Available online at www.sciencedirect.com

ScienceDirect

Journal of Differential Equations 440 (2025) 113445

Journal of Differential Equations

www.elsevier.com/locate/jde

Uniform-in-time estimates for mean-field type SDEs and applications

Jianhai Bao, Jiaqing Hao *

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

Received 8 July 2024; revised 7 March 2025; accepted 11 May 2025

Abstract

Via constructing an asymptotic coupling by reflection, in this paper we establish uniform-in-time estimates on probability distributions for mean-field type SDEs, where the drift terms under consideration are dissipative merely in the long distance. As applications, we (i) explore the long time estimate on probability distributions associated with an SDE and its delay version; (ii) investigate the issue on uniform-in-time propagation of chaos for McKean-Vlasov SDEs, where the drifts might be singular with respect to the spatial variables; (iii) tackle the discretization error bounds in an infinite-time horizon for stochastic algorithms (e.g. backward/tamed/adaptive Euler-Maruyama schemes as three typical candidates) associated with McKean-Vlasov SDEs.

MSC: 60H10; 60F05; 60F15

Keywords: Asymptotic coupling by reflection; Mean-field type SDE; Propagation of chaos; Stochastic algorithm; Uniform-in-time estimate

* Corresponding author.
E-mail addresses: jianhaibao@tju.edu.cn (J. Bao), hjq_0227@tju.edu.cn (J. Hao).

https://doi.org/10.1016/j.jde.2025.113445

1. Introduction, main result and applications

1.1. Introduction

In this paper, we work on the mean-field type SDEs: for $i \in \mathbb{S}_N := \{1, \cdots, N\}$,

$$\mathrm{d}X_t^i = b(X_t^i, \mathscr{L}_{X_t^i})\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^i + \sigma_0(X_t^i)\,\mathrm{d}B_t^i \tag{1.1}$$

and

$$\mathrm{d}X_t^{i,N} = \widetilde{b}(X_{\theta_t}^{i,N}, \widetilde{\mu}_{\overline{\theta}_t}^N)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^i + \sigma_0(X_t^{i,N})\,\mathrm{d}B_t^i \tag{1.2}$$

with the initial datum $(X_{[-r_0,0]}^{1,N}, \cdots, X_{[-r_0,0]}^{N,N})$ for some $r_0 \geq 0$, where, for each $i \in \mathbb{S}_N$, $X_{[-r_0,0]}^{i,N}$ is a $C([-r_0, 0]; \mathbb{R}^d)$-valued random variable. In (1.1) and (1.2), $b, \widetilde{b} : \mathbb{R}^d \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}^d$ with $\mathscr{P}(\mathbb{R}^d)$ being the set of probability measures on $\mathbb{R}^d$; $\sigma \in \mathbb{R}$ and $\sigma_0 : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^m$; $\mathscr{L}_{X_t^i}$ stands for the law of X_t^i; the maps $\theta : [0, \infty) \to [-r_0, \infty)$ and $\overline{\theta} : [0, \infty) \to [0, \infty)$, where concrete expressions will be specified later concerning respective settings; $\widetilde{\mu}_t^N := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}$ means the empirical measure associated with particles $X_t^{1,N}, \cdots, X_t^{N,N}$; $W^1 = (W_t^1)_{t\geq 0}, \cdots, W^N = (W_t^N)_{t\geq 0}$ (resp. $B^1 = (B_t^1)_{t\geq 0}, \cdots, B^N = (B_t^N)_{t\geq 0}$) are mutually independent d-dimensional (resp. m-dimensional) Brownian motions supported on the same filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$; Furthermore, $(W^1, \cdots, W^N)$ is supposed to be independent of $(B^1, \cdots, B^N)$.

Regarding the objects $(X_t^1, \cdots, X_t^N)_{t\geq 0}$ and $(X_t^{1,N}, \cdots, X_t^{N,N})_{t\geq 0}$ solving (1.1) and (1.2), respectively, the central goal in the present paper is to establish the quantitative estimate:

$$\mathbb{W}_1\big(\mathscr{L}_{X_t^i}, \mathscr{L}_{X_t^{i,N}}\big) \leq \varphi(t, N), \quad t \geq 0,\ i \in \mathbb{S}_N, \tag{1.3}$$

where $\mathbb{W}_1$ denotes the L^1-Wasserstein distance and $\varphi : [0, \infty) \times [0, \infty) \to (0, \infty)$ is a decreasing function with respect to the first variable and the second argument, respectively. For the explicit form of φ, we would like to refer to (1.20) below for more details. Hereinafter, we attempt to elaborate why we focus on the framework (1.1) and (1.2), and explore the uniform-in-time estimate (1.3). The interpretations will be expounded based on the following three perspectives.

1.1.1. Uniform-in-time propagation of chaos

In (1.2), once we take $\widetilde{b} = b$, $r_0 = 0$ and $\theta_t = \overline{\theta}_t = t$, (1.2) subsequently becomes

$$\mathrm{d}X_t^{i,N} = b(X_t^{i,N}, \widetilde{\mu}_t^N)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^i + \sigma_0(X_t^{i,N})\,\mathrm{d}B_t^i. \tag{1.4}$$

As we know, (1.1) and (1.4) are the respective non-interacting particle system and interacting particle system corresponding to the following McKean-Vlasov SDE:

$$\mathrm{d}X_t = b(X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma\,\mathrm{d}W_t + \sigma_0(X_t)\,\mathrm{d}B_t, \tag{1.5}$$

where $(W_t)_{t\geq 0}$ is a d-dimensional Brownian motion (a copy of $(W_t^i)_{t\geq 0}$ for each $i \in \mathbb{S}_N$), which is independent of the m-dimensional Brownian motion $(B_t)_{t\geq 0}$ (a copy of $(B_t^i)_{t\geq 0}$ for each $i \in$

$\mathbb{S}_N$). Since the landmark work [29] due to Sznitman, the theory on propagation of chaos in a finite-time horizon has achieved great advancements for various scenarios; see, for instance, [5,14] for McKean-Vlasov SDEs with regular coefficients, and [1] regarding McKean-Vlasov SDEs with irregular drifts or diffusions. Recently, for weakly interacting mean-field particle systems with possibly non-convex confinement and interaction potentials, the uniform-in-time propagation of chaos: for some constants $c, \lambda > 0$ (independent of $t > 0$ and $i \in \mathbb{S}_N$),

$$\mathbb{W}_1\left(\mathscr{L}_{X_t^{i,\mu}}, \mathscr{L}_{X_t^{i,N,\nu}}\right) \le c\left(\mathrm{e}^{-\lambda t}\mathbb{W}_1(\mu,\nu) + N^{-\frac{1}{2}}\right), \quad t > 0,\ i \in \mathbb{S}_N \tag{1.6}$$

was established in the remarkable work [7], where $\mathscr{L}_{X_t^{i,\mu}}$ and $\mathscr{L}_{X_t^{i,N,\nu}}$ stand respectively for the distributions of X_t^i and $X_t^{i,N}$ with $\mathscr{L}_{X_0^i} = \mu$ and $\mathscr{L}_{X_0^{i,N}} = \nu$. For great progress on the uniform-in-time propagation of chaos concerning Langevin dynamics with regular potentials and stochastic particle systems with mean-field singular interactions, we refer to [11,12,27] and references within. As an immediate by-product of the quantitative estimate (see Theorem 1.3 below) derived in this paper, the uniform-in-time propagation of chaos (1.6) will be reproduced right away, where the underlying drift terms might be singular with respect to the spatial variables (see Corollary 1.5 for more details). The proceeding explanations can be viewed as one of our motivations to consider (1.1) and (1.2), and study the estimate (1.3).

1.1.2. *Uniform-in-time probability distance between an SDE and its delay version*

Consider a semi-linear SDE:

$$\mathrm{d}X_t = \beta(\alpha - X_t)\mathrm{d}t + \sigma\,\mathrm{d}W_t + \sigma_0(X_t)\,\mathrm{d}B_t, \tag{1.7}$$

where $\alpha \in \mathbb{R}^d$, $\beta > 0$, $\sigma \in \mathbb{R}$, $\sigma_0 : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$, $(W_t)_{t\ge0}$ and $(B_t)_{t\ge0}$ are mutually independent d-dimensional Brownian motions. In case of $\sigma_0(x) \equiv \mathbf{0}_{d\times d}$ ($d \times d$ zero matrix), (1.7) is a linear SDE solved by the Ornstein–Uhlenbeck (O-U for abbreviation) process. As we know, the O-U process has been applied considerably in financial mathematics and the other related research fields. Whilst, in the real world, the price of an asset or the evolution of population dynamics is influenced inevitably by major events that took place. In turn, the viewpoint above motivates us to consider a memory-dependent version of (1.7) which is described as follows:

$$\mathrm{d}Y_t = \beta(\alpha - Y_{t-r_0})\mathrm{d}t + \sigma\,\mathrm{d}W_t + \sigma_0(X_t)\,\mathrm{d}B_t, \quad t > 0 \tag{1.8}$$

with the initial datum $Y_{[-r_0,0]}$. In (1.8), $(\alpha, \beta, (W_t)_{t\ge0}, (B_t)_{t\ge0})$ is kept untouched as in (1.7), and the positive r_0 is the length of the time lag. Apparently, (1.7) and (1.8) are fit into the framework (1.1) and (1.2) by setting $N = 1$, $W_t^1 = W_t$, and $B_t^1 = B_t$, and taking $\theta_t = t - r_0$ and $\widetilde{b} = b$. Indeed, the quantity r_0 can be regarded as a perturbation. Intuitively speaking, the probability distance between $\mathscr{L}_{X_t}$ and $\mathscr{L}_{Y_t}$ with $\mathscr{L}_{X_0} = \mathscr{L}_{Y_0}$ should be very small in case that the perturbation intensity r_0 is tiny. So, the issue on how to quantify the probability distance between $\mathscr{L}_{X_t}$ and $\mathscr{L}_{Y_t}$ encourages us to pursue the topic (1.3). The above can be regarded as another inspiration to implement the present work.

1.1.3. *Uniform-in-time discretization error bounds for stochastic algorithms*

Our third motivation arises from the long time analysis on stochastic algorithms for McKean-Vlasov SDEs, where the drifts need not to be uniformly dissipative with respect to the spatial variables. As is known to all, the Euler-Maruyama (EM for short) scheme is the simplest and succinctest method to discretize the McKean-Vlasov SDE (1.5) with $\sigma_0(x) \equiv \mathbf{0}_{d\times d}$, that is,

$$\mathrm{d}X_t = b(X_t, \mathscr{L}_{X_t})\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad t > 0. \tag{1.9}$$

Nonetheless, the EM scheme works merely for SDEs with coefficients of linear growth; see, for instance, [16, Theorem 2.1] and [20, Lemma 6.3] for a theoretical support and a counterexample, respectively. Based on this, plenty of variants of the EM scheme were proposed to cope with numerical approximations for SDEs with non-globally Lipschitz continuous coefficients. The finite-time strong convergence of the backward EM scheme, as a typical candidate of EM's variant, related to the McKean-Vlasov SDE (1.9): for a step size $\delta > 0$,

$$\mathrm{d}X_t^{\delta,i,N} = b(X_{t_\delta+\delta}^{\delta,i,N}, \widetilde{\mu}_{t_\delta}^{\delta,N})\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^i, \quad t > 0,\ i \in \mathbb{S}_N \tag{1.10}$$

was explored in [6], where $t_\delta := \lfloor t/\delta \rfloor \delta$ with $\lfloor t/\delta \rfloor$ being the integer part of t/δ, and $\widetilde{\mu}_t^{\delta,N} := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{\delta,i,N}}$. Transparently, (1.9) and (1.10) with $r_0 = 0$ and $\sigma_0(x) \equiv \mathbf{0}_{d\times d}$ are included in (1.1) and (1.2) by taking $\widetilde{b} = b$, $\theta_t = t_\delta + \delta$, and $\overline{\theta}_t = t_\delta$, separately.

Next, inspired by e.g. [15,26], [6] put forward the tamed EM scheme: for $\kappa \in (0, 1/2]$,

$$\mathrm{d}X_t^{\delta,i,N} = \frac{b(X_{t_\delta}^{\delta,i,N}, \widetilde{\mu}_{t_\delta}^{\delta,N})}{1 + \delta^\kappa |b(X_{t_\delta}^{\delta,i,N}, \widetilde{\mu}_{t_\delta}^{\delta,N})|}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^i, \quad t > 0,\ i \in \mathbb{S}_N \tag{1.11}$$

to simulate the McKean-Vlasov SDE (1.9) in a finite time interval. Since, for a fixed step size $\delta > 0$, the modified drift is uniformly bounded, the distribution of $(X_t^{\delta,i,N})_{t\geq 0}$ solving the tamed EM scheme (1.11) is not adequate to approximate the distribution of $(X_t)_{t\geq 0}$ determined by (1.9) in an infinite-time horizon. Enlightened by e.g. [4,18,19], to derive a uniform-in-time estimate between the distributions of the exact solution and the numerical counterpart, we construct the following tamed EM scheme for the McKean-Vlasov SDE (1.9):

$$\mathrm{d}X_t^{\delta,i,N} = \frac{b(X_{t_\delta}^{\delta,i,N}, \widetilde{\mu}_{t_\delta}^{\delta,N})}{1 + \delta^\kappa \|\nabla b(X_{t_\delta}^{\delta,i,N}, \widetilde{\mu}_{t_\delta}^{\delta,N})\|_{\mathrm{HS}}}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^i, \quad t > 0,\ i \in \mathbb{S}_N, \tag{1.12}$$

where, for each fixed $\mu \in \mathscr{P}(\mathbb{R}^d)$, $x \mapsto b(x,\mu)$ is a C^1-function, ∇ means the weak gradient operator with respect to the spatial variables, and $\|\cdot\|_{\mathrm{HS}}$ stipulates the Hilbert-Schmidt norm. Compared (1.12) with (1.11), the tamed drift in (1.12) might not be bounded any more and is at most of linear growth with respect to the spatial variables. Obviously, (1.1) and (1.2) with $r_0 = 0$ and $\sigma_0(x) \equiv \mathbf{0}_{d\times d}$ can cover (1.9) and (1.12) once we choose $\theta_t = \overline{\theta}_t = t_\delta$ and set

$$\widetilde{b}(x,\mu) := \frac{b(x,\mu)}{1 + \delta^\kappa \|\nabla b(x,\mu)\|_{\mathrm{HS}}}.$$

No matter what the backward EM scheme (1.10) or the tamed EM algorithm (1.12), the underlying time step is a uniform constant. In [9], a refined EM scheme with an adaptive step size was initiated to approximate SDEs with super-linear drifts. In the spirit of [9], [24] constructed an adaptive EM scheme associated with (1.9), which is described as below:

$$X_{t_{n+1}}^{\delta,i,N} = X_{t_n}^{\delta,i,N} + b\big(X_{t_n}^{\delta,i,N}, \widetilde{\mu}_{t_n}^{\delta,N}\big)h_n^\delta + \sigma \triangle W_{t_n}^i, \quad n \geq 0,\ i \in \mathbb{S}_N, \tag{1.13}$$

where $t_{n+1} := t_n + h_n^\delta$ for an adaptive time step h_n^δ (see (1.33) below for an alternative of h_n^δ), and $\triangle W_{t_n}^i := W_{t_{n+1}}^i - W_{t_n}^i$. In contrast to (1.10) and (1.12), the time step in (1.13) is not a constant any more but an adaptive process, which is determined by the current approximate solution. Let $\underline{t} = \max\{t_n : t_n \leq t\}$. Then, the continuous version of (1.13) can be formulated as

$$\mathrm{d}X_t^{\delta,i,N} = b\big(X_{\underline{t}}^{\delta,i,N}, \widetilde{\mu}_{\underline{t}}^{\delta,N}\big)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^i, \quad t > 0,\ i \in \mathbb{S}_N. \tag{1.14}$$

Therefore, (1.9) and (1.14) can be incorporated into the framework (1.1) and (1.2) by setting $\overline{\theta}_t = \theta_t = \underline{t}$, $\sigma_0(x) \equiv \mathbf{0}_{d\times d}$, and $b = \widetilde{b}$.

With regard to the backward/tamed/adaptive EM scheme for classical SDEs and McKean-Vlasov SDEs, there is a huge amount of literature concerned with strong/weak convergence in a finite-time interval; see [9,13,17,24,26], to name just a few. Meanwhile, there are still plenty of work handling long time behavior of numerical algorithms when (McKean-Vlasov) SDEs involved are uniformly dissipative with respect to the spatial variables; see e.g. [3,9,21,32,33] and references therein. In the aforementioned papers, the synchronous coupling was employed to analyze the convergence property (in an infinite-time horizon) of the underlying algorithms. Whereas, such an approach does not work any more to deal with the long time behavior of stochastic algorithms when (McKean-Vlasov) SDEs under investigation are not globally dissipative. In the present work, as another direct application of the main result (see Theorem 1.3), concerning McKean-Vlasov SDEs, we quantify the uniform-in-time estimate on the distribution distance between laws of exact solutions and numerical solutions derived via backward/tamed/adaptive EM schemes. Once more, the elaborations above urge us to work on the frameworks (1.1) and (1.2), and conduct a further study on (1.3).

1.2. Main result

Below, we assume that for $x \in \mathbb{R}^d$ and $\mu \in \mathscr{P}(\mathbb{R}^d)$,

$$b(x,\mu) = b_1(x) + (b_0 * \mu)(x) \quad \text{with} \quad (b_0 * \mu)(x) := \int_{\mathbb{R}^d} b_0(x-y)\,\mu(\mathrm{d}y). \tag{1.15}$$

For such setting, we shall assume that the corresponding SDEs (1.1) and (1.2) are strongly well-posed in order to establish a much more general result (i.e., Theorem 1.3). In the sequel, for SDEs under consideration, we shall present explicit conditions on the coefficients to guarantee the strong well-posedness. We further suppose that

($\mathbf{A}_1$) $b_1(x)$ is continuous and locally bounded in $\mathbb{R}^d$ and there exist constants $\ell_0 \geq 0$ and $\lambda > 0$ such that for $x, y \in \mathbb{R}^d$,

$$\langle x-y, b_1(x) - b_1(y)\rangle \leq |x-y|\phi(|x-y|)\mathbb{1}_{\{|x-y|\leq \ell_0\}} - \lambda|x-y|^2\mathbb{1}_{\{|x-y|>\ell_0\}}, \tag{1.16}$$

where $\phi : [0,\infty) \to [0,\infty)$ with $\phi(0) = 0$ is increasing and continuous. Moreover, there exists a constant $K > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$|b_0(x) - b_0(y)| \leq K|x-y|. \tag{1.17}$$

($\mathbf{A}_2$) $\sigma \neq 0$, and there is a constant $L > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$\|\sigma_0(x) - \sigma_0(y)\|_{\mathrm{HS}}^2 \le L|x-y|^2. \tag{1.18}$$

To proceed, we make some comments on Assumptions ($\mathbf{A}_1$) and ($\mathbf{A}_2$).

Remark 1.1. In terms of (1.16), b_1 is dissipative merely in the long range. Particularly, as revealed in Corollary 1.5 below, (1.16) allows the drift term b_1 to be singular. For example, $b_1(x) = \overline{b}_1(x) + \widetilde{b}_1(x)$, in which $\widetilde{b}_1$ is Dini continuous (see Assumption (**H**) below for details), and there exist constants $\lambda_1', \lambda_2', \ell_0' > 0$ such that for any $x, y \in \mathbb{R}^d$,

$$\langle x-y, \overline{b}_1(x) - \overline{b}_1(y)\rangle \le \lambda_1'|x-y|^2 \mathbb{1}_{\{|x-y|\le \ell_0'\}} - \lambda_2'|x-y|^2 \mathbb{1}_{\{|x-y|>\ell_0'\}}.$$

Concerned with the diffusion terms, the additive part is set to be non-degenerate (which plays a crucial role in constructing the asymptotic coupling by reflection), and the multiplicative counterpart might be degenerate. Herein, we would like to stress that the additive intensity σ considered in the present work is a non-zero constant in lieu of a non-degenerate matrix to write merely the prerequisite (1.16) and the asymptotic coupling by reflection (see (1.19) below for more details) in a simple way.

For classical SDEs (with the same drifts and diffusions), it is enough to construct the reflection coupling before the coupling time since two SDEs will merge together afterwards due to strong well-posedness (provided it exists). However, as far as two SDEs with different coefficients are concerned, the coupled processes can diverge once again even though they meet at the coupling time. Therefore, the classical reflection coupling approach no longer works to estimate the probability distance between laws of solutions corresponding to SDEs with different coefficients.

Inspired by [31], where gradient/Hölder estimates as well as the exponential convergence were derived for nonlinear monotone SPDEs, we shall design an asymptotic coupling by reflection to achieve the qualitative estimate (1.3). To describe the asymptotic coupling by reflection, we need to introduce some additional notations. For $\varepsilon \ge 0$, let $h_\varepsilon : [0,\infty) \to [0,1]$ be a C^1-function satisfying

$$h_\varepsilon(r) = \begin{cases} 0, & 0 \le r \le \varepsilon, \\ 1, & r \ge 2\varepsilon, \end{cases}$$

and $h_\varepsilon^* : [0,\infty) \to [0,1]$ be defined by $h_\varepsilon^*(r) = \sqrt{1 - h_\varepsilon^2(r)}$, $r \ge 0$. Set

$$\Pi(x) := I_{d\times d} - 2\mathbf{e}(x) \otimes \mathbf{e}(x), \qquad x \in \mathbb{R}^d,$$

where $I_{d\times d}$ is the $d \times d$ identity matrix, $\mathbf{e}(x) := \frac{x}{|x|}\mathbb{1}_{\{x \neq \mathbf{0}\}}$, and $\mathbf{e}(x) \otimes \mathbf{e}(x)$ means the tensor between $\mathbf{e}(x)$ and $\mathbf{e}(x)$. Thus, Π defined above is an orthogonal matrix. Furthermore, we shall assume that $W^{1,1} := (W_t^{1,1})_{t\ge0}, \cdots, W^{1,N} := (W_t^{1,N})_{t\ge0}$ (resp. $W^{2,1} := (W_t^{2,1})_{t\ge0}, \cdots, W^{2,N} := (W_t^{2,N})_{t\ge0}$) are mutually independent d-dimensional (resp. m-dimensional) Brownian motions carried on the same probability space as that of $B^1, \cdots, B^N$. In addition, we suppose that $(W^{1,1}, \cdots, W^{1,N})$, $(W^{2,1}, \cdots, W^{2,N})$ and $(B^1, \cdots, B^N)$ are mutually independent.

With the proceeding notations at hand, we can write down the asymptotic coupling by reflection associated with (1.1) and (1.2). More precisely, we consider the following coupled interacting particle system: for any $i \in \mathbb{S}_N$ and $t > 0$,

$$\begin{cases} \mathrm{d}Y_t^{i,\varepsilon} = b(Y_t^{i,\varepsilon}, \widehat{\mu}_t^{i,\varepsilon})\,\mathrm{d}t + \sigma h_\varepsilon(|Z_t^{i,N,\varepsilon}|)\mathrm{d}W_t^{1,i} + \sigma h_\varepsilon^*(|Z_t^{i,N,\varepsilon}|)\mathrm{d}W_t^{2,i} + \sigma_0(Y_t^{i,\varepsilon})\,\mathrm{d}B_t^i, \\ \mathrm{d}Y_t^{i,N,\varepsilon} = \widetilde{b}(Y_{\theta_t}^{i,N,\varepsilon}, \widetilde{\mu}_{\overline{\theta}_t}^{N,\varepsilon})\,\mathrm{d}t + \sigma \Pi(Z_t^{i,N,\varepsilon}) h_\varepsilon(|Z_t^{i,N,\varepsilon}|)\mathrm{d}W_t^{1,i} \\ \qquad\qquad + \sigma h_\varepsilon^*(|Z_t^{i,N,\varepsilon}|)\mathrm{d}W_t^{2,i} + \sigma_0(Y_t^{i,N,\varepsilon})\,\mathrm{d}B_t^i \end{cases} \tag{1.19}$$

with the initial condition $\big(Y_0^{i,\varepsilon}, Y_{[-r_0,0]}^{i,N,\varepsilon}\big)_{i\in\mathbb{S}_N} = \big(X_0^i, X_{[-r_0,0]}^{i,N}\big)_{i\in\mathbb{S}_N}$, which are i.i.d. random variables. In (1.19), the quantities $\widehat{\mu}_t^{i,\varepsilon}$, $Z_t^{i,N,\varepsilon}$, and $\widetilde{\mu}_t^{N,\varepsilon}$ are defined respectively by

$$\widehat{\mu}_t^{i,\varepsilon} = \mathscr{L}_{Y_t^{i,\varepsilon}}, \quad Z_t^{i,N,\varepsilon} = Y_t^{i,\varepsilon} - Y_t^{i,N,\varepsilon} \quad \text{and} \quad \widetilde{\mu}_t^{N,\varepsilon} = \frac{1}{N}\sum_{j=1}^N \delta_{Y_t^{j,N,\varepsilon}}.$$

Now, we present some comments on the coupling constructed in (1.19).

Remark 1.2. Note obviously that the noise in (1.1) includes two parts, namely, the additive part and the multiplicative part. In terms of (1.19), for the additive part, which is also non-degenerate, we adopt the asymptotic coupling by reflection; Whereas, for the multiplicative part (might be degenerate), we employ the synchronous coupling, which, in literature, is also named as the coupling of marching soldiers. Moreover, we would like to emphasize that, for the construction of the asymptotic coupling by reflection, the drift term b can be much more general rather than the form in (1.15) as demonstrated in Lemma 2.1.

Furthermore, for the notational brevity, we set for any $t \geq 0$,

$$\mathbf{X}_t^N := \big(X_t^1, \cdots, X_t^N\big), \quad \mathbf{X}_t^{N,N} := \big(X_t^{1,N}, \cdots, X_t^{N,N}\big)$$

and

$$\mathbf{Y}_t^{N,\varepsilon} := \big(Y_t^{1,\varepsilon}, \cdots, Y_t^{N,\varepsilon}\big), \quad \mathbf{Y}_t^{N,N,\varepsilon} := \big(Y_t^{1,N,\varepsilon}, \cdots, Y_t^{N,N,\varepsilon}\big).$$

As claimed in Lemma 2.1 below, for any $\varepsilon > 0$, $(\mathbf{Y}_t^{N,\varepsilon}, \mathbf{Y}_t^{N,N,\varepsilon})_{t\geq 0}$ is a coupling of $(\mathbf{X}_t^N, \mathbf{X}_t^{N,N})_{t\geq 0}$. Additionally, for any $t \geq 0$, $\mu \in \mathscr{P}(\mathbb{R}^d)$, and $\nu \in \mathscr{P}(\mathscr{C})$ with $\mathscr{C} := C([-r_0,0];\mathbb{R}^d)$ (i.e., the set of continuous $\mathbb{R}^d$-valued functions on $[-r_0,0]$), denote μ_t^i and $\nu_t^{i,N}$ by the laws of X_t^i and $X_t^{i,N}$ with $\mathscr{L}_{X_0^i} = \mu$ and $\mathscr{L}_{X_{[-r_0,0]}^{i,N}} = \nu$, separately.

With the aid of previous preliminaries, the quantitative estimate (1.3) can be portrayed precisely as stated in the following theorem.

Theorem 1.3. *Assume Assumptions* $(\mathbf{A}_1)$ *and* $(\mathbf{A}_2)$. *Then, there are constants* $C, K^*, L^*, \lambda^* > 0$ *such that for any* $K \in [0, K^*]$, $L \in (0, L^*]$, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, *and* $\nu \in \mathscr{P}_1(\mathscr{C})$,

$$\mathbb{W}_1(\mu_t^i,\nu_t^{i,N})\le C\bigg(\mathrm{e}^{-\lambda^* t}\mathbb{W}_1(\mu,\nu_0)+N^{-\frac{1}{2}}\mathbb{1}_{\{K>0\}}+\int_0^t \mathrm{e}^{-\lambda^*(t-s)}\mathbb{E}\big|b(X_s^{i,N},\widetilde{\mu}_s^N)-\widetilde{b}(X_{\theta_s}^{i,N},\widetilde{\mu}_{\theta_s}^N)\big|\,\mathrm{d}s\bigg),\quad t>0,\quad i\in\mathbb{S}_N, \tag{1.20}$$

where $\nu_0(\mathrm{d}x):=\nu(\{\eta\in\mathscr{C}\}:\eta_0\in\mathrm{d}x)$.

Before we proceed, let's make some remarks on Theorem 1.3.

Remark 1.4. *Error bounds.* From (1.20), it is easy to see that $\mathbb{W}_1(\mu_t^i,\nu_t^{i,N})$ is dominated by three terms, where the first term is concerned with the L^1-Wasserstein distance between the initial (projection) distributions with the decay prefactor $\mathrm{e}^{-\lambda^* t}$, the second one is related to the decay rate with respect to the particle number, and the third part involves the error among b and $\widetilde{b}$, which, in particular, embodies the dependence of the initial segment and the length of time lag when (1.2) is an SDE with memory; see Theorem 1.6 below for more details. At first sight, the right hand side of (1.20) is not elegant since the third term in the big parenthesis is not explicit. Nevertheless, the third term is much more tractable for applications we shall carry out.

Initial moments. When the drift term is written in the form (1.15) and the associated initial distribution has a finite second-order moment, Theorem 1.3 shows that the decay speed of $\mathbb{W}_1(\mu_t^i,\nu_t^{i,N})$ with respect to the particle number is $N^{-\frac{1}{2}}$. In some scenarios, the drift terms can be allowed to be much more general so the initial distribution necessitates merely a finite lower-order moment. Whereas, for this setting, the decay rate of $\mathbb{W}_1(\mu_t^i,\nu_t^{i,N})$ with respect to the particle number will be dependent on the dimension d and become dramatically worse when, in particular, (2.16) below is tackled by taking advantage of [10, Theorem 1]. Therefore, in the present work, we prefer the former framework rather than the latter one.

Coupling construction. It is worthy to point out that, in [28], another kind of asymptotic coupling by reflection (which was called an approximate reflection coupling therein) was deployed to investigate bounds on the discretization error for Langevin dynamics, where the potential term is a C^1-function and is of linear growth. Compared the asymptotic coupling by reflection in [28] with (1.19), we find that the weak limit process of the coupled process constructed in [28] is a coupling process while, for any $\varepsilon>0$, the coupled process $(\mathbf{Y}_t^{N,\varepsilon},\mathbf{Y}_t^{N,N,\varepsilon})_{t\ge 0}$ determined by (1.19) is a coupling process we desire. It is also worthy to emphasize that in [28] a series of work on tightness need to be implemented in order to examine that the associated weak limit process is a coupling process. Therefore, the asymptotic coupling by reflection built in (1.19) has its own advantages.

Noise terms. It seems to be slightly weird that the noise term in (1.1) encompasses two parts (i.e., the additive part and the multiplicative counterpart). Nevertheless, as long as the diffusive term of the non-interacting particle system under investigation is multiplicative and non-degenerate, we can adopt the noise-decomposition trick (see e.g. [22]) so it can be decomposed equally in the sense of distribution into the format of (1.1). Based on point of view above, the framework (1.1) can make sense very well.

Lipschitz constants. In Theorem 1.3, the constants K^* and L^* are the respective upper bounds of Lipschitz constants concerning b_0 and σ_0. Via a close inspection of the proof for Theorem 1.3, the explicit forms concerning K^* and L^* can be tracked. As far as McKean-Vlasov SDEs are concerned, the Lipschitz constant K is generally small otherwise the phase phenomenon can occur. Furthermore, the multiplicative intensity σ_0 is regarded as a perturbation of σ so the corresponding Lipschitz constant L should also be small provided that one wants to handle through the asymptotic coupling by reflection the uniform-in-time estimate for SDEs with partially dissipative drifts.

As an immediate by-product of Theorem 1.3, we present the following statement, which is concerned with uniform-in-time propagation of chaos for McKean-Vlasov SDEs, where one part of the drifts might be singular in the spatial variables.

Corollary 1.5. *Assume Assumptions* $(\mathbf{A}_1)$ *and* $(\mathbf{A}_2)$*. Then, there are constants* $C, K^*, L^*, \lambda^* > 0$ *such that for any* $K \in (0, K^*]$*,* $L \in [0, L^*]$*,* $\mu \in \mathscr{P}_2(\mathbb{R}^d)$*,* $\nu \in \mathscr{P}_1(\mathbb{R}^d)$*, and* $t \geq 0$*,*

$$\max_{i\in\mathbb{S}_N} \mathbb{W}_1(\mu_t^i, \nu_t^{i,N}) \leq C\big(\mathrm{e}^{-\lambda^* t}\,\mathbb{W}_1(\mu,\nu) + N^{-\frac{1}{2}}\big). \tag{1.21}$$

In particular, (1.21) *holds true for the McKean-Vlasov SDE* (1.5) *with* $b_1 = \overline{b}_1 + \widetilde{b}_1$ *provided that*

(H) *b_0 satisfies (1.17); $\overline{b}_1$ is Lipschitz in $\mathbb{R}^d$ and satisfies (1.16) with $\phi(r) = \lambda_0 r$ for some $\lambda_0 > 0$; $\widetilde{b}_1$ is uniformly bounded and fulfills that*

$$|\widetilde{b}_1(x) - \widetilde{b}_1(y)| \leq \varphi(|x-y|), \quad x, y \in \mathbb{R}^d$$

for some $\varphi \in \mathscr{D}$ with $\lim_{r\to\infty} \varphi(r)/r = 0$. Herein,

$$\mathscr{D} := \Big\{\varphi : \mathbb{R}_+ \to \mathbb{R}_+ \Big| \varphi(0) = 0, \varphi \text{ is increasing, continuous, concave and } \int_0^1 \frac{\varphi(s)}{s}\,\mathrm{d}s < \infty\Big\}.$$

Below, we move forward to dwell on applications of Theorem 1.3, and answer the remaining questions proposed in the introductory subsections, one by one.

1.3. Applications

1.3.1. Uniform-in-time distribution distance between an SDE and its delay version

For convenience, we first recall SDEs (1.7) and (1.8). In this subsection, we focus on the following SDE:

$$\mathrm{d}X_t = \beta(\alpha - X_t)\mathrm{d}t + \sigma\,\mathrm{d}W_t + \sigma_0(X_t)\,\mathrm{d}B_t, \quad t > 0 \tag{1.22}$$

with the initial value $X_0 \in L^1(\Omega \to \mathbb{R}^d; \mathscr{F}_0, \mathbb{P})$, where $\alpha \in \mathbb{R}^d, \beta > 0$, $\sigma \in \mathbb{R}$, $\sigma_0 : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$, which satisfies (1.18); $(W_t)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ are independent d-dimensional Brownian motions. Let $(Y_t)_{t>0}$ be the delay version of $(X_t)_{t\geq 0}$, which is determined by the SDE with memory:

$$\mathrm{d}Y_t = \beta(\alpha - Y_{t-r_0})\mathrm{d}t + \sigma\,\mathrm{d}W_t + \sigma_0(Y_t)\,\mathrm{d}B_t, \quad t > 0 \tag{1.23}$$

with the initial value $Y_{[-r_0,0]} = \xi \in L^1(\Omega \to \mathscr{C}; \mathscr{F}_0, \mathbb{P})$ satisfying that for some $C_\xi > 0$,

$$\mathbb{E}|\xi_t-\xi_s|\le C_\xi|t-s|,\quad t,s\in[-r_0,0]. \tag{1.24}$$

Evidently, both (1.22) and (1.23) are strongly well-posed.

The following statement shows that the distributions $(\mathscr{L}_{X_t})_{t\ge0}$ and $(\mathscr{L}_{Y_t})_{t\ge0}$ associated with (1.22) and (1.23) respectively close to each other when the time lag r_0 approaches zero, and most importantly, provides a quantitative characterization upon the distribution deviation.

Theorem 1.6. *Assume that σ_0 satisfies (1.18) and suppose further $\sigma\neq0$ and $\beta>0$. Then, there exist constants $C^*, L^*, \lambda^*>0$ such that for all $L\in[0,L^*]$ and $t\ge0$,*

$$\mathbb{W}_1\big(\mathscr{L}_{X_t},\mathscr{L}_{Y_t}\big)\le C^*\Big(\mathrm{e}^{-\lambda^*t}\,\mathbb{W}_1\big(\mathscr{L}_{X_0},\mathscr{L}_{Y_0}\big)+(1+r_0)\big(1+\mathbb{E}\|\xi\|_\infty\big)r_0^{\frac12}\Big). \tag{1.25}$$

Remark 1.7. Since the distribution of $(Y_t)_{t\ge0}$ is dependent on the segment $\xi\in L^1(\Omega\to\mathscr{C};\mathscr{F}_0,\mathbb{P})$, it is reasonable that the error bound on the right hand side of (1.25) depends on $\mathbb{E}\|\xi\|_\infty$ rather than $\mathbb{E}|\xi_0|$.

1.3.2. Uniform-in-time discretization error bounds for stochastic algorithms

In this subsection, we focus on the McKean-Vlasov SDE (1.9), where the drift term is of super-linear growth and dissipative in the long distance with respect to the spatial variables. As direct applications of Theorem 1.3, we shall tackle uniform-in-time discretization error bounds for the backward EM scheme, the tamed EM scheme, and the adaptive EM scheme, which are constructed in (1.10), (1.12), and (1.14), respectively.

In addition to Assumption $(\mathbf{A}_1)$, we further need to suppose that the drift term b_1 is smooth and locally Lipschitz, which is stated precisely as below.

$(\mathbf{A}_3)$ $\mathbb{R}^d\ni x\mapsto b_1(x)$ is a C^1-function and there exist constants $l^*\ge0$, $K^\star>0$ such that

$$|b_1(x)-b_1(y)|\le K^\star\big(1+|x|^{l^*}+|y|^{l^*}\big)|x-y|,\quad x,y\in\mathbb{R}^d. \tag{1.26}$$

Under Assumption $(\mathbf{A}_1)$ with $\phi(r)=\lambda_0 r$ for some $\lambda_0>0$, (1.9) admits a unique strong solution for $X_0\in L^1(\Omega\to\mathbb{R}^d;\mathscr{F}_0,\mathbb{P})$; see, for example, [30, Theorem 2.1] for more details. Moreover, note that the discrete time version of (1.10) is indeed an implicit equation. Whereas, under (1.16) with $\phi(r)=\lambda_0 r$ and (1.17), the algorithm (1.10) is well defined as long as the step size $\delta\in(0,1/(2(\lambda_0+K)))$; see, for instance, [13, Lemma 3.4] for related details.

With regard to the backward EM scheme, given in (1.10), the long time error bound under the L^1-Wasserstein distance can be presented as follows.

Theorem 1.8. *Assume $(\mathbf{A}_1)$ with $\phi(r)=\lambda_0 r$ for some $\lambda_0>0$ and $\lambda>2K$, and suppose further $\sigma\neq0$ and $(\mathbf{A}_3)$. Then, there exist constants $C^*,\lambda^*>0$ such that for all $i\in\mathbb{S}_N$, $t\ge0$, and $\delta\in(0,\delta^*]$,*

$$\begin{aligned}\mathbb{W}_1\big(\mathscr{L}_{X_t^i},\mathscr{L}_{X_t^{\delta,i,N}}\big)\le C^*\bigg\{&\mathrm{e}^{-\lambda^*t}\,\mathbb{W}_1\big(\mathscr{L}_{X_0^i},\mathscr{L}_{X_0^{\delta,i,N}}\big)+\frac{1}{\sqrt N}\mathbb{1}_{\{K>0\}}\\&+\Big(1+\mathbb{E}\big|X_0^{\delta,i,N}\big|^{(l^*+1)^2}\Big)\delta^{\frac12}\bigg\}\end{aligned} \tag{1.27}$$

in case of $\mathbb{E}|X_0^i|^2 < \infty$ *and* $\mathbb{E}\big|X_0^{\delta,i,N}\big|^{(l^*+1)^2} < \infty$. *In the above,*

$$\delta^* := 1 \wedge \frac{1}{2(\lambda_0 + K)} \wedge \frac{3(\lambda - 2K)}{4(\lfloor (1 \vee l^*)(1 + l^*) \rfloor + 1)(1 + K)^{\lfloor (1\vee l^*)(1+l^*) \rfloor + 1}}.$$

Remark 1.9. Consider the SDE (1.9) with $b(x,\mu) = b(x)$ (so (1.9) is indeed a distribution-independent SDE). For this setting, via the Banach fixed point theorem, [8, Corollary 2.3] implies that the corresponding solution process $(X_t)_{t\ge 0}$ has a unique invariant probability measure (IPM for short), written as π. Once the time-homogeneous Markov chain $(X^\delta_{k\delta})_{k\ge 0}$ (which is determined by the backward EM scheme) possesses a unique IPM, denoted by $\pi^{(\delta)}$, Theorem 1.8 enables us to deduce that $\mathbb{W}_1(\pi, \pi^{(\delta)}) \le c\,\delta^{\frac{1}{2}}$ for some constant $c > 0$. This reveals the quantitative estimate between the exact IPM and its numerical version for SDEs with partially dissipative drifts.

Next, we apply Theorem 1.3 to the tamed EM scheme. Furthermore, we suppose that

($\mathbf{A}_1'$) there exist constants $\lambda_{b_1}, \widehat{\lambda}_{b_1}, C_{b_1}, \widehat{C}_{b_1} > 0$ such that for any $x \in \mathbb{R}^d$,

$$\langle x, b_1(x)\rangle \le -\lambda_{b_1}|x|^2 \cdot \|\nabla b_1(x)\|_{\mathrm{HS}} + C_{b_1}, \quad |b_1(x)| \le \widehat{\lambda}_{b_1}|x| \cdot \|\nabla b_1(x)\|_{\mathrm{HS}} + \widehat{C}_{b_1}. \tag{1.28}$$

Moreover, for some constant $\alpha > 0$, there is an $R_\alpha > 0$ such that for $x \in \mathbb{R}^d$ with $|x| \ge R_\alpha$,

$$\|\nabla b_1(x)\|_{\mathrm{HS}} \ge \alpha. \tag{1.29}$$

Remark 1.10. The second prerequisite in (1.28) is evidently satisfied when b_1 is of polynomial growth. Obviously, for $b_1(x) = x - x^\ell, x \in \mathbb{R}$, with ℓ being an odd number $\ell \ge 1$, the first technical condition in (1.28) and the one in (1.29) are valid, separately. Moreover, in Assumption ($\mathbf{A}_1'$), the gradient of b_1 is involved based on the construction of the tamed EM scheme presented below.

The tamed EM scheme associated with (1.9) is constructed as follows: for $\delta > 0$,

$$\mathrm{d}X_t^{\delta,i,N} = \big(b_1^\delta(X_{t_\delta}^{\delta,i,N}) + (b_0 * \widetilde{\mu}_{t_\delta}^{\delta,N})(X_{t_\delta}^{\delta,i,N})\big)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^i, \quad i \in \mathbb{S}_N, \quad t > 0, \tag{1.30}$$

where for any $x \in \mathbb{R}^d$,

$$b_1^\delta(x) := \frac{b_1(x)}{1 + \delta^{\frac{1}{2}}\|\nabla b_1(x)\|_{\mathrm{HS}}}.$$

Moreover, for brevity, we set for $\kappa \ge 0$ and $\rho := 4K(K + \widehat{\lambda}_{b_1})$,

$$\delta_\kappa^* := 1 \wedge \alpha^{-\frac{1}{2}} \wedge \frac{\lambda_{b_1}^2}{\widehat{\lambda}_{b_1}^4} \wedge \frac{\kappa^2}{(2\alpha(2K + \rho(1 + 1/\alpha) + \widehat{\lambda}_{b_1}^2)^2} \wedge \frac{1}{\kappa/4 + K + \rho}. \tag{1.31}$$

Concerning the tamed EM scheme (1.30), we have the following discretization error bounds in an infinite-time horizon.

Theorem 1.11. *Assume Assumptions* $(\mathbf{A}_1)$ *with* $\phi(r)=\lambda_0 r$ *for some* $\lambda_0>0$ *and* $(\mathbf{A}_1')$, *and suppose further* $\kappa:=\alpha\lambda_{b_1}-2K>0$, $\lambda>2K$, $\sigma\neq 0$, *as well as* $(\mathbf{A}_3)$. *Then, there exist constants* $C^*,\lambda^*>0$ *such that for all* $\delta\in(0,\delta_\kappa^*]$, $i\in\mathbb{S}_N$, *and* $t\geq 0$,

$$\mathbb{W}_1\big(\mathscr{L}_{X_t^i},\mathscr{L}_{X_t^{\delta,i,N}}\big)\leq C^*\Big\{\mathrm{e}^{-\lambda^* t}\,\mathbb{W}_1\big(\mathscr{L}_{X_0^i},\mathscr{L}_{X_0^{\delta,i,N}}\big)+\frac{1}{\sqrt{N}}\mathbb{1}_{\{K>0\}}+\big(1+\mathbb{E}\big|X_0^{\delta,i,N}\big|^{2l^*+1}\big)\delta^{\frac{1}{2}}\Big\} \tag{1.32}$$

as long as $\mathbb{E}|X_0^i|^2<\infty$ *and* $\mathbb{E}\big|X_0^{\delta,i,N}\big|^{2l^*+1}<\infty$.

Finally, we apply Theorem 1.3 to the adaptive EM scheme (1.13) with the adaptive step size

$$h_n^\delta:=\delta\min\Big\{\frac{1}{1+|b(X_{t_n}^{\delta,1,N},\widetilde{\mu}_{t_n}^{\delta,N})|^2},\cdots,\frac{1}{1+|b(X_{t_n}^{\delta,N,N},\widetilde{\mu}_{t_n}^{\delta,N})|^2}\Big\},\qquad \delta\in(0,1). \tag{1.33}$$

Since, in this paper, we are interested in the error analysis in an infinite-time horizon, the time grid $t_{n+1}=t_n+h_n^\delta$ should go to infinity almost surely. This can be examined in Lemma 4.3 below.

As far as the continuous-time version of (1.13), defined accordingly in (1.14), is concerned, the uniform discretization error bound is revealed as follows.

Theorem 1.12. *Assume Assumptions* $(\mathbf{A}_1)$ *with* $\phi(r)=\lambda_0 r$ *and* $\lambda>2K$, *and suppose further* $(\mathbf{A}_3)$ *and* $\sigma\neq 0$. *Then, there exist constants* $C^*,\lambda^*>0$ *such that for all* $\delta\in(0,1)$, $i\in\mathbb{S}_N$, *and* $t\geq 0$,

$$\mathbb{W}_1\big(\mathscr{L}_{X_t^i},\mathscr{L}_{X_t^{\delta,i,N}}\big)\leq C^*\Big\{\mathrm{e}^{-\lambda^* t}\,\mathbb{W}_1\big(\mathscr{L}_{X_0^i},\mathscr{L}_{X_0^{\delta,i,N}}\big)+\frac{1}{\sqrt{N}}\mathbb{1}_{\{K>0\}}+\big(1+\mathbb{E}\big|X_0^{\delta,i,N}\big|^{l^*}\big)\delta^{\frac{1}{2}}\Big\} \tag{1.34}$$

as long as $\mathbb{E}|X_0^i|^2<\infty$ *and* $\mathbb{E}\big|X_0^{\delta,i,N}\big|^{1\vee l^*}<\infty$.

Remark 1.13. Since the discretization error is investigated under the L^1-Wasserstein distance, it is logical to require that the initial distribution for the algorithm under consideration has a finite moment of the first order. This indeed takes place in case of $l^*=0$ (which corresponds to the globally Lipschitz case for the drift involved) as demonstrated in Theorems 1.8, 1.11 and 1.12. Nevertheless, concerning stochastic algorithms associated with McKean-Vlasov SDEs with drifts of super-linear growth with respect to spatial variables, it is quite natural to enhance the moment order for initial distributions. With contrast to the tamed/adaptive EM scheme, higher order moments need to be imposed on the initial distributions for the backward EM scheme due to the fact that the tamed drift or adaptive step size can offset growth of the original drift in a certain sense.

Furthermore, we would like to say a few words on the noise term σ_0 in (1.1). Once σ_0 does not vanish, the non-interacting particle system and the corresponding numerical version enjoy different noise terms. For this setup, the asymptotic coupling by reflection will no longer work to investigate long time error bounds for stochastic algorithms. Based on this, in the present work, we focus merely on the additive noise case in lieu of the multiplicative noise setting.

The remainder part of this paper is organized as follows. Based on some preliminaries, in Section 2, we complete the proof of Theorem 1.3 by constructing an appropriate asymptotic coupling by reflection, and meanwhile finish the proof of Corollary 1.5. Section 3 is devoted to the proof of Theorem 1.6, where the variation-of-constants formula for semi-linear SDEs with memory plays a crucial role. In the final section, with the aid of uniform-in-time moment estimates for backward/tamed/adaptive EM schemes (where the underlying proofs are rather tricky), we aim to implement proofs of Theorems 1.8, 1.11 and 1.12, respectively.

2. Proofs of Theorem 1.3 and Corollary 1.5

Before the proof of Theorem 1.3, we prepare some warm-up lemmas.

Lemma 2.1. *Assume that the SDEs (1.1) and (1.2) are weakly well-posed. Then, for any $\varepsilon > 0$, the path-valued processes $\mathbf{Y}^{N,\varepsilon}$ and $\mathbf{Y}^{N,N,\varepsilon}$ share the common distributions as those of $\mathbf{X}^N$ and $\mathbf{X}^{N,N}$ on the path spaces $C([0,\infty);\mathbb{R}^d)$ and $C([-r_0,\infty);\mathbb{R}^d)$, respectively.*

Proof. For $i \in \mathbb{S}_N$ and $t \geq 0$, let

$$\overline{W}_t^{1,i} = \int_0^t \Pi(Z_s^{i,N,\varepsilon})\mathrm{d}W_s^{1,i}.$$

With this shorthand notation, the SDE solved by $(Y_t^{i,N,\varepsilon})_{i\in\mathbb{S}_N}$ can be reformulated as

$$\mathrm{d}Y_t^{i,N,\varepsilon} = \widetilde{b}\big(Y_{\theta_t}^{i,N,\varepsilon}, \widetilde{\mu}_{\overline{\theta}_t}^{N,\varepsilon}\big)\,\mathrm{d}t + \sigma h_\varepsilon(|Z_t^{i,N,\varepsilon}|)\mathrm{d}\overline{W}_t^{1,i} + \sigma h_\varepsilon^*(|Z_t^{i,N,\varepsilon}|)\mathrm{d}W_t^{2,i} + \sigma_0(Y_t^{i,N,\varepsilon})\,\mathrm{d}B_t^i \tag{2.1}$$

with the initial value $Y_{[-r_0,0]}^{i,N,\varepsilon} = X_{[-r_0,0]}^{i,N}$. Observe that the SDE (2.1) has the same weak solution as that of the SDE solved by $(Y_t^{1,N,\varepsilon}, \cdots, Y_t^{N,N,\varepsilon})_{t\geq 0}$. Therefore, to complete the proof of Lemma 2.1, it is sufficient to show that the distributions of $(\mathbf{Y}_t^{N,\varepsilon})_{t\geq 0}$ and $(\mathbf{X}_t^N)_{t\geq 0}$ are identical.

Set for $i \in \mathbb{S}_N$ and $t \geq 0$,

$$\widetilde{W}_t^i := \int_0^t h_\varepsilon(|Z_s^{i,N,\varepsilon}|)\,\mathrm{d}W_s^{1,i} + \int_0^t h_\varepsilon^*(|Z_s^{i,N,\varepsilon}|)\,\mathrm{d}W_s^{2,i}.$$

Since $W^{1,i}$ is independent of $W^{2,i}$, besides $h_\varepsilon(r)^2 + h_\varepsilon^*(r)^2 = 1, r \geq 0$, Lévy's characterization shows that $\widetilde{W}^i$ is still a Brownian motion. Then, the SDE solved by $(Y_t^{i,\varepsilon})_{t\geq 0}$ can be rewritten as an SDE driven by $\widetilde{W}^i$. More precisely, we have

$$\mathrm{d}Y_t^{i,\varepsilon} = b(Y_t^{i,\varepsilon}, \widehat{\mu}_t^{i,\varepsilon})\,\mathrm{d}t + \sigma\,\mathrm{d}\widetilde{W}_t^i + \sigma_0(Y_t^{i,\varepsilon})\,\mathrm{d}B_t^i, \quad i \in \mathbb{S}_N,\ t > 0. \tag{2.2}$$

In order to prove that, for any $\varepsilon > 0$, the distribution of $(\mathbf{Y}_t^{N,\varepsilon})_{t\geq 0}$ is equal to that of $(\mathbf{X}_t^N)_{t\geq 0}$, it remains to verify that, for any $i \neq j$, $\widetilde{W}^i$ and $\widetilde{W}^j$ are mutually independent. Indeed, by applying Itô's formula, it follows that for $u, v \in \mathbb{R}^d$, $i, j \in \mathbb{S}_N$ with $i \neq j$, and $t > 0$,

$$\begin{aligned}\mathrm{d}\big(\langle u, \widetilde{W}_t^i\rangle\langle v, \widetilde{W}_t^j\rangle\big) &= \langle v, \widetilde{W}_t^j\rangle\mathrm{d}\langle u, \widetilde{W}_t^i\rangle + \langle u, \widetilde{W}_t^i\rangle\mathrm{d}\langle v, \widetilde{W}_t^j\rangle + \mathrm{d}[\langle u, \widetilde{W}_t^i\rangle, \langle v, \widetilde{W}_t^j\rangle] \\ &= \langle v, \widetilde{W}_t^j\rangle\mathrm{d}\langle u, \widetilde{W}_t^i\rangle + \langle u, \widetilde{W}_t^i\rangle\mathrm{d}\langle v, \widetilde{W}_t^j\rangle,\end{aligned} \tag{2.3}$$

where the second identity holds true due to the quadratic variation $[\langle u, \widetilde{W}_t^i\rangle, \langle v, \widetilde{W}_t^j\rangle] = 0$, which is valid since $W^{1,1}, \cdots, W^{1,N}$ (resp. $W^{2,1}, \cdots, W^{2,N}$) are independent and $(W^{1,1}, \cdots, W^{1,N})$ is also independent of $(W^{2,1}, \cdots, W^{2,N})$. Obviously, (2.3) manifests that $(\langle u, \widetilde{W}_t^i\rangle\langle v, \widetilde{W}_t^j\rangle)_{t\ge 0}$ is a martingale. This results in that the covariance matrix $\mathbb{E}(\widetilde{W}_t^i \otimes \widetilde{W}_t^j), i \neq j$, is a $d \times d$ zero matrix. Hence, we conclude that, for any $i \neq j$, $\widetilde{W}^i$ and $\widetilde{W}^j$ are mutually independent. Next, by following an analogous procedure above, we deduce that $(\widetilde{W}^1, \cdots, \widetilde{W}^N)$ is independent of $(B^1, \cdots, B^N)$. Subsequently, thanks to the weak uniqueness of (1.1), we conclude that $(\mathbf{X}_t^N)_{t\ge 0}$ and $(\mathbf{Y}_t^{N,\varepsilon})_{t\ge 0}$ possess the same distribution. □

Lemma 2.2. *Assume* $(\mathbf{A}_1)$ *and* $(\mathbf{A}_2)$ *with* $\lambda > 2K + \frac{1}{2}(1 + (p-2)^+)L$ *for some* $p \ge 2$. *Then, for any* $i \in \mathbb{S}_N$, *there exists a constant* $C_p^* > 0$ (*independent of* i) *such that*

$$\sup_{t\ge 0} \mathbb{E}\big|Y_t^{i,\varepsilon}\big|^p \le C_p^*\big(1 + \mathbb{E}\big|Y_0^{i,\varepsilon}\big|^p\big) \tag{2.4}$$

as long as $\mathbb{E}|Y_0^{i,\varepsilon}|^p < \infty$.

Proof. For any $p \ge 1$, let

$$V_p(x) = (1 + |x|^2)^{\frac{p}{2}}, \quad x \in \mathbb{R}^d.$$

Performing a direct calculation shows that for any $x \in \mathbb{R}^d$,

$$\begin{aligned}
&\nabla V_p(x) = p(1 + |x|^2)^{\frac{p}{2}-1} x \ \text{ and}\\
&\nabla^2 V_p(x) = p(1 + |x|^2)^{\frac{p}{2}-1} I_{d\times d} + p(p-2)(1 + |x|^2)^{\frac{p}{2}-2}(x \otimes x).
\end{aligned}$$

Next, by virtue of (1.16) and (1.17), it follows that for all $x \in \mathbb{R}^d$ and $\mu \in \mathscr{P}_1(\mathbb{R}^d)$,

$$\langle x, b(x,\mu)\rangle \le \ell_0\phi(\ell_0) + \lambda\ell_0^2 - (\lambda - K)|x|^2 + |x|\big(|b_0(\mathbf{0})| + |b_1(\mathbf{0})| + K\mu(|\cdot|)\big). \tag{2.5}$$

Then, applying Itô's formula to the SDE (2.2), we derive from (1.18) that

$$\begin{aligned}
&\mathrm{d}V_p(Y_t^{i,\varepsilon})\\
&= \Big(p\big(1 + |Y_t^{i,\varepsilon}|^2\big)^{\frac{p}{2}-1}\big\langle Y_t^{i,\varepsilon}, b(Y_t^{i,\varepsilon}, \widehat{\mu}_t^{i,\varepsilon})\big\rangle\\
&\quad + \frac{1}{2}p\sigma^2\big(1 + |Y_t^{i,\varepsilon}|^2\big)^{\frac{p}{2}-1}\big(d + (p-2)(1 + |Y_t^{i,\varepsilon}|^2)^{-1}|Y_t^{i,\varepsilon}|^2\big)\Big)\,\mathrm{d}t\\
&\quad + \frac{1}{2}p\big(1 + |Y_t^{i,\varepsilon}|^2\big)^{\frac{p}{2}-1}\big(\|\sigma_0(Y_t^{i,\varepsilon})\|_{\mathrm{HS}}^2 + (p-2)(1 + |Y_t^{i,\varepsilon}|^2)^{-1}|\sigma_0^*(Y_t^{i,\varepsilon})Y_t^{i,\varepsilon}|^2\big)\,\mathrm{d}t + \mathrm{d}\overline{M}_t^{p,i}\\
&\le \Big(p\big(1 + |Y_t^{i,\varepsilon}|^2\big)^{\frac{p}{2}-1}\big(\ell_0\phi(\ell_0) + \lambda\ell_0^2 - (\lambda - K)|Y_t^{i,\varepsilon}|^2 + |Y_t^{i,\varepsilon}|\big(|b_0(\mathbf{0})| + |b_1(\mathbf{0})| + K\mathbb{E}|Y_t^{i,\varepsilon}|\big)\big)\\
&\quad + \frac{1}{2}p\big(1 + |Y_t^{i,\varepsilon}|^2\big)^{\frac{p}{2}-1}\big(\sigma^2(d + (p-2)^+)\\
&\qquad + \|\sigma_0(Y_t^{i,\varepsilon})\|_{\mathrm{HS}}^2 + (p-2)^+\|\sigma_0(Y_t^{i,\varepsilon})\|_{\mathrm{op}}^2\big)\Big)\,\mathrm{d}t + \mathrm{d}\overline{M}_t^{p,i},
\end{aligned}$$

where $(\overline{M}_t^{p,i})_{t\ge 0}$ is a martingale and $\|\cdot\|_{\rm op}$ means the operator norm. For any $\kappa>0$, it can readily be seen from (1.18) that there exists a constant $C_\kappa>0$ such that

$$\|\sigma_0(x)\|_{\rm HS}^2\le (L+\kappa)|x|^2+C_\kappa,\qquad x\in\mathbb{R}^d.$$

Set $\lambda^\star:=\frac{1}{2}(\lambda-2K-\frac{1}{2}(1+(p-2)^+)L)$, which is positive due to $\lambda>2K+\frac{1}{2}(1+(p-2)^+)L$. Again, by applying Itô's formula, there exists a positive constant $C_\star=C(\lambda^\star)$ such that

$$\begin{aligned}\mathrm{d}\big(\mathrm{e}^{p\lambda^\star t}V_p(Y_t^{i,\varepsilon})\big)&\le \mathrm{e}^{p\lambda^\star t}\big(C_\star-pKV_p(Y_t^{i,\varepsilon})+pK\big(1+|Y_t^{i,\varepsilon}|^2\big)^{\frac{1}{2}(p-1)}\mathbb{E}|Y_t^{i,\varepsilon}|\big)\,\mathrm{d}t+\mathrm{e}^{p\lambda^\star t}\mathrm{d}\overline{M}_t^{p,i}\\&\le \mathrm{e}^{p\lambda^\star t}\big(C_\star-KV_p(Y_t^{i,\varepsilon})+K\mathbb{E}V_p(Y_t^{i,\varepsilon})\big)\,\mathrm{d}t+\mathrm{e}^{p\lambda^\star t}\mathrm{d}\overline{M}_t^{p,i},\end{aligned}\tag{2.6}$$

where we utilized Young's inequality in the first inequality and the second inequality, as well as Jensen's inequality in the second inequality. Next, integrating from 0 to t followed by taking expectations on both sides of (2.6) yields that

$$\begin{aligned}\mathbb{E}V_p(Y_t^{i,\varepsilon})&\le \mathrm{e}^{-p\lambda^\star t}\mathbb{E}V_p(Y_0^{i,\varepsilon})+\int_0^t\mathrm{e}^{-\lambda^\star(t-s)}\mathbb{E}\big(-KV_p(Y_s^{i,\varepsilon})+K\mathbb{E}V_p(Y_s^{i,\varepsilon})\big)\,\mathrm{d}s\\&\quad+C_\star\int_0^t\mathrm{e}^{-\lambda^\star(t-s)}\,\mathrm{d}s\\&=\mathrm{e}^{-p\lambda^\star t}\mathbb{E}V_p(Y_0^{i,\varepsilon})+C_\star\int_0^t\mathrm{e}^{-\lambda^\star(t-s)}\,\mathrm{d}s,\end{aligned}$$

where the identity is valid due to the fact that $\mathbb{E}(-V_p(Y_s^{i,\varepsilon})+\mathbb{E}V_p(Y_s^{i,\varepsilon}))=0, s\in[0,t]$. As a consequence (2.4) is available immediately. □

With preliminary Lemmas 2.1 and 2.2 at hand, we are in position to complete the proof of Theorem 1.3.

Proof of Theorem 1.3. Recall that, for all $t\ge 0$, $(Y_t^{i,\varepsilon},Y_t^{i,N,\varepsilon})_{i\in\mathbb{S}_N}$ solves (1.19) with the initial value $\big(Y_0^{i,\varepsilon},Y_{[-r_0,0]}^{i,N,\varepsilon}\big)_{i\in\mathbb{S}_N}=\big(X_0^i,X_{[-r_0,0]}^{i,N}\big)_{i\in\mathbb{S}_N}$, which are i.i.d. random variables. For any $\mu\in\mathscr{P}_1(\mathbb{R}^d)$ and $\nu\in\mathscr{P}_1(\mathscr{C})$, in the following analysis, we choose $(X_0^i,X_{[-r_0,0]}^{i,N})_{i\in\mathbb{S}_N}$ such that

$$\mathbb{W}_1(\mu,\nu_0)=\mathbb{E}|X_0^i-X_0^{i,N}|,\quad i\in\mathbb{S}_N,\tag{2.7}$$

in which $\nu_0(\mathrm{d}x):=\nu(\{\eta\in\mathscr{C}\}:\eta_0\in\mathrm{d}x)$.

Note that

$$\langle I_{d\times d}-\mathbf{e}(x)\otimes\mathbf{e}(x),\mathbf{e}(x)\otimes\mathbf{e}(x)\rangle_{\rm HS}=0\qquad\text{and}\qquad(\mathbf{e}(x)\otimes\mathbf{e}(x))\mathbf{e}(x)=\mathbf{e}(x),\quad x\neq\mathbf{0}.$$

Next, we shall fix the index $i\in\mathbb{S}_N$. For any $\eta\in(0,1]$, define the function V_η by

$$V_\eta(x) = (\eta + |x|^2)^{1/2}, \quad x \in \mathbb{R}^d,$$

which is indeed a smooth approximation of the function $\mathbb{R}^d \ni x \mapsto |x|$. Applying Itô's formula and utilizing the facts:

$$\nabla V_\eta(x) = \frac{x}{V_\eta(x)} \quad \text{and} \quad \nabla^2 V_\eta(x) = \frac{1}{V_\eta(x)} I_{d\times d} - \frac{x \otimes x}{V_\eta(x)^3},$$

it follows from the (1.19) that

$$\begin{aligned} \mathrm{d}V_\eta(Z_t^{i,N,(\varepsilon)}) \le & \frac{1}{V_\eta(Z_t^{i,N,(\varepsilon)})}\Psi_t^{i,N,\varepsilon}\,\mathrm{d}t + \frac{1}{V_\eta(Z_t^{i,N,(\varepsilon)})}\Big(\big\langle Z_t^{i,N,\varepsilon}, \big(\sigma_0(Y_t^{i,\varepsilon}) - \sigma_0(Y_t^{i,N,\varepsilon})\big)\mathrm{d}B_t^i\big\rangle \\ & + 2\sigma h_\varepsilon(|Z_t^{i,N,\varepsilon}|)\big\langle Z_t^{i,N,\varepsilon}, \big(\mathbf{e}(Z_t^{i,N,\varepsilon}) \otimes \mathbf{e}(Z_t^{i,N,\varepsilon})\big)\mathrm{d}W_t^{1,i}\big\rangle\Big) \\ & + 4\sigma^2 \frac{h_\varepsilon(|Z_t^{i,N,\varepsilon}|)^2}{V_\eta(Z_t^{i,N,(\varepsilon)})}\Big\langle I_{d\times d} - \frac{Z_t^{i,N,\varepsilon} \otimes Z_t^{i,N,\varepsilon}}{|V_\eta(Z_t^{i,N,(\varepsilon)})|^2}, \mathbf{e}(Z_t^{i,N,\varepsilon}) \otimes \mathbf{e}(Z_t^{i,N,\varepsilon})\Big\rangle_{\mathrm{HS}}\,\mathrm{d}t, \end{aligned} \tag{2.8}$$

where $\langle \cdot, \cdot \rangle_{\mathrm{HS}}$ stands for the Hilbert-Schmidt inner product, and

$$\Psi_t^{i,N,\varepsilon} := \big\langle Z_t^{i,N,\varepsilon}, b(Y_t^{i,\varepsilon}, \widehat{\mu}_t^{i,\varepsilon}) - \widetilde{b}(Y_{\theta_t}^{i,N,\varepsilon}, \widetilde{\mu}_{\theta_t}^{N,\varepsilon})\big\rangle + \frac{1}{2}\|\sigma_0(Y_t^{i,\varepsilon}) - \sigma_0(Y_t^{i,N,\varepsilon})\|_{\mathrm{HS}}^2.$$

Note that for any $x \in \mathbb{R}^d$,

$$\frac{x}{V_\eta(x)} \overset{\eta\to 0}{\longrightarrow} \frac{x}{|x|}\mathbb{1}_{\{x\neq \mathbf{0}\}} \quad \text{and} \quad \frac{h_\varepsilon(x)^2}{V_\eta(x)}\Big\langle I_{d\times d} - \frac{x\otimes x}{|V_\eta(x)|^2}, \mathbf{e}(x)\otimes\mathbf{e}(x)\Big\rangle_{\mathrm{HS}} \le \frac{\eta}{(\eta+\varepsilon^2)^{\frac{3}{2}}} \overset{\eta\to 0}{\longrightarrow} 0.$$

Therefore, approaching $\eta \to 0$ in (2.8) leads to the estimates below:

$$\begin{aligned} \mathrm{d}|Z_t^{i,N,\varepsilon}| \le & \frac{1}{|Z_t^{i,N,\varepsilon}|}\mathbb{1}_{\{Z_t^{i,N,\varepsilon}\neq\mathbf{0}\}}\Psi_t^{i,N,\varepsilon}\,\mathrm{d}t + 2\sigma h_\varepsilon(|Z_t^{i,N,\varepsilon}|)\mathbb{1}_{\{Z_t^{i,N,\varepsilon}\neq\mathbf{0}\}}\big\langle \mathbf{e}(Z_t^{i,N,\varepsilon}), \mathrm{d}W_t^{1,i}\big\rangle \\ & + \frac{1}{|Z_t^{i,N,\varepsilon}|}\mathbb{1}_{\{Z_t^{i,N,\varepsilon}\neq\mathbf{0}\}}\big\langle Z_t^{i,N,\varepsilon}, \big(\sigma_0(Y_t^{i,\varepsilon}) - \sigma_0(Y_t^{i,N,\varepsilon})\big)\mathrm{d}B_t^i\big\rangle. \end{aligned} \tag{2.9}$$

By splitting the quantity $b(Y_t^{i,\varepsilon}, \widehat{\mu}_t^{i,\varepsilon}) - \widetilde{b}(Y_{\theta_t}^{i,N,\varepsilon}, \widetilde{\mu}_{\theta_t}^{N,\varepsilon})$ into three terms followed by taking (1.16), (1.17) as well as (1.18) into consideration, we obtain that

$$\begin{aligned} \Psi_t^{i,N,\varepsilon} \le & \langle Z_t^{i,N,\varepsilon}, (b_1(Y_t^{i,\varepsilon}) - b_1(Y_t^{i,N,\varepsilon}) + \widetilde{\psi}_t^{i,N,\varepsilon}\rangle \\ & + \langle Z_t^{i,N,\varepsilon}, (b_0 * \overline{\mu}_t^{N,\varepsilon})(Y_t^{i,\varepsilon}) - (b_0 * \widetilde{\mu}_t^{N,\varepsilon})(Y_t^{i,N,\varepsilon})\rangle + \langle Z_t^{i,N,\varepsilon}, \psi_t^{i,N,\varepsilon}\rangle + \frac{1}{2}L|Z_t^{i,N,\varepsilon}|^2 \\ \le & \big(\phi(|Z_t^{i,N,\varepsilon}|) + \lambda|Z_t^{i,N,\varepsilon}|\big)|Z_t^{i,N,\varepsilon}|\mathbb{1}_{\{|Z_t^{i,N,\varepsilon}|\le \ell_0\}} - (\lambda - K - L/2)|Z_t^{i,N,\varepsilon}|^2 \\ & + |Z_t^{i,N,\varepsilon}|\big(K\,\mathbb{W}_1(\overline{\mu}_t^{N,\varepsilon}, \widetilde{\mu}_t^{N,\varepsilon}) + |\widetilde{\psi}_t^{i,N,\varepsilon}| + |\psi_t^{i,N,\varepsilon}|\big), \end{aligned} \tag{2.10}$$

where $\overline{\mu}_t^{N,\varepsilon} := \frac{1}{N}\sum_{j=1}^N \delta_{Y_t^{j,\varepsilon}}$,

$$\widetilde{\psi}_t^{i,N,\varepsilon} := (b_0 * \widehat{\mu}_t^{i,\varepsilon})(Y_t^{i,\varepsilon}) - (b_0 * \overline{\mu}_t^{N,\varepsilon})(Y_t^{i,\varepsilon}) \text{ and}$$
$$\psi_t^{i,N,\varepsilon} := b(Y_t^{i,N,\varepsilon}, \widetilde{\mu}_t^{N,\varepsilon}) - \widetilde{b}(Y_{\theta_t}^{i,N,\varepsilon}, \widetilde{\mu}_{\overline{\theta}_t}^{N,\varepsilon}).$$

Inserting (2.10) back into (2.9) yields the estimate below:

$$\begin{aligned} \mathrm{d}|Z_t^{i,N,\varepsilon}| \le \mathbb{1}_{\{Z_t^{i,N,\varepsilon}\neq \mathbf{0}\}}\Big((\phi(|Z_t^{i,N,\varepsilon}|) + \lambda|Z_t^{i,N,\varepsilon}|)\mathbb{1}_{\{|Z_t^{i,N,\varepsilon}|\le \ell_0\}} - (\lambda - K - L/2)|Z_t^{i,N,\varepsilon}| \\ + K\,\mathbb{W}_1(\overline{\mu}_t^{N,\varepsilon}, \widetilde{\mu}_t^{N,\varepsilon}) + |\widetilde{\psi}_t^{i,N,\varepsilon}| + |\psi_t^{i,N,\varepsilon}|\Big)\,\mathrm{d}t \\ + 2\sigma h_\varepsilon(|Z_t^{i,N,\varepsilon}|)\mathbb{1}_{\{Z_t^{i,N,\varepsilon}\neq \mathbf{0}\}}\big\langle \mathbf{e}(Z_t^{i,N,\varepsilon}), \mathrm{d}W_t^{1,i}\big\rangle \\ + \frac{1}{|Z_t^{i,N,\varepsilon}|}\mathbb{1}_{\{Z_t^{i,N,\varepsilon}\neq \mathbf{0}\}}\big\langle Z_t^{i,N,\varepsilon}, \big(\sigma_0(Y_t^{i,\varepsilon}) - \sigma_0(Y_t^{i,N,\varepsilon})\big)\mathrm{d}B_t^i\big\rangle. \end{aligned} \tag{2.11}$$

Next, we define the C^2-function

$$f(r) = 1 - \mathrm{e}^{-c_1 r} + c_2 r, \quad r \ge 0,$$

where

$$c_1 := \frac{2(\phi(\ell_0) + (L+2K)\ell/2)}{\sigma^2} \quad \text{and} \quad c_2 := c_1 \mathrm{e}^{-c_1\ell_0}.$$

Applying Itô's formula, we derive from (2.11) and $f'' < 0$ that

$$\begin{aligned} \mathrm{d}f(|Z_t^{i,N,\varepsilon}|) \le \Big(\big(\varphi(|Z_t^{i,N,\varepsilon}|) + 2\sigma^2 f''(|Z_t^{i,N,\varepsilon}|)\big)h_\varepsilon(|Z_t^{i,N,\varepsilon}|)^2\mathbb{1}_{\{Z_t^{i,N,\varepsilon}\neq \mathbf{0}\}} \\ + \varphi(|Z_t^{i,N,\varepsilon}|)\big(1 - h_\varepsilon(|Z_t^{i,N,\varepsilon}|)^2\big) \\ + f'(|Z_t^{i,N,\varepsilon}|)\big(K\,\mathbb{W}_1(\overline{\mu}_t^{N,\varepsilon}, \widetilde{\mu}_t^{N,\varepsilon}) + |\widetilde{\psi}_t^{i,N,\varepsilon}| + |\psi_t^{i,N,\varepsilon}|\big)\Big)\,\mathrm{d}t + \mathrm{d}M_t^{i,\varepsilon} \end{aligned}$$

for some martingale $(M_t^{i,\varepsilon})_{t\ge 0}$, where the function $\varphi : [0,\infty) \to \mathbb{R}$ is defined by

$$\varphi(r) = f'(r)\big((\phi(r) + \lambda r)\mathbb{1}_{\{r\le \ell_0\}} - (\lambda - K - L/2)r\big).$$

Below, for the case $0 \le r \le \ell_0$ and the case $r > \ell_0$, we aim to verify respectively that

$$\varphi^*(r) := \varphi(r) + 2\sigma^2 f''(r) \le -\lambda^* f(r), \quad r \ge 0, \tag{2.12}$$

where for $\lambda > K + L/2$,

$$\lambda^* := \frac{((2\phi(\ell_0) + L\ell_0) \wedge (\lambda - K - L/2))c_2}{1 - \mathrm{e}^{-c_1\ell_0} + c_2\ell_0}.$$

By virtue of

$$f'(r) = c_1\mathrm{e}^{-c_1 r} + c_2 \quad \text{and} \quad f''(r) = -c_1^2\mathrm{e}^{-c_1 r}, \quad r \ge 0,$$

it is easy to ready that for any $r \geq 0$,

$$\varphi^*(r) = (c_1 e^{-c_1 r} + c_2)\big((\phi(r) + \lambda r)\mathbb{1}_{\{r \leq \ell_0\}} - (\lambda - K - L/2)r\big) - 2\sigma^2 c_1^2 e^{-c_1 r}.$$

For the case $0 \leq r \leq \ell_0$, in view of $c_2 = c_1 e^{-c_1 \ell_0} \leq c_1 e^{-c_1 r}$ and the increasing property of ϕ, we find that

$$\begin{aligned} \varphi^*(r) &\leq -\big(2\sigma^2 c_1^2 - 2c_1(\phi(\ell_0) + (L+2K)\ell_0/2)\big)e^{-c_1 r} \\ &\leq -\sigma^2 c_1^2 e^{-c_1 \ell_0} \\ &\leq -\frac{c_1 c_2 \sigma^2}{1 - e^{-c_1 \ell_0} + c_2 \ell_0} f(r), \end{aligned} \tag{2.13}$$

where the second inequality is verifiable owing to $\sigma^2 c_1^2 = 2c_1(\phi(\ell_0) + L\ell_0/2)$, and the last display is valid due to the fact that $r \mapsto f(r)$ is increasing on $[0, \infty)$. On the other hand, once $\lambda > K + L/2$, we derive that for $r > \ell_0$,

$$\varphi^*(r) \leq -c_2(\lambda - K - L/2)r \leq -\frac{c_2(\lambda - K - L/2)}{1 - e^{-c_1 \ell_0} + c_2 \ell_0} f(r), \tag{2.14}$$

where the second inequality is provable since, for $\alpha > 0$, $r \mapsto \frac{r}{1 - e^{-\alpha r} + r}$ is increasing on the interval $[0, \infty)$ by taking the fundamental inequality: $1 - e^{-r} \geq r e^{-r}, r \geq 0$, into consideration. Consequently, (2.12) follows by combining (2.13) with (2.14).

In the sequel, invoking (2.12) and taking $c_2 \leq f'(r) \leq c_1 + c_2, r \geq 0$, and $f(0) = 0$ into account leads to

$$\begin{aligned} d f(|Z_t^{i,N,\varepsilon}|) &\leq \Big(-\lambda^* f(|Z_t^{i,N,\varepsilon}|) h_\varepsilon(|Z_t^{i,N,\varepsilon}|)^2 + \varphi(|Z_t^{i,N,\varepsilon}|)\big(1 - h_\varepsilon(|Z_t^{i,N,\varepsilon}|)^2\big) \\ &\quad + (c_1 + c_2)\big(K\,\mathbb{W}_1(\overline{\mu}_t^{N,\varepsilon}, \widetilde{\mu}_t^{N,\varepsilon}) + |\widetilde{\psi}_t^{i,N,\varepsilon}| + |\psi_t^{i,N,\varepsilon}|\big)\Big) dt + dM_t^{i,\varepsilon} \\ &\leq dM_t^{i,\varepsilon} + \Big(-\lambda^* f(|Z_t^{i,N,\varepsilon}|) + \big(\lambda^* f(|Z_t^{i,N,\varepsilon}|) + \varphi(|Z_t^{i,N,\varepsilon}|)\big)\big(1 - h_\varepsilon(|Z_t^{i,N,\varepsilon}|)^2\big) \\ &\quad + (c_1 + c_2)\Big(|\widetilde{\psi}_t^{i,N,\varepsilon}| + |\psi_t^{i,N,\varepsilon}| + \frac{K}{c_2 N}\sum_{j=1}^N f(|Z_t^{j,N,\varepsilon}|)\Big)\Big) dt. \end{aligned} \tag{2.15}$$

Because $(\widetilde{W}^i)_{i \in \mathbb{S}_N}$ are independent, as demonstrated in the proof of Lemma 2.1, and (2.2) is strongly well-posed, $(Y^{i,\varepsilon})_{i \in \mathbb{S}_N}$ are also independent. As a consequence, by following the line to drive [7, (22), p. 5396], we deduce from Lemma 2.2 with $p = 2$ (in case of $\lambda > 2K + L/2$) that for some constant $C_1 > 0$,

$$\mathbb{E}|\widetilde{\psi}_t^{i,N,\varepsilon}| \leq C_1 N^{-\frac{1}{2}} \mathbb{1}_{\{K>0\}}, \qquad t \geq 0. \tag{2.16}$$

Substituting this estimate into (2.15) yields for some constant $C_2 > 0$,

$$\frac{1}{N}\sum_{j=1}^{N}\mathrm{d}\mathbb{E}f(|Z_t^{j,N,\varepsilon}|)\le\Big(-\frac{\lambda^{**}}{N}\sum_{j=1}^{N}\mathbb{E}f(|Z_t^{j,N,\varepsilon}|)+C_2\mathbb{E}\Lambda_t^{N,\varepsilon}\Big)\,\mathrm{d}t,$$

where $\lambda^{**}:=\lambda^*-K(1+c_1/c_2)$, and

$$\Lambda_t^{N,\varepsilon}:=N^{-\frac{1}{2}}+\frac{1}{N}\sum_{j=1}^{N}\big(\lambda^*f(|Z_t^{j,N,\varepsilon}|)+\varphi(|Z_t^{j,N,\varepsilon}|)\big)\big(1-h_\varepsilon(|Z_t^{j,N,\varepsilon}|)^2\big)+\frac{1}{N}\sum_{j=1}^{N}|\psi_t^{i,N,\varepsilon}|.$$

Obviously, there is a positive constant K^* such that $\lambda^{**}>0$ for any $K\in[0,K^*]$. In what follows, we shall take $K\in[0,K^*]$ so that $\lambda^{**}>0$. Whereafter, an application of Gronwall's inequality yields that

$$\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}f(|Z_t^{j,N,\varepsilon}|)\le\mathrm{e}^{-\lambda^{**}t}\mathbb{E}f(|Z_0^{1,N,\varepsilon}|)+C_2\int_0^t\mathrm{e}^{-\lambda^{**}(t-s)}\mathbb{E}\Lambda_s^{N,\varepsilon}\,\mathrm{d}s,$$

where we also explored the prerequisite that $(X_0^{i,\varepsilon},X_{[-r_0,0]}^{i,N,\varepsilon})_{i\in\mathbb{S}_N}=(Y_0^i,Y_{[-r_0,0]}^{i,N})_{i\in\mathbb{S}_N}$ are distributed identically. Once more, with the help of $c_2r\le f(r)\le(c_1+c_2)r, r\ge0$, along with (2.7), there is a constant $C_3>0$ such that

$$\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}|Z_t^{j,N,\varepsilon}|\le C_3\Big(\mathrm{e}^{-\lambda^{**}t}\mathbb{W}_1(\mu,\nu_0)+\int_0^t\mathrm{e}^{-\lambda^{**}(t-s)}\mathbb{E}\Lambda_s^{N,\varepsilon}\,\mathrm{d}s\Big).\tag{2.17}$$

Next, according to the definition of h_ε, in addition to $f(0)=0$ and $c_2\le f'(r)\le c_1+c_2$ for $r\ge0$, it follows readily that

$$\begin{aligned}\big(\lambda^*f(r)+\varphi(r)\big)\big(1-h_\varepsilon(r)^2\big)&\le2(c_1+c_2)\big((\lambda^*+K+L/2)r+\phi(r)\big)(1-h_\varepsilon(r))\\&\le2(c_1+c_2)\big(2(\lambda^*+K+L/2)\varepsilon+\phi(2\varepsilon)\big):=\rho(\varepsilon).\end{aligned}$$

Whence, we infer that for some constant $C_4>0$,

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}|Z_t^{j,N,\varepsilon}|\le C_4\Big(&\mathrm{e}^{-\lambda^{**}t}\mathbb{W}_1(\mu,\nu_0)+\rho(\varepsilon)+N^{-\frac{1}{2}}\mathbb{1}_{\{K>0\}}\\&+\frac{1}{N}\sum_{j=1}^{N}\int_0^t\mathrm{e}^{-\lambda^{**}(t-s)}\mathbb{E}|\psi_s^{j,N,\varepsilon}|\,\mathrm{d}s\Big).\end{aligned}\tag{2.18}$$

By the aid of Lemma 2.1, $\mu_t^i=\mathscr{L}_{Y_t^{j,\varepsilon}}$ and $\nu_t^{i,N}=\mathscr{L}_{Y_t^{j,N,\varepsilon}}$ for each fixed $i\in\mathbb{S}_N$ and any $j\in\mathbb{S}_N$. Moreover, recall that $(Y_0^{i,\varepsilon},Y_{[-r_0,0]}^{i,N,\varepsilon})_{i\in\mathbb{S}_N}=(X_0^i,X_{[-r_0,0]}^{i,N})_{i\in\mathbb{S}_N}$ are independent and identically distributed. Therefore, we derive that

$$\mathbb{W}_1(\mu_t^i, \nu_t^{i,N}) \le \frac{1}{N}\sum_{j=1}^N \mathbb{E}|Z_t^{j,N,\varepsilon}|, \quad \forall i \in \mathbb{S}_N.$$

Thus, (2.18) enables us to derive that

$$\begin{aligned}&\mathbb{W}_1(\mu_t^i, \nu_t^{i,N})\\&\le C_4\Big(\mathrm{e}^{-\lambda^{**}t}\mathbb{W}_1(\mu, \nu_0) + \rho(\varepsilon) + N^{-\frac{1}{2}}\mathbb{1}_{\{K>0\}} + \frac{1}{N}\sum_{j=1}^N \int_0^t \mathrm{e}^{-\lambda^{**}(t-s)}\mathbb{E}|\psi_s^{j,N,\varepsilon}|\,\mathrm{d}s\Big).\end{aligned}$$

Subsequently, making use of Lemma 2.1 followed by approaching $\varepsilon \downarrow 0$, and applying the prerequisite that $(X^{i,N}_{[-r_0,0]})_{i\in\mathbb{S}_N}$ are independent and identically distributed yields that

$$\mathbb{W}_1(\mu_t^i, \nu_t^{i,N}) \le C_4\Big(\mathrm{e}^{-\lambda^{**}t}\mathbb{W}_1(\mu, \nu_0) + N^{-\frac{1}{2}}\mathbb{1}_{\{K>0\}} + \int_0^t \mathrm{e}^{-\lambda^{**}(t-s)}\mathbb{E}\big|\psi_s^{i,N}\big|\,\mathrm{d}s\Big),$$

in which

$$\psi_s^{i,N} = b(X_s^{i,N}, \widetilde{\mu}_s^N) - \widetilde{b}(X^{i,N}_{\theta_s}, \widetilde{\mu}^N_{\overline{\theta}_s}).$$

Finally, the whole proof is complete by choosing $K^*, L^* > 0$ such that $\lambda^*, \lambda^{**} > 0$ and $\lambda > 2K + L/2$ for all $K \in [0, K^*]$ and $L \in [0, L^*]$. $\square$

Proof of Corollary 1.5. In terms of Theorem 1.3, (1.21) follows immediately by taking $r_0 = 0$, $\widetilde{b} = b$, and $\theta_t = \overline{\theta}_t = t$. To show (1.21) for the McKean-Vlasov SDE (1.5) provided that Assumption (**H**) is imposed, it is sufficient to prove the strong well-posedness and check respectively Assumptions ($\mathbf{A}_1$) and ($\mathbf{A}_2$).

Under Assumption (**H**), (1.5) is strongly well-posed once $X_0 \in L^1(\Omega \to \mathbb{R}^d; \mathscr{F}_0, \mathbb{P})$; see, for instance, [25, Theorem 4.1]. Trivially, Assumption ($\mathbf{A}_2$) holds true. Next, by virtue of (1.16) with $\phi(r) = \lambda_0 r$ and b_1 being replaced by $\overline{b}_1$, it follows readily that for $x, y \in \mathbb{R}^d$ and $\mu \in \mathscr{P}_1(\mathbb{R}^d)$,

$$\begin{aligned}\langle x - y, b_1(x) - b_1(y)\rangle &= \langle x - y, \overline{b}_1(x) - \overline{b}_1(y)\rangle + \langle x - y, \widetilde{b}_1(x) - \widetilde{b}_1(y)\rangle\\&\le |x - y|\varphi(|x - y|) + \big((\lambda_0 + \lambda)\mathbb{1}_{\{|x-y|\le \ell_0\}} - \lambda\big)|x - y|^2.\end{aligned}$$

Owing to $\lim_{r\to\infty}\frac{\varphi(r)}{r} = 0$, there is an $r_0 > \ell_0$ such that $\varphi(r) \le \frac{1}{2}\lambda r, r \ge r_0$. Thus, we derive that

$$\langle x - y, b_1(x) - b_1(y)\rangle \le |x - y|\big(\varphi(|x - y|) + \lambda_0|x - y|\big)\mathbb{1}_{\{|x-y|\le r_0\}} - \frac{1}{2}\lambda|x - y|^2\mathbb{1}_{\{|x-y|>r_0\}}.$$

Therefore, (1.16) is verifiable for $\phi(r) = \varphi(r) + \lambda_0 r$, which obviously is increasing and satisfies $\phi(0) = 0$. This, together with Lipschitz property of b_0, ensures Assumption ($\mathbf{A}_1$). $\square$

3. Proof of Theorem 1.6

First of all, we demonstrate that the moment of the displacement for $(Y_t)_{t\geq 0}$, determined by (1.23), can be bounded by the length of time lag.

Lemma 3.1. *Assume (1.18) and $\beta > 0$. Then, there exist constants $C^*, L^* > 0$ such that for all $L \in [0, L^*]$ and $t \geq 0$,*

$$\mathbb{E}\big|Y_t - Y_{t-r_0}\big| \leq C^*(1+r_0)\big(1+\mathbb{E}\|\xi\|_\infty\big)r_0^{\frac{1}{2}}. \tag{3.1}$$

Proof. To achieve (3.1), we first show that there are constants $C_0^*, L^* > 0$ such that for all $L \in [0, L^*]$ and $t \geq 0$,

$$\mathbb{E}\big(|Y_t|^2\big|\mathscr{F}_0\big) \leq C_0^*\big(1+(1+r_0)\|\xi\|_\infty^2\big). \tag{3.2}$$

According to the variation-of-constants formula (see e.g. [23, Theorem 3.1]), we have for all $t \geq 0$,

$$Y_t = \Gamma_t\xi_0 - \beta\int_{-r_0}^{0}\Gamma_{t-r_0-s}\xi_s\,\mathrm{d}s + \beta\int_0^t\Gamma_{t-s}\alpha\,\mathrm{d}s + \int_0^t\Gamma_{t-s}\sigma\,\mathrm{d}W_s + \int_0^t\Gamma_{t-s}\sigma_0(Y_s)\mathrm{d}B_s, \tag{3.3}$$

where $(\Gamma_t)_{t\geq 0}$ solves the linear ODE with memory

$$\mathrm{d}\Gamma_t = -\beta\Gamma_{t-r_0}\,\mathrm{d}t, \quad t > 0$$

with the initial condition $\Gamma_0 = I_{d\times d}$ and $\Gamma_r = \mathbf{0}_{d\times d}, r \in [-r_0, 0)$. By Hölder's inequality and Itô's isometry, it follows that for any $\varepsilon > 0$ and $t \geq 0$,

$$\begin{aligned}\mathbb{E}\big(|Y_t|^2\big|\mathscr{F}_0\big) &\leq (1+\varepsilon)\int_0^t\|\Gamma_{t-s}\|_{\mathrm{op}}^2\mathbb{E}\big(\|\sigma_0(Y_s)\|_{\mathrm{HS}}^2\big|\mathscr{F}_0\big)\mathrm{d}s\\ &\quad + 8(1+1/\varepsilon)\Bigg(\|\Gamma_t\|_{\mathrm{op}}^2|\xi_0|^2 + \beta^2 r_0\int_{-r_0}^{0}\|\Gamma_{t-r_0-s}\|_{\mathrm{op}}^2|\xi_s|^2\,\mathrm{d}s\\ &\quad + \beta^2|\alpha|^2\Big(\int_0^t\|\Gamma_s\|_{\mathrm{op}}\,\mathrm{d}s\Big)^2 + \|\sigma\|_{\mathrm{HS}}^2\int_0^t\|\Gamma_s\|_{\mathrm{op}}^2\mathrm{d}s\Bigg).\end{aligned}$$

Let

$$\lambda^\star = \sup\big\{\mathrm{Re}(\lambda) : \lambda \in \mathbb{C}, \lambda + \beta\mathrm{e}^{-\lambda r_0} = 0\big\}.$$

It is easy to see that $\lambda^\star < 0$ thanks to $\beta > 0$. By invoking [2, Proposition A.1], for $\lambda_0 \in (0, -\lambda^\star)$, there exists a constant $C_{\lambda_0} > 0$ such that

$$\|\Gamma_t\|_{\mathrm{op}} \le C_{\lambda_0} \mathrm{e}^{-\lambda_0 t}, \quad t \ge 0. \tag{3.4}$$

This, together with (1.18), enables us to deduce that there exists a constant $C^\star_\varepsilon > 0$ such that

$$\mathbb{E}\big(|Y_t|^2\big|\mathscr{F}_0\big) \le (1+\varepsilon)^2 C_{\lambda_0}^2 L \int_0^t \mathrm{e}^{-2\lambda_0(t-s)}\mathbb{E}\big(|Y_s|^2\big|\mathscr{F}_0\big)\mathrm{d}s + C^\star_\varepsilon\big(1+(1+r_0)\|\xi\|_\infty^2\big).$$

Subsequently, the Gronwall inequality yields that

$$\mathbb{E}\big(|Y_t|^2\big|\mathscr{F}_0\big) \le C^\star_\varepsilon\big(1+(1+r_0)\|\xi\|_\infty^2\big)\bigg(1+(1+\varepsilon)^2 C_{\lambda_0}^2 L \int_0^t \mathrm{e}^{-(2\lambda_0-(1+\varepsilon)^2 c_{\lambda_0}^2 L)(t-s)}\,\mathrm{d}s\bigg).$$

Since there exists an $L^* > 0$ such that $2\lambda_0 - C_{\lambda_0}^2 L > 0$ so $2\lambda_0 - (1+\varepsilon)^2 C_{\lambda_0}^2 L > 0$ for all $L \in [0, L^*]$, the assertion (3.2) follows directly.

By invoking Hölder's inequality and Itô's isometry, we infer that for any $t \ge 0$,

$$\begin{aligned}
\mathbb{E}\big(|Y_t - Y_{t-r_0}|\big|\mathscr{F}_0\big) &\le \mathbb{E}\big(|Y_t - Y_{(t-r_0)^+}|\big|\mathscr{F}_0\big) + |\xi_{t-r_0} - \xi_0|\mathbb{1}_{[0,r_0]}(t) \\
&\le |\xi_{t-r_0} - \xi_0|\mathbb{1}_{[0,r_0]}(t) + |\alpha|\beta r_0 + \beta \int_{(t-r_0)^+}^t \mathbb{E}\big(|Y_{s-r_0}|\big|\mathscr{F}_0\big)\,\mathrm{d}s \\
&\quad + |\sigma|\mathbb{E}\big(|W_t - W_{(t-r_0)^+}|\big|\mathscr{F}_0\big) + \mathbb{E}\bigg(\bigg|\int_{(t-r_0)^+}^t \sigma_0(Y_s)\,\mathrm{d}B_s\bigg|\bigg|\mathscr{F}_0\bigg) \\
&\le |\xi_{t-r_0} - \xi_0|\mathbb{1}_{[0,r_0]}(t) + |\alpha|\beta r_0 + |\sigma|(dr_0)^{\frac{1}{2}} \\
&\quad + \beta \int_{(t-r_0)^+}^t \mathbb{E}\big(|Y_{s-r_0}|\big|\mathscr{F}_0\big)\,\mathrm{d}s + \bigg(\int_{(t-r_0)^+}^t \mathbb{E}\big(\|\sigma_0(Y_s)\|_{\mathrm{HS}}^2\big|\mathscr{F}_0\big)\,\mathrm{d}s\bigg)^{\frac{1}{2}}.
\end{aligned}$$

Thus, by taking the Lipschitz property of σ_0, along with (1.24) and (3.2), into account, we find from Hölder's inequality that for some constant $c^\star > 0$ and any $t \ge 0$,

$$\mathbb{E}\big(|Y_t - Y_{t-r_0}|\big|\mathscr{F}_0\big) \le |\xi_{t-r_0} - \xi_0|\mathbb{1}_{[0,r_0]}(t) + c^\star(1+r_0)(1+\|\xi\|_\infty)r_0^{\frac{1}{2}}.$$

Consequently, (3.1) is available by taking advantage of (1.24). □

By invoking Lemma 3.1, it is ready to carry out the

Proof of Theorem 1.6. To apply Theorem 1.3, we set $N = 1$, $W_t^1 = W_t$, $B_t^1 = B_t$, and $\widetilde{b}(x) = b(x) = \beta(\alpha - x)$, $x \in \mathbb{R}^d$, and take $\theta_t = t - r_0$ to fit in the framework (1.1) and (1.2). In this case, $b(Y_t) - \widetilde{b}(Y_{\theta_t}) = \beta(Y_t - Y_{t-r_0})$. Whereafter, by leveraging Theorem 1.3, there exist constants $C^*, \lambda^* > 0$ such that

$$\mathbb{W}_1\big(\mathscr{L}_{X_t},\mathscr{L}_{Y_t}\big)\le C^*\bigg(\mathrm{e}^{-\lambda^* t}\mathbb{W}_1\big(\mathscr{L}_{X_0},\mathscr{L}_{Y_0}\big)+\int_0^t \mathrm{e}^{-\lambda^*(t-s)}\mathbb{E}\big|Y_s-Y_{s-r_0}\big|\,\mathrm{d}s\bigg).$$

Whence, (1.25) is attainable by making use of (3.1) so the proof of Theorem 1.6 is complete. □

4. Proofs of Theorems 1.8, 1.11 and 1.12

In this section, we aim to complete proofs of Theorems 1.8, 1.11 and 1.12, respectively.

4.1. Proof of Theorem 1.8

To finish the proof of Theorem 1.8, we first show that, for any $p\ge 1$, the p-th moment of $(X_{n\delta}^{\delta,1,N},\cdots,X_{n\delta}^{\delta,N,N})_{n\ge 1}$, solving (1.10), is uniformly bounded.

Lemma 4.1. *Assume* $(\mathbf{A}_1)$ *with* $\phi(r)=\lambda_0 r$ *for some* $\lambda_0>0$ *and* $\lambda>2K$*. Then, for any* $p\ge 1$ *and* $\delta\in(0,\delta_p]$ *with*

$$\delta_p:=1\wedge\frac{1}{2(\lambda_0+K)}\wedge\frac{3(\lambda-2K)}{4p(1+K)^{\lfloor p/2\rfloor+1}},\tag{4.1}$$

there exists a constant $C_p^\star>0$ *such that for all* $n\ge 0$ *and* $i\in\mathbb{S}_N$*,*

$$\mathbb{E}\big|X_{n\delta}^{\delta,i,N}\big|^p\le \mathrm{e}^{-\lambda^* n\delta}\mathbb{E}\big|X_0^{\delta,i,N}\big|^p+C_p^*\tag{4.2}$$

in case of $\mathbb{E}\big|X_0^{\delta,i,N}\big|^p<\infty$*, where*

$$\lambda^*:=\frac{3(\lambda-2K)}{2(4+3\lambda-4K)}.$$

Proof. In the sequel, we would like to emphasize that all underlying constants are entirely unrelated to the step size, and let $\mathscr{F}_0^N$ be the σ-algebra generated by $X_0^{1,N},\cdots,X_0^{N,N}$. For any even integer $p\ge 2$ and $\delta\in(0,\delta_p]$, provided that there exists a constant $C_p^*>0$ such that

$$\mathbb{E}\big(|X_{n\delta}^{\delta,i,N}|^p\big|\mathscr{F}_0^N\big)\le \mathrm{e}^{-\lambda^* n\delta}|X_0^{\delta,i,N}|^p+C_p^*,\quad i\in\mathbb{S}_N,\ n\ge 1,\tag{4.3}$$

then, from Hölder's inequality and the inequality: $(a+b)^\theta\le a^\theta+b^\theta$ for $a,b>0$ and $\theta\in(0,1]$, we obtain that for any $p\in[1,2]$,

$$\mathbb{E}\big(|X_{n\delta}^{\delta,i,N}|^p\big|\mathscr{F}_0^N\big)\le\Big(\mathbb{E}\big(|X_{n\delta}^{\delta,i,N}|^2\big|\mathscr{F}_0^N\big)\Big)^{\frac{p}{2}}\le \mathrm{e}^{-\frac{p}{2}\lambda^* n\delta}|X_0^{\delta,i,N}|^p+\big(C_2^*\big)^{\frac{p}{2}},$$

and that for $p>2$ which is not an even number,

$$\begin{aligned}\mathbb{E}\big(\big|X_{n\delta}^{\delta,i,N}\big|^p\big|\mathscr{F}_0^N\big)&\le\Big(\mathbb{E}\big(\big|X_{n\delta}^{\delta,i,N}\big|^{2\lceil p/2\rceil}\big|\mathscr{F}_0^N\big)\Big)^{\frac{p}{2\lceil p/2\rceil}}\\&\le \mathrm{e}^{-\frac{p}{2\lceil p/2\rceil}\lambda^* n\delta}|X_0^{\delta,i,N}|^p+\big(C_{2\lceil p/2\rceil}^*\big)^{\frac{p}{2\lceil p/2\rceil}},\end{aligned}$$

where $\lceil\cdot\rceil$ means the ceiling function. Hence, (4.3) is still valid for the other setting, where the constants λ_p and C_p^* might be different accordingly. Thus, (4.2) follows from (4.3) and the property of conditional expectation.

Let $\triangle W_{n\delta}^i = W_{(n+1)\delta}^i - W_{n\delta}^i$. Based on the preceding analysis, it remains to demonstrate (4.3). It is easy to see (1.10) that

$$\begin{aligned}
\big|X_{(n+1)\delta}^{\delta,i,N}\big|^2 &= \big|X_{n\delta}^{\delta,i,N}\big|^2 - \big|X_{(n+1)\delta}^{\delta,i,N} - X_{n\delta}^{\delta,i,N}\big|^2 + 2\big\langle X_{(n+1)\delta}^{\delta,i,N} - X_{n\delta}^{\delta,i,N}, X_{(n+1)\delta}^{\delta,i,N}\big\rangle \\
&= \big|X_{n\delta}^{\delta,i,N}\big|^2 - \big|X_{(n+1)\delta}^{\delta,i,N} - X_{n\delta}^{\delta,i,N}\big|^2 + 2\delta\big\langle b(X_{(n+1)\delta}^{\delta,i,N}, \widetilde{\mu}_{n\delta}^{\delta,N}), X_{(n+1)\delta}^{\delta,i,N}\big\rangle \\
&\quad + 2\sigma\big\langle \triangle W_{n\delta}^i, X_{(n+1)\delta}^{\delta,i,N} - X_{n\delta}^{\delta,i,N}\big\rangle + 2\sigma\big\langle \triangle W_{n\delta}^i, X_{n\delta}^{\delta,i,N}\big\rangle \\
&\le \big|X_{n\delta}^{\delta,i,N}\big|^2 + 2\delta\big\langle b(X_{(n+1)\delta}^{\delta,i,N}, \widetilde{\mu}_{n\delta}^{\delta,N}), X_{(n+1)\delta}^{\delta,i,N}\big\rangle + \sigma^2\big|\triangle W_{n\delta}^i\big|^2 + 2\sigma\big\langle \triangle W_{n\delta}^i, X_{n\delta}^{\delta,i,N}\big\rangle,
\end{aligned} \tag{4.4}$$

where in the second identity we exploited the fact that

$$X_{(n+1)\delta}^{\delta,i,N} - X_{n\delta}^{\delta,i,N} = \delta b(X_{(n+1)\delta}^{\delta,i,N}, \widetilde{\mu}_{n\delta}^{\delta,N}) + \sigma\triangle W_{n\delta}^i,$$

and the last display is valid by making use of the fundamental inequality: $2ab \le a^2 + b^2$ for any $a, b \in \mathbb{R}$. Next, by means of (2.5), it follows that

$$\langle x, b(x,\mu)\rangle \le C_0 - \frac{1}{4}(3\lambda - 4K)|x|^2 + \frac{1}{2}K\mu(|\cdot|^2), \qquad x \in \mathbb{R}^d, \mu \in \mathscr{P}_1(\mathbb{R}^d), \tag{4.5}$$

for some constant $C_0 > 0$. Hence, we deduce from (4.4) and Jensen's inequality that

$$\begin{aligned}
\big|X_{(n+1)\delta}^{\delta,i,N}\big|^2 &\le \big|X_{n\delta}^{\delta,i,N}\big|^2 + 2\delta\Big(C_0 - \frac{1}{4}(3\lambda - 4K)|X_{(n+1)\delta}^{\delta,i,N}|^2 + \frac{1}{2}K\widetilde{\mu}_{n\delta}^{\delta,N}(|\cdot|^2)\Big) \\
&\quad + \sigma^2\big|\triangle W_{n\delta}^i\big|^2 + 2\sigma\big\langle \triangle W_{n\delta}^i, X_{n\delta}^{\delta,i,N}\big\rangle.
\end{aligned}$$

This obviously implies that

$$(1+\lambda_K\delta)\big|X_{(n+1)\delta}^{\delta,i,N}\big|^2 \le \big|X_{n\delta}^{\delta,i,N}\big|^2 + K\delta\widetilde{\mu}_{n\delta}^{\delta,N}(|\cdot|^2) + 2C_0\delta + \sigma^2\big|\triangle W_{n\delta}\big|^2 + 2\sigma\big\langle \triangle W_{n\delta}, X_{n\delta}^{\delta,i,N}\big\rangle, \tag{4.6}$$

in which $\lambda_K := (3\lambda - 4K)/2$.

According to (4.6), the binomial theorem gives that for any integer $p \ge 1$,

$$\begin{aligned}
(1+\lambda_K\delta)^p\big|X_{(n+1)\delta}^{\delta,i,N}\big|^{2p} &= \big(\big|X_{n\delta}^{\delta,i,N}\big|^2 + K\delta\widetilde{\mu}_{n\delta}^{\delta,N}(|\cdot|^2)\big)^p + p\big(\big|X_{n\delta}^{\delta,i,N}\big|^2 + K\delta\widetilde{\mu}_{n\delta}^{\delta,N}(|\cdot|^2)\big)^{p-1}U_{n\delta}^i \\
&\quad + \mathbb{1}_{\{p\ge 2\}}\sum_{k=0}^{p-2} C_p^k\big(\big|X_{n\delta}^{\delta,i,N}\big|^2 + K\delta\widetilde{\mu}_{n\delta}^{\delta,N}(|\cdot|^2)\big)^k (U_{n\delta}^i)^{p-k} \\
&=: \Gamma_{n\delta}^i(\mathbf{X}_{n\delta}^{\delta,N}) + \widehat{\Gamma}_{n\delta}^i(\mathbf{X}_{n\delta}^{\delta,N}) + \overline{\Gamma}_{n\delta}^i(\mathbf{X}_{n\delta}^{\delta,N}),
\end{aligned}$$

where

$$U_{n\delta}^{i} := 2C_0\delta + \sigma^2 \big|\triangle W_{n\delta}^{i}\big|^2 + 2\sigma\big\langle \triangle W_{n\delta}^{i}, X_{n\delta}^{\delta,i,N}\big\rangle \quad \text{and} \quad \mathbf{X}_{n\delta}^{\delta,N} := \big(X_{n\delta}^{\delta,1,N}, \cdots, X_{n\delta}^{\delta,N,N}\big). \tag{4.7}$$

Below, we attempt to estimate the terms $\Gamma_{n\delta}^{i}$, $\widehat{\Gamma}_{n\delta}^{i}$, as well as $\overline{\Gamma}_{n\delta}^{i}$, one by one. First of all, applying the binomial theorem and invoking Jensen's inequality and Young's inequality yields that

$$\begin{aligned}\Gamma_{n\delta}^{i}(\mathbf{X}_{n\delta}^{\delta,N}) &= \sum_{k=0}^{p} C_p^k \big|X_{n\delta}^{\delta,i,N}\big|^{2k}\big(K\delta\widetilde{\mu}_{n\delta}^{\delta,N}(|\cdot|^2)\big)^{p-k} \\ &\le \sum_{k=0}^{p} C_p^k (K\delta)^{p-k}\bigg(\frac{k}{p}\big|X_{n\delta}^{\delta,i,N}\big|^{2p} + \frac{p-k}{p}\widetilde{\mu}_{n\delta}^{\delta,N}(|\cdot|^{2p})\bigg).\end{aligned}$$

So, by utilizing the fact that $X_{n\delta}^{\delta,i,N}$ and $X_{n\delta}^{\delta,j,N}$ are identically distributed given $\mathscr{F}_0^N$, we conclude

$$\mathbb{E}\big(\Gamma_{n\delta}^{i}(\mathbf{X}_{n\delta}^{\delta,N})\big|\mathscr{F}_0^N\big) \le (1+K\delta)^p \mathbb{E}\big(|X_{n\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big).$$

Next, notice that

$$\mathbb{E}\big(\widehat{\Gamma}_{n\delta}^{i}(\mathbf{X}_{n\delta}^{\delta,N})\big|\mathscr{F}_0^N\big) = p\mathbb{E}\Big(\big(|X_{n\delta}^{\delta,i,N}|^2 + K\delta\widetilde{\mu}_{n\delta}^{\delta,N}(|\cdot|^2)\big)^{p-1}\big(2C_0\delta + \sigma^2\big|\triangle W_{n\delta}^{i}\big|^2\big)\big|\mathscr{F}_0^N\Big).$$

Whence, it is apparent that there exists a constant $C_p > 0$ such that

$$\mathbb{E}\big(\widehat{\Gamma}_{n\delta}^{i}(\mathbf{X}_{n\delta}^{\delta,N})\big|\mathscr{F}_0^N\big) \le \frac{1}{8}(\lambda - 2K)p\delta\,\mathbb{E}\big(|X_{n\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p\delta.$$

Once again, via Young's inequality and by utilizing the fact that $X_{n\delta}^{\delta,i,N}$ and $X_{n\delta}^{\delta,j,N}$ are identically distributed given $\mathscr{F}_0^N$, we infer that for some constant $C_p^{\star} > 0$,

$$\mathbb{E}\big(\overline{\Gamma}_{n\delta}^{i}(\mathbf{X}_{n\delta}^{\delta,N})\big|\mathscr{F}_0^N\big) \le \frac{1}{4}(\lambda - 2K)p\delta\,\mathbb{E}\big(|X_{n\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p^{\star}\delta,$$

where the underlying moment of the polynomial with respect to $|\triangle W_{n\delta}^{i}|$ provides at least the order δ. Now, summing up the previous estimates on $\Gamma_{n\delta}^{i}$, $\widehat{\Gamma}_{n\delta}^{i}$, and $\overline{\Gamma}_{n\delta}^{i}$, in addition to $(1+\lambda_K\delta)^p \ge 1 + p\lambda_K\delta$, enables us to derive that

$$(1+p\lambda_K\delta)\mathbb{E}\big(|X_{(n+1)\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) \le \big((1+K\delta)^p + 3(\lambda-2K)p\delta/8\big)\mathbb{E}\big(|X_{n\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p\delta.$$

The mean value theorem, beside the definition of δ_p given in (4.1), shows that for any $\delta \in (0, \delta_p)$,

$$\begin{aligned}(1+K\delta)^p &\le 1 + pK\delta + \frac{1}{2}(1+K)^{p-2}p(p-1)K^2\delta^2 \\ &\le 1 + pK\delta + \frac{3}{8}(\lambda - 2K)p\delta.\end{aligned}$$

As a consequence, for

$$\lambda_\delta^\star := \frac{1}{1+p\lambda_K\delta}\Big(1+pK\delta+\frac{3}{4}(\lambda-2K)p\delta\Big)\in(0,1)$$

based on the prerequisite $\lambda > 2K$, via an inductive argument, we arrive at

$$\mathbb{E}\big(|X_{(n+1)\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big)\le(\lambda_\delta^\star)^n|X_0^{\delta,i,N}|^{2p}+\frac{C_p^\star}{1-\lambda_\delta^\star}.$$

Subsequently, by invoking the inequality: $a^r \le \mathrm{e}^{-(1-a)r}$ for $a, r > 0$, we derive from $\delta\in(0,1)$ that

$$\mathbb{E}\big(|X_{(n+1)\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big)\le \mathrm{e}^{-\lambda_p^* n\delta}|X_0^{\delta,i,N}|^{2p}+\frac{C_p^\star}{\lambda_p^*},$$

where $\lambda_p^* := \frac{3p(\lambda-2K)}{2(2+p(3\lambda-4K))}$. Consequently, the assertion (4.3) is available by noting that $p\mapsto\lambda_p^*$ is increasing. □

With the help of Lemma 4.1, the proof of Theorem 1.8 can be implemented.

Proof of Theorem 1.8. Below, for nonnegative numbers a, b, we use the shorthand notation $a\lesssim b$ if there exists a constant $c>0$ such that $a\le cb$. Combining (1.17) with ($\mathbf{A}_3$) and $\delta\in(0,1)$, we derive from (1.10) that for any $t\ge0$,

$$\begin{aligned}\mathbb{E}\big(|X_t^{\delta,i,N}|^{2l^*}\big|\mathscr{F}_0^N\big)&\lesssim 1+\mathbb{E}\big(|X_{t_\delta}^{\delta,i,N}|^{2l^*}\big|\mathscr{F}_0^N\big)+\mathbb{E}\big(\big|b(X_{t_\delta+\delta}^{\delta,i,N},\widetilde{\mu}_{t_\delta}^{\delta,N})\big|^{2l^*}\big|\mathscr{F}_0^N\big)\\&\lesssim 1+\mathbb{E}\big(|X_{t_\delta}^{\delta,i,N}|^{2l^*}\big|\mathscr{F}_0^N\big)+\mathbb{E}\big(|X_{t_\delta+\delta}^{\delta,i,N}|^{2l^*(l^*+1)}\big|\mathscr{F}_0^N\big)\\&\quad+\frac{1}{N}\sum_{j=1}^N\mathbb{E}\big(|X_{t_\delta}^{\delta,j,N}|^{2l^*}\big|\mathscr{F}_0^N\big)\\&\lesssim 1+\mathbb{E}\big(|X_{t_\delta}^{\delta,i,N}|^{2l^*}\big|\mathscr{F}_0^N\big)+\mathbb{E}\big(|X_{t_\delta+\delta}^{\delta,i,N}|^{2l^*(l^*+1)}\big|\mathscr{F}_0^N\big)\\&\lesssim 1+|X_0^{\delta,i,N}|^{2l^*(l^*+1)},\end{aligned}\tag{4.8}$$

where in the penultimate inequality we used the fact that $(X_t^{\delta,j,N})_{j\in\mathbb{S}_N}$ are identically distributed given $\mathscr{F}_0^N$, and in the last display we applied Lemma 4.1. Again, by taking (1.17) and ($\mathbf{A}_3$) into consideration, along with $\delta\in(0,1)$, it follows from Lemma 4.1 that

$$\begin{aligned}\mathbb{E}\big(|X_t^{\delta,i,N}-X_{t_\delta}^{\delta,i,N}|^2\big|\mathscr{F}_0^N\big)&\lesssim\Big(1+\mathbb{E}\big(|X_{t_\delta+\delta}^{\delta,i,N}|^{2(l^*+1)}\big|\mathscr{F}_0^N\big)+\mathbb{E}\big(|X_{t_\delta}^{\delta,i,N}|^2\big|\mathscr{F}_0^N\big)\Big)\delta\\&\lesssim\big(1+|X_0^{\delta,i,N}|^{2(l^*+1)}\big)\delta\end{aligned}$$

and subsequently from (4.8) and Hölder's inequality that

$$\begin{aligned}&\mathbb{E}\big(|b(X_t^{\delta,i,N},\widetilde{\mu}_t^{\delta,N})-b(X_{t_\delta+\delta}^{\delta,i,N},\widetilde{\mu}_{t_\delta}^{\delta,N})|\big|\mathscr{F}_0^N\big)\\&\lesssim\Big(1+\big(\mathbb{E}\big(|X_t^{\delta,i,N}|^{2l^*}\big|\mathscr{F}_0^N\big)\big)^{\frac12}+\big(\mathbb{E}\big(|X_{t_\delta+\delta}^{\delta,i,N}|^{2l^*}\big|\mathscr{F}_0^N\big)\big)^{\frac12}\Big)\end{aligned}$$

$$\times \sup_{s-t\le 2\delta}\big(\mathbb{E}(|X_t^{\delta,i,N}-X_s^{\delta,i,N}|^2|\mathscr{F}_0^N)\big)^{\frac12} \tag{4.9}$$
$$+\frac1N\sum_{j=1}^N\mathbb{E}\big(|X_t^{\delta,j,N}-X_{t_\delta}^{\delta,j,N}||\mathscr{F}_0^N\big)$$
$$\lesssim \big(1+|X_0^{\delta,i,N}|^{(l^*+1)^2}\big)\delta^{\frac12}.$$

Next, applying Theorem 1.3 with $r_0=0$, $\theta_t=t_\delta+\delta$, $\overline{\theta}_t=t_\delta$, and $\widetilde{b}=b$ yields that for some constant $\lambda^*>0$,

$$\mathbb{W}_1\big(\mathscr{L}_{X_t^i},\mathscr{L}_{X_t^{\delta,i,N}}\big)\lesssim \mathrm{e}^{-\lambda^* t}\mathbb{W}_1\big(\mathscr{L}_{X_0^i},\mathscr{L}_{X_0^{\delta,i,N}}\big)+N^{-\frac12}$$
$$+\int_0^t \mathrm{e}^{-\lambda^*(t-s)}\mathbb{E}\big|b(X_s^{\delta,i,N},\widetilde{\mu}_s^{\delta,N})-b(X_{s_\delta+\delta}^{\delta,i,N},\widetilde{\mu}_{s_\delta}^{\delta,N})\big|\,\mathrm{d}s.$$

Whence, the desired assertion (1.27) follows from (4.9). □

4.2. *Proof of Theorem 1.11*

By following the line to handle the proof of Theorem 1.8, it is necessary to verify that $(X_{n\delta}^{\delta,1,N},\cdots,X_{n\delta}^{\delta,N,N})_{n\ge1}$, determined by (1.30), is uniformly bounded in the moment sense.

Lemma 4.2. *Assume Assumptions* $(\mathbf{A}_1)$ *with* $\phi(r)=\lambda_0 r$ *for some* $\lambda_0>0$ *and* $(\mathbf{A}_1')$*, and suppose further* $\kappa:=\alpha\lambda_{b_1}-2K>0$ *and* $\lambda>2K$*. Then, for any* $p\ge1$ *and* $\delta\in(0,\delta_\kappa^*]$ *with* δ_κ^* *being defined in* (1.31)*, there is a constant* $C_p^\star>0$ *such that*

$$\mathbb{E}\big|X_{n\delta}^{\delta,i,N}\big|^p\le C_p^\star\big(1+\mathbb{E}\big|X_0^{\delta,i,N}\big|^p\big),\quad n\ge1. \tag{4.10}$$

Proof. By tracing the proof of Lemma 4.1, it suffices to verify that, for any integer $p\ge3$ and $\delta\in(0,\delta_\kappa^*]$, there exists a constant $C_p^{\star\star}>0$ such that

$$\mathbb{E}\big(|X_{(n+1)\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big)\le(1-\delta/16)\mathbb{E}\big(|X_{n\delta}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big)+C_p^{\star\star}\delta \tag{4.11}$$

in order to achieve (4.10).

From (1.30), it can be seen readily that

$$\big|X_{(n+1)\delta}^{\delta,i,N}\big|^2=\big|X_{n\delta}^{\delta,i,N}\big|^2+\big(2\langle X_{n\delta}^{\delta,i,N},b_1^\delta(X_{n\delta}^{\delta,i,N})+\big(b_0*\widetilde{\mu}_{n\delta}^{\delta,N}\big)(X_{n\delta}^{\delta,i,N})\rangle+\delta\big|b_1^\delta(X_{n\delta}^{\delta,i,N})\big|^2\big)\delta$$
$$+\big\langle\big(b_0*\widetilde{\mu}_{n\delta}^{\delta,N}\big)(X_{n\delta}^{\delta,i,N})+2b_1^\delta(X_{n\delta}^{\delta,i,N}),\big(b_0*\widetilde{\mu}_{n\delta}^{\delta,N}\big)(X_{n\delta}^{\delta,i,N})\big\rangle\delta^2$$
$$+\big(\sigma^2\big|\triangle W_{n\delta}^i\big|^2+2\sigma\big\langle X_{n\delta}^{\delta,i,N}+b_1^\delta(X_{n\delta}^{\delta,i,N})\delta+\big(b_0*\widetilde{\mu}_{n\delta}^{\delta,N}\big)(X_{n\delta}^{\delta,i,N})\delta,\triangle W_{n\delta}^i\big\rangle\big)$$
$$=:\big|X_{n\delta}^{\delta,i,N}\big|^2+\Lambda^{i,\delta}(\mathbf{X}_{n\delta}^{\delta,N})\delta+\widehat{\Lambda}^{i,\delta}(\mathbf{X}_{n\delta}^{\delta,N})\delta^2+\overline{\Lambda}^{i,\delta}(\mathbf{X}_{n\delta}^{\delta,N}),$$

where $\mathbf{X}_{n\delta}^{\delta,N}$ was defined as in (4.7).

For any $\mathbf{x} := (x_1, \cdots, x_N) \in (\mathbb{R}^d)^N$, let $\mu_{\mathbf{x}}^N = \frac{1}{N}\sum_{j=1}^N \delta_{x_j}$. By invoking (1.17) and (1.28), we derive that for any $\delta \in (0, \delta_\kappa^*]$,

$$\begin{aligned}\Lambda^{i,\delta}(\mathbf{x}) &\le \frac{1}{1+\delta^{\frac{1}{2}}\|\nabla b_1(x_i)\|_{\mathrm{HS}}}\bigg(2\langle x_i, b_1(x_i)\rangle + \frac{\delta|b_1(x_i)|^2}{1+\delta^{\frac{1}{2}}\|\nabla b_1(x_i)\|_{\mathrm{HS}}}\bigg) + 2|x_i|\,\mu_{\mathbf{x}}^N(|b_0(x_i-\cdot)|)\\ &\le -\frac{2\|\nabla b_1(x_i)\|_{\mathrm{HS}}}{1+\delta^{\frac{1}{2}}\|\nabla b_1(x_i)\|_{\mathrm{HS}}}\big(\lambda_{b_1} - \widehat{\lambda}_{b_1}^2\delta^{\frac{1}{2}}\big)|x_i|^2 + 2\big(C_{b_1} + \widehat{C}_{b_1}^2\big)\\ &\quad + K\big(3|x_i|^2 + \mu_{\mathbf{x}}^N(|\cdot|^2)\big) + 2|x_i|\cdot|b_0(\mathbf{0})|\end{aligned}$$

and that there exists a constant $C_0^* > 0$ such that

$$\begin{aligned}\widehat{\Lambda}^{i,\delta}(\mathbf{x}) &= \big|\mu_{\mathbf{x}}^N(b_0(x_i-\cdot))\big|^2 + \frac{2\langle b_1(x_i), \mu_{\mathbf{x}}^N(b_0(x_i-\cdot))\rangle}{1+\delta^{\frac{1}{2}}\|\nabla b_1(x_i)\|_{\mathrm{HS}}}\\ &\le 4K^2\big(|x_i|^2 + \mu_{\mathbf{x}}^N(|\cdot|^2)\big) + 2|b_0(\mathbf{0})|^2\\ &\quad + 2\big(\widehat{\lambda}_{b_1}\delta^{-\frac{1}{2}}|x_i| + \widehat{C}_{b_1}\big)\big(K|x_i| + K\mu_{\mathbf{x}}^N(|\cdot|) + |b_0(\mathbf{0})|\big)\\ &\le 4K\big(K + \widehat{\lambda}_{b_1}\delta^{-\frac{1}{2}}\big)\big(|x_i|^2 + \mu_{\mathbf{x}}^N(|\cdot|^2)\big) + C_0^*(1+\delta^{-\frac{1}{2}}).\end{aligned}$$

Thus, combining (1.29) with $\delta \in (0,1)$, in addition to the local boundedness of ∇b_1, yields that for some constant $C_1^* > 0$,

$$\begin{aligned}\Lambda^{i,\delta}(\mathbf{x})\delta + \widehat{\Lambda}^{i,\delta}(\mathbf{x})\delta^2 &\le \bigg(-\frac{2\alpha}{1+\delta^{\frac{1}{2}}\alpha}\Big(\lambda_{b_1} - \frac{1}{2\alpha}\big(3K + \rho\delta^{\frac{1}{2}}\big)(1+\delta^{\frac{1}{2}}\alpha) - \widehat{\lambda}_{b_1}^2\delta^{\frac{1}{2}} - \frac{\kappa}{4\alpha}\Big)|x_i|^2\\ &\quad + (K + \rho\delta^{\frac{1}{2}})\mu_{\mathbf{x}}^N(|\cdot|^2)\bigg)\delta + C_1^*\delta\\ &\le \bigg(-\frac{2\alpha}{1+\delta^{\frac{1}{2}}\alpha}\Big(\frac{3\kappa}{4\alpha} - \big(2K + \rho(1+1/\alpha) + \widehat{\lambda}_{b_1}^2\big)\delta^{\frac{1}{2}}\Big)|x_i|^2\\ &\quad + (K + \rho\delta^{\frac{1}{2}})\big(\mu_{\mathbf{x}}^N(|\cdot|^2) - |x_i|^2\big)\bigg)\delta + C_1^*\delta,\end{aligned}$$

where $\rho := 4K(K + \widehat{\lambda}_{b_1})$. In terms of the definition of δ_κ^*, we right now have for any $\delta \in (0, \delta_\kappa^*]$,

$$(2K + \rho(1+1/\alpha) + \widehat{\lambda}_{b_1}^2)\delta^{\frac{1}{2}} \le \frac{\kappa}{2\alpha}.$$

Whence, owing to $\delta^{\frac{1}{2}}\alpha \in (0,1)$ for $\delta \in (0, \delta_\kappa^*]$, we infer that

$$\Lambda^{i,\delta}(\mathbf{x})\delta + \widehat{\Lambda}^{i}(\mathbf{x})\delta^2 \le \big(-(\kappa/4 + \beta_\delta)|x_i|^2 + \beta_\delta\mu_{\mathbf{x}}^N(|\cdot|^2)\big)\delta + C_1^*\delta,$$

where $\beta_\delta := K + \rho\delta^{\frac{1}{2}}$. Whereafter, the preceding estimate enables us to deduce that

$$\big|X^{\delta,i,N}_{(n+1)\delta}\big|^2\le\big(1-(\kappa/4+\beta_\delta)\delta\big)\big|X^{\delta,i,N}_{n\delta}\big|^2+\beta_\delta\widetilde{\mu}^{\delta,N}_{n\delta}(|\cdot|^2)\delta+C_1^*\delta+\overline{\Lambda}^{i,\delta}(\mathbf{X}^\delta_{n\delta}),\quad(4.12)$$

where the factor $1-(\kappa/4+\beta_\delta)\delta$ is positive by taking $\delta\in(0,\delta^*_\kappa]$ into consideration.

With (4.12) at hand, we obtain that for any integer $p\ge3$,

$$\begin{aligned}\big|X^{\delta,i,N}_{(n+1)\delta}\big|^{2p}&\le\Big(\big(1-(\kappa/4+\beta_\delta)\delta\big)\big|X^{\delta,i,N}_{n\delta}\big|^2+\beta_\delta\widetilde{\mu}^{\delta,N}_{n\delta}(|\cdot|^2)\delta\Big)^p\\&\quad+p\Big(\big(1-(\kappa/4+\beta_\delta)\delta\big)\big|X^{\delta,i,N}_{n\delta}\big|^2+\beta_\delta\widetilde{\mu}^{\delta,N}_{n\delta}(|\cdot|^2)\delta\Big)^{p-1}\big(C_1^*\delta+\overline{\Lambda}^{i,\delta}(\mathbf{X}^\delta_{n\delta})\big)\\&\quad+\sum_{k=0}^{p-2}C_p^k\Big(\big(1-(\kappa/4+\beta_\delta)\delta\big)\big|X^{\delta,i,N}_{n\delta}\big|^2+\beta_\delta\widetilde{\mu}^{\delta,N}_{n\delta}(|\cdot|^2)\delta\Big)^k\big(C_1^*\delta+\overline{\Lambda}^{i,\delta}(\mathbf{X}^\delta_{n\delta})\big)^{p-k}\\&=:\Upsilon^i_{n\delta}(\mathbf{X}^{\delta,N}_{n\delta})+\widehat{\Upsilon}^i_{n\delta}(\mathbf{X}^{\delta,N}_{n\delta})+\overline{\Upsilon}^i_{n\delta}(\mathbf{X}^{\delta,N}_{n\delta}).\end{aligned}$$

In the sequel, we aim to estimate separately the conditional expectations of $\Upsilon^i_{n\delta}$, $\widehat{\Upsilon}^i_{n\delta}$, as well as $\overline{\Upsilon}^i_{n\delta}$ given the σ-algebra $\mathscr{F}^N_0$, which is generated by $X^{\delta,1,N}_0,\cdots,X^{\delta,N,N}_0$.

In the first place, the binomial theorem and the Young inequality yield that

$$\begin{aligned}\mathbb{E}\big(\Upsilon^i_{n\delta}(\mathbf{X}^{\delta,N}_{n\delta})\big|\mathscr{F}^N_0\big)&=\sum_{k=0}^pC_p^k\big(1-(\kappa/4+\beta_\delta)\delta\big)^k(\beta_\delta\delta)^{p-k}\big|X^{\delta,i,N}_{n\delta}\big|^{2k}\big(\widetilde{\mu}^{\delta,N}_{n\delta}(|\cdot|^2)\big)^{(p-k)}\\&\le\sum_{k=0}^pC_p^k\big(1-(\kappa/4+\beta_\delta)\delta\big)^k(\beta_\delta\delta)^{p-k}\\&\quad\times\Big(\frac{k}{p}\mathbb{E}\big(|X^{\delta,i,N}_{n\delta}|^{2p}\big|\mathscr{F}^N_0\big)+\frac{p-k}{p}\mathbb{E}\big(\widetilde{\mu}^{\delta,N}_{n\delta}(|\cdot|^{2p})\big|\mathscr{F}^N_0\big)\Big)\\&=(1-\kappa\delta/4)^p\mathbb{E}\big(|X^{\delta,i,N}_{n\delta}|^{2p}\big|\mathscr{F}^N_0\big)\\&\le(1-\kappa\delta/4)\mathbb{E}\big(|X^{\delta,i,N}_{n\delta}|^{2p}\big|\mathscr{F}^N_0\big),\end{aligned}\quad(4.13)$$

where in the second identity we used the fact that $X^{\delta,i,N}_{n\delta}$ and $X^{\delta,j,N}_{n\delta}$ are identically distributed given $\mathscr{F}^N_0$, and the last display is evident thanks to $1-\kappa\delta/4\in(0,1)$. In the next place, due to $\mathbb{E}(\triangle W^i_{n\delta}|\mathscr{F}^N_0)=0$ and $\mathbb{E}(|\triangle W^i_{n\delta}|^2|\mathscr{F}^N_0)=d\delta$, it follows from Young's inequality that there is a constant $C_2^*>0$ such that

$$\begin{aligned}\mathbb{E}\big(\widehat{\Upsilon}^i_{n\delta}(\mathbf{X}^{\delta,N}_{n\delta})\big|\mathscr{F}^N_0\big)&\le pC_2^*\delta\mathbb{E}\Big(\Big(\big(1-(\kappa/4+\beta_\delta)\delta\big)\big|X^{\delta,i,N}_{n\delta}\big|^2+\beta_\delta\widetilde{\mu}^{\delta,N}_{n\delta}(|\cdot|^2)\delta\Big)^{p-1}\Big|\mathscr{F}^N_0\Big)\\&\le\frac{\kappa\delta}{16}\mathbb{E}\big(|X^{\delta,i,N}_{n\delta}|^{2p}\big|\mathscr{F}^N_0\big)+C_2^*\delta.\end{aligned}\quad(4.14)$$

Furthermore, note from (1.28) and (1.17) that

$$\big|b_1^\delta(x_i)+\mu^N_{\mathbf{x}}(b_0(x_i-\cdot))\big|\delta\le\big(\widehat{\lambda}_{b_1}\delta^{-\frac12}|x_i|+\widehat{C}_{b_1}+K(|x_i|+\mu^N_{\mathbf{x}}(|\cdot|))+|b_0(\mathbf{0})|\big)\delta.$$

This, together with the fact that the conditional expectation (given $\mathscr{F}^N_0$) of the increment $\triangle W^i_{n\delta}$ contributes at least the order δ, and the Young inequality, leads to

$$\mathbb{E}\big(\overline{\Upsilon}^i_{n\delta}(\mathbf{X}^{\delta,N}_{n\delta})\big|\mathscr{F}^N_0\big) \le \frac{\kappa\delta}{16}\mathbb{E}\big(|X^{\delta,i,N}_{n\delta}|^{2p}\big|\mathscr{F}^N_0\big) + C^*_3\delta \tag{4.15}$$

for some constant $C^*_3 > 0$. Ultimately, (4.11) is reachable by pulling together (4.13), (4.14) and (4.15). □

Based on Lemma 4.2, it's turn to carry out the

Proof of Theorem 1.11. Applying Theorem 1.3 with $r_0 = 0$, $\theta_t = \overline{\theta}_t = t_\delta$, and $\widetilde{b}(x,\mu) = b^\delta_1(x) + (b_0 * \mu)(x)$, respectively, enables us to derive that for some $\lambda^* > 0$ and any $t \ge 0$,

$$\begin{aligned}\mathbb{W}_1\big(\mathscr{L}_{X^i_t}, \mathscr{L}_{X^{\delta,i,N}_t}\big) \lesssim{}& e^{-\lambda^* t}\mathbb{W}_1\big(\mathscr{L}_{X^i_0}, \mathscr{L}_{X^{\delta,i,N}_0}\big) + N^{-\frac{1}{2}}\mathbb{1}_{\{K>0\}} \\ &+ \int_0^t e^{-\lambda^*(t-s)}\mathbb{E}\big|b(X^{\delta,i,N}_s, \widetilde{\mu}^{\delta,N}_s) - \widetilde{b}(X^{\delta,i,N}_{s_\delta}, \widetilde{\mu}^{\delta,N}_{s_\delta})\big|\,ds.\end{aligned} \tag{4.16}$$

Next, by using (1.17), and (1.26), it is easy to find that for $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathscr{P}_1(\mathbb{R}^d)$,

$$\begin{aligned}|b(x,\mu) - \widetilde{b}(y,\nu)| &\le |b_1(x) - b_1(y)| + |b_1(y) - b^\delta_1(y)| + |(b_0 * \mu)(x) - (b_0 * \nu)(y)| \\ &\le \int_0^1 \langle \nabla b_1(y + s(x-y)), x - y\rangle\,ds + \frac{\delta^{\frac{1}{2}}|b_1(y)| \cdot \|\nabla b_1(y)\|_{\rm HS}}{1 + \delta^{\frac{1}{2}}\|\nabla b_1(y)\|_{\rm HS}} \\ &\quad + K\big(\mathbb{W}_1(\mu,\nu) + |x-y|\big) \\ &\lesssim \big(1 + |x|^{l^*} + |y|^{l^*}\big)|x-y| + \mathbb{W}_1(\mu,\nu) + (1 + |y|^{2l^*+1})\delta^{\frac{1}{2}}.\end{aligned}$$

This, together with (4.16), implies that

$$\begin{aligned}&\mathbb{W}_1\big(\mathscr{L}_{X^i_t}, \mathscr{L}_{X^{\delta,i,N}_t}\big) \\ &\lesssim e^{-\lambda^* t}\mathbb{W}_1\big(\mathscr{L}_{X^i_0}, \mathscr{L}_{X^{\delta,i,N}_0}\big) + N^{-\frac{1}{2}}\mathbb{1}_{\{K>0\}} \\ &\quad + \int_0^t e^{-\lambda^*(t-s)}\mathbb{E}\Big(\mathbb{E}\big((1 + |X^{\delta,i,N}_s|^{l^*} + |X^{\delta,i,N}_{s_\delta}|^{l^*})|X^{\delta,i,N}_s - X^{\delta,i,N}_{s_\delta}|\big|\mathscr{F}^N_0\big)\Big)\,ds \\ &\quad + \int_0^t e^{-\lambda^*(t-s)}\mathbb{E}\big|X^{\delta,i,N}_s - X^{\delta,i,N}_{s_\delta}\big|\,ds + \frac{1}{N}\sum_{j=1}^N \int_0^t e^{-\lambda^*(t-s)}\mathbb{E}\big|X^{\delta,j,N}_s - X^{\delta,j,N}_{s_\delta}\big|\,ds \\ &\quad + \delta^{\frac{1}{2}}\int_0^t e^{-\lambda^*(t-s)}\big(1 + \mathbb{E}|X^{\delta,i,N}_{s_\delta}|^{2l^*+1}\big)\,ds,\end{aligned} \tag{4.17}$$

where σ-algebra $\mathscr{F}^N_0$ is generated by $X^{\delta,1,N}_0, \cdots, X^{\delta,N,N}_0$. Moreover, due to $\mathbb{E}|W^i_t - W^2_{t_\delta}| = d(t - t_\delta)$ and $|b^\delta_1(x)| \lesssim 1 + \delta^{-\frac{1}{2}}|x|, x \in \mathbb{R}^d$, by invoking (1.28), it holds from Lemma 4.2 that

$$\mathbb{E}\big(|X_t^{\delta,i,N}-X_{t_\delta}^{\delta,i,N}|^2\big|\mathscr{F}_0^N\big)\lesssim\big(1+\mathbb{E}\big(|X_{t_\delta}^{\delta,i,N}|^2\big|\mathscr{F}_0^N\big)\big)\delta\lesssim\big(1+|X_0^{\delta,i,N}|^2\big)\delta$$

and

$$\mathbb{E}\big(|X_t^{\delta,i,N}|^{2l^*}|\mathscr{F}_0^N\big)\lesssim 1+\mathbb{E}\big(|X_{t_\delta}^{\delta,i,N}|^{2l^*}|\mathscr{F}_0^N\big)\lesssim 1+|X_0^{\delta,i,N}|^{2l^*}.$$

As a consequence, the assertion (1.32) is verifiable so the proof of Theorem 1.11 is finished from (4.17) and Hölder's inequality. □

4.3. *Proof of Theorem 1.12*

The following lemma addresses the issue that the time grid associated with the adaptive EM scheme tends to infinity almost surely.

Lemma 4.3. *Under Assumptions* ($\mathbf{A}_1$) *and* ($\mathbf{A}_3$),

$$\mathbb{P}\Big\{\omega\in\Omega:\lim_{n\to\infty}t_n(\omega)=+\infty\Big\}=1. \tag{4.18}$$

Proof. Let for all $x\in\mathbb{R}^d$ and $\mu\in\mathscr{P}(\mathbb{R}^d)$,

$$h(x,\mu)=(1+|b(x,\mu)|^2)^{-1}.$$

Then, it is easy to see that for $x\in\mathbb{R}^d$ and $\mu\in\mathscr{P}(\mathbb{R}^d)$,

$$|b(x,\mu)|(1+|b(x,\mu)|)h(x,\mu)\le 3/2.$$

Next, from (1.17) and ($\mathbf{A_3}$), there is a constant $C_1>0$ such that for all $x\in\mathbb{R}^d$ and $\mu\in\mathscr{P}_1(\mathbb{R}^d)$,

$$h(x,\mu)\big(1+|x|^{2l^*+1}+\mu(|\cdot|)^2\big)\le C.$$

Furthermore, note that Assumption (**T2**) in [21, Proposition 4.1] can be weakened as below:

$$\langle x,b(x,\mu)\rangle\le C_1\big(1+|x|^2+\mu(|\cdot|)^2\big),\quad x\in\mathbb{R}^d,\ \mu\in\mathscr{P}_1(\mathbb{R}^d).$$

Consequently, the assertion (4.18) follows from [21, Proposition 4.1] and (4.5). □

Unlike the backward/tamed EM scheme, the step size involved is a constant so an inductive argument can be used to establish the uniform moment bound. Nevertheless, with regard to the adaptive EM scheme (1.14), the underlying step size is an adaptive stochastic process. Hence, the approach adopted in treating Lemmas 4.1 and 4.2 no longer works to handle the uniform moment bound of the adaptive EM scheme (1.14), which is stated as a lemma below.

Lemma 4.4. *Assume* ($\mathbf{A}_1$) *with* $\phi(r)=\lambda_0 r$ *for some* $\lambda_0>0$ *and* $\lambda>2K$. *Then, for any* $p\ge1$, $i\in\mathbb{S}_N$ *and* $\delta\in(0,1)$, *there is a constant* $C_p^*>0$ *such that*

$$\mathbb{E}\big|X_t^{\delta,i,N}\big|^p\le C_p^*\big(1+\mathbb{E}\big|X_0^{\delta,i,N}\big|^p\big),\quad t\ge0. \tag{4.19}$$

Proof. By following partly the proof of Lemma 4.1, for each integer $p \geq 3$, it is sufficient to show that there exist constants $\lambda_p, C_p^* > 0$ such that

$$\mathbb{E}\big(|X_t^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) \leq C_p^*\big(1 + \mathrm{e}^{-\lambda_p t}\big|X_0^{\delta,i,N}\big|^{2p}\big), \quad i \in \mathbb{S}_N,\ t \geq 0 \tag{4.20}$$

for the sake of validity of (4.19).

Applying Itô's formula, we obtain from (1.14) that for $\lambda_p := p(\lambda - 2K)/6$ and integer $p \geq 3$,

$$\mathrm{d}\big(\mathrm{e}^{\lambda_p t}|X_t^{\delta,i,N}|^{2p}\big) = \mathrm{e}^{\lambda_p t}\big(\lambda_p|X_t^{\delta,i,N}|^{2p} + \Phi^{i,p}(\mathbf{X}_t^{\delta,N}) + c_p|X_t^{\delta,i,N}|^{2(p-1)}\big)\mathrm{d}t + \mathrm{d}M_t^{(p)}, \tag{4.21}$$

where $(M_t^{(p)})_{t\geq 0}$ is a martingale, $c_p := \sigma^2 p(d + 2(p-1))$, and

$$\Phi^{i,p}\big(\mathbf{X}_t^{\delta,N}\big) := 2p|X_t^{\delta,i,N}|^{2(p-1)}\big\langle X_t^{\delta,i,N}, b(X_{\underline{t}}^{\delta,i,N}, \widetilde{\mu}_{\underline{t}}^{\delta,N})\big\rangle.$$

By taking the structure of the adaptive step size defined in (1.33) into consideration, besides $\delta \in (0,1)$, it is easy to see that

$$\big|b(X_{\underline{t}}^{\delta,i,N}, \widetilde{\mu}_{\underline{t}}^{\delta,N})\big|(t - \underline{t}) \leq \delta^{\frac{1}{2}}(t - \underline{t})^{\frac{1}{2}} \leq 1. \tag{4.22}$$

Whence, by making use of the strong Markov property of $(W_t^i)_{t\geq 0}$ and the tower property of conditional expectations, in addition to $t - \underline{t} \leq 1$, there exist constants $C_p^{*,1}, C_p^{*,2} > 0$ such that

$$\begin{aligned}\mathbb{E}\big(|X_t^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) &\leq \frac{3}{2}\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p^{*,1}\big(1 + \mathbb{E}\big(|W_t^i - W_{\underline{t}}^i|^{2p}\big|\mathscr{F}_0^N\big)\big)\\ &= \frac{3}{2}\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p^{*,1}\big(1 + \mathbb{E}\big(\mathbb{E}\big(|W_t^i - W_{\underline{t}}^i|^{2p}\big|\mathscr{F}_{\underline{t}}^N\big)\big|\mathscr{F}_0^N\big)\big)\\ &\leq \frac{3}{2}\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p^{*,2},\end{aligned} \tag{4.23}$$

where $\mathscr{F}_{\underline{t}}^N$ is the σ-algebra of $\underline{t}$-past with $(\mathscr{F}_t^N)_{t\geq 0}$ being the σ-algebra generated by $(X_0^{\delta,1,N}, \cdots, X_0^{\delta,N,N})$ and $(W_t^1, \cdots, W_t^N)_{t\geq 0}$. Then, provided that we claim that there is a constant $C_p^{*,3} > 0$ satisfying

$$\mathbb{E}\big(\Phi^{i,p}(\mathbf{X}_t^{\delta,N})\big|\mathscr{F}_0^N\big) \leq -\frac{1}{2}p(\lambda - 2K)\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p^{*,3}, \tag{4.24}$$

combining (4.21) with (4.23) yields that for some positive constants $C_p^{*,4}, C_p^{*,5}$,

$$\begin{aligned}\mathrm{e}^{\lambda_p t}\mathbb{E}\big(|X_t^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) &\leq \big|X_0^{\delta,i,N}\big|^{2p} + \int_0^t \mathrm{e}^{\lambda_p s}\Big(2\lambda_p\mathbb{E}\big(|X_s^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big)\\ &\qquad - \frac{1}{2}p(\lambda - 2K)\mathbb{E}\big(|X_{\underline{s}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p^{*,4}\Big)\,\mathrm{d}s\\ &\leq \big|X_0^{\delta,i,N}\big|^{2p} + C_p^{*,5}(\mathrm{e}^{\lambda_p t} - 1).\end{aligned}$$

Therefore, (4.20) follows directly.

Below, we attempt to verify (4.24). By invoking (1.14) once more, it is apparent to see that

$$
\begin{aligned}
|X_t^{\delta,i,N}|^{2(p-1)} &= \big|X_{\underline{t}}^{\delta,i,N} + b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})(t-\underline{t}) + \sigma(W_t^i - W_{\underline{t}}^i)\big|^{2(p-1)} \\
&= \big|X_{\underline{t}}^{\delta,i,N}\big|^{2(p-1)} + \sum_{k=0}^{p-2} C_{p-1}^k \big|X_{\underline{t}}^{\delta,i,N}\big|^{2k}\big(\Psi^i(\mathbf{X}_{\underline{t}}^{\delta,N})\big)^{p-1-k}
\end{aligned}
\tag{4.25}
$$

and that

$$
\big\langle X_t^{\delta,i,N}, b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big\rangle = \big\langle X_{\underline{t}}^{\delta,i,N}, b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big\rangle + \Upsilon^i(\mathbf{X}_{\underline{t}}^{\delta,N}),
\tag{4.26}
$$

where

$$
\begin{aligned}
\Psi^i(\mathbf{X}_{\underline{t}}^{\delta,N}) :&= \big|b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big|^2(t-\underline{t})^2 + \sigma^2\big|W_t^i - W_{\underline{t}}^i\big|^2 \\
&\quad + 2\big\langle X_{\underline{t}}^{\delta,i,N}, b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big\rangle(t-\underline{t}) + 2\sigma\big\langle X_{\underline{t}}^{\delta,i,N}, W_t^i - W_{\underline{t}}^i\big\rangle \\
&\quad + 2\sigma\big\langle b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N}), W_t^i - W_{\underline{t}}^i\big\rangle(t-\underline{t}), \\
\Upsilon^i(\mathbf{X}_{\underline{t}}^{\delta,N}) :&= \big|b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big|^2(t-\underline{t}) + \sigma\big\langle W_t^i - W_{\underline{t}}^i, b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big\rangle.
\end{aligned}
$$

Now, plugging (4.25) and (4.26) back into (4.21) gives that

$$
\begin{aligned}
\Phi^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N}) &= 2p\big|X_{\underline{t}}^{\delta,i,N}\big|^{2(p-1)}\big\langle X_{\underline{t}}^{\delta,i,N}, b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big\rangle + 2p\big|X_{\underline{t}}^{\delta,i,N}\big|^{2(p-1)}\Upsilon^i(\mathbf{X}_t^{\delta,N}) \\
&\quad + 2p\big(\big\langle X_{\underline{t}}^{\delta,i,N}, b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big\rangle + \Upsilon^i(\mathbf{X}_t^{\delta,N})\big)\sum_{k=0}^{p-2} C_{p-1}^k\big|X_{\underline{t}}^{\delta,i,N}\big|^{2k}\big(\Psi^i(\mathbf{X}_t^{\delta,N})\big)^{p-1-k} \\
&=: \Theta^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N}) + \widehat{\Theta}^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N}) + \overline{\Theta}^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N}).
\end{aligned}
$$

By utilizing (4.5), the leading term $\Theta^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N})$ can be tackled as follows:

$$
\begin{aligned}
\mathbb{E}\big(\Theta^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N})\big|\mathscr{F}_0^N\big) &\le 2p\Big(-\frac{1}{4}(3\lambda - 4K)\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) \\
&\quad + \frac{1}{2}K\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2(p-1)}\widetilde{\mu}_{\underline{t}}^{\delta,N}(|\cdot|^2)\big|\mathscr{F}_0^N\big) + C_0\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2(p-1)}\big|\mathscr{F}_0^N\big)\Big) \\
&\le -p(\lambda - 2K)\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big) + C_p^{*,6}
\end{aligned}
\tag{4.27}
$$

for some constant $C_p^{*,6} > 0$, where we also employed that $X_{\underline{t}}^{\delta,i,N}$ and $X_{\underline{t}}^{\delta,j,N}$ are distributed identically given $\mathscr{F}_0^N$. Subsequently, in view of

$$
\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2(p-1)}\big\langle W_t^i - W_{\underline{t}}^i, b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big\rangle\big|\mathscr{F}_0^N\big) = 0,
$$

the Young inequality implies that for some constant $C_p^{*,7} > 0$,

$$\mathbb{E}\big(\widehat{\Theta}^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N})\big|\mathscr{F}_0^N\big) \le \frac{1}{4}p(\lambda-2K)\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big)+C_p^{*,7}>0. \tag{4.28}$$

Furthermore, by virtue of (4.22), we obviously obtain that

$$\Psi^i(\mathbf{X}_{\underline{t}}^{\delta,N}) \le t-\underline{t}+\sigma^2\big|W_t^i-W_{\underline{t}}^i\big|^2+2(t-\underline{t})^{\frac{1}{2}}\big|X_{\underline{t}}^{\delta,i,N}\big|+2|\sigma|\big(1+\big|X_{\underline{t}}^{\delta,i,N}\big|\big)\big|W_t^i-W_{\underline{t}}^i\big|$$

and

$$\Upsilon^i(\mathbf{X}_{\underline{t}}^{\delta,N}) \le 1+|\sigma|\cdot\big|b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big|\cdot\big|W_t^i-W_{\underline{t}}^i\big|.$$

Thus, we find that

$$\begin{aligned}\mathbb{E}\big(\overline{\Theta}^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N})\big|\mathscr{F}_0^N\big) &\le 2p\big(1+\big(|X_{\underline{t}}^{\delta,i,N}|+|\sigma||W_t^i-W_{\underline{t}}^i|\big)\big|b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big|\big)\\ &\quad\times\sum_{k=0}^{p-2}C_{p-1}^k\big|X_{\underline{t}}^{\delta,i,N}\big|^{2k}\Big(t-\underline{t}+\sigma^2|W_t^i-W_{\underline{t}}^i|^2+2\delta|X_{\underline{t}}^{\delta,i,N}|(t-\underline{t})^{\frac{1}{2}}\\ &\quad+2|\sigma|\big(1+|X_{\underline{t}}^{\delta,i,N}|\big)\big|W_t^i-W_{\underline{t}}^i\big|\Big)^{p-1-k}.\end{aligned} \tag{4.29}$$

Notice that the degree of the polynomial (on the right hand side of (4.29)) with respect to $\big|X_{\underline{t}}^{\delta,i,N}\big|$ is $2(p-1)$. In addition, the conditional expectation of the polynomial with respect to $|W_t^i-W_{\underline{t}}^i|$ given the σ-algebra $\mathscr{F}_{\underline{t}}^N$ offers at least the order $(t-\underline{t})^{\frac{1}{2}}$ so that the term $|b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})|(t-\underline{t})^{\frac{1}{2}}$ can be uniformly bounded by taking advantage of (4.22). Once again, with the aid of Young's inequality, there exists a constant $C_p^{*,8}>0$ such that

$$\mathbb{E}\big(\overline{\Theta}^{i,p}(\mathbf{X}_{\underline{t}}^{\delta,N})\big|\mathscr{F}_0^N\big) \le \frac{1}{4}p(\lambda-2K)\mathbb{E}\big(|X_{\underline{t}}^{\delta,i,N}|^{2p}\big|\mathscr{F}_0^N\big)+C_p^{*,8}>0.$$

This, together with (4.27) and (4.28), enables us to achieve (4.24). □

Before the end of this work, we accomplish the

Proof of Theorem 1.12. Once again, an application of Theorem 1.3 with $r_0=0$, $\theta_t=\overline{\theta}_t=\underline{t}$, and $\widetilde{b}=b$, respectively, yields that for some $\lambda^*>0$ and any $t\ge0$,

$$\begin{aligned}\mathbb{W}_1\big(\mathscr{L}_{X_t^i},\mathscr{L}_{X_t^{\delta,i,N}}\big) &\lesssim \mathrm{e}^{-\lambda^* t}\mathbb{W}_1\big(\mathscr{L}_{X_0^i},\mathscr{L}_{X_0^{\delta,i,N}}\big)+N^{-\frac{1}{2}}\mathbb{1}_{\{K>0\}}\\ &\quad+\int_0^t \mathrm{e}^{-\lambda^*(t-s)}\mathbb{E}\big|b(X_s^{\delta,i,N},\widetilde{\mu}_s^{\delta,N})-b(X_{\underline{s}}^{\delta,i,N},\widetilde{\mu}_{\underline{s}}^{\delta,N})\big|\,\mathrm{d}s.\end{aligned} \tag{4.30}$$

Next, from (1.17) and ($\mathbf{A}_3$), it is easy to see that

$$\mathbb{E}\big|b(X_t^{\delta,i,N},\widetilde{\mu}_t^{\delta,N})-b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})\big|\lesssim\mathbb{E}\big((1+\big|X_t^{\delta,i,N}\big|^{l^*}+\big|X_{\underline{t}}^{\delta,i,N}\big|^{l^*})\big|X_t^{\delta,i,N}-X_{\underline{t}}^{\delta,i,N}\big|\big)$$
$$+\mathbb{E}\big|X_t^{\delta,i,N}-X_{\underline{t}}^{\delta,i,N}\big|+\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\big|X_t^{\delta,j,N}-X_{\underline{t}}^{\delta,j,N}\big|. \tag{4.31}$$

Moreover, by taking the definition of h^δ given in (1.33) into account, one has

$$|X_t^{\delta,i,N}-X_{\underline{t}}^{\delta,i,N}|\le|b(X_{\underline{t}}^{\delta,i,N},\widetilde{\mu}_{\underline{t}}^{\delta,N})|(t-\underline{t})+|\sigma(W_t^i-W_{\underline{t}}^i)|\le\delta+|\sigma(W_t^i-W_{\underline{t}}^i)|. \tag{4.32}$$

Note that from Hölder's inequality, we deduce that

$$\begin{aligned}&\mathbb{E}\big((1+|X_t^{\delta,i,N}|^{l^*})\big|X_t^{\delta,i,N}-X_{\underline{t}}^{\delta,i,N}\big|\big)\\ &\lesssim\mathbb{E}\Big(\big(1+\big(\mathbb{E}\big(|X_t^{\delta,i,N}|^{2l^*}\big|\mathscr{F}_0^N\big)\big)^{\frac{1}{2}}\big)\big(\mathbb{E}\big(\big|X_t^{\delta,i,N}-X_{\underline{t}}^{\delta,i,N}\big|^2\big|\mathscr{F}_0^N\big)\big)^{\frac{1}{2}}\Big).\end{aligned} \tag{4.33}$$

Whereafter, the assertion (1.34) can be attainable by plugging (4.31) back into (4.30) followed by combining (4.32) and (4.33) with Lemma 4.4. Consequently, the proof of Theorem 1.12 is complete. □

Acknowledgments

We would like to thank the referee for the constructive comments and suggestions, which improve greatly our manuscript. The research of Bao is supported by the National Key R&D Program of China (2022YFA1006004).

Data availability

No data was used for the research described in the article.

References

[1] J. Bao, X. Huang, Approximations of McKean-Vlasov stochastic differential equations with irregular coefficients, J. Theor. Probab. 35 (2022) 1187–1215.
[2] J. Bao, F.-Y. Wang, C. Yuan, Hypercontractivity for functional stochastic differential equations, Stoch. Process. Appl. 125 (2015) 3636–3656.
[3] U. Botija-Munoz, C. Yuan, Explicit numerical approximations for SDDEs in finite and infinite horizons using the adaptive EM method: strong convergence and almost sure exponential stability, Appl. Math. Comput. 478 (2024) 128853, 26 pp.
[4] N. Brosse, A. Durmus, É. Moulines, S. Sabanis, The tamed unadjusted Langevin algorithm, Stoch. Process. Appl. 129 (2019) 3638–3663.
[5] R. Carmona, F. Delarue, Probabilistic Theory of Mean Field Games with Applications I: Mean Field FBSDEs, Control, and Games, Springer International Publishing, Switzerland, 2018.
[6] G. dos Reis, S. Engelhardt, G. Smith, Simulation of McKean-Vlasov SDEs with super-linear growth, IMA J. Numer. Anal. 42 (2022) 874–922.
[7] A. Durmus, A. Eberle, A. Guillin, R. Zimmer, An elementary approach to uniform in time propagation of chaos, Proc. Am. Math. Soc. 148 (2020) 5387–5398.
[8] A. Eberle, Reflection couplings and contraction rates for diffusions, Probab. Theory Relat. Fields 166 (2016) 851–886.

[9] W. Fang, Michael B. Giles, Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift, Ann. Appl. Probab. 30 (2020) 526–560.
[10] N. Fournier, A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, Probab. Theory Relat. Fields 162 (2015) 707–738.
[11] A. Guillin, P. Le Bris, P. Monmarché, On systems of particles in singular repulsive interaction in dimension one: log and Riesz gas, J. Éc. Polytech. Math. 10 (2023) 867–916.
[12] A. Guillin, P. Le Bris, P. Monmarché, Uniform in time propagation of chaos for the 2D vortex model and other singular stochastic systems, J. Eur. Math. Soc. 27 (2025) 2359–2386.
[13] Desmond J. Higham, X. Mao, Andrew M. Stuart, Strong convergence of Euler-type methods for nonlinear stochastic differential equations, SIAM J. Numer. Anal. 40 (2002) 1041–1063.
[14] X. Huang, Propagation of chaos for mean field interacting particle system with multiplicative noise, arXiv:2308.15181v5.
[15] M. Hutzenthaler, A. Jentzen, Peter E. Kloeden, Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients, Ann. Appl. Probab. 22 (2012) 1611–1641.
[16] M. Hutzenthaler, A. Jentzen, Peter E. Kloeden, Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients, Proc. R. Soc. Lond. Ser. A, Math. Phys. Eng. Sci. 467 (2011) 1563–1576.
[17] C. Kumar, C. Reisinger, Neelima, W. Stockinger, Well-posedness and tamed schemes for McKean-Vlasov equations with common noise, Ann. Appl. Probab. 32 (2022) 3283–3330.
[18] D.-Y. Lim, A. Neufeld, S. Sabanis, Y. Zhang, Langevin dynamics based algorithm e-THεO POULA for stochastic optimization problems with discontinuous stochastic gradient, Math. Oper. Res. (2025), in press, https://doi.org/10.1287/moor.2022.0307.
[19] D.-Y. Lim, A. Neufeld, S. Sabanis, Y. Zhang, Non-asymptotic estimates for TUSLA algorithm for non-convex learning with applications to neural networks with ReLU activation function, IMA J. Numer. Anal. 44 (2024) 1464–1559.
[20] J.C. Mattingly, A.M. Stuart, D.J. Higham, Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise, Stoch. Process. Appl. 101 (2002) 185–232.
[21] N.-k. Tran, T.-T. Kieu, D.-T. Luong, H.-L. Ngo, On the infinite time horizon approximation for Lévy-driven McKean-Vlasov SDEs with non-globally Lipschitz continuous and super-linearly growth drift and diffusion coefficients, J. Math. Anal. Appl. 543 (2, Part 2) (2025) 128982, 38 pp.
[22] E. Priola, F.-Y. Wang, Gradient estimates for diffusion semigroups with singular coefficients, J. Funct. Anal. 236 (2006) 244–264.
[23] M. Reiß, M. Riedle, O. van Gaans, Delay differential equations driven by Lévy processes: stationarity and Feller properties, Stoch. Process. Appl. 116 (2006) 1409–1432.
[24] C. Reisinger, W. Stockinger, An adaptive Euler-Maruyama scheme for McKean-Vlasov SDEs with super-linear growth and application to the mean-field FitzHugh-Nagumo model, J. Comput. Appl. Math. 400 (2022) 113725, 23 pp.
[25] P. Ren, F.-Y. Wang, Entropy estimate between diffusion processes with application to nonlinear Fokker-Planck equations, arXiv:2302.13500.
[26] S. Sabanis, Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients, Ann. Appl. Probab. 26 (2016) 2083–2105.
[27] K. Schuh, Global contractivity for Langevin dynamics with distribution-dependent forces and uniform in time propagation of chaos, Ann. Inst. Henri Poincaré Probab. Stat. 60 (2024) 753–789.
[28] K. Suzuki, Weak convergence of approximate reflection coupling and its application to non-convex optimization, arXiv:2205.11970.
[29] A.S. Sznitman, Topics in Propagation of Chaos, Springer, Berlin, 1991.
[30] F.-Y. Wang, Distribution dependent SDEs for Landau type equations, Stoch. Process. Appl. 128 (2018) 595–621.
[31] F.-Y. Wang, Asymptotic couplings by reflection and applications for nonlinear monotone SPDEs, Nonlinear Anal. 117 (2015) 169–188.
[32] H. Yang, X. Li, Explicit approximations for nonlinear switching diffusion systems in finite and infinite horizons, J. Differ. Equ. 265 (2018) 2921–2967.
[33] C. Yuan, X. Mao, Stability in distribution of numerical solutions for stochastic differential equations, Stoch. Anal. Appl. 22 (2004) 1133–1150.