

# First order convergence of Milstein schemes for McKean–Vlasov equations and interacting particle systems

Jianhai Bao<sup>\*</sup>, Christoph Reisinger<sup>†</sup>, Panpan Ren<sup>‡</sup>, Wolfgang Stockinger<sup>§</sup>

## Abstract

In this paper, we derive fully implementable first order time-stepping schemes for McKean–Vlasov stochastic differential equations (McKean–Vlasov SDEs), allowing for a drift term with super-linear growth in the state component. We propose Milstein schemes for a time-discretised interacting particle system associated with the McKean–Vlasov equation and prove strong convergence of order 1 and moment stability, taming the drift if only a one-sided Lipschitz condition holds. To derive our main results on strong convergence rates, we make use of calculus on the space of probability measures with finite second order moments. In addition, numerical examples are presented which support our theoretical findings.

## 1 Introduction

A McKean–Vlasov equation (introduced by H. McKean [21]) for a  $d$ -dimensional process  $X$  is an 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  $W$  is an  $m$ -dimensional standard Brownian motion,  $\mathcal{L}_{X_t}$  denotes the marginal law of the process  $X$  at time  $t \geq 0$  and  $\xi$  is an  $\mathbb{R}^d$ -valued random variable. We omit an explicit dependence of the coefficients on  $t$  for brevity, but our results easily generalise to this case.

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., [29]). For further existence and uniqueness results on weak and strong solutions of McKean–Vlasov SDEs we refer the reader to [3, 22] and the references cited therein. Also, in the case of super-linear growth it is known that a McKean–Vlasov SDE admits a unique strong solution [25], assuming a so-called one-sided Lipschitz condition for the drift term, see item (1) of assumption  $(\mathbf{A}_b^1)$  in Section 2.1.

McKean–Vlasov SDEs have numerous applications, for instance, in the social and natural sciences. These include fundamental models in neuroscience, such as the Hodgkin-Huxley model (see [2]) for neuron activation, or in biology and chemistry, such as the Patlak-Keller-Segel equations describing, e.g., chemotactic interactions (see [15]) and long-chain polymers (see [23]). The aforementioned examples all have drift terms which do not exhibit the classical global Lipschitz conditions.

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  $\mu_t^{X^{i,N}}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{X_t^{j,N}}(dx)$ , where  $\delta_x$  denotes the Dirac measure at point  $x$  and  $(X^{i,N})_{i=1,\dots,N}$  (so-called interacting particles) is the solution to the  $\mathbb{R}^{dN}$ -dimensional SDE

$$dX_t^{i,N} = b(X_t^{i,N}, \mu_t^{X^{i,N}}) dt + \sigma(X_t^{i,N}, \mu_t^{X^{i,N}}) dW_t^i, \quad X_0^{i,N} = X_0^i.$$

Here,  $W^i$  and  $X_0^i$ ,  $i = 1, \dots, N$  are independent Brownian motions (also independent of  $W$ ) and i. i. d. random initial values with  $\mathcal{L}_{X_0^i} = \mathcal{L}_{X_0}$ , respectively. In the second step, one needs to introduce a reasonable time-stepping method to discretise the particle system  $(X^{i,N})_{i=1,\dots,N}$  over an interval  $[0, T]$ .

<sup>\*</sup>Center for Applied Mathematics, Tianjin University, 300072 Tianjin, China, E-mail: jianhaibao13@gmail.com.

<sup>†</sup>Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Road, Oxford, OX2 6GG, UK, E-mail: christoph.reisinger@maths.ox.ac.uk.

<sup>‡</sup>Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong, China, E-mail: rppzoe@gmail.com.

<sup>§</sup>Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Road, Oxford, OX2 6GG, UK, E-mail: wolfgang.stockinger@maths.ox.ac.uk.

An Euler scheme for the particle system is introduced in [4], and the strong convergence to the solution of the McKean–Vlasov equation, of order  $1/2$  in the time-step and also  $1/2$  in the number of particles, is shown under global Lipschitz assumptions for the coefficients with respect to the state and measure (for  $d = m = 1$ ). The convergence in  $N$  is often referred to as propagation of chaos.

The results in [4] have recently been extended to the case where only a one-sided Lipschitz condition for the drift holds with respect to the state, while a global Lipschitz condition is still assumed for all other dependencies, by using so-called tamed and implicit schemes (in [24]) or adaptive schemes (in [26]).

An approximation of order  $1/2$  in the timestep is also given in [4] for the density and cumulative distribution function, and this is improved to order 1 in [1]. In this paper, in contrast, we are concerned with the strong convergence of the approximations to the process itself.

We complement the work in [4] and [24] by introducing stable first order time-stepping schemes (Milstein schemes) for a particle system associated with McKean–Vlasov SDEs, allowing for a drift coefficient which grows super-linearly in the state component. Here, we will prove moment stability of the time-discretised particle system and strong convergence of order 1.

Our new Milstein scheme reveals that a term involving the Lions derivative (abbreviated by  $L$ -derivative) of the diffusion coefficient with respect to the (empirical) measure is necessary to obtain the strong convergence result. The  $L$ -derivative of functions on  $\mathcal{P}_2(\mathbb{R}^d)$  was introduced by P. -L. Lions in his lectures [5] at Collège de France. This is of theoretical interest and demonstrates the inherent difference of McKean–Vlasov SDEs to classical SDEs with regard to higher order time-stepping schemes.

The main difficulty presented by the super-linearity is to prove the stability of the proposed scheme, and we adopt here the taming approach given in [13] for standard SDEs. In the present context, we require pathwise estimates for each particle, which are challenging to obtain as the particles interact through the appearance of the empirical distribution of the particle system in the coefficients.

The convergence results for the time-stepping scheme hold for a fixed dimension  $N$ , and are robust in  $N$ . To obtain error bounds with respect to the solution of (1.1), these have to be supplemented by propagation of chaos results from [24] to bound the error in  $N$  from the particle approximation. In practice, many McKean–Vlasov equations are motivated by large but finite particle systems, and our time-stepping schemes are directly applicable in that case.

We will demonstrate that the terms involving the measure derivative are only significant for fixed finite  $N$ , but vanish with the same order of  $N$  as the terms from the particle approximation. Omitting these terms to obtain a simplified scheme for large  $N$  is practically useful as they require the simulation of Lévy areas and are the computationally most expensive part of the scheme.

On a side note, in the special case  $N = 1$  our work gives the first order convergence of tamed Milstein schemes for standard SDEs with super-linear drift. It is observed in [14] that already in this setting of classical SDEs (i.e., where the coefficients have no measure-dependence), the explicit Euler–Maruyama scheme (see, e.g., [16]) is not appropriate in the presence of drift terms with super-linear growth. To overcome this problem, several stable time-discretisation methods, including a tamed explicit Euler and Milstein scheme [13, 27, 11], an explicit adaptive Euler–Maruyama method [8], a truncated Euler method [20] and an implicit Euler scheme [12], have been introduced. In [11], a tamed Milstein scheme is introduced and convergence is proven under a commutativity assumption for the diffusion matrix, such that the Lévy area vanishes. Our technique of handling the Lévy area terms in a pathwise sense allows the analysis of the general case without this assumption. While [19] introduces a tamed Milstein scheme for SDEs with super-linearly growing drift and diffusion without assuming a commutativity assumption, the taming there is required to be stronger than the taming approach proposed in [13], which makes the stability analysis of the schemes easier, but can result in inferior numerical performance (see Remark 2.7 and Section 4 for details).

In summary, the main contributions of this paper are the following:

- derivation of a Milstein scheme for particle systems associated with McKean–Vlasov equations using Lions calculus on measure space;
- proof of moment stability and first order uniform (in time) convergence in the time-step using a tamed scheme if only a one-sided Lipschitz condition holds for the drift;
- estimates of the terms involving measure derivatives, showing that these are essential for small particle systems but negligible to approximate the mean-field limit;
- detailed numerical tests supporting the theoretical findings.

- The main results extend those on tamed Milstein schemes (for certain taming approaches) for standard SDEs by eliminating commutativity conditions.

A tamed Milstein scheme, with stronger taming than used in our work, was developed under a similar set of assumptions independently and in parallel in [17]. The authors only state a pointwise (in time) strong  $L_2$  convergence rate of order 1, albeit under slightly weaker differentiability conditions. The present work also goes beyond [17] by analysing the asymptotic behaviour (in terms of the number of particles) of the  $L$ -derivative terms in the schemes and by providing detailed numerical illustrations.

The remainder of this article is organised as follows: In Section 2 we formulate the problem set-up and introduce two different tamed Milstein schemes, in which we use two different taming factors for the drift. The proofs of the main convergence results are deferred to Section 3. Section 4 illustrates the numerical performance of the proposed time-stepping schemes and reveals that the schemes can also successfully be applied to equations which do not satisfy all imposed model assumptions. The electronic appendix to this paper introduces a notion which allows to consider derivatives on the Wasserstein space. We end this section by fixing the notation and by introducing several notions needed throughout this paper.

*Notations.* Let  $(\mathbb{R}^d, \langle \cdot, \cdot \rangle, |\cdot|)$  represent the  $d$ -dimensional Euclidean space and  $\mathbb{R}^d \otimes \mathbb{R}^m$  be the collection of all  $d \times m$ -matrices. The transpose of a matrix  $A$  will be denoted by  $A^*$ . In addition, we use  $\mathcal{P}(\mathbb{R}^d)$  to denote the family of all probability measures on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , where  $\mathcal{B}(\mathbb{R}^d)$  denotes the Borel  $\sigma$ -field over  $\mathbb{R}^d$ , and define the subset of probability measures with finite second moment by

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

For all linear (e.g., matrices), and bilinear operators appearing in this article, we will use the standard Hilbert-Schmidt norm denoted by  $\|\cdot\|$ .

For  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ , the Wasserstein distance between  $\mu$  and  $\nu$  in  $\mathcal{P}_2(\mathbb{R}^d)$  is defined as

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

where  $\mathcal{C}(\mu, \nu)$  is the set of all couplings of  $\mu$  and  $\nu$ , i.e.,  $\pi \in \mathcal{C}(\mu, \nu)$  if and only if  $\pi(\cdot, \mathbb{R}^d) = \mu(\cdot)$  and  $\pi(\mathbb{R}^d, \cdot) = \nu(\cdot)$ . Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  be a filtered probability space satisfying the usual assumptions. For a given  $p \geq 1$ ,  $L_p^0(\mathbb{R}^d)$  will denote the space of  $\mathbb{R}^d$ -valued,  $\mathcal{F}_0$ -measurable random variables  $X$  satisfying  $\mathbb{E}|X|^p < \infty$ . Further,  $\mathcal{S}^p([0, T])$  refers to the space of  $\mathbb{R}^d$ -valued continuous,  $\mathcal{F}$ -adapted processes, defined on the interval  $[0, T]$ , for  $T > 0$ , with finite  $p$ -th moments (uniform in time).

## 2 Tamed Milstein schemes for non-Lipschitz McKean–Vlasov SDEs

In this section, we define the model set-up with its assumptions (subsection 2.1), and define the Milstein schemes (subsection 2.2). We focus in the analysis on the schemes with tamed drift coefficients, where the super-linear drifts are approximated by functions which are bounded depending on the mesh size (similar to the tamed Euler schemes in [24]). A simplified version of the proofs gives the corresponding results for the Milstein scheme without drift approximation in the global Lipschitz case, and we only state the assumptions required in subsection 2.3.

### 2.1 Assumptions and interacting particle system

For a given  $T > 0$ , we consider the following  $\mathbb{R}^d$ -valued McKean–Vlasov SDE on  $[0, T]$ ,

$$dX_t = b(X_t, \mathcal{L}_{X_t}) dt + \sigma(X_t, \mathcal{L}_{X_t}) dW_t, \quad X_0 = \xi, \quad (2.1)$$

where we recall that  $\mathcal{L}_{X_t}$  denotes the marginal law of  $X$  at the time  $t \in [0, T]$ ,  $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ ,  $\sigma : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d \otimes \mathbb{R}^m$  are measurable functions,  $\xi \in L_p^0(\mathbb{R}^d)$ , for all  $p \geq 1$ , and  $(W_t)_{t \geq 0}$  is an  $m$ -dimensional Brownian motion on the filtered, atomless probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ , where  $(\mathcal{F}_t)_{t \geq 0}$  is the natural filtration of  $(W_t)_{t \geq 0}$  augmented with an independent  $\sigma$ -algebra  $\mathcal{F}_0$ .

We assume, for any  $x, y \in \mathbb{R}^d$  and  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$  the following:

( $\mathbf{A}_b^1$ ) There exist constants  $L_b^1, \alpha_1 > 0$  such that

$$\langle x - y, b(x, \mu) - b(y, \mu) \rangle \leq L_b^1 |x - y|^2, \quad (1)$$

$$|b(x, \mu) - b(y, \mu)| \leq L_b^1 (1 + |x|^{\alpha_1} + |y|^{\alpha_1}) |x - y|, \quad (2)$$

$$|b(x, \mu) - b(x, \nu)| \leq L_b^1 \mathbb{W}_2(\mu, \nu). \quad (3)$$

( $\mathbf{A}_\sigma^1$ ) There exists a constant  $L_\sigma^1 > 0$  such that

$$\|\sigma(x, \mu) - \sigma(y, \nu)\| \leq L_\sigma^1 (|x - y| + \mathbb{W}_2(\mu, \nu)).$$

The analysis presented below can be readily extended to the case where  $b$  and  $\sigma$  depend explicitly on  $t$  in a Lipschitz continuous way.

We recall that under these assumptions [25, Theorem 3.3] guarantees that (2.1) has a unique strong solution with bounded  $p$ -th moments, i.e., we have  $\mathbb{E}\|X\|_{\infty, T}^p < \infty$ , where  $\|X\|_{\infty, t}^p := \sup_{0 \leq s \leq t} |X_s|^p$ , for  $t \geq 0$  and  $p \geq 1$ .

For each  $i \in \mathbb{S}_N := \{1, \dots, N\}$ , let  $(W^i, X_0^i)$  be independent copies of  $(W, X_0)$ . Consider first the following *non-interacting* particle system associated with (2.1),

$$dX_t^i = b(X_t^i, \mathcal{L}_{X_t^i}) dt + \sigma(X_t^i, \mathcal{L}_{X_t^i}) dW_t^i, \quad \mathcal{L}_{X_0^i} = \mathcal{L}_{X_0}, \quad i \in \mathbb{S}_N. \quad (2.2)$$

One obviously has  $\mathcal{L}_{X_t} = \mathcal{L}_{X_t^i}$ ,  $i \in \mathbb{S}_N$ . Compared to the simulation of classical SDEs, the key difference is the need to approximate the measure  $\mathcal{L}_{X_t}$ , for each  $t \geq 0$ . To do so, we introduce the following *interacting* particle system (see e.g., [4])

$$dX_t^{i,N} = b(X_t^{i,N}, \mu_t^{X^{\cdot,N}}) dt + \sigma(X_t^{i,N}, \mu_t^{X^{\cdot,N}}) dW_t^i, \quad X_0^{i,N} = X_0^i, \quad i \in \mathbb{S}_N, \quad (2.3)$$

where  $\mu_t^{X^{\cdot,N}}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{X_t^{j,N}}(dx)$ .

It is easy to deduce from ( $\mathbf{A}_b^1$ ) and ( $\mathbf{A}_\sigma^1$ ) and the estimate

$$\mathbb{W}_2(\hat{\mu}^{\mathbf{x},N}, \hat{\mu}^{\mathbf{y},N})^2 \leq \frac{1}{N} |\mathbf{x} - \mathbf{y}|^2, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^{dN}, \quad (2.4)$$

where  $\hat{\mu}^{\mathbf{x},N}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{x_j}(dx)$  for  $\mathbf{x} := (x_1, x_2, \dots, x_N)$  that the stochastic interacting particle system (2.3) is well-posed.

## 2.2 Time-stepping schemes and main results

Since the drift term  $b$  is not assumed to be Lipschitz with respect to the spatial argument, the standard Euler scheme is, in general, not suitable for (2.3), due to the potential moment-unboundedness of the time-discretised interacting particle system [14]. Here, we propose two novel stable time-stepping schemes which achieve a first order strong convergence rate.

We partition a given time interval  $[0, T]$  into  $M \in \mathbb{N}$  steps of equal length  $\delta := T/M$ . In the sequel, we set  $t_n := n\delta$  and for any  $t \in [0, T]$ , we define  $t_\delta := \max\{t_n | t_n \leq t\} = \lfloor t/\delta \rfloor \delta$ . Now, for  $n \in \{0, \dots, M-1\}$ , we introduce the following tamed Milstein schemes: For each  $i \in \mathbb{S}_N$ ,  $Y_{t_{n+1}}^{i,N}$  is computed by

$$\begin{aligned} Y_{t_{n+1}}^{i,N} &= Y_{t_n}^{i,N} + b_\delta(Y_{t_n}^{i,N}, \mu_{t_n}^{Y^{\cdot,N}}) \delta + \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y^{\cdot,N}}) \Delta W_n^i \\ &\quad + \int_{t_n}^{t_{n+1}} \nabla \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y^{\cdot,N}}) \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y^{\cdot,N}}) \int_{t_n}^s dW_u^i dW_s^i \\ &\quad + \int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^N D^L \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y^{\cdot,N}}) (Y_{t_n}^{j,N}) \sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{Y^{\cdot,N}}) \int_{t_n}^s dW_u^j dW_s^i, \end{aligned} \quad (2.5)$$

where the driving Brownian motions  $W^i$  and initial values  $Y_0^{i,N} = X_0^{i,N}$  are assumed to be independent and  $\Delta W_n^i := W_{t_{n+1}}^i - W_{t_n}^i$ . In addition, we used the notation  $\mu_{t_n}^{Y^{\cdot,N}}(dx) := \frac{1}{N} \sum_{j=1}^N \delta_{Y_{t_n}^{j,N}}(dx)$ . Above,  $\nabla \sigma$  denotes the first order gradient operator (applied to each column of  $\sigma$ ) with respect to the state variable of  $\sigma$  and  $D^L \sigma$  is the Lions derivative operator (see electronic appendix for details). Note that

for  $x, y \in \mathbb{R}^d$ ,  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\nabla \sigma(x, \mu)$  and  $D^L \sigma(x, \mu)(y)$  are tensors and can be viewed as linear operators from  $\mathbb{R}^d$  to  $\mathbb{R}^d \otimes \mathbb{R}^m$ .

We define  $b_\delta$  in two different ways yielding two schemes, which will subsequently be denoted by Scheme 1 and Scheme 2, respectively: For Scheme 1, we use

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

and for Scheme 2, we define

$$b_\delta(x, \mu) := \frac{b(x, \mu)}{1 + \delta |b(x, \mu)|^2}, \quad x \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

Note that the following bounds hold for the two different choices of  $b_\delta$ :

$$\frac{|b(x, \mu)|}{1 + \delta |b(x, \mu)|} \leq |b(x, \mu)| \wedge \frac{M}{T}, \quad \frac{|b(x, \mu)|}{1 + \delta |b(x, \mu)|^2} \leq |b(x, \mu)| \wedge \sqrt{\frac{M}{T}}. \quad (2.6)$$

**Remark 2.1.** If the drift  $b$  is also globally Lipschitz in the state component, taming is not necessary and we can instead introduce a standard Milstein scheme by simply replacing  $b_\delta$  in (2.5) with  $b$ ; see subsection 2.3 for details of this case.

The continuous time version of (2.5) reads

$$\begin{aligned} dY_t^{i,N} &= b_\delta(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}}) dt + \left( \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}}) + \int_{t_\delta}^t (\nabla \sigma) \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^i \right. \\ &\quad \left. + \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t D^L \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y_{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^j \right) dW_t^i, \quad Y_0^{i,N} = X_0^{i,N}. \end{aligned} \quad (2.7)$$

**Remark 2.2.** Compared to the classical Milstein scheme for standard SDEs without measure dependence, a term involving the  $L$ -derivative appears. Although this term is crucial for the theoretical analysis of the scheme for fixed  $N$ , it can be dropped in practice for large  $N$  and when an approximation to the limiting McKean–Vlasov equation is sought. We will provide a theoretical justification in the Lipschitz case at the end of subsection 2.3, and give a numerical illustration in Section 4.

In what follows, we show that the fully discretised particle system converges, in a strong sense, to a solution of the limit McKean–Vlasov SDE if  $N \rightarrow \infty$  and  $\delta \rightarrow 0$  and establish the order of convergence. To establish the following main result on moment stability and strong convergence of the above proposed time-stepping schemes, we need further assumptions on the coefficients  $b$  and  $\sigma$ . We refer the reader to the online appendix of this paper for the precise definitions of the function spaces used in the assumptions listed below, in particular the class  $C^{2,(2,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ .

For any  $x, x', y, y' \in \mathbb{R}^d$  and  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ , we require the following:

(A<sub>b</sub><sup>2</sup>) Set  $b = (b_1, \dots, b_d)^*$ , let  $b_i \in C^{2,(2,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$  and assume that there exist constants  $L_b^2, \alpha_2 > 0$  such that for all  $i \in \{1, \dots, d\}$

$$\begin{aligned} \|\nabla^2 b_i(x, \mu)\| \vee \|\nabla \{D^L b_i(x, \mu)(\cdot)\}(y)\| \vee \|(D^L)^2 b_i(x, \mu)(y, y)\| \\ \leq L_b^2 (1 + |x|^{1+\alpha_2} + |y|^{1+\alpha_2} + \mu(|\cdot|^2)^{\frac{1+\alpha_2}{2}}), \end{aligned}$$

where  $\nabla^2$  is the second order gradient operator with respect to the first argument and  $(D^L)^2$  the second order  $L$ -derivative operator.

(A<sub>b</sub><sup>3</sup>) There exist constants  $L_b^3, \alpha_3 > 0$  such that

$$\begin{aligned} \|\nabla b(x, \mu) - \nabla b(y, \nu)\| &\leq L_b^3 (|x - y| + \mathbb{W}_2(\mu, \nu)) \\ &\quad \times (1 + |x|^{\alpha_3} + |y|^{\alpha_3} + \mu(|\cdot|^2)^{\frac{\alpha_3}{2}} + \nu(|\cdot|^2)^{\frac{\alpha_3}{2}}), \\ \|D^L b(x, \mu)(y) - D^L b(x', \nu)(y')\| &\leq L_b^3 (|x - x'| + |y - y'| + \mathbb{W}_2(\mu, \nu)) \\ &\quad \times (1 + |x|^{\alpha_3} + |y|^{\alpha_3} + |x'|^{\alpha_3} + |y'|^{\alpha_3} + \mu(|\cdot|^2)^{\frac{\alpha_3}{2}} + \nu(|\cdot|^2)^{\frac{\alpha_3}{2}}). \end{aligned}$$

( $\mathbf{A}_b^4$ ) There exists a constant  $L_b^4 > 0$  such that  $|b(0, \mu)| \leq L_b^4$ .

Concerning the diffusion coefficient  $\sigma$ , we further impose, for all  $x, x', y, y' \in \mathbb{R}^d$  and  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ :

( $\mathbf{A}_\sigma^2$ ) Let  $\sigma_{ij} \in C^{2,(2,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ , where  $\sigma_{ij}$  denotes the  $(i, j)$ -th component of  $\sigma$ , and assume that there exists a constant  $L_\sigma^2 > 0$  such that for all  $i \in \{1, \dots, m\}$ ,  $j \in \{1, \dots, d\}$

$$\|D^L \sigma_{ij}(x, \mu)(y)\| \vee \|\nabla\{D^L \sigma_{ij}(x, \mu)(\cdot)\}(y)\| \vee \|(D^L)^2 \sigma_{ij}(x, \mu)(y, y)\| \leq L_\sigma^2.$$

( $\mathbf{A}_\sigma^3$ ) There exists a constant  $L_\sigma^3 > 0$  such that

$$\|\nabla \sigma(x, \mu) - \nabla \sigma(y, \nu)\| \leq L_\sigma^3(|x - y| + \mathbb{W}_2(\mu, \nu)), \quad (1)$$

$$\|D^L \sigma(x, \mu)(y) - D^L \sigma(x', \nu)(y')\| \leq L_\sigma^3(|x - x'| + |y - y'| + \mathbb{W}_2(\mu, \nu)). \quad (2)$$

( $\mathbf{A}_\sigma^4$ ) There exists a constant  $L_\sigma^4 > 0$  such that  $\|\sigma(0, \mu)\| + \|D^L \sigma(x, \mu)(y)\sigma(y, \mu)\| \leq L_\sigma^4$ .

**Remark 2.3.** The first inequality in ( $\mathbf{A}_b^1$ ) is the so called one-sided Lipschitz condition which is needed to control the polynomial growth of the drift in the state variable (uniformly with respect to the measure variable). The third inequality in ( $\mathbf{A}_b^1$ ) and assumption ( $\mathbf{A}_\sigma^1$ ) express that both coefficients are globally Lipschitz continuous in the measure component (uniformly in the state variable). In ( $\mathbf{A}_b^2$ )-( $\mathbf{A}_b^3$ ) and ( $\mathbf{A}_\sigma^2$ )-( $\mathbf{A}_\sigma^3$ ), we require growth and Lipschitz conditions on the derivatives of  $b$  and  $\sigma$ , which are necessary for the strong convergence analysis. Further, we assume uniform boundedness of the coefficients in the measure component in ( $\mathbf{A}_b^4$ ) and ( $\mathbf{A}_\sigma^4$ ), which is essential to achieve moment boundedness of Scheme 1.

Now we are in a position to present our first main result (concerned with Scheme 1) and we remark that in Section 3 we give more details on the generic constants used in the statement of the results.

**Lemma 2.4.** Let  $Y_{t_n}^{i,N}$ ,  $n \in \{0, \dots, M\}$ , be defined as in (2.5) with  $b_\delta(x, \mu) = \frac{b(x, \mu)}{1 + \delta|b(x, \mu)|}$  and  $p \geq 1$ . Then, under assumptions ( $\mathbf{A}_b^1$ ), ( $\mathbf{A}_b^4$ ), ( $\mathbf{A}_\sigma^1$ ), ( $\mathbf{A}_\sigma^2$ ) and ( $\mathbf{A}_\sigma^4$ ), there exists a constant  $C > 0$  independent of  $M$  ( $\delta$ , respectively) such that

$$\sup_{i \in \mathbb{S}_N} \sup_{n \in \{0, \dots, M\}} \mathbb{E}|Y_{t_n}^{i,N}|^p \leq C.$$

*Proof.* The proof is deferred to Section 3, subsection 3.1.  $\square$

**Theorem 2.5.** Let  $p \geq 1$ . Assume ( $\mathbf{A}_b^1$ )-( $\mathbf{A}_b^4$ ), and ( $\mathbf{A}_\sigma^1$ )-( $\mathbf{A}_\sigma^4$ ). Let  $(Y_t^{i,N})_{t \in [0, T]}$  be defined by (2.7) with  $b_\delta(x, \mu) = \frac{b(x, \mu)}{1 + \delta|b(x, \mu)|}$ . Then there exists a constant  $C > 0$  independent of  $M$  ( $\delta$ , respectively) and  $N$  such that

$$\mathbb{E}\|X^{i,N} - Y^{i,N}\|_{\infty, T}^p \leq C\delta^p, \quad i \in \mathbb{S}_N. \quad (2.8)$$

*Proof.* The proof is deferred to Section 3.  $\square$

The following corollary is an immediate consequence of the above theorem and the pathwise propagation of chaos result in [24, Proposition 3.1].

**Corollary 2.6.** Let the assumptions of Theorem 2.5 hold for  $p = 2$ . Then there exists a constant  $C > 0$  independent of  $M$  ( $\delta$ , respectively) and  $N$  such that

$$\mathbb{E}\|X^i - Y^{i,N}\|_{\infty, T}^2 \leq C(\delta^2 + \phi(N)), \quad i \in \mathbb{S}_N,$$

where

$$\phi(N) = \begin{cases} N^{-1/2}, & \text{for } d < 4, \\ N^{-1/2} \log N, & \text{for } d = 4, \\ N^{-2/d}, & \text{for } d > 4. \end{cases} \quad (2.9)$$

**Remark 2.7.** Let  $(Y_t^{i,N})_{t \in [0,T]}$  be defined by Scheme 2, i.e., (2.7) with  $b_\delta(x, \mu) = b(x, \mu)/(1 + \delta|b(x, \mu)|^2)$ . Assume, for some  $p \geq 1$ ,  $(\mathbf{A}_b^1) - (\mathbf{A}_b^3)$ ,  $(\mathbf{A}_\sigma^1) - (\mathbf{A}_\sigma^3)$  and  $X_0 \in L_{4p(1+\alpha)}^0(\mathbb{R}^d)$ , where  $\alpha = \alpha_1 \vee \alpha_2 \vee \alpha_3$ . Then the statement of Theorem 2.5 still holds.

Due to the choice of taming, we have the stronger bound in (2.6) and the moment boundedness of Scheme 2 follows immediately from [18, Lemma 4.3] (see also Remark 2.7 below). The proof of the strong convergence rate is then analogous to Scheme 1 and is therefore omitted.

The crucial difference between these two results is that for Scheme 2 we do not need to require  $(\mathbf{A}_b^4)$  and  $(\mathbf{A}_\sigma^4)$ , i.e., that the coefficients are uniformly bounded in the measure component. The reason is that the taming for Scheme 2 is stronger than for Scheme 1. The numerical tests in Section 4 show, however, that Scheme 1 is significantly more accurate in all cases studied, even when  $(\mathbf{A}_b^4)$  and  $(\mathbf{A}_\sigma^4)$  are violated.

**Remark 2.8.** To prove moment boundedness for Scheme 1, a pathwise estimate of the time-discretised particle system is required. In order to obtain such an estimate the mean-field terms need to be controlled. Here, our main contribution is to show how the Lévy area terms in above schemes can be handled in a pathwise sense (see Lemma 2.4). In [11], a commutativity assumption for the diffusion matrix is imposed, such that the Lévy areas vanish. The techniques used to prove our result can also be employed to relax the assumptions imposed in [11].

### 2.3 Milstein scheme for globally Lipschitz drift

For the sake of completeness, we give a set of model assumptions which allow to derive an analogous convergence result for a standard Milstein scheme, i.e., without taming the drift. We assume  $(\mathbf{A}_\sigma^1) - (\mathbf{A}_\sigma^3)$  for the diffusion coefficient, and for any  $x, x', y, y' \in \mathbb{R}^d$  and  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$  we impose:

$(\mathbf{AA}_b^1)$  There exists a constant  $L_b^1 > 0$  such that

$$\|b(x, \mu) - b(y, \nu)\| \leq L_b^1(|x - y| + \mathbb{W}_2(\mu, \nu)).$$

$(\mathbf{AA}_b^2)$  Set  $b = (b_1, \dots, b_d)^*$ , let  $b_i \in C^{2,(2,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$  and assume that there exists a constant  $L_b^2 > 0$  such that for all  $i \in \{1, \dots, d\}$

$$\|\nabla\{D^L b_i(x, \mu)(\cdot)\}(y)\| \vee \|(D^L)^2 b_i(x, \mu)(y, y)\| \leq L_b^2.$$

$(\mathbf{AA}_b^3)$  There exists a constant  $L_b^3 > 0$  such that

$$\|\nabla b(x, \mu) - \nabla b(y, \nu)\| \leq L_b^3(|x - y| + \mathbb{W}_2(\mu, \nu)), \quad (1)$$

$$\|D^L b(x, \mu)(y) - D^L b(x', \nu)(y')\| \leq L_b^3(|x - x'| + |y - y'| + \mathbb{W}_2(\mu, \nu)). \quad (2)$$

Under the assumptions listed above, the standard Milstein scheme defined as in (2.5), with  $b_\delta$  replaced by  $b$ , is stable (i.e., has bounded moments) and converges with strong order 1. As  $b$  and  $\sigma$  have linear growth in both components and the particles are identically distributed, the claim concerning the stability of the scheme follows by a standard Gronwall type argument. The proof for the strong convergence order is a simplified version of the proof of Theorem 2.5 and is therefore omitted.

We end by estimating the term involving the  $L$ -derivative in (2.5),

$$\int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t D^L \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y_{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^j dW_t^i, \quad (2.10)$$

which is expected to be close to zero for large  $N$ , as explained by the following heuristic observation.

Instead of discretising the particle system, we can directly discretise (2.1) in time using a Milstein scheme without employing the particle approximation for the measure. Itô's formula applied to a function from  $C^{2,(1,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ , in our case to the components  $\sigma_{ij}$  of the diffusion coefficient,  $i \in \{1, \dots, m\}$ ,  $j \in \{1, \dots, d\}$ , gives

$$d\sigma_{ij}(X_s, \mathcal{L}_{X_s}) = \left[ \frac{1}{2} \text{tr}(\sigma \sigma^* \nabla^2 \sigma_{ij})(X_s, \mathcal{L}_{X_s}) + \langle b, \partial_x \sigma_{ij} \rangle(X_s, \mathcal{L}_{X_s}) \right]$$

$$\begin{aligned}
& + \int_{\mathbb{R}^d} \left[ \frac{1}{2} \text{tr} \left\{ \sigma \sigma^*(y, \mathcal{L}_{X_s}) \nabla \{ D^L \sigma_{ij}(X_s, \mathcal{L}_{X_s})(\cdot) \}(y) \right\} \right. \\
& \left. + \langle b(y, \mathcal{L}_{X_s}), D^L \sigma_{ij}(X_s, \mathcal{L}_{X_s})(y) \rangle \right] \mathcal{L}_{X_s}(dy) \Big] ds \\
& + \langle (\sigma^* \nabla \sigma_{ij})(X_s, \mathcal{L}_{X_s}), dW_s \rangle,
\end{aligned}$$

where  $\text{tr}$  denotes the standard trace operator (see [6, Proposition 5.102]). Using this expansion, we can introduce a Milstein scheme for (2.1)

$$Y_{t_{n+1}} = Y_{t_n} + b(Y_{t_n}, \mathcal{L}_{Y_{t_n}}) \delta + \sigma(Y_{t_n}, \mathcal{L}_{Y_{t_n}}) \Delta W_n + \int_{t_n}^{t_{n+1}} \nabla \sigma(Y_{t_n}, \mathcal{L}_{Y_{t_n}}) \sigma(Y_{t_n}, \mathcal{L}_{Y_{t_n}}) \int_{t_n}^s dW_u dW_s, \quad (2.11)$$

for  $n \in \{0, \dots, M-1\}$ , with  $Y_0 = X_0$ . Therefore, it is reasonable to expect that the term (2.10) vanishes as  $N \rightarrow \infty$ , since (2.11) and (2.5) (with  $b_\delta$  replaced by  $b$ , as we restrict this discussion to a global Lipschitz setting) should coincide for  $N \rightarrow \infty$ , i.e., a propagation of chaos result for the time-discrete system.

More precisely, we give the following proposition. We only prove this statement for globally Lipschitz coefficients, but expect a similar result to hold in the general setting of this paper. To prove the claim we additionally require the following for any  $x, x', y, y' \in \mathbb{R}^d$  and  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ :

(**AA** $^1_\sigma$ ) Let  $\sigma$  be continuously differentiable in both components and assume that there exists a constant  $L^1_\sigma > 0$  such that

$$\|(\nabla \sigma)\sigma(x, \mu) - (\nabla \sigma)\sigma(x', \nu)\| \leq L^1_\sigma(|x - x'| + \mathbb{W}_2(\mu, \nu)), \quad (1)$$

$$\|D^L \sigma(x, \mu)(y)\sigma(y, \mu) - D^L \sigma(x', \nu)(y')\sigma(y', \nu)\| \leq L^1_\sigma(|x - x'| + |y - y'| + \mathbb{W}_2(\mu, \nu)). \quad (2)$$

**Proposition 2.9.** Assume (**AA** $^1_b$ ), (**A** $^1_\sigma$ ), (**AA** $^1_\sigma$ ) and  $X_0 \in L^0_2(\mathbb{R}^d)$ . Further, for  $i \in \mathbb{S}_N$  and  $n \in \{0, \dots, M\}$ , let  $Y_{t_n}^i$  be given by (2.11), with independent copies  $(W^i, X_0^i)$  of  $(W, X_0)$ , and  $Y_{t_n}^{i,N}$  be defined by (2.5), with  $b_\delta$  replaced by  $b$ . Then, there exists a constant  $C > 0$  such that

$$\sup_{i \in \mathbb{S}_N} \sup_{n \in \{0, \dots, M\}} \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 \leq C\phi(N).$$

*Proof.* The proof is provided in Section 3.4. □

### 3 Proofs of results

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  $a, b \in \mathbb{R}$ . The implied constant  $C > 0$  may depend on the parameters  $p, \varepsilon, T, m, d$ , the constants appearing in above assumptions and the moments of the initial data, but is independent of  $M$  (and  $\delta$ , respectively) and  $N$ . Also implied constants may change their values from line to line in a sequence of inequalities. To make the presentation clearer, we split the proof of Theorem 2.5 into several auxiliary lemmata.

For the sake of readability, we also set, for any  $i \in \mathbb{S}_N$ ,

$$\begin{aligned}
\Upsilon_t^i &:= \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}}) + \int_{t_\delta}^t \nabla \sigma(\cdot, \mu_{r_\delta}^{Y_{\cdot,N}})(Y_{r_\delta}^{i,N}) \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^i \\
&\quad + \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t D^L \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y_{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^j,
\end{aligned} \quad (3.1)$$

$$\Gamma_t^i := \sigma(Y_t^{i,N}, \mu_t^{Y_{\cdot,N}}) - \Upsilon_t^i, \quad M_t^i := \int_0^t \{ \sigma(X_s^{i,N}, \mu_s^{X_{\cdot,N}}) - \Upsilon_s^i \} dW_s^i, \quad Z^{i,N} := X^{i,N} - Y^{i,N}.$$

#### 3.1 Proof of Lemma 2.4

*Proof.* We utilise ideas developed in [13] for tamed schemes for standard SDEs. We will point these out and focus on the key differences to [13].



For  $i \in \mathbb{S}_N$ , we set

$$A_n^i := \sum_{r,q=1}^m \left| \int_{t_n}^{t_{n+1}} \Delta W_s^{i,r} dW_s^{i,q} \right|, \quad A_n^{i,N} := \frac{1}{N} \sum_{j=1}^N \sum_{r,q=1}^m \left| \int_{t_n}^{t_{n+1}} \Delta W_s^{j,r} dW_s^{i,q} \right|,$$

where  $W_s^{i,r}$  denotes the  $p$ -th component of the Brownian motion  $W^i$  and  $\Delta W_s^i = W_s^i - W_{t_n}^i$ . Define

$$\Omega_n^i := \left\{ \omega \in \Omega \mid \sup_{k \in \{0, \dots, n-1\}} A_k^i(\omega) \leq \frac{1}{\sqrt{M}}, \sup_{k \in \{0, \dots, n-1\}} A_k^{i,N}(\omega) \leq \frac{1}{\sqrt{M}} \right\},$$

$$\Omega_{n,\lambda}^i := \left\{ \omega \in \Omega \mid \sup_{k \in \{0, \dots, n-1\}} (D_{k,\lambda}^i(\omega) + D_{k,\lambda}(\omega)) \leq M^{1/\alpha_1 \wedge 1}, \sup_{k \in \{0, \dots, n-1\}} |\Delta W_k^i(\omega)| \leq 1 \right\},$$

the set of events such that the processes  $D_{n,\lambda}^i : \Omega \rightarrow [0, \infty)$ , given by

$$D_{n,\lambda}^i := \left( \lambda^2 + |Y_0^{i,N}|^2 \right) \exp \left[ \lambda + \sup_{u \in \{0, \dots, n\}} \sum_{k=u}^{n-1} \lambda \left( |\Delta W_k^i|^2 + \tilde{A}_k^i + \alpha_k^i \right) \right],$$

and  $D_{n,\lambda} : \Omega \rightarrow [0, \infty)$  (precisely defined at the end of the proof), for  $i \in \mathbb{S}_N$ ,  $n \in \{0, \dots, M\}$ , are small enough. Further,  $\lambda \geq 1$  is some sufficiently large constant and  $\tilde{A}_n^i : \Omega \rightarrow [0, \infty)$ , for  $i \in \mathbb{S}_N$ ,  $n \in \{0, \dots, M\}$ , will be defined at a later stage of the proof.

Also, we introduce the quantity

$$\alpha_n^i := \mathbf{I}_{\{|Y_{t_n}^{i,N}| \geq c\}} \left\langle \frac{Y_{t_n}^{i,N}}{|Y_{t_n}^{i,N}|}, \frac{\sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}}) \Delta W_n^i}{|Y_{t_n}^{i,N}|} \right\rangle,$$

where  $c \geq 1$  and its role will become clearer at a later stage of the proof. Let  $Y_{t_n}^{i,N}$  be given by (2.5), then we aim at showing that

$$\mathbf{I}_{\Omega_{n,\lambda}^i \cap \Omega_n^i} |Y_{t_n}^{i,N}|^2 \leq D_{n,\lambda}^i + D_{n,\lambda}, \quad (3.2)$$

for all  $i \in \mathbb{S}_N$ ,  $n \in \{0, \dots, M\}$  and  $M \in \mathbb{N}$ . Inequality (3.2) is a pathwise estimate of each particle on the set of events  $\Omega_{n,\lambda}^i \cap \Omega_n^i$  and is a version of [13, Lemma 3.1] adapted to particle systems. This lemma is the crucial result to obtain  $p$ -th moment bounds of tamed schemes.

Note that on  $\Omega_{n+1}^i \cap \Omega_{n+1,\lambda}^i \cap \{\omega \mid |Y_{t_n}^{i,N}(\omega)| \leq c\}$  we have due to  $(\mathbf{A}_b^1)$ ,  $(\mathbf{A}_b^4)$ ,  $(\mathbf{A}_\sigma^1)$ ,  $(\mathbf{A}_\sigma^2)$  and  $(\mathbf{A}_\sigma^4)$  that there exists some constant  $C > 0$  such that

$$\begin{aligned} |Y_{t_{n+1}}^{i,N}| &\leq |Y_{t_n}^{i,N}| + \delta |b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}})| + C \|\sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}})\| + C \|(\nabla \sigma) \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}})\| \\ &\quad + C \frac{1}{N} \sum_{j=1}^N \|D^L \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}})(Y_{t_n}^{j,N}) \sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{Y_{t_n}^{j,N}})\| \\ &\leq c + C \delta (1 + |Y_{t_n}^{i,N}|^{\alpha_1}) |Y_{t_n}^{i,N}| + C |Y_{t_n}^{i,N}| + C \leq \lambda, \end{aligned}$$

for all  $n \in \{0, \dots, M-1\}$  and  $M \in \mathbb{N}$ , where  $\lambda \geq 1$  is chosen large enough and depends on the constants appearing in the assumptions for  $b$  and  $\sigma$  and the terminal time  $T > 0$ .

Further, we obtain from standard inequalities,

$$\begin{aligned} |Y_{t_{n+1}}^{i,N}|^2 &= \left| Y_{t_n}^{i,N} + b_\delta(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}}) \delta + \int_{t_n}^{t_{n+1}} \Upsilon_s^i dW_s^i \right|^2 \\ &\leq |Y_{t_n}^{i,N}|^2 + 2\delta^2 |b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}})|^2 + 2 \left| \int_{t_n}^{t_{n+1}} \Upsilon_s^i dW_s^i \right|^2 \\ &\quad + \frac{2\delta}{1 + \delta |b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}})|} \left\langle Y_{t_n}^{i,N}, b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{t_n}^{i,N}}) \right\rangle + 2 \left\langle Y_{t_n}^{i,N}, \int_{t_n}^{t_{n+1}} \Upsilon_s^i dW_s^i \right\rangle. \end{aligned}$$

In the sequel, we will need the set of events, for  $c$  chosen appropriately,

$$\tilde{\Omega}_{n+1}^i := \{\omega \in \Omega \mid M^{1/2\alpha_1 \wedge 1/2} \geq |Y_{t_n}^{i,N}(\omega)| \geq c\}.$$

Using the global Lipschitz assumption  $(\mathbf{A}_\sigma^1)$  for  $\sigma$  and the growth condition on  $(\nabla\sigma)\sigma$ , see  $(\mathbf{A}_\sigma^4)$ , allows us to deduce that

$$\|\sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}})\|^2 \lesssim |Y_{t_n}^{i,N}|^2, \quad \|\nabla\sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}})\sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}})\|^2 \lesssim |Y_{t_n}^{i,N}|^2. \quad (3.3)$$

From the one-sided Lipschitz assumption and the polynomial growth in the state variable in  $(\mathbf{A}_b^1)$ ,

$$\left\langle Y_{t_n}^{i,N}, b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) \right\rangle \lesssim |Y_{t_n}^{i,N}|^2, \quad |b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}})|^2 \lesssim M|Y_{t_n}^{i,N}|^2.$$

By virtue of (3.3), we obtain the estimate

$$\left| \int_{t_n}^{t_{n+1}} \Upsilon_s^i dW_s^i \right|^2 \lesssim |Y_{t_n}^{i,N}|^2 |\Delta W_n^i|^2 + |Y_{t_n}^{i,N}|^2 A_n^i + (A_n^{i,N})^2 \quad \text{on} \quad \tilde{\Omega}_{n+1}^i \cap \Omega_{n+1}^i.$$

Note that on  $\tilde{\Omega}_{n+1}^i$

$$\left\langle Y_{t_n}^{i,N}, \int_{t_n}^{t_{n+1}} \Upsilon_s^i dW_s^i \right\rangle \lesssim \left\langle Y_{t_n}^{i,N}, \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) \Delta W_n^i \right\rangle + |Y_{t_n}^{i,N}|^2 A_n^i + |Y_{t_n}^{i,N}| A_n^{i,N}.$$

Further, for  $r \neq q$  we have the following decomposition of the Lévy area

$$\int_{t_n}^{t_{n+1}} \Delta W_s^{i,r} dW_s^{i,q} = \Delta W_n^{i,r} \Delta H_n^q - \Delta W_n^{i,q} \Delta H_n^r + L_n^{r,q},$$

where  $L_n^{r,q}$  is a logistic random variable with zero mean and variance  $\frac{1}{12}\delta^2$  and  $\Delta H_n^q$  (the space time Lévy area) is normally distributed with zero mean and variance  $\frac{1}{12}\delta^2$  (see [10, 9] and references therein for details). Using this, we may write

$$\begin{aligned} |Y_{t_n}^{i,N}|^2 A_n^i &\leq |Y_{t_n}^{i,N}|^2 \left( \sum_{r,q=1}^m |\Delta W_n^{i,r} \Delta H_n^q| + \sum_{r,q=1}^m |\Delta W_n^{i,q} \Delta H_n^r| + \sum_{r,q=1}^m |L_n^{r,q}| \right) \\ &\lesssim |Y_{t_n}^{i,N}|^2 \left( \sum_{r=1}^m |(\Delta W_n^{i,r})^2| + \sum_{q=1}^m |(\Delta H_n^q)^2| + \sum_{q=1}^m |(\Delta W_n^{i,q})^2| + \sum_{r=1}^m |(\Delta H_n^r)^2| + \sum_{r,q=1}^m |L_n^{r,q}| \right) \\ &= |Y_{t_n}^{i,N}|^2 \tilde{A}_n^i. \end{aligned}$$

Therefore, on  $\tilde{\Omega}_{n+1}^i \cap \Omega_{n+1}^i$  we have, for some constant  $C > 0$ ,

$$\begin{aligned} |Y_{t_{n+1}}^{i,N}|^2 &\leq |Y_{t_n}^{i,N}|^2 + C\delta(T+1)|Y_{t_n}^{i,N}|^2 \\ &\quad + C|Y_{t_n}^{i,N}|^2(\delta + |\Delta W_n^i|^2) + 2 \left\langle Y_{t_n}^{i,N}, \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) \Delta W_n^i \right\rangle + C|Y_{t_n}^{i,N}|^2 \tilde{A}_n^i + \frac{C}{c} |Y_{t_n}^{i,N}|^2 A_n^{i,N}, \end{aligned}$$

where  $c$  will be set to  $C$ . Hence, it can be shown that there is a sufficiently large  $\lambda \geq 1$  such that

$$|Y_{t_{n+1}}^{i,N}|^2 \leq |Y_{t_n}^{i,N}|^2 \left( \exp\left(\frac{\lambda}{M} + \lambda(|\Delta W_n^i|^2 + \tilde{A}_n^i + \alpha_n^i)\right) + A_n^{i,N} \right) \quad (3.4)$$

on  $\tilde{\Omega}_{n+1}^i \cap \Omega_{n+1}^i$ . For a sufficiently large  $M$  and some positive real parameter  $a$  we have

$$\mathbb{E}(\exp(a|L_n^{p,q}|)) = \frac{2\pi + a\delta \left( H\left(-\frac{a\delta}{4\pi}\right) - H\left(-\frac{a\delta}{4\pi} - \frac{1}{2}\right) \right)}{2\pi} \leq 1 + C\delta \leq \exp(C\delta), \quad (3.5)$$

where  $C > 0$  is independent of  $\delta$  and  $H(\cdot)$  is the harmonic number for real values.

For the additional summand  $A_n^{i,N}$  in (3.4), which does not appear in [13, Lemma 3.1], defining

$$I_n^i := \exp\left(\frac{\lambda}{M} + \lambda(|\Delta W_n^i|^2 + \tilde{A}_n^i + \alpha_n^i)\right)$$

and iterating (3.4) gives, for an integer  $q < n$ ,

$$|Y_{t_{n+1}}^{i,N}|^2 \leq |Y_{t_n}^{i,N}|^2 I_n^i + |Y_{t_n}^{i,N}|^2 A_n^{i,N} \leq \dots \leq |Y_{t_q}^{i,N}|^2 \prod_{l=q}^n (I_l^i + A_l^{i,N}).$$

It remains to analyse

$$\prod_{l=q}^n (I_l^i + A_l^{i,N}) = \prod_{l=q}^n I_l^i + \prod_{l=q}^n A_l^{i,N} + \text{mixed terms},$$

where the first product can be estimated by  $D_{n+1,\lambda}^i$ , which can be treated as in [13, Lemma 3.5]. The second product and the mixed terms, in total at most  $2^M$  terms, have the form

$$A_n^{i,N} \dots I_{l_1}^i \dots A_{l_2}^{i,N} I_{l_2-1}^i \dots A_{l_3}^{i,N} \dots A_q^{i,N}. \quad (3.6)$$

Let now  $\{l_1, \tilde{l}_2, \dots, \tilde{l}_u\} =: L$ , for  $u \in \{1, \dots, n\}$ , denote the set of indices indicating an appearance of  $I^i$  in (3.6). We can estimate

$$\begin{aligned} & A_n^{i,N} \dots I_{l_1}^i \dots A_{l_2}^{i,N} I_{l_2-1}^i \dots A_{l_3}^{i,N} \dots A_q^{i,N} \\ & \leq \left( \lambda^2 + |Y_0^{i,N}|^2 \right) \exp \left[ \lambda + \sup_{\tilde{L} \subseteq L} \sum_{k \in \tilde{L}} \lambda \left( |\Delta W_k^i|^2 + \tilde{A}_k^i + \alpha_k^i \right) \right] A_n^{i,N} \dots A_{l_2}^{i,N} \dots A_{l_3}^{i,N} \dots A_q^{i,N}. \end{aligned} \quad (3.7)$$

This allows us to introduce  $D_{n+1,\lambda}^i := \prod_{l=0}^n A_l^{i,N} + \text{mixed terms}$ , where the mixed terms are defined by (3.7). Further, note that all factors of the form (3.7) are in expectation of order  $\mathcal{O}(\delta^r)$ , where  $r$  denotes the number of appearances of a factor of the form  $A_l^{i,N}$ , due to Hölder's inequality and  $\mathbb{E} \left( A_l^{i,N} \right)^q \lesssim \delta^q$  for  $q \geq 1$ . Hence, in total, we obtain a term of order  $\sum_{k=0}^M \binom{M}{k} \delta^{M-k}$ , which tends to  $e^T$  (Euler's constant), as  $M \rightarrow \infty$ .

An inductive argument as in [13, Lemma 3.1] allows us to deduce (3.2). Furthermore, due to (3.5) and (3.7), analogous statements to [13, Lemmas 3.2–3.8] still hold true  $\square$

### 3.2 Some auxiliary lemmata

**Lemma 3.1.** Assume  $(\mathbf{A}_b^1)$  and  $(\mathbf{A}_\sigma^1)$ – $(\mathbf{A}_\sigma^3)$ . Then, for all  $p \geq 1$ , there is a constant  $C > 0$  such that

$$\Lambda_t^{i,p} := \int_0^t (\mathbb{E} \|\sigma(X_s^{i,N}, \mu_s^{X^{i,N}}) - \Upsilon_s^i\|^{2p})^{\frac{1}{p}} ds \leq C \left\{ \delta^2 + \int_0^t (\mathbb{E} |Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds \right\}, \quad t \geq 0. \quad (3.8)$$

*Proof.* From  $(\mathbf{A}_\sigma^1)$  and Minkowski's inequality, we derive that

$$\begin{aligned} \Lambda_t^{i,p} & \lesssim \int_0^t (\mathbb{E} \|\sigma(X_s^{i,N}, \mu_s^{X^{i,N}}) - \sigma(Y_s^{i,N}, \mu_s^{Y^{i,N}})\|^{2p})^{\frac{1}{p}} ds + \int_0^t (\mathbb{E} \|\Gamma_s^i\|^{2p})^{\frac{1}{p}} ds \\ & \lesssim \int_0^t (\mathbb{E} |Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds + \frac{1}{N} \sum_{j=1}^N \int_0^t (\mathbb{E} |Z_s^{j,N}|^{2p})^{\frac{1}{p}} ds + \int_0^t (\mathbb{E} \|\Gamma_s^i\|^{2p})^{\frac{1}{p}} ds \\ & \lesssim \int_0^t (\mathbb{E} |Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds + \int_0^t (\mathbb{E} \|\Gamma_s^i\|^{2p})^{\frac{1}{p}} ds, \end{aligned}$$

where in the last display we used the fact that  $Z_t^{i,N}, i \in \mathbb{S}_N, t \in [0, T]$ , are identically distributed. Consequently, to derive (3.8), it is sufficient to show that

$$(\mathbb{E} \|\Gamma_t^i\|^{2p})^{\frac{1}{p}} \lesssim \delta^2. \quad (3.9)$$

In the sequel, we aim at verifying (3.9). According to [7, Proposition 3.1], we have

$$\begin{aligned} \frac{\partial \sigma_{kl}}{\partial x_j}(x_i, \hat{\mu}^{\mathbf{x},N}) &= \frac{\partial \sigma_{kl}(\cdot, \hat{\mu}^{\mathbf{x},N})}{\partial x_j}(x_i) \delta_{j,i} + \frac{1}{N} D^L \sigma_{kl}(x_i, \hat{\mu}^{\mathbf{x},N})(x_j), \\ \frac{\partial^2 \sigma_{kl}}{\partial x_j^2}(x_i, \hat{\mu}^{\mathbf{x},N}) &= \left\{ \frac{\partial^2 \sigma_{kl}(\cdot, \hat{\mu}^{\mathbf{x},N})}{\partial x_j^2}(x_i) + \frac{2}{N} \frac{\partial \{D^L \sigma_{kl}(\cdot, \hat{\mu}^{\mathbf{x},N})(x_i)\}}{\partial x_j}(x_i) \right\} \delta_{j,i} \\ & \quad + \frac{1}{N^2} (D^L)^2 \sigma_{kl}(x_i, \hat{\mu}^{\mathbf{x},N})(x_j, x_j) + \frac{1}{N} \nabla \{D^L \sigma_{kl}(x_i, \hat{\mu}^{\mathbf{x},N})(\cdot)\}(x_j), \quad i, j \in \mathbb{S}_N, \end{aligned} \quad (3.10)$$

where  $\sigma_{kl}$  is a component of  $\sigma$ , and  $k \in \{1, \dots, m\}$ ,  $l \in \{1, \dots, d\}$ . Also,  $\delta_{j,i} = 1$ , for  $j = i$  and  $\delta_{j,i} = 0$ , for  $j \neq i$ . Itô's formula in combination with (3.10) yields

$$\begin{aligned}
\Gamma_t^i &= \int_{t_\delta}^t \nabla \sigma(\cdot, \mu_s^{Y_{\cdot,N}})(Y_s^{i,N}) b_\delta(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y_{\cdot,N}}) ds + \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t D^L \sigma(Y_s^{i,N}, \mu_s^{Y_{\cdot,N}})(Y_s^{j,N}) b_\delta(Y_{s_\delta}^{j,N}, \mu_{s_\delta}^{Y_{\cdot,N}}) ds \\
&+ \frac{1}{2} \sum_{l=1}^m \int_{t_\delta}^t \left\{ \nabla^2 \sigma(\cdot, \mu_s^{Y_{\cdot,N}})(Y_s^{i,N}) + \frac{2}{N} \nabla \{ D^L \sigma(\cdot, \mu_s^{Y_{\cdot,N}})(Y_s^{i,N}) \} (Y_s^{i,N}) \right. \\
&+ \frac{1}{2} \sum_{k=1}^N \left( \frac{1}{N^2} (D^L)^2 \sigma(Y_s^{i,N}, \mu_s^{Y_{\cdot,N}})(Y_s^{k,N}, Y_s^{k,N}) \right. \\
&+ \frac{1}{N} \nabla \{ D^L \sigma(Y_s^{i,N}, \mu_s^{Y_{\cdot,N}})(\cdot) \} (Y_s^{k,N}) \left. \left. \right\} (\Upsilon_s^i e_l, \Upsilon_s^i e_l) ds \right. \\
&+ \int_{t_\delta}^t \{ \nabla \sigma(\cdot, \mu_s^{Y_{\cdot,N}})(Y_s^{i,N}) - \nabla \sigma(\cdot, \mu_{s_\delta}^{Y_{\cdot,N}})(Y_{s_\delta}^{i,N}) \} \sigma(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y_{\cdot,N}}) dW_s^i \\
&+ \int_{t_\delta}^t \nabla \sigma(\cdot, \mu_s^{Y_{\cdot,N}})(Y_s^{i,N}) \int_{s_\delta}^s (\nabla \sigma(\cdot, \mu_{r_\delta}^{Y_{\cdot,N}})(Y_{r_\delta}^{i,N}) \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^i dW_s^i \\
&+ \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t D^L \sigma(Y_s^{i,N}, \mu_s^{Y_{\cdot,N}})(Y_s^{j,N}) \int_{s_\delta}^s \nabla \sigma(\cdot, \mu_{r_\delta}^{Y_{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^j dW_s^j \\
&+ \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t \nabla \sigma(\cdot, \mu_s^{Y_{\cdot,N}})(Y_s^{i,N}) \int_{s_\delta}^s D^L \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y_{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^j dW_s^i \\
&+ \frac{1}{N^2} \sum_{j=1}^N \int_{t_\delta}^t D^L \sigma(Y_s^{i,N}, \mu_s^{Y_{\cdot,N}})(Y_s^{j,N}) \int_{s_\delta}^s D^L \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y_{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y_{\cdot,N}}) dW_r^j dW_s^i \\
&+ \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t \{ D^L \sigma(Y_s^{i,N}, \mu_s^{Y_{\cdot,N}})(Y_s^{j,N}) - D^L \sigma(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y_{\cdot,N}})(Y_{s_\delta}^{j,N}) \} \sigma(Y_{s_\delta}^{j,N}, \mu_{s_\delta}^{Y_{\cdot,N}}) dW_s^j \\
&=: \sum_{k=1}^9 \Theta_t^{k,i},
\end{aligned}$$

where  $(e_l)_{1 \leq l \leq m}$  is the standard orthogonal basis of  $\mathbb{R}^m$ . Note that Lemma 2.4 implies that there is a constant  $C > 0$  such that

$$\sup_{i \in \mathbb{S}_N} \sup_{0 \leq t \leq T} \mathbb{E} |Y_t^{i,N}|^p \leq C. \quad (3.11)$$

Due to (2) and (3) of  $(\mathbf{A}_b^1)$  and  $(\mathbf{A}_\sigma^1)$ , one has

$$|b(x, \mu)| + \|\sigma(x, \mu)\| \lesssim 1 + |x|^{1+\alpha_1} + \mu(|\cdot|^2)^{\frac{1}{2}}, \quad x, y \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d). \quad (3.12)$$

Further, note that  $(\mathbf{A}_\sigma^1)$  implies

$$\|\nabla_y \sigma(x, \mu)\| \lesssim |y|, \quad x, y \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d). \quad (3.13)$$

Obviously, (3.11) and (3.12) yield

$$\sup_{0 \leq t \leq T} \{ \mathbb{E} |b(Y_t^{i,N}, \mu_t^{Y_{\cdot,N}})|^p + \mathbb{E} \|\sigma(Y_t^{i,N}, \mu_t^{Y_{\cdot,N}})\|^p \} \leq C. \quad (3.14)$$

By Hölder's inequality and Minkowski's inequality, it follows from (3.11), (3.13), (3.14) and  $(\mathbf{A}_\sigma^2)$  that

$$\begin{aligned}
&\sum_{k=1}^3 (\mathbb{E} \|\Theta_t^{k,i}\|^{2p})^{\frac{1}{p}} \\
&\lesssim \delta \int_{t_\delta}^t (\mathbb{E} \|\nabla \sigma(\cdot, \mu_s^{Y_{\cdot,N}})(Y_s^{i,N})\|^{4p})^{\frac{1}{2p}} ds (\mathbb{E} |b(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}})|^{4p})^{\frac{1}{2p}}
\end{aligned}$$

$$\begin{aligned}
& + \frac{\delta}{N} \sum_{j=1}^N \int_{t_\delta}^t (\mathbb{E} \|D^L \sigma(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(Y_s^{j,N})\|^{4p})^{\frac{1}{2p}} ds (\mathbb{E} |b(Y_{t_\delta}^{j,N}, \mu_{t_\delta}^{Y^{\cdot,N}})|^{4p})^{\frac{1}{2p}} \\
& + \delta \int_{t_\delta}^t \left\{ (\mathbb{E} \|\nabla^2 \sigma(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N})\|^{4p})^{\frac{1}{2p}} + (\mathbb{E} \|\nabla \{D^L \sigma(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N})\}\|^{4p})^{\frac{1}{2p}} \right. \\
& + \frac{1}{N^2} \sum_{k=1}^N (\mathbb{E} \|(D^L)^2 \sigma(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(Y_s^{k,N}, Y_s^{k,N})\|^{4p})^{\frac{1}{2p}} \\
& \left. + \frac{1}{N} \sum_{k=1}^N (\mathbb{E} \|\nabla \{D^L \sigma(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(\cdot)\}(Y_s^{k,N})\|^{4p})^{\frac{1}{2p}} \right\} (\mathbb{E} \|\Upsilon_s^i\|^{4p})^{\frac{1}{2p}} ds \\
& \lesssim \delta^2.
\end{aligned} \tag{3.15}$$

Taking (3.11)-(3.14), and  $(\mathbf{A}_\sigma^4)$  into consideration yields

$$\sup_{0 \leq t \leq T} \mathbb{E} |Y_t^{i,N} - Y_{t_\delta}^{i,N}|^q \lesssim \delta^{\frac{q}{2}}, \quad q \geq 1, \quad i \in \mathbb{S}_N.$$

Therefore, employing BDG's inequality in combination with Hölder's inequality, gives that

$$\begin{aligned}
& \sum_{k=4}^9 (\mathbb{E} \|\Theta_t^{k,i}\|^{2p})^{\frac{1}{p}} \\
& \lesssim \delta^2 (\mathbb{E} \|\nabla \sigma(\cdot, \mu_{t_\delta}^{Y^{\cdot,N}})(Y_{t_\delta}^{i,N}) \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})\|^{4p})^{\frac{1}{2p}} \\
& + (\mathbb{E} \|\sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})\|^{4p})^{\frac{1}{2p}} \int_{t_\delta}^t (\mathbb{E} \|\nabla \sigma(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N}) - \nabla \sigma(\cdot, \mu_{t_\delta}^{Y^{\cdot,N}})(Y_{t_\delta}^{i,N})\|^{4p})^{\frac{1}{2p}} ds \\
& + \frac{\delta^2}{N} \sum_{j=1}^N (\mathbb{E} \|\nabla \sigma(\cdot, \mu_{t_\delta}^{Y^{\cdot,N}})(Y_{t_\delta}^{i,N}) \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})\|^{4p})^{\frac{1}{2p}} \\
& + \frac{\delta^2}{N} \sum_{j=1}^N (\mathbb{E} \|D^L \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})(Y_{t_\delta}^{j,N}) \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})\|^{4p})^{\frac{1}{2p}} \\
& + \frac{\delta^2}{N^2} \sum_{j=1}^N (\mathbb{E} \|D^L \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})(Y_{t_\delta}^{j,N}) \sigma(Y_{t_\delta}^{j,N}, \mu_{t_\delta}^{Y^{\cdot,N}})\|^{4p})^{\frac{1}{2p}} \\
& + \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t (\mathbb{E} \|D^L \sigma(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(Y_s^{j,N}) - D^L \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})(Y_{t_\delta}^{j,N})\|^{4p})^{\frac{1}{2p}} ds \\
& \times (\mathbb{E} \|\sigma(Y_{t_\delta}^{j,N}, \mu_{t_\delta}^{Y^{\cdot,N}})\|^{4p})^{\frac{1}{2p}} \lesssim \delta^2,
\end{aligned} \tag{3.16}$$

where we exploited  $(\mathbf{A}_\sigma^2)$ ,  $(\mathbf{A}_\sigma^3)$ , and (3.11)-(3.13) in the last display. Consequently, (3.9) follows from (3.15) and (3.16).  $\square$

**Lemma 3.2.** Assume  $(\mathbf{A}_b^1)$  and  $(\mathbf{A}_\sigma^1)$ – $(\mathbf{A}_\sigma^2)$ . Then, for all  $\varepsilon > 0$ ,  $p \geq 1$ , there exists a constant  $C > 0$  such that

$$\left( \mathbb{E} \left| \int_0^t \langle Z_s^{i,N}, \hat{\Upsilon}_s^i \rangle ds \right|^p \right)^{\frac{1}{p}} \leq \varepsilon (\mathbb{E} \|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + C \left\{ \int_0^t (\mathbb{E} |Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds + \delta^2 \right\}, \quad t \geq 0, \tag{3.17}$$

where

$$\hat{\Upsilon}_t^i := \nabla b(\cdot, \mu_{t_\delta}^{Y^{\cdot,N}})(Y_{t_\delta}^{i,N}) \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})(W_t^i - W_{t_\delta}^i).$$

The same result can be shown if  $\hat{\Upsilon}_t^i$  is replaced by

$$\frac{1}{N} \sum_{j=1}^N D^L b(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}})(Y_{t_\delta}^{j,N}) \sigma(Y_{t_\delta}^{j,N}, \mu_{t_\delta}^{Y^{\cdot,N}})(W_t^j - W_{t_\delta}^j).$$

*Proof.* Since

$$\begin{aligned} Z_t^{i,N} &= Z_{t_\delta}^{i,N} + \{b(X_{t_\delta}^{i,N}, \mu_{t_\delta}^{X_{\cdot,N}}) - b(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}})\}(t - t_\delta) + \{b(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}}) - b_\delta(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}})\}(t - t_\delta) \\ &\quad + \int_{t_\delta}^t \{b(X_s^{i,N}, \mu_s^{X_{\cdot,N}}) - b(X_{t_\delta}^{i,N}, \mu_{t_\delta}^{X_{\cdot,N}})\} ds \\ &\quad + \int_{t_\delta}^t \{\sigma(X_s^{i,N}, \mu_s^{X_{\cdot,N}}) - \sigma(Y_s^{i,N}, \mu_s^{Y_{\cdot,N}})\} dW_s^i + \int_{t_\delta}^t \Gamma_s^i dW_s^i, \end{aligned}$$

where  $\Gamma^i$  was introduced in (3.1), we deduce from Minkowski's inequality that

$$\begin{aligned} \left( \mathbb{E} \left| \int_0^t \langle Z_s^{i,N}, \hat{\Upsilon}_s^i \rangle ds \right|^p \right)^{\frac{1}{p}} &\leq \left( \mathbb{E} \left| \int_0^t \langle Z_{s_\delta}^{i,N}, \hat{\Upsilon}_s^i \rangle ds \right|^p \right)^{\frac{1}{p}} + \left( \mathbb{E} \left| \int_0^t \left\langle \int_{s_\delta}^s \Gamma_r^i dW_r^i, \hat{\Upsilon}_s^i \right\rangle ds \right|^p \right)^{\frac{1}{p}} \\ &\quad + \left( \mathbb{E} \left| \int_0^t \left\langle \int_{s_\delta}^s \{b(X_r^{i,N}, \mu_r^{X_{\cdot,N}}) - b(X_{s_\delta}^{i,N}, \mu_{s_\delta}^{X_{\cdot,N}})\} dr, \hat{\Upsilon}_s^i \right\rangle ds \right|^p \right)^{\frac{1}{p}} \\ &\quad + \left( \mathbb{E} \left| \int_0^t \langle \{b(X_{s_\delta}^{i,N}, \mu_{s_\delta}^{X_{\cdot,N}}) - b(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y_{\cdot,N}})\}(s - s_\delta), \hat{\Upsilon}_s^i \rangle ds \right|^p \right)^{\frac{1}{p}} \\ &\quad + \left( \mathbb{E} \left| \int_0^t \langle \{b(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y_{\cdot,N}}) - b_\delta(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y_{\cdot,N}})\}(s - s_\delta), \hat{\Upsilon}_s^i \rangle ds \right|^p \right)^{\frac{1}{p}} \\ &\quad + \left( \mathbb{E} \left| \int_0^t \left\langle \int_{s_\delta}^s \{\sigma(X_r^{i,N}, \mu_r^{X_{\cdot,N}}) - \sigma(Y_r^{i,N}, \mu_r^{Y_{\cdot,N}})\} dW_r^i, \hat{\Upsilon}_s^i \right\rangle ds \right|^p \right)^{\frac{1}{p}} \\ &=: \sum_{k=1}^6 \hat{\Lambda}_t^{k,i}. \end{aligned}$$

In what follows, we will estimate  $\hat{\Lambda}_t^{k,i}$ ,  $k = 1, \dots, 6$ , one-by-one. By Itô's formula, one has

$$d((t - t_{k+1})(W_t^i - W_{t_k}^i)) = (W_t^i - W_{t_k}^i) dt + (t - t_{k+1}) dW_t^i, \quad t \in [t_k, t_{k+1}],$$

which implies that

$$\begin{aligned} &\int_0^t \langle Z_{s_\delta}^{i,N}, \hat{\Upsilon}_s^i \rangle ds \\ &= \sum_{k=0}^{\lfloor t/\delta \rfloor} \int_{t_k}^{t_{k+1} \wedge t} \langle Z_{t_k}^{i,N}, \nabla b(\cdot, \mu_{t_k}^{Y_{\cdot,N}})(Y_{t_k}^{i,N}) \sigma(Y_{t_k}^{i,N}, \mu_{t_k}^{Y_{\cdot,N}}) d((s - t_{k+1})(W_s^i - W_{t_k}^i)) \rangle \\ &\quad - \sum_{k=0}^{\lfloor t/\delta \rfloor} \int_{t_k}^{t_{k+1} \wedge t} (s - t_{k+1}) \langle Z_{t_k}^{i,N}, \nabla b(\cdot, \mu_{t_k}^{Y_{\cdot,N}})(Y_{t_k}^{i,N}) \sigma(Y_{t_k}^{i,N}, \mu_{t_k}^{Y_{\cdot,N}}) dW_s^i \rangle \\ &= (t - (t_\delta + \delta)) \langle Z_{t_\delta}^{i,N}, \nabla b(\cdot, \mu_{t_\delta}^{Y_{\cdot,N}})(Y_{t_\delta}^{i,N}) \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}})(W_t^i - W_{t_\delta}^i) \rangle \\ &\quad - \int_0^t (s - (s_\delta + \delta)) \langle Z_{s_\delta}^{i,N}, \nabla b(\cdot, \mu_{s_\delta}^{Y_{\cdot,N}})(Y_{s_\delta}^{i,N}) \sigma(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y_{\cdot,N}}) dW_s^i \rangle. \end{aligned}$$

Hence, we deduce from Young's inequality and BDG's inequality that

$$\begin{aligned} \hat{\Lambda}_t^{1,i} &\leq \frac{2\varepsilon}{5} (\mathbb{E} \|Z_{t_\delta}^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + \frac{5\delta^2}{4\varepsilon} (\mathbb{E} \|\nabla b(\cdot, \mu_{t_\delta}^{Y_{\cdot,N}})(Y_{t_\delta}^{i,N}) \sigma(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y_{\cdot,N}})\|^{4p})^{\frac{1}{2p}} \\ &\quad + \frac{5\delta^2}{4\varepsilon} \int_0^t (\mathbb{E} \|\nabla b(\cdot, \mu_{s_\delta}^{Y_{\cdot,N}})(Y_{s_\delta}^{i,N}) \sigma(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y_{\cdot,N}})\|^{2p})^{\frac{1}{p}} ds \\ &\leq \frac{2\varepsilon}{5} (\mathbb{E} \|Z_{t_\delta}^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + C(\varepsilon)\delta^2, \end{aligned} \tag{3.18}$$

for some constant  $C(\varepsilon) > 0$ , where in the last display we utilised (3.11), (3.14) and

$$\|\nabla b(\cdot, \mu)(x)\| \lesssim 1 + |x|^{\alpha_1}, \tag{3.19}$$

due to (2) of  $(\mathbf{A}_b^1)$ . Via Hölder's inequality, we find from (3.11) and (3.12), together with (3.19) that

$$(\mathbb{E}|\hat{\Upsilon}_t^i|^{2p})^{\frac{1}{2p}} \lesssim \delta^{1/2}. \quad (3.20)$$

Next, Hölder's inequality followed by BDG's inequality, gives

$$\hat{\Lambda}_t^{2,i} \leq \int_0^t \left( \mathbb{E} \left| \int_{s_\delta}^s \Gamma_r^i dW_r^i \right|^{2p} \right)^{\frac{1}{2p}} (\mathbb{E}|\hat{\Upsilon}_s^i|^{2p})^{\frac{1}{2p}} ds \lesssim \delta^{1/2} \int_0^t \left( \delta^{p-1} \int_{s_\delta}^s \mathbb{E} \|\Gamma_r^i\|^{2p} dr \right)^{\frac{1}{2p}} ds \lesssim \delta^2, \quad (3.21)$$

where we used (3.20) in the second step, and (3.9) in the last step. Using Minkowski's inequality and Hölder's inequality, we derive from (2) and (3) of  $(\mathbf{A}_b^1)$ , (2.4), and (3.20) that

$$\begin{aligned} \hat{\Lambda}_t^{3,i} &\leq \int_0^t \int_{s_\delta}^s (\mathbb{E}|b(X_r^{i,N}, \mu_r^{X^{i,N}}) - b(X_{s_\delta}^{i,N}, \mu_{s_\delta}^{X^{i,N}})|^{2p})^{\frac{1}{2p}} dr (\mathbb{E}|\hat{\Upsilon}_s^i|^{2p})^{\frac{1}{2p}} ds \\ &\lesssim \delta^{1/2} \int_0^t \int_{s_\delta}^s \left\{ (1 + (\mathbb{E}|X_r^{i,N}|^{4p\alpha_1})^{\frac{1}{4p}} + (\mathbb{E}|X_{s_\delta}^{i,N}|^{4p\alpha_1})^{\frac{1}{4p}}) (\mathbb{E}|X_r^{i,N} - X_{s_\delta}^{i,N}|^{4p})^{\frac{1}{4p}} \right. \\ &\quad \left. + \frac{1}{N} \sum_{j=1}^N (\mathbb{E}|X_r^{j,N} - X_{s_\delta}^{j,N}|^{2p})^{\frac{1}{2p}} \right\} dr ds \lesssim \delta^2, \end{aligned} \quad (3.22)$$

where in the last estimate we utilised that  $\mathbb{E}\|X^{i,N}\|_{\infty,T}^q \leq C$  and  $\mathbb{E}|X_t^{i,N} - X_{t_\delta}^{i,N}|^q \lesssim \delta^{q/2}$ , for  $q \geq 1$  and  $i \in \mathbb{S}_N$ . Applying Hölder's inequality and Young's inequality, we deduce from (3.11) and (3.20) that

$$\begin{aligned} \hat{\Lambda}_t^{4,i} &\leq L_b^1 \delta \int_0^t (\mathbb{E}(|Z_{s_\delta}^{i,N}|(1 + |X_{s_\delta}^{i,N}|^{\alpha_1} + |Y_{s_\delta}^{i,N}|^{\alpha_1})|\hat{\Upsilon}_s^i|)^p)^{\frac{1}{p}} ds \\ &\quad + \frac{L_b^1 \delta}{N} \sum_{j=1}^N \int_0^t (\mathbb{E}(|Z_{s_\delta}^{j,N}||\hat{\Upsilon}_s^i|)^p)^{\frac{1}{p}} ds \\ &\leq \frac{2\varepsilon}{5} (\mathbb{E}\|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + \frac{5(L_b^1)^2 t \delta^2}{4\varepsilon} \int_0^t (\mathbb{E}|\hat{\Upsilon}_s^i|^{2p})^{\frac{1}{p}} ds \\ &\quad + \frac{5(L_b^1)^2 t \delta^2}{4\varepsilon} \int_0^t (1 + (\mathbb{E}|X_{s_\delta}^{i,N}|^{2p\alpha_1})^{\frac{1}{2p}} + (\mathbb{E}|Y_{s_\delta}^{i,N}|^{2p\alpha_1})^{\frac{1}{2p}}) (\mathbb{E}|\hat{\Upsilon}_s^i|^{2p})^{\frac{1}{2p}} ds \\ &\leq \frac{2\varepsilon}{5} (\mathbb{E}\|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + C_2(\varepsilon) \delta^2 \end{aligned} \quad (3.23)$$

for some constant  $C(\varepsilon) > 0$ . Taking advantage of (3.14) and (3.20) gives

$$\hat{\Lambda}_t^{5,i} \leq \delta^2 \int_0^t (\mathbb{E}(|b(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y^{i,N}})| |\hat{\Upsilon}_s^i|)^p)^{\frac{1}{p}} ds \lesssim \delta^2. \quad (3.24)$$

Finally, applying BDG's, Minkowski's, and Young's inequalities, and taking  $(\mathbf{A}_\sigma^1)$ , (2.4) and (3.20) into account, leads to

$$\begin{aligned} \hat{\Lambda}_t^{6,i} &\leq \frac{\varepsilon}{10\delta t} \int_0^t \int_{s_\delta}^s (\mathbb{E}\|\sigma(X_r^{i,N}, \mu_r^{X^{i,N}}) - \sigma(Y_r^{i,N}, \mu_r^{Y^{i,N}})\|^{2p})^{\frac{1}{p}} dr ds + \frac{5t\delta}{2\varepsilon} \int_0^t (\mathbb{E}|\hat{\Upsilon}_s^i|^{2p})^{\frac{1}{p}} ds \\ &\leq \frac{\varepsilon}{5} (\mathbb{E}\|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + C(\varepsilon) \delta^2 \end{aligned} \quad (3.25)$$

for some constant  $C(\varepsilon) > 0$ . Consequently, (3.17) follows from (3.18) and (3.21)-(3.25).  $\square$

**Lemma 3.3.** Assume  $(\mathbf{A}_b^1)$ – $(\mathbf{A}_b^3)$  and  $(\mathbf{A}_\sigma^1)$ – $(\mathbf{A}_\sigma^2)$ . Then, for all  $\varepsilon > 0$ ,  $p \geq 1$ , there exists a constant  $C > 0$  such that

$$\begin{aligned} \hat{\Gamma}_t^{i,p} &:= \int_0^t (\mathbb{E}|\langle Z_s^{i,N}, b(Y_s^{i,N}, \mu_s^{Y^{i,N}}) - b(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y^{i,N}}) \rangle|^p)^{\frac{1}{p}} ds \\ &\leq \varepsilon (\mathbb{E}\|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + C \left\{ \int_0^t (\mathbb{E}|Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds + \delta^2 \right\}, \quad t \geq 0. \end{aligned} \quad (3.26)$$

*Proof.* Using (3.10) with  $\sigma$  therein replaced by  $b$ , we obtain from Itô's formula that

$$\begin{aligned}
& b(Y_t^{i,N}, \mu_t^{Y^{\cdot,N}}) - b(Y_{t_\delta}^{i,N}, \mu_{t_\delta}^{Y^{\cdot,N}}) \\
&= \int_{t_\delta}^t \nabla b(\cdot, \mu_{s_\delta}^{Y^{\cdot,N}})(Y_{s_\delta}^{i,N}) \sigma(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y^{\cdot,N}}) dW_s^i \\
&+ \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t D^L b(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y^{\cdot,N}})(Y_{s_\delta}^{j,N}) \sigma(Y_{s_\delta}^{j,N}, \mu_{s_\delta}^{Y^{\cdot,N}}) dW_s^j + \int_{t_\delta}^t \nabla b(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N}) b_\delta(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y^{\cdot,N}}) ds \\
&+ \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t D^L b(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(Y_s^{j,N}) b_\delta(Y_{s_\delta}^{j,N}, \mu_{s_\delta}^{Y^{\cdot,N}}) ds \\
&+ \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t \{D^L b(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(Y_s^{j,N}) - D^L b(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y^{\cdot,N}})(Y_s^{j,N})\} \sigma(Y_{s_\delta}^{j,N}, \mu_{s_\delta}^{Y^{\cdot,N}}) dW_s^j \\
&+ \frac{1}{2} \sum_{l=1}^m \int_{t_\delta}^t \left\{ \nabla^2 b(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N}) + \frac{2}{N} \nabla \{D^L b(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N})\} (Y_s^{i,N}) \right. \\
&+ \left. \frac{1}{2} \sum_{k=1}^N \left( \frac{1}{N^2} (D^L)^2 b(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(Y_s^{k,N}, Y_s^{k,N}) + \frac{1}{N} \nabla \{D^L b(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(\cdot)\} (Y_s^{k,N}) \right) \right\} (\Upsilon_{s_{el}}^i, \Upsilon_{s_{el}}^i) ds \\
&+ \int_{t_\delta}^t \left\{ \nabla b(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N}) - \nabla b(\cdot, \mu_{s_\delta}^{Y^{\cdot,N}})(Y_{s_\delta}^{i,N}) \right\} \sigma(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y^{\cdot,N}}) dW_s^i \\
&+ \int_{t_\delta}^t \nabla b(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N}) \int_{s_\delta}^s (\nabla \sigma(\cdot, \mu_{r_\delta}^{Y^{\cdot,N}})(Y_{r_\delta}^{i,N}) \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y^{\cdot,N}})) dW_r^i dW_s^i \\
&+ \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t \nabla b(\cdot, \mu_s^{Y^{\cdot,N}})(Y_s^{i,N}) \int_{s_\delta}^s D^L \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y^{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y^{\cdot,N}}) dW_r^j dW_s^i \\
&+ \frac{1}{N} \sum_{j=1}^N \int_{t_\delta}^t D^L b(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(Y_s^{j,N}) \int_{s_\delta}^s \nabla \sigma(\cdot, \mu_{r_\delta}^{Y^{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y^{\cdot,N}}) dW_r^j dW_s^j \\
&+ \frac{1}{N^2} \sum_{j=1}^N \int_{t_\delta}^t D^L b(Y_s^{i,N}, \mu_s^{Y^{\cdot,N}})(Y_s^{j,N}) \int_{s_\delta}^s D^L \sigma(Y_{r_\delta}^{i,N}, \mu_{r_\delta}^{Y^{\cdot,N}})(Y_{r_\delta}^{j,N}) \sigma(Y_{r_\delta}^{j,N}, \mu_{r_\delta}^{Y^{\cdot,N}}) dW_r^j dW_s^i \\
&=: \sum_{l=1}^{11} \Psi_t^{l,i},
\end{aligned}$$

where  $\Upsilon^i$  was given in (3.1). Therefore, we have

$$\hat{\Gamma}_t^{i,p} \leq \sum_{l=1}^{11} \int_0^t (\mathbb{E} |\langle Z_s^{i,N}, \Psi_s^{l,i} \rangle|^p)^{1/p} ds.$$

By virtue of Lemma 3.2, one has

$$\int_0^t (\mathbb{E} |\langle Z_s^{i,N}, \Psi_s^{1,i} \rangle|^p)^{\frac{1}{p}} ds \leq \frac{\varepsilon}{8} (\mathbb{E} \|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + C \left\{ \int_0^t (\mathbb{E} |Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds + \delta^2 \right\}, \quad (3.27)$$

for some constant  $C > 0$  and  $\epsilon > 0$ . On the other hand, following the steps of the proof of Lemma 3.2, there exists a constant  $C > 0$  such that

$$\int_0^t \mathbb{E} |\langle Z_s^{i,N}, \Psi_s^{2,i} \rangle|^p)^{\frac{1}{p}} ds \leq \frac{\varepsilon}{8} (\mathbb{E} \|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + C \left\{ \int_0^t (\mathbb{E} |Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds + \delta^2 \right\}. \quad (3.28)$$

Next, by Young's inequality, it follows that

$$\sum_{l=3}^{11} \int_0^t (\mathbb{E} (|Z_s^{i,N}|^p |\Psi_s^{l,i}|^p))^{\frac{1}{p}} ds \leq \frac{\varepsilon}{2} (\mathbb{E} \|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + \frac{1}{2\varepsilon} \sum_{l=3}^{11} \int_0^t (\mathbb{E} |\Psi_s^{l,i}|^{2p})^{\frac{1}{p}} ds.$$



Following analogous arguments to those for (3.15) and (3.16), we deduce from  $(\mathbf{A}_b^1)$ – $(\mathbf{A}_b^3)$  and  $(\mathbf{A}_\sigma^1)$ – $(\mathbf{A}_\sigma^2)$

$$\sum_{l=3}^{11} \int_0^t (\mathbb{E}(|\Psi_s^{l,i}|^{2p}))^{\frac{1}{p}} ds \lesssim \delta^2. \quad (3.29)$$

As a consequence, (3.26) follows directly from (3.27), (3.28) and (3.29).  $\square$

### 3.3 Proof of Theorem 2.5

With Lemmas 3.1 and 3.3 at hand, we are in a position to complete the proof of Theorem 2.5:

*Proof.* In the sequel, we fix  $i \in \mathbb{S}_N$  and recall  $t \in [0, T]$ . We aim at showing that

$$\mathbb{E} \|Z^{i,N}\|_{\infty,T}^p \lesssim \delta^{2p}, \quad (3.30)$$

for  $p \geq 1$ . By Itô's formula, along with  $Y_0^{i,N} = X_0^{i,N}$ , it follows that

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

where  $M_s^i$  and  $\Upsilon_s^i$  were defined in (3.1). Obviously, we have

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

From the one-sided Lipschitz condition in  $(\mathbf{A}_b^1)$ , one obviously has

$$\|\Pi^{1,i}\|_{\infty,t} \leq L_1^b \int_0^t |Z_s^{i,N}|^2 ds. \quad (3.31)$$

Due to (3) of  $(\mathbf{A}_b^1)$  and Young's inequality, we obtain from (2.4)

$$\begin{aligned} \|\Pi^{2,i}\|_{\infty,t} &\leq \frac{1}{2} \int_0^t |Z_s^{i,N}|^2 ds + \frac{(L_b^1)^2}{2} \int_0^t \mathbb{W}_2(\mu_s^{X^{i,N}}, \mu_s^{Y^{i,N}})^2 ds \\ &\leq \frac{1}{2} \int_0^t |Z_s^{i,N}|^2 ds + \frac{(L_b^1)^2}{2N} \sum_{j=1}^N \int_0^t |Z_s^{j,N}|^2 ds. \end{aligned} \quad (3.32)$$

In terms of (2) of  $(\mathbf{A}_b^1)$ , it follows immediately from Young's inequality that

$$\begin{aligned} \|\Pi^{3,i}\|_{\infty,t} &\lesssim \int_0^t |Z_s^{i,N}|^2 ds + \delta^2 \int_0^t |b(Y_{s_\delta}^{i,N}, \mu_{s_\delta}^{Y^{i,N}})|^2 ds \\ &\lesssim \int_0^t |Z_s^{i,N}|^2 ds + \delta^2 \int_0^t \left\{ 1 + |Y_{s_\delta}^{i,N}|^{2(1+\alpha_1)} + \frac{1}{N} \sum_{j=1}^N |Y_{s_\delta}^{j,N}|^2 \right\} ds. \end{aligned} \quad (3.33)$$

Thus, combining (3.31), (3.32) with (3.33) and taking advantage of Lemmas 3.1 and 3.3 (with  $\epsilon$  small enough), besides (3.11) and (3.14), we infer from Minkowski's inequality that, for some constant  $C > 0$

$$(\mathbb{E} \|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} \leq C \left\{ \int_0^t (\mathbb{E} |Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds + \frac{1}{N} \sum_{j=1}^N \int_0^t (\mathbb{E} |Z_s^{j,N}|^{2p})^{\frac{1}{p}} ds \right\}$$

$$\begin{aligned}
& + \delta^2 \int_0^t \left\{ 1 + (\mathbb{E}|Y_{s\delta}^{i,N}|^{2p(1+\alpha_1)})^{\frac{1}{p}} + \frac{1}{N} \sum_{j=1}^N (\mathbb{E}|Y_{s\delta}^{j,N}|^{2p})^{\frac{1}{p}} \right\} ds \\
& + \hat{\Gamma}_t^{i,p} + \Lambda_t^{i,p} + \left( \mathbb{E} \left\| \int_0^\cdot \langle Z_s^{i,N}, dM_s^i \rangle \right\|_{\infty,t}^p \right)^{\frac{1}{p}} \Big\} \\
& \leq \frac{1}{2} (\mathbb{E} \|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + C \left\{ \delta^2 + \int_0^t (\mathbb{E} |Z_s^{i,N}|^{2p})^{\frac{1}{p}} ds \right\},
\end{aligned} \tag{3.34}$$

where we used that  $Z_t^{i,N}, i \in \mathbb{S}_N$ , are identically distributed and, for some constant  $C > 0$ ,

$$\left( \mathbb{E} \left\| \int_0^\cdot \langle Z_s^{i,N}, dM_s^i \rangle \right\|_{\infty,t}^p \right)^{\frac{1}{p}} \leq \frac{\epsilon}{2} (\mathbb{E} \|Z^{i,N}\|_{\infty,t}^{2p})^{\frac{1}{p}} + \frac{C}{2\epsilon} \Lambda_t^{i,p}.$$

Thus, (3.30) follows from (3.34) and Grönwall's inequality.  $\square$

### 3.4 Proof of 2.9

*Proof.* From the definitions of  $Y_{t_{n+1}}^i$  and  $Y_{t_{n+1}}^{i,N}$ , we obtain directly

$$\begin{aligned}
& \left| Y_{t_{n+1}}^{i,N} - Y_{t_{n+1}}^i \right|^2 \\
& = \left| Y_{t_n}^{i,N} - Y_{t_n}^i + b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) \delta - b(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \delta + \left( \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) - \sigma(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \right) \Delta W_n^i \right. \\
& \quad + \int_{t_n}^{t_{n+1}} \nabla \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) \sigma(Y_{t_n}^i, \mu_{t_n}^{Y_{\cdot,N}}) \int_{t_n}^s dW_u^i dW_s^i - \int_{t_n}^{t_{n+1}} \nabla \sigma(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \sigma(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \int_{t_n}^s dW_u^i dW_s^i \\
& \quad \left. + \int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^N D^L \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) (Y_{t_n}^{j,N}) \sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{Y_{\cdot,N}}) \int_{t_n}^s dW_u^j dW_s^i \right|^2.
\end{aligned}$$

Squaring out the right side of the previous equality and taking the expectation on both sides gives

$$\begin{aligned}
& \mathbb{E} \left| Y_{t_{n+1}}^{i,N} - Y_{t_{n+1}}^i \right|^2 \\
& \leq \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + 4\mathbb{E} \left| b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) \delta - b(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \delta \right|^2 + 4\mathbb{E} \left| \left( \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) - \sigma(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \right) \Delta W_n^i \right|^2 \\
& \quad + 4\mathbb{E} \left| \int_{t_n}^{t_{n+1}} \nabla \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) \sigma(Y_{t_n}^i, \mu_{t_n}^{Y_{\cdot,N}}) \int_{t_n}^s dW_u^i dW_s^i - \int_{t_n}^{t_{n+1}} \nabla \sigma(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \sigma(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \int_{t_n}^s dW_u^i dW_s^i \right|^2 \\
& \quad + 4\mathbb{E} \left| \int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^N D^L \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) (Y_{t_n}^{j,N}) \sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{Y_{\cdot,N}}) \int_{t_n}^s dW_u^j dW_s^i \right|^2 \\
& \quad + 2\mathbb{E} \left\langle Y_{t_n}^{i,N} - Y_{t_n}^i, b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) \delta - b(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \delta \right\rangle = \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + \sum_{i=1}^5 \Pi_i.
\end{aligned}$$

In the sequel, we estimate each of these expectations one-by-one.

First, using the imposed Lipschitz assumptions on  $b(\cdot, \cdot)$ ,  $\sigma(\cdot, \cdot)$  and  $(\nabla \sigma)\sigma(\cdot, \cdot)$ , we can readily estimate, for some constant  $C > 0$ ,

$$\Pi_1 + \Pi_2 + \Pi_3 \leq C\delta \left( \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + \mathbb{E} \mathbb{W}_2 \left( \mathcal{L}_{Y_{t_n}^i}, \tilde{\mu}_{t_n}^{\tilde{Y}_{\cdot,N}} \right)^2 \right),$$

where  $\tilde{\mu}_{t_n}^{\tilde{Y}_{\cdot,N}}$  is the empirical distribution of non-interacting particle system  $(Y_{t_n}^i)_{i \in \mathbb{S}_N}$ . The term  $\Pi_4$  can be treated as follows: First, we compute

$$\mathbb{E} \left| \int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^N D^L \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y_{\cdot,N}}) (Y_{t_n}^{j,N}) \sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{Y_{\cdot,N}}) \int_{t_n}^s dW_u^j dW_s^i \right|^2$$

$$\begin{aligned}
&\leq 2\mathbb{E}\left|\int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^N \left( D^L \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{Y^{\cdot,N}})(Y_{t_n}^{j,N}) \sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{Y^{\cdot,N}}) - D^L \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j})(Y_{t_n}^j) \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j}) \right) \int_{t_n}^s dW_u^j dW_s^i \right|^2 \\
&\quad + 2\mathbb{E}\left|\int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^N D^L \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j})(Y_{t_n}^j) \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j}) \int_{t_n}^s dW_u^j dW_s^i \right|^2 \\
&=: \Pi_{41} + \Pi_{42}.
\end{aligned}$$

Also, due to assumption  $(\mathbf{AA}_\sigma)$  (2), we obtain

$$\Pi_{41} \leq C\delta^2 \left( \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + \mathbb{E} \mathbb{W}_2 \left( \mathcal{L}_{Y_{t_n}^i}, \tilde{\mu}_{t_n}^{\tilde{Y}^{\cdot,N}} \right)^2 \right).$$

Further, for ease of notation, we restrict the analysis to  $d = 1$  for the subsequent analysis:

$$\begin{aligned}
\Pi_{42} &\leq C\mathbb{E} \left( \int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^N D^L \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j})(Y_{t_n}^j) \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j}) \int_{t_n}^s dW_u^j dW_s^i \right)^2 \\
&\leq C \frac{1}{N^2} \sum_{j,k=1}^N \mathbb{E} \left[ \left( \int_{t_n}^{t_{n+1}} D^L \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j})(Y_{t_n}^j) \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j}) \int_{t_n}^s dW_u^j dW_s^i \right) \right. \\
&\quad \times \left. \left( \int_{t_n}^{t_{n+1}} D^L \sigma(Y_{t_n}^k, \mathcal{L}_{Y_{t_n}^k})(Y_{t_n}^k) \sigma(Y_{t_n}^k, \mathcal{L}_{Y_{t_n}^k}) \int_{t_n}^s dW_u^k dW_s^i \right) \right] \\
&= C \frac{1}{N^2} \sum_{j,k=1}^N \mathbb{E} \left[ \int_{t_n}^{t_{n+1}} \left( D^L \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j})(Y_{t_n}^j) \sigma(Y_{t_n}^j, \mathcal{L}_{Y_{t_n}^j}) \int_{t_n}^s dW_u^j \right) \right. \\
&\quad \times \left. \left( D^L \sigma(Y_{t_n}^k, \mathcal{L}_{Y_{t_n}^k})(Y_{t_n}^k) \sigma(Y_{t_n}^k, \mathcal{L}_{Y_{t_n}^k}) \int_{t_n}^s dW_u^k \right) ds \right] \leq C \frac{\delta^2}{N},
\end{aligned}$$

where in the last display, we employed the independence and moment stability of the random variables  $(Y_{t_n}^i)_{i \in \mathbb{S}_N}$  and the Brownian motions, respectively. Hence, using above estimates for  $\Pi_{41}$  and  $\Pi_{42}$ , we get

$$\Pi_4 \leq C\delta \left( \frac{\delta}{N} + \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + \mathbb{E} \mathbb{W}_2 \left( \mathcal{L}_{Y_{t_n}^i}, \tilde{\mu}_{t_n}^{\tilde{Y}^{\cdot,N}} \right)^2 \right).$$

Next, we compute, using the fact that  $b(\cdot, \cdot)$  is Lipschitz continuous in both components,

$$\begin{aligned}
\Pi_5 &= 2\mathbb{E} \left\langle Y_{t_n}^{i,N} - Y_{t_n}^i, b(Y_{t_n}^{i,N}, \mu_{t_n}^{Y^{\cdot,N}}) \delta - b(Y_{t_n}^i, \mathcal{L}_{Y_{t_n}^i}) \delta \right\rangle \\
&\leq C\delta \left( \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + \mathbb{E} \mathbb{W}_2 \left( \mathcal{L}_{Y_{t_n}^i}, \tilde{\mu}_{t_n}^{\tilde{Y}^{\cdot,N}} \right)^2 \right).
\end{aligned}$$

Hence, putting everything together, we obtain

$$\begin{aligned}
\mathbb{E} \left| Y_{t_{n+1}}^{i,N} - Y_{t_n}^i \right|^2 &\leq \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + C\delta \left( \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + \frac{\delta}{N} + \mathbb{E} \mathbb{W}_2 \left( \mathcal{L}_{Y_{t_n}^i}, \tilde{\mu}_{t_n}^{\tilde{Y}^{\cdot,N}} \right)^2 \right) \\
&\leq (1 + C\delta) \mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 + C\delta\phi(N),
\end{aligned}$$

where  $\phi(N)$  is defined in the statement of Corollary 2.6. The constant  $C$  is independent of the present time-step  $n \in \{0, \dots, M\}$ . Consequently, recalling that  $Y_{t_0}^{i,N} = Y_{t_0}^i$ , one can inductively prove that  $\mathbb{E} \left| Y_{t_n}^{i,N} - Y_{t_n}^i \right|^2 \leq C\phi(N)$ , for any  $n \in \{0, \dots, M\}$ .  $\square$

## 4 Numerical results and implementation details

We now present a number of numerical tests to illustrate our theoretical results. For the following Examples 1 to 5, we employ the particle method to approximate the law  $\mathcal{L}_{Y_{t_n}}$  at each time-step  $t_n$ ,

$n \in \{0, \dots, M\}$ , by its empirical distribution. For our numerical experiments we used  $N = 10^4$  (and  $N = 10^3$  in Example 3), unless otherwise stated.

As we do not know the exact solution in the considered examples, the strong convergence with respect to the number of time-steps is assessed by comparing two solutions computed on a fine and coarse time grid, respectively, using the same samples of Brownian motion. Specifically, we compute the root-mean-square error (RMSE) between the numerical solution  $Y_T^{i,N,l}$  at a level  $l$  of the time-discretisation with  $2^l T$  time steps and  $N$  particles, and the solution at level  $l-1$ , at the final time  $T = 1$ ,

$$\text{RMSE} := \sqrt{\frac{1}{N} \sum_{i=1}^N \left( Y_T^{i,N,l} - Y_T^{i,N,l-1} \right)^2},$$

where  $i$  refers to the  $i$ -th particle,  $i \in \mathbb{S}_N$ .

In Examples 1 to 5 below, we study the convergence order in  $\delta$ , the relevance of the  $L$ -derivative for a small number of particles and the asymptotic behaviour of the  $L$ -derivative terms of the Milstein scheme, and propagation of chaos.

## 4.1 Examples

In the examples presented below, we choose the parameter values  $\sigma = 1.5$ ,  $c = 0.5$  and set  $X_0 = 1$ . These non-globally Lipschitz SDEs satisfy all assumptions needed to guarantee a unique strong solution.

While Examples 1 to 3 have a linear measure dependence, i.e.,  $\sigma$  (or  $b$ ) are of the form  $\sigma(x, \mu) = \int_{\mathbb{R}} \tilde{\sigma}(x, x') \mu(dx')$ , for some function  $\tilde{\sigma} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ , Examples 4 and 5 have coefficients with non-linear measure dependence.

**Example 1:** 
$$dX_t = \left( \frac{\sigma^2}{2} X_t - X_t^3 + c \mathbb{E} X_t \right) dt + \mathbb{E} X_t dW_t.$$

**Example 2:** 
$$dX_t = \left( \frac{\sigma^2}{2} X_t - X_t^3 + c \mathbb{E} X_t \right) dt + X_t dW_t.$$

**Example 3:** 
$$dX_t = \left( \frac{\sigma^2}{2} X_t - X_t^3 + \int_{\mathbb{R}} \sin(X_t - y) \mathcal{L}_{X_t}(dy) \right) dt + \left( X_t + \int_{\mathbb{R}} \sin(X_t - y) \mathcal{L}_{X_t}(dy) \right) dW_t.$$

**Example 4:** 
$$dX_t = \left( \frac{\sigma^2}{2} X_t - X_t^3 + \mathbb{E} X_t \right) dt + \int_{\mathbb{R}} \int_{\mathbb{R}} \sin(x + y) \mathcal{L}_{X_t}(dy) \mathcal{L}_{X_t}(dx) dW_t.$$

**Example 5:** 
$$dX_t = \left( \frac{\sigma^2}{2} X_t - X_t^3 + \mathbb{E} X_t \right) dt + e^{-\mathbb{V}\left[\frac{X_t}{1+X_t^2}\right]} dW_t.$$

Note that not all model assumptions needed to guarantee moment stability for Scheme 1 are satisfied in the case of Examples 1, 2, 4 and 5, as  $(\mathbf{A}_b^4)$  and/or  $(\mathbf{A}_\sigma^4)$  are violated. However, as illustrated numerically, it seems reasonable that the established theory is also valid in a more general setting. Furthermore, we remark that in Example 1 the diffusion coefficient only contains an expectation and does not explicitly depend on the current state  $X_t$  (i.e., the derivatives with respect to the state component vanishes). Similarly, in Examples 4 and 5 we also do not have an explicit state dependence. Hence, we expect the tamed Euler scheme to converge with order 1. This is confirmed numerically below as well. The drift  $b_\delta$  of the tamed Euler scheme is chosen as in Scheme 1 for the Milstein scheme.

## 4.2 Time-stepping schemes and $L$ -derivative

We first use the two different tamed Milstein schemes (i.e., Scheme 1 and Scheme 2) without the  $L$ -derivative terms for Examples 1 to 3. For all these examples, we observe in Fig. 1a and Fig. 1b strong convergence of order 1.

From these numerical results, it is also apparent that Scheme 1 consistently outperforms Scheme 2 in terms of accuracy. We further note that the strong convergence order for Scheme 2 seems to be observable for extremely fine time-grids (i.e.,  $M \geq 2^{12}$ ) only, which might be due to a large implied constant appearing in the strong convergence analysis of Remark 2.7.

Note that in the above examples the  $L$ -derivative of the diffusion term can be explicitly computed, which allows us to compute the term (2.10). In Example 1, the  $L$ -derivative of the diffusion term is simply  $D^L \sigma(x, \mu)(y) = 1$ , for all  $y \in \mathbb{R}$  (see [6, Example 1 on page 385]). Similarly, the  $L$ -derivative of

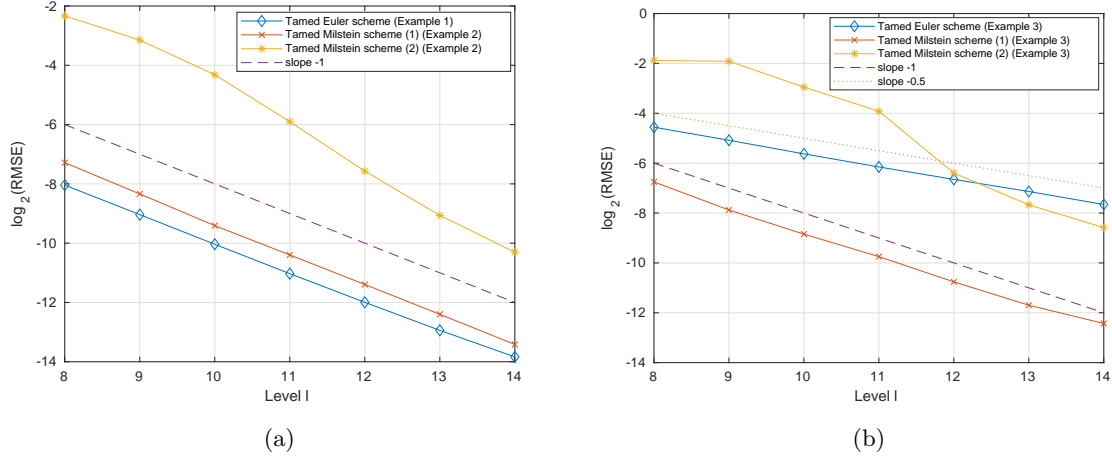


Figure 1: Numerical illustration for the strong convergence of Example 1, Example 2 (both left), and Example 3 (right) using the tamed Euler scheme and both tamed Milstein schemes, Scheme 1 and Scheme 2 (without the  $L$ -derivative terms), respectively. In Example 1, the tamed Milstein scheme without the  $L$ -derivative terms then coincides with the tamed Euler scheme.

the diffusion term appearing in Example 3 can be computed as  $D^L \sigma(x, \mu)(y) = \partial_y \sin(x-y) = -\cos(x-y)$ , for all  $y \in \mathbb{R}$  (see [6, Example 4 on page 389]), and for Example 4 as  $D^L \sigma(x, \mu)(y) = 2 \int_{\mathbb{R}} \cos(x+y) \mu(dx)$ , for all  $y \in \mathbb{R}$  (see [6, Example 3 on page 387]). Finally, for Example 5, the chain rule shows that the  $L$ -derivative of  $f : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  defined as follows:  $f(\mu) := \exp \left( - \int_{\mathbb{R}} g^2(x) \mu(dx) + \left( \int_{\mathbb{R}} g(x) \mu(dx) \right)^2 \right)$ , for some bounded and continuously differentiable function  $g : \mathbb{R} \rightarrow \mathbb{R}$  (with bounded derivative), is

$$D^L f(\mu)(y) = \exp \left( - \int_{\mathbb{R}} g^2(x) \mu(dx) + \left( \int_{\mathbb{R}} g(x) \mu(dx) \right)^2 \right) \left( -2g(y)g'(y) + 2g'(y) \int_{\mathbb{R}} g(x) \mu(dx) \right),$$

which is then applied with  $g(x) = \frac{x}{1+x^2}$ .

We now state, for illustration purposes, the (full) tamed Milstein schemes for Examples 1 and 3. The tamed Milstein scheme (Scheme 1) for Example 1 reads as

$$\begin{aligned} Y_{t_{n+1}}^{i,N} &= Y_{t_n}^{i,N} + \frac{\frac{\sigma^2}{2} Y_{t_n}^{i,N} - (Y_{t_n}^{i,N})^3 + \frac{c}{N} \sum_{j=1}^N Y_{t_n}^{j,N}}{1 + \delta \left| \frac{\sigma^2}{2} Y_{t_n}^{i,N} - (Y_{t_n}^{i,N})^3 + \frac{c}{N} \sum_{j=1}^N Y_{t_n}^{j,N} \right|} \delta + \frac{1}{N} \sum_{j=1}^N Y_{t_n}^{j,N} \Delta W_n^i \\ &\quad + \frac{1}{N} \sum_{l=1}^N Y_{t_n}^{l,N} \frac{1}{N} \sum_{j=1}^N \int_{t_n}^{t_{n+1}} \int_{t_n}^s dW_u^j dW_s^i. \end{aligned}$$

The tamed Milstein scheme (Scheme 1) for Example 3 has the form

$$\begin{aligned} Y_{t_{n+1}}^{i,N} &= Y_{t_n}^{i,N} + \frac{\frac{\sigma^2}{2} Y_{t_n}^{i,N} - (Y_{t_n}^{i,N})^3 + \frac{1}{N} \sum_{j=1}^N \sin(Y_{t_n}^{i,N} - Y_{t_n}^{j,N})}{1 + \delta \left| \frac{\sigma^2}{2} Y_{t_n}^{i,N} - (Y_{t_n}^{i,N})^3 + \frac{1}{N} \sum_{j=1}^N \sin(Y_{t_n}^{i,N} - Y_{t_n}^{j,N}) \right|} \delta \\ &\quad + \left( Y_{t_n}^{i,N} + \frac{1}{N} \sum_{j=1}^N \sin(Y_{t_n}^{i,N} - Y_{t_n}^{j,N}) \right) \Delta W_n^i \\ &\quad + \left( 1 + \frac{1}{N} \sum_{j=1}^N \cos(Y_{t_n}^{i,N} - Y_{t_n}^{j,N}) \right) \left( Y_{t_n}^{i,N} + \frac{1}{N} \sum_{j=1}^N \sin(Y_{t_n}^{i,N} - Y_{t_n}^{j,N}) \right) \frac{(\Delta W_n^i)^2 - \delta}{2} \\ &\quad - \frac{1}{N} \sum_{j=1}^N \cos(Y_{t_n}^{i,N} - Y_{t_n}^{j,N}) \left( Y_{t_n}^{j,N} + \frac{1}{N} \sum_{l=1}^N \sin(Y_{t_n}^{j,N} - Y_{t_n}^{l,N}) \right) \int_{t_n}^{t_{n+1}} \int_{t_n}^s dW_u^j dW_s^i. \end{aligned}$$

To illustrate the effect of the  $L$ -derivative, in Fig. 2a, we compare for Example 3 the strong convergence rate of the tamed Milstein scheme (Scheme 1) without the terms involving the  $L$ -derivative, with a full-tamed Milstein scheme, i.e., including the  $L$ -derivative terms, for  $N = 20$ . We employ the approximation to the Lévy areas proposed in [31]. The complexity for computing the Lévy areas for all time-steps is  $\mathcal{O}(N^2 M^{3/2})$ , where the additional factor  $M^{1/2}$  comes from the choice for the truncation level of the so-called Karhunen-Loève expansion of certain Brownian Bridge processes.

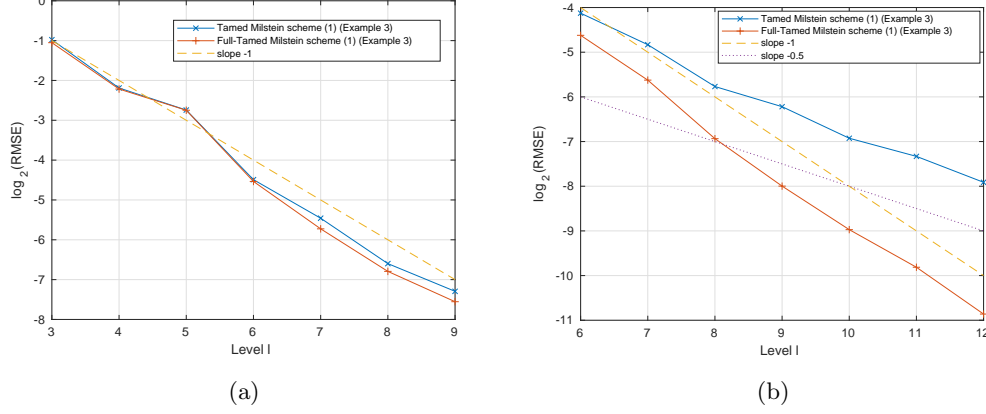


Figure 2: Numerical illustration of the strong convergence in Example 3 for the tamed Milstein scheme (Scheme 1) with and without the  $L$ -derivative terms, with  $N = 20$  particles (left) and  $N = 3$  (right).

Both schemes converge strongly with order 1. Although the full-tamed Milstein scheme shows a slightly better accuracy, we note that this additional gain is not significant taking the complexity of computing the Lévy areas into account.<sup>1</sup> In Fig. 2b we show results for the same test with  $N = 3$ ; also performed for Examples 4 and 5 in Fig. 3a. This illustrates the necessity of implementing the  $L$ -derivative terms to achieve the desired strong convergence order of 1 for fixed  $N$ , since the scheme without the  $L$ -derivative terms only converges with order 1/2. Fig. 3b explores this further by comparing, in the case of

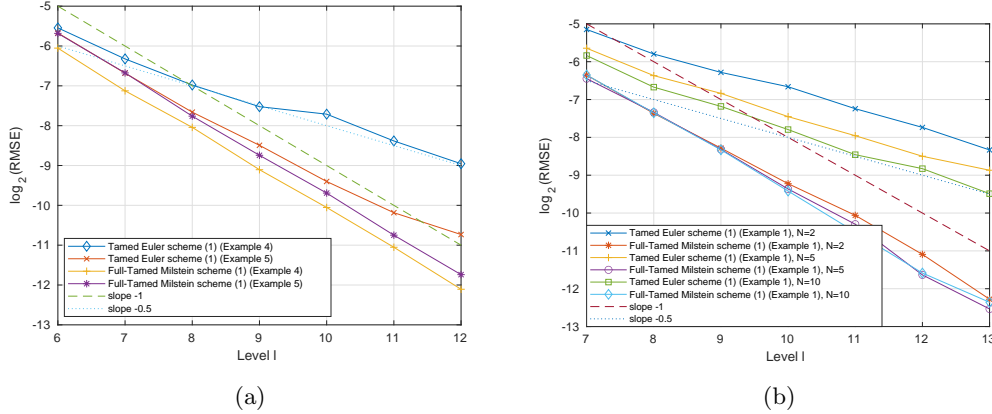


Figure 3: Illustration of the effect of the  $L$ -derivative terms in the tamed Milstein scheme (Scheme 1) for Examples 4 and 5 (left, for  $N = 3$  particles) and Example 1 (right, for three different choices of  $N$ ).

Example 1 (with  $c = 1$ ), the strong convergence rate of a tamed Euler scheme with a full-tamed Milstein scheme including the  $L$ -derivative terms for different small  $N$ .

In Fig. 4a, we compute for a given step size  $\delta = 1/M$  the root-mean-square difference (for several choices of  $N$ ) between a numerical approximation of the particle system associated with the McKean-Vlasov SDE in Example 5, obtained using a tamed Euler scheme (i.e., ignoring the  $L$ -derivative terms)

<sup>1</sup>Due to the high computational complexity, the time discretisation is coarser than the one chosen in Fig. 3b below.

and one obtained by the full-tamed Milstein scheme, at the final time  $T = 1$ . We choose  $M = 2^4, 2^5, 2^6$  for our tests. For a fixed choice of  $M$ , we then compute the RMSE for  $N_l = 2^l$ , for  $l = 2, \dots, 6$ , where

$$\text{RMSE} = \sqrt{\frac{1}{N_l} \sum_{i=1}^{N_l} \left( Y_T^{i, N_l, M} - \tilde{Y}_T^{i, N_l, M} \right)^2},$$

where  $\tilde{Y}_T^{i, N_l, M}$  denotes the numerical approximation using a tamed Euler scheme of  $X$  at time  $T$  computed with  $N_l$  particles and  $M$  time steps, while  $Y_T^{i, N_l, M}$  was computed using a full-tamed Milstein scheme.

We observe a strong convergence rate  $1/2$  with respect to the number of particles. For Examples 1 and 4 similar results were obtained and are therefore omitted.

### 4.3 Propagation of chaos

Fig. 4b depicts the strong propagation of chaos convergence rate for Example 1. For a fixed number of time steps  $M = 2^6$ , we compute for different sizes of the particle system,  $N_l = 2^l$ , for  $l = 3, \dots, 7$ , a RMSE between two particle systems,

$$\text{RMSE} = \sqrt{\frac{1}{N_l} \sum_{i=1}^{N_l} \left( Y_T^{i, N_l, M} - \tilde{Y}_T^{i, N_l, M} \right)^2},$$

where the particle system  $\tilde{Y}^{i, N_l, M}$ ,  $1 \leq i \leq N_l$ , is obtained by splitting the set of Brownian motions driving the particle system  $Y^{i, N_l, M}$ ,  $1 \leq i \leq N_l$ , in two sets and simulating two independent particle systems, each of size  $N_l/2$ :  $(\tilde{Y}^{i, N_l, M, (1)})_{1 \leq i \leq N_l/2}$  is a particle system obtained by using the Brownian motions  $(W^i)_{1 \leq i \leq N_l/2}$  and  $(\tilde{Y}^{i, N_l, M, (2)})_{N_l/2+1 \leq i \leq N_l}$  is obtained from the set  $(W^i)_{N_l/2+1 \leq i \leq N_l}$ . In particular, this means that for these two smaller particle systems only  $N_l/2$  particles are used to approximate the mean-field term. As a time discretisation scheme, we employ a tamed Euler scheme and the full-tamed Milstein scheme, respectively. We observe a strong convergence order of  $1/2$  in terms of number of particles. Corollary 2.6 only proves a strong convergence order of  $1/4$  (see equation (2.9) for  $d = 1$ ). However, recent results (see, e.g., [30, Theorem 2.4]) suggest that the optimal rate is  $1/2$ . These results do not apply to our problem description, as they rely on strong regularity assumptions on the coefficients of the underlying McKean–Vlasov SDE (in particular, the drift needs to be globally Lipschitz continuous in the state component).

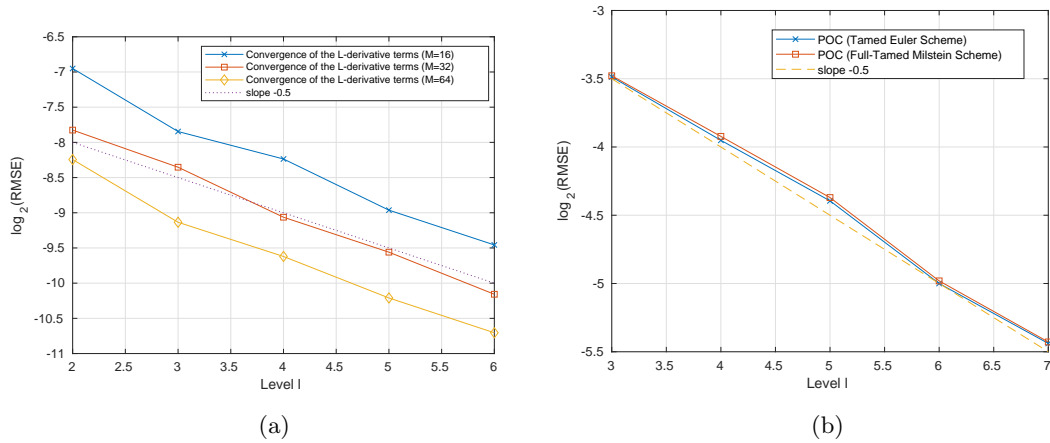


Figure 4: Convergence of the terms involving the  $L$ -derivatives with an increasing number of particles for Example 5 (left). Illustration of the propagation of chaos property (POC) for Example 1 (right).

All figures in this section were generated by the authors' MATLAB code found here [32].

## References

- [1] F. Antonelli and A. Kohatsu-Higa, *Rate of convergence of a particle method to the solution of the McKean–Vlasov equation*, The Annals of Applied Probability, Vol. 12(2), pp. 424-476, 2002.
- [2] J. Baladron, D. Fasoli, O. Faugeras and J. Touboul, *Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons*, The Journal of Mathematical Neuroscience, Vol. 2(10), 2012.
- [3] M. Bauer, T. Meyer-Brandis and F. Proske, *Strong solutions of mean-field stochastic differential equations with irregular drift*, Electronic Journal of Probability, Vol. 23(32), 2018.
- [4] M. Bossy and D. Talay, *A stochastic particle method for the McKean–Vlasov and the Burgers equation*, Mathematics of Computation, Vol. 66(217), pp. 157-192, 1997.
- [5] P. Cardaliaguet, *Notes on mean field games*, P. -L. Lions lectures at Collège de France. Online at <https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf>.
- [6] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I*, vol. 84 of Probability Theory and Stochastic Modelling, Springer International Publishing, 1st ed., 2017.
- [7] J.-F. Chassagneux, D. Crisan and F. Delarue, *A probabilistic approach to classical solutions of the master equation for large population equilibria*, arXiv:1411.3009, 2014.
- [8] W. Fang and M. Giles, *Adaptive Euler–Maruyama method for SDEs with nonglobally Lipschitz drift*, Annals of Applied Probability, Vol. 30(2), pp. 526-560, 2020.
- [9] G. Flint and T. Lyons, *Pathwise approximation of SDEs by coupling piecewise Abelian rough paths*, arXiv:1505.01298, 2015.
- [10] J. Foster, T. Lyons and H. Oberhauser, *An optimal polynomial approximation of Brownian motion*, SIAM Journal on Numerical Analysis, Vol. 58(3), pp. 1393-1421, 2020.
- [11] S. Gan and X. Wang, *The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients*, Journal of Difference Equations and Applications, Vol. 19(3), pp. 466-490, 2013.
- [12] D. J. Higham, X. Mao and A. M. Stuart, *Strong convergence of Euler-type methods for nonlinear stochastic differential equations*, SIAM Journal of Numerical Analysis, Vol. 40(3), pp. 1041-1063, 2002.
- [13] M. Hutzenhaler, A. Jentzen and P. E. Kloeden, *Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients*, The Annals of Applied Probability, Vol. 22(4), pp. 1611-1641, 2012.
- [14] M. Hutzenhaler, A. Jentzen and P. E. Kloeden, *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, pp. 1563-1576, 2011.
- [15] E. F. Keller and L. A. Segel, *Initiation of slime mold aggregation viewed as an instability*. Journal of Theoretical Biology, Vol. 26(3), pp. 399-415, 1970.
- [16] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.
- [17] C. Kumar and Neelima, *On explicit Milstein-type scheme for McKean–Vlasov stochastic differential equations with super-linear drift coefficient*, arXiv:2004.01266, 2020.
- [18] C. Kumar, Neelima, C. Reisinger and W. Stockinger, *Well-posedness and tamed schemes for McKean–Vlasov equations with common noise*, arXiv:2006.00463, 2020.



- [19] C. Kumar and S. Sabanis, *On Milstein approximations with varying coefficients: the case of super-linear diffusion coefficients*, BIT Numerical Mathematics, Vol. 59, pp. 929-968, 2019.
- [20] X. Mao, *The truncated Euler–Maruyama method for stochastic differential equations*, Journal of Computational and Applied Mathematics, Vol. 290, pp. 370-384, 2015.
- [21] H. McKean, *A class of Markov processes associated with nonlinear parabolic equations*, Proceedings of the National Academy of Sciences of the USA, Vol. 56(6), pp. 1907-1911, 1966.
- [22] Y. S. Mishura and A. Y. Veretennikov, *Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations*, arXiv:1603.02212, 2018.
- [23] C. S. Patlak, *Random walk with persistence and external bias*, The Bulletin of Mathematical Biophysics, Vol. 15(3), pp. 311-338, 1953.
- [24] G. d. Reis, S. Engelhardt and G. Smith, *Simulation of McKean–Vlasov SDEs with super linear drift*, arXiv:1808.05530, 2018.
- [25] G. d. Reis, W. Salkeld and J. Tugaut, *Freidlin–Wentzell LDPs in path space for McKean–Vlasov equations and the functional iterated logarithm law*, The Annals of Applied Probability, Vol. 29(3), 2019.
- [26] C. Reisinger and W. Stockinger, *An adaptive Euler–Maruyama scheme for McKean–Vlasov SDEs with super-linear growth and application to the mean-field FitzHugh–Nagumo model*, arXiv:2005.06034, 2020.
- [27] S. Sabanis, *A note on tamed Euler approximations*, Electronic Communications in Probability, Vol. 18(47), pp. 1-10, 2013.
- [28] S. Sabanis, *Euler approximation with varying coefficients: The case of superlinearly growing diffusion coefficients*, The Annals of Applied Probability, Vol. 26(4), pp. 2083-2105, 2016.
- [29] A. S. Sznitman, *Topics in Propagation of Chaos*, Ecole d’été de probabilités de Saint-Flour XIX - 1989, vol. 1464 of Lecture notes in Mathematics, Springer-Verlag, 1991.
- [30] L. Szpruch and A. Tse, *Antithetic multilevel particle system sampling method for McKean–Vlasov SDEs*, arXiv:1903.07063v2, 2019.
- [31] M. Wiktorsson, *Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple Brownian motions*, The Annals of Applied Probability, Vol. 11(2), pp. 470-487, 2001.
- [32] J. Bao, C. Reisinger, P. Ren and W. Stockinger, *Code used in Section 4* <https://github.com/Wstockinger/Milstein>, 2020.

**Acknowledgement:** Jianhai Bao is supported by National Natural Science Foundation of China (11801406). Wolfgang Stockinger is supported by an Upper Austrian Government fund.