Multilevel Picard approximations for McKean-Vlasov stochastic differential equations

Martin Hutzenthaler* Thomas Kruse† Tuan Anh Nguyen‡

Monday 22nd March, 2021

Abstract

In the literature there exist approximation methods for McKean-Vlasov stochastic differential equations which have a computational effort of order 3. In this article we introduce full-history recursive multilevel Picard approximations for McKean-Vlasov stochastic differential equations. We prove that these MLP approximations have computational effort of order 2+ which is essentially optimal in high dimensions.

1 Introduction

McKean [19] introduced stochastic differential equations (SDEs) whose coefficients depend on the distribution of the solution. These McKean-Vlasov SDEs allow a stochastic representation of solutions of nonlinear, possibly non-local partial parabolic differential equations (PDEs) such as Vlasov's equation, Boltzmann's equation, or Burgers' equation. Moreover, weakly dependent diffusions converge to independent solutions of McKean-Vlasov SDEs as the system size tends to infinity. This phenomenon was termed *propagation of chaos* by Kac [16] and is well studied in the literature; see, e.g., [18, 20, 8, 22, 11, 24, 15].

For simplicity we consider in this article the McKean-Vlasov SDE in (3) below with additive noise whose drift coefficient depends linearly on the distribution of the solution. In the literature there exist a number of approximation methods for the solution of (3). A direct approach approximates the spatial integral in (3) with an average over weakly dependent versions of the solution (resulting in weakly interacting diffusions) and the temporal integral in (3) with suitable Rieman sums (Euler method). The L^2-error of this approximation is of order $1/N$ as $\mathbb{N} = \{1, 2, \ldots\} \ni N \to \infty$ if we use N^2 interacting diffusions and N time intervals resulting

*Faculty of Mathematics, University of Duisburg-Essen, Essen, Germany; e-mail: martin.hutzenthaler@uni-due.de

†Institute of Mathematics, University of Gießen, Gießen, Germany; e-mail: thomas.kruse@math.uni-giessen.de

‡Faculty of Mathematics, University of Duisburg-Essen, Essen, Germany; e-mail: tuan.nguyen@uni-due.de

Key words and phrases: curse of dimensionality, high-dimensional SDEs, high-dimensional semilinear McKean-Vlasov SDEs, multilevel Picard approximations, multilevel Monte Carlo method.

AMS 2010 subject classification: 65M75

in N^5 function evaluations of the drift coefficient. Thus the computational effort for achieving L^2-error $\varepsilon \in (0,1)$ is of order ε^{-5} as $\varepsilon \to 0$; cf., e.g., [1, 5, 6]. This computational effort for achieving error $\varepsilon \in (0,1)$ can be reduced to order $\varepsilon^{-4}|\log(\varepsilon)|^3$ by replacing averages by the multilevel Monte Carlo method (cf. [23, Theorem 4.5] and, e.g., [10, 9, 17]) and to order ε^{-3} by the antithetic multilevel Monte Carlo method (cf. [24, Theorem 4.3]). In the very special case of ordinary differential equations with an expectation in the driving function, where the plain vanilla Monte Carlo method has computational effort of order ε^{-3}, [3, Theorem 1.1] shows that this computational effort can be reduced to order ε^{-2+}. In low dimensions, the spatial integral in (3) can also be approximated e.g. by projections on function spaces and then the computational effort can be reduced to order $\varepsilon^{-2}|\log(\varepsilon)|^4$ (or better); cf., e.g., [4, Theorem 4]. In high dimensions, the numerical approximation of Lebesgue integrals with (deterministic) quadrature rules suffers from the curse of dimensionality; see [21]. The Monte Carlo method overcomes this curse and achieves a L^2-error $\varepsilon \in (0,1)$ with computational effort of order ε^{-2} in the numerical approximation of Lebesgue integrals without the curse of dimensionality. Thus, in high dimensions, the computational effort for approximating the spatial integral on the right-hand side of (3) has optimal order ε^{-2} and this is clearly a lower bound for the approximation of the full McKean-Vlasov SDE. It remained an open question in the literature whether McKean-Vlasov SDEs can be approximated up to L^2-error $\varepsilon \in (0,1)$ with computational effort of order ε^{-2} (or whether an higher effort such as ε^{-3} is required in general).

In this article we partially answer this question positively. In other words, we show that the computational problem of approximating the solution of the McKean-Vlasov SDE in (3) has up to logarithmic factors the same computational complexity as the numerical approximation of the spatial integral in (3). More specifically, we view (3) as fixed point equation and adapt the full-history multilevel Picard (MLP) method, which was introduced in [7], to this fixed point equation. This MLP method was already successfully applied to overcome the curse of dimensionality in the numerical approximation of semilinear PDEs; see, e.g., [14, 12, 2, 13]. Our MLP approximation method (2) below is, roughly speaking, based on the idea to (a) reformulate the McKean-Vlasov SDE in (3) as a stochastic fixed point problem $X = \Phi(X)$ with a suitable function Φ, to (b) approximate the fixed point X through Picard iterates $(X_k)_{k\in\{0,1,2,\ldots\}}$, to (c) write X as telescoping series over this sequence, that is,

$$X = X_1 + \sum_{k=1}^{\infty}(X_{k+1} - X_k) = X_1 + \sum_{k=1}^{\infty}\Big(\Phi(X_k) - \Phi(X_{k-1})\Big), \tag{1}$$

and to (d) approximate the series by a finite sum and the temporal and spatial integrals in the summands by Monte Carlo averages with fewer and fewer independent samples as k increases. Roughly speaking, the rationale behind this approach is that $X_{k+1} - X_k$ converges exponentially fast (or even factorially fast) to 0 as $k \to \infty$ and the mean squared error of the Monte Carlo average is bounded by the second moment of the involved random variable divided by the number of independent samples in the average. This motivates our MLP approximations in (2).

The main result of this article, Theorem 3.1 in Section 3 below, implies that the MLP approximation method approximates solutions of McKean-Vlasov SDEs with additive noise whose drift coefficients depend linearly on the distribution of the solution up to an L^2-error $\varepsilon \in (0,1)$ with computational effort ε^{-2+} without suffering from the curse of dimensionality.

To illustrate our main results, we now present in Theorem 1.1 a special case of Theorem 3.1.

Theorem 1.1. *Let $\delta, T \in (0,\infty)$, $d \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\Theta = \bigcup_{n\in\mathbb{N}} (\mathbb{N}_0)^n$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be a norm, let $\mu\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be globally Lipschitz continuous, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathfrak{u}^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, be i.i.d. random variables, assume for all $t \in [0,1]$ that $\mathbb{P}(\mathfrak{u}^0 \le t) = t$, let $W^\theta\colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta\in\Theta$, be i.i.d. standard Brownian motions with continuous sample paths, assume that $(\mathfrak{u}^\theta)_{\theta\in\Theta}$ and $(W^\theta)_{\theta\in\Theta}$ are independent, let $X^\theta_{n,m}\colon [0,T]\times\Omega\to\mathbb{R}^d$, $\theta\in\Theta$, $n,m\in\mathbb{N}_0$, satisfy for all $\theta\in\Theta$, $m\in\mathbb{N}$, $n\in\mathbb{N}_0$, $t\in[0,T]$ that*

$$X^\theta_{n,m}(t) = \big(\xi + W^\theta\big(\max(\{\tfrac{kT}{m^n}\colon k\in\mathbb{N}_0\}\cap[0,t])\big) + t\mu(0,0)\big)\,\mathbb{1}_{\mathbb{N}}(n) + \sum_{\ell=1}^{n-1}\sum_{k=1}^{m^{n-\ell}} \frac{t\Big[\mu\Big(X^\theta_{\ell,m}(\mathfrak{u}^{(\theta,n,k,\ell)}t), X^{(\theta,n,k,\ell)}_{\ell,m}(\mathfrak{u}^{(\theta,n,k,\ell)}t)\Big) - \mu\Big(X^\theta_{\ell-1,m}(\mathfrak{u}^{(\theta,n,k,\ell)}t), X^{(\theta,n,k,\ell)}_{\ell-1,m}(\mathfrak{u}^{(\theta,n,k,\ell)}t)\Big)\Big]}{m^{n-\ell}}, \tag{2}$$

let $X\colon [0,T]\times\Omega\to\mathbb{R}^d$ be a $(\sigma(\{W^0(s)\colon s\in[0,t]\}))_{t\in[0,T]}$-adapted stochastic process with continuous sample paths, assume for all $t\in[0,T]$ that $\int_0^T (\mathbb{E}[\|X(s)\|^2])^{1/2}\,ds < \infty$ and

$$X(t) = \xi + \int_0^t \int \mu(X(s),x)\mathbb{P}\big(X(s)\in dx\big)\,ds + W^0(t), \tag{3}$$

and for every $n,m\in\mathbb{N}$ let $C_{n,m}\in\mathbb{N}_0$ be the number of function evaluations of μ and the number of scalar random variables which are used to compute one realization of $X^0_{n,m}(T)$ (cf. (19) in Theorem 3.1 below). Then there exist $c\in\mathbb{R}$ and $\mathrm{n} = (\mathrm{n}_\varepsilon)_{\varepsilon\in(0,1]}\colon (0,1]\to\mathbb{N}$, such that for all $\varepsilon\in(0,1]$ it holds that $\sup_{k\in[\mathrm{n}_\varepsilon,\infty)\cap\mathbb{N}}\sup_{t\in[0,T]}\big(\mathbb{E}\big[\|X^0_{k,k}(t)-X(t)\|^2\big]\big)^{1/2}\le\varepsilon$ and $C_{\mathrm{n}_\varepsilon,\mathrm{n}_\varepsilon}\le c\,\varepsilon^{-(2+\delta)}$.

In the following we add further comments on our approximation method. The MLP approximations $X^\theta_{n,m}$, $n\in\mathbb{N}_0$, $m\in\mathbb{N}$, $\theta\in\Theta$, in (2) are indexed by the number $n\in\mathbb{N}_0$ of fixed point iterates, by a parameter $m\in\mathbb{N}$ which is fixed in the recursion in (2) and is the basis of the number of Monte Carlo averages, and by a parameter $\theta\in\Theta$ which is used to distinguish independent MLP approximations in (2). We note for every θ that all $X^\theta_{n,m}$, $n,m\in\mathbb{N}$, depend on the same Brownian path W^θ so that for all $n,m\in\mathbb{N}$, $\theta\in\Theta$, $t\in[0,T]$ we need to have $(W^\theta(\frac{kT}{m^n}))_{k\in\{0,1,\ldots,m^n\}}$ as argument of the function call which calculates $X^\theta_{n,m}(t)$.

The remainder of this article is organized as follows. In Section 2 we solve recursions of Gronwall-type. In particular, Corollary 2.3 will be applied to obtain an upper bound for the computational effort which satisfies the recursion in (19) in Theorem 3.1. Moreover, in Theorem 3.1 in Section 3 we estimate the L^2-error between the solution of the McKean-Vlasov SDE and our MLP approximations and we estimate the computational effort for computing one realization of our MLP approximation.

2 Discrete Gronwall-type recursions

In this section we solve recursions of Gronwall-type. The following result, Lemma 2.1, provides the exact solutions of certain linear recurrence relations of second order.

Lemma 2.1 (Two-step recursions)**.** *Let $\kappa,\lambda,x_1,x_2\in\mathbb{C}$, $(a_k)_{k\in\mathbb{N}_0},(b_k)_{k\in\mathbb{N}_0}\subseteq\mathbb{C}$ satisfy for all $k\in\mathbb{N}_0$, $i\in\{1,2\}$ that*

$$a_0 = b_0, \quad a_1 = b_1 + \kappa b_0, \quad a_{k+2} = b_{k+2} + \kappa a_{k+1} + \lambda a_k, \quad x_i^2 = \kappa x_i + \lambda, \quad \text{and} \quad x_1 \ne x_2. \tag{4}$$

Then it holds for all $k \in \mathbb{N}_0$ *that* $a_k = \frac{1}{x_2-x_1}\sum_{\ell=0}^{k} b_\ell (x_2^{k-\ell+1} - x_1^{k-\ell+1})$.

Proof of Lemma 2.1. Throughout this proof let $(z_k)_{k\in\mathbb{N}_0} \subseteq \mathbb{C}$ satisfy for all $k \in \mathbb{N}_0$ that $z_k = \frac{1}{x_2-x_1}\sum_{\ell=0}^{k} b_\ell \left(x_2^{k-\ell+1} - x_1^{k-\ell+1}\right)$. We consider the two cases $\lambda = 0$ and $\lambda \neq 0$.

Case 1. $\lambda = 0$. The fact that $x_1 \neq x_2$ and the fact that $\forall\, i \in \{1,2\}\colon x_i^2 = \kappa x_i$ prove that $\kappa \neq 0$ and $(x_1, x_2) \in \{(0,\kappa),(\kappa,0)\}$. This proves for all $k \in \mathbb{N}_0$ that $z_k = \frac{1}{\kappa}\sum_{\ell=0}^{k} b_\ell \kappa^{k-\ell+1} = \sum_{\ell=0}^{k} b_\ell \kappa^{k-\ell}$. This and the fact that $\lambda = 0$ show for all $k \in \mathbb{N}_0$ that $z_0 = b_0$, $z_1 = b_1 + \kappa b_0$, and $z_{k+2} = \sum_{\ell=0}^{k+2} b_\ell \kappa^{k+2-\ell} = b_{k+2} + \kappa z_{k+1} = b_{k+2} + \kappa z_{k+1} + \lambda z_k$. This, (4), and induction prove for all $k \in \mathbb{N}_0$ that $z_k = a_k$.

Case 2. $\lambda \neq 0$. The fact that $\forall\, i \in \{1,2\}\colon x_i^2 = \kappa x_i + \lambda$ implies that $x_1 \neq 0$ and $x_2 \neq 0$. Moreover, the fact that $0 = (x_1^2 - \kappa x_1 - \lambda) - (x_2^2 - \kappa x_2 - \lambda) = (x_1 - x_2)(x_1 + x_2 - \kappa)$ and the fact that $x_1 \neq x_2$ imply that $x_1 + x_2 = \kappa$. Next, the fact that $0 = \frac{x_1^2-\kappa x_1-\lambda}{x_1} - \frac{x_2^2-\kappa x_2-\lambda}{x_2} = x_1 - x_2 + \lambda(\frac{1}{x_2} - \frac{1}{x_1})$ imply that $\frac{\lambda}{x_2-x_1}(\frac{1}{x_2} - \frac{1}{x_1}) = 1$. This, the definition of $(z_k)_{k\in\mathbb{N}_0}$, and the fact that $x_1 + x_2 = \kappa$ imply for all $k \in \mathbb{N}_0$ that $z_0 = b_0 = a_0$, $z_1 = \frac{1}{x_2-x_1}\left(b_0(x_2^2 - x_1^2) + b_1(x_2 - x_1)\right) = b_0(x_1 + x_2) + b_1 = b_0\kappa + b_1 = a_1$, and

$$
\begin{aligned}
z_{k+2} &= \sum_{\ell=0}^{k+2} \frac{b_\ell\left(x_2^{k-\ell+3} - x_1^{k-\ell+3}\right)}{x_2-x_1} = \sum_{\ell=0}^{k+2} \frac{b_\ell\left(x_2^{k-\ell+1}x_2^2 - x_1^{k-\ell+1}x_1^2\right)}{x_2-x_1} = \sum_{\ell=0}^{k+2} \frac{b_\ell\left(x_2^{k-\ell+1}(\kappa x_2+\lambda) - x_1^{k-\ell+1}(\kappa x_1+\lambda)\right)}{x_2-x_1} \\
&= \kappa\left[\sum_{\ell=0}^{k+2} \frac{b_\ell\left(x_2^{k-\ell+2} - x_1^{k-\ell+2}\right)}{x_2-x_1}\right] + \lambda\left[\sum_{\ell=0}^{k+2} \frac{b_\ell\left(x_2^{k-\ell+1} - x_1^{k-\ell+1}\right)}{x_2-x_1}\right] \\
&= \kappa\left[\sum_{\ell=0}^{k+1} \frac{b_\ell\left(x_2^{(k+1)-\ell+1} - x_1^{(k+1)-\ell+1}\right)}{x_2-x_1}\right] + \lambda\left[\sum_{\ell=0}^{k} \frac{b_\ell\left(x_2^{k-\ell+1} - x_1^{k-\ell+1}\right)}{x_2-x_1}\right] + \lambda\left[\sum_{\ell=k+1}^{k+2} \frac{b_\ell\left(x_2^{k-\ell+1} - x_1^{k-\ell+1}\right)}{x_2-x_1}\right] \\
&= \kappa z_{k+1} + \lambda z_k + \frac{\lambda b_{k+2}}{x_2-x_1}\left(\frac{1}{x_2} - \frac{1}{x_1}\right) = \kappa z_{k+1} + \lambda z_k + b_{k+2}.
\end{aligned}
\tag{5}
$$

This, (4), and induction show for all $k \in \mathbb{N}_0$ that $z_k = a_k$. Combining the two cases $\lambda = 0$ and $\lambda \neq 0$ completes the proof of Lemma 2.1. ☐

The following result, Lemma 2.2, generalizes the discrete Gronwall inequality which is the special case $\lambda = 0$ of Lemma 2.2.

Lemma 2.2 (Discrete Gronwall-type recursion)**.** *Let* $\kappa, \lambda, x_1, x_2 \in \mathbb{C}$, $(a_n)_{n\in\mathbb{N}_0}, (b_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{C}$ *satisfy for all* $n \in \mathbb{N}_0$, $i \in \{1,2\}$ *that*

$$
a_n = b_n + \sum_{k=0}^{n-1}\left[\kappa a_k + \mathbb{1}_{\mathbb{N}}(k)\lambda a_{|k-1|}\right], \quad x_i^2 = (1+\kappa)x_i + \lambda, \quad \text{and} \quad x_1 \neq x_2. \tag{6}
$$

Then it holds for all $n \in \mathbb{N}_0$ *that*

$$
a_n = \sum_{k=0}^{n}\left[\frac{b_k - \mathbb{1}_{\mathbb{N}}(k) b_{|k-1|}}{x_2-x_1}\left(x_2^{n-k+1} - x_1^{n-k+1}\right)\right]. \tag{7}
$$

Proof of Lemma 2.2. Throughout this proof let $(z_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{C}$ satisfy for all $n \in \mathbb{N}_0$ that $z_n = \sum_{k=0}^{n} a_k$. This and (6) show for all $n \in \mathbb{N}_0$ that $z_0 = a_0 = b_0$, $z_1 = a_0 + a_1 = a_0 + b_1 + \kappa a_0 = b_1 + (1+\kappa)b_0$, $z_{n+2} - z_{n+1} = a_{n+2} = b_{n+2} + \kappa\left(\sum_{k=0}^{n+1} a_k\right) + \lambda\left(\sum_{k=0}^{n} a_k\right) = b_{n+2} + \kappa z_{n+1} + \lambda z_n$, and therefore $z_{n+2} = b_{n+2} + (1+\kappa)z_{n+1} + \lambda z_n$. This, Lemma 2.1 (applied with $\kappa \curvearrowleft (1+\kappa)$,

$(a_n)_{n\in\mathbb{N}_0} \curvearrowleft (z_n)_{n\in\mathbb{N}_0}$ in the notation of Lemma 2.1), and the assumptions on x_1, x_2 prove for all $n \in \mathbb{N}_0$ that $z_n = \sum_{k=0}^{n} \frac{b_k(x_2^{n-k+1}-x_1^{n-k+1})}{x_2-x_1}$. Therefore, it holds for all $n \in \mathbb{N}$ that

$$\begin{aligned} a_n = z_n - z_{n-1} &= \sum_{k=0}^{n} \tfrac{b_k(x_2^{n-k+1}-x_1^{n-k+1})}{x_2-x_1} - \sum_{k=0}^{n-1} \tfrac{b_k(x_2^{n-k}-x_1^{n-k})}{x_2-x_1} \\ &= \sum_{k=0}^{n} \tfrac{b_k(x_2^{n-k+1}-x_1^{n-k+1})}{x_2-x_1} - \sum_{k=1}^{n} \tfrac{b_{k-1}(x_2^{n-k+1}-x_1^{n-k+1})}{x_2-x_1} = \sum_{k=0}^{n} \tfrac{(b_k - \mathbb{1}_{\mathbb{N}}(k) b_{|k-1|})(x_2^{n-k+1}-x_1^{n-k+1})}{x_2-x_1}. \end{aligned} \tag{8}$$

This and the fact that $a_0 = b_0$ (see (6)) complete the proof of Lemma 2.2. □

The following result, Corollary 2.3, generalizes, e.g., [14, Lemma 3.6].

Corollary 2.3 (Discrete Gronwall-type inequality)**.** *Let $(a_n)_{n\in\mathbb{N}_0} \subseteq [0,\infty)$, $\kappa, \lambda, c_1, c_2, c_3, c_4, \beta \in [0,\infty)$ satisfy for all $n \in \mathbb{N}_0$ that*

$$a_n \leq c_1 + c_2 n + c_3 \sum_{k=1}^{n} c_4^k + \sum_{k=0}^{n-1} \left[\kappa a_k + \lambda \mathbb{1}_{\mathbb{N}}(k) a_{|k-1|}\right] \quad \textit{and} \quad \beta = \tfrac{(1+\kappa)+\sqrt{(1+\kappa)^2+4\lambda}}{2} > 1. \tag{9}$$

Then it holds for all $n \in \mathbb{N}_0$ that

$$a_n \leq \begin{cases} \frac{3}{2}\beta^n c_1 + \frac{3c_2(\beta^n-1)}{2(\beta-1)} + \frac{3}{2} c_3 n \beta^n & \textit{if } c_4 = \beta, \\ \frac{3}{2}\beta^n c_1 + \frac{3c_2(\beta^n-1)}{2(\beta-1)} + \frac{3c_3(c_4^{n+1}-c_4\beta^n)}{2(c_4-\beta)} & \textit{else.} \end{cases} \tag{10}$$

Proof of Corollary 2.3. Throughout this proof let $(\tilde{a}_n)_{n\in\mathbb{N}_0}, (b_n)_{n\in\mathbb{N}_0} \subseteq [0,\infty)$, $x_1, x_2 \in \mathbb{R}$ satisfy for all $n \in \mathbb{N}_0$ that $x_1 = \frac{(1+\kappa)-\sqrt{(1+\kappa)^2+4\lambda}}{2}$, $x_2 = \frac{(1+\kappa)+\sqrt{(1+\kappa)^2+4\lambda}}{2}$,

$$\tilde{a}_n = b_n + \sum_{k=0}^{n-1} \left[\kappa \tilde{a}_k + \lambda \mathbb{1}_{\mathbb{N}}(k) \tilde{a}_{|k-1|}\right], \quad \text{and} \quad b_n = c_1 + c_2 n + c_3 \sum_{k=1}^{n} c_4^k. \tag{11}$$

This and the quadratic formula show for all $n \in \mathbb{N}_0$, $k \in \{0, 1, \ldots, n\}$, $i \in \{1, 2\}$ that

$$\begin{aligned} &\tfrac{|x_1|}{|x_2-x_1|} = \tfrac{\sqrt{(1+\kappa)^2+4\lambda}-(1+\kappa)}{2\sqrt{(1+\kappa)^2+4\lambda}} \leq \tfrac{1}{2}, \qquad \tfrac{|x_2|}{|x_2-x_1|} = \tfrac{(1+\kappa)+\sqrt{(1+\kappa)^2+4\lambda}}{2\sqrt{(1+\kappa)^2+4\lambda}} \leq 1, \\ &|x_1| \leq |x_2|, \quad \text{and} \quad \left|\tfrac{x_2^{n-k+1}-x_1^{n-k+1}}{x_2-x_1}\right| \leq \left[\tfrac{|x_2|}{|x_2-x_1|}|x_2|^{n-k} + \tfrac{1}{|x_2-x_1|}|x_1|^{n-k+1}\right] \leq \tfrac{3}{2}|x_2|^{n-k}, \end{aligned} \tag{12}$$

$x_1 \neq x_2$, $x_i^2 = (1+\kappa)x_i + \lambda$, $b_0 = c_1$, and $b_{n+1} - b_n = c_2 + c_3 c_4^{n+1}$. This, Lemma 2.2 (applied with $(a_n)_{n\in\mathbb{N}_0} \curvearrowleft (\tilde{a}_n)_{n\in\mathbb{N}_0}$ in the notation of Lemma 2.2), the fact that $\lambda, \kappa, c_1, c_2, c_3, c_4 \geq 0$, and the definition of x_2 show for all $n \in \mathbb{N}_0$ that

$$\begin{aligned} \tilde{a}_n &= \sum_{k=0}^{n} \tfrac{(b_k - \mathbb{1}_{\mathbb{N}}(k) b_{|k-1|})(x_2^{n-k+1}-x_1^{n-k+1})}{x_2-x_1} \leq \tfrac{3}{2} c_1 |x_2|^n + \sum_{k=1}^{n} \left[\tfrac{3}{2}(c_2 + c_3 c_4^k)|x_2|^{n-k}\right] \\ &= \tfrac{3}{2}|x_2|^n c_1 + \tfrac{3}{2} c_2 \left[\sum_{k=0}^{n-1} |x_2|^k\right] + \tfrac{3}{2} c_3 |x_2|^n \left[\sum_{k=1}^{n} \tfrac{|c_4|^k}{|x_2|^k}\right]. \end{aligned} \tag{13}$$

This and the fact that $x_2 = \beta > 1$ show for all $n \in \mathbb{N}_0$ that if $c_4 \neq x_2$, then

$$\begin{aligned} \tilde{a}_n &\leq \tfrac{3}{2}|x_2|^n c_1 + \tfrac{3}{2} c_2 \tfrac{|x_2|^n - 1}{|x_2|-1} + \tfrac{3}{2} c_3 |x_2|^n \frac{\left|\frac{c_4}{x_2}\right|^{n+1} - \left|\frac{c_4}{x_2}\right|}{\left|\frac{c_4}{x_2}\right| - 1} \\ &\leq \tfrac{3}{2}|x_2|^n c_1 + \tfrac{3}{2} c_2 \tfrac{|x_2|^n - 1}{|x_2|-1} + \tfrac{3}{2} c_3 \tfrac{|c_4|^{n+1} - c_4 |x_2|^n}{|c_4| - |x_2|} \end{aligned} \tag{14}$$

and if $c_4 = x_2$, then

$$\tilde{a}_n \leq \tfrac{3}{2}|x_2|^n c_1 + \tfrac{3}{2} c_2 \tfrac{|x_2|^n - 1}{|x_2| - 1} + \tfrac{3}{2} c_3 n |x_2|^n. \tag{15}$$

Furthermore, (9), (11), and induction prove for all $n \in \mathbb{N}_0$ that $a_n \leq \tilde{a}_n$. This, (14), (15), and the fact that $x_2 = \beta$ complete the proof of Corollary 2.3. □

3 Multilevel Picard approximations of McKean-Vlasov SDEs

The following theorem, Theorem 3.1, shows that the computational effort of MLP approximations of McKean-Vlasov SDEs is of order 2+ if the noise is additive and if the drift coefficients depend linearly on the distributions. In Theorem 3.1, for every $n, M \in \mathbb{N}$ we think of $C_{n,M}$ as an upper bound for the sum of the number of scalar random variables and the number of function evaluations of the drift coefficient which are used to compute one realization of $X^0_{n,m}(T)$. Let us comment on the recursion (19) which describes this computational effort. The binary variables $\mathfrak{v}, \mathfrak{f} \in \{0, 1\}$ indicate whether we want to count the number of scalar random variables ($\mathfrak{v} = 1$) and whether we want to count the number of function evaluations of the drift coefficient ($\mathfrak{f} = 1$). For every $n, M \in \mathbb{N}$ to compute one realization of $X^0_{n,m}(T)$ the scheme in (18) first has to generate a realization of $(W^0(\frac{kT}{m^n}))_{k \in \{0,1,\ldots,m^n\}}$ which corresponds to the generation of $m^n d$ scalar random variables. Additionally, the scheme evaluates the drift coefficient once at $(0, 0) \in \mathbb{R}^d \times \mathbb{R}^d$. Next, for every $l \in \{1, 2, \ldots, n-1\}$ the scheme does m^{n-l} times the following: it evaluates the drift coefficient twice, it generates a continuously uniformly on $[0, 1]$ distributed random variable, it generates a realization of $(W^\theta(\frac{kT}{m^l}))_{k \in \{0,1,\ldots,m^l\}}$ (corresponding to $m^l d$ scalar random variables), and for suitable $s \in [0, T]$, $\theta \in \Theta$ it calls twice the functions which calculate $X^\theta_{l,m}(s)$ and $X^\theta_{n,m}(s)$.

Theorem 3.1. *Let $T, L \in (0, \infty)$, $d \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in C(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)$, $\Theta = \bigcup_{n \in \mathbb{N}} (\mathbb{N}_0)^n$, let $\|\cdot\| \colon \mathbb{R}^d \to [0, \infty)$ be the standard norm, assume for all $x_1, y_1, x_2, y_2 \in \mathbb{R}^d$ that*

$$\big\|\mu(x_1, y_1) - \mu(x_2, y_2)\big\| \leq \tfrac{L}{2}\|x_1 - x_2\| + \tfrac{L}{2}\|y_1 - y_2\|, \tag{16}$$

let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathfrak{u}^\theta \colon \Omega \to [0, 1]$, $\theta \in \Theta$, be i.i.d. random variables, assume for all $t \in [0, 1]$ that $\mathbb{P}(\mathfrak{u}^0 \leq t) = t$, let $W^\theta \colon [0, T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be i.i.d. standard Brownian motions with continuous sample paths, assume that $(\mathfrak{u}^\theta)_{\theta \in \Theta}$ and $(W^\theta)_{\theta \in \Theta}$ are independent, let $X \colon [0, T] \times \Omega \to \mathbb{R}^d$ be a $(\sigma(\{W^0(s) \colon s \in [0, t]\}))_{t \in [0,T]}$-adapted stochastic process with continuous sample paths, assume for all $t \in [0, T]$ that $\int_0^T (\mathbb{E}[\|X(s)\|^2])^{1/2} \, ds < \infty$ and

$$X(t) = \xi + \int_0^t \int \mu(X(s), x) \mathbb{P}\big(X(s) \in dx\big) \, ds + W^0(t), \tag{17}$$

let $X^\theta_{n,m} \colon [0, T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, $m \in \mathbb{N}$, $n \in \mathbb{N}_0$, satisfy for all $\theta \in \Theta$, $m, n \in \mathbb{N}$, $t \in [0, T]$ that $X^\theta_{0,m}(t) = 0$ and

$$\begin{aligned} X^\theta_{n,m}(t) &= \xi + W^\theta\left(\sup\left(\{\tfrac{kT}{m^n} \colon k \in \mathbb{N}_0\} \cap [0, t]\right)\right) + t\mu(0, 0) \\ &+ \sum_{\ell=1}^{n-1} \sum_{k=1}^{m^{n-\ell}} \frac{t\Big[\mu\Big(X^\theta_{\ell,m}(\mathfrak{u}^{(\theta,n,k,\ell)}t), X^{(\theta,n,k,\ell)}_{\ell,m}(\mathfrak{u}^{(\theta,n,k,\ell)}t)\Big) - \mu\Big(X^\theta_{\ell-1,m}(\mathfrak{u}^{(\theta,n,k,\ell)}t), X^{(\theta,n,k,\ell)}_{\ell-1,m}(\mathfrak{u}^{(\theta,n,k,\ell)}t)\Big)\Big]}{m^{n-\ell}}, \end{aligned} \tag{18}$$

let $\mathfrak{v},\mathfrak{f}\in\{0,1\}$, *and let* $C_{n,m}\in\mathbb{N}_0$, $m,n\in\mathbb{N}_0$, *satisfy for all* $m,n\in\mathbb{N}$ *that* $C_{0,m}=0$ *and*

$$C_{n,m}\leq \mathfrak{v}m^n d+\mathfrak{f}+\sum_{\ell=1}^{n-1}\Big[m^{n-\ell}\Big(\mathfrak{v}(m^\ell d+1)+2\mathfrak{f}+2C_{\ell,m}+2C_{\ell-1,m}\Big)\Big]. \tag{19}$$

Then

(i) it holds for all $t\in[0,T]$ *that* $(\mathbb{E}[\|X(t)\|^2])^{1/2}\leq\Big[\|\xi\|+\|\mu(0,0)\|t+\sqrt{td}\Big]\,e^{Lt}$,

(ii) it holds for all $t\in[0,T]$, $m,n\in\mathbb{N}$ *that* $X^0_{n,m}(t)$ *is measurable and*

$$\Big(\mathbb{E}\Big[\big\|X^0_{n,m}(t)-X(t)\big\|^2\Big]\Big)^{1/2}\leq m^{-n/2}e^{m/2}\Big[\|\xi\|+\|\mu(0,0)\|t+\sqrt{Td}\Big]\,e^{Lt}\,(1+2Lt)^n, \tag{20}$$

and

(iii) there exists $\mathrm{n}=(\mathrm{n}_\varepsilon)_{\varepsilon\in(0,1)}\colon(0,1)\to\mathbb{N}$ *such that for all* $\delta,\varepsilon\in(0,1)$ *it holds that* $\sup_{k\in[\mathrm{n}_\varepsilon,\infty)\cap\mathbb{N}}\sup_{t\in[0,T]}\mathbb{E}\big[\|X^0_{k,k}(t)-X(t)\|^2\big]\leq\varepsilon^2$ *and*

$$C_{\mathrm{n}_\varepsilon,\mathrm{n}_\varepsilon}\varepsilon^{2+\delta}\leq(\mathfrak{v}d+\mathfrak{f})\sup_{k\in\mathbb{N}}\left[(4k+4)^{k+1}\left[\frac{e^{k/2}\big[1+\|\xi\|+\|\mu(0,0)\|T+\sqrt{Td}\big]e^{LT}(1+2LT)^k}{k^{k/2}}\right]^{2+\delta}\right]<\infty. \tag{21}$$

Proof of Theorem 3.1. Throughout this proof for every random variable $\mathfrak{X}\colon\Omega\to\mathbb{R}^d$ with $\mathbb{E}[\|\mathfrak{X}\|^2]<\infty$ and every σ-algebra $\mathcal{A}\subseteq\mathcal{F}$ let $\mathrm{Var}[\mathfrak{X}|\mathcal{A}]\colon\Omega\to[0,\infty)$ be a random variable which satisfies that a.s. it holds that $\mathrm{Var}[\mathfrak{X}|\mathcal{A}]=\mathbb{E}\big[\|\mathfrak{X}-\mathbb{E}[\mathfrak{X}|\mathcal{A}]\|^2|\mathcal{A}\big]$ and let $\mathcal{G}_{n,m}\subseteq\mathcal{F}$, $n\in\mathbb{N}_0$, $m\in\mathbb{N}$, be the σ-algebras which satisfy for all $n\in\mathbb{N}_0$, $m\in\mathbb{N}$ that $\mathcal{G}_{n,m}=\sigma\big(\{W^0(t),X^0_{\ell,m}(t)\colon\ell\in\{0,1,\dots,n\},t\in[0,T]\}\big)$.

This proof is organized as follows. In Step 1 we prove the upper bound of the exact solution X in (i). In Steps 2 and 3 we establish distributional, measurablility, and integrability properties for the MLP approximations in (18). In Step 4 we consider *the bias*. In Step 5 we consider *the statistical error*. In Step 6 we combine Steps 4 and 5 to obtain a recursive bound of the approximation error, which, together with a Gronwall-type inequality and the upper bound of the exact solution in (i), establishes (ii). In Step 7 we estimate the computational complexity and obtain (iii).

Step 1. We prove the upper bound of the exact solution X in (i). Observe that Jensen's inequality, the triangle inequality, and (16) show for all $s\in[0,T]$ that

$$\begin{aligned}&\left(\mathbb{E}\left[\left\|\int\mu(X(s),x)\mathbb{P}\big(X(s)\in dx\big)\right\|^2\right]\right)^{1/2}\leq\left(\mathbb{E}\Big[\int\|\mu(X(s),x)\|^2\,\mathbb{P}\big(X(s)\in dx\big)\Big]\right)^{1/2}\\&\leq\|\mu(0,0)\|+\tfrac{L}{2}\left(\mathbb{E}\big[\|X(s)\|^2\big]\right)^{1/2}+\tfrac{L}{2}\left(\mathbb{E}\big[\|X(s)\|^2\big]\right)^{1/2}\\&=\|\mu(0,0)\|+L\left(\mathbb{E}\big[\|X(s)\|^2\big]\right)^{1/2}.\end{aligned}\tag{22}$$

This, (17), the triangle inequality, and the fact that $\forall\,t\in[0,T]\colon\mathbb{E}\big[\|W^0(t)\|^2\big]=td$ prove for all $t \in [0,T]$ that

$$\begin{aligned}
\left(\mathbb{E}\left[\|X(t)\|^2\right]\right)^{1/2} &= \left(\mathbb{E}\left[\left\|\xi + \int_0^t \int \mu(X(s),x)\mathbb{P}\big(X(s)\in dx\big)\,ds + W^0(t)\right\|^2\right]\right)^{1/2} \\
&\le \|\xi\| + \int_0^t \left(\mathbb{E}\left[\left\|\int \mu(X(s),x)\mathbb{P}\big(X(s)\in dx\big)\right\|^2\right]\right)^{1/2} ds + \left(\mathbb{E}\left[\left\|W^0(t)\right\|^2\right]\right)^{1/2} \\
&\le \|\xi\| + \int_0^t \|\mu(0,0)\| + L\left(\mathbb{E}\left[\|X(s)\|^2\right]\right)^{1/2} ds + \sqrt{td} \\
&= \int_0^t L\left(\mathbb{E}\left[\|X(s)\|^2\right]\right)^{1/2} ds + \|\xi\| + \|\mu(0,0)\|t + \sqrt{td}.
\end{aligned} \tag{23}$$

This, the fact that $\forall\, m \in \mathbb{N}\colon X^0_{0,m} = 0$, Gronwall's lemma, and the fact that $\int_0^T \left(\mathbb{E}[\|X(t)\|^2]\right)^{1/2} dt < \infty$ prove for all $t \in [0,T]$ that

$$\left(\mathbb{E}\left[\left\|X^0_{0,m}(t) - X(t)\right\|^2\right]\right)^{1/2} = \left(\mathbb{E}\left[\|X(t)\|^2\right]\right)^{1/2} \le \left[\|\xi\| + \|\mu(0,0)\|t + \sqrt{td}\right] e^{Lt}. \tag{24}$$

This proves (i).

Step 2. We establish measurability and distributional properties. First, the assumptions on measurablity, (18), induction, and the fact that $\forall\, m \in \mathbb{N}, \theta \in \Theta\colon X^\theta_{0,m} = 0$ prove for all $m \in \mathbb{N}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$ that $X^\theta_{n,m}$ is measurable. Next, the fact that $\forall\, m \in \mathbb{N}, \theta \in \Theta\colon X^\theta_{0,m} = 0$, (18), and induction prove for all $n \in \mathbb{N}_0$, $m \in \mathbb{N}$, $\theta \in \Theta$ that

$$\begin{aligned}
&\sigma\big(\{W^\theta(t), X^\theta_{\ell,m}(t)\colon \ell \in \{0,1,\ldots,n\}, t \in [0,T]\}\big) \\
&\subseteq \sigma\big(\{W^\theta(t), W^{(\theta,i,\nu)}(t), \mathfrak{u}^{(\theta,i,\nu)}\colon i \in \{0,1,\ldots,n\}, \nu \in \Theta, t \in [0,T]\}\big).
\end{aligned} \tag{25}$$

This and the fact that $\forall\, m \in \mathbb{N}, \theta \in \Theta\colon X^\theta_{0,m} = 0$ prove for all $n,m \in \mathbb{N}$, $\theta \in \Theta$, $k,\ell \in \mathbb{N}$, $j \in \{\ell-1,\ell\}$ that

$$\begin{aligned}
&\sigma\Big(\Big\{X^{(\theta,n,k,\ell)}_{j,m}(t)\colon t \in [0,T]\Big\}\Big) \\
&\subseteq \sigma\big(\{W^{(\theta,n,k,\ell)}(t), W^{(\theta,n,k,\ell,\nu)}(t), \mathfrak{u}^{(\theta,n,k,\ell,\nu)}\colon \nu \in \Theta, t \in [0,T]\}\big).
\end{aligned} \tag{26}$$

This, (25), and the independence assumptions show for all $n,m \in \mathbb{N}$, $\theta \in \Theta$ that

$$(W^\theta, (X^\theta_{j,m})_{j\in\{0,1,\ldots,n-1\}}), \quad (X^{(\theta,n,k,\ell)}_{\ell,m}, X^{(\theta,n,k,\ell)}_{\ell-1,m}), \quad \mathfrak{u}^{(\theta,n,k,\ell)}, \quad k,\ell \in \mathbb{N}, \tag{27}$$

are independent. This, the fact that $\forall\, m \in \mathbb{N}, \theta \in \Theta\colon X^\theta_{0,m} = 0$, (18), the disintegration theorem (see, e.g., [14, Lemma 2.2]), and induction show for all $m \in \mathbb{N}$, $n \in \mathbb{N}_0$ that $(W^\theta, (X^\theta_{\ell,m})_{\ell\in\{0,1,\ldots,n\}})$, $\theta \in \Theta$, are identically distributed.

Step 3. We establish that the approximations are square-integrable. Observe that the triangle inequality, (16), distributional properties (see Step 2), and the disintegration theorem (see, e.g., [14, Lemma 2.2]) prove for all $n,m,k \in \mathbb{N}$, $\ell \in \{1,2,\ldots,n-1\}$, $j \in \{\ell-1,\ell\}$, $t \in [0,T]$ that

$$t\left(\mathbb{E}\left[\left\|\mu\big(X^0_{j,m}(t\mathfrak{u}^{(0,n,k,\ell)}), X^{(0,n,k,\ell)}_{j,m}(t\mathfrak{u}^{(0,n,k,\ell)})\big)\right\|^2\right]\right)^{1/2}$$

$$\leq t\,\|\mu(0,0)\| + \tfrac{tL}{2}\left(\mathbb{E}\Big[\big\|X^0_{j,m}(t\mathfrak{u}^{(0,n,k,\ell)})\big\|^2\Big]\right)^{1/2} + \tfrac{tL}{2}\left(\mathbb{E}\Big[\big\|X^{(0,n,k,\ell)}_{j,m}(t\mathfrak{u}^{(0,n,k,\ell)})\big\|^2\Big]\right)^{1/2}$$
$$= t\,\|\mu(0,0)\| + L\sqrt{t}\left(t\mathbb{E}\Big[\big\|X^0_{j,m}(t\mathfrak{u}^{(0,n,k,\ell)})\big\|^2\Big]\right)^{1/2}$$
$$= t\,\|\mu(0,0)\| + L\sqrt{t}\left(\int_0^t \mathbb{E}\Big[\big\|X^0_{j,m}(s)\big\|^2\Big]\,ds\right)^{1/2}. \tag{28}$$

This, the fact that $\forall\, m\in\mathbb{N},\theta\in\Theta\colon X^\theta_{0,m}=0$, (18), the triangle inequality, and induction yield for all $n,m,k\in\mathbb{N}$, $\ell\in\{1,2,\dots,n-1\}$, $j\in\{\ell-1,\ell\}$ that

$$\sup_{t\in[0,T]}\left[\left(\mathbb{E}\Big[\|X^0_{n,m}(t)\|^2\Big]\right)^{1/2} + t\left(\mathbb{E}\Big[\big\|\mu\big(X^0_{j,m}(t\mathfrak{u}^{(0,n,k,\ell)}),X^{(0,n,k,\ell)}_{j,m}(t\mathfrak{u}^{(0,n,k,\ell)})\big)\big\|^2\Big]\right)^{1/2}\right]<\infty. \tag{29}$$

Step 4. We consider the bias. Observe that (18), (29), linearity of conditional expectations, the disintegration theorem (see, e.g., [14, Lemma 2.2]), distributional properties (cf. Step 2), a telescoping sum argument, the fact that $\forall\, m\in\mathbb{N}\colon X^0_{0,m}=0$, and the substitution rule imply that for all $n,m\in\mathbb{N}$, $t\in[0,T]$ it holds a.s. that

$$\mathbb{E}\Big[X^0_{n,m}(t)-\xi-W^0\left(\sup(\{\tfrac{kT}{m^n}\colon k\in\mathbb{N}_0\}\cap[0,t])\right)\Big|\mathcal{G}_{n-1,m}\Big]$$
$$= t\mu(0,0) + \sum_{\ell=1}^{n-1}\Bigg(\frac{t}{m^{n-\ell}}\sum_{k=1}^{m^{n-\ell}}\Bigg(\mathbb{E}\Big[\mu\big(X^0_{\ell,m}(\mathfrak{u}^{(0,n,k,\ell)}t),X^{(0,n,k,\ell)}_{\ell,m}(\mathfrak{u}^{(0,n,k,\ell)}t)\big)\Big|\mathcal{G}_{n-1,m}\Big]$$
$$- \mathbb{E}\Big[\mu\big(X^0_{\ell-1,m}(\mathfrak{u}^{(0,n,k,\ell)}t),X^{(0,n,k,\ell)}_{\ell-1,m}(\mathfrak{u}^{(0,n,k,\ell)}t)\big)\Big|\mathcal{G}_{n-1,m}\Big]\Bigg)\Bigg)$$
$$= t\mu(0,0) + \sum_{\ell=1}^{n-1}\Bigg(\frac{1}{m^{n-\ell}}\sum_{k=1}^{m^{n-\ell}}\int_0^t \mathbb{E}\Big[\mu\big(X^0_{\ell,m}(s),X^{(0,n,k,\ell)}_{\ell,m}(s)\big)\Big|\mathcal{G}_{n-1,m}\Big]$$
$$- \mathbb{E}\Big[\mu\big(X^0_{\ell-1,m}(s),X^{(0,n,k,\ell)}_{\ell-1,m}(s)\big)\Big|\mathcal{G}_{n-1,m}\Big]\,ds\Bigg)$$
$$= t\mu(0,0) + \sum_{\ell=1}^{n-1}\int_0^t \int \mu\big(X^0_{\ell,m}(s),y\big)\mathbb{P}\big(X^0_{\ell,m}(s)\in dy\big)$$
$$- \int \mu\big(X^0_{\ell-1,m}(s),y\big)\mathbb{P}\big(X^0_{\ell-1,m}(s)\in dy\big)\,ds$$
$$= \int_0^t \int \mu\big(X^0_{n-1,m}(s),y\big)\mathbb{P}\big(X^0_{n-1,m}(s)\in dy\big)\,ds. \tag{30}$$

Next, (17) shows that for all $t\in[0,T]$ it holds a.s. that

$$\mathbb{E}\big[X(t)-\xi-W^0(t)\big|\mathcal{G}_{n-1,m}\big] = X(t)-\xi-W^0(t) = \int_0^t \int \mu(X(s),x)\mathbb{P}\big(X(s)\in dx\big)\,ds. \tag{31}$$

This, the triangle inequality, the fact that $\forall\, s,t\in[0,T]\colon \mathbb{E}\big[\|W^0(t)-W^0(s)\|^2\big]=d|t-s|$, (30), Jensen's inequality, (16), and Tonelli's theorem show for all $n,m\in\mathbb{N}$, $t\in[0,T]$ that

$$\left(\mathbb{E}\Big[\|\mathbb{E}[X^0_{n,m}(t)-X(t)|\mathcal{G}_{n-1,m}]\|^2\Big]\right)^{1/2} \leq \left(\mathbb{E}\Big[\big\|W^0\left(\sup(\{\tfrac{kT}{m^n}\colon k\in\mathbb{N}_0\}\cap[0,t])\right)-W^0(t)\big\|^2\Big]\right)^{1/2}$$
$$+\left(\mathbb{E}\Big[\Big\|\mathbb{E}\Big[\Big(X^0_{n,m}(t)-\xi-W^0\left(\sup(\{\tfrac{kT}{m^n}\colon k\in\mathbb{N}_0\}\cap[0,t])\right)\Big)-\Big(X(t)-\xi-W^0(t)\Big)\Big|\mathcal{G}_{n-1,m}\Big]\Big\|^2\Big]\right)^{1/2}$$

$$\leq \tfrac{\sqrt{Td}}{\sqrt{m^n}} + \left(\mathbb{E}\left[\left\|\int_0^t \int \left[\mu\big(X^0_{n-1,m}(s),y\big) - \mu\big(X(s),x\big)\right] \mathbb{P}\big(X^0_{n-1,m}(s)\in dy, X(s)\in dx\big)\, ds\right\|^2\right]\right)^{1/2}$$
$$\leq \tfrac{\sqrt{Td}}{\sqrt{m^n}} + \left(\mathbb{E}\left[t\int_0^t \int \left\|\mu\big(X^0_{n-1,m}(s),y\big) - \mu\big(X(s),x\big)\right\|^2 \mathbb{P}\big(X^0_{n-1,m}(s)\in dy, X(s)\in dx\big)\, ds\right]\right)^{1/2}$$
$$\leq \tfrac{\sqrt{Td}}{\sqrt{m^n}} + \tfrac{L\sqrt{t}}{2}\left(\mathbb{E}\left[\int_0^t \left\|X^0_{n-1,m}(s) - X(s)\right\|^2 ds\right]\right)^{1/2} + \tfrac{L\sqrt{t}}{2}\left(\mathbb{E}\left[\int_0^t \mathbb{E}\left[\left\|X^0_{n-1,m}(s) - X(s)\right\|^2\right] ds\right]\right)^{1/2}$$
$$= \tfrac{\sqrt{Td}}{\sqrt{m^n}} + L\sqrt{t}\left(\int_0^t \mathbb{E}\left[\left\|X^0_{n-1,m}(s) - X(s)\right\|^2\right] ds\right)^{1/2}. \tag{32}$$

Step 5. We consider the statistical error. Distributional properties (cf. Step 2) imply for all $n,m\in\mathbb{N}$, $\ell\in\{1,2,\ldots,n-1\}$ that

a) it holds for all $k\in\mathbb{N}$ that $(X^{(0,n,k,\ell)}_{\ell,m}, X^{(0,n,k,\ell)}_{\ell-1,m}, \mathfrak{u}^{(0,n,k,\ell)})$ and $\mathcal{G}_{n-1,m}$ are independent,

b) it holds that $(X^{(0,n,k,\ell)}_{\ell,m}, X^{(0,n,k,\ell)}_{\ell-1,m}, \mathfrak{u}^{(0,n,k,\ell)})$, $k\in\mathbb{N}$, are i.i.d., and

c) it holds that $(X^{(0,n,1,\ell)}_{\ell,m}, X^{(0,n,1,\ell)}_{\ell-1,m})$ and $(X^0_{\ell,m}, X^0_{\ell-1,m})$ are identically distributed.

This, (18), the triangle inequality, Biennaymé's identity, the assumptions on distributions, and the disintegration theorem (see, e.g., [14, Lemma 2.2]) prove that for all $n,m\in\mathbb{N}$, $t\in[0,T]$ it holds a.s. that

$$\begin{aligned}
\left(\mathrm{Var}\left[X^0_{n,m}(t)\middle|\mathcal{G}_{n-1,m}\right]\right)^{1/2} &= \Bigg(\mathrm{Var}\Bigg[\sum_{\ell=1}^{n-1}\Bigg[\tfrac{t}{m^{n-\ell}}\sum_{k=1}^{m^{n-\ell}}\Big[\mu\big(X^0_{\ell,m}(\mathfrak{u}^{(0,n,k,\ell)}t), X^{(0,n,k,\ell)}_{\ell,m}(\mathfrak{u}^{(0,n,k,\ell)}t)\big)\\
&\qquad - \mu\big(X^0_{\ell-1,m}(\mathfrak{u}^{(0,n,k,\ell)}t), X^{(0,n,k,\ell)}_{\ell-1,m}(\mathfrak{u}^{(0,n,k,\ell)}t)\big)\Big]\Bigg]\Bigg|\mathcal{G}_{n-1,m}\Bigg]\Bigg)^{1/2}\\
&\leq \sum_{\ell=1}^{n-1}\Bigg[\tfrac{t}{m^{n-\ell}}\Bigg(\mathrm{Var}\Bigg[\sum_{k=1}^{m^{n-\ell}}\Big[\mu\big(X^0_{\ell,m}(\mathfrak{u}^{(0,n,k,\ell)}t), X^{(0,n,k,\ell)}_{\ell,m}(\mathfrak{u}^{(0,n,k,\ell)}t)\big)\\
&\qquad - \mu\big(X^0_{\ell-1,m}(\mathfrak{u}^{(0,n,k,\ell)}t), X^{(0,n,k,\ell)}_{\ell-1,m}(\mathfrak{u}^{(0,n,k,\ell)}t)\big)\Big]\Bigg|\mathcal{G}_{n-1,m}\Bigg]\Bigg)^{1/2}\Bigg]\\
&\leq \sum_{\ell=1}^{n-1}\Bigg[\tfrac{\sqrt{t}}{\sqrt{m^{n-\ell}}}\Bigg(t\mathbb{E}\Bigg[\Big\|\mu\big(X^0_{\ell,m}(\mathfrak{u}^{(0,n,1,\ell)}t), X^{(0,n,1,\ell)}_{\ell,m}(\mathfrak{u}^{(0,n,1,\ell)}t)\big)\\
&\qquad - \mu\big(X^0_{\ell-1,m}(\mathfrak{u}^{(0,n,1,\ell)}t), X^{(0,n,1,\ell)}_{\ell-1,m}(\mathfrak{u}^{(0,n,1,\ell)}t)\big)\Big\|^2\Bigg|\mathcal{G}_{n-1,m}\Bigg]\Bigg)^{1/2}\Bigg]
\end{aligned}\tag{33}$$

and

$$\begin{aligned}
&\left(\mathrm{Var}\left[X^0_{n,m}(t)\middle|\mathcal{G}_{n-1,m}\right]\right)^{1/2}\\
&\leq \sum_{\ell=1}^{n-1}\Bigg[\tfrac{\sqrt{t}}{\sqrt{m^{n-\ell}}}\Bigg(\int_0^t\int \left\|\mu\big(X^0_{\ell,m}(s),x\big) - \mu\big(X^0_{\ell-1,m}(s),y\big)\right\|^2\\
&\qquad\qquad \mathbb{P}\Big(X^{(0,n,1,\ell)}_{\ell,m}(s)\in dx, X^{(0,n,1,\ell)}_{\ell-1,m}(s)\in dy\Big)\, ds\Bigg)^{1/2}\Bigg]\\
&= \sum_{\ell=1}^{n-1}\Bigg[\tfrac{\sqrt{t}}{\sqrt{m^{n-\ell}}}\Bigg(\int_0^t\int \left\|\mu\big(X^0_{\ell,m}(s),x\big) - \mu\big(X^0_{\ell-1,m}(s),y\big)\right\|^2
\end{aligned}$$

$$\mathbb{P}\big(X^0_{\ell,m}(s)\in dx, X^0_{\ell-1,m}(s)\in dy\big)\,ds\bigg)^{1/2}\Bigg]$$

$$\leq \sum_{\ell=1}^{n-1}\Bigg[\tfrac{L\sqrt{t}}{2\sqrt{m^{n-\ell}}}\bigg(\int_0^t \big\|X^0_{\ell,m}(s)-X^0_{\ell-1,m}(s)\big\|^2\,ds\bigg)^{1/2}$$

$$+\tfrac{L\sqrt{t}}{2\sqrt{m^{n-\ell}}}\bigg(\int_0^t \mathbb{E}\Big[\big\|X^0_{\ell,m}(s)-X^0_{\ell-1,m}(s)\big\|^2\Big]\,ds\bigg)^{1/2}\Bigg]. \tag{34}$$

This, the tower property, the definition of conditional variances, the triangle inequality, Jensen's inequality, and Tonelli's theorem show for all $n,m\in\mathbb{N}$, $t\in[0,T]$ that

$$\begin{aligned}&\Big(\mathbb{E}\Big[\big\|X^0_{n,m}(t)-\mathbb{E}\big[X^0_{n,m}(t)\big|\mathcal{G}_{n-1,m}\big]\big\|^2\Big]\Big)^{1/2}\\&=\Big(\mathbb{E}\Big[\mathbb{E}\Big[\big\|X^0_{n,m}(t)-\mathbb{E}\big[X^0_{n,m}(t)\big|\mathcal{G}_{n-1,m}\big]\big\|^2\Big|\mathcal{G}_{n-1,m}\Big]\Big]\Big)^{1/2}=\Big(\mathbb{E}\big[\mathrm{Var}\big[X^0_{n,m}(t)\big|\mathcal{G}_{n-1,m}\big]\big]\Big)^{1/2}\\&\leq\sum_{\ell=1}^{n-1}\Bigg[\frac{L\sqrt{t}}{\sqrt{m^{n-\ell}}}\bigg(\int_0^t\mathbb{E}\Big[\big\|X^0_{\ell,m}(s)-X^0_{\ell-1,m}(s)\big\|^2\Big]ds\bigg)^{1/2}\Bigg]\\&\leq\sum_{\ell=0}^{n-1}\frac{(2-\mathbb{1}_{\{n-1\}}(\ell))L\sqrt{t}}{\sqrt{m^{n-\ell-1}}}\bigg(\int_0^t\mathbb{E}\Big[\big\|X^0_{\ell,m}(s)-X(s)\big\|^2\Big]ds\bigg)^{1/2}.\end{aligned}\tag{35}$$

Step 6. We now prove (ii). Observe that the definition of $\mathcal{G}_{n,m}$, $n\in\mathbb{N}_0$, $m\in\mathbb{N}$, and the fact that X is $(\sigma(\{W^0(s)\colon s\in[0,t]\}))_{t\in[0,T]}$-adapted show for all $n,m\in\mathbb{N}$, $t\in[0,T]$ that

$$X^0_{n,m}(t)-X(t)=\mathbb{E}\big[X^0_{n,m}(t)-X(t)\big|\mathcal{G}_{n-1,m}\big]+X^0_{n,m}(t)-\mathbb{E}\big[X^0_{n,m}(t)\big|\mathcal{G}_{n-1,m}\big]. \tag{36}$$

This, the triangle inequality, (32), and (35) show for all $n,m\in\mathbb{N}$, $t\in[0,T]$ that

$$\begin{aligned}&\Big(\mathbb{E}\Big[\big\|X^0_{n,m}(t)-X(t)\big\|^2\Big]\Big)^{1/2}\\&\leq\Big(\mathbb{E}\Big[\big\|\mathbb{E}\big[X^0_{n,m}(t)-X(t)\big|\mathcal{G}_{n-1,m}\big]\big\|^2\Big]\Big)^{1/2}+\Big(\mathbb{E}\Big[\big\|X^0_{n,m}(t)-\mathbb{E}\big[X^0_{n,m}(t)\big|\mathcal{G}_{n-1,m}\big]\big\|^2\Big]\Big)^{1/2}\\&\leq\tfrac{\sqrt{Td}}{\sqrt{m^n}}+L\sqrt{t}\bigg(\int_0^t\mathbb{E}\Big[\big\|X^0_{n-1,m}(s)-X(s)\big\|^2\Big]ds\bigg)^{1/2}\\&\quad+\sum_{\ell=0}^{n-1}\Bigg[\tfrac{(2-\mathbb{1}_{\{n-1\}}(\ell))L\sqrt{t}}{\sqrt{m^{n-\ell-1}}}\bigg(\int_0^t\mathbb{E}\Big[\big\|X^0_{\ell,m}(s)-X(s)\big\|^2\Big]ds\bigg)^{1/2}\Bigg]\\&=\tfrac{\sqrt{Td}}{\sqrt{m^n}}+\sum_{\ell=0}^{n-1}\Bigg[\tfrac{2L\sqrt{t}}{\sqrt{m^{n-\ell-1}}}\bigg(\int_0^t\mathbb{E}\Big[\big\|X^0_{\ell,m}(s)-X(s)\big\|^2\Big]ds\bigg)^{1/2}\Bigg].\end{aligned}\tag{37}$$

This, [13, Lemma 3.9] (applied for every $m,N\in\mathbb{N}$, $t\in[0,T]$ with $a\curvearrowleft\sqrt{Td}$, $b\curvearrowleft 2L\sqrt{t}$, $c\curvearrowleft 1/\sqrt{m}$, $\alpha\curvearrowleft 0$, $\beta\curvearrowleft t$, $p\curvearrowleft 2$, $(f_n)_{n\in\mathbb{N}_0}\curvearrowleft\big([0,t]\ni s\mapsto\big(\mathbb{E}\big[\|X^0_{n,m}(s)-X(s)\|^2\big]\big)^{1/2}\in[0,\infty]\big)_{n\in\mathbb{N}_0}$ in the notation of [13, Lemma 3.9]), (24), and the fact that for all $m,N\in\mathbb{N}$ it holds that

$$\max_{k\in\{0,1,\ldots,N\}}\tfrac{1}{\sqrt{m^{N-k}k!}}=m^{-N/2}\max_{k\in\{0,1,\ldots,N\}}\tfrac{\sqrt{m^k}}{\sqrt{k!}}\leq m^{-N/2}e^{m/2} \tag{38}$$

show for all $t \in [0,T]$, $m, N \in \mathbb{N}$ that

$$\begin{aligned}
&\left(\mathbb{E}\Big[\big\|X^0_{N,m}(t) - X(t)\big\|^2\Big]\right)^{1/2} \leq \left[\sqrt{Td} + 2L\sqrt{t}\cdot\sqrt{t}\sup_{s\in[0,t]}\left(\mathbb{E}\Big[\big\|X^0_{0,m}(s) - X(s)\big\|^2\Big]\right)^{1/2}\right] \\
&\qquad\cdot\left[\max_{k\in\{0,1,\ldots,N\}}\tfrac{1}{\sqrt{m^{N-k}k!}}\right]\left(1 + 2L\sqrt{t}\cdot\sqrt{t}\right)^{N-1} \\
&\leq \left[\sqrt{Td} + 2Lt\left[\|\xi\| + \|\mu(0,0)\|t + \sqrt{td}\right]e^{Lt}\right]m^{-N/2}e^{m/2}\left(1+2Lt\right)^{N-1} \\
&\leq \left[\sqrt{Td} + 2Lt\left[\|\xi\| + \|\mu(0,0)\|t + \sqrt{td}\right]\right]e^{Lt}m^{-N/2}e^{m/2}\left(1+2Lt\right)^{N-1} \\
&\leq m^{-N/2}e^{m/2}\left[\|\xi\| + \|\mu(0,0)\|t + \sqrt{Td}\right]e^{Lt}\left(1+2Lt\right)^{N}.
\end{aligned} \tag{39}$$

This proves (ii).

Step 7. We estimate the computational complexity. Let $\mathrm{n} = (\mathrm{n}_\varepsilon)_{\varepsilon\in(0,1)}\colon (0,1) \to \mathbb{N}\cup\{\infty\}$ satisfy for all $\varepsilon \in (0,1)$ that

$$\mathrm{n}_\varepsilon = \inf\left(\left\{n\in\mathbb{N}\colon \sup_{k\in[n,\infty)\cap\mathbb{N}}\sup_{t\in[0,T]}\left(\mathbb{E}\Big[\big\|X^0_{k,k}(t) - X(t)\big\|^2\Big]\right)^{1/2} < \varepsilon\right\}\cup\{\infty\}\right). \tag{40}$$

This, (ii), and the fact that $\forall\,\delta \in (0,1), s,t \in (0,\infty)\colon \lim_{n\to\infty}(s^n n^t n^{-n/2}) = 0$ prove for all $\varepsilon \in (0,1)$ that $\lim_{n\to\infty}\sup_{t\in[0,T]}\mathbb{E}\big[\|X^0_{n,n}(t) - X(t)\|^2\big] = 0$ and $\mathrm{n}_\varepsilon \in \mathbb{N}$. Next, (19) and the fact that $\forall\,\ell, m \in \mathbb{N}\colon m^{-\ell} \leq 1$ imply for all $m, n \in \mathbb{N}$ that $C_{0,m} = 0$ and

$$\begin{aligned}
m^{-n}C_{n,m} &\leq \mathfrak{v}d + \mathfrak{f}m^{-n} + \sum_{\ell=1}^{n-1}\Big[m^{-\ell}\Big(\mathfrak{v}(m^\ell d + 1) + 2\mathfrak{f} + 2C_{\ell,m} + 2C_{\ell-1,m}\Big)\Big] \\
&\leq 2n(\mathfrak{v}d + \mathfrak{f}) + \sum_{\ell=0}^{n-1}\Big(2m^{-\ell}C_{\ell,m} + \mathbb{1}_{\mathbb{N}}(\ell)2m^{-(\ell-1)}C_{|\ell-1|,m}\Big).
\end{aligned} \tag{41}$$

This, Corollary 2.3 (applied for every $m \in \mathbb{N}$ with $(a_n)_{n\in\mathbb{N}_0} \curvearrowleft (m^{-n}C_{n,m})_{n\in\mathbb{N}_0}$, $\kappa \curvearrowleft 2$, $\lambda \curvearrowleft 2$, $c_1 \curvearrowleft 0$, $c_2 \curvearrowleft 2(\mathfrak{v}d + \mathfrak{f})$, $c_3 \curvearrowleft 0$, $c_4 \curvearrowleft 1$, $\beta \curvearrowleft \frac{(1+2)+\sqrt{(1+2)^2+4\cdot 2}}{2}$ in the notation of Corollary 2.3), and the fact that $\frac{(1+2)+\sqrt{(1+2)^2+4\cdot 2}}{2} \leq 4$ imply for all $n, m \in \mathbb{N}$ that $m^{-n}C_{n,m} \leq 1.5\cdot 2(\mathfrak{v}d + \mathfrak{f})\frac{4^n-1}{4-1} \leq (\mathfrak{v}d+\mathfrak{f})4^n$. This and (ii) imply for all $n, m \in \mathbb{N}$ that $C_{n,m} \leq (\mathfrak{v}d+\mathfrak{f})(4m)^n$ and

$$\begin{aligned}
&C_{n+1,n+1}\sup_{t\in[0,T]}\left(\mathbb{E}\Big[\big\|X^0_{n,n}(t) - X(t)\big\|^2\Big]\right)^{\frac{2+\delta}{2}} \\
&\leq (\mathfrak{v}d+\mathfrak{f})(4n+4)^{n+1}\left[n^{-n/2}e^{n/2}\left[\|\xi\| + \|\mu(0,0)\|T + \sqrt{dT}\right]e^{LT}\left(1+2LT\right)^n\right]^{2+\delta} \\
&\leq (\mathfrak{v}d+\mathfrak{f})\sup_{k\in\mathbb{N}}\left[(4k+4)^{k+1}\left[k^{-k/2}e^{k/2}\left[\|\xi\| + \|\mu(0,0)\|T + \sqrt{dT}\right]e^{LT}\left(1+2LT\right)^k\right]^{2+\delta}\right].
\end{aligned} \tag{42}$$

This and the fact that $\forall\,\delta \in (0,1), s,t \in (0,\infty)\colon \lim_{n\to\infty}(s^n(n+1)^t n^{-n\delta/2}) = 0$ imply for all $\delta, \varepsilon \in (0,1)$ that

$$\begin{aligned}
&C_{\mathrm{n}_\varepsilon,\mathrm{n}_\varepsilon}\varepsilon^{2+\delta} \leq \mathbb{1}_{\{1\}}(\mathrm{n}_\varepsilon)C_{1,1} + \mathbb{1}_{[2,\infty)}(\mathrm{n}_\varepsilon)C_{\mathrm{n}_\varepsilon,\mathrm{n}_\varepsilon}\sup_{t\in[0,T]}\left(\mathbb{E}\Big[\big\|X^0_{\mathrm{n}_\varepsilon-1,\mathrm{n}_\varepsilon-1}(t) - X(t)\big\|^2\Big]\right)^{\frac{2+\delta}{2}} \\
&\leq (\mathfrak{v}d+\mathfrak{f})\sup_{k\in\mathbb{N}}\left[(4k+4)^{k+1}\left[\tfrac{e^{k/2}\left[1+\|\xi\|+\|\mu(0,0)\|T+\sqrt{Td}\right]e^{LT}(1+2LT)^k}{k^{k/2}}\right]^{2+\delta}\right] < \infty.
\end{aligned} \tag{43}$$

This proves (iii). The proof of Theorem 3.1 is thus completed. □

Acknowledgements

We thank Arnulf Jentzen for very helpful comments and suggestions. This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the research grant HU1889/6-2.

References

[1] Antonelli, F., and Kohatsu-Higa, A. Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *The Annals of Applied Probability 12*, 2 (2002), 423–476.

[2] Beck, C., Hornung, F., Hutzenthaler, M., Jentzen, A., and Kruse, T. Overcoming the curse of dimensionality in the numerical approximation of Allen-Cahn partial differential equations via truncated full-history recursive multilevel Picard approximations. *Journal of Numerical Mathematics 28*, 4 (2020), 197–222.

[3] Beck, C., Hutzenthaler, M., Jentzen, A., and Magnani, E. Full history recursive multilevel picard approximations for ordinary differential equations with expectations. *arXiv:2103.02350* (2021).

[4] Belomestny, D., Szpruch, L., and Tan, S. Iterative Multilevel density estimation for McKean-Vlasov SDEs via projections. *arXiv:1909.11717v1* (2019).

[5] Bossy, M., and Jourdain, B. Rate of convergeance of a particle method for the solution of a 1D viscous scalar conservation law in a bounded interval. *The Annals of Probability 30*, 4 (2002), 1797–1832.

[6] Bossy, M., and Talay, D. Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation. *The Annals of Applied Probability 6*, 3 (1996), 818–861.

[7] E, W., Hutzenthaler, M., Jentzen, A., and Kruse, T. Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. *arXiv:1607.03295* (2016). Accepted in Springer Nature Partial Differential Equations and Applications.

[8] Gärtner, J. On the McKean-Vlasov limit for interacting diffusions. *Math. Nachr. 137* (1988), 197–248.

[9] Giles, M. B. Multilevel Monte Carlo path simulation. *Oper. Res. 56*, 3 (2008), 607–617.

[10] Heinrich, S. Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing*, vol. 2179 of *Lect. Notes Comput. Sci.* Springer, Berlin, 2001, pp. 58–67.

[11] Hutzenthaler, M. Interacting diffusions and trees of excursions: convergence and comparison. *Electron. J. Probab. 17*, 71 (2012), 1–49.

[12] Hutzenthaler, M., Jentzen, A., and Kruse, T. Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities. *arXiv:1912.02571* (2019). Accepted in Foundations of Computational Mathematics.

[13] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. Multilevel Picard approximations of high-dimensional semilinear partial differential equations with gradient-independent nonlinearities. *arXiv:2009.02484v4* (2020).

[14] Hutzenthaler, M., Jentzen, A., Kruse, T., Nguyen, T. A., and von Wurstemberger, P. Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *Proceeding of the Royal Society A 476*, 20190630 (2020).

[15] Hutzenthaler, M., and Pieper, D. Propagation of chaos and the many-demes limit for weakly interacting diffusions in the sparse regime. *Annals of Applied Probability 30*, 5 (2020), 2311–2354.

[16] Kac, M. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III* (Berkeley and Los Angeles, 1956), University of California Press, pp. 171–197.

[17] Lemaire, V., and Pagès, G. Multilevel Richardson-Romberg extrapolation. *Bernoulli 23*, 4A (2017), 2643–2692.

[18] McKean, Jr., H. P. Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*. Air Force Office Sci. Res., Arlington, VA, 1967, pp. 41–57.

[19] McKean Jr, H. P. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America 56*, 6 (1966), 1907.

[20] Oelschläger, K. A law of large numbers for moderately interacting diffusion processes. *Z. Wahrsch. Verw. Gebiete 69*, 2 (1985), 279–322.

[21] Sukharev, A. G. Optimal numerical integration formulas for some classes of functions of several variables. In *Doklady Akademii Nauk* (1979), vol. 246, Russian Academy of Sciences, pp. 282–285.

[22] Sznitman, A.-S. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, vol. 1464 of *Lecture Notes in Math.* Springer, Berlin, 1991, pp. 165–251.

[23] Szpruch, L., Tan, S., and Tse, A. Iterative Multilevel Particle Approximation for McKean-Vlasov SDEs. *Annals of Applied Probability 29*, 4 (2019), 2230–2265.

[24] Szpruch, L., and Tse, A. Antithetic multilevel particle system sampling method for McKean-Vlasov SDEs. *arXiv:1903.07063v2* (2019).