ell^{(i,\ell)} = P((Y^{M_\ell})^{(i)})$ are independent samples at level ℓ.

- But for MLMC variance for particle systems decays as $(N^{-1} + h)$

Theorem (Giles, 2006)

If there exist independent estimators Y_ℓ based on N_ℓ Monte Carlo samples, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2}\,\min(\beta,\gamma)$ and

i) $|\mathbb{E}[P_\ell - P]| \leq c_1\, 2^{-\alpha\,\ell}$

ii) $\mathbb{E}[Y_\ell] = \begin{cases} \mathbb{E}[P_0], & \ell = 0 \\ \mathbb{E}[P_\ell - P_{\ell-1}], & \ell > 0 \end{cases}$

iii) $\mathbb{V}[Y_\ell] \leq c_2\, N_\ell^{-1} 2^{-\beta\,\ell}$

iv) $C_\ell \leq c_3\, N_\ell\, 2^{\gamma\,\ell}$, *where C_ℓ is the computational complexity of Y_ℓ*

then there exists a positive constant c_4 such that for any $\varepsilon < e^{-1}$ there are values L and N_ℓ for which the multilevel estimator

$$Y = \sum_{\ell=0}^{L} Y_\ell,$$

has a mean-square-error with bound

$$MSE \equiv \mathbb{E}\left[(Y - \mathbb{E}[P])^2\right] < \varepsilon^2$$

with a computational complexity C with bound

$$C \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > \gamma, \\ c_4\, \varepsilon^{-2} (\log \varepsilon)^2, & \beta = \gamma, \\ c_4\, \varepsilon^{-2-(\gamma-\beta)/\alpha}, & 0 < \beta < \gamma. \end{cases}$$

MLMC - multicloud

- generate R_l *independent* clouds of particles $\{\mathcal{C}_j^{N_l}\}_{j=1,\dots,R_l}$ with N_l interacting particles in each cloud i.e. $\{X_t^{((N_l,i);j)} : i = 1,\dots,N_l\}_{j=1,\dots,R_l}$.
- *propagation of chaos* property suggests to consider $R = 1$!
- Define estimator:

$$\frac{1}{R_0}\sum_{j=1}^{R_0}\frac{1}{N_0}\sum_{i=1}^{N_0} f(X_T^{((N_0,i);j)})$$
$$+\sum_{l=1}^{L}\frac{1}{R_l}\sum_{j=1}^{R_l}\left(\frac{1}{N_l}\sum_{i=1}^{N_l} f(X_T^{((N_l,i);j)}) - \frac{1}{N_{l-1}}\sum_{i=1}^{N_{l-1}} f(X_T^{((N_{l-1},i);j)})\right),$$

- coupling is introduced by considering particles $X_T^{((N_{l-1},i);j)}$ and $X_T^{((N_{l-1},i);j)}$ from the same cloud.
- MLMC complexity theorem by Giles with $\beta = 1$, $\gamma = 2$, $\alpha = 1/2$ we obtain computational complexity ϵ^{-5}, which is the same as for the propagation of chaos estimator. But see Hajj-Ali and Tempone and Multi-index approach.

Sznitman's iteration proof

$$\begin{cases} X_t &= X_0 + W_t + \int_0^t \int_C b(X_s, y)\mathbb{P}_t(dy)ds, \quad 0 \le t \le T \\ \mathbb{P}_t &= Law(X_t) \end{cases}$$

Pick a measure $\mu \in \mathcal{P}(C[0,T],\mathbb{R}^d)$. Define an operator $\Phi : \mathcal{P}(C[0,T],\mathbb{R}^d) \mapsto \mathcal{P}(C[0,T],\mathbb{R}^d)$ that returns $Law(X^\mu)$

$$X_t^\mu = X_0 + W_t + \int_0^t \int_C b(X_s^\mu, y)\mu_t(dy)ds, \quad 0 \le t \le T.$$

Theorem (Sznitman)

Let $T > 0$, and $\mu \in \mathcal{P}(C[0,T],\mathbb{R}^d))$. There exists $C > 0$ st.

$$W_2(\Phi^{k+1}(\mu), \Phi^k(\mu)) \le C\frac{T^k}{k!}W_2(\Phi(\mu), \mu).$$

Picard's Particle system

Let m be the index corresponding to the Picard step.

$$dX_t^m = \int_C b(X_t^m, y) Law_t(X^{m-1})(dy)\, dt + \sigma\, dW_t^m, \quad X_0^m = X_0,$$

Let $(Y_t^{0,n,N_0})_{1 \le n \le N_0}$ be an i.i.d. sample with law $\mathbb{P}_0$. For $m \ge 1$, we define

$$dY_t^{m,n,N_m} = \frac{1}{N_{m-1}} \sum_{j=1}^{N_{m-1}} b(Y_t^{m,n,N_m}, Y_t^{m-1,j,N_{m-1}})\, dt + \sigma\, dW_t^{m,n}, \quad 1 \le n \le N_m,$$

Key idea:

- Use next to last Picard steps to approximate $\int_C b(X_s^\mu, y)\mu_t(dy)$
- Use the final Picard step to approximate the quantity of interest.

Picard's Particle system

Theorem

We assume that the interacting kernel $b : \mathbb{R}^2 \to \mathbb{R}$ *satisifes*

(H1) $b \in C_b^2(\mathbb{R}^2)$,

(H2) b *is globally Lipschitz*

$$\sup_{t\in[0,T]} \left|\mathbb{E}\phi(X_t) - \mathbb{E}\phi(Y_t^{m,h_m,N_m})\right|$$

$$\leq K\left[\left(\frac{T^m}{m!}\right) + \frac{1}{\sqrt{N_m}} + h_m + \sum_{s=1}^{m} \frac{1}{(s-1)!}\left(\frac{1}{\sqrt{N_{m-s}}} + h_{m-s}\right)\right].$$

$$\sup_{t\in[0,T]} \mathbb{E}\left|\phi(X_t) - \phi(Y_t^{m,h_m,N_m})\right|^2$$

$$\leq K\left[\left(\frac{T^m}{m!}\right) + \frac{1}{N_m} + (h_m)^2 + \sum_{s=1}^{m} \frac{1}{(s-1)!}\left(\frac{1}{N_{m-s}} + (h_{m-s})^2\right)\right].$$

Even simpler example

$$dX_t^m = b(X_t^m, \mathbb{E}[f(X_t^{m-1})])dt + \sigma\, dW_t^m, \quad X_0^m = X_0,$$

Euler scheme:

$$dZ_t^{m,\ell} = b(Z_{\eta(t)}^{m,\ell}, \mathbb{E}[f(Z_{\eta(t)}^{m-1,\ell})])dt + \sigma\, dW_t^m,$$

where $\eta(t) = t_k$ for $t \in [t_k, t_{k+1})$. We can now write a telescopic MLMC sum

$$\mathbb{E}[f(Z_{\eta(t)}^{m-1,L})] = \mathbb{E}[f(Z_{\eta(t)}^{m-1,0})] + \sum_{\ell=1}^{L} \mathbb{E}[f(Z_{\eta(t)}^{m-1,\ell}) - f(Z_{\eta(t)}^{m-1,\ell-1}]$$

and resulting MC estimator is given by

$$[\mathcal{M}_{t_k}^m(Z)](G) = \frac{1}{N_{0,m}} \sum_{i=1}^{N_{0,m}} G(Z_{t_k}^{i,m,\ell}) + \sum_{\ell=1}^{L_m} \frac{1}{N_{\ell,m}} \sum_{i=1}^{N_{\ell,m}} \Big(G(Z_{t_k}^{i,m,\ell}) - G(Z_{t_k}^{i,m,\ell-1}) \Big)$$

Picard's Particle system

$$dY_t^{i,N_m,\ell} = b(Y_{\eta(t)}^{i,N_m,\ell}, [\mathcal{M}_{\eta(t)}^{m-1}(Y)](f))dt + \sigma dW_t^i.$$

- We interpolate MLMC on $[0, T]$
- Picard approach ensures telescoping sum is preserved - the same mean-filed approximation used across the levels.
- Standard error analysis plus control of the term

$$[\mathcal{M}_{t_k}^{m-1}(Y)](f) - [\mathcal{M}_{t_{k+1})}^{m-1}(Y)](f)$$

Picard's Particle system

Error decomposition:

$$\mathbb{E}\left[\left(\mathbb{E}[G(X_T)] - [\mathcal{M}^m_T(Y)](G)\right)^2\right] \leq 2\left(\mathbb{E}[G(X_T)] - \mathbb{E}[G(Y^m_T)]\right)^2 + \mathbb{V}^{MLMC,m}_T$$

Now the term $\mathbb{V}^{MLMC,m}_T := \mathbb{E}(\mathbb{E}[G(Y^m_t)]) - [\mathcal{M}^m_{t_k}(Y)](G)))^2$ (not exactly a variance). Consequently

$$\mathbb{E}\left[\left(\mathbb{E}[G(X_T)] - [\mathcal{M}^m_T(Y)](G)\right)^2\right]$$
$$\leq C\left(\frac{1}{N_m} + h_L + \int_0^T \sup_{x\in\mathbb{R}^d} \mathbb{E}\left(b(x, \mathcal{M}^{m-1}_{\eta(s)}(Y)](f)) - B(\eta(s), x)\right)^2 ds\right) + \mathbb{V}^{MLMC,m}_T$$

where $B(t,x) = b(x, \mathbb{E}[f(X_t)])$.

Nested MLMC

But by the same argument

$$\left(\mathbb{E}[b(x,\mathcal{M}^{m-1}_{\eta(s)}(Y)](f)) - B(\eta(s),x)\right)^2$$
$$\leq C\left(\frac{1}{N_{m-1}} + h_L + \int_0^{\eta(s)} \sup_{x\in\mathbb{R}^d} \mathbb{E}\left(b(x,\mathcal{M}^{m-2}_{\eta(\theta)}(Y)](f)) - B(\eta(\theta),x)\right)^2 d\theta\right)$$
$$+ \mathbb{V}^{MLMC,m}_{\eta(s)}$$

For $\mathbb{R}^k$-valued random variable $\mathcal{M}$ by

$$\mathbb{V}[\mathcal{M}] = \mathbb{E}[\|\mathcal{M} - \mathbb{E}[\mathcal{M}]\|^2_\infty].$$

It can be further shown

$$\mathbb{V}[\mathcal{M}] \leq c(k)\sum_{i=1}^n \mathbb{V}(\mathcal{M}_i),$$

$$c_1\log(k+1) \leq c(k) \leq c_2\log(k+1).$$

[M Ledoux, M Talagrand, 1991]

Interacting kernel

$$dX_t^m = \int_C b(X_t^m, y) Law_t(X^{m-1})(dy)\, dt + \sigma\, dW_t^m, \quad X_0^m = X_0,$$

Euler scheme:

$$dZ_t^{m,\ell} = \mathbb{E}[b(x, Z_{\eta(t)}^{m-1,\ell})]_{x=Z_{\eta(t)}^{m,\ell}} dt + \sigma\, dW_t^m,$$

where $\eta(t) = t_k$ for $t \in [t_k, t_{k+1})$. We can now write a telescopic MLMC sum

$$\mathbb{E}[b(x, Z_{\eta(t)}^{m-1,L})] = \mathbb{E}[b(x, Z_{\eta(t)}^{m-1,0})] + \sum_{\ell=1}^{L} \mathbb{E}[b(x, Z_{\eta(t)}^{m-1,\ell}) - b(x, Z_{\eta(t)}^{m-1,\ell-1}]$$

and resulting MC estimator is given by

$$\begin{aligned}[\mathcal{M}_{t_k}^m(x, Z)](b) =& \frac{1}{N_{0,m}} \sum_{i=1}^{N_{0,m}} b(x, Z_{t_k}^{i,m,\ell}) \\ &+ \sum_{\ell=1}^{L_m} \frac{1}{N_{\ell,m}} \sum_{i=1}^{N_{\ell,m}} \Big(b(x, Z_{t_k}^{i,m,\ell}) - b(x, Z_{t_k}^{i,m,\ell-1}) \Big)\end{aligned}$$

Non-interacting kernel

The target stochastic differential equation is

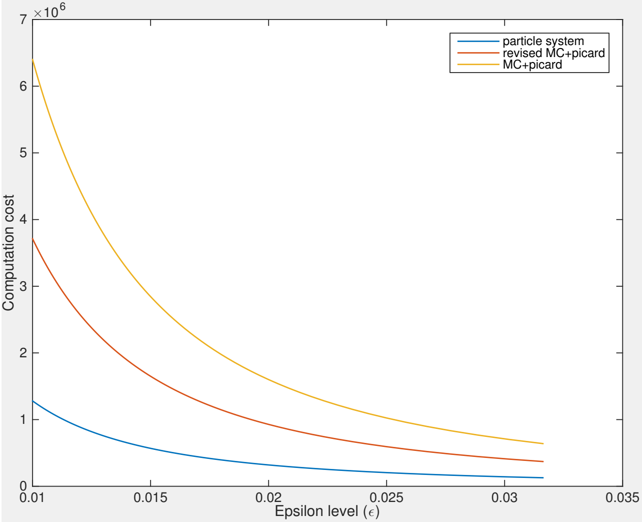
$$dX_t = \sin(X_t - \mathbb{E}[X_t])dt + \sigma dW_t \ , \quad X_0 = 0.$$

The testing payoff function is

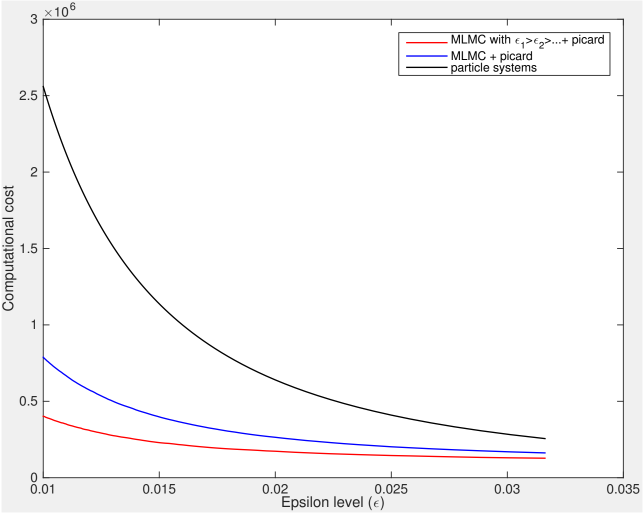
$$P(x) = \max(x - K, 0),$$

where strike K is set to 0.1.

Non-interacting kernel



Non-interacting case



Interacting kernel

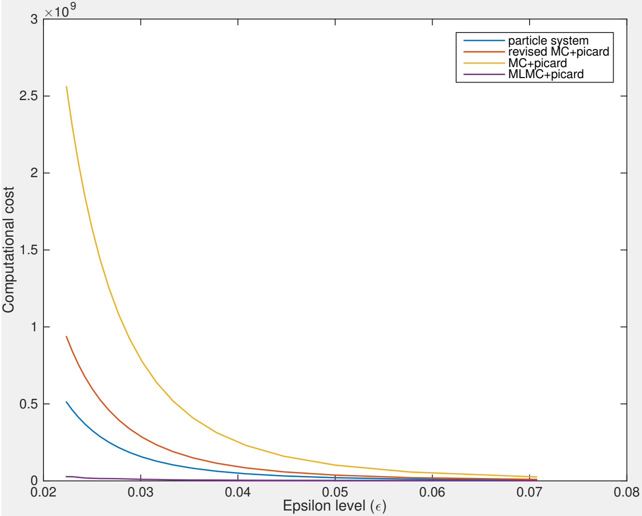
The target stochastic differential equation is

$$dX_t = \mathbb{E}[\sin(x - X_t)]|_{x=X_t} dt + \sigma dW_t \ , \quad X_0 = 0,$$

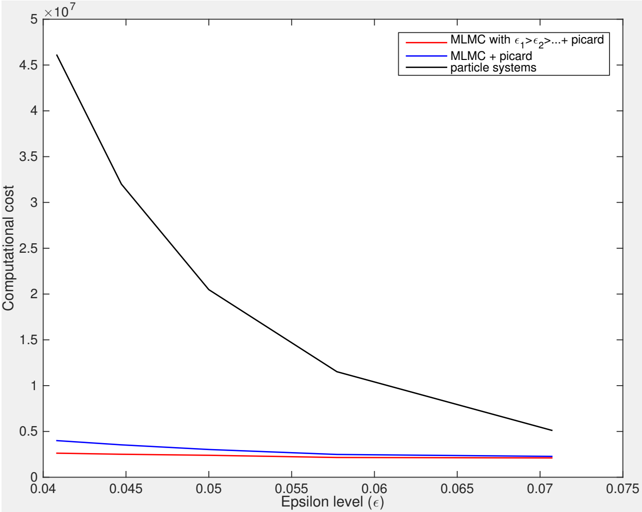
where Y_t is an independent copy of X_t. The payoff

$$P(x) = \sqrt{1 + x^2}.$$

Interacting kernel



Interacting kernel



Thank You