

NEW PARTICLE REPRESENTATIONS FOR ERGODIC MCKEAN-VLASOV SDES

HOUSSAM ALRACHID¹, MIREILLE BOSSY², CRISTIANO RICCI³ AND LUKASZ SZPRUCH^{4, 5}

Abstract. The aim of this paper is to introduce several new particle representations for *ergodic* McKean-Vlasov SDEs. We construct new algorithms by leveraging recent progress in weak convergence analysis of interacting particle system. We present detailed analysis of errors and associated costs of various estimators, highlighting key differences between long-time simulations of linear (classical SDEs) versus non-linear (McKean-Vlasov SDEs) process.

1. INTRODUCTION

Diffusion processes are at the core of many algorithms used in statistics to sample from, typically high-dimensional, distributions. These algorithms are often based on some variant of Langevin stochastic dynamics [LRS10]. Given a probability measure π (possibly known up to a normalising constant), the key idea is to construct a diffusion process which admits π as its invariant measure. Then one can run long-time simulations of that diffusion to obtain samples from π . This idea has been extensively studied in the context of classical SDEs [Tal90, Tal02, LP⁺02, LP⁺10, MST10, PP⁺12, Dal17, DM⁺17].

Recently new promising classes of algorithms based on the theory of gradient flows takes the form of McKean-Vlasov ODEs or SDEs [SLMD18, Ber18, Liu17]. To turn them into practical algorithms one needs to approximate them with systems of interacting diffusions also called stochastic interacting particle systems. The key challenge is that, typically, with the increase of the dimension of the problem one needs to consider large number N of particles. Because, for most models, the cost of particle samples grows as N^2 (as each particle interacts with others), the computational cost for simulating the particle systems is prohibitive. Another complication is that when using a single ensemble of particles the statistical error due to the approximation of the measure creates biased dynamics. Put differently bias is a non-linear function of the statistical error. In addition particles are not independent. All of that renders classical variance reduction techniques not directly applicable and consequently simulations of particle systems challenging. The high computational cost is even more pronounced when the aim is to simulate particle systems over a long-time horizon. This should not come a surprise as particle systems give rise to probabilistic numerical methods for highly non-linear PDEs e.g Burgers or Navier-Stokes PDEs.

In this work we leverage recent progress in weak convergence analysis of interacting diffusions [JFC, CCD, Kol10, CD18]. With this new insight we propose several new algorithms and analyse their errors and costs. The emphasis of the work is on algorithmic side and we gloss over some theoretical bounds that will require

¹ IDP Laboratory, Orleans University, Orleans, France.

² Université Côte dAzur, Inria, France.

³ University of Florence, Italy

⁴ School of Mathematics, University of Edinburgh – Maxwell Institute for Mathematical Sciences, Edinburgh

⁵ The Alan Turing Institute, London.

further research in future. As such we see this work as beacon that helps to identify the most promising research directions in the area of simulations of the ergodic particle systems.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a filtered probability space endowed with an \mathbb{R}^k -valued Wiener process $w = (w_t)_{t \geq 0}$. Let $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times k}$. We consider, for $t \geq 0$, the McKean-Vlasov SDEs (McKV-SDE)

$$\begin{cases} x_t = x_0 + \int_0^t b(x_s, \mu_s) ds + \int_0^t \sigma(x_s, \mu_s) dw_s \\ \mu_t \text{ is the law of } x_t, \end{cases} \quad (1)$$

where x_0 is distributed according to a given \mathbb{R}^d -measure μ_0 . The nonlinearity in the McKean-Vlasov SDEs (1) appears through the dependency of its coefficients on the law of the process. Existence of the unique solution to (1) has been established under various conditions on (b, σ) . See [Szn91, Mel96] classical results on that topic that mainly cover the case of finite time interval. For the infinite time horizon we refer to [Ver06, Hs18].

Furthermore, [BGT08] gives conditions for the existence and uniqueness of the invariant measure π for the equation (1). We refer to [EGZ16] for more complete theory. In particular [EGZ16] gives fairly general conditions that guarantee that the convergence to the invariant measure in the L^2 -Wasserstein distance is exponentially fast for some $\lambda > 0$, i.e

$$W_2(\mu_t, \pi) \leq \exp(-\lambda t) W_2(\mu_0, \pi). \quad (2)$$

One can also control the bias of ergodic averages

$$\left| \mathbb{E} \left[\frac{1}{t} \int_0^t f(x_s) ds - \int_{\mathbb{R}^d} f(x) \pi(dx) \right] \right| \lesssim t^{-1}. \quad (3)$$

Consider the following system of N particles $(x^{1,N}, x^{2,N}, \dots, x^{N,N})$ defined as

$$\begin{cases} x_t^{i,N} = x_0^{i,N} + \int_0^t b(x_s^{i,N}, \mu_s^N) ds + \int_0^t \sigma(x_s^{i,N}, \mu_s^N) dw_s^i \\ \mu_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{x_t^{i,N}}, \end{cases} \quad (4)$$

where $(w^i)_i$ are independent k -dimensional Brownian motions, $(x_0^i)_i$ are initial i.i.d. variables independent of $(w^i)_i$. The measure valued random variable μ_t^N is an empirical measure of the system at time t . For the purpose of computer simulations one needs to introduce time discretisation to simulate (4). We will do that in the forthcoming section. Under classical Lipschitz continuity conditions the law, seen as element of $\mathcal{P}([0, T], \mathbb{R}^d)$, of every fixed subsystem of k particles from $(x^{i,N})$ converges, when N tends to infinity, to the law $\mu^{\otimes k}$. This property is called propagation of chaos phenomena. Under strong convexity of the drift, a time-uniform version of propagation of chaos has been established in [BGT08]. In a recent work, [DEGZ18] it has been shown that in general only convexity at infinity is needed.

Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$. The objective of this work is to derive, analyse and numerically investigate, several novel particle representations that will allow to approximate:

$$\int_{\mathbb{R}^d} f(x) \pi(dx). \quad (5)$$

To motivate our work, let's temporarily assume that (b, σ) do not depend on measure, i.e we are dealing with a classical SDEs. Then a typical strategy in obtaining an approximation to (5) would be to set a finite time t

and take N i.i.d. trajectories to compute $\frac{1}{t} \int_0^t \frac{1}{N} \sum_{i=1}^N f(x_s^i) ds$ ($N = 1$ corresponds to ergodic estimator). The error of this estimator can be decomposed as follows

$$\begin{aligned} & \mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x) \pi(dx) - \frac{1}{t} \int_0^t \frac{1}{N} \sum_{i=1}^N f(x_s^i) ds \right)^2 \right]^{1/2} \\ & \leq \mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x) (\pi(dx) - \mu_t(dx)) \right)^2 \right]^{1/2} + \mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x) \mu_t(dx) - \frac{1}{t} \int_0^t \frac{1}{N} \sum_{i=1}^N f(x_s^i) ds \right)^2 \right]^{1/2}. \end{aligned}$$

The first term in the right hand side is the (weak) error of approximating the invariant measure which decays to zero as $\exp(-\lambda t)$ due to (2). The second term is CLT type result and can be shown to decay to zero as $(N \cdot t)^{-1/2}$, see e.g [KV86, CCG12]. We see that both t and N have the same impact on the variance. In the case of SDEs the cost of simulations is linear in t and N and hence one may be indifferent whether to simulate one long trajectory (ergodic estimator) and many shorter ones (space average). Of course if one uses parallel computer architecture, taking more samples is much more efficient.

The situation of McKean-Vlasov SDE (1) is dramatically different. The cost of simulating interacting particles (4) is N^2 while it still increases linearly with time. As we will show it is possible to construct estimator that has one-order of magnitude lower cost while maintains the same accuracy. Furthermore, we will investigate *ensemble* version of interacting diffusions where we generate M independent systems of particles with N particles in each system (ensemble). More precisely we define

$$\begin{cases} x_t^{(i,N),(j,M)} = x_0^{(i,N),(j,M)} + \int_0^t b(x_s^{(i,N),(j,M)}, \mu_s^{N,j}) ds + \int_0^t \sigma(x_s^{(i,N),(j,M)}, \mu_s^{N,j}) dw_s^{i,j}, \\ \mu_t^{N,j} = \frac{1}{N} \sum_i \delta_{x_t^{(i,N),(j,M)}}, \end{cases} \quad (6)$$

where $(w^{i,j}, i, j)$ are independent Brownian motions. That way particles within each ensemble j^* driven by $(w^{i,j^*})_{i,j^*}$ are interacting and are not independent. The particle systems j and j' , $j \neq j'$, driven by $(w^{i,j})_{i,j}$ and $(w^{i,j'})_{i,j}$ respectively, are independent. This idea for finite time simulations has been proposed in [HAT16]. Another approach that we investigate are self-interacting diffusion

$$z_t = x_0 + \int_0^t \left(\frac{1}{s} \int_0^s b(z_s, z_r) dr \right) ds + \int_0^t \left(\frac{1}{s} \int_0^s \sigma(z_s, z_r) dr \right) dw_s \quad (7)$$

We expect that the law of z_t approximates the law of x_t for large t due to ergodic property (2,3). This gives an alternative to the particle system. We will show that the structure of the equation seems to play a crucial role in this set up.

The rest of the paper is organised as follows. In Section 2, we recall some classical methods to approximate (5) and give error estimations and computational costs of the associated algorithms. In Section 3 we study several variants of algorithms for ergodic interacting particle systems. In Section 4, we present the ensemble algorithm with ergodic average particle system. We end this paper with a general conclusion and some perspectives in Section 5.

2. SETUP

2.1. Algorithms

As in this work we are interested in designing implementable algorithms for (1), we need to introduce time discretisation for (4). Let us define $t_k^n := \frac{k}{n}$, $k = 0, 1, \dots$ and $\kappa_n(t) = t_k^n$ for $t \in [t_k^n, t_{k+1}^n)$. We introduce (continuous time) Euler approximations for each $i \leq N$, $(y_t^{i,N}, t \geq 0)$, $n \in \mathbb{N}$,

$$\begin{cases} y_t^{i,N} = y_0^{i,N} + \int_0^t b(y_{\kappa_n(s)}^{i,N}, \bar{\mu}_{\kappa_n(s)}^N) ds + \int_0^t \sigma(y_{\kappa_n(s)}^{i,N}, \bar{\mu}_{\kappa_n(s)}^N) dw_s^i, \\ \bar{\mu}_{\kappa_n(t)}^N = \frac{1}{N} \sum_{i=1}^N \delta_{y_{\kappa_n(t)}^{i,N}}. \end{cases} \quad (8)$$

One may approximate (5) by ergodic average estimator with fixed t ,

$$\frac{1}{t} \int_0^t f(y_{\kappa_n(s)}^{1,N}) ds. \quad (\text{EA})$$

For error analysis one needs to choose N (number of particles) and n (number of timesteps) to control the bias of the approximation of (1). While (EA) estimator is a reasonable choice for computing approximation of (5) for the invariant measures induced by classical SDEs as we already argued, the case of McKean-Vlasov SDEs approximated with a particle system, (EA) estimator does not seem to be the best choice. This is because when using particle system (4), one computes N particles and therefore calculating ergodic average along one trajectory is not efficient. See sections 3.1, 3.2 for more details. Hence, improvement can be obtained by computing, averaged ergodic average estimator

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{t'} \int_0^{t'} f(y_{\kappa_n(s)}^{i,N}) ds \right). \quad (\text{AEA})$$

Of course one expects that t' in (AEA) to be smaller than t in (EA) for fixed accuracy. Alternative strategy for approximating (5) is to resort to the standard Monte Carlo estimator where the average is taken only "over the space". More precisely we compute Monte Carlo average

$$\frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}). \quad (\text{MCA})$$

Of course, the above estimator is less efficient than (AEA), as we explain in the coming section. The only reason we study it here is to warn practitioners that if not careful with setting up particle estimators the cost might be huge.

Computational Cost.

By $\mathcal{A}(\eta)$ we denote an algorithm that outputs the approximation for the quantity (5), where η denotes the set of all the parameters we need to choose to implement it. To be able to compare the algorithms we need to fix a measure of error. For simplicity, we restore to the mean-square-error:

$$mse(\mathcal{A}(\eta)) := \mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x) \pi(dx) - \mathcal{A}(\eta) \right)^2 \right]^{1/2}. \quad (9)$$

With the measure of error of a given estimator set up, the second equally important quantity is the computation cost of algorithm \mathcal{A} , denoted by $cost(\mathcal{A})$. With both quantities in place, we can wonder about the optimal choice

of parameters achieving a prescribed tolerance. More precisely, for fixed error tolerance $\epsilon > 0$, we need to solve the following optimisation problem:

$$\begin{cases} mse(\mathcal{A}(\eta)) < \epsilon, \\ \min_{\eta} cost(\mathcal{A}(\eta)). \end{cases}$$

2.2. Assumptions

In this section we list all the assumptions needed for our considerations. The only assumption that has not been yet established in the literature is uniform in time particle error (HW) estimation below. To the best of authors knowledge only finite time weak particle error has been studied. It is clear how to extend the weak convergence to be uniform in time but this would require lengthy introduction of heavy machinery of PDEs on measure spaces. This falls outside this paper. All other assumptions are established in literature under various level of generality and we point out reader to the corresponding papers.

We label by x^i a McKean-Vlasov SDE driven by i th Brownian motion, that is

$$x_t^i = x_0^i + \int_0^t b(x_s^i, \mu_s) ds + \int_0^t \sigma(x_s^i, \mu_s) dw_s^i. \quad (10)$$

Assumption 2.1. *Convergence rate to ergodic measure: there exists $\lambda > 0$ such that*

$$W_2(\mu_t, \pi) \lesssim \exp(-\lambda t) W_2(\mu_0, \pi). \quad (HE)$$

As we already mentioned this has been proved in [BGT08] and [EGZ16] under fairly general conditions.

Assumption 2.2. *Convergence rate of ergodic average*

$$\left(\mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x) \mu_t(dx) - \frac{1}{t} \int_0^t f(x_s) ds \right)^2 \right] \right)^{1/2} \lesssim \frac{(\sup_{s \in [0, t]} \text{Var}[x_s])^{1/2}}{\sqrt{t}}. \quad (HEA)$$

This is classical CLT result. See [CCG12].

Assumption 2.3. *Uniform in time weak convergence of the particle system: for sufficiently smooth f ,*

$$\sup_{t \geq 0} |\mathbb{E}f(x_t^1) - \mathbb{E}f(x_t^{1, N})| \lesssim \frac{1}{N}. \quad (HW)$$

This type of bound is new in the literature. We refer a reader to [JFC, CCD, Kol10, CD18] for more details.

Assumption 2.4. *Uniform in time strong propagation of chaos*

$$\sup_{t \geq 0} (\mathbb{E}|x_t^1 - x_t^{1, N}|^2)^{1/2} \lesssim \frac{1}{\sqrt{N}}. \quad (HS)$$

See [DEGZ18] for details.

Assumption 2.5. *Uniform in time weak discretisation error: for sufficiently smooth f ,*

$$\sup_{t \geq 0} |\mathbb{E}f(x_t^1) - \mathbb{E}f(y_t^{1, N})| \lesssim \frac{1}{n}. \quad (HDW)$$

Uniform in time strong discretisation error

$$\sup_{t \geq 0} (\mathbb{E}|x_t^1 - y_t^{1, N}|^2)^{1/2} \lesssim \frac{1}{\sqrt{n}}. \quad (HDS)$$

One can refer to [BGT08], where such results are proved.

3. ALGORITHMS FOR ERGODIC INTERACTING PARTICLE SYSTEMS

3.1. Monte Carlo Average

For the Monte Carlo Average estimator we introduce the following notation $\mathcal{A}^{MCA}(t, n, N)$. The aim is to find the optimal allocation of the parameters (t, n, N) for fixed mean-square-error. We have

$$\begin{aligned} mse(\mathcal{A}^{MCA}(t, n, N)) &= \mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x) \pi(dx) - \frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}) \right)^2 \right]^{1/2} \\ &\lesssim \left| \int_{\mathbb{R}^d} f(x) (\pi(dx) - \mu_t(dx)) \right| + \left| \mathbb{E}[f(x_t)] - \mathbb{E}[f(x_t^{1,N})] \right| \\ &\quad + \left| \mathbb{E}[f(x_t^{1,N})] - \mathbb{E}[f(y_t^{1,N})] \right| + \mathbb{E} \left[\left(\mathbb{E}[f(y_t^{1,N})] - \frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}) \right)^2 \right]^{1/2}. \end{aligned}$$

The four error terms are in order: bias (due to finite time simulation); weak particle approximation error; weak time discretisation error; variance/propagation of chaos. The first three error terms can be estimated directly from the Assumptions in Section 2.2. The last variance error term requires extra comment, as the fact that particles are not i.i.d does not allow to use classical central limit theorem (CLT). Indeed

$$\begin{aligned} \mathbb{E}[(\mathbb{E}[f(y_t^{1,N})] - \frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}))^2]^{1/2} &= \mathbb{E}[(\mathbb{E}[\frac{1}{N} \sum_{i=1}^N f(y_t^{i,N})] - \frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}))^2]^{1/2} \\ &\leq \mathbb{E}[(\frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}))^2]^{1/2}. \end{aligned}$$

Next, define $(\tilde{y}^i)_i$ as the solution of the continuous Euler scheme:

$$\begin{cases} \tilde{y}_t^i = \tilde{y}_0^i + \int_0^t b(\tilde{y}_{\kappa_n(s)}^i, \mathcal{L}(\tilde{y}_{\kappa_n(s)})) ds + \int_0^t \sigma(\tilde{y}_{\kappa_n(s)}^i, \mathcal{L}(\tilde{y}_{\kappa_n(s)})) dw_s^i, \\ \mathcal{L}(\tilde{y}_t) = Law(\tilde{y}_t). \end{cases}$$

It is an easy exercise to show that strong propagation of chaos (HS) can be established on the level of Euler discretisation. This together with Cauchy-Schwarz inequality gives

$$\begin{aligned} &\mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}) - f(\tilde{y}_t^i) \right)^2 \right] \\ &= \mathbb{E} \left[\frac{1}{N^2} \sum_{i=1}^N (f(y_t^{i,N}) - f(\tilde{y}_t^i))^2 + \frac{1}{N^2} \sum_{i < j} (f(y_t^{i,N}) - f(\tilde{y}_t^i)) (f(y_t^{j,N}) - f(\tilde{y}_t^j)) \right] \\ &\leq \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left[(f(y_t^{i,N}) - f(\tilde{y}_t^i))^2 \right] \\ &\quad + \frac{1}{N^2} \sum_{i < j} (\mathbb{E}[(f(y_t^{i,N}) - f(\tilde{y}_t^{i,N}))^2])^{1/2} (\mathbb{E}[(f(y_t^j) - f(\tilde{y}_t^j))^2])^{1/2} \lesssim \frac{1}{N}. \end{aligned} \tag{11}$$

This, and the fact that $(\tilde{y}_t^i)_i$ are i.i.d. allows to conclude that

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}) \right)^2 \right] &= \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N f(\tilde{y}_t^i) + (f(y_t^{i,N}) - f(\tilde{y}_t^i)) \right)^2 \right] \\ &\lesssim \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N f(\tilde{y}_t^i) \right)^2 \right] + \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N f(y_t^{i,N}) - f(\tilde{y}_t^i) \right)^2 \right] \lesssim \frac{1}{N}. \end{aligned}$$

From here and Assumptions in Section 2.2 we have

$$mse(\mathcal{A}^{MCA}(t, n, N)) \lesssim e^{-\lambda t} + \frac{1}{N} + \frac{1}{n} + \frac{1}{\sqrt{N}}.$$

Notice that because of the term $1/\sqrt{N}$, it is not clear how we can take advantage of the assumption (HW). Fix $\epsilon > 0$ and set $mse(\mathcal{A}^{MCA}(t, n, N)) \lesssim \epsilon$. This leads to the following choice of the parameters $t \approx \lambda^{-1} \log(\epsilon^{-1})$, $N \approx \epsilon^{-2}$, $n \approx \epsilon^{-1}$. As the cost of simulating particle system at every step of the Euler scheme is N^2 we have

$$cost(\mathcal{A}^{MCA}(t, n, N)) = tnN^2 \approx \log(\epsilon^{-1})\epsilon^{-5}.$$

This should be compared to $tnN = \log(\epsilon^{-1})\epsilon^{-3}$ for the simulation of standard SDEs.

3.2. Averaged Ergodic Average

As before, we denote by $\mathcal{A}^{AEA}(t, n, N)$ the averaged ergodic average estimator in (AEA). We have

$$\begin{aligned} mse(\mathcal{A}^{AEA}(t, n, N)) &= \mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x)\pi(dx) - \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{t} \int_0^t f(y_{\kappa_n(s)}^{i,N}) ds \right) \right)^2 \right]^{1/2} \\ &\lesssim \left| \int_{\mathbb{R}^d} f(x)(\pi(dx) - \mu_t(dx)) \right| + \left| \mathbb{E}[f(x_t)] - \mathbb{E}[f(x_t^{1,N})] \right| \\ &\quad + \left| \mathbb{E}[f(x_t^{1,N})] - \mathbb{E}[f(y_t^{1,N})] \right| + \mathbb{E} \left[\left(\mathbb{E}[f(y_t^{1,N})] - \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{t} \int_0^t f(y_{\kappa_n(s)}^{i,N}) ds \right) \right)^2 \right]^{1/2}. \end{aligned}$$

To estimate the variance term we note that

$$\mathbb{E}[f(y_t^{1,N})] - \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{t} \int_0^t f(y_{\kappa_n(s)}^{i,N}) ds \right) = \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N (y_t^{i,N}) \right] - \left(\frac{1}{t} \int_0^t \frac{1}{N} \sum_{i=1}^N f(y_{\kappa_n(s)}^{i,N}) ds \right).$$

Hence to estimate the variance we use (HEA) combined with the computation (11). Therefore, by the Assumptions in Section 2.2, we have

$$mse(\mathcal{A}^{AEA}(t, n, N)) \lesssim e^{-\lambda t} + \frac{1}{N} + \frac{1}{n} + \frac{1}{\sqrt{tN}}.$$

We notice that comparing to the (MCA) case, the last term is multiplied by $1/\sqrt{t}$. The (asymptotic) cost of the algorithm is the same as before. Again we fix ϵ . The following choice of the parameters ensures that

$mse(\mathcal{A}^{AEA}(t, n, N)) \lesssim \epsilon^2$ is $t \approx \epsilon^{-1}$, $N \approx \epsilon^{-1}$, $n \approx \epsilon^{-1}$. The cost consists of two parts: the cost of simulating particle system and the cost of computing averaged ergodic estimator. We have

$$cost(\mathcal{A}^{AEA}(t, n, N)) = tnN^2 + tN \approx \epsilon^{-4}.$$

Which is an order of magnitude lower than for Monte Carlo average!

Notice that similar computation for ergodic average estimator gives $mse(\mathcal{A}^{EA}(t, n, N)) \lesssim e^{-\lambda t} + \frac{1}{N} + \frac{1}{n} + \frac{1}{\sqrt{t}}$, leading to the same cost than Monte Carlo average.

3.3. ensemble AEA

For the ensemble version of the algorithm we generate M independent systems of particles with N particles in each system. More precisely we define,

$$\begin{cases} x_t^{(i,N),(j,M)} = x_0^{(i,N),(j,M)} + \int_0^t b(x_s^{(i,N),(j,M)}, \mu_s^{N,j}) ds + \int_0^t \sigma(x_s^{(i,N),(j,M)}, \mu_s^{N,j}) dw_s^{i,j}, \\ \mu_t^{N,j} = \frac{1}{N} \sum_i \delta_{x_t^{(i,N),(j,M)}}, \end{cases} \quad (12)$$

where $(w^{i,j}, i, j)$ are independent Brownian motions. That way particle within each cloud j^* driven by $(w^{i,j^*})_{i,j^*}$ are interacting and are not independent. The particle systems j and j' , $j \neq j'$, driven by $(w^{i,j})_{i,j}$ and $(w^{i,j'})_{i,j}$ respectively, are independent. This idea has been proposed in [HAT16] for the finite time simulations.

We consider ensemble version of AEA.

$$\frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{t} \int_0^t f(y_{\kappa_n(s)}^{(i,N),(j,M)}) ds \right). \quad (\text{C-AEA})$$

In fact, all algorithms that we study can have their ensemble versions. By denoting \mathcal{A}^{C-AEA} the new method

$$\begin{aligned} mse(\mathcal{A}^{C-AEA}(t, n, N, M)) &= \mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x) \pi(dx) - \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{t} \int_0^t f(y_{\kappa_n(s)}^{(i,N),(j,N)}) ds \right) \right)^2 \right]^{1/2} \\ &\lesssim \left| \int_{\mathbb{R}^d} f(x) (\pi(dx) - \mu_t(dx)) \right| + \left| \mathbb{E}[f(x_t)] - \mathbb{E}[f(x_t^{1,N})] \right| \\ &\quad + \left| \mathbb{E}[f(x_t^{1,N})] - \mathbb{E}[f(y_t^{1,N})] \right| + \mathbb{E} \left[\left(\mathbb{E}[f(y_t^{1,N})] - \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{t} \int_0^t f(y_{\kappa_n(s)}^{(i,N),(j,M)}) ds \right) \right)^2 \right]^{1/2}. \end{aligned}$$

The Assumptions in Section 2.2 yield

$$mse(\mathcal{A}^{C-AEA}(t, n, N, M)) \lesssim e^{-\lambda t} + 1/N + 1/n + \frac{1}{\sqrt{tNM}}.$$

We notice that comparing to the previous case the last term is multiplied by $1/\sqrt{M}$. Crucially, the cost of the algorithm grows linearly in M . As the cost grows as N^2 , we are better off taking $M \approx N$ to balance the

error in the last term (instead of taking $M = 1$ and N^2). To make it precise we fix ϵ . The following choice of the parameters ensures that $mse(\mathcal{A}^{C-AEA}(t, n, N)) \lesssim \epsilon$.

$$t \approx \lambda^{-1} \log(\epsilon^{-1}), \quad N \approx \epsilon^{-1}, \quad n \approx \epsilon^{-1}, \quad M = (\lambda^{-1} \log(\epsilon^{-1}))^{-1} \epsilon^{-1}.$$

The cost of simulating particles and computing the estimator is

$$cost(\mathcal{A}^{C-AEA}(t, n, N)) = tnN^2M + tNM \approx \epsilon^{-4}.$$

This is the same as for averaged ergodic estimator. However the above computations do not take under consideration the fact that ensemble algorithms can take full advantage from the parallel computer architecture and therefore will be superior in practice.

4. ALGORITHMS FOR SEF-INTERACTING PARTICLE SYSTEMS

In this section, we present the key ideas improvement in the definition of more efficient algorithm. From the decomposition of the mean square error, we see that different algorithms that we considered only affected the “variance” of the final estimator. Therefore, to improve the efficiency of the algorithm we need to either modified particle system itself or consider different simulation strategies such as Multilevel-Monte Carlo. Here we focus on the former.

Significance of the ergodic theorem is that one can approximate the integral (5) by simulating only one path of the process (1) rather than the whole particle system. From now on, we will keep the structural assumptions on the coefficients of (1), namely

$$\begin{cases} x_t = x_0 + \int_0^t b(x_s, \mathcal{L}(x_s)) ds + \int_0^t \sigma(x_s, \mathcal{L}(x_s)) dw_s \\ \mathcal{L}(x_t) \text{ is the law of } x_t. \end{cases}$$

One may consider

$$z_t = x_0 + \int_0^t \left(\frac{1}{s} \int_0^s b(z_r, z_r) dr \right) ds + \int_0^t \left(\frac{1}{s} \int_0^s \sigma(z_r, z_r) dr \right) dw_s. \quad (13)$$

We expect that the law of z_t approximates the law of x_t for large t due to ergodic property (2)

Processes of the form (13) are known in literature as self interacting diffusions. We refer to [KK12] where the convergence to the invariant measure has been established. Notice that there is no need for the particle system any more as one could simply simulate one path of the process to calculate ergodic integral (5).

However, motivated by computations in the previous section where mixed ergodic/Monte Carlo average we introduce the corresponding mean self-integrated SDE

$$z_t = x_0 + \int_0^t \left(\frac{1}{s} \int_0^s b(z_r, \mathcal{L}(z_r)) dr \right) ds + \int_0^t \left(\frac{1}{s} \int_0^s \mathbb{E} \sigma(z_r, \mathcal{L}(z_r)) dr \right) dw_s \quad (14)$$

and its independent copies (z^i) driven by Brownian motion (w^i). Note that “one-particle” approximation of (14) is precisely a self-interacting diffusion.

To gain better insight, into the idea of using self-interacting diffusions to approximate McKean-Vlasov SDEs we consider a simple example first.

Example 4.1. Consider a simple scalar McKean-Vlasov SDE x , together with its mean self-integrated version z ,

$$\begin{aligned} x_t &= x_0 - \int_0^t \alpha x_s ds + \int_0^t \beta \mathbb{E}[x_s] ds + w_t, \\ z_t &= x_0 - \int_0^t \alpha z_s ds + \int_0^t \frac{1}{s} \left(\int_0^s \beta \mathbb{E}[z_\theta] d\theta \right) ds + w_t. \end{aligned}$$

We stress out that dissipativity comes from the part of the drift that do not depend on measure. We assume $\alpha > \beta$. To estimate the convergence rate to the invariant measure we analyse the evolution of the difference of two solutions to the above x SDE initiated at L^2 random variables ξ_1 and ξ_2 . With

$$\mathbb{E}[(x_t^{\xi_1} - x_t^{\xi_2})] = e^{-(\alpha-\beta)t} \mathbb{E}[(\xi_1 - \xi_2)], \quad (15)$$

we have

$$\begin{aligned} e^{2\alpha t} \mathbb{E}[(x_t^{\xi_1} - x_t^{\xi_2})^2] &= \mathbb{E}[(\xi_1 - \xi_2)^2] + 2\beta \int_0^t e^{2\alpha s} (\mathbb{E}[x_s^{\xi_1} - x_s^{\xi_2}])^2 ds \\ &= \mathbb{E}[(\xi_1 - \xi_2)^2] + 2(\mathbb{E}[\xi_1 - \xi_2])^2 \beta \int_0^t e^{2\alpha s} e^{-2(\alpha-\beta)s} ds \\ &= \mathbb{E}[(\xi_1 - \xi_2)^2] + (\mathbb{E}[\xi_1 - \xi_2])^2 (e^{2\beta t} - 1). \end{aligned}$$

Due to properties of W_2 distance and the fact that the above calculation does not depend on a particular choice of random variables ξ_1 and ξ_2 , we have

$$W_2^2(\mathcal{L}(x_t^{\xi_1}), \mathcal{L}(x_t^{\xi_2})) \lesssim e^{-2(\alpha-\beta)t} W_2^2(\mathcal{L}(\xi_1), \mathcal{L}(\xi_2)).$$

Furthermore, in this simple example we can take advantage from the explicit solutions to calculate

$$\mathbb{E}[x_t] - \frac{1}{t} \int_0^t \mathbb{E}[x_s] ds = \mathbb{E}[x_0] \left(e^{-(\alpha-\beta)t} - \frac{1}{t(\alpha-\beta)} (1 - e^{-(\alpha-\beta)t}) \right).$$

Hence if $\log t < (\alpha - \beta)t$ then we have that

$$|\mathbb{E}[x_t] - \frac{1}{t} \int_0^t \mathbb{E}[x_s] ds| \leq c1/t.$$

Let us consider now process z . By integration by part we have that

$$\mathbb{E}[z_t] = \mathbb{E}[z_0] - \int_0^t \alpha \mathbb{E}[z_\theta] d\theta + \beta \int_0^t \log(t/\theta) \mathbb{E}z_\theta d\theta$$

and

$$e^{\alpha t} \mathbb{E}[z_t] = \mathbb{E}[z_0] + \int_0^t e^{\alpha s} \frac{1}{s} \int_0^s \mathbb{E}z_\theta d\theta.$$

Then we observe that if $\mathbb{E}[z_0] > 0$, then $\mathbb{E}[z_t]$ stays non-negative and do not cross 0. We observe also that $t \mapsto \mathbb{E}[z_t]$ is decreasing as for all $t \geq 0$, for all $s \in [te^{-\alpha/\beta}, t]$

$$\mathbb{E}[z_t] - \mathbb{E}[z_s] = \int_s^t (\beta \log(t/\theta) - \alpha) \mathbb{E}z_\theta d\theta \leq 0.$$

In particular

$$\frac{\mathbb{E}[z_t] - \mathbb{E}[z_s]}{t - s} = \frac{1}{t - s} \int_s^t (\beta \log(t/\theta) - \alpha) \mathbb{E}[z_\theta] d\theta \leq \frac{\sup_{u \geq 0} \mathbb{E}[z_u]}{t - s} \int_s^t (\beta \log(t/\theta) - \alpha) d\theta.$$

Taking limit $s \rightarrow t$, we obtain after integration that

$$\mathbb{E}[z_t] \leq \mathbb{E}[z_0] \exp(-(\alpha - \beta)t). \quad (16)$$

The following computation on z is a tentative to evidence the rate's gain in the convergence rate to equilibrium. To this aim, we analyse the evolution of the difference of two solutions z initiated at L^2 random variables ξ_1 and ξ_2 .

Repeating the previous computation for $\mathbb{E}[\xi_1 - \xi_2] \geq 0$, we also obtain

$$0 \leq \mathbb{E}[(z_t^{\xi_1} - z_t^{\xi_2})] \leq e^{-(\alpha - \beta)t} \mathbb{E}[(\xi_1 - \xi_2)],$$

and we use this weak estimation to derive the L^2 -norm bound. We have first that

$$\begin{aligned} e^{2\alpha t} \mathbb{E}[(z_t^{\xi_1} - z_t^{\xi_2})^2] &= \mathbb{E}[(\xi_1 - \xi_2)^2] + 2\beta \int_0^t e^{2\alpha s} \mathbb{E}[z_s^{\xi_1} - z_s^{\xi_2}] \frac{1}{s} \left(\int_0^s \mathbb{E}[z_\theta^{\xi_1} - z_\theta^{\xi_2}] d\theta \right) ds \\ &\leq \mathbb{E}[(\xi_1 - \xi_2)^2] + 2\beta (\mathbb{E}[\xi_1 - \xi_2])^2 \int_0^t e^{2\alpha s} e^{-(\alpha - \beta)s} \frac{1}{(\alpha - \beta)s} (1 - e^{-(\alpha - \beta)s}) ds \\ &\leq \mathbb{E}[(\xi_1 - \xi_2)^2] + 2\beta (\mathbb{E}[\xi_1 - \xi_2])^2 \int_0^t \frac{1}{(\alpha - \beta)s} (e^{(\alpha + \beta)s} - e^{2\beta s}) ds. \end{aligned}$$

As $\alpha > \beta$, it is not difficult to check that for t big enough,

$$0 \leq \int_0^t \frac{1}{s} (e^{(\alpha + \beta)s} - e^{2\beta s}) ds \leq \frac{2}{t} \left(\frac{e^{(\alpha + \beta)t}}{\alpha + \beta} - \frac{e^{2\beta t}}{2\beta} \right) - \frac{2}{t} \left(\frac{1}{\alpha + \beta} - \frac{1}{2\beta} \right).$$

And we obtain the contraction inequality,

$$\mathbb{E}[(z_t^{\xi_1} - z_t^{\xi_2})^2] \leq e^{-2\alpha t} \mathbb{E}[(\xi_1 - \xi_2)^2] + c \frac{e^{-(\alpha - \beta)t}}{t} (\mathbb{E}[\xi_1 - \xi_2])^2,$$

which means that after a time $t \geq t_0$, the convergence in L^2 is exponentially fast with a rate $\alpha \wedge (\alpha + \log(t) - \beta)/2$, that accelerates and becomes with time better than the rate for process x (in $\alpha - \beta$).

We have also to show that x and z have the same equilibrium measure. A sufficient condition is the L^2 -convergence for z toward x in time. To not spend too much time on this example, we make the following assumption from the previous Wasserstein contraction for z and from what we obtain for x , that

$$\left| \mathbb{E}[z_t] - \frac{1}{t} \int_0^t \mathbb{E}[z_s] ds \right| \leq c1/t.$$

We consider now

$$\begin{aligned} e^{2(\alpha - \beta)t} \mathbb{E}(x_t - z_t)^2 &\leq e^{2(\alpha - \beta)t} \mathbb{E}(x_1 - z_1)^2 - 2\beta \int_1^t e^{2(\alpha - \beta)s} \mathbb{E}(x_s - z_s)^2 ds \\ &\quad + 2\beta \int_1^t e^{2(\alpha - \beta)s} \mathbb{E}(x_s - z_s) \left(\mathbb{E}[x_s] - \frac{1}{s} \int_0^s \mathbb{E}[z_\theta] d\theta \right) ds. \end{aligned}$$

But since $|\mathbb{E}(x_s - z_s)| \leq |\mathbb{E}(x_s)| + |\mathbb{E}(z_s)| \leq ce^{-(\alpha-\beta)s}$ from (15) and (16),

$$\begin{aligned} & \int_1^t e^{2(\alpha-\beta)s} \mathbb{E}(x_s - z_s) \left(\mathbb{E}[x_s] - \frac{1}{s} \int_0^s \mathbb{E}[z_\theta] d\theta \right) ds \\ & \leq \int_1^t e^{2(\alpha-\beta)s} \mathbb{E}(x_s - z_s)^2 + \int_1^t e^{2(\alpha-\beta)s} |\mathbb{E}(x_s - z_s)| \frac{c}{s} ds \leq \int_1^t e^{2(\alpha-\beta)s} \mathbb{E}(x_s - z_s)^2 + \int_1^t e^{(\alpha-\beta)s} \frac{c'}{s} ds \end{aligned}$$

and we have obtained

$$\begin{aligned} e^{2(\alpha-\beta)t} \mathbb{E}(x_t - z_t)^2 & \leq e^{2(\alpha-\beta)} \mathbb{E}(x_1 - z_1)^2 + 2\beta \int_1^t e^{(\alpha-\beta)s} \frac{c'}{s} ds \\ \mathbb{E}(x_t - z_t)^2 & \leq e^{-2(\alpha-\beta)(t-1)} \mathbb{E}(x_1 - z_1)^2 + Ce^{-(\alpha-\beta)t} \log(t). \end{aligned}$$

The convergence is then ensured. Hence, mean self-interacting version z of x is converging to x in L^2 , which means that z convergence to the equilibrium measure π and hence can be used as an alternative model for sampling.

Self-averaged ergodic averaged algorithm

By considering error decomposition studies in the previous section, we see that the key ingredient that we ought to understand is the following rate of convergence

Key ideas: here we want to take part of a special structure, i.e

$$\begin{cases} x_t = x_0 + \int_0^t V(x_s) + W(x_s, \mathcal{L}(x_s)) ds + \int_0^t \sigma dw_s \\ \mathcal{L}(x_t) \text{ is the law of } x_t, \end{cases}$$

such that potential V is say convex and W has a small Lipschitz constant we should obtain exponential "forgetting property", as we observed in Example 4.1.

For small time this will be bad approximation and because we average the error the bad approximation from the initial time will prevail.

Further, we consider particle system of the form

$$z_t^{i,N} = x_0^{i,N} + \int_0^t \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{s} \int_0^s b(z_s^{i,N}, z_r^{j,N}) dr \right) ds + \int_0^t \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{s} \int_0^s \sigma(z_s^{i,N}, z_r^{j,N}) dr \right) dw_s^i \quad (17)$$

for which the error to be investigate is

$$\left| \mathbb{E} [f(z_t^1)] - \mathbb{E} [f(z_t^{1,N})] \right|,$$

to have in place of (HW). According to our computation on example 4.1, we impose the following assumption

The following bound gives a leading error term. We chose λ as a exponent for simplicity as its particular value does not affect asymptotic cost/error analysis.

Assumption 4.2.

$$\left| \mathbb{E} f(z_t^1) - \mathbb{E} [f(z_t^{1,N_t})] \right| \lesssim \frac{e^{-\lambda t}}{N_t}. \quad (\text{HEW})$$

The time-discretisation of the equation (17) reads as

$$r_t^{i,N_t} = x_0^{i,N_t} + \int_0^t \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\frac{1}{s} \int_0^s b(r_{\kappa_n(s)}^{i,N_t}, r_{\kappa_n(\theta)}^{j,N_t}) d\theta \right) ds + \int_0^t \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\frac{1}{s} \int_0^s \sigma(r_{\kappa_n(s)}^{i,N_t}, r_{\kappa_n(\theta)}^{j,N_t}) d\theta \right) dw_s^i. \quad (18)$$

Let's introduce the corresponding estimator, an ensemble self-integrated version of AEA.

$$\frac{1}{M} \sum_{j=1}^M \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\frac{1}{t} \int_0^t f(r_{\kappa_n}^{(i, N_t), (j, M)}) ds \right). \quad (\text{CS-AEA})$$

In what follows, we test (HEW) in the cost analysis. Notice that we cannot test it directly, as we need to do Monte-Carlo approximation for the expectation. We first observe that by exchangeability of the law of the particle systems,

$$\mathbb{E}[f(z_t^1)] - \mathbb{E}[f(z_t^{1, N})] = \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N f(z_t^i) \right] - \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N f(z_t^{i, N}) \right].$$

Next we consider M independent ensembles to approximate expectations that is

$$\begin{aligned} & \left| \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N f(z_t^{ij}) - \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N f(z_t^{(i, N), (j, M)}) \right| \\ & \leq \left| \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N f(z_t^{ij}) - \mathbb{E}[f(z_t^1)] \right| \\ & \quad + \left| \mathbb{E}[f(z_t^1)] - \mathbb{E}[f(z_t^{1, N})] \right| + \left| \mathbb{E}[f(z_t^{1, N})] - \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N f(z_t^{(i, N), (j, M)}) \right| \\ & \lesssim (\sqrt{MN})^{-1} + (Nt)^{-1} + (\sqrt{MN})^{-1} \end{aligned}$$

where the first and last error are standard MC estimates while the middle one is given by (HEW).

4.1. Cost analysis

We analyse the cost of the self-averaged ergodic averaged estimator

$$mse(\mathcal{A}^{ES-AEA}(t, n, N)) = \mathbb{E} \left[\left(\int_{\mathbb{R}^d} f(x) \pi(dx) - \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\frac{1}{t} \int_0^t f(r_{\kappa_n}^{i, N_t}) ds \right) \right)^2 \right]^{1/2}.$$

The mean-square error decomposition reads

$$\begin{aligned} & mse(\mathcal{A}^{ES-AEA}(t, n, N_t)) \\ & \lesssim \left| \int_{\mathbb{R}^d} f(x) (\pi(dx) - \mu_t(dx)) \right| + \left| \mathbb{E}[f(y_t)] - \mathbb{E}[f(z_t^{1, N_t})] \right| \\ & \quad + \left| \mathbb{E}[f(z_t^{1, N_t})] - \mathbb{E}[f(r_t^{1, N_t})] \right| + \left(\mathbb{E} \left[\left(\mathbb{E}[f(r_t^{1, N_t})] - \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\frac{1}{t} \int_0^t f(r_{\kappa_n}^{i, N_t}) ds \right) \right)^2 \right] \right)^{1/2} \end{aligned}$$

Reasoning as before, with (HW) replaced by (HEW), we have

$$mse(\mathcal{A}^{ES-AEA}(t, n, N_t)) \lesssim e^{-\lambda t} + e^{-\lambda t} (N_t)^{-1} + 1/n + 1/\sqrt{tN_t}.$$

Note that because of the variance (last term) there is no benefit of exponential decay in t assumed in HEW. Again we fix ϵ . The following choice of the parameters ensures that $mse(\mathcal{A}^{ES-AEA}(t, n, N_t)) \lesssim \epsilon$:

$$t \approx \lambda^{-1} \log(\epsilon^{-1}), \quad tN_t \approx \epsilon^{-2}, \quad n \approx \epsilon^{-1}.$$

Notice that the choice implies that tN_t to be "constant". Hence we chose $N_t = Nt^{-1}$. Which in the case of $t \approx \lambda^{-1} \log(\epsilon^{-1})$ implies that $N \approx \epsilon^{-2}/\log(\epsilon^{-1})$.

Now we study the computational cost of simulating self-interacting diffusions (for the nonlinear interacting kernel). Note that, due nonlinear interactions at every step of the Euler scheme, we have N_t particles and each particle interacts with itself from all the past times-steps. Recall also that we take n time steps in each unit time interval so that overall number of steps on the interval $[0, t]$ is tn . With that in mind and the fact that we take $N_t = Nt^{-1}$ we have

$$\begin{aligned} cost(\mathcal{A}^{ES-AEA}(t, n, N)) &= N_{1/n}^2 1 + N_{2/n}^2 2 + N_{3/n}^2 3 + \dots + N_{tn/n}^2 tn \\ &= (nN)^2 (1 + 1/2 + 1/3 + \dots + 1/tn) = (nN)^2 \sum_{k=1}^{tn} 1/k \approx (nN)^2 \log(1 + tn). \end{aligned}$$

Hence the cost for the set of parameters $\{t \approx \lambda^{-1} \log(\epsilon^{-1}), N \approx \epsilon^{-2}/\log(\epsilon^{-1}), n \approx \epsilon^{-1}\}$ is

$$cost(\mathcal{A}^{ES-AEA}(t, n, N)) \approx \epsilon^{-6}$$

Note also that one can also take

$$t \approx \epsilon^{-2}, \quad (N_t = 1 \implies tN_t = \epsilon^{-2}), \quad n \approx \epsilon^{-1}$$

to ensure $mse(\mathcal{A}^{ES-AEA}(t, n, N_t)) \lesssim \epsilon$. This second choice, with N_t not varying with time, leads to

$$\begin{aligned} cost(\mathcal{A}^{ES-AEA}(t, n, N)) &= N^2 1 + N^2 2 + N^2 3 + \dots + N^2 tn = N^2 \frac{1}{2} tn(1 + tn) \approx \epsilon^{-6}. \end{aligned}$$

Let us consider ensemble implementation of the above algorithm. Reasoning as before, by the Assumption HEW, we have

$$mse(\mathcal{A}^{ES-AEA}(t, n, N_t)) \lesssim e^{-\lambda t} + e^{-\lambda t} (N_t)^{-1} + 1/n + 1/\sqrt{tN_t M}.$$

To balance the first two terms on the right hand side we take $N_t = 1$. With that choice we then chose M so that $e^{-\lambda t} = (tM)^{-1/2}$, i.e $M = e^{2\lambda t} t^{-1}$. With this choices we have

$$mse(\mathcal{A}^{ES-AEA}(t, n, N_t)) \lesssim e^{-\lambda t} + 1/n,$$

to make that error to be less then ϵ we take $t = \lambda^{-1} \log \epsilon^{-1}$ and $n = \epsilon^{-1}$. Note that this leads to $M = e^{\log(\epsilon^{-2})} (\log(\epsilon^{-1}))^{-1} = \epsilon^{-2} (\log(\epsilon^{-1}))^{-1}$. Reasoning as before the cost with $N_t = 1$ is

$$cost(\mathcal{A}^{CS-AEA}(t, n, N_t, M)) \approx M(tn)^2 \approx \epsilon^{-4} \log(\epsilon^{-1}).$$

Hence one more time we achieved order better computational cost in comparison to naive estimator. Note that the presented analysis implies that ensemble of M independent self-interacting diffusions yields best result.

5. CONCLUSION AND PERSPECTIVES

We presented a number of different algorithms for the approximation of the invariant measure of the McKean-Vlasov SDE. We have achieved one order better computational costs comparing to naive particle based estimator. On algorithmic side possible extensions are that may consider fixed length window for self-interacting diffusion and it also possible to study Multilevel Monte Carlo strategies in this setup. Overall we anticipate that it will be possible to bring the cost of simulating particle system to the same level as for standard independent copies of SDEs.

REFERENCES

- [Ber18] Espen Bernton. Langevin Monte Carlo and JKO splitting. *arXiv preprint arXiv:1802.08671*, 2018.
- [BGT08] M Bossy, A Ganz, and D Talay. Equilibrium measure approximation for diffusion processes with a McKean-Vlasov perturbation of a monotone drift. From A. Ganz PhD Thesis, Inria, 2008.
- [CCD] Jean-François Chassagneux, Dan Crisan, and François Delarue. A probabilistic approach to classical solutions of the master equation for large population equilibria.
- [CCG12] P. Cattiaux, D. Chafai, and A. Guillin. Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA*, 9(2):337–382, 2012.
- [CD18] René Carmona and François Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II*. Springer, 2018.
- [Dal17] Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):651–676, 2017.
- [DEGZ18] Alain Durmus, Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. An Elementary Approach To Uniform In Time Propagation Of Chaos. *arXiv preprint arXiv:1805.11387*, 2018.
- [DM⁺17] Alain Durmus, Eric Moulines, et al. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3):1551–1587, 2017.
- [EGZ16] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative Harris type theorems for diffusions and McKean-Vlasov processes. *arXiv preprint arXiv:1606.06012*, 2016.
- [HAT16] Abdul-Lateef Haji-Ali and Raul Tempone. Multilevel and Multi-index Monte Carlo methods for McKean-Vlasov equations. *arXiv preprint arXiv:1610.09934*, 2016.
- [HŠS18] William Hammersley, David Šiška, and Lukasz Szpruch. McKean-Vlasov SDEs under Measure Dependent Lyapunov Conditions. *arXiv preprint arXiv:1802.03974*, 2018.
- [JFC] Alvin Tse Jean-François Chassagneux, Lukasz Szpruch. Weak quantitative propagation of chaos via differential calculus on the space of measures. *arXiv preprint arXiv: 1901.02556*.
- [KK12] Victor Kleptsyn and Aline Kurtzmann. Ergodicity of self-attracting motion. *Electron. J. Probab.*, 17:no. 50, 37, 2012.
- [Kol10] Vassili N Kolokoltsov. *Nonlinear Markov processes and kinetic equations*, volume 182. Cambridge University Press, 2010.
- [KV86] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.
- [Liu17] Qiang Liu. Stein variational gradient descent as gradient flow. In *Advances in neural information processing systems*, pages 3115–3123, 2017.
- [LP⁺02] Damien Lambertson, Gilles Pages, et al. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002.
- [LP⁺10] Vincent Lemaire, Gilles Pagès, et al. Unconstrained recursive importance sampling. *The Annals of Applied Probability*, 20(3):1029–1067, 2010.
- [LRS10] Tony Lelièvre, Mathias Rousset, and Gabriel Stoltz. *Free energy computations: A mathematical perspective*. World Scientific, 2010.
- [Mél96] Sylvie Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations*, pages 42–95. Springer, 1996.
- [MST10] Jonathan C Mattingly, Andrew M Stuart, and Michael V Tretyakov. Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM Journal on Numerical Analysis*, 48(2):552–577, 2010.
- [PP⁺12] Gilles Pagès, Fabien Panloup, et al. Ergodic approximation of the distribution of a stationary diffusion: rate of convergence. *The Annals of Applied Probability*, 22(3):1059–1100, 2012.
- [ŠLMD18] Umut Şimşekli, Antoine Liutkus, Szymon Majewski, and Alain Durmus. Sliced-Wasserstein Flows: Nonparametric Generative Modeling via Optimal Transport and Diffusions. *arXiv preprint arXiv:1806.08141*, 2018.
- [Szn91] Alain-Sol Sznitman. *Topics in propagation of chaos*. Springer, 1991.
- [Tal90] Denis Talay. Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics: An International Journal of Probability and Stochastic Processes*, 29(1):13–36, 1990.

- [Tal02] D Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002.
- [Ver06] A Yu Veretennikov. On ergodic measures for McKean-Vlasov stochastic equations. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 471–486. Springer, 2006.