

Milstein schemes and antithetic multilevel Monte Carlo sampling for delay McKean–Vlasov equations and interacting particle systems

JIANHAI BAO

Center for Applied Mathematics, Tianjin University, 300072 Tianjin, China

CHRISTOPH REISINGER

Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Road, Oxford, OX2 6GG, UK

PANPAN REN

Mathematics Department, City University of Hong Kong, 83 Tat Chee Ave, Kowloon Tong, Hongkong

AND

WOLFGANG STOCKINGER*

Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Road, Oxford, OX2 6GG, UK

*Corresponding author: wolfgang.stockinger@maths.ox.ac.uk.

[Received on 29 November 2021; revised on 6 July 2023]

In this paper, we first derive Milstein schemes for an interacting particle system associated with point delay McKean–Vlasov stochastic differential equations, possibly with a drift term exhibiting super-linear growth in the state component. We prove strong convergence of order one and moment stability, making use of techniques from variational calculus on the space of probability measures with finite second-order moments. Then, we introduce an antithetic multilevel Milstein scheme, which leads to optimal complexity estimators for expected functionals of solutions to delay McKean–Vlasov equations without the need to simulate Lévy areas.

1. Introduction

A McKean–Vlasov equation (introduced by McKean, 1966) for a d -dimensional process $X = (X(t))_{t \in [0, T]}$, for some given $T > 0$, is a stochastic differential equation (SDE) where the underlying coefficients depend on the current state $X(t)$ and, additionally, on the law of $X(t)$, i.e.,

$$dX(t) = b(X(t), \mathcal{L}_{X(t)}) dt + \sigma(X(t), \mathcal{L}_{X(t)}) dW(t), \quad X_0 = \xi, \quad (1.1)$$

where $W = (W(t))_{t \in [0, T]}$ is an m -dimensional standard Brownian motion, $\mathcal{L}_{X(t)}$ denotes the marginal law of the process X at time $t \in [0, T]$ and ξ is an \mathbb{R}^d -valued random variable. The existence and uniqueness theory for strong solutions of McKean–Vlasov SDEs with coefficients of linear growth and Lipschitz type conditions (with respect to the state and the measure) is well-established (see, e.g., Sznitman, 1991; Méléard, 1996). For further existence and uniqueness results on weak and strong solutions of McKean–Vlasov SDEs, we refer to Bauer *et al.* (2018); Chaudrude Raynal *et al.* (2018); Lacker (2018); Hammersley *et al.* (2021); Mishura & Veretennikov (2021) and references cited therein. It is also known from Reis *et al.* (2019) that McKean–Vlasov SDEs with super-linear growth of the drift term with respect

to the state variable admit a unique strong solution provided a one-sided Lipschitz condition holds. Existence and uniqueness for path-dependent McKean–Vlasov SDEs (where the drift has super-linear growth) were studied in [Huang et al. \(2017\)](#).

An illustrative example for path-dependent McKean–Vlasov SDEs, which will be the focus of this work, is the one-point delay McKean–Vlasov SDE

$$dX(t) = b(X(t), X(t - \tau), \mathcal{L}_{X(t)}, \mathcal{L}_{X(t-\tau)}) dt + \sigma(X(t), X(t - \tau), \mathcal{L}_{X(t)}, \mathcal{L}_{X(t-\tau)}) dW(t),$$

with $\tau > 0$ a given deterministic delay parameter and the initial datum ξ is assumed to be a continuous (random) function $\xi : \Omega \times [-\tau, 0] \rightarrow \mathbb{R}^d$.

Such (McKean–Vlasov) SDEs with a time delay are relevant for applications, for instance, in medicine, biology and finance, where the effect of some actions or causes is not immediately visible or significant. For example, infectious diseases have a certain incubation time, so that the currently observed and infectious cases were themselves infected at a prior time. Furthermore, neural networks can be modelled using delay mean-field equations to account for the delay in the transmission of signals amongst large numbers of neurons; see [Touboul \(2012\)](#) for details. In finance, a collapse of a bank or an investment might not immediately have an impact on the entire system, but only after some time lag. Delay McKean–Vlasov SDEs are also employed in computational finance to describe path-dependent volatility models and the associated calibration problems ([Bergomi, 2009](#); [Guyon, 2014](#)). The latter models are characterized solely by a diffusion coefficient (i.e., zero drift) and, compared to standard local (stochastic) volatility models, can capture prominent historical patterns of volatility.

In addition, several mean-field control problems described by stochastic delayed differential equations of McKean–Vlasov type can be found in the literature; see, e.g., [Fouque & Zhang \(2019\)](#) and references cited therein. Further background on the theory and applications of delay equations is found, e.g., in [Kolmanovskii & Myshkis \(1992\)](#); [Hale & Lunel \(1993\)](#). For examples of McKean–Vlasov SDEs with nonglobally Lipschitz drift, without delay, we refer the reader to [Bao et al. \(2021\)](#) and the references cited therein.

The simulation of McKean–Vlasov SDEs typically involves two steps: first, at each time t , the true measure $\mathcal{L}_{X(t)}$ is approximated by the empirical measure

$$\mu_t^{X,N}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{X^{j,N}(t)}(dx),$$

where δ_x denotes the Dirac measure at point x and, for $\mathbb{S}_N := \{1, \dots, N\}$, $(X^{i,N})_{i \in \mathbb{S}_N}$ (a so-called *interacting particle system*) is the solution to the \mathbb{R}^{dN} -dimensional SDE

$$dX^{i,N}(t) = b(X^{i,N}(t), X^{i,N}(t - \tau), \mu_t^{X,N}, \mu_{t-\tau}^{X,N}) dt + \sigma(X^{i,N}(t), X^{i,N}(t - \tau), \mu_t^{X,N}, \mu_{t-\tau}^{X,N}) dW^i(t), \quad (1.2)$$

with $X_0^{i,N} = \xi^i$. Here, W^i and ξ^i , for $i \in \mathbb{S}_N$, are independent Brownian motions (also independent of W) and i.i.d. copies of ξ , respectively. We will be able to refer to [Neelima et al. \(2021\)](#) for an approximation result, referred to as *propagation of chaos*, in the delay setting. In a next step, one typically needs to introduce a reasonable time-stepping method to discretize the particle system $(X^{i,N})_{i \in \mathbb{S}_N}$ over some finite time horizon $[0, T]$. This second step is the focus of the present work.

The literature on higher order numerical schemes (i.e., beyond the basic Euler–Maruyama method of order 1/2) for classical delay SDEs is sparse and restricted to Lipschitz continuous coefficients. In [Hu *et al.* \(2004\)](#), the strong convergence of a Milstein scheme for point-delay SDEs was proven using an Itô formula for so-called tame functions and techniques from anticipative calculus. For the analysis of a Milstein scheme without relying on methods from anticipative calculus, we refer to [Kloeden & Shardlow \(2012\)](#). To the best of our knowledge, there is no published Milstein scheme for classical point-delay SDEs where the coefficients have super-linear growth.

It is also well-documented that standard schemes, e.g., of Euler-type with uniform step-size, are not suitable either for standard SDEs (see [Hutzenthaler *et al.*, 2011](#)) or (nondelay) McKean–Vlasov SDEs (see [Bao *et al.*, 2021](#); [Reis *et al.*, 2021](#)) with super-linear growth in the drift, i.e., the moments of the discretized process explode as the mesh-size tends to zero (even though a unique solution with bounded moments to the nondiscretized process exists). For remedies in these cases using adaptive, truncated, implicit or tamed schemes, see the references given in the above mentioned works.

For tamed Euler–Maruyama schemes applied to certain delay SDEs, we refer to [Dareiotis *et al.* \(2016\)](#) and the references cited therein. A Euler–Maruyama type scheme for the approximation of delay McKean–Vlasov SDEs was introduced in [Ren & Wu \(2018\)](#), but without an implementable (particle) approximation of the true measure. We are not aware of any published higher order time-stepping schemes in this context, not even for globally Lipschitz continuous coefficients. In this work, we provide a strong convergence analysis of first-order time-stepping schemes for delay McKean–Vlasov equations, possibly with super-linear drift.

A first contribution of the current article is the following:

- We derive a Milstein scheme for point delay McKean–Vlasov SDEs, i.e., path-dependent equations where the coefficients depend on the values of the process at a finite number of time points, and its law at these points (but not on the entire path).
- We prove first-order strong convergence for such problems, where the drift is allowed to grow super-linearly in the current state, but does not depend on delayed values (Theorem 2.4 for a ‘tamed’ scheme). We will also require Lipschitz dependence of all coefficients on the measure.¹

We hereby extend ideas from [Bao *et al.* \(2021\)](#); [Kumar & Neelima \(2021\)](#) on tamed Milstein schemes for nondelay equations. To derive our main result, we make use of the Lions derivative of functionals acting on $\mathcal{P}_2(\mathbb{R}^d)$. As we do not employ an Itô formula for the analysis of the scheme, we only need to require the coefficients to be once continuously differentiable in all variables (state and measure component).

The second, and main, novelty of the article concerns efficient estimators of expected functionals of the solution to McKean–Vlasov SDEs with delay, on the basis of the aforementioned Milstein schemes. The crucial computational difficulty for such equations is that already in the one-dimensional case the simulation of iterated stochastic integrals (see [Kloeden & Platen, 1992](#)) is required, as a double stochastic integral of an earlier segment of the Wiener path with respect to the ‘current’ Wiener path appears in the delay setting. Hence, for a direct application of the scheme, appropriate approximation techniques for such integrals (e.g., as proposed in [Wiktorsson, 2001](#)) need to be used and add significant computational complexity. Therefore, for the estimation of expected (path) functionals, we propose an antithetic approach that avoids the simulation of these iterated integrals, but still allows us to obtain higher order

¹ We refer the reader to Remark 5.6 for a discussion on technical difficulties which prevented us from considering delayed components in the drift term for the super-linear case.

convergence of the correction terms within the multilevel Monte Carlo estimator (MLMC; see, Giles, 2008; Giles & Szpruch, 2014). Consequently, we recover optimal complexity of the MLMC sampling scheme. Here, we adapt ideas from Giles (2015) for classical multidimensional SDEs without delay or mean-field interaction. Our results are new also for delay equations without mean-field interaction.

A natural MLMC estimator for $\mathbb{E}[P(X(T))]$, where P is a regular enough function (see Corollary 3.4) and X is the solution to the (delay) McKean–Vlasov SDE, can be constructed as follows (see Haji-Ali & Tempone, 2018):

First, for a given integer L and $l \in \{0, 1, \dots, L\}$, we consider numerical approximations $(Y^{j,N,l}(T))_{j \in \mathbb{S}_N}$ to the particle system at time T , obtained by a time-stepping scheme with mesh-size $\delta_l > 0$. Then, we introduce the quantity

$$\phi_N^l := \frac{1}{N} \sum_{j=1}^N P(Y^{j,N,l}(T)), \quad (1.3)$$

on the time discretization level l . In the sequel, we fix the number of particles N across all levels and construct independent realizations of ϕ_N^l .

The set of random variables needed to generate the k th realization of ϕ_N^l is denoted by $\omega_{1:N}^{(l,k)}$, $k \in \{1, \dots, K_l\}$ for some given integer K_l ; this is the set of N independent copies $(W^i, \xi^i)_{i \in \mathbb{S}_N}$. For $L \in \mathbb{N}$ levels, the MLMC estimator for $\mathbb{E}[P(X(T))]$ can be written as

$$\sum_{l=0}^L \frac{1}{K_l} \sum_{k=1}^{K_l} (\phi_N^l - \phi_N^{l-1})(\omega_{1:N}^{(l,k)}), \quad (1.4)$$

where ϕ_N^{l-1} has to be chosen appropriately. In particular, if $\varphi_N^{-1} = 0$ and $\mathbb{E}[\phi_N^{l-1}] = \mathbb{E}[\varphi_N^{l-1}]$, then a telescoping property applies, which ensures the bias of the MLMC estimator for given L is the same as that of the standard MC estimator ϕ_N^L with mesh-size δ_L . The key to reducing the computational complexity of the MLMC estimator is to introduce a strong correlation between φ and ϕ to reduce the variance. A simple choice to achieve this is $\varphi_N^{l-1}(\omega_{1:N}^{(l,k)}) = \phi_N^{l-1}(\omega_{1:N}^{(l,k)})$, i.e., the same Brownian motions and initial data are used for the two particle systems (with the same N) on the fine grid and on the coarse grid.

In the present paper,

- we show (in Theorem 3.2) that a certain antithetic Milstein scheme for ϕ_N^l gives strong order one in δ_l for the multilevel corrections $\phi_N^l - \phi_N^{l-1}$, without the need to compute Lévy areas,

and hence at the same cost per path as the lower-order Euler–Maruyama scheme. For simplicity, we restrict the analysis to the case where the drift coefficient depends only on the state and the measure, and the diffusion coefficient only on the state and a single delayed value, but the results extend straightforwardly to measure dependent volatility. In addition, it is assumed $d = m = 1$ for notational convenience.

Optimal complexity results for multilevel simulation of interacting particle systems are found in Haji-Ali & Tempone (2018), which are based on the assumption of a multiplicative bound for the variance of the correction terms,

$$\mathbb{V}[\phi_N^l - \phi_N^{l-1}] \lesssim \frac{\delta_l^s}{N}. \quad (1.5)$$

Here, $s/2$ corresponds to the strong order of the time-stepping scheme. Although the assumption (1.5) seems reasonable, it is an open problem to prove such multiplicative estimates (in N and δ_l^s). The analysis in Haji-Ali & Tempone (2018) subsequently reveals that higher order time-stepping schemes (e.g., $s = 2$) can reduce the overall computational complexity.

For the following discussion, we assume that (1.5) holds with $s = 2$, i.e., in addition to the first-order convergence in δ_l , which we prove, we assume the factor $1/N$. We also assume that the complexity of computing all interaction terms for N particles due to the presence of the empirical measure is $\mathcal{O}(N^{p+1})$ per time-step, where $p = 0, 1$. To be more explicit, for $p = 0$, the cost is of identical order to standard systems of SDEs, which is the case, for instance, if the coefficients involve the expectation of $X(t)$, which can be approximated by the empirical average over all particles in a single computation. The case $p = 1$ would be relevant, e.g., if different expectations needed to be computed for each particle.

Then, Haji-Ali & Tempone (2018, Theorem 3.1) allows us to conclude that the work complexity required to compute $\mathbb{E}[P(X(T))]$ with RMSE of order $\varepsilon > 0$ is $\mathcal{O}(\varepsilon^{-2-p})$ when our antithetic Milstein scheme is used as a time-stepping method. For a tamed Euler–Maruyama scheme, the multilevel scheme outlined above would give a work complexity of order $\mathcal{O}(\varepsilon^{-2-p} \log^2(\varepsilon))$. A standard MC estimator based on $K = \mathcal{O}(\varepsilon^{-1})$ samples of the entire particle system would have complexity $\mathcal{O}(\varepsilon^{-3-p})$; see Haji-Ali & Tempone (2018); Szpruch & Tse (2021).

As noted in Haji-Ali & Tempone (2018) and further analysed in Szpruch & Tse (2021), one can additionally vary the number of particles with the levels (see also earlier Bujok *et al.*, 2015, for the case with interaction through common noise). If we want to work with two particle systems of size N_l and N_{l-1} on level l , where $N_l = \beta N_{l-1}$ for some integer β (e.g., $\beta = 2$), we split the set of N_l underlying Brownian motions into β sets of Brownian motions and simulate independently β particle systems each of size N_{l-1} . Hence, we can define

$$\phi_{N_{l-1}}^{l-1}(\omega_{1:N_l}^{(l,k)}) := \frac{1}{\beta} \sum_{i=1}^{\beta} \phi_{N_{l-1}}^{l-1}(\omega_{(i-1)N_{l-1}+1:iN_{l-1}}^{(l,k)}).$$

Using this definition in conjunction with (1.4) gives then the MLMC estimator for $\mathbb{E}[P(X(T))]$ where both the number of particles and number of time-steps vary with the level. In Szpruch & Tse (2021, Theorem 6.2), for the setting without delay and with a constant diffusion coefficient (such that a standard Euler–Maruyama scheme already yields strong convergence of order one; see Szpruch & Tse, 2021, Lemma 6.1), an additive error bound for the variance of the correction terms, of order one with respect to both M (number of time-steps) and N , was proven.

Our main result on the strong convergence of the multilevel correction terms using the antithetic Milstein scheme for the delay case complements (Szpruch & Tse, 2021, Theorem 6.2) in the sense that we prove order two convergence of the variance of the correction terms in the setting with nonconstant diffusion term without increasing the computational complexity compared to a standard (tamed) Euler–Maruyama scheme. We achieve this through an antithetic approach by avoiding the simulation of iterated stochastic integrals required for the direct Milstein scheme even for $d = m = 1$ in the delay case.

The analysis in (Szpruch & Tse, 2021, Theorem 6.3) then reveals that the complexity of such a multilevel scheme is at most $\mathcal{O}(\varepsilon^{-2-p})$ with RMSE of order ε . This coincides with the result of the earlier discussion, where the number of particles was fixed across all time discretization levels. However, this holds now also without the assumption of a multiplicative error bound as in (1.5). This means that if both, the number of time-steps and number of particles, are varied with the level, then an additive bound for the variance decay yields the same computational complexity as the MLMC scheme with a multiplicative

rate (1.5), but where only the number of time-steps are varied across the levels. Also, according to Table 1 in Haji-Ali & Tempone (2018), if $s = 2$ and the computation of the interaction terms is of order N^2 , then both MLMC estimators, i.e., where either only the number of time-steps or where both the number of particles and the number of time-steps vary with the levels, yield the same computational complexity. However, if the complexity in terms of N is only of order one and $s = 2$, then Table 1 in Haji-Ali & Tempone (2018) reveals that it is more efficient to keep the number of particles constant across all levels; see Section 4.2 for details.

The remainder of this article is organized as follows: in Section 2, we introduce the precise class of delay equations considered here and state the first-order convergence result for the tamed Milstein schemes. Section 3 is concerned with the analysis of the antithetic approach. This allows us to construct a numerical scheme without simulating the iterated stochastic integrals, which still achieves convergence order one in the MLMC framework. Section 4 illustrates the numerical performance of the proposed time-stepping schemes. The proofs of the main convergence results are deferred to Sections 5 and 6.

Preliminaries: We end this section by introducing some notations and concepts that will be needed throughout this article. Let $T > 0$ be a given terminal time and $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ will denote a complete filtered probability space satisfying the usual assumptions, where $(\Omega, \mathcal{F}, \mathbb{P})$ is assumed to be atomless.

Here, $(\mathbb{R}^d, \langle \cdot, \cdot \rangle, |\cdot|)$ will represent the d -dimensional Euclidean space and $\mathbb{R}^d \otimes \mathbb{R}^m$ be the collection of all $d \times m$ -matrices.

In addition, we use $\mathcal{P}(\mathbb{R}^d)$ to denote the family of all probability measures on \mathbb{R}^d and define the subset of probability measures with finite second-order moment by

$$\mathcal{P}_2(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \mu(|\cdot|^2) = \int_{\mathbb{R}^d} |x|^2 \mu(dx) < \infty \right\}.$$

For matrices, we will use the standard Hilbert–Schmidt norm denoted by $\|\cdot\|$.

As metric on the space $\mathcal{P}_2(\mathbb{R}^d)$, we use the Wasserstein distance. For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the Wasserstein distance between μ and ν is defined as

$$\mathcal{W}_2(\mu, \nu) := \inf_{\pi \in \mathcal{C}(\mu, \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(dx, dy) \right)^{1/2},$$

where $\mathcal{C}(\mu, \nu)$ is the set of all couplings of μ and ν , i.e., $\pi \in \mathcal{C}(\mu, \nu)$ if and only if $\pi(\cdot, \mathbb{R}^d) = \mu(\cdot)$ and $\pi(\mathbb{R}^d, \cdot) = \nu(\cdot)$.

We briefly introduce the L -derivative of a functional $f : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, as it will appear in the proofs presented in the main section. For further information on this concept, we refer to (Cardaliaguet, 2013) or Buckdahn *et al.* (2017); Hammersley *et al.* (2021). Here, we follow the exposition of Carmona & Delarue (2017). We will associate to the function f a lifted function \tilde{f} by $\tilde{f}(X) = f(\mathcal{L}_X)$, where \mathcal{L}_X is the law of X , for $X \in L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$.

A function f on $\mathcal{P}_2(\mathbb{R}^d)$ is said to be L -differentiable at $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists a random variable X_0 with law μ_0 , such that the lifted function \tilde{f} is Fréchet differentiable at X_0 .

Now, the Riesz representation theorem implies that there is a (\mathbb{P} -a.s.) unique $\Phi \in L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ with

$$\tilde{f}(X) = \tilde{f}(X_0) + \langle \Phi, X - X_0 \rangle_{L_2} + o(\|X - X_0\|_{L_2}), \text{ as } \|X - X_0\|_{L_2} \rightarrow 0,$$

with the standard inner product and norm on $L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$. If f is L -differentiable for all $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, then we say that f is L -differentiable.

It is known (see, e.g., Carmona & Delarue, 2017, Proposition 5.25) that there exists a Borel measurable function $\kappa : \mathbb{R}^d \rightarrow \mathbb{R}^d$, such that $\Phi = \kappa(X_0)$ almost surely, and hence

$$f(\mathcal{L}_X) = f(\mathcal{L}_{X_0}) + \mathbb{E} \langle \kappa(X_0), X - X_0 \rangle + o(\|X - X_0\|_{L_2}).$$

Note that Φ can be expressed as a function (κ) of any random variable X_0 with distribution μ_0 , giving a meaning to the L -derivative independently of the lifting chosen to construct it, see Carmona & Delarue (2017, Remark 5.26). We define $\partial_\mu f(\mathcal{L}_{X_0})(y) := \kappa(y)$, $y \in \mathbb{R}^d$, as the L -derivative of f at μ_0 . For a vector-valued (or matrix-valued) function f , these definitions have to be understood componentwise.

For some real $\tau > 0$, let $\mathcal{C} = C([-\tau, 0]; \mathbb{R}^d)$ be the set of all continuous functions $f : [-\tau, 0] \rightarrow \mathbb{R}^d$. Further, for $\eta \in \mathcal{C}$ and $\theta \in [-\tau, 0]$, let $\Pi_\theta : \mathcal{C} \rightarrow \mathbb{R}^d$ be the projection operator, i.e., $\Pi_\theta(\eta) = \eta(\theta)$. For $f \in C([-\tau, T]; \mathbb{R}^d)$ and for $t \in [0, T]$, let $f_t \in \mathcal{C}$ be defined by $f_t(\theta) = f(t + \theta)$, $\theta \in [-\tau, 0]$. In the literature (see e.g., Mao, 1997), $(f_t)_{t \geq 0}$ is called the segment process associated with $(f(t))_{t \geq -\tau}$. We will also need the norms $\|f\|_\infty := \sup_{v \in [-\tau, 0]} |f(v)|$ for $f \in \mathcal{C}$ and $\|X\|_{\infty, t} := \sup_{s \in [0, t]} |X(s)|$, for $t \geq 0$ (and zero otherwise) and a process X precisely defined below.

2. Delay McKean–Vlasov SDEs and Milstein schemes

2.1 Problem formulation

In this section, we consider the following SDE for $t \in [0, T]$

$$dX(t) = b(\Pi(X_t), \mathcal{L}_{\Pi(X_t)}) dt + \sigma(\Pi(X_t), \mathcal{L}_{\Pi(X_t)}) dW(t), \quad X_0 = \xi \in L_p^0(\mathcal{C}), \quad (2.1)$$

where $L_p^0(\mathcal{C})$ is the space \mathcal{F}_0 -measurable \mathcal{C} -valued random variables with $\mathbb{E}[\|\xi\|_\infty^p] < \infty$, for a given $p \geq 2$. Further, for $-\tau \leq s_k < \dots < s_2 < s_1 \leq 0$, with $\tau < T$ and a \mathcal{C} -valued random variable χ

$$\Pi(\chi) := (\Pi_{s_1}(\chi), \dots, \Pi_{s_k}(\chi)), \quad \mathcal{L}_{\Pi(\chi)} := (\mathcal{L}_{\Pi_{s_1}(\chi)}, \dots, \mathcal{L}_{\Pi_{s_k}(\chi)}),$$

$b : \mathbb{R}^{dk} \times \mathcal{P}_2(\mathbb{R}^d)^k \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^{dk} \times \mathcal{P}_2(\mathbb{R}^d)^k \rightarrow \mathbb{R}^d \otimes \mathbb{R}^m$ are given measurable functions, with $\mathcal{P}_2(\mathbb{R}^d)^k := \mathcal{P}_2(\mathbb{R}^d) \times \dots \times \mathcal{P}_2(\mathbb{R}^d)$ (k -times), and $W = (W(t))_{t \in [0, T]}$ is an m -dimensional Brownian motion on some complete filtered atomless probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$, where $(\mathcal{F}_t)_{t \in [0, T]}$ is the natural filtration of W augmented with an independent σ -algebra \mathcal{F}_0 . In what follows, we will set $s_1 = 0$ and $s_k = -\tau$. In the sequel, we refer to these equations as point-delay McKean–Vlasov SDEs. In Hu *et al.* (2004), the same type of delays was considered for classical SDEs.

In the following, we impose several conditions on the drift and diffusion coefficient which guarantee well-posedness of the considered point-delay McKean–Vlasov SDE.

To do so, we introduce the notation $\mathbf{y}_k := (y_1, \dots, y_k)$, $\tilde{\mathbf{y}}_k := (\tilde{y}_1, \dots, \tilde{y}_k)$, $\bar{\mathbf{y}}_k := (y'_1, y_2, \dots, y_k)$, for $y'_1, y_i, \tilde{y}_i \in \mathbb{R}^d$, and $\boldsymbol{\mu}_k := (\mu_1, \dots, \mu_k)$, $\tilde{\boldsymbol{\mu}}_k := (\tilde{\mu}_1, \dots, \tilde{\mu}_k)$, for $\tilde{\mu}_i, \mu_i \in \mathcal{P}_2(\mathbb{R}^d)$. Further, we set $\mathbf{0}_k := (\mathbf{0}, \dots, \mathbf{0})$, (k components), where $\mathbf{0} \in \mathbb{R}^d$.

For the drift term b , we assume that:

(\mathbf{AD}_b^1) There exist constants $L_b^1, q_1 \geq 0$ such that

$$\begin{aligned} \langle y_1 - y'_1, b(\mathbf{y}_k, \boldsymbol{\mu}_k) - b(\tilde{\mathbf{y}}_k, \boldsymbol{\mu}_k) \rangle &\leq L_b^1 |y_1 - y'_1|^2, \\ |b(\mathbf{y}_k, \boldsymbol{\mu}_k) - b(\tilde{\mathbf{y}}_k, \boldsymbol{\mu}_k)| &\leq L_b^1 \sum_{i=1}^k |y_i - \tilde{y}_i| (1 + |y_i|^{q_1} + |\tilde{y}_i|^{q_1}), \\ |b(\mathbf{0}_k, \boldsymbol{\mu}_k) - b(\mathbf{0}_k, \tilde{\boldsymbol{\mu}}_k)| &\leq L_b^1 \sum_{i=1}^k \mathcal{W}_2(\mu_i, \tilde{\mu}_i), \end{aligned}$$

for all $\tilde{\mathbf{y}}_k, \mathbf{y}_k, \tilde{\mathbf{y}}_k \in \mathbb{R}^{dk}$ and $\boldsymbol{\mu}_k, \tilde{\boldsymbol{\mu}}_k \in \mathcal{P}_2(\mathbb{R}^d)^k$.

Concerning the diffusion coefficient σ , we assume:

(\mathbf{AD}_σ^1) There exist constants $L_\sigma^1, q_2 \geq 0$ such that

$$\begin{aligned} \|\sigma(\mathbf{y}_k, \boldsymbol{\mu}_k) - \sigma(\tilde{\mathbf{y}}_k, \boldsymbol{\mu}_k)\| &\leq L_\sigma^1 \left\{ |y_1 - \tilde{y}_1| + \sum_{i=2}^k |y_i - \tilde{y}_i| (1 + |y_i|^{q_2} + |\tilde{y}_i|^{q_2}) \right\}, \\ \|\sigma(\mathbf{0}_k, \boldsymbol{\mu}_k) - \sigma(\mathbf{0}_k, \tilde{\boldsymbol{\mu}}_k)\| &\leq L_\sigma^1 \sum_{i=1}^k \mathcal{W}_2(\mu_i, \tilde{\mu}_i), \end{aligned}$$

for all $\mathbf{y}_k, \tilde{\mathbf{y}}_k \in \mathbb{R}^{dk}$ and $\boldsymbol{\mu}_k, \tilde{\boldsymbol{\mu}}_k \in \mathcal{P}_2(\mathbb{R}^d)^k$.

We note that under these conditions, the SDE (2.1) has a unique strong solution. To present an inductive argument for this assertion, we assume, for ease of notation that there is only one delay, i.e., $k = 2$ and $s_1 = 0, s_2 = -\tau$. On the interval $[0, \tau]$, SDE (2.1) can be written as a nondelay SDE with random coefficients

$$dX(t) = b(X(t), \xi(t - \tau), \mathcal{L}_{X(t)}, \mathcal{L}_{\xi(t - \tau)}) dt + \sigma(X(t), \xi(t - \tau), \mathcal{L}_{X(t)}, \mathcal{L}_{\xi(t - \tau)}) dW(t),$$

with initial value $X(0) = \xi(0)$. This SDE has a unique strong solution under (\mathbf{AD}_b^1) and (\mathbf{AD}_σ^1), see, e.g., Reis *et al.* (2019). Requiring the initial data to be in $L_{p(1+q)}^0(\mathcal{C})$, where $q = q_1 \vee q_2$, the solution also has finite moments up to order p . A similar argument can be employed on the interval $[\tau, 2\tau]$, and so forth up to the final time $T > 0$.

PROPOSITION 2.1 Let $p \geq 2$ be given. Let Assumptions (\mathbf{AD}_b^1) and (\mathbf{AD}_σ^1) hold and suppose $X_0 = \xi \in L_{k_1}^0(\mathcal{C})$, where $k_1 \geq p(1+q)$, with $q = q_1 \vee q_2$, will be defined in the proof. Then, there exists a constant $C > 0$ such that

$$\mathbb{E} \left[\sup_{-\tau \leq t \leq T} |X_t|^p \right] \leq C(1 + \mathbb{E}[\|\xi\|_\infty^{k_1}]).$$

Proof. We prove the result for $k = 2$ and $s_1 = 0, s_2 = -\tau$ only. Itô's formula implies, for $t \geq 0$ and $p \geq 2$,

$$\begin{aligned} |X(t)|^p &\leq |\xi(0)|^p + p \int_0^t |X(s)|^{p-2} \langle X(s), b(X(s), X(s-\tau), \mathcal{L}_{X(s)}, \mathcal{L}_{X(s-\tau)}) \rangle ds \\ &\quad + p \int_0^t |X(s)|^{p-2} \langle X(s), \sigma(X(s), X(s-\tau), \mathcal{L}_{X(s)}, \mathcal{L}_{X(s-\tau)}) dW(s) \rangle \\ &\quad + \frac{p(p-1)}{2} \int_0^t |X(s)|^{p-2} \|\sigma(X(s), X(s-\tau), \mathcal{L}_{X(s)}, \mathcal{L}_{X(s-\tau)})\|^2 ds. \end{aligned}$$

Using assumptions (\mathbf{AD}_b^1) , (\mathbf{AD}_σ^1) and the moment boundedness of the initial process, it is standard to show that there exists a constant $C > 0$, such that

$$\mathbb{E}[\|X\|_{\infty,t}^p] \leq C + C \int_0^t \mathbb{E}[\|X\|_{\infty,s}^p] ds + C \int_0^{0 \vee (t-\tau)} \mathbb{E}[\|X\|_{\infty,s}^p (1 + \|X\|_{\infty,s}^{pq})] ds.$$

We define, for the given $p \geq 2$,

$$k_i := p([T/\tau] + 2 - i)(1 + q)^{[T/\tau] + 1 - i}, \quad \text{for } i \in \{1, \dots, [T/\tau] + 1\}.$$

Thus,

$$(1 + q)k_{i+1} < k_i, \quad k_{[T/\tau] + 1} = p, \quad i \in \{1, \dots, [T/\tau]\}.$$

Hence, we have on $[0, \tau]$

$$\mathbb{E}[\|X\|_{\infty,\tau}^{k_1}] < \infty.$$

Consequently, by Hölder's inequality,

$$\begin{aligned} \mathbb{E}[\|X\|_{\infty,2\tau}^{k_2}] &\leq C + C \int_0^{2\tau} \mathbb{E}[\|X\|_{\infty,s}^{k_2}] ds + C \int_0^\tau \mathbb{E}[\|X\|_{\infty,s}^{(1+q)k_2}] ds \\ &\leq C + C \int_0^{2\tau} \mathbb{E}[\|X\|_{\infty,s}^{k_2}] ds + C \int_0^\tau \left(\mathbb{E}[\|X\|_{\infty,s}^{k_1}] \right)^{\frac{(1+q)k_2}{k_1}} ds < \infty. \end{aligned}$$

This procedure can be repeated up to time $T > 0$ and the claim follows. \square

We now introduce a particle system associated with the point-delay McKean–Vlasov SDE (2.1)

$$dX^{i,N}(t) = b(\Pi(X_t^{i,N}), \Pi(\mu_t^{X,N})) dt + \sigma(\Pi(X_t^i), \Pi(\mu_t^{X,N})) dW^i(t), \quad (2.2)$$

where

$$\Pi(\mu_t^{X,N}) := (\mu_{t+s_1}^{X,N}, \dots, \mu_{t+s_k}^{X,N}), \quad \mu_{t+s_l}^{X,N}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{X^{j,N}(t+s_l)}(dx), \text{ for } l \in \{1, \dots, k\},$$

and $(W^i, X_0^{i,N})_{i \in \mathbb{S}_N}$ are independent copies of (W, ξ) . The following proposition provides a strong propagation of chaos result, which is proven in Neelima *et al.* (2021, Proposition 5.2). Note that Assumptions \bar{A}_1 – \bar{A}_5 in Neelima *et al.* (2021) are satisfied if Assumptions (\mathbf{AD}_b^1) and (\mathbf{AD}_σ^1) hold (taking without loss of generality $q_1 = q_2$).

PROPOSITION 2.2 (Proposition 5.2 in Neelima *et al.*, 2021). Let Assumptions (\mathbf{AD}_b^1) and (\mathbf{AD}_σ^1) hold and suppose $X_0 = \xi \in L_{k_1}^0(\mathcal{C})$, with k_1 as in Proposition 2.1 such that $k_1 > 4$. Then, the interacting particle system (2.2) is strongly well-posed and its solution $X^{i,N}$ converges to the solution X^i of (2.1) with W replaced by W^i (a ‘noninteracting particle system’),

$$\max_{i \in \mathbb{S}_N} \sup_{t \in [0, T]} \mathbb{E}[|X_t^i - X_t^{i,N}|^2] \leq C \begin{cases} N^{-1/2}, & \text{if } d < 4, \\ N^{-1/2} \ln(N), & \text{if } d = 4, \\ N^{-2/d}, & \text{if } d > 4 \end{cases}$$

for any $N \in \mathbb{N}$, where the constant $C > 0$ does not depend on N .

REMARK 2.3 It is possible to rewrite the particle system as a sequence of $(N \times d \times k)$ -dimensional SDEs without explicit delay dependence, see K  chler & Platen (2000). Due to the dimensionality and the structure of the coefficients of the enlarged system, classical results are not directly applicable, as it is crucial in our analysis that implied constants are independent of the number of particles N and the extra difficulties arising from the delayed components, which we address in the following, cannot be avoided.

2.2 Milstein schemes for delay McKean–Vlasov equations

We propose the following tamed Milstein schemes for the numerical approximation of the particle system (2.2), which are motivated by the scheme presented in Hu *et al.* (2004). Considering a uniform time mesh on $[-\tau, T]$ with mesh-size $\delta = T/M = \tau/M'$, for integers $M, M' > 1$, we define the following approximation of (2.2) (note that we assume the delayed time-points $t_n + s_l$ to be contained on the grid):

$$\begin{aligned} Y^{i,N}(t_{n+1}) &= Y^{i,N}(t_n) + b_\delta(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N}))\delta + \sigma(\Pi(Y_{t_n}^i), \Pi(\mu_{t_n}^{Y,N}))\Delta W_n^i \\ &+ \sum_{l=1}^k \partial_{x_l} \sigma(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N})) \tilde{\sigma}(t_n + s_l, \Pi(Y_{t_n+s_l}^{i,N}), \Pi(\mu_{t_n+s_l}^{Y,N})) \int_{t_n}^{t_{n+1}} \int_{t_n+s_l}^{s+s_l} dW^i(u) dW^i(s) \\ &+ \sum_{l=1}^k \frac{1}{N} \sum_{j=1}^N \partial_{\mu_l} \sigma(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N}))(Y^{j,N}(t_n + s_l)) \\ &\quad \times \tilde{\sigma}(t_n + s_l, \Pi(Y_{t_n+s_l}^{j,N}), \Pi(\mu_{t_n+s_l}^{Y,N})) \int_{t_n}^{t_{n+1}} \int_{t_n+s_l}^{s+s_l} dW^j(u) dW^j(s), \end{aligned} \quad (2.3)$$

for $t_n := n\delta \geq 0$, $n \in \{0, \dots, M-1\}$. Furthermore,

$$\tilde{\sigma}(t, \Pi(Y_t^{i,N}), \Pi(\mu_t^{Y,N})) := \begin{cases} \sigma(\Pi(Y_t^{i,N}), \Pi(\mu_t^{Y,N})), & t \geq 0, \\ 0, & -\tau \leq t < 0, \end{cases}$$

and we used the notation

$$\Pi(\mu_{t_n}^{Y,N}) := (\mu_{t_n+s_1}^{Y,N}, \dots, \mu_{t_n+s_k}^{Y,N}), \quad \mu_{t_n+s_l}^{Y,N}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^N \delta_{Y^{j,N}(t_n+s_l)}(\mathrm{d}x), \text{ for } l \in \{1, \dots, k\}.$$

We denote by ∂_{x_l} the gradient with respect to the l th state component and by ∂_{μ_l} the L -derivative with respect to l th measure component. Observe that ∂_{x_l} and ∂_{μ_l} when applied to σ have to be understood in an operator sense. Note that for $t \leq 0$, we have $Y^{i,N}(t) = \xi^i(t)$; an independent copy of ξ . As in the nondelay case (Bao *et al.*, 2021), we define b_δ in two different ways yielding schemes that will be denoted by Scheme 1 and Scheme 2, respectively: for Scheme 1, we use

$$b_\delta(x, \mu) := \frac{b(x, \mu)}{1 + \delta|b(x, \mu)|}, \quad x \in \mathbb{R}^{dk}, \quad \mu \in \mathcal{P}_2(\mathbb{R}^d)^k,$$

and for Scheme 2, we define

$$b_\delta(x, \mu) := \frac{b(x, \mu)}{1 + \delta|b(x, \mu)|\bar{q}}, \quad x \in \mathbb{R}^{dk}, \quad \mu \in \mathcal{P}_2(\mathbb{R}^d)^k,$$

where $\bar{q} = \frac{2q_1}{q_1+1}$. For any $t \in [0, T]$, let $t_\delta := \lfloor t/\delta \rfloor \delta$. The continuous-time version of (2.3) is

$$\begin{aligned} \mathrm{d}Y^{i,N}(t) &= b_\delta(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \mathrm{d}t + \left(\sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \right. \\ &\quad + \sum_{l=1}^k \partial_{x_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta+s_l}^{i,N}), \Pi(\mu_{t_\delta+s_l}^{Y,N})) \int_{t_\delta+s_l}^{t+s_l} \mathrm{d}W^i(u) \\ &\quad + \sum_{l=1}^k \frac{1}{N} \sum_{j=1}^N \partial_{\mu_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) (Y^{j,N}(t_\delta + s_l)) \\ &\quad \left. \times \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta+s_l}^{i,N}), \Pi(\mu_{t_\delta+s_l}^{Y,N})) \int_{t_\delta+s_l}^{t+s_l} \mathrm{d}W^j(u) \right) \mathrm{d}W^i(t). \end{aligned}$$

A Milstein scheme for globally Lipschitz continuous drift and diffusion coefficients can be readily formulated by replacing b_δ with b in (2.3).

In the sequel, we restrict the discussion to $d = m = 1$, and introduce

$$I^i(t_n + s_l, t_{n+1} + s_l; s_l) := \int_{t_n}^{t_{n+1}} \int_{t_n+s_l}^{t_{n+1}+s_l} \mathrm{d}W^i(u) \mathrm{d}W^i(s).$$

For $s_l = 0$, this double Itô integral simplifies to

$$I^i(t_n, t_{n+1}; 0) = \frac{1}{2}((\Delta W_n^i)^2 - \delta).$$

To simulate all the double stochastic integrals for $s_l \neq 0$, which appear for delay equations already for $m = d = 1$, one can employ an approximation of $I^i(t_n + s_l, t_{n+1} + s_l; s_l)$; see Kloeden & Platen (1992) and Gaines & Lyons (1994); Wiktorsson (2001); Malham & Wiese (2014) for further developments on this topic. Based on the definitions $\bar{W}^i(s) := W^i(s + t_n + s_l) - W^i(t_n + s_l)$ (delayed Brownian increment), and $\tilde{B}^i(s) := \bar{W}^i(s - s_l) - \bar{W}^i(-s_l)$ (present increment), for $s \geq 0$, Hu *et al.* (2004) introduces an approximation for $I^i(t_n + s_l, t_{n+1} + s_l; s_l)$ based on an infinite series representation for Brownian bridges (truncated at level $r \in \mathbb{N}$) as in Kloeden & Platen (1992). In our tests, we add a term for approximating the tail of the Lévy area as in Wiktorsson (2001), to improve the mean-square error of the approximation to the order $\mathcal{O}(\delta^2/r^2)$. This justifies the choice $r = \mathcal{O}(\delta^{-1/2})$, compared to $1/r \leq \delta \wedge \min\{|s_l| : l \in \{1, \dots, k\}\}$ given in Hu *et al.* (2004, Lemma 4.2), and hence reduces the complexity. The tamed Milstein scheme combined with this approximation technique of the double stochastic integrals has computational complexity of order $\varepsilon^{-3/2}$ per sample for a root-mean-square error of order $\varepsilon > 0$, compared to the standard Euler–Maruyama scheme, which has complexity of order ε^{-2} . The following additional assumptions will be needed for the subsequent presentation and are motivated by the assumptions formulated in Bao *et al.* (2021). We impose the following to hold uniformly in $x_1, x'_1, z_1, z'_1 \in \mathbb{R}^d$, $\mu_1, \tilde{\mu}_1 \in \mathcal{P}_2(\mathbb{R}^d)$, $y_k, \tilde{y}_k, \hat{y}_k \in \mathbb{R}^{dk}$ and $v_k, \tilde{v}_k, \hat{v}_k \in \mathcal{P}_2(\mathbb{R}^d)^k$:

- (AD_b²) The drift b does not depend on delay variables and $b \in C^{1,1}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$, i.e., b is continuously differentiable in both components.
- (AD_b³) There exist constants $L_b^3, q_3 \geq 0$ such that

$$\begin{aligned} \|\partial_x b(x_1, \mu_1) - \partial_x b(x'_1, \tilde{\mu}_1)\| &\leq L_b^3 \{ |x_1 - x'_1| (1 + |x_1|^{q_3} + |x'_1|^{q_3}) + \mathcal{W}_2(\mu_1, \tilde{\mu}_1) \}, \\ \|\partial_{\mu} b(x_1, \mu_1)(z_1) - \partial_{\mu} b(x'_1, \tilde{\mu}_1)(z'_1)\| &\leq L_b^3 \{ |x_1 - x'_1| + |z_1 - z'_1| + \mathcal{W}_2(\mu_1, \tilde{\mu}_1) \}. \end{aligned}$$

The reason why b cannot depend on delay variables will be made clearer in the proof of Theorem 2.4. Concerning the diffusion term σ with columns denoted by σ_u for $u \in \{1, \dots, m\}$, we further require:

- (AD_σ²) For all $u \in \{1, \dots, m\}$, we have $\sigma_u \in C^{1,1}(\mathbb{R}^{dk} \times \mathcal{P}_2(\mathbb{R}^d)^k)$ and, in particular, there exist constants $L_\sigma^2, q_2 \geq 0$ such that for all u

$$\begin{aligned} \|\partial_{\mu_1} \sigma_u(y_k, v_k)(z_1) \sigma(y_k, v_k)\| \vee \|\partial_{x_1} \sigma_u(y_k, v_k) \sigma(y_k, v_k)\| \\ \leq L_\sigma^2 \left\{ 1 + |z_1| + |y_1| + v_1(|\cdot|^2)^{\frac{1}{2}} + \sum_{i=2}^k (|y_i|^{q_2+1} + v_i(|\cdot|^2)^{\frac{1}{2}}) \right\}. \end{aligned}$$

For the derivatives with respect to the delay variables, we require that one of the following holds (the reason and structure for these assumptions will become clearer in Section 5):

- a. The derivatives (in the state and measure component) with respect to the delay variables are uniformly bounded. In particular, we impose $q_2 = 0$ in (AD_σ¹).

- b. For all $u \in \{1, \dots, m\}$ and for any $l > 1$, and $z_l \in \mathbb{R}^d$, the derivatives $\partial_{\mu_l} \sigma_u(\mathbf{y}_k, \mathbf{v}_k)(z_l)$ and $\partial_{x_l} \sigma_u(\mathbf{y}_k, \mathbf{v}_k)$ do not depend on y_1, \dots, y_{l-1} and v_1, \dots, v_{l-1} , and for a constant $q_3 \geq 0$

$$\begin{aligned} & \|\partial_{\mu_l} \sigma_u(\mathbf{y}_k, \mathbf{v}_k)(z_l) \sigma(\hat{\mathbf{y}}_k, \hat{\mathbf{v}}_k)\| \\ & \leq L_\sigma^2 \left\{ 1 + |z_l| + \sum_{i=l}^k \left(|y_i|^{q_3+1} + v_i(|\cdot|^2)^{\frac{1}{2}} \right) + \sum_{i=1}^k \left(|\hat{y}_i|^{q_3+1} + \hat{v}_i(|\cdot|^2)^{\frac{1}{2}} \right) \right\}, \\ & \|\partial_{x_l} \sigma_u(\mathbf{y}_k, \mathbf{v}_k) \sigma(\hat{\mathbf{y}}_k, \hat{\mathbf{v}}_k)\| \\ & \leq L_\sigma^2 \left\{ 1 + \sum_{i=l}^k \left(|y_i|^{q_3+1} + v_i(|\cdot|^2)^{\frac{1}{2}} \right) + \sum_{i=1}^k \left(|\hat{y}_i|^{q_3+1} + \hat{v}_i(|\cdot|^2)^{\frac{1}{2}} \right) \right\}. \end{aligned}$$

(**AD** _{σ} ³) There exists a constant $L_\sigma^3 \geq 0$ such that for all $u \in \{1, \dots, m\}$

$$\begin{aligned} \|\partial_{x_1} \sigma_u(\mathbf{y}_k, \mathbf{v}_k) - \partial_{x_1} \sigma_u(\tilde{\mathbf{y}}_k, \tilde{\mathbf{v}}_k)\| & \leq L_\sigma^3 \left\{ \sum_{i=1}^k (|y_i - \tilde{y}_i| + \mathcal{W}_2(v_i, \tilde{v}_i)) \right\}, \\ \|\partial_{\mu_1} \sigma_u(\mathbf{y}_k, \mathbf{v}_k)(z_1) - \partial_{\mu_1} \sigma_u(\tilde{\mathbf{y}}_k, \tilde{\mathbf{v}}_k)(z'_1)\| & \leq L_\sigma^3 \left\{ |z_1 - z'_1| + \sum_{i=1}^k (|y_i - \tilde{y}_i| + \mathcal{W}_2(v_i, \tilde{v}_i)) \right\}, \end{aligned}$$

and analogously for the derivatives in the delay components.

In addition to above assumptions, we demand:

- (**H**₁) Regularity on the initial data ξ : ξ is a deterministic function and there exists a constant $C > 0$, such that for any $s, t \in [-\tau, 0]$

$$|\xi(t) - \xi(s)| \leq C|t - s|. \quad (2.4)$$

The assumption that ξ is deterministic is made to simplify the formulation of the results and could be relaxed. We give a stability and strong convergence analysis for Scheme 2 only, as less assumptions are required for this scheme (compared to Scheme 1); see [Bao et al. \(2021\)](#) for a discussion on this. We prove the following statement:

THEOREM 2.4 Let Assumptions (**AD** _{b} ¹)–(**AD** _{b} ³), (**AD** _{σ} ¹)–(**AD** _{σ} ³) and (**H**₁) hold. Then, for any $p \geq 2$, there exist constants $C_1, C_2 > 0$ (independent of M and N) such that

$$\max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M\}} \mathbb{E}[|Y^{i,N}(t_n)|^p] \leq C_1,$$

and

$$\max_{i \in \mathbb{S}_N} \mathbb{E}[\|X^{i,N} - Y^{i,N}\|_{\infty, T}^p] \leq C_2 \delta^p, \quad (2.5)$$

where $X^{i,N}$ is given by (1.2) and $Y^{i,N}$ is the continuous time version of (2.3) with $b_\delta(x, \mu) := \frac{b(x, \mu)}{1 + \delta|b(x, \mu)|^q}$.

Proof. The proof is deferred to Section 5. □

REMARK 2.5 Since the implied constant in Theorem 2.4 is independent of N , we can easily complete our statement with Proposition 2.2 to obtain an approximation of (2.1) in an L_p -sense. However, note that the rate provided in Proposition 2.2 is, in general, not optimal and does not yield the $1/N$ -rate given in (1.5). To achieve such a convergence order, further structural or regularity assumption on the coefficients of (2.1) are required, see e.g., Méléard (1996); Szpruch & Tse (2021).

3. Antithetic MLMC approach for delay equations

As explained in the introduction and detailed in the previous section, Milstein schemes for delay equations require computationally costly simulation of iterated stochastic integrals already for $d = m = 1$. This motivates the study of an antithetic approach for such delay McKean–Vlasov equations in order to improve the efficiency of the simulation.

3.1 Problem formulation

We consider the following one-dimensional, one-point delay McKean–Vlasov SDE

$$dX(t) = b(X(t), \mathcal{L}_{X(t)}) dt + \sigma(X(t), X(t - \tau)) dW(t), \quad X_0 = \xi \in L_p^0(\mathcal{C}), \quad (3.1)$$

for a given $p \geq 2$, with measurable functions $b : \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \rightarrow \mathbb{R}$, $\sigma : \mathbb{R}^2 \rightarrow \mathbb{R}$, and where $(W(t))_{t \in [0, T]}$ is a one-dimensional Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$. As earlier, we will restrict the discussion to deterministic initial data with time regularity specified in (2.4).

Above and in the analysis, we assume that the diffusion coefficient has no law dependence. We present the extension of the scheme to the law dependent case below, but for the main part avoid adding to the already lengthy notation. The scheme is directly applicable to equations with delay in the drift b , but the analysis of this case does not follow directly and is left for future work.

The associated particle system has the form

$$dX^{i,N}(t) = b(X^{i,N}(t), \mu_t^{X,N}) dt + \sigma(X^{i,N}(t), X^{i,N}(t - \tau)) dW^i(t), \quad X_0^{i,N} = \xi^i \in \mathcal{C}, \quad (3.2)$$

with (W^i, ξ^i) an independent copy of (W, ξ) for $i \in \mathbb{S}_N$, and

$$\mu_t^{X,N}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{X^{j,N}(t)}(dx), \quad t \in [0, T].$$

In the sequel, we will present model assumptions that guarantee well-posedness of the one point-delay McKean–Vlasov SDE (and the associated particle system) and will also be employed for the analysis of the subsequent numerical scheme. We impose the following to hold uniformly in $x, y, x', y' \in \mathbb{R}$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$.

(AAD_b¹) There exist constants $L_b^1, q_1 \geq 0$ such that

$$\begin{aligned} \langle x - y, b(x, \mu) - b(y, \mu) \rangle &\leq L_b^1 |x - y|^2, \\ |b(x, \mu) - b(y, \mu)| &\leq L_b^1 (1 + |x|^{q_1} + |y|^{q_1}) |x - y|, \\ |b(x, \mu) - b(x, \nu)| &\leq L_b^1 \mathcal{W}_2(\mu, \nu). \end{aligned}$$

(AAD_b²) The drift satisfies assumption (AD_b³).

(AAD_b³) The function $\mathbb{R} \ni x \mapsto b(x, \mu)$ belongs to $C^2(\mathbb{R}; \mathbb{R})$ and the second-order derivative is (possibly) of polynomial growth, i.e., there exist constants $L_b^3 \geq 0$ and $q_2 \geq 0$ such that

$$|\partial_x^2 b(x, \mu)| \leq L_b^3 (1 + |x|^{1+q_2}).$$

(AAD_σ¹) There exists an $L_\sigma^1 \geq 0$ such that

$$|\sigma(x, y) - \sigma(x', y')| \leq L_\sigma^1 (|x - x'| + |y - y'|).$$

(AAD_σ²) We have $\mathbb{R} \ni x \mapsto \sigma(x, y), \mathbb{R} \ni y \mapsto \sigma(x, y) \in C^2(\mathbb{R}; \mathbb{R})$ and their second-order derivatives are uniformly bounded.

3.2 Antithetic sampling scheme

Throughout the remaining sections, we denote by $\partial_{x_1} \sigma$ and $\partial_{x_2} \sigma$ the derivatives of σ with respect to the first and second component, respectively.

The tamed Milstein scheme for this system of particles has the form (for ease of notation, we refrain from using to notation $\tilde{\sigma}$ below)

$$\begin{aligned} Y^{i,N}(t_{n+1}) &= Y^{i,N}(t_n) + b_\delta(Y^{i,N}(t_n), \mu_{t_n}^{Y,N})\delta + \sigma(Y^{i,N}(t_n), Y^{i,N}(t_n - \tau))\Delta W_n^i \\ &\quad + \sigma(Y^{i,N}(t_n), Y^{i,N}(t_n - \tau))\partial_{x_1} \sigma(Y^{i,N}(t_n), Y^{i,N}(t_n - \tau)) \frac{(\Delta W_n^i)^2 - \delta}{2} \\ &\quad + \sigma(Y^{i,N}(t_n - \tau), Y^{i,N}(t_n - 2\tau))\partial_{x_2} \sigma(Y^{i,N}(t_n), Y^{i,N}(t_n - \tau)) \int_{t_n}^{t_{n+1}} \int_{t_n}^s dW_{u-\tau}^i dW_s^i \end{aligned}$$

for $t_m := m\delta, m \in \{0, \dots, M\}$, with $\delta = T/M = \tau/M'$ for $M, M' > 1$, and where

$$Y^{i,N}(\theta) = \xi^i(\theta), \theta \in [-\tau, 0], \quad \mu_{t_n}^{Y,N}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{Y^{j,N}(t_n)}(dx).$$

For simplicity, we assume that b_δ is defined by Scheme 2 introduced in Section 2.2, as less assumptions are needed to guarantee moment boundedness in that case.

Now, we propose a modified scheme that does not require the simulation of iterated stochastic integrals, but still achieves a higher order convergence rate for the variance of multilevel correction

terms: we define for $n \in \{0, \dots, M-1\}$ what will be the coarse grid solution

$$\begin{aligned}
 Y^{i,N,c}(t_{n+1}) &= Y^{i,N,c}(t_n) + b_\delta(Y^{i,N,c}(t_n), \mu_{t_n}^{Y,N,c})\delta + \sigma(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau))\Delta W_n^i \\
 &\quad + \sigma(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau))\partial_{x_1}\sigma(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau))\frac{(\Delta W_n^i)^2 - \delta}{2} \\
 &\quad + \sigma(Y^{i,N,c}(t_n - \tau), Y^{i,N,c}(t_n - 2\tau))\partial_{x_2}\sigma(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau))\frac{\Delta W_n^i\Delta W_{n-\tau}^i}{2}, \\
 &=: \mathcal{S}(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau), Y^{i,N,c}(t_n - 2\tau), \mu_{t_n}^{Y,N,c}, \delta, \Delta W_n^i, \Delta W_{n-\tau}^i\Delta W_{n-\tau}^i), \quad (3.3)
 \end{aligned}$$

where

$$Y^{i,N,c}(\theta) := \xi^i(\theta), \theta \in [-\tau, 0], \quad \mu_{t_n}^{Y,N,c}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{Y^{j,N,c}(t_n)}(dx),$$

and $\Delta W_{n-\tau}^i := W^i(t_{n+1} - \tau) - W^i(t_n - \tau)$.

The above proposed modified scheme on the coarse grid can be readily extended to the setting where the diffusion is law dependent, i.e., as in (2.1),

$$dX(t) = b(\Pi(X_t), \mathcal{L}_{\Pi(X_t)}) dt + \sigma(\Pi(X_t), \mathcal{L}_{\Pi(X_t)}) dW(t), \quad X_0 = \xi \in \mathcal{C}.$$

To be precise, consider a diffusion depending explicitly on $\mathcal{L}_{X(t)}$ and $\mathcal{L}_{X(t-\tau)}$ (i.e., $k = 2$), then this would yield the following modified scheme: for $n \in \{0, \dots, M-1\}$

$$\begin{aligned}
 Y^{i,N,c}(t_{n+1}) &= Y^{i,N,c}(t_n) + b_\delta(Y^{i,N,c}(t_n), \mu_{t_n}^{Y,N,c})\delta + \sigma(\Pi(Y_{t_n}^{i,N,c}), \Pi(\mu_{t_n}^{Y,N,c}))\Delta W_n^i \\
 &\quad + \sigma(\Pi(Y_{t_n}^{i,N,c}), \Pi(\mu_{t_n}^{Y,N,c}))\partial_{x_1}\sigma(\Pi(Y_{t_n}^{i,N,c}), \Pi(\mu_{t_n}^{Y,N,c}))\frac{(\Delta W_n^i)^2 - \delta}{2} \\
 &\quad + \sigma(\Pi(Y_{t_n-\tau}^{i,N,c}), \Pi(\mu_{t_n-\tau}^{Y,N,c}))\partial_{x_2}\sigma(\Pi(Y_{t_n}^{i,N,c}), \Pi(\mu_{t_n}^{Y,N,c}))\frac{\Delta W_n^i\Delta W_{n-\tau}^i}{2}, \\
 &\quad + \frac{1}{N} \sum_{j=1}^N \sigma(\Pi(Y_{t_n}^{j,N,c}), \Pi(\mu_{t_n}^{Y,N,c}))\partial_{\mu_1}\sigma(\Pi(Y_{t_n}^{i,N,c}), \Pi(\mu_{t_n}^{Y,N,c}))(Y_{t_n}^{j,N,c}(t_n)) \\
 &\quad \times \frac{\Delta W_n^i\Delta W_{n-\tau}^j - \mathbf{I}_{\{i=j\}}\delta}{2} \\
 &\quad + \frac{1}{N} \sum_{j=1}^N \sigma(\Pi(Y_{t_n-\tau}^{j,N,c}), \Pi(\mu_{t_n-\tau}^{Y,N,c}))\partial_{\mu_2}\sigma(\Pi(Y_{t_n}^{i,N,c}), \Pi(\mu_{t_n}^{Y,N,c}))(Y_{t_n}^{j,N,c}(t_n - \tau)) \\
 &\quad \times \frac{\Delta W_n^i\Delta W_{n-\tau}^j}{2},
 \end{aligned}$$

where

$$Y^{i,N,c}(\theta) := \xi^i(\theta), \theta \in [-\tau, 0], \quad \mu_{t_n}^{Y,N,c}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^N \delta_{Y^{j,N,c}(t_n)}(\mathrm{d}x)$$

(and similarly the processes $Y^{i,N,f}$ and $Y^{i,N,a}$, precisely defined below). The presence of L -derivative terms and associated iterated integrals can be accounted for in the proofs as the terms involving spatial derivatives without providing new mathematical challenges.

Using above assumptions, it follows that the moments of $Y^{i,N,c}$ are bounded and that $Y^{i,N,c}$ converges strongly to the solution of SDE (3.2), with strong order 1/2; see Lemma 3.1 below.

In the sequel, δ will correspond to the mesh-size of a coarse time discretization, where the elements of the time-grid are denoted by t_0, t_1, \dots, t_M . The corresponding fine time-grid will additionally include the elements $t_{1/2}, t_{3/2}, \dots, t_{M-1/2}$, where $t_{n+1/2} := (n + 1/2)\delta$ for $n \in \{0, \dots, M-1\}$. Hence, we will make use of the notation $\delta W_n^i := W(t_{n+1/2}) - W(t_n)$ and $\delta W_{n+1/2}^i := W(t_{n+1}) - W(t_{n+1/2})$. Analogously, we can define $\delta W_{n-\tau}^i$ and $\delta W_{n+1/2-\tau}^i$.

On a refined mesh with step-size $\delta/2$, we first introduce two discrete processes $Y^{i,N,f}$ and $Y^{i,N,a}$

$$\begin{aligned} Y^{i,N,f}(t_{n+1/2}) &= \mathcal{S} \left(Y^{i,N,f}(t_n), Y^{i,N,f}(t_n - \tau), Y^{i,N,f}(t_n - 2\tau), \mu_{t_n}^{Y,N,f}, \delta/2, \delta W_n^i, \delta W_n^i \delta W_{n-\tau}^i \right), \\ Y^{i,N,f}(t_{n+1}) &= \mathcal{S} \left(Y^{i,N,f}(t_{n+1/2}), Y^{i,N,f}(t_{n+1/2} - \tau), Y^{i,N,f}(t_{n+1/2} - 2\tau), \mu_{t_{n+1/2}}^{Y,N,f}, \delta/2, \delta W_{n+1/2}^i, \delta W_{n+1/2}^i \delta W_{n+1/2-\tau}^i \right) \end{aligned} \quad (3.4)$$

for $n \in \{0, \dots, M-1\}$, with

$$\mu_{t'}^{Y,N,f}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^N \delta_{Y^{j,N,f}(t')}(\mathrm{d}x), \text{ where } t' = t_n, t_{n+1/2},$$

and then the antithetic version of this process

$$\begin{aligned} Y^{i,N,a}(t_{n+1/2}) &= \mathcal{S} \left(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau), Y^{i,N,a}(t_n - 2\tau), \mu_{t_n}^{Y,N,a}, \delta/2, \delta W_{n+1/2}^i, \delta W_{n+1/2}^i \delta W_{n+1/2-\tau}^i \right), \\ Y^{i,N,a}(t_{n+1}) &= \mathcal{S} \left(Y^{i,N,a}(t_{n+1/2}), Y^{i,N,a}(t_{n+1/2} - \tau), Y^{i,N,a}(t_{n+1/2} - 2\tau), \mu_{t_{n+1/2}}^{Y,N,a}, \delta/2, \delta W_n^i, \delta W_n^i \delta W_{n-\tau}^i \right) \end{aligned} \quad (3.5)$$

with

$$\mu_{t'}^{Y,N,a}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^N \delta_{Y^{j,N,a}(t')}(\mathrm{d}x), \text{ where } t' = t_n, t_{n+1/2}.$$

We now take the average of $Y^{i,N,f}$ and $Y^{i,N,a}$ and define for all $n \in \{0, \dots, M\}$

$$\bar{Y}^{i,N,f}(t_n) := \frac{Y^{i,N,f}(t_n) + Y^{i,N,a}(t_n)}{2}. \quad (3.6)$$

Similarly, we define estimators to be used in conjunction with (1.4) as follows:

$$\phi_N^l := \frac{1}{N} \sum_{j=1}^N \frac{P(Y^{j,N,f}(T)) + P(Y^{j,N,a}(T))}{2}, \quad \varphi_N^{l-1} := \frac{1}{N} \sum_{j=1}^N P(Y^{j,N,c}(T)).$$

First, we give a lemma on stability and strong convergence of the coarse grid solution as a consequence of the proofs of the results stated in Section 2.

LEMMA 3.1 Let Assumptions (\mathbf{AAD}_b^1) – (\mathbf{AAD}_b^3) and (\mathbf{AAD}_σ^1) – (\mathbf{AAD}_σ^2) hold. For $n \in \{0, \dots, M\}$, let $Y^{i,N,c}(t_n)$ be defined as above and $X^{i,N}(t_n)$ be given by (3.2). Then, for any $p \geq 2$, there exists a constant $C > 0$ such that

$$\max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M\}} \mathbb{E}[|Y^{i,N,c}(t_n)|^p] \leq C, \quad \max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M\}} \mathbb{E}[|Y^{i,N,c}(t_n) - X^{i,N}(t_n)|^p] \leq C\delta^{p/2}.$$

The following main result reveals the first-order convergence of the difference between fine and coarse grid approximation for the antithetic scheme.

THEOREM 3.2 Let Assumptions (\mathbf{AAD}_b^1) – (\mathbf{AAD}_b^3) and (\mathbf{AAD}_σ^1) – (\mathbf{AAD}_σ^2) hold. For $n \in \{0, \dots, M\}$, let $\bar{Y}^{i,N,f}(t_n)$ be defined as in (3.6) and $Y^{i,N,c}(t_n)$ be the coarse path approximation, defined as in (3.3). Then, there exists a constant $C > 0$ such that

$$\max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M\}} \mathbb{E}[|\bar{Y}^{i,N,f}(t_n) - Y^{i,N,c}(t_n)|^2] \leq C\delta^2. \quad (3.7)$$

Proof. The proof is deferred to Section 6. \square

3.3 MLMC complexity analysis

To investigate the cost of the proposed MLMC estimator in combination with the antithetic approach, we state for the reader's convenience an adaptation of the complexity result of [Haji-Ali & Tempone \(2018\)](#) to our setting.

PROPOSITION 3.3 (cf. Theorem 3.1 in [Haji-Ali & Tempone, 2018](#)). For every $(\tilde{L}, l) \in (\mathbb{N} \cup \{0\})^2$, let $N = \tilde{\beta}^{\tilde{L}}$ for some $\tilde{\beta} > 0$, and $\phi_N^l, \varphi_N^{l-1}$ be approximations of the random variable P . For $k \in \{0, \dots, K_l\}$, let $\phi_N^l(\omega_{1:N}^{(l,k)})$ and $\varphi_N^{l-1}(\omega_{1:N}^{(l,k)})$ be the k th realizations of ϕ_N^l and φ_N^{l-1} , respectively. Consider the MLMC estimator

$$\mathcal{A}_{\text{MLMC}}(\tilde{L}, L) = \sum_{l=0}^L \frac{1}{K_l} \sum_{k=1}^{K_l} \left(\phi_N^l(\omega_{1:N}^{(l,k)}) - \varphi_N^{l-1}(\omega_{1:N}^{(l,k)}) \right),$$

with $\varphi_N^{-1} = 0$ and for $\beta, w, \gamma, s, \tilde{\gamma}, \tilde{w}, \tilde{c} > 0$, where $s \leq 2w$, assume the following:

- (i) $|\mathbb{E}[P - \phi_N^l]| = \mathcal{O}(\tilde{\beta}^{-\tilde{w}\tilde{L}} + \beta^{-wl})$
- (ii) $\mathbb{E}[\phi_N^l] = \mathbb{E}[\varphi_N^l]$
- (iii) $\mathbb{V}[\phi_N^l - \varphi_N^{l-1}] = \mathcal{O}(\tilde{\beta}^{-\tilde{c}\tilde{L}}\beta^{-sl})$
- (iv) $\text{Work}[\phi_N^l - \varphi_N^{l-1}] = \mathcal{O}(\tilde{\beta}^{\tilde{\gamma}\tilde{L}}\beta^{\gamma l})$.

Then, for any $\varepsilon < e^{-1}$, there exist \tilde{L}, L and a sequence of $(K_l)_{l \in \{0, \dots, L\}}$, such that

$$\text{MSE} := \mathbb{E}[(\mathcal{A}_{\text{MLMC}}(\tilde{L}, L) - \mathbb{E}[P])^2] \leq \varepsilon^2,$$

and

$$\text{Work}[\mathcal{A}_{\text{MLMC}}(\tilde{L}, L)] := \sum_{l=0}^L K_l \text{Work}[\phi_N^l - \varphi_N^{l-1}] = \begin{cases} \mathcal{O}\left(\varepsilon^{-2-\frac{\tilde{\gamma}-\tilde{c}}{w}}\right), & \text{if } s > \gamma, \\ \mathcal{O}\left(\varepsilon^{-2-\frac{\tilde{\gamma}-\tilde{c}}{w}} \log^2(\varepsilon)\right), & \text{if } s = \gamma, \\ \mathcal{O}\left(\varepsilon^{-2-\frac{\tilde{\gamma}-\tilde{c}}{w}-\frac{\gamma-s}{w}}\right), & \text{if } s < \gamma. \end{cases}$$

We apply this result in the setting of the antithetic approach with a fixed number of particles $N = 2^{\tilde{L}}$ across all levels and $M_l = 2^l$ time-steps on level l for $l \in \{0, \dots, L\}$.

For sufficiently regular payoff functions P (e.g., twice continuously differentiable with bounded first and second derivatives), it follows as in Theorem 3.2 that $\mathbb{E}[|\phi_N^l - \varphi_N^{l-1}|^2] \leq C\delta_l^2$ (see also Giles, 2015, Lemma 2.2). We now assume instead the stronger, multiplicative bound

$$\mathbb{E}\left[|\phi_N^l - \varphi_N^{l-1}|^2\right] \leq C \frac{\delta_l^2}{N}. \quad (3.8)$$

This bound is consistent with our numerical tests (see Fig. 1, top right) and is also the one assumed in Haji-Ali & Tempone (2018), but the proof (in the present context) is elusive so far.

Then, in the setting of Proposition 3.3, $s = 2$ and $\gamma = 1$, assuming P to be regular enough. Hence, for a given $\varepsilon > 0$, from $w = 1$, we choose the number of levels $L = \mathcal{O}(\log(\varepsilon^{-1})/\log(2))$, and assuming $\tilde{w} = 1$, we choose the number of particles across all levels $N = 2^{\tilde{L}}$ (i.e., $\tilde{L} = L$).

Following Szpruch & Tse (2021), we denote now by $p = 0, 1$ the order of interaction in the particle system, i.e., the cost required to compute all interaction terms is of order N^{p+1} , and $\tilde{\gamma} = p + 1$, $\gamma = 1$ above. The proof of Theorem 3.1 in Haji-Ali & Tempone (2018) then reveals that the optimal number of particle systems per level is $K_l = \mathcal{O}(\varepsilon^{-1}2^{-3l/2})$. Note that the implied constants do not depend on l .

From Proposition 3.3, we hence deduce the overall computational cost as follows.

COROLLARY 3.4 Under the assumption of (3.8), the optimal complexity of the antithetic multilevel estimator for MSE ε^2 is of order ε^{-2-p} , for interaction terms of order p .

For standard MLMC without the antithetic technique (i.e., $s = 1$), setting $K_l = \mathcal{O}(\varepsilon^{-1}(L+1)2^{-l})$ (with the same choices of L, M_l and N), we derive that the overall cost is of order $\varepsilon^{-2-p} \log^2(\varepsilon)$. A plain

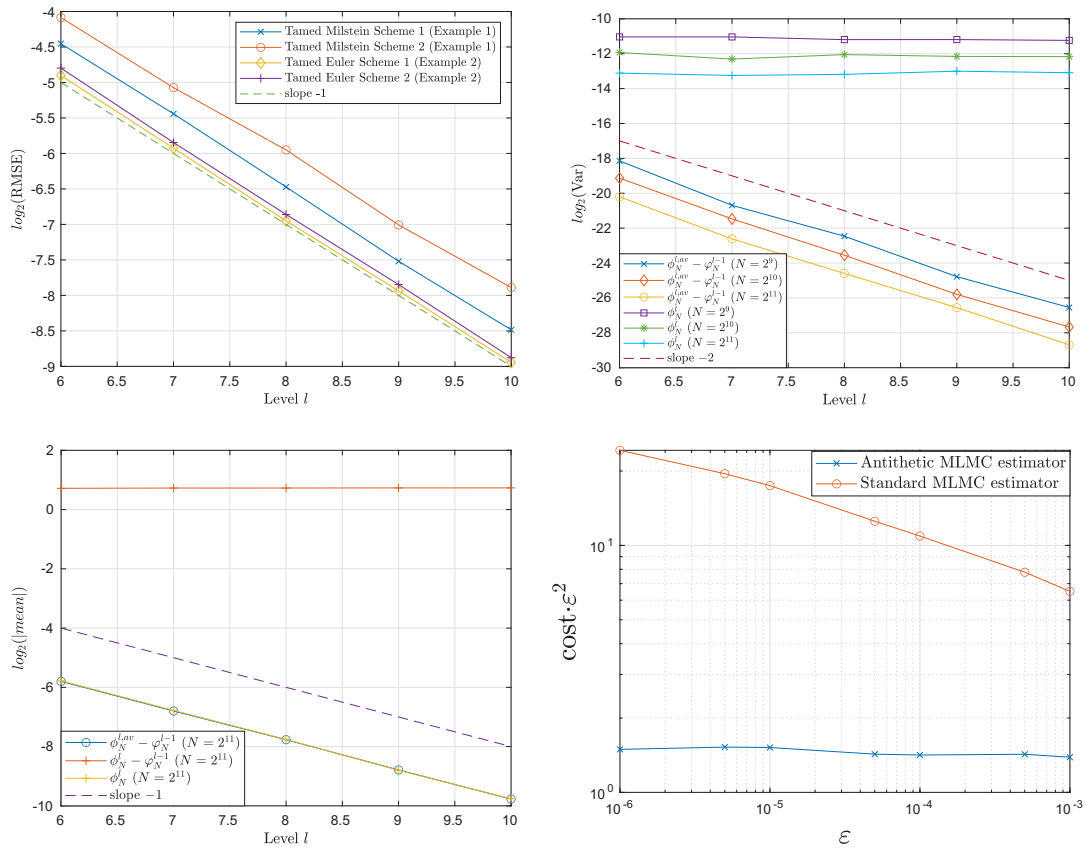


FIG. 1. Left top: strong convergence of the tamed Milstein schemes (Scheme 1 and 2) for Example 1 and the tamed Euler–Maruyama schemes for Example 2 (equivalent to Milstein in this case). The following are for Scheme 2. Right top: variance decay of the multilevel correction terms using the antithetic approach compared to standard MC. Left bottom: decay of the expected value of the multilevel correction terms using the antithetic approach compared to standard MC. Right bottom: computational complexity.

MC approach gives order ε^{-3-p} , as N, K and M need to be chosen of order $\mathcal{O}(\varepsilon^{-1})$ to obtain the desired accuracy.

Proposition 3.3 contrasts with the classical MLMC setting of Giles & Szpruch (2014) in that the multilevel decomposition of the estimator only acts in the index related to l , but not the one related to \tilde{L} . That is to say, in our setting, we vary the number of time-steps M across levels, but not the number of particles N . This leads to optimal complexity because the variance is assumed to decay in both M and N due to (3.8).

In the setting without delay and for a constant diffusion coefficient, Szpruch & Tse (2021) proposes an antithetic scheme with respect to the number of particles—as opposed to antithetic in the discrete-time paths as we do here—and proves an additive variance bound of the form $\mathcal{O}(\delta^2 + 1/N^2)$. The order $1/N^2$ is an improvement over the order $1/N$, which is expected for propagation of chaos without antithetic sampling, while the order δ^2 holds for the Euler–Maruyama scheme because it coincides with

the Milstein scheme for constant diffusion coefficient. This allows a similar complexity analysis as presented above; in particular, [Szpruch & Tse \(2021, Theorem 6.3\)](#) shows how to choose a sequence K_l in order to obtain a result analogous to Corollary 3.4.

To obtain a combined MLMC estimator, i.e., employing the antithetic approach proposed in our paper together with the antithetic scheme with respect to the number of particles, we would choose $M_l = N_l = 2^l$ and define

$$\phi_{N_l}^l := \frac{1}{N_l} \sum_{j=1}^{N_l} \frac{P(Y^{j,N_l,f}(T)) + P(Y^{j,N_l,a}(T))}{2}, \quad \phi_{N_l}^{l-1} := \frac{1}{N_l} \sum_{j=1}^{N_l} \frac{P(Y^{j,N_l,(1),c}(T)) + P(Y^{j,N_l,(2),c}(T))}{2},$$

where $Y^{j,N_l,(1),c}(T)$ is a particle system of size $N_l/2$ and uses the first $N_l/2$ Brownian motions and initial data from the set $(W^i, \xi^i)_{i \in \{1, \dots, N_l\}}$, while $Y^{j,N_l,(2),c}(T)$ is the system associated with the other half.

As commented above, such an estimator is expected to give a variance decay $\mathcal{O}(2^{-2l})$. The optimal complexity under this assumption is the same as derived in Corollary 3.4.

4. Numerical results

We now present a number of numerical tests to illustrate the practical behaviour of the schemes proposed in this article. We use the canonical particle approximation ([Bossy & Talay, 1997](#)) to the law $\mathcal{L}_{Y_{t_n}}$ at each time-step t_n by its empirical distribution. For our numerical experiments, we used $N = 10^3$, unless stated otherwise.

As we do not know the exact solution in the considered examples, the convergence rates with respect to the number of time-steps were determined by comparing two solutions (at time $T = 1$) computed on a fine and coarse time grid, respectively, where the same Brownian increments were used for both. In order to assess the strong convergence in δ , we thus compute the root-mean-square error (RMSE)

$$\text{RMSE} := \sqrt{\frac{1}{N} \sum_{j=1}^N (Y^{j,N,l}(T) - Y^{j,N,l-1}(T))^2},$$

where $Y^{j,N,l}(T)$ denotes the approximation of X at time T computed with N particles and $2^l T$ time-steps.

For simplicity, we assume that s_1, \dots, s_k are contained in the considered time-grid, i.e., that they are of the form $-n\delta_{l_0}$ (bluewhere l_0 refers to the coarsest levels) for some non-negative integer n .

4.1 First-order strong convergence of the Milstein scheme

This section numerically illustrates the convergence of the tamed Milstein schemes for point-delay McKean–Vlasov SDEs for two test cases. Specifically, we compare Schemes 1 and 2 from Section 2 in Fig. 1, top left. We also give implementation details of the scheme including the computation of the iterated stochastic integrals.

Example 1: Here, we consider the point-delay McKean–Vlasov SDE

$$\begin{aligned} dX(t) = & \left(1 - (X(t))^3 + X(t) + \frac{1}{k} \sum_{l=1}^k X(t + s_l) + \frac{1}{k} \sum_{l=1}^k \mathbb{E}[X(t + s_l)] \right) dt \\ & + \frac{1}{k} \sum_{l=1}^k X(t + s_l) dW(t), \quad t \in [0, T], \end{aligned}$$

with the initial value $X_0 = \xi = \mathbf{0} \in \mathcal{C}$, $s_1, \dots, s_k \in [-\tau, 0]$. Hence, the tamed Milstein approximation for this example results in the particle system

$$\begin{aligned} Y^{i,N}(t_{n+1}) = & Y^{i,N}(t_n) + b_\delta(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N}))\delta + \sigma(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N}))\Delta W_n^i \\ & + \sum_{l=1}^k \partial_{x_l} \sigma(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N}))\tilde{\sigma}(t_n + s_l, \Pi(Y_{t_n+s_l}^{i,N}), \Pi(\mu_{t_n}^{Y,N}))I^i(t_n + s_l, t_{n+1} + s_l; s_l), \end{aligned}$$

where the driving Brownian motions W^i , $i \in \mathbb{S}_N$, are independent, and

$$\begin{aligned} b(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N})) &= 1 - (Y^{i,N}(t_n))^3 + Y^{i,N}(t_n) + \frac{1}{k} \sum_{l=1}^k Y^{i,N}(t_n + s_l) + \frac{1}{N} \sum_{j=1}^N \frac{1}{k} \sum_{l=1}^k Y^{j,N}(t_n + s_l), \\ \sigma(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N})) &= \frac{1}{k} \sum_{l=1}^k Y^{i,N}(t_n + s_l). \end{aligned}$$

In addition, for Scheme 1, we have

$$b_\delta(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N})) = \frac{b(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N}))}{1 + \delta |b(\Pi(Y_{t_n}^{i,N}), \Pi(\mu_{t_n}^{Y,N}))|},$$

and with the additional exponent \bar{q} for Scheme 2. Further, recall that ∂_{x_l} denotes the derivative with respect to the l th state component.

Above, $I^i(t_n + s_l, t_{n+1} + s_l; s_l)$ denotes the iterated stochastic integrals, with are approximated using a truncation parameter $r = M^{1/2}$ as described in Section 2.2.

In the test, the delay parameters were $\tau = 1/8$ and $k = 2$ (i.e., $s_1 = 0$ and $s_2 = -\tau$). The strong convergence of the discretized particle system is depicted in Fig. 1, top left, where we observe the expected order one.

Example 2: In analogy to above, we additionally study the example

$$\begin{aligned} dX(t) = & \left(1 - (X(t))^3 + X(t) + \frac{1}{k} \sum_{l=1}^k X(t + s_l) + \frac{1}{k} \sum_{l=1}^k \mathbb{E}[X(t + s_l)] \right) dt \\ & + \left(\frac{1}{k} \sum_{l=1}^k \mathbb{E}[X(t + s_l)] \right) dW(t), \quad t \in [0, T], \end{aligned}$$

with the initial value $X_0 = \xi = \mathbf{0} \in \mathcal{C}$, $s_1, \dots, s_k \in [-\tau, 0]$. As the diffusion coefficient does not depend on the state, the Milstein scheme reduces to the Euler–Maruyama scheme. We confirm numerically in Fig. 1, top left, that the tamed Euler–Maruyama scheme converges in this case with order one.

4.2 Convergence and complexity of the antithetic multilevel sampling scheme

Here, we numerically illustrate the performance of the MLMC Milstein method that uses the antithetic approach presented in Section 3 applied to Example 1 from above. We focus on Scheme 2, which was found to be slightly more accurate numerically in the previous section. In particular, we will demonstrate the variance decay of the multilevel correction terms described in equation (1.5) and investigate the overall computational complexity to estimate $\mathbb{E}[P(X_T)]$ with given accuracy $\varepsilon > 0$ when the estimator (1.4) is used.

Using the same notation as in Section 2, the scheme has the form

$$\begin{aligned} Y^{i,N}_{t_{n+1}} = & Y^{i,N}_{t_n} + b_\delta(\Pi(Y^{i,N}_{t_n}), \Pi(\mu^{Y,N}_{t_n}))\delta + \sigma(\Pi(Y^{i,N}_{t_n}), \Pi(\mu^{Y,N}_{t_n}))\Delta W_n^i \\ & + \frac{1}{2} \sum_{l=1}^k \partial_{x_l} \sigma(\Pi(Y^{i,N}_{t_n}), \Pi(\mu^{Y,N}_{t_n})) \tilde{\sigma}(t_n + s_l, \Pi(Y_{t_n+s_l}), \Pi(\mu^{Y,N}_{t_n+s_l})) (\bar{W}^i(\delta) \bar{B}^i(\delta) - \mathbf{I}_{\{s_l=0\}} \delta). \end{aligned}$$

As payoff function, we choose for simplicity $P(x) = x$, but a similar behaviour is expected for any smooth enough P .

Now, on a fine time-grid, we compute two approximate solutions to the particle system at time T , denoted by $Y^{i,N,f}(T)$ and $Y^{i,N,a}(T)$, where the simulation of $Y^{i,N,a}(T)$ is based on the antithetic Brownian paths of the paths used to compute $Y^{i,N,f}(T)$, i.e., the odd and even Brownian increments are interchanged. We then determine the variance (1.5), where $\phi_N^{l,av}$ and ϕ_N^{l-1} are estimated using the average $(Y^{i,N,f}(T) + Y^{i,N,a}(T))/2$ and $Y^{i,N,c}(T)$, a coarse path solution, respectively. The quantity $\phi_N^{l,av}$ is estimated using the samples $Y^{i,N,f}(T)$ only. We choose $N \in \{2^9, 2^{10}, 2^{11}\}$, to numerically confirm the additional factor $1/N$ in (1.5).

The computations are based on the choices $k = 2$ and $\tau = 1/8$ using the SDE in Example 1 given in the previous section. Figure 1 shows that the strong convergence rate of the proposed scheme is one (i.e., $s = 2$). The variance of the standard MC estimator barely varies with the levels. Additionally, in Fig. 1, we depict the decay of the the expected value of the MLMC correction terms. We observe that this decay is of order one. As anticipated, the antithetic MLMC and standard MLMC estimator have the same expected value.

To investigate the required complexity of the proposed MLMC estimator to achieve an accuracy $\varepsilon > 0$ (in the RMSE sense), we plot $\varepsilon^{-2} \cdot \text{cost}$ against the accuracy ε , see Fig. 1. Here, cost denotes

the complexity to compute (1.4) with given $L, K_l, M_l = 2^l$ and N , i.e., $\text{cost} = \mathcal{O}(N \sum_{l=0}^L K_l M_l)$. Since, in Example 1, the law dependence is given by an expectation, the empirical distribution only has to be computed once (at cost N) at each time-step, so $p = 0$ in Section 3.3.

5. Proof of Theorem 2.4

Here, and throughout the remaining article, we write $a \lesssim b$ to express that there exists a constant $C > 0$ such that $a \leq Cb$, where $a, b \in \mathbb{R}$. We remark that the implied constant $C > 0$ might be dependent on p, ε, T, m, d , i.e., $C = C(\varepsilon, p, T, m, d)$, but is independent of M and N ; also, they may change their values from line to line in a sequence of inequalities. We restrict the theoretical investigation to Scheme 2, as proving a moment bound for Scheme 1 is very involved and requires a lengthy (nonstandard) analysis.

5.1 Proof of Theorem 2.4

For the proof of Theorem 2.4, we will focus on the key differences to the nondelay case (see Bao *et al.*, 2021) and omit other details. We again consider the one point-delay case for most parts of the subsequent discussion. We define, for $t \geq 0$ and $i \in \mathbb{S}_N$,

$$\begin{aligned} \Psi_t^{i,k} &:= \sum_{l=1}^k \partial_{x_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \tilde{\sigma}(\Pi(Y_{t_\delta+s_l}^{i,N}), \Pi(\Pi(Y_{t_\delta+s_l}^{i,N}), \Pi(\mu_{t_\delta+s_l}^{Y,N}))(\mu_{t_\delta+s_l}^{Y,N})) \int_{t_\delta+s_l}^{t+s_l} dW^i(r) \\ &\quad + \sum_{l=1}^k \frac{1}{N} \sum_{j=1}^N \partial_{\mu_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) (Y_{t_\delta+s_l}^{j,N}) \tilde{\sigma}(t_\delta+s_l, \Pi(Y_{t_\delta+s_l}^{i,N}), \Pi(\mu_{t_\delta+s_l}^{Y,N})) \int_{t_\delta+s_l}^{t+s_l} dW^j(r), \\ \Upsilon_t^{i,k} &:= \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) + \Psi_t^{i,k}, \\ \Gamma_t^{i,k} &:= \sigma(\Pi(Y_t^{i,N}), \Pi(\mu_t^{Y,N})) - \Upsilon_t^{i,k} \\ M_t^{i,k} &:= \int_0^t (\sigma(\Pi(X_s^{i,N}), \Pi(\mu_s^{X,N})) - \Upsilon_s^{i,k}) dW^i(s). \end{aligned} \quad (5.1)$$

LEMMA 5.1 Let $Y^{i,N}(t_n)$ for $n \in \{0, \dots, M\}$, be defined as in (2.3). Then, under Assumptions (\mathbf{AD}_b^1) , (\mathbf{AD}_σ^1) – (\mathbf{AD}_σ^2) and (\mathbf{H}_1) , for any $p \geq 2$, there exists a constant $C > 0$ (independent of N and M) such that

$$\max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M\}} \mathbb{E}[|Y^{i,N}(t_n)|^p] \leq C.$$

Proof. The proof can be carried out using Itô's formula along with a standard Gronwall-type argument (see, e.g., Kumar & Neelima, 2021, Lemma 4) in combination with an inductive procedure outlined in the proof of Proposition 2.1. We note that item (a) in (\mathbf{AD}_σ^2) on the uniform boundedness of derivatives

of σ with respect to the delay variables implies that

$$\begin{aligned} & \mathbb{E} \left[\left| \partial_{x_2} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \tilde{\sigma}(t_\delta + s_2, \Pi(Y_{t_\delta+s_2}^{i,N}), \Pi(\mu_{t_\delta+s_2}^{Y,N})) \int_{t_\delta+s_2}^{t+s_2} dW^i(r) \right|^p \right] \\ & \lesssim \mathbb{E} \left[\left| \int_{t_\delta+s_2}^{t+s_2} \tilde{\sigma}(t_\delta + s_2, \Pi(Y_{t_\delta+s_2}^{i,N}), \Pi(\mu_{t_\delta+s_2}^{Y,N})) dW^i(r) \right|^p \right] \\ & \lesssim \delta^{p/2} \left(1 + \mathbb{E}[|Y^{i,N}(t_\delta - \tau)|^p] + \mathbb{E}[|Y^{i,N}(t_\delta - 2\tau)|^p] \right), \end{aligned}$$

where we used the growth assumptions on σ (\mathbf{AD}_σ^1) with $q_2 = 0$, and for simplicity consider $k = 2$ and $s_2 = -\tau$.

Observe that the uniform boundedness of derivatives with respect to the delay variables is crucial, as in the term

$$\partial_{x_2} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \int_{t_\delta+s_2}^{t+s_2} dW^i(r),$$

$\partial_{x_2} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N}))$ is anticipative.

However, if the second assumption item (b) in (\mathbf{AD}_σ^2) holds, then it becomes nonanticipative. The growth assumption on $\partial_{x_2} \sigma_u \sigma$ in (\mathbf{AD}_σ^2) gives

$$\begin{aligned} & \mathbb{E} \left[\left| \partial_{x_2} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \tilde{\sigma}(t_\delta + s_2, \Pi(Y_{t_\delta+s_2}^{i,N}), \Pi(\mu_{t_\delta+s_2}^{Y,N})) \int_{t_\delta+s_2}^{t+s_2} dW^i(r) \right|^p \right] \\ & \lesssim \delta^{p/2} \left(1 + \mathbb{E}[|Y^{i,N}(t_\delta - \tau)|^{p(1+q_3)}] + \mathbb{E}[|Y^{i,N}(t_\delta - 2\tau)|^{p(1+q_3)}] + \mathbb{E}[|Y^{i,N}(t_\delta - \tau)|^p] \right. \\ & \quad \left. + \mathbb{E}[|Y^{i,N}(t_\delta - 2\tau)|^p] \right). \end{aligned}$$

□

REMARK 5.2 It seems that techniques from anticipative calculus, see [Nualart \(2006\)](#), are necessary if one wishes to allow (unbounded) mixed terms involving delay and nondelay variables in the diffusion coefficient.

In the sequel, we set $Z^{i,N} := X^{i,N} - Y^{i,N}$, where $(X^{i,N})_{i \in \mathbb{S}_N}$ is the particle system defined in (2.2) and $(Y^{i,N})_{i \in \mathbb{S}_N}$ is defined by (2.3).

LEMMA 5.3 Let Assumptions (\mathbf{AD}_b^1), (\mathbf{AD}_σ^1)–(\mathbf{AD}_σ^3), (\mathbf{H}_1) hold, and $k = 2$ (for simplicity). Then, for any $p \geq 2$, there is a constant $C > 0$ such that, for all $t \in [0, T]$

$$\begin{aligned} \Lambda_t^{i,2,p} &:= \int_0^t \left(\mathbb{E}[\|\sigma(\Pi(X_s^{i,N}), \Pi(\mu_s^{X,N})) - \gamma_s^{i,2}\|^{2p}] \right)^{\frac{1}{p}} ds \\ &\leq C \left\{ \delta^2 + \int_0^t (\mathbb{E}[|Z^{i,N}(s)|^{2p}])^{\frac{1}{p}} ds \right. \\ &\quad \left. + \int_0^{0 \vee (t-\tau)} (\mathbb{E}[|Z^{i,N}(s)|^{2p} (1 + |X^{i,N}(s)|^{2pq_2} + |Y^{i,N}(s)|^{2pq_2})])^{\frac{1}{p}} ds \right\}. \end{aligned} \quad (5.2)$$

Proof. From (\mathbf{AD}_σ^1) (possibly with $q_2 = 0$) and Minkowski's inequality, we derive that

$$\begin{aligned}
 \Lambda_t^{i,2,p} &\lesssim \int_0^t (\mathbb{E}[\|\sigma(\Pi(X_s^{i,N}), \Pi(\mu_s^{X,N})) - \sigma(\Pi(Y_s^{i,N}), \Pi(\mu_s^{Y,N}))\|^{2p}])^{\frac{1}{p}} ds + \int_0^t (\mathbb{E}[\|\Gamma_s^{i,2}\|^{2p}])^{\frac{1}{p}} ds \\
 &\lesssim \int_0^t (\mathbb{E}[|Z^{i,N}(s)|^{2p}])^{\frac{1}{p}} ds + \frac{1}{N} \sum_{j=1}^N \int_0^t (\mathbb{E}[|Z^{j,N}(s)|^{2p}])^{\frac{1}{p}} ds \\
 &\quad + \int_0^{0 \vee (t-\tau)} (\mathbb{E}[|Z^{i,N}(s)|^{2p} (1 + |X^{i,N}(s)|^{2pq_2} + |Y^{i,N}(s)|^{2pq_2})])^{\frac{1}{p}} ds \\
 &\quad + \frac{1}{N} \sum_{j=1}^N \int_0^{0 \vee (t-\tau)} (\mathbb{E}[|Z^{j,N}(s)|^{2p}])^{\frac{1}{p}} ds + \int_0^t (\mathbb{E}[\|\Gamma_s^{i,k}\|^{2p}])^{\frac{1}{p}} ds \\
 &\lesssim \int_0^t (\mathbb{E}[|Z^{i,N}(s)|^{2p}])^{\frac{1}{p}} ds + \int_0^t (\mathbb{E}[\|\Gamma_s^{i,k}\|^{2p}])^{\frac{1}{p}} ds \\
 &\quad + \int_0^{0 \vee (t-\tau)} (\mathbb{E}[|Z^{i,N}(s)|^{2p} (1 + |X^{i,N}(s)|^{2pq_2} + |Y^{i,N}(s)|^{2pq_2})])^{\frac{1}{p}} ds,
 \end{aligned}$$

where in the last display we used the fact that $Z^{j,N}, j \in \mathbb{S}_N$, are identically distributed. Consequently, to derive (5.2), it is sufficient to show that for $t \in [0, T]$,

$$(\mathbb{E}[\|\Gamma_t^{i,2}\|^{2p}])^{\frac{1}{p}} \lesssim \delta^2. \quad (5.3)$$

In the sequel, we will restrict the discussion to $d = m = 1$ for ease of notation. First, we write

$$\begin{aligned}
 &\sigma(Y^{i,N}(t), Y^{i,N}(t-\tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) - \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta-\tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) \\
 &= \sigma(Y^{i,N}(t), Y^{i,N}(t-\tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) - \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t-\tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) \\
 &\quad + \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t-\tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) - \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta-\tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) \\
 &\quad + \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta-\tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) - \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta-\tau), \mu_t^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) \\
 &\quad + \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta-\tau), \mu_t^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) - \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta-\tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) \\
 &=: \sum_{i=1}^4 \mathcal{E}_i.
 \end{aligned}$$

Next, using the definition $\Delta Y^{i,N}(t) := Y^{i,N}(t) - Y^{i,N}(t_\delta)$, we observe that

$$\begin{aligned}
 \mathcal{E}_1 &= \int_0^1 \frac{d}{d\lambda} \sigma(Y^{i,N}(t_\delta) + \lambda \Delta Y^{i,N}(t), Y^{i,N}(t-\tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) d\lambda \\
 &= \int_0^1 \partial_{x_1} \sigma(Y^{i,N}(t_\delta) + \lambda \Delta Y^{i,N}(t), Y^{i,N}(t-\tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) \Delta Y^{i,N}(t) d\lambda.
 \end{aligned}$$

Therefore, simple computations show, due to assumption (\mathbf{AD}_σ^3)

$$\begin{aligned}
 & \mathbb{E} \left[\left| \mathcal{E}_1 - \partial_{x_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N}) \Delta Y^{i,N}(t) \right|^2 \right] \\
 &= \mathbb{E} \left[\left| \int_0^1 \left(\partial_{x_1} \sigma(Y^{i,N}(t_\delta) + \lambda \Delta Y^{i,N}(t), Y^{i,N}(t - \tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) \right. \right. \right. \\
 &\quad \left. \left. \left. - \partial_{x_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N}) \right) \Delta Y^{i,N}(t) \, d\lambda \right|^2 \right] \\
 &\lesssim \mathbb{E}[|Y^{i,N}(t) - Y^{i,N}(t_\delta)|^4] + (\mathbb{E}[|Y^{i,N}(t) - Y^{i,N}(t_\delta)|^4])^{1/2} (\mathbb{E}[|Y^{i,N}(t - \tau) - Y^{i,N}(t_\delta - \tau)|^4])^{1/2} \\
 &\quad + \left(\frac{1}{N} \sum_{j=1}^N \mathbb{E}[|Y^{j,N}(t) - Y^{j,N}(t_\delta)|^4] \right)^{1/2} (\mathbb{E}[|Y^{i,N}(t) - Y^{i,N}(t_\delta)|^4])^{1/2} \lesssim \delta^2, \tag{5.4}
 \end{aligned}$$

where we used (\mathbf{H}_1) and $\mathbb{E}[|Y^{i,N}(t) - Y^{i,N}(t_\delta)|^4] \lesssim \delta^2$. The claim regarding the one-step error is a consequence of the identity

$$\begin{aligned}
 Y^{i,N}(t) - Y_{t_\delta}^{i,N} &= b_\delta(Y^{i,N}(t_\delta), \mu_{t_\delta}^{Y,N})(t - t_\delta) + \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \int_{t_\delta}^t dW^i(s) \\
 &\quad + \sum_{l=1}^2 \partial_{x_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta + s_l}^{i,N}), \Pi(\mu_{t_\delta + s_l}^{Y,N})) \int_{t_\delta}^t \int_{t_\delta + s_l}^{s+s_l} dW^i(u) dW^i(s) \\
 &\quad + \sum_{l=1}^2 \frac{1}{N} \sum_{j=1}^N \partial_{\mu_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N}))(Y_{t_\delta + s_l}^{j,N}) \\
 &\quad \times \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta + s_l}^{j,N}), \Pi(\mu_{t_\delta + s_l}^{Y,N})) \int_{t_\delta}^t \int_{t_\delta + s_l}^{s+s_l} dW^j(u) dW^i(s), \tag{5.5}
 \end{aligned}$$

and standard inequalities along with the moment stability of $Y^{i,N}$. A similar estimate holds if \mathcal{E}_1 is replaced by \mathcal{E}_2 .

Using [Carmona & Delarue \(2017, Proposition 5.35\)](#), we get that

$$\begin{aligned}
 \mathcal{E}_4 &= \int_0^1 \frac{d}{d\lambda} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_t^{\lambda, Y, N}, \mu_{t_\delta - \tau}^{Y, N}) \, d\lambda \\
 &= \int_0^1 \frac{1}{N} \sum_{j=1}^N \partial_{\mu_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_t^{\lambda, Y, N}, \mu_{t_\delta - \tau}^{Y, N})(Y_{t_\delta}^{j,N} + \lambda \Delta Y^{j,N}(t)) \Delta Y^{j,N}(t) \, d\lambda,
 \end{aligned}$$

with the definition

$$\mu_t^{\lambda, Y, N}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{Y_{t_\delta}^{j,N} + \lambda \Delta Y^{j,N}(t)}(dx).$$

By virtue of (\mathbf{AD}_σ^3) and employing similar estimates as in (5.4), we derive

$$\begin{aligned}
& \mathbb{E} \left[\left| \mathcal{E}_4 - \frac{1}{N} \sum_{j=1}^N \partial_{\mu_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N})(Y^{j,N}(t_\delta)) \Delta Y^{j,N}(t) \right|^2 \right] \\
&= \mathbb{E} \left[\left| \int_0^1 \frac{1}{N} \sum_{j=1}^N \partial_{\mu_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_t^{Y,N}, \mu_{t_\delta - \tau}^{Y,N})(Y^{j,N}(t_\delta) + \lambda \Delta Y^{j,N}(t)) \Delta Y^{j,N}(t) d\lambda \right. \right. \\
&\quad \left. \left. - \int_0^1 \frac{1}{N} \sum_{j=1}^N \partial_{\mu_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N})(Y^{j,N}(t_\delta)) \Delta Y^{j,N}(t) d\lambda \right|^2 \right] \\
&\lesssim \delta^2.
\end{aligned} \tag{5.6}$$

Analogous computations hold true if \mathcal{E}_4 is replaced by \mathcal{E}_3 . Above considerations motivate to express

$$\begin{aligned}
\Gamma_t^{i,k} &= \sigma(Y^{i,N}(t), Y^{i,N}(t - \tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) - \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N}) \\
&\quad - \sum_{l=1}^2 \partial_{x_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta + s_l}^{i,N}), \Pi(\mu_{t_\delta + s_l}^{Y,N})) \int_{t_\delta + s_l}^{t + s_l} dW^i(r) \\
&\quad - \sum_{l=1}^2 \frac{1}{N} \sum_{j=1}^N \partial_{\mu_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N}))(Y^{j,N}(t_\delta + s_l)) \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta + s_l}^{j,N}), \Pi(\mu_{t_\delta + s_l}^{Y,N})) \\
&\quad \int_{t_\delta + s_l}^{t + s_l} dW^j(r) \\
&= \sigma(Y^{i,N}(t), Y^{i,N}(t - \tau), \mu_t^{Y,N}, \mu_{t-\tau}^{Y,N}) - \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N}) \\
&\quad - \partial_{x_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N}) \Delta Y^{i,N}(t) \\
&\quad - \partial_{x_2} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N}) \Delta Y^{i,N}(t - \tau) \\
&\quad - \frac{1}{N} \sum_{j=1}^N \partial_{\mu_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N})(Y^{j,N}(t_\delta)) \Delta Y^{j,N}(t) \\
&\quad - \frac{1}{N} \sum_{j=1}^N \partial_{\mu_2} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta - \tau}^{Y,N})(Y^{j,N}(t_\delta - \tau)) \Delta Y^{j,N}(t - \tau) \\
&\quad - \sum_{l=1}^2 \partial_{x_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta + s_l}^{i,N}), \Pi(\mu_{t_\delta + s_l}^{Y,N})) \int_{t_\delta + s_l}^{t + s_l} dW^i(r)
\end{aligned} \tag{5.7}$$

$$\begin{aligned}
 & - \sum_{l=1}^2 \frac{1}{N} \sum_{j=1}^N \partial_{\mu_l} \sigma(\Pi(Y^{i,N}(t_\delta)), \Pi(\mu_{t_\delta}^{Y,N})) (Y^{j,N}(t_\delta + s_l)) \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta+s_l}^{j,N}), \Pi(\mu_{t_\delta+s_l}^{Y,N})) \\
 & \times \int_{t_\delta+s_l}^{t+s_l} dW^j(r)
 \end{aligned} \tag{5.8}$$

$$+ \partial_{x_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) \Delta Y^{i,N}(t) \tag{5.9}$$

$$+ \partial_{x_2} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) \Delta Y^{i,N}(t - \tau) \tag{5.10}$$

$$+ \frac{1}{N} \sum_{j=1}^N \partial_{\mu_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) (Y^{j,N}(t_\delta)) \Delta Y^{j,N}(t) \tag{5.11}$$

$$+ \frac{1}{N} \sum_{j=1}^N \partial_{\mu_2} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) (Y^{j,N}(t_\delta - \tau)) \Delta Y^{j,N}(t - \tau). \tag{5.12}$$

Notice that (5.5) implies

$$\begin{aligned}
 & \partial_{x_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) \Delta Y^{i,N}(t) \\
 & = \partial_{x_1} \sigma(Y^{i,N}(t_\delta), Y^{i,N}(t_\delta - \tau), \mu_{t_\delta}^{Y,N}, \mu_{t_\delta-\tau}^{Y,N}) \left(b_\delta(Y^{i,N}(t_\delta), \mu_{t_\delta}^{Y,N})(t - t_\delta) \right. \\
 & \quad + \sigma(\Pi(Y_{t_\delta}^i), \Pi(\mu_{t_\delta}^{Y,N})) \int_{t_\delta}^t dW^i(s) \\
 & \quad + \sum_{l=1}^2 \partial_{x_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta+s_l}^{i,N}), \Pi(\mu_{t_\delta+s_l}^{Y,N})) \int_{t_\delta}^t \int_{t_\delta+s_l}^{s+s_l} dW^i(u) dW^i(s) \\
 & \quad + \sum_{l=1}^2 \frac{1}{N} \sum_{j=1}^N \partial_{\mu_l} \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) (Y_{t_\delta+s_l}^{j,N}) \\
 & \quad \left. \times \tilde{\sigma}(t_\delta + s_l, \Pi(Y_{t_\delta+s_l}^{j,N}), \Pi(\mu_{t_\delta+s_l}^{Y,N})) \int_{t_\delta}^t \int_{t_\delta+s_l}^{s+s_l} dW^j(u) dW^i(s) \right).
 \end{aligned}$$

Similar expressions can be derived for equations (5.10), (5.11) and (5.12) given above. From this, we observe that (5.7) and (5.8) cancel with nonhigher order terms appearing in (5.9), (5.10), (5.11) and (5.12). The remaining terms are readily proven to be of order δ . Recalling the estimates (5.6) and (5.4) allows us to deduce the claim. \square

LEMMA 5.4 We define for $t \in [0, T]$

$$\hat{\gamma}_t^i := \partial_x b(Y^{i,N}(t_\delta), \mu_{t_\delta}^{Y,N}) \sigma(\Pi(Y_{t_\delta}^{i,N}), \Pi(\mu_{t_\delta}^{Y,N})) \int_{t_\delta}^t dW^i(r).$$

Let Assumptions (\mathbf{AD}_b^1) , $(\mathbf{AD}_\sigma^1)-(\mathbf{AD}_\sigma^3)$, (\mathbf{H}_1) hold (and set $k = 2$ for simplicity). Then, for all $\varepsilon > 0$ and $p \geq 2$, there exists $C > 0$ such that

$$\begin{aligned} & \left(\mathbb{E} \left[\left\| \int_0^\cdot \langle Z^{i,N}(s), \hat{\gamma}_s^i \rangle \, ds \right\|_{\infty,t}^p \right] \right)^{\frac{1}{p}} \leq \varepsilon \left((\mathbb{E}[\|Z^{i,N}\|_{\infty,t}^{2p}])^{\frac{1}{p}} \right. \\ & \quad + (\mathbb{E}[\|Z^{i,N}\|_{\infty,t-\tau}^{2p} (1 + \|X^{i,N}\|_{\infty,t-\tau}^{2pq_2} + \|Y^{i,N}\|_{\infty,t-\tau}^{2pq_2})])^{\frac{1}{p}} \\ & \quad \left. + C \left\{ \int_0^t (\mathbb{E}[|Z^{i,N}(s)|^{2p}])^{\frac{1}{p}} \, ds + \int_0^{0 \vee (t-\tau)} (\mathbb{E}[|Z^{i,N}(s)|^{2p} (1 + |X^{i,N}(s)|^{2pq_2} + |Y^{i,N}(s)|^{2pq_2})])^{\frac{1}{p}} \, ds + \delta^2 \right\} \right), \end{aligned} \quad (5.13)$$

for $t \in [0, T]$. The same holds if $\hat{\gamma}_t^i$ is replaced by

$$\frac{1}{N} \sum_{j=1}^N \partial_\mu b(Y^{i,N}(t_\delta), \mu_{t_\delta}^{Y,N})(Y^{j,N}(t_\delta)) \sigma(\Pi(Y_{t_\delta}^{j,N}), \Pi(\mu_{t_\delta}^{Y,N})) \int_{t_\delta}^t dW^j(r).$$

Proof. The result can be proven following the same steps as Bao *et al.* (2021, Lemma 3.2) and its proof therefore omitted. \square

LEMMA 5.5 Let Assumptions $(\mathbf{AD}_b^1)-(\mathbf{AD}_b^3)$, $(\mathbf{AD}_\sigma^1)-(\mathbf{AD}_\sigma^3)$, (\mathbf{H}_1) hold (and set $k = 2$ for simplicity). Then, for all $\varepsilon > 0$ and $p \geq 2$, there exists a constant $C > 0$ such that

$$\begin{aligned} & \int_0^t (\mathbb{E}[|Z^{i,N}(s), b(Y^{i,N}(s), \mu_s^{Y,N}) - b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N})|)^p])^{\frac{1}{p}} \, ds \\ & \leq \varepsilon \left((\mathbb{E}[\|Z^{i,N}\|_{\infty,t}^{2p}])^{\frac{1}{p}} + (\mathbb{E}[\|Z^{i,N}\|_{\infty,t-\tau}^{2p} (1 + \|X^{i,N}\|_{\infty,t-\tau}^{2pq_2} + \|Y^{i,N}\|_{\infty,t-\tau}^{2pq_2})])^{\frac{1}{p}} \right) \\ & \quad + C \left\{ \int_0^t (\mathbb{E}[|Z^{i,N}(s)|^{2p}])^{\frac{1}{p}} \, ds + \int_0^{0 \vee (t-\tau)} (\mathbb{E}[|Z^{i,N}(s)|^{2p} (1 + |X^{i,N}(s)|^{2pq_2} + |Y^{i,N}(s)|^{2pq_2})])^{\frac{1}{p}} \, ds + \delta^2 \right\}, \end{aligned} \quad (5.14)$$

for $t \in [0, T]$.

Proof. We will restrict the subsequent discussion to $d = m = 1$ for ease of notation. We may write

$$\begin{aligned} & Z^{i,N}(s)(b(Y^{i,N}(s), \mu_s^{Y,N}) - b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N})) \\ & = Z^{i,N}(s)(b(Y^{i,N}(s), \mu_s^{Y,N}) - b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N})) \end{aligned} \quad (5.15)$$

$$- Z^{i,N}(s) \partial_x b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) \Delta Y^{i,N}(s) \quad (5.16)$$

$$- Z^{i,N}(s) \frac{1}{N} \sum_{j=1}^N \partial_\mu b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N})(Y^{j,N}(s_\delta)) \Delta Y^{j,N}(s) \quad (5.17)$$

$$+ Z^{i,N}(s) \partial_x b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) \Delta Y^{i,N}(s) \quad (5.18)$$

$$+ Z^{i,N}(s) \frac{1}{N} \sum_{j=1}^N \partial_\mu b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N})(Y^{j,N}(s_\delta)) \Delta Y^{j,N}(s). \quad (5.19)$$

The terms (5.15)–(5.17) can be estimated using the techniques employed in the proof of Lemma 5.3. In particular, we have due to Young's inequality

$$\begin{aligned} & \int_0^t \mathbb{E} \left[\left| Z^{i,N}(s) \left(b(Y^{i,N}(s), \mu_s^{Y,N}) - b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) \right. \right. \right. \\ & \quad \left. \left. \left. - \partial_x b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) \Delta Y^{i,N}(s) \right. \right. \right. \\ & \quad \left. \left. \left. - \frac{1}{N} \sum_{j=1}^N \partial_\mu b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N})(Y^{j,N}(s_\delta)) \Delta Y^{j,N}(s) \right) \right|^p \right] ds \\ & \leq \varepsilon \mathbb{E}[\|Z^{i,N}\|_{\infty,t}^{2p}] + C\delta^{2p}. \end{aligned}$$

The remaining two terms (5.18) and (5.19) can be estimated similar to Kumar & Neelima (2021, Lemma 10) in combination with Lemma 5.4. \square

REMARK 5.6 The above proof reveals why we assumed the drift to be independent of the delay variables. In case of explicit delay dependence, we would have to investigate, for $l > 1$, a term of the form

$$Z^{i,N}(s) \partial_{x_l} b(\Pi(Y^{i,N}(s_\delta)), \Pi(\mu_{s_\delta}^{Y,N}))(Y^{i,N}(s + s_l) - Y^{i,N}(s_\delta + s_l)).$$

However, the equation for the difference $Y^{i,N}(s - \tau) - Y^{i,N}(s_\delta - \tau)$ depends on the delayed Brownian increment $W^i(s - \tau) - W^i(s_\delta - \tau)$, which makes $Z^{i,N}(s)$ anticipative. This difficulty does not appear in the case of globally Lipschitz continuous coefficients. The reason is the following: in the super-linear growth setting, one applies Itô's formula to $|Z^{i,N}(s)|^2$ in order to employ the one-sided Lipschitz assumption. The strong convergence analysis for globally Lipschitz continuous coefficients does not require this and therefore the problematic term discussed in this remark does not appear.

Proof of Theorem 2.4:

Proof. We aim to show that

$$\mathbb{E}[\|Z^{i,N}\|_{\infty,T}^p] \lesssim \delta^p,$$

for any $p \geq 2$. By Itô's formula together with $Z^{i,N}(0) = 0$, it follows, for $t \geq 0$, that

$$\begin{aligned} |Z^{i,N}(t)|^2 &= 2 \int_0^t \langle Z^{i,N}(s), b(X^{i,N}(s), \mu_s^{X,N}) - b_\delta(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) \rangle ds + 2 \int_0^t \langle Z^{i,N}(s), dM_s^{i,k} \rangle \\ &\quad + \int_0^t \|\sigma(\Pi(X_s^{i,N}), \Pi(\mu_s^{X,N})) - \gamma_s^{i,k}\|^2 ds, \end{aligned}$$

where $M_t^{i,k}$ and $\Upsilon_t^{i,k}$ for $t \in [0, T]$ are defined in (5.1). The first term from above will be estimated based on the decomposition

$$\int_0^t \langle Z^{i,N}(s), b(X^{i,N}(s), \mu_s^{X,N}) - b_\delta(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) \rangle ds = \Pi_t^{1,i} + \Pi_t^{2,i} + \Pi_t^{3,i} + \Pi_t^{4,i},$$

where we defined

$$\begin{aligned} \Pi_t^{1,i} &:= \int_0^t \langle Z^{i,N}(s), b(X^{i,N}(s), \mu_s^{X,N}) - b(Y^{i,N}(s), \mu_s^{X,N}) \rangle ds, \\ \Pi_t^{2,i} &:= \int_0^t \langle Z^{i,N}(s), b(Y^{i,N}(s), \mu_s^{X,N}) - b(Y^{i,N}(s), \mu_s^{Y,N}) \rangle ds, \\ \Pi_t^{3,i} &:= \int_0^t \langle Z^{i,N}(s), b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) - b_\delta(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) \rangle ds, \\ \Pi_t^{4,i} &:= \int_0^t \langle Z^{i,N}(s), b(Y^{i,N}(s), \mu_s^{Y,N}) - b(Y^{i,N}(s_\delta), \mu_{s_\delta}^{Y,N}) \rangle ds. \end{aligned}$$

Employing (\mathbf{AD}_b^1) , Lemma 5.5 and Lemma 5.1 shows that

$$\begin{aligned} \sum_{l=1}^4 \mathbb{E}[\|\Pi^{l,i}\|_{\infty,t}^{2p}] &\leq \varepsilon (\mathbb{E}[\|Z^{i,N}\|_{\infty,t}^{2p}] + \mathbb{E}[\|Z^{i,N}\|_{\infty,t-\tau}^{2p} (1 + \|X^{i,N}\|_{\infty,t-\tau}^{2pq_2} + \|Y^{i,N}\|_{\infty,t-\tau}^{2pq_2})]) \\ &\quad + C \left\{ \delta^{2p} + \mathbb{E} \int_0^t |Z^{i,N}(s)|^{2p} ds + \mathbb{E} \int_0^{0 \vee (t-\tau)} |Z_s^{i,N}|^{2p} (1 + |X^{i,N}(s)|^{2pq_2} + |Y^{i,N}(s)|^{2pq_2}) ds \right\}. \end{aligned}$$

The remaining parts of Theorem 2.4 can be readily shown using above auxiliary results and the techniques employed in the proof of Bao *et al.* (2021, Theorem 2.1) combined with an inductive argument (using subintervals of length τ), illustrated in Proposition 2.1. \square

6. Proof of Theorem 3.2

6.1 Some auxiliary results

We start by giving some standard results for stability and one-step error of the processes $Y^{i,N,a}$ and $Y^{i,N,f}$.

LEMMA 6.1 Let Assumptions (\mathbf{AAD}_b^1) – (\mathbf{AAD}_b^3) and (\mathbf{AAD}_σ^1) – (\mathbf{AAD}_σ^2) hold. Let $Y^{i,N,a}$ and $Y^{i,N,f}$ be defined as above. Then, for any $p \geq 2$, there exists a constant $C > 0$ such that

$$\begin{aligned} \max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M\}} \mathbb{E}[|Y^{i,N,f}(t_n)|^p] &\leq C, & \max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M-1\}} \mathbb{E}[|Y^{i,N,f}(t_{n+1/2}) - Y^{i,N,f}(t_n)|^p] &\leq C\delta^{p/2}, \\ \max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M\}} \mathbb{E}[|Y^{i,N,a}(t_n)|^p] &\leq C, & \max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M-1\}} \mathbb{E}[|Y^{i,N,a}(t_{n+1/2}) - Y^{i,N,a}(t_n)|^p] &\leq C\delta^{p/2}. \end{aligned}$$

Proof. The moment stability of $Y^{i,Nf}$ and $Y^{i,N,a}$ is a consequence of the main results stated in Section 2. For the one-step errors, we remark that Hölder's inequality yields

$$\begin{aligned} & \left| Y^{i,Nf}(t_{n+1/2}) - Y^{i,Nf}(t_n) \right|^p \\ & \leq C \left\{ \left| b_\delta(Y^{i,Nf}(t_n), \mu_{t_n}^{Y^{i,Nf}}) \delta/2 \right|^p + \left| \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \delta W_n^i \right|^p \right. \\ & \quad + \left| \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \frac{(\delta W_n^i)^2 - \delta/2}{2} \right|^p \\ & \quad \left. + \left| \sigma(Y^{i,Nf}(t_n - \tau), Y^{i,Nf}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \frac{\delta W_n^i \delta W_{n-\tau}^i}{2} \right|^p \right\}. \end{aligned}$$

Then, taking expectations on both sides, employing the growth assumptions on the coefficients along with the moment stability of $Y^{i,Nf}$ and standard estimates for Brownian increments allows to deduce the claim. \square

We proceed by stating another standard lemma, which bounds the difference of $Y^{i,Nf}$ and $Y^{i,N,a}$ over a coarse time-step and will be needed throughout this section:

LEMMA 6.2 Let Assumptions (\mathbf{AAD}_b^1) – (\mathbf{AAD}_b^3) and (\mathbf{AAD}_σ^1) – (\mathbf{AAD}_σ^2) hold. Then, for any $p \geq 2$, there exists a constant $C > 0$ such that

$$\max_{i \in \mathbb{S}_N} \max_{n \in \{0, \dots, M\}} \mathbb{E}[|Y^{i,Nf}(t_n) - Y^{i,N,a}(t_n)|^p] \leq C \delta^{p/2}.$$

Proof. The proof follows analogous steps to Giles (2015, Lemma 4.6) in combination with Lemma 3.1 and is therefore omitted. \square

In a next step, we will represent $Y^{i,Nf}$ and $Y^{i,N,a}$, defined by (3.4) and (3.5), respectively, over a single coarse time-step. This is necessary in order to give an approximation result for $\bar{Y}^{i,Nf}$, see (3.6), on a coarse grid, which consequently will enable us to derive an estimate for the difference between $\bar{Y}^{i,Nf}$ and $Y^{i,N,c}$, see (3.3), over a coarse grid in an L_2 -sense:

LEMMA 6.3 Let Assumptions (\mathbf{AAD}_b^1) – (\mathbf{AAD}_b^3) and (\mathbf{AAD}_σ^1) – (\mathbf{AAD}_σ^2) hold and let $p \geq 2$. Let $x \in \{f, a\}$ and we use the convention $\text{sign}(f) := -$ and $\text{sign}(a) := +$. For all $n \in \{0, \dots, M-1\}$, there exist $N_{i,n,x}$ and $M_{i,n,x}$, where $\mathbb{E}[M_{i,n,x} | \mathcal{F}_{t_n}] = 0$ and for some constant $C > 0$

$$\max_{n \in \{0, \dots, M-1\}} \mathbb{E}[|M_{i,n,x}|^p] \leq C \delta^{3p/2}, \quad \max_{n \in \{0, \dots, M-1\}} \mathbb{E}[|N_{i,n,x}|^p] \leq C \delta^{2p},$$

such that the difference equation for $Y^{i,N,x}$ can be written as

$$\begin{aligned} Y^{i,N,x}(t_{n+1}) = & \mathcal{S} \left(Y^{i,N,x}(t_n), Y^{i,N,x}(t_n - \tau), Y^{i,N,x}(t_n - 2\tau), \mu_{t_n}^{Y,N,x}, \delta, \Delta W_n^i, \Delta W_n^i \Delta W_{n-\tau}^i \right) \\ & \text{sign}(x) \sigma(Y^{i,N,x}(t_n - \tau), Y^{i,N,x}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,x}(t_n), Y^{i,N,x}(t_n - \tau)) \\ & \times \frac{1}{2} \left(\delta W_n^i \delta W_{n+1/2-\tau}^i - \delta W_{n-\tau}^i \delta W_{n+1/2}^i \right) + N_{i,n,x} + M_{i,n,x}. \end{aligned}$$

Proof. We observe that $N_{i,n,f} + M_{i,n,f}$ can be expressed as the sum of the following terms indexed by f (and analogously for $N_{i,n,a} + M_{i,n,a}$):

$$\begin{aligned} R_{i,n,x} &:= (b(Y^{i,N,x}(t_{n+1/2}), \mu_{t_{n+1/2}}^{Y,N,x}) - b(Y^{i,N,x}(t_n), \mu_{t_n}^{Y,N,x})) \frac{\delta}{2} \\ B_{i,n,x} &:= (b(Y^{i,N,x}(t_n), \mu_{t_n}^{Y,N,x}) - b_\delta(Y^{i,N,x}(t_n), \mu_{t_n}^{Y,N,x})) \frac{\delta}{2} \\ &\quad + (b_\delta(Y^{i,N,x}(t_{n+1/2}), \mu_{t_{n+1/2}}^{Y,N,x}) - b(Y^{i,N,x}(t_{n+1/2}), \mu_{t_{n+1/2}}^{Y,N,x})) \frac{\delta}{2} \\ M_{i,n,f}^{(2)} &:= \left(\sigma(Y^{i,N,f}(t_{n+1/2}), Y^{i,N,f}(t_{n+1/2} - \tau)) - \sigma(Y^{i,N,f}(t_n), Y^{i,N,f}(t_n - \tau)) \right. \\ &\quad \left. - \sigma(Y^{i,N,f}(t_n), Y^{i,N,f}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,N,f}(t_n), Y^{i,N,f}(t_n - \tau)) \delta W_n^i \right. \\ &\quad \left. - \sigma(Y^{i,N,f}(t_n - \tau), Y^{i,N,f}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,f}(t_n), Y^{i,N,f}(t_n - \tau)) \delta W_{n-\tau}^i \right) \delta W_{n+1/2}^i \\ M_{i,n,a}^{(2)} &:= \left(\sigma(Y^{i,N,a}(t_{n+1/2}), Y^{i,N,a}(t_{n+1/2} - \tau)) - \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \right. \\ &\quad \left. - \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \delta W_{n+1/2}^i \right. \\ &\quad \left. - \sigma(Y^{i,N,a}(t_n - \tau), Y^{i,N,a}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \delta W_{n+1/2-\tau}^i \right) \delta W_n^i \\ M_{i,n,f}^{(3)} &:= \left(\sigma(Y^{i,N,f}(t_{n+1/2}), Y^{i,N,f}(t_{n+1/2} - \tau)) \partial_{x_1} \sigma(Y^{i,N,f}(t_{n+1/2}), Y^{i,N,f}(t_{n+1/2} - \tau)) \right. \\ &\quad \left. - \sigma(Y^{i,N,f}(t_n), Y^{i,N,f}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,N,f}(t_n), Y^{i,N,f}(t_n - \tau)) \right) \frac{(\delta W_{n+1/2}^i)^2 - \delta/2}{2} \\ M_{i,n,a}^{(3)} &:= \left(\sigma(Y^{i,N,a}(t_{n+1/2}), Y^{i,N,a}(t_{n+1/2} - \tau)) \partial_{x_1} \sigma(Y^{i,N,a}(t_{n+1/2}), Y^{i,N,a}(t_{n+1/2} - \tau)) \right. \\ &\quad \left. - \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \right) \frac{(\delta W_n^i)^2 - \delta/2}{2} \\ M_{i,n,f}^{(4)} &:= \left(\sigma(Y^{i,N,f}(t_{n+1/2} - \tau), Y^{i,N,f}(t_{n+1/2} - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,f}(t_{n+1/2}), Y^{i,N,f}(t_{n+1/2} - \tau)) \right. \\ &\quad \left. - \sigma(Y^{i,N,f}(t_n - \tau), Y^{i,N,f}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,f}(t_n), Y^{i,N,f}(t_n - \tau)) \right) \frac{\delta W_{n+1/2}^i \delta W_{n+1/2-\tau}^i}{2} \\ M_{i,n,a}^{(4)} &:= \left(\sigma(Y^{i,N,a}(t_{n+1/2} - \tau), Y^{i,N,a}(t_{n+1/2} - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,a}(t_{n+1/2}), Y^{i,N,a}(t_{n+1/2} - \tau)) \right. \\ &\quad \left. - \sigma(Y^{i,N,a}(t_n - \tau), Y^{i,N,a}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \right) \frac{\delta W_n^i \delta W_{n-\tau}^i}{2}. \end{aligned}$$

In the following, for simplicity, we will restrict the discussion to $x = f$. We first observe that

$$B_{i,n,f} = \left(\frac{\delta |b(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf})|^{\bar{q}} b(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf})}{1 + \delta |b(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf})|^{\bar{q}}} \right) \frac{\delta}{2} \\ - \frac{\delta |b(Y^{i,Nf}(t_{n+1/2}), \mu_{t_{n+1/2}}^{Y,Nf})|^{\bar{q}} b(Y^{i,Nf}(t_{n+1/2}), \mu_{t_{n+1/2}}^{Y,Nf})}{1 + \delta |b(Y^{i,Nf}(t_{n+1/2}), \mu_{t_{n+1/2}}^{Y,Nf})|^{\bar{q}}} \frac{\delta}{2},$$

which implies $\max_{\{0, \dots, M-1\}} \mathbb{E}[|B_{i,n,f}|^p] \leq C\delta^{2p}$, due to Lemma 6.1 and (\mathbf{AAD}_b^1) .

Next, we aim to analyse the term $R_{i,n,f}$ and assume for ease of notation that $b(x, \mu) = b_1(x) + b_2(\mu)$ for any $x \in \mathbb{R}$, $\mu \in \mathcal{P}_2(\mathbb{R})$. Then, we may write

$$b(Y^{i,Nf}(t_{n+1/2}), \mu_{t_{n+1/2}}^{Y,Nf}) - b(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf}) \\ = b_1(Y^{i,Nf}(t_{n+1/2})) - b_1(Y^{i,Nf}(t_n)) + b_2(\mu_{t_{n+1/2}}^{Y,Nf}) - b_2(\mu_{t_n}^{Y,Nf}),$$

and compute the expansion

$$b_1(Y^{i,Nf}(t_{n+1/2})) - b_1(Y^{i,Nf}(t_n)) = \partial_x b_1(Y^{i,Nf}(t_n))(Y^{i,Nf}(t_{n+1/2}) - Y^{i,Nf}(t_n)) \\ + \frac{1}{2} \partial_x^2 b_1(\xi^{i,N})(Y^{i,Nf}(t_{n+1/2}) - Y^{i,Nf}(t_n))^2,$$

where $\xi^{i,N}$ lies on the line between $Y^{i,Nf}(t_{n+1/2})$ and $Y^{i,Nf}(t_n)$. In case the drift is not decomposable, we have to additionally analyse

$$\partial_x b(Y^{i,Nf}(t_n), \mu_{t_{n+1/2}}^{Y,Nf})(Y^{i,Nf}(t_{n+1/2}) - Y^{i,Nf}(t_n)) \\ = \partial_x b(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf})(Y^{i,Nf}(t_{n+1/2}) - Y^{i,Nf}(t_n)) \\ + (\partial_x b(Y^{i,Nf}(t_n), \mu_{t_{n+1/2}}^{Y,Nf}) - \partial_x b(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf}))(Y^{i,Nf}(t_{n+1/2}) - Y^{i,Nf}(t_n)),$$

which along with (\mathbf{AAD}_b^2) allows to treat this term analogously to $N_{i,n,f}^{(1)}$ precisely introduced below.

We continue by writing

$$b_1(Y^{i,Nf}(t_{n+1/2})) - b_1(Y^{i,Nf}(t_n)) \frac{\delta}{2} = M_{i,n,f}^{(1)} + N_{i,n,f}^{(1)},$$

where we introduced

$$\begin{aligned}
 M_{i,nf}^{(1)} &:= \partial_x b_1(Y^{i,Nf}(t_n)) \sigma(Y^{i,N}(t_n), Y^{i,N}(t_n - \tau)) \delta W_n^i \frac{\delta}{2} \\
 N_{i,nf}^{(1)} &:= \partial_x b_1(Y^{i,Nf}(t_n)) \left(b_\delta(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf}) \frac{\delta}{2} \right. \\
 &\quad + \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \frac{(\delta W_n^i)^2 - \delta/2}{2} \\
 &\quad + \sigma(Y^{i,Nf}(t_n - \tau), Y^{i,Nf}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \frac{\delta W_n^i \delta W_{n-\tau}^i}{2} \left. \right) \frac{\delta}{2} \\
 &\quad + \frac{1}{2} \partial_x^2 b_1(\xi^{i,N})(Y^{i,Nf}(t_{n+1/2}) - Y^{i,Nf}(t_n))^2 \frac{\delta}{2}.
 \end{aligned}$$

Using the notation $\tilde{\mu}_{t_n}^{Y,Nf,\lambda}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^N \delta_{\lambda Y^{j,Nf}(t_{n+1/2}) + (1-\lambda)Y^{j,Nf}(t_n)}(\mathrm{d}x)$, where $\lambda \in [0, 1]$, and

$$\begin{aligned}
 \Delta Y^{i,N}(t_{n+1/2}) &:= Y^{i,Nf}(t_{n+1/2}) - Y^{i,Nf}(t_n) \\
 &= b_\delta(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf}) \frac{\delta}{2} \\
 &\quad + \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \delta W_n^i \\
 &\quad + \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \frac{(\delta W_n^i)^2 - \delta/2}{2} \\
 &\quad + \sigma(Y^{i,Nf}(t_n - \tau), Y^{i,Nf}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \frac{\delta W_n^i \delta W_{n-\tau}^i}{2},
 \end{aligned}$$

we obtain, as $\mathcal{P}_2(\mathbb{R}) \ni \mu \mapsto b(x, \mu)$ is continuously L -differentiable,

$$\begin{aligned}
 &(b_2(\mu_{t_{n+1/2}}^{Y,Nf}) - b_2(\mu_{t_n}^{Y,Nf})) \frac{\delta}{2} \\
 &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\lambda} b_2(\tilde{\mu}_{t_n}^{Y,Nf,\lambda}) \mathrm{d}\lambda \frac{\delta}{2} \\
 &= \int_0^1 \frac{\delta}{2N} \sum_{j=1}^N \partial_\mu b_2(\tilde{\mu}_{t_n}^{Y,Nf,\lambda}) (\lambda Y^{j,Nf}(t_{n+1/2}) + (1-\lambda)Y^{j,Nf}(t_n)) (Y^{j,Nf}(t_{n+1/2}) - Y^{j,Nf}(t_n)) \mathrm{d}\lambda \\
 &= \int_0^1 \frac{\delta}{2N} \sum_{j=1}^N \partial_\mu b_2(\tilde{\mu}_{t_n}^{Y,Nf,\lambda}) (Y^{j,Nf}(t_n) + \lambda \Delta Y^{j,N}(t_{n+1/2})) \Delta Y^{j,N}(t_{n+1/2}) \mathrm{d}\lambda \\
 &= \frac{\delta}{2N} \sum_{j=1}^N \partial_\mu b_2(\mu_{t_n}^{Y,Nf}) (Y^{i,Nf}(t_n)) \Delta Y^{j,N}(t_{n+1/2}) \\
 &\quad + \frac{\delta}{2N} \sum_{j=1}^N \int_0^1 \left(\partial_\mu b_2(\tilde{\mu}_{t_n}^{Y,Nf,\lambda}) (Y^{j,Nf}(t_n) + \lambda \Delta Y^{j,N}(t_{n+1/2})) - \partial_\mu b_2(\mu_{t_n}^{Y,Nf}) (Y^{j,Nf}(t_n)) \right) \\
 &\quad \times \Delta Y^{j,N}(t_{n+1/2}) \mathrm{d}\lambda \\
 &=: R_{i,nf}^{(1)} + R_{i,nf}^{(2)}.
 \end{aligned}$$

Note that $R_{i,nf}^{(1)}$ can be decomposed as

$$R_{i,nf}^{(1)} = R_{i,nf}^{(1,1)} + R_{i,nf}^{(1,2)},$$

where

$$\mathbb{E}[R_{i,nf}^{(1,1)} | \mathcal{F}_{t_n}] = 0, \quad \max_{n \in \{0, \dots, M-1\}} \mathbb{E}[|R_{i,nf}^{(1,1)}|^p] \leq C\delta^{3p/2}, \quad \max_{n \in \{0, \dots, M-1\}} \mathbb{E}[|R_{i,nf}^{(1,2)}|^p] \leq C\delta^{2p}.$$

The terms $N_{i,nx}$ and $M_{i,nx}$ introduced in Lemma 6.3 can be written as

$$\begin{aligned} N_{i,nx} &= N_{i,nx}^{(1)} + B_{i,nx} + R_{i,nx}^{(2)} + R_{i,nx}^{(1,2)} \\ M_{i,nx} &= M_{i,nx}^{(1)} + M_{i,nx}^{(2)} + M_{i,nx}^{(3)} + M_{i,nx}^{(4)} + R_{i,nx}^{(1,1)}. \end{aligned}$$

We continue to show the claimed moment estimates for $R_{i,nf}^{(2)}$ and the martingale terms. Due to assumption (\mathbf{AAD}_b^2) , we obtain

$$\begin{aligned} |R_{i,nf}^{(2)}| &\leq \frac{\delta}{2N} \sum_{j=1}^N \int_0^1 \left| \partial_\mu b_2(\tilde{\mu}_{t_n}^{Y,Nf,\lambda})(Y^{j,Nf}(t_n) + \lambda \Delta Y^{j,N}(t_{n+1/2})) - \partial_\mu b_2(\mu_{t_n}^{Y,Nf})(Y^{j,Nf}(t_n)) \right| \\ &\quad \times |\Delta Y^{j,N}(t_{n+1/2})| d\lambda \\ &\leq \frac{C\delta}{2N} \sum_{j=1}^N \int_0^1 |\mathcal{W}_2(\tilde{\mu}_{t_n}^{Y,Nf,\lambda}, \mu_{t_n}^{Y,Nf}) + \lambda \Delta Y^{j,N}(t_{n+1/2})| |\Delta Y^{j,N}(t_{n+1/2})| d\lambda \\ &\leq \frac{C\delta}{2N} \sum_{j=1}^N \int_0^1 \left| \left(\frac{1}{N} \sum_{l=1}^N |Y^{l,Nf}(t_{n+1/2}) - Y^{l,Nf}(t_n)|^2 \right)^{1/2} + \lambda \Delta Y^{j,N}(t_{n+1/2}) \right| |\Delta Y^{j,N}(t_{n+1/2})| d\lambda. \end{aligned}$$

Using the fact that all particles are identically distributed, Lemma 6.1 gives, for $p \geq 2$,

$$\max_{n \in \{0, \dots, M-1\}} \mathbb{E}[|R_{i,nf}^{(2)}|^p] \leq C\delta^{2p}.$$

We continue with estimating the martingale terms. The above expression for $M_{i,nf}^{(2)}$ can be rewritten as

$$\begin{aligned} M_{i,nf}^{(2)} &= \partial_{x_1} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) Q^{if}(t_n) \delta W_{n+1/2}^i \\ &\quad + \partial_{x_2} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) Q^{if}(t_n - \tau) \delta W_{n+1/2}^i \\ &\quad + \left(\frac{1}{2} \sum_{j,k=1}^2 \partial_{x_j x_k}^2 \sigma(\xi_2^{i,N}, \xi_3^{i,N}) \Delta Y^{if}(t_{n+1/2} - (j-1)\tau) \Delta Y^{if}(t_{n+1/2} - (k-1)\tau) \right) \delta W_{n+1/2}^i, \end{aligned}$$

where

$$\begin{aligned} Q^{if}(t_n) &:= b_\delta(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf}) \frac{\delta}{2} \\ &\quad + \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \frac{(\delta W_n^i)^2 - \delta/2}{2} \\ &\quad + \sigma(Y^{i,Nf}(t_n - \tau), Y^{i,Nf}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \frac{\delta W_n^i \delta W_{n-\tau}^i}{2}, \end{aligned}$$

and $\xi_2^{i,N}, \xi_3^{i,N}$ on the line between $Y^{i,Nf}(t_n)$ and $Y^{i,Nf}(t_{n+1/2})$ and $Y^{i,Nf}(t_n - \tau)$ and $Y^{i,Nf}(t_{n+1/2} - \tau)$, respectively.

Further, considering $M_{i,nf}^{(3)}$ and $M_{i,nf}^{(4)}$, we have using a first-order Taylor series expansion

$$M_{i,nf}^{(3)} = \left(\sum_{j=1}^2 \partial_{x_j} \bar{\sigma}(\xi_4^{i,N}, \xi_5^{i,N}) \Delta Y^{if}(t_{n+1/2} - (j-1)\tau) \right) \frac{(\delta W_{n+1/2}^i)^2 - \delta/2}{2},$$

where

$$\bar{\sigma}(x, y) = \sigma(x, y) \partial_x \sigma(x, y),$$

and $\xi_4^{i,N}, \xi_5^{i,N}$ on the line between $Y^{i,Nf}(t_n)$ and $Y^{i,Nf}(t_{n+1/2})$ and $Y^{i,Nf}(t_n - \tau)$ and $Y^{i,Nf}(t_{n+1/2} - \tau)$, respectively. Additionally, we have

$$M_{i,nf}^{(4)} = \left(\sum_{j=1}^2 \partial_{x_j} \hat{\sigma}(\xi_6^{i,N}, \xi_7^{i,N}) \Delta Y^{if}(t_{n+1/2} - (j-1)\tau) \right) \frac{\delta W_{n+1/2}^i \delta W_{n+1/2-\tau}^i}{2},$$

where

$$\hat{\sigma}(x, y) = \sigma(x, y) \partial_y \sigma(x, y),$$

and $\xi_6^{i,N}, \xi_7^{i,N}$ on the line between $Y^{i,Nf}(t_n - \tau)$ and $Y^{i,Nf}(t_{n+1/2} - \tau)$ and $Y^{i,Nf}(t_n - 2\tau)$ and $Y^{i,Nf}(t_{n+1/2} - 2\tau)$, respectively. From here, the claimed estimates for the moments of $M_{i,nf}^{(2)}$, $M_{i,nf}^{(3)}$ and $M_{i,nf}^{(4)}$ follow. \square

We have the following approximation of the antithetic scheme:

LEMMA 6.4 Let Assumptions $(\mathbf{AAD}_b^1) - (\mathbf{AAD}_b^3)$ and $(\mathbf{AAD}_\sigma^1) - (\mathbf{AAD}_\sigma^2)$ hold and let $p \geq 2$. Then for each $n \in \{0, \dots, M-1\}$, there exist $\tilde{M}_{i,n}$ with $\mathbb{E}[\tilde{M}_{i,n} | \mathcal{F}_n] = 0$ and p th moments of order $\delta^{3p/2}$, and $\tilde{B}_{i,n}$ and $\tilde{N}_{i,n}$ with p th moments of order δ^{2p} , such that the difference equation for $\bar{Y}^{i,Nf}$ reads

$$\begin{aligned} \bar{Y}^{i,Nf}(t_{n+1}) &= \mathcal{S} \left(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau), \bar{Y}^{i,Nf}(t_n - 2\tau), \mu_{t_n}^{\bar{Y},N}, \delta, \Delta W_n^i, \Delta W_n^i \Delta W_{n-\tau}^i \right) \\ &\quad + \frac{1}{2} (N_{i,nf} + N_{i,n,a} + M_{i,nf} + M_{i,n,a}) + \tilde{B}_{i,n} + \tilde{N}_{i,n} + \tilde{M}_{i,n}, \end{aligned}$$

where $N_{i,n,x}$ and $M_{i,n,x}$ are defined as in Lemma 6.3 and we set

$$\mu_{t_n}^{\bar{Y},N}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^N \delta_{\bar{Y}^{i,Nf}(t_n)}(\mathrm{d}x).$$

Proof. In Lemma 6.4, we introduced the abbreviations

$$\begin{aligned} \tilde{N}_{i,n} &:= \left(\frac{1}{2} (b(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf}) + b(Y^{i,N,a}(t_n), \mu_{t_n}^{Y,N,a})) - b(\bar{Y}^{i,Nf}(t_n), \mu_{t_n}^{\bar{Y},N})) \right) \delta \\ \tilde{B}_{i,n} &:= \left(\frac{1}{2} ((b_\delta - b)(Y^{i,Nf}(t_n), \mu_{t_n}^{Y,Nf}) + (b_\delta - b)(Y^{i,N,a}(t_n), \mu_{t_n}^{Y,N,a})) - (b - b_\delta)(\bar{Y}^{i,Nf}(t_n), \mu_{t_n}^{\bar{Y},N})) \right) \delta, \end{aligned}$$

and $\tilde{M}_{i,n}$ can be decomposed in the form

$$\begin{aligned} \tilde{M}_{i,n}^{(1)} &:= \left(\frac{1}{2} \left(\sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) + \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \right) \right. \\ &\quad \left. - \sigma(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau)) \right) \Delta W_n^i \\ \tilde{M}_{i,n}^{(2)} &:= \left(\frac{1}{2} \left(\sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \right. \right. \\ &\quad \left. \left. + \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \right) \right. \\ &\quad \left. - \sigma(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau)) \partial_{x_1} \sigma(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau)) \right) \frac{(\Delta W_n^i)^2 - \delta}{2} \\ \tilde{M}_{i,n}^{(3)} &:= \left(\frac{1}{2} \left(\sigma(Y^{i,Nf}(t_n - \tau), Y^{i,Nf}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \right. \right. \\ &\quad \left. \left. + \sigma(Y^{i,N,a}(t_n - \tau), Y^{i,N,a}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \right) \right. \\ &\quad \left. - \sigma(\bar{Y}^{i,N,a}(t_n - \tau), \bar{Y}^{i,N,a}(t_n - 2\tau)) \partial_{x_2} \sigma(\bar{Y}^{i,N,a}(t_n), \bar{Y}^{i,N,a}(t_n - \tau)) \right) \frac{\Delta W_n^i \Delta W_{n-\tau}^i}{2} \\ \tilde{M}_{i,n}^{(4)} &:= \left(\sigma(Y^{i,N,a}(t_n - \tau), Y^{i,N,a}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,a}(t_n), Y^{i,N,a}(t_n - \tau)) \right. \\ &\quad \left. - \sigma(Y^{i,Nf}(t_n - \tau), Y^{i,Nf}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,Nf}(t_n), Y^{i,Nf}(t_n - \tau)) \right) \\ &\quad \times \frac{1}{2} (\delta W_n^i \delta W_{n+1/2-\tau}^i - \delta W_{n-\tau}^i \delta W_{n+1/2}^i). \end{aligned}$$

In what follows, we analyse the term $\tilde{N}_{i,n}$ assuming for ease of notation the drift term to be decomposable as in the previous Lemma (the assertion for $\tilde{B}_{i,n}$ follows employing similar arguments to the ones used

for the term $B_{i,n,x}$ in Lemma 6.3). We define, for $\lambda \in [0, 1]$ and $n \in \{0, \dots, M\}$

$$m_{t_n}^{f,\lambda}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^N \delta_{(1-\lambda)\bar{Y}^{j,Nf}(t_n) + \lambda Y^{j,Nf}(t_n)}(\mathrm{d}x),$$

$$m_{t_n}^{a,\lambda}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^N \delta_{(1-\lambda)\bar{Y}^{j,Nf}(t_n) + \lambda Y^{j,N,a}(t_n)}(\mathrm{d}x),$$

and observe that for $x \in \{f, a\}$

$$\begin{aligned} & b_2(\mu_{t_n}^{\bar{Y},N} + (\mu_{t_n}^{Y,N,x} - \mu_{t_n}^{\bar{Y},N})) - b_2(\mu_{t_n}^{\bar{Y},N}) \\ &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\lambda} b_2(m_{t_n}^{x,\lambda}) \mathrm{d}\lambda \\ &= \int_0^1 \frac{1}{N} \sum_{j=1}^N \partial_\mu b_2(m_{t_n}^{x,\lambda}) (\lambda Y^{j,N,x}(t_n) + (1-\lambda)\bar{Y}^{j,Nf}(t_n)) (Y^{j,N,x}(t_n) - \bar{Y}^{j,Nf}(t_n)) \mathrm{d}\lambda \\ &= \frac{1}{N} \sum_{j=1}^N \partial_\mu b_2(\mu_{t_n}^{\bar{Y},N}) (\bar{Y}^{j,Nf}(t_n)) (Y^{j,N,x}(t_n) - \bar{Y}^{j,Nf}(t_n)) \\ &\quad + \frac{1}{N} \sum_{j=1}^N \int_0^1 \left(\partial_\mu b_2(m_{t_n}^{x,\lambda}) (\bar{Y}^{j,Nf}(t_n) + \lambda(Y^{j,N,x}(t_n) - \bar{Y}^{j,Nf}(t_n))) \right. \\ &\quad \left. - \partial_\mu b_2(\mu_{t_n}^{\bar{Y},N}) (\bar{Y}^{j,Nf}(t_n)) \right) (Y^{j,N,x}(t_n) - \bar{Y}^{j,Nf}(t_n)) \mathrm{d}\lambda. \end{aligned}$$

Therefore, performing similar estimates as for $R_{i,n,f}^{(2)}$ in Lemma 6.3 and Lemma 6.2 the moments of

$$\left(\frac{1}{2} \left(b_2(\mu_{t_n}^{Y,N,f}) + b_2(\mu_{t_n}^{Y,N,a}) \right) - b_2(\mu_{t_n}^{\bar{Y},N}) \right) \delta,$$

achieve the asserted order. A second-order Taylor series expansion in the state component further gives

$$\begin{aligned} & \left(\frac{1}{2} \left(b_1(Y^{i,Nf}(t_n)) + b_1(Y^{i,N,a}(t_n)) \right) - b_1(\bar{Y}^{i,Nf}(t_n)) \right) \delta \\ &= \frac{1}{16} \left(\partial_x^2 b_1(\xi_8^{i,N}) + \partial_x^2 b_1(\xi_9^{i,N}) \right) \left(Y^{j,Nf}(t_n) - Y^{j,N,a}(t_n) \right)^2 \delta, \end{aligned}$$

where $\xi_8^{i,N}$ and $\xi_9^{i,N}$ are on the line between $\bar{Y}^{i,Nf}(t_n)$ and $Y^{i,Nf}(t_n)$ and $\bar{Y}^{i,Nf}(t_n)$ and $Y^{i,N,a}(t_n)$, respectively. This allows to deduce the claim for $\tilde{N}_{i,n}$ taking Lemma 6.2 into account. Here, we only

remark that $\tilde{M}_{i,n}^{(1)}$ can be treated using a second-order Taylor series expansions about $(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau))$ and $\tilde{M}_{i,n}^{(2)}, \tilde{M}_{i,n}^{(3)}$ by employing first-order Taylor series expansions about $(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau))$ or $(\bar{Y}^{i,Nf}(t_n - \tau), \bar{Y}^{i,Nf}(t_n - 2\tau))$, respectively. To further analyse $\tilde{M}_{i,n}^{(4)}$, one can also employ a Taylor series expansion argument and Lemma 6.2. Note that all of these expressions are martingales and satisfy $\mathbb{E}[|\tilde{M}_{i,n}^{(k)}|^p] = \delta^{3p/2}$, for $k \in \{1, \dots, 4\}$. \square

6.2 Proof of Theorem 3.2

Proof. We analyse now the difference of the antithetic approximation and the coarse-path approximation, i.e., we obtain

$$\begin{aligned} & \mathbb{E} \left[\left| \bar{Y}^{i,Nf}(t_{n+1}) - Y^{i,N,c}(t_{n+1}) \right|^2 \right] = \mathbb{E} \left[\left| \bar{Y}^{i,Nf}(t_n) - Y^{i,N,c}(t_n) \right. \right. \\ & \quad + (b_\delta(\bar{Y}^{i,Nf}(t_n)), \mu_{t_n}^{\bar{Y}^N}) - b_\delta(Y^{i,N,c}(t_n), \mu_{t_n}^{Y,N,c}) \delta \\ & \quad + (\sigma(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau)) - \sigma(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau))) \Delta W_n^i \\ & \quad + \sigma(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau)) \partial_{x_1} \sigma(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau)) \frac{(\Delta W_n^i)^2 - \delta}{2} \\ & \quad - \sigma(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau)) \partial_{x_1} \sigma(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau)) \frac{(\Delta W_n^i)^2 - \delta}{2} \\ & \quad + \sigma(\bar{Y}^{i,Nf}(t_n - \tau), \bar{Y}^{i,Nf}(t_n - 2\tau)) \partial_{x_2} \sigma(\bar{Y}^{i,Nf}(t_n), \bar{Y}^{i,Nf}(t_n - \tau)) \frac{\Delta W_n^i \Delta W_{n-\tau}^i}{2} \\ & \quad - \sigma(Y^{i,N,c}(t_n - \tau), Y^{i,N,c}(t_n - 2\tau)) \partial_{x_2} \sigma(Y^{i,N,c}(t_n), Y^{i,N,c}(t_n - \tau)) \frac{\Delta W_n^i \Delta W_{n-\tau}^i}{2} \\ & \quad \left. + \frac{1}{2} (N_{i,nf} + N_{i,n,a} + M_{i,nf} + M_{i,n,a}) + \tilde{B}_{i,n} + \tilde{N}_{i,n} + \tilde{M}_{i,n} \right|^2 \Big] \\ & \leq (1 + C\delta) \mathbb{E} \left[\left| \bar{Y}^{i,Nf}(t_n) - Y^{i,N,c}(t_n) \right|^2 \right] + C\delta \mathbb{E} \left[\left| \bar{Y}^{i,Nf}(t_n - \tau) - Y^{i,N,c}(t_n - \tau) \right|^2 \right] + C\delta^3, \end{aligned}$$

where the above square was explicitly computed and each term was estimated, using either a martingale property or standard techniques in combination with the auxiliary Lemma 6.2.

Iterating the above recursion yields

$$\mathbb{E} \left[\left| \bar{Y}^{i,Nf}(t_{n+1}) - Y^{i,N,c}(t_{n+1}) \right|^2 \right] \lesssim \delta^2 + \delta \sum_{l=0}^n \mathbb{E} \left[\left| \bar{Y}^{i,Nf}(t_l - \tau) - Y^{i,N,c}(t_l - \tau) \right|^2 \right].$$

Considering above estimate on the interval $[0, \tau]$ allows us to deduce the claim as

$$\sum_{l=0}^n \mathbb{E} \left[\left| \bar{Y}^{i,Nf}(t_l - \tau) - Y^{i,N,c}(t_l - \tau) \right|^2 \right]$$

vanishes on this subinterval. In a next step, we investigate the subinterval $[\tau, 2\tau]$. Here, we can employ the fact that $\mathbb{E}[|\bar{Y}^{i,N,f}(t_l - \tau) - Y^{i,N,c}(t_l - \tau)|^2] \lesssim \delta^2$, and hence the assertion follows in this case as well. An inductive argument can then be used up to the final time T . \square

Funding

Upper Austrian Government fund; National Natural Science Foundation of China (11801406).

REFERENCES

- BAO, J., REISINGER, C., REN, P. & STOCKINGER, W. (2021) First order convergence of Milstein schemes for McKean–Vlasov equations and interacting particle systems. *Proc. R. Soc. A*, **477**, 27.
- BAUER, M., MEYER-BRANDIS, T. & PROSKE, F. (2018) Strong solutions of mean-field stochastic differential equations with irregular drift. *Electron. J. Probab.*, **23**, 1–35.
- BERGOMI, L. (2009) *Smile Dynamics II*. Available at SSRN, <http://dx.doi.org/10.2139/ssrn.1493302>.
- BOSSY, M. & TALAY, D. (1997) A stochastic particle method for the McKean–Vlasov and the Burgers equation. *Math. Comp.*, **66**, 157–193.
- BUCKDAHN, R., LI, J., PENG, S. & RAINER, C. (2017) Mean-field stochastic differential equations and associated PDEs. *Ann. Probab.*, **45**, 824–878.
- BUJOK, K., HAMBLY, B. M. & REISINGER, C. (2015) Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives. *Methodol. Comput. Appl. Probab.*, **17**, 579–604.
- CARDALIAGUET, P. (2013) *Notes on Mean Field Games*. P.-L. Lions Lectures at Collège de France. Online at <https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf>.
- CARMONA, R. & DELARUE, F. (2017) *Probabilistic Theory of Mean Field Games With Applications I*. Probability Theory and Stochastic Modelling, vol. 83, 1st edn. Springer International Publishing.
- CHAUDRUDE RAYNAL, P.-E. & FRIKHA, N. (2018) Well-posedness for some non-linear diffusion processes and related PDE on the Wasserstein space. arXiv:1811.06904.
- DAREIOTIS, K., KUMAR, C. & SABANIS, S. (2016) On tamed Euler approximations of SDEs driven by Lévy noise with application to delay equations. *SIAM J. Numer. Anal.*, **54**, 1840–1872.
- FOUQUE, J.-P. & ZHANG, Z. (2019) Deep learning methods for mean field control problems with delay. *Front. Appl. Math. Stat.*, **6**, 11.
- GAINES, J. & LYONS, T. (1994) Random generation of stochastic area integrals. *SIAM J. Appl. Math.*, **54**, 1132–1146.
- GILES, M. & SZPRUCH, L. (2014) Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.*, **24**, 1585–1620.
- GILES, M. B. (2008) Multilevel Monte Carlo path simulation. *Oper. Res.*, **56**, 607–617.
- GILES, M. B. (2015) Multilevel Monte Carlo methods. *Acta Numer.*, **24**, 259–328.
- GUYON, J. (2014) *Path-Dependent Volatility*. Available at SSRN, <http://dx.doi.org/10.2139/ssrn.2425048>.
- HAJI-ALI, A.-L. & TEMPONE, R. (2018) Multilevel and multi-index Monte Carlo methods for the McKean–Vlasov equation. *Statist. Comput.*, **28**, 923–935.
- HALE, J. K. & LUNEL, V. S. M. (1993) *Introduction to Functional-Differential Equations*. Berlin: Springer.
- HAMMERSLEY, W., SISK, D. & SZPRUCH, L. (2021) McKean–Vlasov SDE under measure dependent Lyapunov conditions. *Ann. Inst. Henri Poincaré Probab. Stat.*, **57**, 1032–1057.
- HU, Y., MOHAMMED, S.-E. A. & YAN, F. (2004) Discrete-time approximations of stochastic delay equations: the Milstein scheme. *Ann. Probab.*, **32**, 265–314.
- HUANG, X., RÖCKNER, M. & WANG, F. Y. (2017) Nonlinear Fokker–Planck equations for probability measures on path space and path-distribution dependent SDEs. *Discrete Contin. Dyn. Syst. B*, **39**, 3017–3035.

- HUTZENTHALER, M., JENTZEN, A. & KLOEDEN, P. E. (2011) Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 467. The Royal Society, pp. 1563–1576.
- KLOEDEN, P. E. & PLATEN, E. (1992) *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer.
- KLOEDEN, P. E. & SHARDLOW, T. (2012) The Milstein scheme for stochastic delay differential equations without using anticipative calculus. *Stochastic Anal. Appl.*, **30**, 181–202.
- KOLMANOVSKII, V. & MYSHKIS, A. (1992) *Applied Theory of Fundamental Differential Equations*. Dordrecht: Kluwer Academic Publishers.
- KÜCHLER, U. & PLATEN, E. (2000) Strong discrete time approximation of stochastic differential equations with time delay. *Math. Comput. Simulation*, **54**, 189–205.
- KUMAR, C. & NEELIMA (2021) On explicit Milstein-type scheme for McKean–Vlasov stochastic differential equations with super-linear drift coefficient. *Electron. J. Probab.*, **26**, 1–32.
- LACKER, D. (2018) On a strong form of propagation of chaos for McKean–Vlasov equations. *Electron. Comm. Probab.*, **23**, 11.
- MALHAM, S. J. A. & WIESE, A. (2014) Efficient almost-exact Lévy area sampling. *Stat. Probab. Lett.*, **88**, 50–55.
- MAO, X. (1997) *Stochastic Differential Equations and Applications*. Cambridge: Horwood Publishers.
- MCKEAN, H. (1966) A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. U S A*, **56**, 1907–1911.
- MÉLÉARD, S. (1996) Asymptotic behaviour of some particle systems: McKean–Vlasov and Boltzmann models. *Probabilistic Models for Nonlinear Partial Differential Equations*. Lecture Notes in Math., vol. 1627, pp. 42–95 (D. Talay & L. Tubaro eds). Berlin: Springer, pp. 42–95.
- MISHURA, Y. S. & VERETENNIKOV, A. Y. (2021) Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations. *Theory Probab. Math. Statist.*, **103**, 59–101.
- NEELIMA, BISWAS, S., KUMAR, C., REIS, G. D. & REISINGER, C. (2021) Well-posedness and tamed Euler schemes for McKean–Vlasov equations driven by Lévy noise. arXiv:2010.08585.
- NUALART, D. (2006) *The Malliavin Calculus and Related Topics*, 2nd edn. Berlin: Springer.
- REIS, G. D., ENGELHARDT, S. & SMITH, G. (2021) Simulation of McKean–Vlasov SDEs with super linear drift. *IMA J. Numer. Anal.*, **42**, 874, 922.
- REIS, G. D., SALKELD, W. & TUGAUT, J. (2019) Freidlin–Wentzell LDPs in path space for McKean–Vlasov equations and the functional iterated logarithm law. *Ann. Appl. Probab.*, **29**, 1487–1540.
- REN, P. & WU, J.-L. (2018) Least squares estimation for path-distribution dependent stochastic differential equations. arXiv:1802.00820.
- SZNITMAN, A. S. (1991) Topics in propagation of chaos. *Ecole d'été de Probabilités de Saint-Flour XIX—1989*. Lecture Notes in Mathematics, vol. 1464. Springer.
- SZPRUCH, L. & TSE, A. (2021) Antithetic multilevel sampling method for nonlinear functionals of measure. *Ann. Appl. Probab.*, **31**, 1100–1139.
- TOUBOUL, J. (2012) Limits and dynamics of stochastic neuronal networks with random heterogeneous delays. *J. Statist. Phys.*, **149**, 569–597.
- WIKTORSSON, M. (2001) Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple Brownian motions. *Ann. Appl. Probab.*, **11**, 470–487.