

# Multilevel Monte Carlo for Stochastic McKean-Vlasov Equations

Lukasz Szpruch

School of Mathematics  
University of Edinburgh

joint work with Shuren Tan and Alvin Tse (Edinburgh)



## MKV-SDEs

McKean-Vlasov SDEs on  $[0, T]$ , of the form

$$\begin{cases} dX_t = \int_{\mathbb{R}^d} b(X_t, y) \mu_t(dy) dt + \int_{\mathbb{R}^d} \sigma(X_t, y) \mu_t(dy) dW_t, \\ \mu_t = \text{Law}(X_t), \quad X_0 \in \mathbb{R}^d. \end{cases}$$

- $b \in C_b^{2,1}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)$  and  $\sigma \in C_b^{2,1}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^{d \otimes r})$ .
- $\implies$  For all  $x_1, x_2, y_1, y_2 \in \mathbb{R}^d$ , there exists a constant  $L$  such that

$$\begin{aligned} |b(x_1, y_1) - b(x_2, y_2)| + \|\sigma(x_1, y_1) - \sigma(x_2, y_2)\| &\leq L(|x_1 - x_2| + |y_1 - y_2|), \\ |b(x_1, y_1)| + \|\sigma(x_1, y_1)\| &\leq L(1 + |x_1| + |y_1|). \end{aligned}$$

- $\implies$  (can be relaxed)

$$\sup_{(x_1, y_1) \in \mathbb{R}^d \times \mathbb{R}^d} \|b(x_1, y_1)\| + \sup_{(x_2, y_2) \in \mathbb{R}^d \times \mathbb{R}^d} \|\sigma(x_2, y_2)\| \leq K < \infty.$$

- The initial law  $\mu_0$  satisfies:
  - ▶  $\mu_0$  is a Dirac measure at  $X_0$ , or
  - ▶  $p \geq 2$ ,  $\int_{\mathbb{R}^d} |x|^p \mu_0(dx) < \infty$ .

## Motivation

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

$$\begin{cases} \frac{\partial}{\partial t} \langle \mu_t, f \rangle &= \langle \mu_t, \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x, \mu_t) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^d b_i(x, \mu_t) \frac{\partial f}{\partial x_i}(x) \rangle, \\ \mu_0 &= \mathbb{P} \circ X_0^{-1} = \text{Law}(X_0), \end{cases}$$

where  $a(x, \mu_t) = \sigma(X_t, \mu_t)^T \sigma(X_t, \mu_t)$  and  $b(x, \mu_t) := \int_{\mathbb{R}^d} b(x, y) \mu_t(dy)$ .

- Example:  $dY_t = \int_{\mathbb{R}^d} b(Y_t, y) \mu_t(dy) dt + dW_t$ .
- 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,Reygnier,Talay...)
  - ▶ Mean-Field Games (Lasry, Lions, 2007, Chassagneux, Crisan, Delarue, 2015)
  - ▶ Stochastic Local Volatility Models (Gyongy, 1996; Guyon, Henry-Labordere 2011; Guyon 2014,2015; Jourdain ,Zhou 2016)

## Propagation of chaos

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

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

where  $\{X_0^{i,N}\}_{i=1,\dots,N}$  are i.i.d samples with the law  $\mu_0$  and  $\{W_t^i\}_{i=1,\dots,N}$  are independent Brownian motions.

- $X_t^{i,N} \Rightarrow X_t^i$  when  $N \rightarrow \infty$ .

## Time discretization - Euler Scheme

Consider MV-SDEs

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

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

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

- 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  $\mathbb{E}[G(X_T)]$ .

## Computational 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

$$\begin{aligned} \mathbb{E}[f(X_T)] - \frac{1}{N} \sum_{i=1}^N f(Y_T^{i,N}) &= (\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]) \\ &\quad + (\mathbb{E}[f(\bar{X}_T)] - \frac{1}{N} \sum_{i=1}^N f((\bar{X}_T^i))) \\ &\quad + \frac{1}{N} \sum_{i=1}^N (f((\bar{X}_T^i)) - f(Y_T^{i,N})). \end{aligned}$$

## Cost

- Typical mean-square error

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

- Cost  $\mathcal{C}_\gamma = N^\gamma h^{-1}$ ,  $\gamma = 1$  no-interacting Kernel,  $\gamma = 2$  interacting Kernel.
- For the root-mean-square-error  $\epsilon$  the cost is  $\mathcal{C}_1 = \mathcal{O}(\epsilon^{-3})$  or  $\mathcal{C}_2 = \mathcal{O}(\epsilon^{-5})$
- 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  $P : C([0, T], \mathbb{R}^d) \rightarrow \mathbb{R}$  and  $\{Y^{M_\ell}\}$ ,  $\ell = 0 \dots L$ , being discrete time approximation of process  $X$  with  $M_\ell$  number of time steps.

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

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

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

- But for MLMC variance for particle systems, typically decays as  $\mathcal{O}(N^{-1} + h)$
- In the case of particle systems, one need to be careful to ensure telescopic sum is preserved.

## Sznitman's iteration proof

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

- Step 1: Pick a measure  $\mu \in \mathcal{P}(C[0, T], \mathbb{R}^d)$
- Step 2: Define an operator  $\Phi : \mathcal{P}(C[0, T], \mathbb{R}^d) \mapsto \mathcal{P}(C[0, T], \mathbb{R}^d); \Phi(\mu) = \text{Law}(X^\mu)$

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

- Iterate

### Theorem (Sznitman)

Let  $T > 0$ , and  $\mu \in \mathcal{P}_2(C[0, T], \mathbb{R}^d)$ . There exists  $C > 0$  st.

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

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

## Picard's Particle system

Consider a sequence of SDEs linear (in a sense of McKean) defined as

$$dX_t^m = \int_{\mathcal{C}} b(X_t^m, y) \mu_t^{m-1}(dy) dt + \int_{\mathcal{C}} \sigma(X_t^m, y) \mu_t^{m-1}(dy) dW_t^m, \quad \text{Law}(X_0^m) = \text{Law}(X_0),$$

where  $\mu_t^{m-1} = \text{Law}(X_t^{m-1})$ . Let  $(Y_t^{0,n,N_0})_{1 \leq n \leq N_0}$  be an i.i.d. sample with law  $\mu_0$ . For  $m \geq 1$ , we define

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

Key idea:

- Use  $m - 1$  steps to approximate  $\int_{\mathbb{R}^d} b(X_s^\mu, y) \mu_t(dy)$
- Use the final  $m$  Picard step to approximate the quantity of interest.
- Iterated Particle system is less efficient than original Particle system!

## Iterative Particle System

Consider a sequence of SDEs linear (in a sense of McKean) defined as

$$dX_t^m = \int_{\mathbb{R}^d} b(X_t^m, y) \mu_t^{m-1}(dy) dt + \int_{\mathbb{R}^d} \sigma(X_t^m, y) \mu_t^{m-1}(dy) dW_t^m, \quad \text{Law}(X_0^m) = \text{Law}(X_0),$$

where  $\mu_t^{m-1} = \text{Law}(X_t^{m-1})$  and  $W^m$  and  $X_0^m$  are independent.

Iterative particle system

$$Y_{k+1}^{i,m,\ell} = Y_k^{i,m,\ell} + [\mathcal{M}_{t_k^\ell}(\tilde{\mu}_Y^{m-1})](b(Y^{i,m,\ell}, \cdot)) h_\ell + [\mathcal{M}_{t_k^\ell}(\tilde{\mu}_Y^{m-1})](\sigma(Y^{i,m,\ell}, \cdot)) \Delta W_{k+1}^{i,m,\ell}.$$

- No-interacting Kernels are much easier to analyse.

## Iterative Particle System

For any borel function  $G = (G^{(1)}, \dots, G^{(d)})$ , where  $G^{(i)} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ , we define

$$[\mathcal{M}_t(\mu)](G(x, \cdot)) := \sum_{\ell=0}^L \left( \mu_t^\ell - \mu_t^{\ell-1} \right) (G(x, \cdot)), \quad \mu \in \mathcal{P}_2,$$

where  $\mu_t^\ell(G(x, \cdot)) = \int_{\mathbb{R}^d} G(x, y) \mu_t^\ell(dy)$ .

$$\mu_t^\ell := \begin{cases} \left[ \frac{t - \eta_l(t)}{h_l} \right] \mu_{\eta_\ell(t) + h_l}^\ell + \left[ 1 - \frac{t - \eta_l(t)}{h_l} \right] \mu_{\eta_\ell(t)}^\ell & , t \notin \Pi^\ell, \\ \frac{1}{N_{\ell,m}} \sum_{i=1}^{N_{\ell,m}} \delta_{Y_t^{i,m,\ell}} & , t \in \Pi^\ell. \end{cases}$$

## Further assumptions

Define  $B(t, x) = \mathbb{E}[b(x, X_t)]$  and  $\Sigma(t, x) = \mathbb{E}[\sigma(x, X_t)]$  ( $B(\cdot, \cdot) \in C^{1,2}$  and  $\Sigma(\cdot, \cdot) \in C^{1,2}$ ). Consider stochastic flow  $(X_t^{s,x}, t \in [s, T])$

$$X_t^{s,x} = x + \int_s^t B(\theta, X_\theta^{s,x}) d\theta + \int_s^t \Sigma(\theta, X_\theta^{s,x}) dW_\theta, \quad t \in [s, T].$$

We consider

$$v_y(s, x) := \mathbb{E}[G(y, X_t^{s,x})], \quad y \in \mathbb{R}^d \text{ and } (s, x) \in [0, t] \times \mathbb{R}^d,$$

and associated (family of) PDEs reads

$$\begin{cases} \frac{\partial v_y}{\partial s}(s, x) + \frac{1}{2} \sum_{i,j=1}^d A_{ij}(s, x) \frac{\partial^2 v_y}{\partial x_i \partial x_j}(s, x) + \sum_{j=1}^d B_j(s, x) \frac{\partial v_y}{\partial x_j}(s, x) = 0, & (s, x) \in [0, t] \times \mathbb{R}^d \\ v_y(t, x) = G(y, x), \end{cases}$$

where  $A = \Sigma(s, x)\Sigma(s, x)^T$ . We have

$$\sup_{y \in \mathbb{R}^d} \left| \left| \frac{\partial v_y}{\partial x}(t, x') \right| \right|_\infty \leq L_v \text{ and } \sup_{y \in \mathbb{R}^d} \left| \left| \frac{\partial^2 v_y}{\partial x^2}(t, x') \right| \right|_\infty \leq L_v \quad \forall (t, x') \in [0, T] \times \mathbb{R}^d.$$

# Main Result

## Theorem

Assume regularity of the coefficients and the initial law. Then

$$\begin{aligned} & \sup_{x \in \mathbb{R}^d} \sup_{0 \leq t \leq T} \mathbb{E} \left[ ([\mathcal{M}_{\eta_L(t)}(\mu_Y^M)](G(x, \cdot)) - \mathbb{E}[G(x, X_{\eta_L(t)})])^2 \right] \\ & \leq c \left\{ h_L^2 + \sum_{m=1}^M \frac{c^{M-m}}{(M-m)!} \cdot \sum_{\ell=0}^L \frac{h_\ell}{N_{m,\ell}} + \frac{c^M}{M!} \right\}, \end{aligned}$$

- We can reduce computational cost of approximating expectation by the order of magnitude.

## Glimpse into the analysis

SDEs with random coefficients

$$dU_t = \bar{B}(U_t, \mathcal{V}_{\eta_L(t)})dt + \bar{\Sigma}(U_t, \mathcal{V}_{\eta_L(t)})dW_t, \quad \text{Law}(U_0) = \text{Law}(X_0),$$

where  $\mathcal{V}_{\eta_L(t)} \in \mathcal{P}(\mathbb{R}^d)$  is random measure. Euler scheme is given

$$dZ_t^\ell = \bar{B}(Z_{\eta(t)}^\ell, \mathcal{V}_{\eta(t)})dt + \bar{\Sigma}(Z_{\eta(t)}^\ell, \mathcal{V}_{\eta(t)})dW_t, \quad \text{Law}(Z_0^\ell) = \text{Law}(X_0),$$

- (Conditional) Independence: The random measure  $\mathcal{V}$  is independent of  $W$  and  $Z_0$ .
- Integrability: Let  $G$  be Lipschitz continuous, then

$$\sup_{x \in \mathbb{R}^d} \sup_{0 \leq s \leq T} \mathbb{E}\left[\int_{\mathbb{R}^d} |G(x, y)|^p \mathcal{V}_s(dy)\right] \leq c \quad \forall p \geq 1.$$

- Regularity: Let  $G$  be Lipschitz continuous, then

$$\sup_{x \in \mathbb{R}^d} \sup_{0 \leq s \leq t \leq T} \mathbb{E}\left[\left|\int_{\mathbb{R}^d} G(x, y)(\mathcal{V}_t - \mathcal{V}_s)(dy)\right|^2\right] \leq c(t-s).$$

## Glimpse into the analysis

### Theorem

Assume regularity of the coefficients and the initial law. There exists a constant  $c$  such that

$$\sup_{x \in \mathbb{R}^d} \sup_{0 \leq t \leq T} \mathbb{E} |\mathcal{V}_{\eta_L(t)}^{Z^\ell}(G(x, \cdot)) - \mathcal{V}_{\eta_L(t)}^{Z^{\ell-1}}(G(x, \cdot))|^2 \leq ch_\ell.$$

### Theorem

Assume regularity of the coefficients and the initial law. Then there exists a constant  $c$  such that  $\forall t \in [0, T]$ ,

$$\begin{aligned} & \sup_{x \in \mathbb{R}^d} \sup_{0 \leq s \leq t} |\mathbb{E}[G(x, Z_s^L)] - \mathbb{E}[G(x, X_s)]| \\ & \leq c \left( h_L + \int_0^t \sup_{x \in \mathbb{R}^d} \mathbb{E} |\bar{B}(x, \mathcal{V}_{\eta_L(s)}) - \mathbb{E}[b(x, X_{\eta_L(s)})]| ds \right. \\ & \quad \left. + \int_0^t \sup_{x \in \mathbb{R}^d} \mathbb{E} \|\bar{\Sigma}(x, \mathcal{V}_{\eta_L(s)}) - \mathbb{E}[\sigma(x, X_{\eta_L(s)})]\| ds \right). \end{aligned}$$

## New idea

The idea is to use previous Iteration as Control Variates i.e

$$\mathbb{E}[G(x, X^M)] = \mathbb{E}[G(x, X^0)] + \sum_{m=1}^M \mathbb{E}[G(x, X^m) - G(x, X^{m-1})]$$

- Coupling is obtained by simulating both iterations with the same BM
- Inner expected value can be further expanded into a telescopic sum
- First numerical examples are very promising.

## Numerical Experiment: Non-interacting kernel

The target stochastic differential equation is

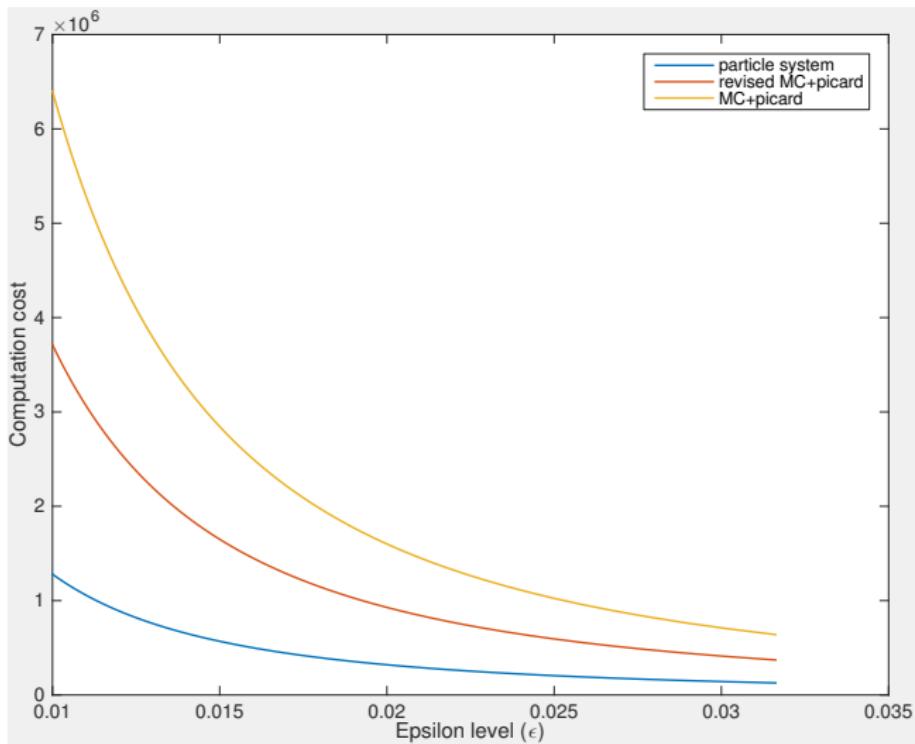
$$dX_t = \sin(X_t - \mathbb{E}[X_t])dt + \sigma dW_t, \quad X_0 = 0.$$

The testing payoff function is

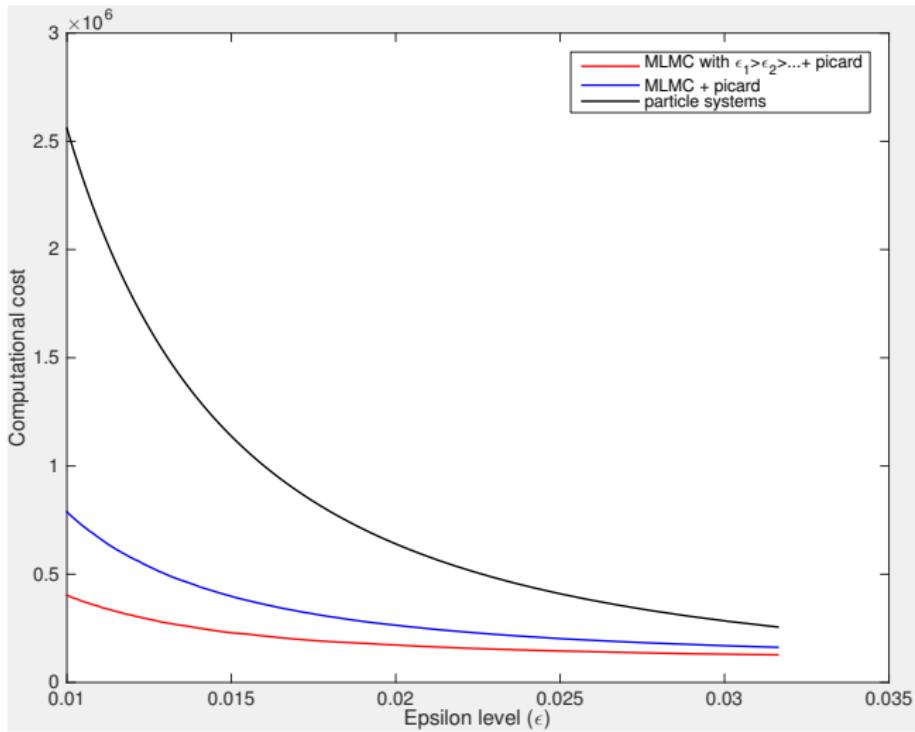
$$P(x) = \max(x - K, 0),$$

where strike  $K$  is set to 0.1.

## Non-interacting kernel



## Non-interacting case



## Interacting kernel

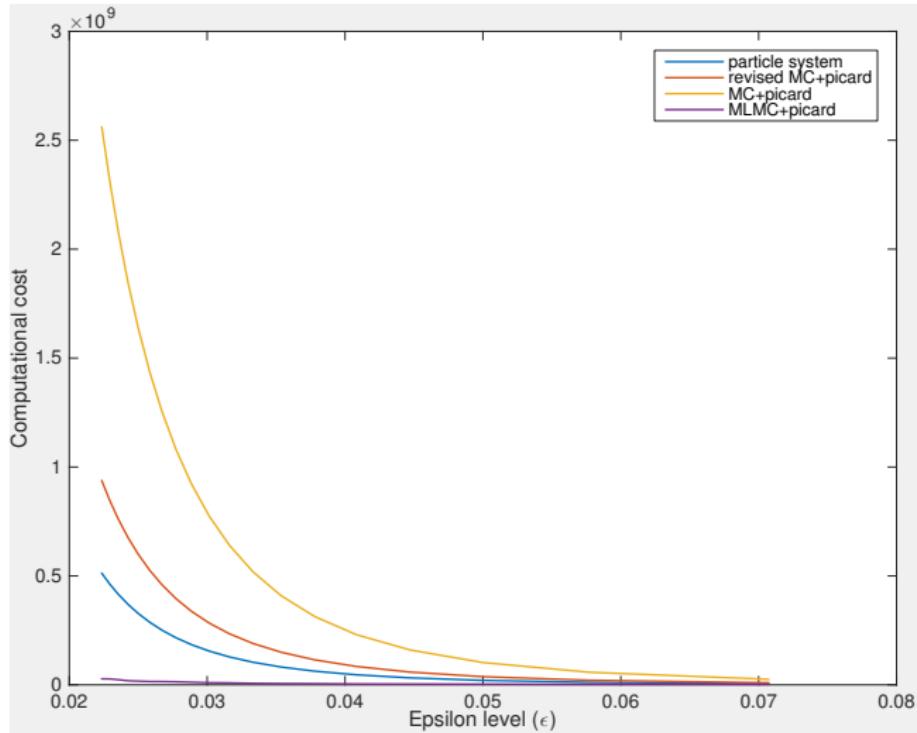
The target stochastic differential equation is

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

where  $Y_t$  is an independent copy of  $X_t$ . The payoff

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

## Interacting kernel



## Interacting kernel

