OPEN ACCESS
Mathematical Models and Methods in Applied Sciences
Vol. 33, No. 2 (2023) 289–339

DOI: 10.1142/S0218202523500082

Consensus-based optimization via jump-diffusion stochastic differential equations

Dante Kalise

Department of Mathematics, Imperial College London,
South Kensington Campus, SW7 2AZ London, UK
dkaliseb@imperial.ac.uk

Akash Sharma* and Michael V. Tretyakov†

School of Mathematical Sciences,
University of Nottingham, UK
*akash.sharma1@nottingham.ac.uk
†*Michael.Tretyakov@nottingham.ac.uk*

Received 23 May 2022
Revised 15 November 2022
Accepted 14 December 2022
Published 17 February 2023
Communicated by N. Bellomo

We introduce a new consensus-based optimization (CBO) method where an interacting particle system is driven by jump-diffusion stochastic differential equations (SDEs). We study well-posedness of the particle system as well as of its mean-field limit. The major contributions of this paper are proofs of convergence of the interacting particle system towards the mean-field limit and convergence of a discretized particle system towards the continuous-time dynamics in the mean-square sense. We also prove convergence of the mean-field jump-diffusion SDEs towards global minimizer for a large class of objective functions. We demonstrate improved performance of the proposed CBO method over earlier CBO methods in numerical simulations on benchmark objective functions.

Keywords: Global non-convex optimization; interacting particle systems; mean-field jump-diffusion SDEs; McKean–Vlasov SDEs with jumps.

AMS Subject Classification: 60H10, 90C26, 65C30, 65C35, 60J76

1. Introduction

Large-scale individual-based models have become a well-established modeling tool in modern social science, natural science and engineering, with applications including social networks, crowd dynamics, epidemics, pedestrian motion, collective

†Corresponding author

animal behavior, swarm robotics and molecular dynamics, among many others (see e.g. Refs. 2, 4, 7, 9 and 26). Through the iteration of basic interactions forces found in nature and society such as attraction, repulsion, alignment, flocking, swarming, synchronization, polarization, fragmentation, competition and cooperation, these complex systems exhibit a rich self-organization behavior (see e.g. Refs. 6, 10, 12, 17, 36, 44 and 50).

Over the last decades, individual-based models have also entered the field of global optimization and its many applications in operations research, control, engineering, economics, finance and machine learning. In many applied problems arising in the aforementioned fields, the objective function to be optimized can be non-convex and/or non-smooth, disabling the use of traditional continuous/convex optimization technique. In such scenarios, individual-based metaheuristic models have been proven surprisingly effective. Examples include genetic algorithms, ant colony optimization, particle swarm optimization, simulated annealing, etc. (see Refs. 15, 20, 37 and 40 and the references therein). These methods are probabilistic in nature which set them apart from other derivative-free algorithms.[16] Unlike many convex optimization methods, metaheuristic algorithms are relatively simple to implement and easily parallelizable. This combination of simplicity and effectiveness has fueled the application of metaheuristic in complex engineering problems such as shape optimization, scheduling problems and hyper-parameter tuning in machine learning models. However, it is often the case that metaheuristics lack rigorous convergence results, a question which has become an active area of research.[30, 31]

In Ref. 46, the authors introduced an optimization algorithm which employs an individual-based model to frame a global minimization

$$\min_{x\in\mathbb{R}^d} f(x),$$

where $f(x)$ is a positive function from $\mathbb{R}^d$ to $\mathbb{R}$, as a consensus problem. In this model, each individual particle explores the energy landscape given by $f(x)$, broadcasting its current value to the rest of the ensemble through a weighted average. This iterated interaction generates trajectories which flock towards a consensus point corresponding to a global minimizer of $f(x)$, hence the name *Consensus-Based Optimization* (CBO). We refer to Refs. 30 and 51 for the two recent surveys on the topic. The dynamics of existing CBO models are governed by stochastic differential equations (SDEs) with Wiener noise.[11, 13, 46] Hence, we can resort to a toolbox from stochastic calculus and stochastic numerics to perform analysis of these models. This amenability of CBO models to theoretical as well as numerical analysis differentiates them from other agent-based optimization algorithms.

In this paper, we propose a new CBO model which is governed by jump-diffusion stochastic differential equations. This means randomness in the dynamics of the proposed CBO model comes from Wiener process as well as compound Poisson process. The following are the main contributions of this paper:

(i) We prove well-posedness of the interacting-particle system and of its mean-field limit driven by jump-diffusion SDEs and convergence of the mean-field SDEs to the global minimum. The approach to study well-posedness and convergence to the global minimum is similar to Ref. 11 but adapted to the jump-diffusion case with time-dependent coefficients.

(ii) The major contribution of the paper is that we prove mean-square convergence of the interacting particle system to the mean-field limit when the number of particles, N, tends to ∞, i.e. for all $t \in [0, T]$,

$$\lim_{N\to\infty} \sup_{i=1,\dots,N} \mathbb{E}|X_N^i(t) - X^i(t)|^2 = 0, \tag{1.1}$$

where X_N^i represent interacting particles (see (2.11)) and X^i denote their mean-field limit (see (2.12)). This also implies convergence of the particle system towards the mean-field limit in 2-Wasserstein metric. Let us emphasize that we prove this result for quadratically growing objective function. We also study uniform in N convergence of the implementable discretized particle system towards the jump-diffusion SDEs as the discretization step, h, goes to 0.

(iii) It is illustrated in the presented numerical experiments that the addition of jumps to the particle system leads to more effective exploration of the energy landscape. This is particularly relevant when a good prior knowledge of the optimal solution for initialization of the CBO is not available.

As was highlighted in Remark 3.2 of Ref. 11 and Remark 2 of Ref. 24, it is not straightforward to prove mean-square convergence of the CBO particle system towards its mean-field limit, even after proving uniform in N moment bounds of the solutions of the SDEs driving particles system. Convergence results of this type have been proved for special cases of compact manifolds (see Ref. 22 for compact hypersurfaces and Ref. 35 for Stiefel manifolds) and globally Lipschitz continuous objective functions. In this case, not only the objective function is bounded but also particles are evolving on a compact set. Under the assumptions on the objective function as in our paper, in the diffusion case weak convergence of the empirical measure of CBO particle system to the law of the corresponding mean-field SDEs is proved in Refs. 30 and 38 exploiting Prokhorov's theorem. In Ref. 24, the authors proved convergence in probability of the CBO particle system with diffusion to the mean-field limit.

A propagation of chaos result in the weak sense for non-linear jump-diffusions, with globally Lipschitz coefficients, has been studied in Refs. 28 and 29. The authors of Ref. 45 proved convergence of the particle system driven by Lévy noise in 2-Wasserstein distance for one-sided Lipschitz drift coefficient in spatial variable but uniformly Lipschitz in measure. The mean-square convergence has also been established in Ref. 21 for locally Lipschitz drift coefficient in measure, bounded diffusion coefficient and particular choice of the jump coefficient. Here we prove L^2 convergence of the CBO particle system to the mean-field SDEs where drift, diffusion and jump coefficients are locally Lipschitz in measure. We consider quadratically growing locally Lipschitz objective function defined on $\mathbb{R}^d$. We note that our convergence results hold for the earlier CBO models[11, 13, 46] as well. Although our main focus is on CBO models and the global optimization problem, our work also contributes to the currently very active research in mean-filed SDEs and their particles approximations.

Furthermore, practical implementation of the particle system corresponding to a CBO model needs a numerical approximation in the mean-square sense. We utilize an explicit Euler scheme to implement the proposed jump-diffusion CBO model. This leads to the question whether the Euler scheme converges to the CBO model taking into account that the coefficients of the particle system are not globally Lipschitz and the Lipschitz constants grow exponentially when the objective function is not bounded. At the same time, the coefficients of the particle system have linear growth at infinity. In the case of jump-diffusion SDEs, earlier works either showed convergence of the Euler scheme in the case of globally Lipschitz coefficients[47] or proposed special schemes in the case of non-globally Lipschitz coefficients with super-linear growth, e.g. a tamed Euler scheme.[18] Here we prove mean-square convergence of the Euler scheme and we show that this convergence is uniform in the number of particles N, i.e. the choice of a discretization time-step h is independent of N. Our results can be utilized for the earlier CBO models.[11, 13, 46]

In Sec. 2, we first present a review of existing CBO models and then describe the proposed jump-diffusion CBO model. We also formally introduce mean-field limit of the new CBO model. In Sec. 3, we focus on well-posedness of the interacting particle system behind the new CBO model and its mean-field limit. In Sec. 4, we discuss convergence of the mean-field limit towards a point in $\mathbb{R}^d$, which approximates the global minimum, convergence of the interacting particle system towards the mean-field limit, and convergence of the implementable discretized particle system towards the continuous-time particle system. We present results of numerical experiments in Sec. 5 to compare performance of our model and the existing CBO models.

Throughout the paper, C is a floating constant which may vary at different places. We denote $(a \cdot b)$ as dot product between two vectors, $a, b \in \mathbb{R}^d$. We will omit brackets () wherever it does not lead to any confusion.

2. CBO Models: Existing and New

In Sec. 2.1, we review the existing CBO models. In Sec. 2.2, we introduce a new CBO model driven by jump-diffusion and discuss potential advantages of adding jumps to CBO models which are confirmed by numerical experiments in Sec. 5. The numerical experiments of Sec. 5 are conducted using the Euler scheme presented in Sec. 2.2.

2.1. *Review of the existing CBO models*

Let $N \in \mathbb{N}$ denote the number of agents with position vector, $X_N^i(t) \in \mathbb{R}^d$, $i = 1, \dots, N$. The following model was proposed in Ref. 46:

$$dX_N^i(t) = -\beta(X_N^i(t) - \bar{X}_N^{\alpha,f}(t))H^\epsilon(f(X_N^i(t)) - f(\bar{X}_N^{\alpha,f}(t)))dt + \sqrt{2}\sigma|X_N^i(t) - \bar{X}_N^{\alpha,f}(t)|dW^i(t), \quad i = 1, \dots, N, \tag{2.1}$$

where $H^\epsilon : \mathbb{R} \to \mathbb{R}$ is a smooth regularization of the Heaviside function, $W^i(t)$, $i = 1, \dots, N$, represent N-independent d-dimensional standard Wiener processes, $\beta > 0$, $\sigma > 0$ and $\bar{X}_N^{\alpha,f}(t)$ is given by

$$\bar{X}_N^{\alpha,f}(t) = \frac{\sum_{i=1}^N X_N^i(t) w_f^\alpha(X_N^i(t))}{\sum_{i=1}^N w_f^\alpha(X_N^i(t))} \tag{2.2}$$

with $w_f^\alpha(x) = \exp(-\alpha f(x))$, $\alpha > 0$.

Each particle X_N^i at time t is assigned an opinion $f(X_N^i(t))$. The lesser the value of f for a particle, the more is the influence of that particle, i.e. the more weight is assigned to that particle at that time as can be seen in (2.2) of the instantaneous weighted average. If the value $f(X_N^i(t))$ of a particle X_N^i at time t is greater than the value $f(\bar{X}_N^{\alpha,f}(t))$ at the instantaneous weighted average $\bar{X}_N^{\alpha,f}(t)$ then the regularized Heaviside function forces the particle X_N^i to drift towards $\bar{X}_N^{\alpha,f}$. If the opinion of ith particle matters more among the interacting particles, i.e. the value $f(X_N^i(t))$ is less than $f(\bar{X}_N^i(t))$, then it is not beneficial for it to move towards $\bar{X}_N^{\alpha,f}$. The noise term is added to explore the space $\mathbb{R}^d$ and to avoid non-uniform consensus. The noise intensity induced in the dynamics of the ith particle at time t takes into account the distance of the particle from the instantaneous weighted average, $\bar{X}_N^{\alpha,f}(t)$. Over a period of time as the particles start moving towards a consensus opinion, the coefficients in (2.1) go to zero.

One can observe that the more influential opinion a particular particle has, the higher is the weight assigned to that particle in the instantaneous weighted average (2.2). Based on this logic, in Ref. 11, the authors dropped the regularized Heaviside function in the drift coefficient and the model (2.1) was simplified as follows:

$$dX_N^i(t) = -\beta(X_N^i(t) - \bar{X}_N^{\alpha,f}(t))dt + \sqrt{2}\sigma|X_N^i(t) - \bar{X}_N^{\alpha,f}(t)|dW^i(t), \quad i = 1, \dots, N \tag{2.3}$$

with β, σ, $\bar{X}_N^{\alpha,f}$ as in (2.1) and (2.2).

The major drawback of the consensus-based models (2.1) and (2.3) is that the parameters β and σ are dependent on the dimension d. To illustrate this fact, we replace $\bar{X}_N^{\alpha,f}$ in (2.3) by a fixed vector $V \in \mathbb{R}^d$. Then, using Ito's formula, we have

$$\frac{d}{dt}\mathbb{E}|X_N^i(t) - V|^2 = (-2\beta + \sigma^2 d)\mathbb{E}|X_N^i(t) - V|^2, \quad i = 1, \dots, N. \tag{2.4}$$

As one can notice, for particles to reach the consensus point whose position vector is V, one needs $2\beta > d\sigma^2$. To overcome this deficiency, the authors of Ref. 13 proposed the following model which is based on component-wise noise intensity instead of isotropic noise used in (2.1) and (2.3):

$$\begin{aligned} dX_N^i(t) &= -\beta(X_N^i(t) - \bar{X}_N^{\alpha,f}(t))dt \\ &\quad + \sqrt{2}\sigma\,\mathrm{Diag}(X_N^i(t) - \bar{X}_N^{\alpha,f}(t))dW^i(t), \quad i = 1,\ldots,N, \end{aligned} \tag{2.5}$$

where β, σ and $\bar{X}_N^{\alpha,f}$ are as in (2.1) and (2.2), and $\mathrm{Diag}(U)$ is a diagonal matrix whose diagonal is a vector $U \in \mathbb{R}^d$. Now, if we replace $\bar{X}_N^{\alpha,f}$ by a fixed vector V and then use Ito's formula for (2.5), we get

$$\begin{aligned} \frac{d}{dt}\mathbb{E}|X_N^i(t) - V|^2 &= -2\beta\mathbb{E}|X_N^i(t) - V|^2 + \sigma^2\mathbb{E}\sum_{j=1}^{d}(X_N^i(t) - V)_j^2 \\ &= (-2\beta + \sigma^2)\mathbb{E}|X_N^i(t) - V|^2, \quad i = 1,\ldots,N, \end{aligned} \tag{2.6}$$

where $(X_N^i(t) - V)_j$ denotes the jth component of $(X_N^i(t) - V)$. It is clear that in this model, there is no dimensional restriction on β and σ.

Other CBO models[33, 34] are based on interacting particles driven by common noise. Since the same noise drives all the particles, the exploration is usually not effective. They are not scalable with respect to dimension and typically less effective than the CBO models (2.1), (2.3), (2.5) and the model introduced in Sec. 2.2. This fact is demonstrated in experiments in Sec. 5.

2.2. *Jump-diffusion CBO models*

Let us consider the following jump-diffusion model:

$$\begin{aligned} dX_N^i(t) &= -\beta(t)(X_N^i(t) - \bar{X}_N(t))dt + \sqrt{2}\sigma(t)\,\mathrm{Diag}(X_N^i(t) - \bar{X}_N(t))dW^i(t) \\ &\quad + \gamma(t)\,\mathrm{Diag}(X_N^i(t^-) - \bar{X}_N(t^-))dJ^i(t), \quad i = 1,\ldots,N \end{aligned} \tag{2.7}$$

with

$$J^i(t) = \sum_{j=1}^{N^i(t)} Z_j^i, \tag{2.8}$$

where $N^i(t)$, $i = 1\ldots,N$ are N-independent Poisson processes with jump intensity λ and $Z_j^i = (Z_{j,1}^i,\ldots,Z_{j,d}^i)^\top$ are i.i.d. d-dimensional random variables denoting jth jump by ith particle and $Z_j^i \sim Z$. We denote the lth component of the vector Z by Z_l. We assume that the components Z_l of Z are also i.i.d. random variables distributed as

$$Z_l \sim \mathcal{Z}, \tag{2.9}$$

where $\mathcal{Z}$ is an $\mathbb{R}$-valued random variable whose probability density is given by $\rho_{\mathfrak{z}}(\mathfrak{z})$ such that $\mathbb{E}(\mathcal{Z}) = \int_{\mathbb{R}} \mathfrak{z}\rho_{\mathfrak{z}}(\mathfrak{z})d\mathfrak{z} = 0$. We also denote the probability density of Z as $\rho_z(z) = \prod_{l=1}^d \rho_{\mathcal{Z}}(z_l)$. Note that $\mathbb{E}(Z)$ is a d-dimensional zero vector, since each Z_l is distributed as $\mathcal{Z}$. The Wiener processes $W^i(t)$, the Poisson processes $N^i(t)$, $i = 1, \ldots, N$, and the jump sizes Z are assumed to be mutually independent (see further theoretical details concerning Lévy-driven SDEs in Ref. 3). Also, $\beta(t)$, $\sigma(t), \gamma(t)$ are continuous functions and

$$\bar{X}_N(t) = (\bar{X}_N^1(t), \ldots, \bar{X}_N^d(t)) := \frac{\sum_{i=1}^N X_N^i(t) e^{-\alpha f(X_N^i(t))}}{\sum_{i=1}^N e^{-\alpha f(X_N^i(t))}} \tag{2.10}$$

with $\alpha > 0$. Note that we have omitted α and f of $\bar{X}_N^{\alpha,f}$ in the notation used in (2.7) for the simplicity of writing.

We recall the meaning of the jump term

$$\int_0^t \gamma(s)\,\mathrm{Diag}(X^i(s^-) - \bar{X}_N(s^-))dJ^i(s)$$
$$= \sum_{j=1}^{N^i(t)} \gamma(\tau_j^i)\,\mathrm{Diag}(X^i(\tau_j^{i-}) - \bar{X}_N(\tau_j^{i-}))Z_j^i,$$

where τ_j^i denotes the time of jth jump of the Poisson process $N^i(t)$. Thanks to the assumption that $\mathbb{E}(\mathcal{Z}) = 0$ (which in turn implies $\mathbb{E}(Z_{j,l}^i) = 0$, $j = 1, \ldots, N^i(t)$, $i = 1, \ldots, N$, $l = 1, \ldots, d$), the above integral is a martingale, and hence (similarly to Ito's integral term in (2.7)) it does not bias trajectories of $X_N^i(t)$, $i = 1, \ldots, N$.

The jump diffusion SDEs (2.7) are different from (2.5) in the two ways:

- The SDEs (2.7) are a consequence of interlacing of Ito's diffusion by jumps arriving according to the Poisson processes whose jump intensity is given by λ.
- We take $\beta(t)$ as a continuous positive non-decreasing function of t such that $\beta(t) \to \beta > 0$ as $t \to \infty$, $\sigma(t)$ as a continuous positive non-increasing function of t such that $\sigma(t) \to \sigma > 0$ as $t \to \infty$ and $\gamma(t)$ as a continuous non-negative non-increasing function of t such that $\gamma(t) \to \gamma \geq 0$ as $t \to \infty$.

Although we analyze the CBO model (2.7) with time-dependent parameters, a decision to take parameters time-dependent or not is problem specific. Note that the particles driven by SDEs (2.7) jump at different times with different jump sizes and jumps arrive according to the Poisson processes with the same intensity λ.

We can also write the jump-diffusion SDEs (2.7) in terms of the Poisson random measure[3] as

$$dX_N^i(t) = -\beta(t)(X_N^i(t) - \bar{X}_N(t))dt + \sqrt{2}\sigma(t)\,\mathrm{Diag}(X_N^i(t) - \bar{X}_N(t))dW^i(t)$$
$$+ \int_{\mathbb{R}^d} \gamma(t)\,\mathrm{Diag}(X_N^i(t^-) - \bar{X}_N(t^-))z\mathcal{N}^i(dt, dz), \tag{2.11}$$

where $\mathcal{N}^i(dt, dz)$, $i = 1, \ldots, N$, represent the independent Poisson random measures with intensity measure $\nu(dz)dt$. Here $\nu(dz) = \lambda\rho_z(z)dz$ is a finite Lévy measure.

Although for simplicity we introduced our model as (2.7), in proving well-posedness and convergence results we will make use of (2.11).

We can formally write the mean-field limit of the model (2.7) as the following McKean–Vlasov SDEs:

$$\begin{aligned} dX(t) &= -\beta(t)(X(t) - \bar{X}(t))dt + \sqrt{2}\sigma(t)\,\mathrm{Diag}(X(t) - \bar{X}(t))dW(t) \\ &\quad + \gamma(t)\,\mathrm{Diag}(X(t^-) - \bar{X}(t^-))dJ(t), \end{aligned} \tag{2.12}$$

where $J(t) = \sum_{j=1}^{N(t)} Z_j$, $N(t)$ is a Poisson process with intensity λ and

$$\bar{X}(t) := \bar{X}^{\mathcal{L}_{X(t)}} = \frac{\int_{\mathbb{R}^d} x e^{-\alpha f(x)} \mathcal{L}_{X(t)}(dx)}{\int_{\mathbb{R}^d} e^{-\alpha f(x)} \mathcal{L}_{X(t)}(dx)} = \frac{\mathbb{E}\big(X(t) e^{-\alpha f(X(t))}\big)}{\mathbb{E}\big(e^{-\alpha f(X(t))}\big)} \tag{2.13}$$

with $\mathcal{L}_{X(t)} := \mathrm{Law}(X(t))$. We can rewrite the mean-field jump-diffusion SDEs (2.12) in terms of the Poisson random measure as

$$\begin{aligned} dX(t) &= -\beta(t)(X(t) - \bar{X}(t))dt + \sqrt{2}\sigma(t)\,\mathrm{Diag}(X(t) - \bar{X}(t))dW(t) \\ &\quad + \gamma(t) \int_{\mathbb{R}^d} \mathrm{Diag}(X(t^-) - \bar{X}(t^-)) z \mathcal{N}(dt, dz). \end{aligned} \tag{2.14}$$

2.2.1. *Other jump-diffusion CBO models*

Although the aim of the paper is it to analyze the CBO model (2.11), we discuss three other jump-diffusion CBO models of interest in this section.

Additional Model 1: Writing (2.7) in terms of Poisson random measure suggests that we can also consider a CBO model with an infinite activity Lévy process, e.g. an α-stable process, to introduce jumps in dynamics of particles:

$$\begin{aligned} dX_N^i(t) &= -\beta(t)(X_N^i(t) - \bar{X}_N(t))dt + \sqrt{2}\sigma(t)\,\mathrm{Diag}(X_N^i(t) - \bar{X}_N(t))dW^i(t) \\ &\quad + \int_{\mathbb{R}^d} \gamma(t)\,\mathrm{Diag}(X_N^i(t^-) - \bar{X}_N(t^-)) z \mathcal{N}^i(dt, dz), \end{aligned} \tag{2.15}$$

where the Lévy measure corresponding to $\mathcal{N}^i(dt, dz)$ can be infinite. However, numerical approximation of SDEs driven by infinite activity Lévy processes is computationally more expensive (see e.g. Refs. 19 and 47), hence it can be detrimental for the overall CBO performance.

Additional Model 2: In the SDEs (2.7), the intensity of Poisson process λ is constant. If we take jump intensity as $\lambda(t)$, i.e. a function of t, then the corresponding SDEs will be as follows:

$$\begin{aligned} dX_N^i(t) &= -\beta(t)(X_N^i(t) - \bar{X}_N(t))dt + \sqrt{2}\sigma(t)\,\mathrm{Diag}(X_N^i(t) - \bar{X}_N(t))dW^i(t) \\ &\quad + \mathrm{Diag}(X_N^i(t^-) - \bar{X}_N(t^-))dJ^i(t), \quad i = 1, \ldots, N, \end{aligned} \tag{2.16}$$

where all the notation are as in (2.7) and (2.10) except here the intensity of the Poisson processes $N^i(t)$ is a time-dependent function $\lambda(t)$. It is assumed that $\lambda(t)$ is a decreasing function such that $\lambda(t) \to 0$ as $t \to \infty$. Also, in comparison with (2.7), there is no $\gamma(t)$ in the jump component of (2.16). Note that the compound Poisson process with constant jump intensity λ is a Lévy process but with time-dependent jump intensity $\lambda(t)$, it is not a Lévy process, rather it is an additive process. Additive process is a generalization of Lévy process which satisfies all conditions of Lévy process except stationarity of increments.[39] The SDEs (2.16) present a jump-diffusion CBO model driven by additive process. The analysis of model (2.16) follows similar arguments as used in the paper for the model (2.11), since the jump-diffusion SDEs (2.16) can also be written in terms of the Poisson random measure with intensity measure $\nu_t(dz)dt$, where $(\nu_t)_{t\geq 0}$ is a family of Lévy measures.

Additional Model 3: In the model (2.11), the particles have idiosyncratic noise, which means they are driven by different Wiener processes and different compound Poisson processes. Instead, we can have another jump-diffusion model in which the same Poisson noise drives the particle system but Wiener processes stay different and jumps sizes still independently vary for all particles. This means jumps arrive at the same times for all particles, but particles jump with different random jump-sizes. We can write this CBO model as

$$dX_N^i(t) = -\beta(t)(X_N^i(t) - \bar{X}_N(t))dt + \sqrt{2}\sigma(t)\,\mathrm{Diag}(X_N^i(t) - \bar{X}_N(t))dW^i(t)$$
$$+ \int_{\mathbb{R}^d} \gamma(t)\,\mathrm{Diag}(X_N^i(t^-) - \bar{X}_N(t^-))z\mathcal{N}(dt, dz). \tag{2.17}$$

We compare performance of the jump-diffusion CBO models (2.11) and (2.17) in Sec. 5.

2.2.2. *Discussion*

First, we will discuss dependence of the parameters $\beta(t)$, $\sigma(t)$, $\gamma(t)$ and λ on dimension d. The independence and identical distribution of Z_l, which denotes the lth component of Z, result in the non-dependency of parameters on dimension in the similar manner as for the model (2.5). We illustrate this fact by fixing a vector $V \in \mathbb{R}^d$ and replacing $\bar{X}_N$ in (2.11) by V, then using Ito's formula and the assumption made on $\rho_{z}(z)$, we have

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}|X_N^i(t) - V|^2 &= -2\beta(t)\mathbb{E}|X_N^i(t) - V|^2 + \sigma^2(t)\sum_{j=1}^{d}\mathbb{E}(X_N^i(t) - V)_j^2 \\
&\quad + \lambda\int_{\mathbb{R}^d}\mathbb{E}\big(|X_N^i(t) - V + \gamma(t)\,\mathrm{Diag}(X_N^i(t) - V)z|^2 \\
&\quad - |X_N^i(t) - V|^2\big)\rho_z(z)dz \\
&= (-2\beta(t) + \sigma^2(t))\mathbb{E}|X_N^i(t) - V|^2 \\
&\quad + \lambda\int_{\mathbb{R}^d}\gamma^2(t)\mathbb{E}|\mathrm{Diag}(X_N^i(t) - V)z|^2\rho_z(z)dz
\end{aligned}$$

$$\begin{aligned}
&= (-2\beta(t) + \sigma^2(t))\mathbb{E}|X_N^i(t) - V|^2 \\
&\quad + \lambda\gamma^2(t)\sum_{j=1}^{d}\int_{\mathbb{R}^d}\mathbb{E}(X_N^i(t) - V)_j^2 z_j^2 \prod_{l=1}^{d}\rho_{\mathcal{Z}}(z_l)dz \\
&= (-2\beta(t) + \sigma^2(t) + \lambda\gamma^2(t)\mathbb{E}(\mathcal{Z}^2))\mathbb{E}|X_N^i(t) - V|^2, \quad i = 1\ldots, N.
\end{aligned} \tag{2.18}$$

We can choose $\beta(t)$, $\sigma(t)$, $\gamma(t)$, λ and the distribution of $\mathcal{Z}$ guaranteeing that there is a $t_* \geq 0$ such that $-2\beta(t)+\sigma^2(t)+\lambda\gamma^2(t)\mathbb{E}(\mathcal{Z}^2) < 0$ for all $t \geq t_*$ and such a choice is independent of d. It is clear from (2.18) that with this choice, $\mathbb{E}|X_N^i(t) - V|^2$, $i = 1, \ldots, N$, decay in time as $t \to \infty$.

In the previous CBO models, there were only two terms namely, the drift term and the diffusion term. The drift tries to take the particles towards their instantaneous weighted average. The diffusion term helps in exploration of the state space with the aim to find a state with better weighted average than the current one. The model (2.7) contains one extra term, which we call the jump term. Jumps help in intensifying the search in a search space and aid in avoiding premature convergence or trapping in local minima. This results in more effective use of the interaction of particles.

Moreover, the effect of jumps decays with time in (2.7) by virtue of decreasing $\gamma(t)$. The reason for considering the model (2.7) where jumps affect only the initial period of time is that we want particles to explore more space faster at the beginning of simulation and as soon as the weighted average of particles is in a vicinity of the global minimum, we do not want jumps to affect convergence of particles towards that consensus point lying in a close neighborhood of the global minimum. Therefore, time dependence of the parameters and degeneracy of the coefficients in front of the noises help in exploiting the searched space.

As a consequence, the jump-diffusion noise and time-dependent coefficients in the model (2.7) may help in keeping the balance of ***exploration*** and ***exploitation*** by interacting particles over a period of time. We will continue this discussion on exploration and exploitation in Sec. 5, where the proposed CBO model is tested.

2.2.3. *Implementation*

Let $0 = t_0 < \cdots < t_n = T$ be a uniform partition of the time interval $[0, T]$ into n sub-intervals such that $h := t_{k+1} - t_k$, $k = 0, \ldots, n-1$ and $T = nh$. To approximate (2.7), we construct a Markov chain $(Y_N^i(t_k))$, $k = 1, \ldots, n$, using the following Euler scheme:

$$\begin{aligned}
Y_N^i(t_{k+1}) &= Y_N^i(t_k) - \beta(t_k)(Y_N^i(t_k) - \bar{Y}_N(t_k))h + \sigma(t_k)\operatorname{Diag}\big(Y_N^i(t_k) \\
&\quad - \bar{Y}_N(t_k)\big)\Delta W^i(t_k) + \gamma(t_k)\sum_{j=N^i(t_k)+1}^{N^i(t_{k+1})}\operatorname{Diag}(Y_N^i(t_k) - \bar{Y}_N(t_k))Z_j^i,
\end{aligned} \tag{2.19}$$

where $\Delta W_l^i(t_k) = W_l^i(t_{k+1}) - W_l^i(t_k)$, $i = 1, \ldots, N$, $l = 1, \ldots, d$ are independent random variables having Gaussian distribution with mean 0 and variance h, $W_l^i(t) \in \mathbb{R}$ denotes lth component of $W^i(t)$, $Z_j^i \in \mathbb{R}^d$ denotes jth jump size of the ith particle, $N^i(t)$ are independent Poisson processes with jump intensity λ and

$$\bar{Y}_N(t) = \sum_{i=1}^{N} Y_N^i(t) \frac{e^{-\alpha f(Y_N^i(t))}}{\sum_{j=1}^{N} e^{-\alpha f(Y_N^j(t))}}. \tag{2.20}$$

To implement the discretization scheme, we initialize the $N \times d$ matrix Y at time $t_0 = 0$ and update it according to (2.19) and (2.20) at each iteration. We will consider mean-square convergence of the scheme (2.19) in Sec. 4.3.

The Python code for the above numerical scheme is available on github. It uses a matrix formulation of the corresponding algorithm to save memory and time in computations.

3. Well-posedness Results

In Sec. 3.1, we discuss well-posedness of the interacting particle system (2.11) and prove a moment bound for this system. In Sec. 3.2, we prove well-posedness of and a moment bound for the mean-field limit (2.14) of the particle system (2.11).

3.1. *Well-posedness of the jump-diffusion particle system*

This section is focused on showing existence and uniqueness of the solution of (2.11). We first introduce the notation which are required in this section.

We denote $\mathbf{x}_N := (x_N^1, \ldots, x_N^N)^\top \in \mathbb{R}^{Nd}$, $\bar{\mathbf{x}}_N := \sum_{i=1}^N x_N^i e^{-\alpha f(x_N^i)} / \sum_{j=1}^N e^{-\alpha f(x_N^j)}$, $\mathbf{W}(t) := (W^1(t), \ldots, W_N(t))^\top$, $\mathbf{F}_N(\mathbf{x}_N) := \big(F_N^1(\mathbf{x}_N), \ldots, F_N^N(\mathbf{x}_N)\big)^\top \in \mathbb{R}^{Nd}$ with $F_N^i(\mathbf{x}_N) = (x_N^i - \bar{\mathbf{x}}_N) \in \mathbb{R}^d$ for all $i = 1, \ldots, N$, $\mathbf{G}_N(\mathbf{x}_N) := \mathrm{Diag}(\mathbf{F}_N(\mathbf{x}_N)) \in \mathbb{R}^{Nd \times Nd}$ and $\mathbf{J}(t) = (J^1(t), \ldots, J^N(t))$, where $J^i(t)$ is from (2.8) which implies $\int_0^t \gamma(s)\,\mathrm{Diag}(F_N^i(\mathbf{X}_N(s^-)))dJ^i(s) = \int_0^t \int_{\mathbb{R}^d} \gamma(s)\,\mathrm{Diag}(F_N^i(\mathbf{X}_N(s^-))) z \mathcal{N}^i(ds, dz)$. Let us represent $\ell(dz)$ as the Lebesgue measure of dz, and for the sake of convenience we will use dz in place of $\ell(dz)$ whenever there is no confusion. We can write the particle system (2.11) using the above notation as

$$\begin{aligned} d\mathbf{X}_N(t) &= \beta(t)\mathbf{F}_N(\mathbf{X}_N(t^-))dt + \sqrt{2}\sigma(t)\mathbf{G}_N(\mathbf{X}_N(t^-))d\mathbf{W}(t) \\ &\quad + \gamma(t)\mathbf{G}_N(\mathbf{X}_N(t^-))d\mathbf{J}(t). \end{aligned} \tag{3.1}$$

In order to show well-posedness of (3.1), we need the following natural assumptions on the objective function f. Let

$$f_m := \inf f. \tag{3.2}$$

Assumption 3.1. $f_m > 0$.

Assumption 3.2. $f : \mathbb{R}^d \to \mathbb{R}$ is locally Lipschtiz continuous, i.e. there exists a positive function $L(R)$ such that

$$|f(x) - f(y)| \leq L(R)|x - y|,$$

whenever $|x|,\ |y| \leq R,\ x,\ y \in \mathbb{R}^d,\ R > 0$.

Assumption 3.2 is used for proving local Lipschitz continuity and linear growth of F_N^i and G_N^i, $i = 1, \ldots, N$. Let $B(R) = \{x \in \mathbb{R}^d\ :\ |x| \leq R\}$.

Lemma 3.1. *Under Assumptions* 3.1 *and* 3.2, *the following inequalities hold for any* $\boldsymbol{x}_N,\ \boldsymbol{y}_N \in \mathbb{R}^{Nd}$ *satisfying* $\sup_{i=1,\ldots,N}|x_N^i|,\ \sup_{i=1,\ldots,N}|y_N^i| \leq R$ *and for all* $i = 1, \ldots, N$:

(1) $|F_N^i(\boldsymbol{x}_N) - F_N^i(\boldsymbol{y}_N)| \leq |x_N^i - y_N^i| + \frac{C(R)}{N^{1/2}}|\boldsymbol{x}_N - \boldsymbol{y}_N|$,
(2) $|F_N^i(\boldsymbol{x}_N)|^2 \leq 2(|x_N^i|^2 + |\boldsymbol{x}_N|^2)$,

where $C(R) = e^{\alpha(|f|_{L_\infty(B(R))} - f_m)}\big(1 + \alpha R L(R) + \alpha R L(R) e^{\alpha(|f|_{L_\infty(B(R))} - f_m)}\big)$.

Proof. We can write

$$\begin{aligned}
&|F_N^i(\mathbf{x}_N) - F_N^i(\mathbf{y}_N)| \\
&\quad \leq |x_N^i - y_N^i| + \frac{1}{\sum_{j=1}^N e^{-\alpha f(x_N^j)}}\left(\left|\sum_{i=1}^N (x_N^i - y_N^i)e^{-\alpha f(x_N^i)}\right|\right. \\
&\qquad \left. + \left|\sum_{i=1}^N y_N^i\big(e^{-\alpha f(x_N^i)} - e^{-\alpha f(y_N^i)}\big)\right|\right) \\
&\qquad + \sum_{i=1}^N |y_N^i| e^{-\alpha f(y_N^i)}\left|\frac{1}{\sum_{j=1}^N e^{-\alpha f(x_N^j)}} - \frac{1}{\sum_{j=1}^N e^{-\alpha f(y_N^j)}}\right|.
\end{aligned} \tag{3.3}$$

Using discrete Jensen's inequality, we have

$$\frac{1}{\frac{1}{N}\sum_{i=1}^N e^{-\alpha f(x_N^i)}} \leq e^{\alpha \frac{1}{N}\sum_{i=1}^N f(x_N^i)}.$$

Using the above estimate and then the Cauchy–Bunyakovsky–Shwartz inequality in (3.3), we get the desired result. □

Theorem 3.1. *Let the initial condition* $\boldsymbol{X}_N(0)$ *of the jump-diffusion SDEs* (2.7) *satisfy* $\mathbb{E}|\boldsymbol{X}_N(0)|^2 < \infty$ *and* $\mathbb{E}|Z|^2 < \infty$, *then the* Nd-*dimensional system* (2.7) *has a unique strong solution* $\boldsymbol{X}_N(t)$ *under Assumptions* 3.1 *and* 3.2 *for each* $N \in \mathbb{N}$.

Proof. Note that $|G_N^i(\mathbf{x}_N) - G_N^i(\mathbf{y}_N)| = |F_N^i(\mathbf{x}_N) - F_N^i(\mathbf{y}_N)|$ and for all $i = 1,\dots,N$,

$$\begin{aligned}
&\int_{\mathbb{R}^d} |\mathrm{Diag}(F_N^i(\mathbf{x}_N) - F_N^i(\mathbf{y}_N))z|^2 \rho_z(z)dz \\
&\quad = \int_{\mathbb{R}^d} \sum_{l=1}^{d} |(F_N^i(\mathbf{x}_N))_l - (F_N^i(\mathbf{y}_N))_l|^2 |z_l|^2 \prod_{k=1}^{d} \rho_{\mathfrak{z}}(z_k)dz \\
&\quad = \sum_{l=1}^{d} |(F_N^i(\mathbf{x}_N))_l - (F_N^i(\mathbf{y}_N))_l|^2 \int_{\mathbb{R}^d} |z_l|^2 \prod_{k=1}^{d} \rho_{\mathfrak{z}}(z_k)dz \\
&\quad = |F_N^i(\mathbf{x}_N) - F_N^i(\mathbf{y}_N)|^2 \mathbb{E}(\mathcal{Z})^2,
\end{aligned}$$

where $(F_N^i(\mathbf{x}_N))_l$ means the lth component of d-dimensional vector $F_N^i(\mathbf{x}_N)$ and z_l means the lth component of d-dimensional vector z. Therefore, from Lemma 3.1, there is a positive function $K(R)$ of $R > 0$ such that

$$\begin{aligned}
&|\mathbf{F}_N(\mathbf{x}_N) - \mathbf{F}_N(\mathbf{y}_N)|^2 + |\mathbf{G}_N(\mathbf{x}_N) - \mathbf{G}_N(\mathbf{y}_N)|^2 \\
&\quad + \sum_{i=1}^{N} \int_{\mathbb{R}^d} |\mathrm{Diag}(F_N^i(\mathbf{x}_N) - F_N^i(\mathbf{y}_N))z|^2 \rho_z(z)dz \le K(R)|\mathbf{x}_N - \mathbf{y}_N|,
\end{aligned}$$

whenever $|\mathbf{x}_N|$, $|\mathbf{y}_N| \le R$. Moreover,

$$|\mathbf{F}_N(\mathbf{x}_N)|^2 + |\mathbf{G}_N(\mathbf{x}_N)|^2 + \sum_{i=1}^{N} \int_{\mathbb{R}^d} |\mathrm{Diag}(F_N^i(x_N))z|^2 \rho_z(z)dz \le C|\mathbf{x}_N|^2,$$

where C is some positive constant independent of $|\mathbf{x}_N|$. Then the proof immediately follows from Theorem 1 from Ref. 32.

Consequently, by Lemma 2.3 from Ref. 18, the following moment bound, provided $\mathbb{E}|\mathbf{X}_N(0)|^{2p} < \infty$ and $\mathbb{E}|\mathbf{Z}|^{2p} < \infty$, holds:

$$\mathbb{E} \sup_{0\le t\le T} |\mathbf{X}_N(t)|^{2p} \le C_N, \tag{3.4}$$

where C_N may depend on N and $p \ge 1$. $\square$

In the last step of the proof above, we highlighted that C_N may depend on N. However, for convergence analysis in later sections, we need a uniform in N bound for $\sup_{i=1,\dots,N} \mathbb{E}\big(\sup_{t\in[0,T]} |X_N^i(t)|^{2p}\big)$, $p \ge 1$, which we prove under the following assumptions (cf. Ref. 11).

Assumption 3.3. There exists a positive constant K_f such that

$$|f(x) - f(y)| \le K_f(1 + |x| + |y|)|x - y| \quad \text{for all } x, y \in \mathbb{R}^d.$$

Assumption 3.4. There is a constant $K_u > 0$,

$$f(x) - f_m \leq K_u(1 + |x|^2) \quad \text{for all } x \in \mathbb{R}^d.$$

Assumption 3.5. There exist constants $R > 0$ and $K_l > 0$ such that

$$f(x) - f_m \geq K_l|x|^2 \quad \text{for } |x| \geq R.$$

As one can see, we need a stronger Assumption 3.3 compared to Assumption 3.2 to obtain a moment bound uniform in N. Assumptions 3.4 and 3.5 are to make sure that objective function f has quadratic growth at infinity.

From Lemma 3.3 in Ref. 11, we have the following result under Assumptions 3.1 and 3.3–3.5:

$$\sum_{i=1}^{N} |x_N^i|^2 \frac{e^{-\alpha f(x_N^i)}}{\sum_{j=1}^{N} e^{-\alpha f(x_N^j)}} \leq L_1 + L_2 \frac{1}{N} \sum_{i=1}^{N} |x_N^i|^2, \tag{3.5}$$

where $L_1 = R^2 + L_2$ and $L_2 = 2\frac{K_u}{K_l}\left(1 + \frac{1}{\alpha K_l R^2}\right)$ with R from Assumption 3.5.

Lemma 3.2. *Let Assumptions* 3.1 *and* 3.3–3.5 *be satisfied. Let* $p \geq 1$, $\sup_{i=1,\ldots,N} \mathbb{E}|X_N^i(0)|^{2p} < \infty$ *and* $\mathbb{E}|Z|^{2p} < \infty$. *Then*

$$\sup_{i\in\{1,\ldots,N\}} \mathbb{E} \sup_{0\leq t\leq T} |X_N^i(t)|^{2p} \leq K_m,$$

where $X_N^i(t)$ *is from* (2.11) *and* K_m *is a positive constant independent of* N.

Proof. Let p be a positive integer. Using Ito's formula, we have

$$\begin{aligned}
|X_N^i(t)|^{2p} &= |X_N^i(0)|^{2p} - 2p\mathbb{E}\int_0^t \beta(s)|X_N^i(s)|^{2p-2}\big(X_N^i(s)\cdot(X_N^i(s) - \bar{X}_N(s))\big)ds \\
&\quad + 2\sqrt{2}p\int_0^t \sigma(s)|X_N^i(s)|^{2p-2}\big(X_N^i(s)\cdot \mathrm{Diag}(X_N^i(s) - \bar{X}_N(s))dW^i(s)\big) \\
&\quad + 4p(p-1)\int_0^t \sigma^2(s)|X_N^i(s)|^{2p-4}|\mathrm{Diag}(X_N^i(s) - \bar{X}_N(s))X_N^i(s)|^2 ds \\
&\quad + 2p\int_0^t \sigma^2(s)|X_N^i(s)|^{2p-2}|\mathrm{Diag}(X_N^i(s) - \bar{X}_N(s))|^2 ds \\
&\quad + \int_0^t\int_{\mathbb{R}^d}\big(|X_N^i(s^-) + \gamma(s)\,\mathrm{Diag}(X_N^i(s^-) - \bar{X}_N(s^-))z|^{2p} \\
&\quad - |X_N^i(s^-)|^{2p}\big)\mathcal{N}^i(ds,dz).
\end{aligned}$$

First taking supremum over $0 \leq t \leq T$ and then taking expectation, we get

$$\begin{aligned}
\mathbb{E}\sup_{0\leq t\leq T}|X_N^i(t)|^{2p} &\leq \mathbb{E}|X_N^i(0)|^{2p} + C\mathbb{E}\int_0^T |X_N^i(s)|^{2p-2}\big|X_N^i(s)\cdot(X_N^i(s) \\
&\quad - \bar{X}_N(s))\big|ds + 2\sqrt{2}p\mathbb{E}\sup_{0\leq t\leq T}\bigg|\int_0^t \sigma(s)|X_N^i(s)|^{2p-2}\big(X_N^i(s)
\end{aligned}$$

$$
\begin{aligned}
&\cdot \operatorname{Diag}(X_N^i(s) - \bar{X}_N(s))dW^i(s)\Big) \Bigg| \\
&+ \mathbb{E}\int_0^T |X_N^i(s)|^{2p-4} |\operatorname{Diag}(X_N^i(s) - \bar{X}_N(s))X_N^i(s)|^2 ds \\
&+ C\mathbb{E}\int_0^T |X_N^i(s)|^{2p-2} |\operatorname{Diag}(X_N^i(s) - \bar{X}_N(s))|^2 ds \\
&+ C\mathbb{E}\sup_{0\le t\le T}\int_0^t \int_{\mathbb{R}^d} \big(|X_N^i(s^-) + \gamma(s)\operatorname{Diag}(X_N^i(s^-) \\
&- \bar{X}_N(s^-))z|^{2p} - |X_N^i(s^-)|^{2p}\big)\mathcal{N}^i(ds, dz).
\end{aligned} \tag{3.6}
$$

To deal with the second term in (3.6), we use Young's inequality and obtain

$$
\begin{aligned}
|X_N^i(s)|^{2p-2}\big|X_N^i(s)\cdot(X_N^i(s) - \bar{X}_N(s))\big| &\le |X_N^i(s)|^{2p} + |X_N^i(s)|^{2p-1}|\bar{X}_N(s)| \\
&\le \frac{4p-1}{2p}|X_N^i(s)|^{2p} + \frac{1}{2p}|\bar{X}_N(s)|^{2p}.
\end{aligned}
$$

To ascertain a bound on $|\bar{X}_N(s)|^{2p}$, we first apply Jensen's inequality to $|\bar{X}_N(s)|^2$ to get

$$
\begin{aligned}
|\bar{X}_N(s)|^2 &= \left|\sum_{i=1}^N X_N^i(s)\frac{e^{-\alpha f(X_N^i(s))}}{\sum_{j=1}^N e^{-\alpha f(X_N^j(s))}}\right|^2 \\
&\le \sum_{i=1}^N |X_N^i(s)|^2 \frac{e^{-\alpha f(X_N^i(s))}}{\sum_{j=1}^N e^{-\alpha f(X_N^j(s))}},
\end{aligned}
$$

then using (3.5), we obtain $|\bar{X}_N(s)|^2 \le L_1 + L_2\frac{1}{N}\sum_{i=1}^N |X_N^i(s)|^2$, which on applying the elementary inequality, $(a+b)^p \le 2^{p-1}(a^p + b^p)$, $a, b \in \mathbb{R}_+$ and Jensen's inequality, gives

$$
|\bar{X}_N(s)|^{2p} \le 2^{p-1}\left(L_1^p + L_2^p\frac{1}{N}\sum_{i=1}^N |X_N^i(s)|^{2p}\right).
$$

As a consequence of the above calculations, we get

$$
\begin{aligned}
&|X_N^i(s)|^{2p-2}\big|X_N^i(s)\cdot(X_N^i(s) - \bar{X}_N(s))\big| \\
&\quad \le C\left(1 + |X_N^i(s)|^{2p} + \frac{1}{N}\sum_{i=1}^N |X_N^i(s)|^{2p}\right),
\end{aligned} \tag{3.7}
$$

where C is a positive constant independent of N.

Using the Burkholder–Davis–Gundy inequality, we get

$$\begin{aligned}
&\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t \sigma(s)|X_N^i(s)|^{2p-2}\big(X_N^i(s)\cdot \mathrm{Diag}(X_N^i(s)-\bar{X}_N(s))dW^i(s)\big)\right| \\
&\quad\le \mathbb{E}\left(\int_0^T \big(\sigma(s)|X_N^i(s)|^{2p-2}\big(X_N^i(s)\cdot(X_N^i(s)-\bar{X}_N(s))\big)\big)^2 ds\right)^{1/2} \\
&\quad\le C\mathbb{E}\left(\sup_{0\le t\le T}|X_N^i(t)|^{2p-1}\left(\int_0^T |X_N^i(s)-\bar{X}_N(s)|^2 ds\right)^{1/2}\right),
\end{aligned}$$

which on applying generalized Young's inequality $(ab \le (\epsilon a^{q_1})/q_1 + b^{q_2}/(\epsilon^{q_2/q_1}q_2)$, $\epsilon, q_1, q_2 > 0, 1/q_1 + 1/q_2 = 1)$ yields

$$\begin{aligned}
&2\sqrt{2}p\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t \sigma(s)|X_N^i(s)|^{2p-2}\big(X_N^i(s)\cdot \mathrm{Diag}(X_N^i(s)-\bar{X}_N(s))dW^i(s)\big)\right| \\
&\quad\le \frac{1}{2}\mathbb{E}\sup_{0\le t\le T}|X_N^i(t)|^{2p} + C\mathbb{E}\left(\int_0^T |X_N^i(s)-\bar{X}_N(s)|^2 ds\right)^p \\
&\quad\le \frac{1}{2}\mathbb{E}\sup_{0\le t\le T}|X_N^i(t)|^{2p} + C\mathbb{E}\left(\int_0^T |X_N^i(s)-\bar{X}_N(s)|^{2p} ds\right), \qquad (3.8)
\end{aligned}$$

where in the last step we have utilized Hölder's inequality.

Now, we move on to obtain estimates which are required to deal with the fourth and fifth terms in (3.6). Using Young's inequality, we have

$$\begin{aligned}
A_1 &:= |X_N^i(s)|^{2p-4}(|X_N^i(s)|^2 - (X_N^i(s)\cdot\bar{X}_N(s)))^2 \\
&\le 2|X_N^i(s)|^{2p} + 2|X_N^i(s)|^{2p-2}|\bar{X}_N(s)|^2 \\
&\le \frac{4p-2}{p}|X_N^i(s)|^{2p} + \frac{2}{p}|\bar{X}_N(s)|^{2p}. \qquad (3.9)
\end{aligned}$$

In the same way, applying Young's inequality, we obtain

$$\begin{aligned}
A_2 &:= |X_N^i(s)|^{2p-2}|\mathrm{Diag}(X_N^i(s)-\bar{X}_N(s))|^2 \\
&\le 2|X_N^i(s)|^{2p} + 2|X_N^i(s)|^{2p-2}|\bar{X}_N(s)|^2 \\
&\le \frac{4p-2}{p}|X_N^i(s)|^{2p} + \frac{2}{p}|\bar{X}_N(s)|^{2p}. \qquad (3.10)
\end{aligned}$$

Following the same procedure based on (3.5), which we followed to obtain the bound (3.7), we also get

$$A_1 + A_2 \le C\left(1 + |X_N^i(s)|^{2p} + \frac{1}{N}\sum_{i=1}^N |X_N^i(s)|^{2p}\right), \qquad (3.11)$$

where C is a positive constant independent of N.

It is left to deal with the last term in (3.6). Using the Cauchy–Bunyakovsky–Schwartz inequality, we get

$$\begin{aligned}
&\mathbb{E}\sup_{0\le t\le T}\int_0^t\int_{\mathbb{R}^d}\big(|X_N^i(s^-)+\gamma(s)\operatorname{Diag}(X_N^i(s^-)-\bar{X}_N(s^-))z|^{2p}\\
&\qquad -|X_N^i(s^-)|^{2p}\big)\mathcal{N}^i(ds,dz)\\
&\quad\le \mathbb{E}\sup_{0\le t\le T}\int_0^t\int_{\mathbb{R}^d}\big(2^{2p-1}\big(|X_N^i(s^-)|^{2p}+|\gamma(s)\operatorname{Diag}(X_N^i(s^-)-\bar{X}_N(s^-))z|^{2p}\big)\\
&\qquad -|X_N^i(s^-)|^{2p}\big)\mathcal{N}^i(ds,dz)\\
&\quad\le C\mathbb{E}\int_0^T\int_{\mathbb{R}^d}\big(|X_N^i(s^-)|^{2p}+|\gamma(s)\operatorname{Diag}(X_N^i(s^-)-\bar{X}_N(s^-))z|^{2p}\big)\mathcal{N}^i(ds,dz)\\
&\quad=\lambda C\mathbb{E}\int_0^T\int_{\mathbb{R}^d}(|X_N^i(s)|^{2p}+|\gamma(s)\operatorname{Diag}(X_N^i(s)-\bar{X}_N(s))z|^{2p}\big)\rho_z(z)dzds\\
&\quad\le C\mathbb{E}\int_0^T\left(|X_N^i(s)|^{2p}+|X_N^i(s)-\bar{X}_N(s)|^{2p}\int_{\mathbb{R}^d}|z|^{2p}\rho_z(z)dz\right)ds.
\end{aligned}$$

We have $|X_N^i(s)-\bar{X}_N(s)|^{2p}\le 2^{2p-1}\big(|X_N^i(s)|^{2p}+|\bar{X}_N(s)|^{2p}\big)\le C(1+|X_N^i(s)|^{2p}+\frac{1}{N}\sum_{i=1}^N|X_N^i(s)|^{2p})$, and hence

$$\begin{aligned}
&\mathbb{E}\sup_{0\le t\le T}\int_0^t\int_{\mathbb{R}^d}\big(|X_N^i(s^-)+\gamma(s)\operatorname{Diag}(X_N^i(s^-)-\bar{X}_N(s^-))z|^{2p}\\
&\qquad -|X_N^i(s^-)|^{2p}\big)\mathcal{N}^i(ds,dz)\\
&\quad\le C\mathbb{E}\int_0^T\left(1+|X_N^i(s)|^{2p}+\frac{1}{N}\sum_{i=1}^N|X_N^i(s)|^{2p}\right)ds,
\end{aligned}\tag{3.12}$$

where $C>0$ does not depend on N.

Using (3.7), (3.8), (3.11) and (3.12) in (3.6), we get

$$\begin{aligned}
\frac{1}{2}\mathbb{E}\sup_{0\le t\le T}|X_N^i(t)|^{2p}&\le\mathbb{E}|X_N^i(0)|^{2p}\\
&\quad+C\mathbb{E}\int_0^T\left(1+|X_N^i(s)|^{2p}+\frac{1}{N}\sum_{i=1}^N|X_N^i(s)|^{2p}\right)ds
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\sup_{0\le t\le T}|X_N^i(t)|^{2p}&\le 2\mathbb{E}|X_N^i(0)|^{2p}+C\mathbb{E}\int_0^T\left(1+\sup_{0\le u\le s}|X_N^i(u)|^{2p}\right.\\
&\quad\left.+\frac{1}{N}\sum_{i=1}^N\sup_{0\le u\le s}|X_N^i(u)|^{2p}\right)ds.
\end{aligned}$$

Taking supremum over $\{1, \ldots, N\}$, we obtain

$$\sup_{i=1,\ldots,N} \mathbb{E} \sup_{0\leq t\leq T} |X_N^i(t)|^{2p} \leq 2 \sup_{i=\{1,\ldots,N\}} \mathbb{E}|X_N^i(0)|^{2p} + C\left(1 + \int_0^T \sup_{i=1,\ldots,N} \mathbb{E} \sup_{0\leq u\leq s} |X_N^i(u)|^{2p} ds\right),$$

which gives the required result for positive integers p by applying Grönwall's lemma (note that we can apply Grönwall's lemma due to (3.4)). We can extend the result to non-integer values of $p \geq 1$ using Hölder's inequality. □

3.2. *Well-posedness of the mean-field jump-diffusion SDEs*

In this section, we first introduce Wasserstein metric and state Lemma 3.3 which is crucial for establishing well-posedness of the mean-field limit. Then, we prove existence and uniqueness of the McKean–Vlasov jump-diffusion SDEs (2.12) in Theorem 3.2.

Let $\mathbb{D}([0,T];\mathbb{R}^d)$ be the space of $\mathbb{R}^d$ valued cádlág functions and $\mathcal{P}_p(\mathbb{R}^d)$, $p \geq 1$, be the space of probability measures on the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that for any $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty$, and which is equipped with the p-Wasserstein metric

$$\mathcal{W}_p(\mu, \vartheta) := \inf_{\pi \in \prod(\mu,\vartheta)} \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \pi(dx, dy)\right)^{\frac{1}{p}},$$

where $\prod(\mu, \vartheta)$ is the set of couplings of $\mu, \vartheta \in \mathcal{P}_p(\mathbb{R}^d)$.[52]

Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} |x|^2\mu(dx) \leq K$. Then, using Jensen's inequality, we have

$$e^{-\alpha \int_{\mathbb{R}^d} f(x)\mu(dx)} \leq \int_{\mathbb{R}^d} e^{-\alpha f(x)}\mu(dx)$$

and the simple rearrangement together with Assumption 3.4 gives

$$\frac{e^{-\alpha f_m}}{\int_{\mathbb{R}^d} e^{-\alpha f(x)}\mu(dx)} \leq e^{\alpha(\int_{\mathbb{R}^d} f(x)\mu(dx) - f_m)} \leq e^{\alpha K_u \int_{\mathbb{R}^d}(1+|x|^2)\mu(dx)} \leq C_K, \tag{3.13}$$

where $C_K > 0$ is a constant. We will also need the following notation:

$$\bar{X}^\mu = \frac{\int_{\mathbb{R}^d} x e^{-\alpha f(x)}\mu(dx)}{\int_{\mathbb{R}^d} e^{-\alpha f(x)}\mu(dx)}, \tag{3.14}$$

where $\mu \in \mathcal{P}_4(\mathbb{R}^d)$.

The next lemma is required for proving well-posedness of the McKean–Vlasov SDEs (2.14). Its proof is available in Ref. 11 (see its Lemma 3.2).

Lemma 3.3. *Let Assumptions* 3.1 *and* 3.3–3.5 *hold and there exists a constant* $K > 0$ *such that* $\int |x|^4\mu(dx) \leq K$ *and* $\int |y|^4\vartheta(dy) \leq K$ *for all* $\mu, \vartheta \in \mathcal{P}_4(\mathbb{R}^d)$*, then* *the following inequality is satisfied*:

$$|\bar{X}^{\mu} - \bar{X}^{\vartheta}| \leq C\mathcal{W}_2(\mu, \vartheta),$$

where $C > 0$ *is independent of* μ *and* ϑ.

Theorem 3.2. *Let Assumptions* 3.1 *and* 3.3–3.5 *hold, and let* $\mathbb{E}|X(0)|^4 < \infty$ *and* $\int_{\mathbb{R}^d} |z|^4 \rho_z(z)dz < \infty$. *Then, there exists a unique non-linear process* $X \in \mathbb{D}([0,T];\mathbb{R}^d)$, $T > 0$, *which satisfies the McKean–Vlasov SDEs* (2.14) *in the strong sense.*

Proof. Let $v \in C([0,T];\mathbb{R}^d)$. Consider the following SDEs:

$$\begin{aligned} dX_v(t) &= -\beta(t)(X_v(t) - v(t))dt + \sigma(t)\operatorname{Diag}(X_v(t) - v(t))dW(t) \\ &\quad + \gamma(t)\int_{\mathbb{R}^d} \operatorname{Diag}(X_v(t^-) - v(t))z\mathcal{N}(dt, dz) \end{aligned} \tag{3.15}$$

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

Note that $v(t)$ is a deterministic function of t, therefore the coefficients of SDEs (3.15) only depend on x and t. The coefficients are globally Lipschitz continuous and have linear growth in x. The existence and uniqueness of a process $X_v \in \mathbb{D}([0,T];\mathbb{R}^d)$ satisfying SDEs with Lévy noise (3.15) follows from pp. 311−312 in Ref. 3. We also have $\int_{\mathbb{R}^d} |x|^4 \mathcal{L}_{X_v(t)}(dx) = \mathbb{E}|X_v(t)|^4 \leq \sup_{t\in[0,T]} \mathbb{E}|X_v(t)|^4 \leq K$, where K is a positive constant depending on v and T and $\mathcal{L}_{X_v(t)}$ represents the law of $X_v(t)$.

We define a mapping

$$\mathbb{T} : C([0,T];\mathbb{R}^d) \to C([0,T];\mathbb{R}^d), \quad \mathbb{T}(v) = \bar{X}_v, \tag{3.16}$$

such that

$$\begin{aligned} \mathbb{T}v(t) &= \bar{X}_v(t) = \mathbb{E}(X_v(t)e^{-\alpha f(X_v(t))})/\mathbb{E}(e^{-\alpha f(X_v(t))}) \\ &= \int_{\mathbb{R}^d} xe^{-\alpha f(x)}\mathcal{L}_{X_v(t)}(dx) \Big/ \int_{\mathbb{R}^d} e^{-\alpha f(x)}\mathcal{L}_{X_v(t)}(dx) = \bar{X}^{\mathcal{L}_{X_v(t)}}(t), \end{aligned}$$

where the last equality is due to (3.14).

Let $\delta \in (0,1)$. For all $t, t+\delta \in (0,T)$, Ito's isometry provides

$$\begin{aligned} \mathbb{E}|X_v(t+\delta) - X_v(t)|^2 &\leq C\int_t^{t+\delta} \mathbb{E}|X_v(s) - v(s)|^2 ds \\ &\quad + C\int_t^{t+\delta}\int_{\mathbb{R}^d} \mathbb{E}|X_v(s) - v(s)|^2|z|^2\rho(z)dzds \leq C\delta, \end{aligned} \tag{3.17}$$

where C is a positive constant independent of δ. Using Lemma 3.3 and (3.17), we obtain

$$\begin{aligned} |\bar{X}_v(t+\delta) - \bar{X}_v(t)| &= |\bar{X}^{\mathcal{L}_{X_v(t+\delta)}}(t+\delta) - \bar{X}^{\mathcal{L}_{X_v(t)}}(t)| \leq C\mathcal{W}_2(\mathcal{L}_{X_v(t+\delta)}, \mathcal{L}_{X_v(t)}) \\ &\leq C\big(\mathbb{E}|X_v(t+\delta) - X_v(t)|^2\big)^{1/2} \leq C|\delta|^{1/2}, \end{aligned}$$

where C is a positive constant independent δ. This implies the Hölder continuity of the map $t \to \bar{X}_v(t)$. Therefore, the compactness of $\mathbb{T}$ follows from the compact embedding $C^{0,\frac{1}{2}}([0,T];\mathbb{R}^d) \hookrightarrow C([0,T];\mathbb{R}^d)$.

Using Ito's isometry, we have

$$\begin{aligned}\mathbb{E}|X_v(t)|^2 &\leq 4\bigg(\mathbb{E}|X_v(0)|^2 + \mathbb{E}\bigg|\int_0^t \beta(s)(X_v(s)-v(s))ds\bigg|^2 \\ &\quad + \mathbb{E}\bigg|\int_0^t \sigma(s)\,\mathrm{Diag}(X_v(s)-v(s))dW(s)\bigg|^2 \\ &\quad + \mathbb{E}\bigg|\int_0^t \gamma(s)\,\mathrm{Diag}(X_v(s^-)-v(s))z\mathcal{N}(ds,dz)\bigg|^2\bigg) \\ &\leq C\bigg(1+\int_0^t \mathbb{E}|X_v(s)-v(s)|^2 ds\bigg) \\ &\leq C\bigg(1+\int_0^t (\mathbb{E}|X_v(s)|^2 + |v(s)|^2)ds\bigg), \end{aligned} \tag{3.18}$$

where C is a positive constant independent of v. Moreover, we have the following result under Assumptions 3.1 and 3.3–3.5 (see Lemma 3.3 in Ref. 11):

$$|\bar{X}_v(t)|^2 \leq L_1 + L_2\mathbb{E}|X_v(t)|^2, \tag{3.19}$$

where L_1 and L_2 are from (3.5). Consider a set $\mathcal{S} = \{v \in C([0,T];\mathbb{R}^d) : v = \epsilon\mathbb{T}v,\ 0 \leq \epsilon \leq 1\}$. The set $\mathcal{S}$ is non-empty due to the fact that $\mathbb{T}$ is compact (see the remark after Theorem 10.3 in Ref. 27). Therefore, for any $v \in \mathcal{S}$, we have the corresponding unique process $X_v(t) \in \mathbb{D}([0,T];\mathbb{R}^d)$ satisfying (3.15), and $\mathcal{L}_{X_v(t)}$ represents the law of $X_v(t)$, such that the following holds due to (3.19):

$$|v(s)|^2 = \epsilon^2|\mathbb{T}v(s)|^2 = \epsilon^2|\bar{X}_v(s)|^2 \leq \epsilon^2\big(L_1 + L_2\mathbb{E}|X_v(s)|^2\big) \tag{3.20}$$

for all $s \in [0,T]$. Substituting (3.20) in (3.18), we get

$$\mathbb{E}|X_v(t)|^2 \leq C\bigg(1+\int_0^t \mathbb{E}|X_v(s)|^2 ds\bigg),$$

which on applying Grönwall's lemma gives

$$\mathbb{E}|X_v(t)|^2 \leq C, \tag{3.21}$$

where C is independent of v. Due to (3.20) and (3.21), we can claim the boundedness of the set $\mathcal{S}$. Therefore, from the Leray–Schauder theorem (see Theorem 10.3 in Ref. 27) there exists a fixed point of the mapping $\mathbb{T}$. This proves existence of the solution of (2.14).

Let v_1 and v_2 be two fixed points of the mapping $\mathbb{T}$ and let us denote the corresponding solutions of (3.15) as X_{v_1} and X_{v_2}. Using Ito's isometry, we can get

$$\mathbb{E}|X_{v_1}(t) - X_{v_2}(t)|^2 \leq \mathbb{E}|X_{v_1}(0) - X_{v_2}(0)|^2 + C\int_0^t \big(\mathbb{E}|X_{v_1}(s) - X_{v_2}(s)|^2 + |v_1(s) - v_2(s)|^2\big)ds. \tag{3.22}$$

Note that $\mathcal{S}$ is a bounded set and by definition v_1 and v_2 belong to $\mathcal{S}$. Then, we can apply Lemma 3.3 to ascertain

$$\begin{aligned}|v_1(s) - v_2(s)|^2 = |\bar{X}_{v_1}(s) - \bar{X}_{v_2}(s)|^2 &\leq C\mathcal{W}_2(\mathcal{L}_{X_{v_1}(s)}, \mathcal{L}_{X_{v_2}(s)}) \\ &\leq C\mathbb{E}|X_{v_1}(s) - X_{v_2}(s)|^2.\end{aligned}$$

Using the above estimate, Grönwall's lemma and the fact $X_{v_1}(0) = X_{v_2}(0)$ in (3.22), we get uniqueness of the solution of (2.14). □

The strong existence and uniqueness of the mean-field SDEs (2.14) also implies existence of the solution of the Fokker–Planck equation in weak sense (cf. Theorem 3.1 in Ref. 11). This implication follows from application of Ito's formula as discussed below.

Let $\varphi \in C_b^2(\mathbb{R}^d)$. Applying Ito's formula to $\varphi(X(t))$, where $X(t)$ is from (2.14), we have

$$\begin{aligned}\varphi(X(t)) = \varphi(X(0)) &- \int_0^t \beta(s)\big(\nabla\varphi(X(s)) \cdot (X(s) - \bar{X}(s))\big)ds \\ &+ \int_0^t \sigma^2(s)\sum_{j=1}^d (X(s) - \bar{X}(s))_j^2 \frac{\partial^2}{\partial x_j^2}\varphi(X(s))ds \\ &+ \sqrt{2}\int_0^t \sigma(s)\big(\nabla\varphi(X(s)) \cdot \mathrm{Diag}(X(s) - \bar{X}(s))dW(s)\big) \\ &+ \int_0^t \int_{\mathbb{R}^d} (\varphi(X(s^-) + \gamma(s)\,\mathrm{Diag}(X(s^-) - \bar{X}(s^-))z) \\ &- \varphi(X(s^-)))\mathcal{N}(ds, dz),\end{aligned}$$

which on taking expectation and writing in the differential form becomes the Fokker–Planck equation in weak sense:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\big(\varphi(X(t))\big) = &-\beta(t)\mathbb{E}\big(\nabla\varphi(X(t)) \cdot (X(t) - \bar{X}(t))\big) \\ &+ \sigma^2(t)\sum_{j=1}^d \mathbb{E}\bigg((X(t) - \bar{X}(t))_j^2 \frac{\partial^2}{\partial x_j^2}\varphi(X(t))\bigg) \\ &+ \lambda\int_{\mathbb{R}^d} \mathbb{E}\big(\varphi(X(t) + \gamma(t)\,\mathrm{Diag}(X(t) - \bar{X}(t))z) \\ &- \varphi(X(t))\big)\rho_z(z)dz.\end{aligned} \tag{3.23}$$

This equation can be also written in the following compact form. For notational convenience, let $\mu_t = \mathcal{L}_{X(t)}$, where $X(t)$ is from (2.14). Define $\mu_t^{(\gamma)}$ as $\int_{\mathbb{R}^d} \varphi(x)\mu_t^{(\gamma)}(dx) = \int_{\mathbb{R}^d} \varphi(x+\gamma(t)\,\mathrm{Diag}(x-\bar{X}^{\mu_t})z)\mu_t(dx)$ for all $\varphi \in C_b^2(\mathbb{R}^d)$, where $\bar{X}^{\mu_t}$ is from (3.14). Then we can say, based on (3.23), that $\mu_t \in \mathcal{P}_4(\mathbb{R}^d)$ satisfies the Fokker–Planck equation associated with (2.14) in the weak sense for all $t \in [0,T]$ (see e.g. Ref. 1):

$$\frac{\partial \mu_t}{\partial t} = \sigma^2(t)\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}((x_j - (\bar{X}^{\mu_t})_j)^2\mu_t) + \beta(t)\sum_{j=1}^d \frac{\partial}{\partial x_j}((x_j-(\bar{X}^{\mu_t})_j)\mu_t)$$
$$+ \lambda\int_{\mathbb{R}^d}(\mu_t^{(\gamma)} - \mu_t)\rho_z(z)dz. \tag{3.24}$$

One can notice that (3.24) is a degenerate non-linear partial integral differential equation.

Theorem 3.3. *Let Assumptions* 3.1 *and* 3.3–3.5 *are satisfied. Let* $p \geq 1$, $\mathbb{E}|X(0)|^{2p} < \infty$ *and* $\mathbb{E}|Z|^{2p} < \infty$, *then*

$$\mathbb{E}\sup_{0\leq t\leq T}|X(t)|^{2p} \leq K_p,$$

where $X(t)$ *satisfies* (2.14) *and* K_p *is a positive constant.*

Proof. Recall that under the assumptions of this theorem, Theorem 3.2 guarantees existence of a strong solution $X(t)$ of (2.14).

Let p be a positive integer. Let us denote $\theta_R = \inf\{s \geq 0; |X(s)| \geq R\}$. Using Ito's formula and then taking suprema over $0 \leq t \leq T\wedge\theta_R$ and expectations, we obtain

$$\begin{aligned}
\mathbb{E}\sup_{0\leq t\leq T\wedge\theta_R}|X(t)|^{2p} &\leq \mathbb{E}|X(0)|^{2p} + C\mathbb{E}\int_0^{T\wedge\theta_R}|X(s)|^{2p-2}\big|X(s)\cdot(X(s)-\bar{X}(s))\big|ds \\
&\quad + 2\sqrt{2}p\mathbb{E}\sup_{0\leq t\leq T\wedge\theta_R}\bigg|\int_0^t \sigma(s)|X(s)|^{2p-2}\big(X(s) \\
&\quad \cdot(\mathrm{Diag}(X(s)-\bar{X}(s))dW(s))\big)\bigg| \\
&\quad + C\mathbb{E}\int_0^{T\wedge\theta_R}|X(s)|^{2p-4}|\mathrm{Diag}(X(s)-\bar{X}(s))X(s)|^2ds \\
&\quad + C\mathbb{E}\int_0^{T\wedge\theta_R}|X(s)|^{2p-2}|\mathrm{Diag}(X(s)-\bar{X}(s))|^2ds \\
&\quad + \mathbb{E}\sup_{0\leq t\leq T\wedge\theta_R}\int_0^t\int_{\mathbb{R}^d}(|X(s^-)+\gamma(s)\,\mathrm{Diag}(X(s^-) \\
&\quad - \bar{X}(s^-))z|^{2p} - |X(s^-)|^{2p})\mathcal{N}(ds,dz).
\end{aligned} \tag{3.25}$$

To deal with the second term in (3.25), we use Young's inequality and ascertain

$$|X(s)|^{2p-2}\big|X(s)\cdot(X(s)-\bar{X}(s))\big| \le C(|X(s)|^{2p}+|\bar{X}(s)|^{2p}). \tag{3.26}$$

Using the Burkholder–Davis–Gundy inequality, we have

$$\begin{aligned} &\mathbb{E}\sup_{0\le t\le T\wedge\theta_R}\left|\int_0^t \sigma(s)|X(s)|^{2p-2}\big(X(s)\cdot(\mathrm{Diag}(X(s)-\bar{X}(s))dW(s))\big)\right| \\ &\quad\le \mathbb{E}\left(\int_0^{T\wedge\theta_R}\sigma^2(s)|X(s)|^{4p-2}|X(s)-\bar{X}(s)|^2ds\right)^{1/2} \\ &\quad\le C\mathbb{E}\left(\sup_{0\le t\le T\wedge\theta_R}|X(t)|^{2p-1}\left(\int_0^{T\wedge\theta_R}|X(s)-\bar{X}(s)|^2ds\right)^{1/2}\right). \end{aligned} \tag{3.27}$$

We apply generalized Young's inequality $\big(ab\le(\epsilon a^{q_1})/q_1+b^{q_2}/(\epsilon^{q_2/q_1}q_2),\ \epsilon,q_1,q_2>0, 1/q_1+1/q_2=1\big)$ and Hölder's inequality on the right-hand side of (3.27) to get

$$\begin{aligned} &2\sqrt{2}p\mathbb{E}\sup_{0\le t\le T\wedge\theta_R}\left|\int_0^t\sigma(s)|X(s)|^{2p-2}\big(X(s)\cdot\mathrm{Diag}(X(s)-\bar{X}(s))dW(s)\big)\right| \\ &\quad\le\frac{1}{2}\mathbb{E}\sup_{0\le t\le T\wedge\theta_R}|X(t)|^{2p}+C\mathbb{E}\left(\int_0^{T\wedge\theta_R}|X(s)-\bar{X}(s)|^2ds\right)^p \\ &\quad\le\frac{1}{2}\mathbb{E}\sup_{0\le t\le T\wedge\theta_R}|X(t)|^{2p}+C\mathbb{E}\left(\int_0^{T\wedge\theta_R}\big(|X(s)|^{2p}+|\bar{X}(s)|^{2p}\big)ds\right). \end{aligned} \tag{3.28}$$

We have the following estimate to use for the fourth term in (3.25):

$$|X(s)|^{2p-4}|\mathrm{Diag}(X(s)-\bar{X}(s))X(s)|^2\le C\big(|X(s)|^{2p}+|\bar{X}(s)|^{2p}\big). \tag{3.29}$$

We make use of Minkowski's and Young's inequalities to get

$$\begin{aligned} |X(s)|^{2p-2}|\mathrm{Diag}(X(s)-\bar{X}(s))|^2 &\le 2|X(s)|^{2p}+2|X(s)|^{2p-2}|\bar{X}(s)|^2, \\ &\le C(|X(s)|^{2p}+|\bar{X}(s)|^{2p}). \end{aligned} \tag{3.30}$$

Now, we find an estimate for the last term in (3.25). Using the Cauchy–Bunyakovsky–Schwartz inequality, we obtain

$$\begin{aligned} &\mathbb{E}\sup_{0\le t\le T\wedge\theta_R}\int_0^t\int_{\mathbb{R}^d}(|X(s^-)+\gamma(s)\,\mathrm{Diag}(X(s^-)-\bar{X}(s^-))z|^{2p} \\ &\qquad-|X(s^-)|^{2p})\mathcal{N}(ds,dz) \end{aligned}$$

$$\leq \mathbb{E} \sup_{0\leq t\leq T\wedge\theta_R} \int_0^t \int_{\mathbb{R}^d} 2^{2p-1}(|X(s^-)|^{2p} + |\gamma(s)\operatorname{Diag}(X(s^-) - \bar{X}(s^-))z|^{2p})$$
$$- |X(s^-)|^{2p}\mathcal{N}(ds, dz)$$
$$\leq C\mathbb{E}\int_0^{T\wedge\theta_R} \int_{\mathbb{R}^d} (|X(s^-)|^{2p} + |\gamma(s)\operatorname{Diag}(X(s^-) - \bar{X}(s^-))z|^{2p})\mathcal{N}(ds, dz).$$

Since $\int_0^t \int_{\mathbb{R}^d} (|X(s^-)|^{2p} + |\gamma(s)\operatorname{Diag}(X(s^-) - \bar{X}(s^-))z|^{2p})\mathcal{N}(ds, dz) - \lambda \int_0^t \int_{\mathbb{R}^d} \times (|X(s^-)|^{2p} + |\gamma(s)\operatorname{Diag}(X(s^-) - \bar{X}(s^-))z|^{2p})\rho_z(z)dzds$ is a martingale, using Doob's optional stopping theorem (see e.g. Theorem 2.2.1 in Ref. 3), we get

$$\mathbb{E} \sup_{0\leq t\leq T\wedge\theta_R} \int_0^t \int_{\mathbb{R}^d} (|X(s^-) + \gamma(s)\operatorname{Diag}(X(s^-) - \bar{X}(s^-))z|^{2p}$$
$$- |X(s^-)|^{2p})\mathcal{N}(ds, dz)$$
$$\leq C\mathbb{E}\int_0^{T\wedge\theta_R} \int_{\mathbb{R}^d} (|X(s)|^{2p} + |\gamma(s)\operatorname{Diag}(X(s) - \bar{X}(s))z|^{2p})\rho_z(z)dzds$$
$$\leq C\mathbb{E}\int_0^{T\wedge\theta_R} \left(|X(s)|^{2p} + |\bar{X}(s)|^{2p}\right)\left(1 + \int_{\mathbb{R}^d} |z|^{2p}\rho_z(z)dz\right)ds$$
$$\leq C\mathbb{E}\int_0^{T\wedge\theta_R} \left(|X(s)|^{2p} + |\bar{X}(s)|^{2p}\right)ds. \tag{3.31}$$

We have the following result under Assumptions 3.1 and 3.3–3.5 (see Lemma 3.3 in Ref. 11):

$$|\bar{X}(s)|^2 \leq L_1 + L_2\mathbb{E}|X(s)|^2, \tag{3.32}$$

where L_1 and L_2 are from (3.5).

Substituting (3.26) and (3.28)–(3.32) in (3.25) and using Hölder's inequality, we arrive at the following bound:

$$\mathbb{E} \sup_{0\leq t\leq T\wedge\theta_R} |X(t)|^{2p} \leq 2\mathbb{E}|X(0)|^{2p} + C\mathbb{E}\int_0^{T\wedge\theta_R} (|X(s)|^{2p} + |\bar{X}(s)|^{2p})ds$$
$$\leq C + C\mathbb{E}\int_0^{T\wedge\theta_R} (1 + |X(s)|^{2p} + \mathbb{E}|X(s)|^{2p})ds$$
$$\leq C + C\int_0^T \mathbb{E} \sup_{0\leq u\leq s\wedge\theta_R} |X(u)|^{2p} ds$$

and using Grönwall's lemma, we obtain

$$\mathbb{E} \sup_{0\leq t\leq T\wedge\theta_R} |X(t)|^{2p} \leq C,$$

where $C > 0$ is independent of R. Then, tending $R \to \infty$ and applying Fatou's lemma give the desired result. $\square$

4. Three Convergence Results

In Sec. 4.1, we prove convergence of $X(t)$, which is the mean-field limit of the particle system (2.11), towards the global minimizer of the considered optimization problem. This convergence proof is based on the Laplace principle. In Sec. 4.2, we prove convergence of the interacting particle system (2.11) to the mean-field limit (2.14) as $N \to \infty$. In Sec. 4.3, we prove uniform in N convergence of the Euler scheme (2.19) to (2.11) as $h \to 0$, where h is the discretization time step.

4.1. *Convergence towards the global minimum*

The aim of this section is to show that the non-linear process $X(t)$ driven by the distribution dependent SDEs (2.12) converges to a point x^* which lies in a close vicinity of the global minimum which we denote $x_{\min}$. To this end, we will first prove that $\mathrm{Var}(t) := \mathbb{E}|X(t) - \mathbb{E}(X(t))|^2$ satisfies a differential inequality which, with particular choice of parameters, implies exponential decay of $\mathrm{Var}(t)$ as $t \to \infty$. We also obtain a differential inequality for $M(t) := \mathbb{E}\big(e^{-\alpha f(X(t))}\big)$.

The approach that we follow in this subsection is along the lines of Refs. 11 and 13. The main result (Theorem 4.1) of this subsection differs from Refs. 11 and 13 in two respects. First, in our model (2.11), the parameters are time-dependent. Second, we need to treat the jump part of (2.11).

Lemma 4.1. *Under Assumptions* 3.1 *and* 3.3–3.5, *the following inequality is satisfied for* $\mathrm{Var}(t)$:

$$\frac{d}{dt}\,\mathrm{Var}(t) \le -\left(2\beta(t) - \big(2\sigma^2(t) + \lambda\gamma^2(t)\mathbb{E}|Z|^2\big)\left(1 + \frac{e^{-\alpha f_m}}{M(t)}\right)\right)\mathrm{Var}(t). \quad (4.1)$$

Proof. Using Ito's formula, we have

$$\begin{aligned}|X(t) - \mathbb{E}X(t)|^2 = {} & |X(0) - \mathbb{E}X(0)|^2 - 2\int_0^t \beta(s)(X(s) - \mathbb{E}X(s))\cdot(X(s) - \bar{X}(s))ds \\ & - 2\int_0^t (X(s) - \mathbb{E}X(s))\cdot d\mathbb{E}X(s) + 2\int_0^t \sigma^2(s)|X(s) - \bar{X}(s)|^2 ds \\ & + 2\sqrt{2}\int_0^t \sigma(s)\big((X(s) - \mathbb{E}X(s))\cdot \mathrm{Diag}(X(s) - \bar{X}(s))dW(s)\big) \\ & + \int_0^t\int_{\mathbb{R}^d} \big\{|X(s^-) - \mathbb{E}X(s^-) + \gamma(s)\,\mathrm{Diag}(X(s^-) - \bar{X}(s^-))z|^2 \\ & - |X(s^-) - \mathbb{E}(X(s^-))|^2\big\}\mathcal{N}(ds, dz).\end{aligned}$$

Taking expectation on both sides, we get

$$\begin{aligned}
\mathrm{Var}(t) &= \mathrm{Var}(0) - 2\mathbb{E}\int_0^t \beta(s)\mathbb{E}\big((X(s)-\mathbb{E}X(s))\cdot(X(s)-\bar{X}(s))\big)ds \\
&\quad + 2\int_0^t \sigma^2(s)\mathbb{E}|X(s)-\bar{X}(s)|^2 ds \\
&\quad + \lambda\int_0^t\int_{\mathbb{R}^d}\gamma^2(s)\big(\mathbb{E}|\mathrm{Diag}(X(s)-\bar{X}(s))z|^2\rho_z(z)\big)dzds \\
&= \mathrm{Var}(0) + \int_0^t \big(-2\beta(s)\,\mathrm{Var}(s) + 2\sigma^2(s)\mathbb{E}|X(s)-\bar{X}(s)|^2 \\
&\quad + \lambda\gamma^2(s)\mathbb{E}|\mathcal{Z}|^2\mathbb{E}|X(s)-\bar{X}(s)|^2\big)ds, \qquad (4.2)
\end{aligned}$$

since

$$\begin{aligned}
&\mathbb{E}\big((X(t)-\mathbb{E}X(t))\cdot(\mathbb{E}X(t)-\bar{X}(t))\big) = 0, \\
&|X(t)-\mathbb{E}X(t)+\mathrm{Diag}(X(t)-\bar{X}(t))z|^2 \\
&\qquad = |X(t)-\mathbb{E}X(t)|^2 + |\mathrm{Diag}(X(t)-\bar{X}(t))z|^2 + 2\big((X(t)-\mathbb{E}X(t)) \\
&\qquad\quad \cdot\,\mathrm{Diag}(X(t)-\bar{X}(t))z\big), \\
&\int_{\mathbb{R}^d}\big((X(t)-\mathbb{E}X(t))\cdot\mathrm{Diag}(X(t)-\bar{X}(t))z\big)\rho_z(z)dz = 0.
\end{aligned}$$

Moreover, $\int_{\mathbb{R}^d}\sum_{l=1}^d (X(t)-\bar{X}(t))_l^2 z_l^2\rho_z(z)dz = \sum_{l=1}^d (X(t)-\bar{X}(t))_l^2\int_{\mathbb{R}^d} z_l^2 \times \prod_{i=1}^d \rho_{\mathcal{z}}(z_i)dz = |X(t)-\bar{X}(t)|^2\mathbb{E}|\mathcal{Z}|^2$, since each component Z_l of Z is distributed as $\mathcal{Z}$.

We also have

$$\mathbb{E}|X(t)-\bar{X}(t)|^2 = \mathrm{Var}(t) + |\mathbb{E}X(t)-\bar{X}(t)|^2. \qquad (4.3)$$

We estimate the term $|\mathbb{E}(X(t))-\bar{X}(t)|^2$ using Jensen's inequality as

$$\begin{aligned}
|\mathbb{E}X(t)-\bar{X}(t)|^2 &= \left|\mathbb{E}X(t) - \frac{\mathbb{E}X(t)e^{-\alpha f(X(t))}}{\mathbb{E}e^{-\alpha f(X(t))}}\right|^2 \\
&= \left|\mathbb{E}\left((\mathbb{E}X(t)-X(t))\frac{e^{-\alpha f(X(t))}}{\mathbb{E}e^{-\alpha f(X(t))}}\right)\right|^2 \\
&= \left|\int_{\mathbb{R}^d}(\mathbb{E}X(t)-x)\vartheta_{X(t)}(dx)\right|^2 \le \int_{\mathbb{R}^d}\left|\mathbb{E}X(t)-x\right|^2\vartheta_{X(t)}(dx) \\
&= \mathbb{E}\left(|X(t)-\mathbb{E}(X(t))|^2\frac{e^{-\alpha f(X(t))}}{\mathbb{E}e^{-\alpha f(X(t))}}\right) \le \frac{e^{-\alpha f_m}}{M(t)}\mathrm{Var}(t), \qquad (4.4)
\end{aligned}$$

where $\vartheta_{X(t)}(dx) = e^{-\alpha f(x)}/\mathbb{E}(e^{-\alpha f(X(t))})\mathcal{L}_{X(t)}(dx)$ which implies $\int_{\mathbb{R}^d}\vartheta_{X(t)}(dx) = 1$. Using (4.3) and (4.4) in (4.2) gives the targeted result. □

To prove the main result of this subsection, we need an additional inequality, which is obtained under the following assumption.

Assumption 4.1. $f \in C^2(\mathbb{R}^d)$ and there exist three constants K_1, $K_2, K_3 > 0$ such that the following inequalities are satisfied for sufficiently large α:

(i) $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq -K_1|x - y|^2$ for all x, $y \in \mathbb{R}^d$.
(ii) $\alpha(\frac{\partial f}{\partial x_i})^2 - \frac{\partial^2 f}{\partial x_i^2} \geq -K_2$ for all $i = 1, \ldots, d$ and $x \in \mathbb{R}^d$.
(iii) $|\mathbb{E}f(x + \mathrm{Diag}(x)Z) - f(x)| \leq K_3|x|^2\mathbb{E}|\mathcal{Z}|^2$,
where Z is a d-dimensional random vector and $\mathcal{Z}$ is a real-valued random variable introduced in Sec. 2.2.

We note that for $f(x) = 1 + |x|^2$, $x \in \mathbb{R}^d$, we have $\mathbb{E}|x + \mathrm{Diag}(x)Z|^2 - |x|^2 = \mathbb{E}|\mathrm{Diag}(x)Z|^2 = \sum_{l=1}^d \mathbb{E}(x_l Z_l)^2$. However, each Z_l is distributed as $\mathcal{Z}$. Hence, $\mathbb{E}|x+\mathrm{Diag}(x)Z|^2 - |x|^2 = |x|^2\mathbb{E}|\mathcal{Z}|^2$. The conditions (i) and (ii) are straightforward to verify for $1 + |x|^2$. This implies the existence of a function satisfying the above assumption. This ensures that the class of functions satisfying the above assumption is not empty and is consistent with Assumptions 3.1 and 3.3–3.5. The most important implication is that the above assumption allows f to have quadratic growth at infinity which is important for several loss functions in machine learning problems.

In Ref. 11, the authors assumed $f \in C^2(\mathbb{R}^d)$, the norm of Hessian of f being bounded by a constant, and the norm of gradient and Laplacian of f satisfying the inequality, $\Delta f \leq c_0 + c_1|\nabla f|^2$, where c_0 and c_1 are positive constants. Therefore, in Assumption 4.1, we have imposed restrictions on f similar to Ref. 11 in the essence of regularity but adapted to our jump-diffusion case with component-wise Wiener noise.

Lemma 4.2. *The following inequality holds under Assumptions* 3.1, 3.3–3.5 *and* 4.1:

$$\frac{d}{dt}\, M^2(t) \geq -4\alpha e^{-\alpha f_m}\left(\beta(t)K_1 + \sigma^2(t)K_2 + \lambda\gamma^2(t)K_3\mathbb{E}|\mathcal{Z}|^2\right)\mathrm{Var}(t), \tag{4.5}$$

where the constants K_1, K_2 *and* K_3 *are from Assumption* 4.1.

Proof. Using Ito's formula, we get

$$\begin{aligned} e^{-\alpha f(X(t))} = &\int_0^t \alpha\beta(s)e^{-\alpha f(X(s))}\big(\nabla f(X(s)) \cdot (X(s) - \bar{X}(s))\big)ds \\ &- \sqrt{2}\int_0^t \alpha\sigma(s)e^{-\alpha f(X(s))}\big(\nabla f(X(s)) \cdot \mathrm{Diag}(X(s) - \bar{X}(s))dW(s)\big) \\ &+ \int_0^t \sigma^2(s)e^{-\alpha f(X(s))}\sum_{j=1}^d \bigg(\big(X(s) - \bar{X}(s)\big)_j^2 \end{aligned}$$

$$\times\left(\alpha^2\left(\frac{\partial f(X(s))}{\partial x_j}\right)^2-\alpha\frac{\partial^2 f(X(s))}{\partial x_j^2}\right)\Bigg)ds$$
$$+\int_0^t\int_{\mathbb{R}^d}\left(e^{-\alpha f(X(s^-)+\gamma(s)\,\mathrm{Diag}(X(s^-)-\bar{X}(s^-))z)}\right.$$
$$\left.-e^{-\alpha f(X(s^-))}\right)\mathcal{N}(ds,dz).$$

Taking expectation on both sides and writing the equation in the differential form yield

$$\begin{aligned}d\mathbb{E}e^{-\alpha f(X(t))}&=\alpha\beta(t)\mathbb{E}\big(e^{-\alpha f(X(t))}(\nabla f(X(t))-\nabla f(\bar{X}(t)))\cdot(X(t)-\bar{X}(t))\big)dt\\&\quad+\sigma^2(t)\mathbb{E}\left(e^{-\alpha f(X(t))}\sum_{j=1}^d\left((X(t)-\bar{X}(t))_j^2\right.\right.\\&\quad\times\left.\left.\left(\alpha^2\left(\frac{\partial f(X(t))}{\partial x_j}\right)^2-\alpha\frac{\partial^2 f(X(t))}{\partial x_j^2}\right)\right)\right)dt\\&\quad+\lambda\int_{\mathbb{R}^d}\mathbb{E}\Big(e^{-\alpha f(X(t)+\gamma(t)\,\mathrm{Diag}(X(t)-\bar{X}(t))z)}-e^{-\alpha f(X(t))}\Big)\rho_z(z)dzdt,\end{aligned}$$

where we have used the fact $\mathbb{E}\big[e^{-\alpha f(X(t))}(\nabla f(\bar{X}(t))\cdot(X(t)-\bar{X}(t)))\big]=0$.

Note that $|e^{-\alpha f(x)}-e^{-\alpha f(y)}|\le\alpha e^{-\alpha f_m}|f(x)-f(y)|$ which implies $e^{-\alpha f(x)}-e^{-\alpha f(y)}\ge-\alpha e^{-\alpha f_m}|f(x)-f(y)|$. Using Assumption 4.1, we get

$$\begin{aligned}d\mathbb{E}e^{-\alpha f(X(t))}&\ge-\alpha e^{-\alpha f_m}\left(\beta(t)K_1+\sigma^2(t)K_2+\lambda\gamma^2(t)K_3\mathbb{E}|Z|^2\right)\\&\quad\times\mathbb{E}|X(t)-\bar{X}(t)|^2dt.\end{aligned}$$

From (4.3) and (4.4), we have

$$\mathbb{E}|X(t)-\bar{X}(t)|^2\le\mathrm{Var}(t)+\frac{e^{-\alpha f_m}}{M(t)}\mathrm{Var}(t)\le 2\frac{e^{-\alpha f_m}}{M(t)}\mathrm{Var}(t).$$

This implies

$$d\,M(t)\ge-2\alpha e^{-\alpha f_m}\left(\beta(t)K_1+\sigma^2(t)K_2+\lambda\gamma^2(t)K_3\mathbb{E}|Z|^2\right)\frac{e^{-\alpha f_m}}{M(t)}\mathrm{Var}(t)dt,$$

which is what we aimed to prove in this lemma. □

Our next objective is to show that $\mathbb{E}(X(t))$ converges to x^* as $t\to\infty$, where x^* is close to $x_{\min}$, i.e. the point at which $f(x)$ attains its minimum value, f_m. Applying Laplace's method (see e.g. Chap. 3 in Ref. 25 and also Refs. 11 and 46), we can obtain the following asymptotics: for any compactly supported probability measure $\rho\in\mathcal{P}(\mathbb{R}^d)$ with $x_{\min}\in\mathrm{supp}(\rho)$, we have

$$\lim_{\alpha\to\infty}\left(-\frac{1}{\alpha}\log\left(\int_{\mathbb{R}^d}e^{-\alpha f(x)}d\rho(x)\right)\right)=f_m>0. \tag{4.6}$$

Based on the above asymptotics, we aim to prove that

$$f(x^*) \leq f_m + \Gamma(\alpha) + \mathcal{O}\left(\frac{1}{\alpha}\right),$$

where the function $\Gamma(\alpha) \to 0$ as $\alpha \to \infty$.

We introduce the following function:

$$\chi(t) = 2\beta(t) - \left(2\sigma^2(t) + \lambda\gamma^2(t)\mathbb{E}|\mathcal{Z}|^2\right)\left(1 + \frac{2e^{-\alpha f_m}}{M(0)}\right).$$

We choose α, $\beta(t)$, $\sigma(t)$, $\gamma(t)$, λ, distribution of $\mathcal{Z}$ such that

(i) $\chi(t)$ is a continuous function of time t,
(ii) $\chi(t) > 0$ for all $t \geq 0$ and
(iii) $\chi(t)$ attains its minimum which we denote as $\chi_{\min}$.

We also introduce

$$\eta := 4\alpha e^{-\alpha f_m} \operatorname{Var}(0) \frac{K_1\beta + K_2\sigma^2(0) + K_3\lambda\gamma^2(0)\mathbb{E}|\mathcal{Z}|^2}{M^2(0)\chi_{\min}},$$

where β is introduced in Sec. 2.2 and K_1, K_2 and K_3 are from Assumption 4.1.

The next theorem is the main result of this subsection. We will be assuming that $\eta \leq 3/4$ which can always be achieved by choosing sufficiently small $\operatorname{Var}(0)$.

Theorem 4.1. *Let Assumptions* 3.1, 3.3–3.5 *and* 4.1 *hold. Let us also assume that* $\mathcal{L}_{X(0)}$ *is compactly supported and* $x_{\min} \in \operatorname{supp}(\mathcal{L}_{X(0)})$. *If* $\eta \leq 3/4$, *then* $\operatorname{Var}(t)$ *exponentially decays to zero as* $t \to \infty$. *Further, there exists an* $x^* \in \mathbb{R}^d$ *such that* $X(t) \to x^*$ *a.s.*, $\mathbb{E}(X(t)) \to x^*$, $\bar{X}(t) \to x^*$ *as* $t \to \infty$ *and the following inequality holds*:

$$f(x^*) \leq f_m + \Gamma(\alpha) + \frac{3\log 2}{2\alpha},$$

where function $\Gamma(\alpha) \to 0$ *as* $\alpha \to \infty$.

Proof. Let $T^* = \sup\{t : M(s) > \frac{M(0)}{2}$ for all $s \in [0,t]\}$. Observe that $T^* > 0$ by definition.

Let us assume that $T^* < \infty$. We can deduce that the following holds by definition of T^* for all $t \in [0, T^*]$:

$$\begin{aligned}&2\beta(t) - \left(2\sigma^2(t) + \lambda\gamma^2(t)\mathbb{E}|\mathcal{Z}|^2\right)\left(1 + \frac{e^{-\alpha f_m}}{M(t)}\right)\\ &\quad\geq 2\beta(t) - \left(2\sigma^2(t) + \lambda\gamma^2(t)\mathbb{E}|\mathcal{Z}|^2\right)\left(1 + \frac{2e^{-\alpha f_m}}{M(0)}\right) = \chi(t),\end{aligned}$$

where the left-hand side of the above inequality is from (4.1). Using Lemma 4.1, the fact that $\chi(t)$ is continuous and $\chi(t) > 0$ for all $t \geq 0$, we get for all $t \in [0, T^*]$:

$$\operatorname{Var}(t) \leq \operatorname{Var}(0)e^{-\chi(t)t} \leq \operatorname{Var}(0)e^{-\chi_{\min}t}.$$

We have from Lemma 4.2 for all $t \in (0, T^*]$:

$$\begin{aligned} M^2(t) &\geq M^2(0) - 4\alpha e^{-\alpha f_m} \int_0^t \left(K_1\beta(s) + K_2\sigma^2(s) + K_3\lambda\gamma^2(s)\mathbb{E}|\mathcal{Z}|^2\right) \mathrm{Var}(s)ds \\ &\geq M^2(0) - 4\alpha e^{-\alpha f_m}\left(K_1\beta + K_2\sigma^2(0) + K_3\lambda\gamma^2(0)\mathbb{E}|\mathcal{Z}|^2\right)\frac{\mathrm{Var}(0)}{\chi_{\min}}\left(1 - e^{-\chi_{\min}t}\right) \\ &> M^2(0) - 4\alpha e^{-\alpha f_m}\left(K_1\beta + K_2\sigma^2(0) + K_3\lambda\gamma^2(0)\mathbb{E}|\mathcal{Z}|^2\right)\frac{\mathrm{Var}(0)}{\chi_{\min}} \geq \frac{M^2(0)}{4}, \end{aligned}$$

where in the last step we have used the fact that $\eta \leq 3/4$. This shows $M(t) > M(0)/2$ which implies $M(t) - M(0)/2 > 0$ on the set $(0, T^*]$. Also, note that $M(t)$ is continuous in t, therefore there exists an $\epsilon > 0$ such that $M(t) > M(0)/2$ for all $t \in [T^*, T^* + \epsilon)$. This creates a contradiction which implies $T^* = \infty$. Hence,

$$\mathrm{Var}(t) \leq \mathrm{Var}(0)e^{-\chi_{\min}t} \quad \text{and} \quad M(t) > M(0)/2 \quad \text{for all } t > 0. \tag{4.7}$$

Therefore, $\mathrm{Var}(t)$ exponentially decays to zero as $t \to \infty$. From (4.4) and (4.7), we get

$$|\mathbb{E}X(t) - \bar{X}(t)|^2 \leq e^{-\alpha f_m}\frac{\mathrm{Var}(t)}{M(t)} \leq Ce^{-\chi_{\min}t}, \quad t > 0, \tag{4.8}$$

where C is a positive constant independent of t.

Taking expectation on both sides of (2.14) (recall that $\mathbb{E}\mathcal{Z} = 0$), applying Hölder's inequality and using (4.3) gives

$$\begin{aligned} \left|\frac{d}{dt}\mathbb{E}X(t)\right| &\leq \beta\mathbb{E}|X(t) - \bar{X}(t)| \leq \beta(\mathbb{E}|X(t) - \bar{X}(t)|^2)^{1/2} \\ &\leq \beta\left(\mathrm{Var}(t) + |\mathbb{E}X(t) - \bar{X}(t)|^2\right)^{1/2} \\ &\leq Ce^{-\chi_{\min}t/2}, \quad t > 0, \end{aligned} \tag{4.9}$$

where C is a positive constant independent of t.

It is clear from (4.9) that there exists an $x^* \in \mathbb{R}^d$ such that $\mathbb{E}(X(t)) \to x^*$ as $t \to \infty$. Further, $\bar{X}(t) \to x^*$ as $t \to \infty$ due to (4.8).

Let $\ell > 0$. Using Chebyshev's inequality, we have

$$\mathbb{P}(|X(t) - \mathbb{E}X(t)| \geq e^{-\ell t}) \leq \frac{\mathrm{Var}\,(t)}{e^{-2\ell t}} \leq Ce^{-(\chi_{\min} - 2\ell)t},$$

where $C > 0$ is independent of t. If we choose $\ell < \chi_{\min}/2$, then we can say $|X(t) - \mathbb{E}X(t)| \to 0$ as $t \to 0$ a.s. due to the Borel–Cantelli lemma. This implies $X(t) \to x^*$ a.s. Consequently, application of the bounded convergence theorem gives the convergence result: $\mathbb{E}e^{-\alpha f(X(t))} \to e^{-\alpha f(x^*)}$ as $t \to \infty$. Hence, using (4.7), we obtain for sufficiently large t

$$e^{-2\alpha f(x^*)} \geq M^2(t) - M^2(0)/8 > M^2(0)/8$$

and hence

$$f(x^*) \leq -\frac{1}{\alpha}\log(M(0)) + \frac{3}{2\alpha}\log 2.$$

Then, using the asymptotics (4.6), we get

$$f(x^*) \leq f_m + \Gamma(\alpha) + \frac{3}{2\alpha}\log 2, \tag{4.10}$$

where the function $\Gamma(\alpha) \to 0$ as $\alpha \to \infty$. □

As mentioned before, the approach that we have followed in this subsection is analogous to the one of Refs. 11 and 13. Alternatively, the improved approach of Ref. 24 can be adopted instead.

4.2. *Convergence to the mean-field SDEs*

In the previous subsection, we showed convergence of the non-linear process $X(t)$ from (2.14) towards the global minimizer. However, the CBO method is based on the system (2.7) of finite particles. This means there is a missing link in the theoretical analysis which we fill in this subsection by showing convergence of the particle system (2.7) to the mean-field limit in the mean-square sense (2.14) as the number of particles tends to infinity. The proof of this result has some ingredients inspired from Ref. 42 (see also Ref. 43), precisely where we partition the sample space (cf. Theorem 4.2). Further, we need here the stronger moment bound results of Lemmas 3.2 and 3.3 compared to Lemma 3.4 in Ref. 11.

We first discuss some concepts necessary for later use in this subsection. We introduce the following notation for the empirical measure of i.i.d. particles driven by the McKean–Vlasov SDEs (2.14):

$$\mathcal{E}_t := \frac{1}{N}\sum_{i=1}^{N}\delta_{X^i(t)}, \tag{4.11}$$

where δ_x is the Dirac measure at $x \in \mathbb{R}^d$. We will also need the following notation:

$$\bar{X}^{\mathcal{E}_t}(t) = \frac{\int_{\mathbb{R}^d} x e^{-\alpha f(x)}\mathcal{E}_t(dx)}{\int_{\mathbb{R}^d} e^{-\alpha f(x)}\mathcal{E}_t(dx)} = \frac{\sum_{i=1}^{N} X^i(t)e^{-\alpha f(X^i(t))}}{\sum_{i=1}^{N} e^{-\alpha f(X^i(t))}}. \tag{4.12}$$

Using discrete Jensen's inequality, we have

$$\exp\left(-\alpha\frac{1}{N}\sum_{i=1}^{N} f(X^i(t))\right) \leq \frac{1}{N}\sum_{i=1}^{N}\exp\left(-\alpha f(X^i(t))\right),$$

which, on rearrangement and multiplying both sides by $e^{-\alpha f_m}$, gives

$$\frac{e^{-\alpha f_m}}{\frac{1}{N}\sum_{i=1}^{N} e^{-\alpha f(X^i(t))}} \leq e^{\alpha K_u}\exp\left(\frac{\alpha K_u}{N}\sum_{i=1}^{N}|X^i(t)|^2\right), \tag{4.13}$$

where we have used Assumption 3.4.

Let $R > 0$ be a sufficiently large real number. Let us fix a $t \in [0, T]$. We introduce the stopping times

$$\begin{aligned} \tau_{1,R} &= \inf\left\{s \geq 0 : \frac{1}{N}\sum_{i=1}^{N} |X_N^i(s)|^4 \geq R\right\}, \\ \tau_{2,R} &= \inf\left\{s \geq 0 : \frac{1}{N}\sum_{i=1}^{N} |X^i(s)|^4 \geq R\right\}, \\ \tau_R &= \tau_{1,R} \wedge \tau_{2,R} \end{aligned} \tag{4.14}$$

and the events

$$\Omega_1(t) = \{\tau_{1,R} \leq t\} \cup \{\tau_{2,R} \leq t\}, \tag{4.15}$$

$$\Omega_2(t) = \Omega \backslash \Omega_1(t) = \{\tau_{1,R} > t\} \cap \{\tau_{2,R} > t\}. \tag{4.16}$$

Lemma 4.3. *Let Assumptions* 3.1 *and* 3.3–3.5 *be satisfied. Then, the following inequality holds for all* $t \in [0, T]$:

$$\begin{aligned} &\mathbb{E}\int_0^{t\wedge\tau_R} |\bar{X}_N(s) - \bar{X}^{\mathcal{E}_s}(s)|^2 ds \\ &\quad \leq CRe^{4\alpha K_u \sqrt{R}} \int_0^t \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}|X_N^i(s\wedge\tau_R) - X^i(s\wedge\tau_R)|^2 ds, \end{aligned} \tag{4.17}$$

where τ_R *is from* (4.14), $\bar{X}_N(s)$ *is from* (2.10), $\bar{X}^{\mathcal{E}_s}(s)$ *is from* (4.12) *and* $C > 0$ *is a constant independent of* N *and* R.

Proof. We have

$$\begin{aligned} &|\bar{X}_N(s) - \bar{X}^{\mathcal{E}_s}(s)| \\ &\quad \leq \left|\frac{1}{N}\sum_{i=1}^{N}\big(X_N^i(s) - X^i(s)\big)\frac{e^{-\alpha f(X_N^i(s))}}{\frac{1}{N}\sum_{j=1}^{N} e^{-\alpha f(X_N^j(s))}}\right| \\ &\qquad + \left|\frac{\frac{1}{N}\sum_{i=1}^{N} X^i(s)\big(e^{-\alpha f(X_N^i(s))} - e^{-\alpha f(X^i(s))}\big)}{\frac{1}{N}\sum_{j=1}^{N} e^{-\alpha f(X_N^j(s))}}\right| \\ &\qquad + \left|\frac{1}{N}\sum_{i=1}^{N} X^i(s)e^{-\alpha f(X^i(s))}\left(\frac{1}{\frac{1}{N}\sum_{j=1}^{N} e^{-\alpha f(X_N^j(s))}} - \frac{1}{\frac{1}{N}\sum_{j=1}^{N} e^{-\alpha f(X^j(s))}}\right)\right|. \end{aligned}$$

Using the discrete Jensen inequality, we get

$$\begin{aligned} |\bar{X}_N(s) - \bar{X}^{\mathcal{E}_s}(s)| \leq C\Bigg(& e^{\frac{\alpha}{N}\sum_{j=1}^{N} f(X_N^j(s))}\frac{1}{N}\sum_{i=1}^{N}|X_N^i(s) - X^i(s)| \\ &+ e^{\frac{\alpha}{N}\sum_{j=1}^{N} f(X_N^j(s))}\frac{1}{N}\sum_{i=1}^{N}|X^i(s)||e^{-\alpha f(X_N^i(s))} - e^{-\alpha f(X^i(s))}| \end{aligned}$$

$$+ e^{\frac{\alpha}{N}\sum_{j=1}^{N}(f(X_N^j(s))+f(X^j(s)))}\frac{1}{N}\sum_{i=1}^{N}|X^i(s)|\frac{1}{N}\sum_{j=1}^{N}|e^{-\alpha f(X_N^j(s))} - e^{-\alpha f(X^j(s))}|\Bigg), \tag{4.18}$$

where C is a positive constant independent of N. Applying Assumptions 3.3 and 3.4, the Cauchy–Bunyakovsky–Schwartz inequality and Young's inequality, $ab \leq a^2/2 + b^2/2$, $a, b > 0$, we obtain

$$\begin{aligned}
&|\bar{X}_N(s) - \bar{X}^{\mathcal{E}_s}(s)| \\
&\quad \leq C\Bigg(e^{\frac{\alpha K_u}{N}\sum_{j=1}^{N}|X_N^j(s)|^2}\frac{1}{N}\sum_{i=1}^{N}|X_N^i(s) - X^i(s)| + e^{\frac{\alpha K_u}{N}\sum_{j=1}^{N}|X_N^j(s)|^2} \\
&\qquad \times \frac{1}{N}\sum_{i=1}^{N}|X^i(s)|\big(1 + |X_N^i(s)| + |X^i(s)|\big)|X_N^i(s) - X^i(s)| \\
&\qquad + e^{\frac{\alpha K_u}{N}\sum_{j=1}^{N}(|X_N^j(s)|^2+|X^j(s)|^2)}\frac{1}{N}\sum_{i=1}^{N}|X^i(s)| \\
&\qquad \times \frac{1}{N}\sum_{j=1}^{N}\big(1 + |X_N^j(s)| + |X^j(s)|\big)|X_N^j(s) - X^j(s)|\Bigg) \\
&\quad \leq C\Bigg(e^{\frac{\alpha K_u}{N}\sum_{j=1}^{N}|X_N^j(s)|^2}\frac{1}{N}\sum_{i=1}^{N}|X_N^i(s) - X^i(s)| + e^{\frac{\alpha K_u}{N}\sum_{j=1}^{N}(|X_N^j(s)|^2+|X^j(s)|^2)} \\
&\qquad \times \frac{1}{N}\sum_{i=1}^{N}\big(1 + |X_N^i(s)|^2 + |X^i(s)|^2\big)|X_N^i(s) - X^i(s)| \\
&\qquad + e^{\frac{\alpha K_u}{N}\sum_{j=1}^{N}(|X_N^j(s)|^2+|X^j(s)|^2)}\frac{1}{N}\sum_{i=1}^{N}|X^i(s)|^2\frac{1}{N}\sum_{j=1}^{N}|X_N^j(s) - X^j(s)|\Bigg) \\
&\quad \leq C\Bigg(e^{\frac{\alpha K_u}{N}\sum_{j=1}^{N}|X_N^j(s)|^2}\frac{1}{N}\sum_{i=1}^{N}|X_N^i(s) - X^i(s)| + e^{\frac{\alpha K_u}{N}\sum_{j=1}^{N}(|X_N^j(s)|^2+|X^j(s)|^2)} \\
&\qquad \times \left(\frac{1}{N}\sum_{i=1}^{N}\big(1 + |X_N^i(s)|^2 + |X^i(s)|^2\big)^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}|X_N^i(s) - X^i(s)|^2\right)^{1/2}\Bigg).
\end{aligned} \tag{4.19}$$

On squaring both sides, we ascertain

$$\begin{aligned}
&|\bar{X}_N(s) - \bar{X}^{\mathcal{E}_s}(s)|^2 \\
&\quad \le C\Bigg(e^{\frac{2\alpha K_u}{N}\sum_{j=1}^N |X_N^j(s)|^2}\frac{1}{N}\sum_{i=1}^N |X_N^i(s) - X^i(s)|^2 \\
&\qquad + e^{\frac{2\alpha K_u}{N}\sum_{j=1}^N (|X_N^j(s)|^2 + |X^j(s)|^2)}\left(\frac{1}{N}\sum_{i=1}^N \left(1 + |X_N^i(s)|^2 + |X^i(s)|^2\right)^2\right) \\
&\qquad \times \left(\frac{1}{N}\sum_{i=1}^N |X_N^i(s) - X^i(s)|^2\right)\Bigg). \qquad (4.20)
\end{aligned}$$

Using Hölder's inequality, we have

$$\frac{1}{N}\sum_{j=1}^N (|X_N^j(s)|^2 + |X^j(s)|^2) \le \frac{\sqrt{2}}{N^{1/2}}\left(\sum_{j=1}^N (|X_N^j(s)|^4 + |X^j(s)|^4)\right)^{1/2}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\int_0^{t\wedge\tau_R} |\bar{X}_N(s) - \bar{X}^{\mathcal{E}_s}(s)|^2 ds \\
&\qquad \le CRe^{4\alpha K_u\sqrt{R}}\int_0^t \frac{1}{N}\sum_{i=1}^N \mathbb{E}|X_N^i(s\wedge\tau_R) - X^i(s\wedge\tau_R)|^2 ds,
\end{aligned}$$

where $C > 0$ is independent of N and R. □

Lemma 4.4. *Let Assumptions* 3.1 *and* 3.3–3.5 *be satisfied. Then, the following inequality holds for all* $t \in [0, T]$:

$$\mathbb{E}\int_0^{t\wedge\tau_R} |\bar{X}^{\mathcal{E}_s}(s) - \bar{X}(s)|^2 ds \le C\frac{e^{2\alpha K_u\sqrt{R}}}{N}, \qquad (4.21)$$

where τ_R *is from* (4.14), $\bar{X}^{\mathcal{E}_s}(s)$ *is from* (4.12), $\bar{X}(s)$ *is from* (2.13), *and* $C > 0$ *is independent of* N *and* R.

Proof. We have

$$\begin{aligned}
&|\bar{X}^{\mathcal{E}_s}(s) - \bar{X}(s)| \\
&\quad \le \frac{1}{\frac{1}{N}\sum_{j=1}^N e^{-\alpha f(X^j(s))}}\left|\frac{1}{N}\sum_{i=1}^N \left(X^i(s)e^{-\alpha f(X^i(s))} - \int_{\mathbb{R}^d} xe^{-\alpha f(x)}\mathcal{L}_{X(s)}(dx)\right)\right| \\
&\qquad + \left|\int_{\mathbb{R}^d} xe^{-\alpha f(x)}\mathcal{L}_{X(s)}(dx)\frac{\frac{1}{N}\sum_{j=1}^N \left(e^{-\alpha f(X^j(s))} - \int_{\mathbb{R}^d} e^{-\alpha f(x)}\mathcal{L}_{X(s)}(dx)\right)}{\frac{1}{N}\sum_{j=1}^N e^{-\alpha f(X^j(s))}\int_{\mathbb{R}^d} e^{-\alpha f(x)}\mathcal{L}_{X(s)}(dx)}\right|.
\end{aligned}$$

Using Jensen's inequality and squaring both sides, we get

$$
\begin{aligned}
&|\bar{X}^{\mathcal{E}_s}(s) - \bar{X}(s)|^2 \\
&\quad \le C\Bigg(\Bigg|\frac{1}{N}\sum_{i=1}^{N}(X^i(s)e^{-\alpha f(X^i(s))} - \mathbb{E}\big(X(s)e^{-\alpha f(X(s))}\big))\Bigg|^2 \\
&\qquad + e^{2\alpha\mathbb{E}f(X(s))}(\mathbb{E}|X(s)|)^2\Bigg|\frac{1}{N}\sum_{j=1}^{N}\Big(e^{-\alpha f(X^j(s))} - \mathbb{E}\big(e^{-\alpha f(X(s))}\big)\Big)\Bigg|^2\Bigg) \\
&\qquad \times e^{\frac{2\alpha}{N}\sum_{j=1}^{N} f(X^j(s))},
\end{aligned}
$$

where C is a positive constant independent of N. Applying Assumption 3.4, we ascertain

$$
\begin{aligned}
&|\bar{X}^{\mathcal{E}_s}(s) - \bar{X}(s)|^2 \\
&\quad \le C\Bigg(\Bigg|\frac{1}{N}\sum_{i=1}^{N}\Big(X^i(s)e^{-\alpha f(X^i(s))} - \mathbb{E}\big(X(s)e^{-\alpha f(X(s))}\big)\Big)\Bigg|^2 \\
&\qquad + e^{2\alpha K_u\mathbb{E}|X(s)|^2}(\mathbb{E}|X(s)|)^2\Bigg|\frac{1}{N}\sum_{j=1}^{N}\Big(e^{-\alpha f(X^j(s))} - \mathbb{E}\big(e^{-\alpha f(X(s))}\big)\Big)\Bigg|^2\Bigg) \\
&\qquad \times e^{\frac{2\alpha K_u}{N}\sum_{j=1}^{N} |X^j(s)|^2}.
\end{aligned}
$$

Hence, using Theorem 3.3, we obtain

$$
\begin{aligned}
&\mathbb{E}\int_0^{t\wedge\tau_R} |\bar{X}^{\mathcal{E}_s}(s) - \bar{X}(s)|^2 ds \\
&\quad \le Ce^{2\alpha K_u\sqrt{R}}\mathbb{E}\int_0^{t\wedge\tau_R}\Bigg|\frac{1}{N}\sum_{i=1}^{N}\Big(X^i(s)e^{-\alpha f(X^i(s))} - \mathbb{E}\big(X(s)e^{-\alpha f(X(s))}\big)\Big)\Bigg|^2 ds \\
&\qquad + Ce^{2\alpha K_u\sqrt{R}}\mathbb{E}\int_0^{t\wedge\tau_R}\Bigg|\frac{1}{N}\sum_{j=1}^{N}\Big(e^{-\alpha f(X^j(s))} - \mathbb{E}\big(e^{-\alpha f(X(s))}\big)\Big)\Bigg|^2 ds \\
&\quad \le Ce^{2\alpha K_u\sqrt{R}}\int_0^t \mathbb{E}\Bigg|\frac{1}{N}\sum_{i=1}^{N} U_1^i(s\wedge\tau_R)\Bigg|^2 ds \\
&\qquad + Ce^{2\alpha K_u\sqrt{R}}\int_0^t \mathbb{E}\Bigg|\frac{1}{N}\sum_{i=1}^{N} U_2^i(s\wedge\tau_R)\Bigg|^2 ds,
\end{aligned}
$$

where $U_1^i(s\wedge\tau_R) = X^i(s\wedge\tau_R)e^{-\alpha f(X^i(s\wedge\tau_R))} - \mathbb{E}\big(X(s\wedge\tau_R)e^{-\alpha f(X(s\wedge\tau_R))}\big)$, $U_2^i(s\wedge\tau_R) = e^{-\alpha f(X^i(s))} - \mathbb{E}\big(e^{-\alpha f(X(s))}\big)$ and C is independent of N and R.

We have

$$\begin{aligned}\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N}U_1^i(s\wedge\tau_R)\right|^2 &= \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}|U_1^i(s\wedge\tau_R)|^2\\ &\quad+\frac{1}{N^2}\sum_{i,j=1,i\neq j}^{N}\mathbb{E}\big(U_1^i(s\wedge\tau_R)\cdot U_1^j(s\wedge\tau_R)\big).\end{aligned}$$

Note that $\mathbb{E}\big(U_1^i(s)\cdot U_1^j(s)\big)=0$ for $i\neq j$ and $s\wedge\tau_R$ is a bounded stopping time then $\mathbb{E}\big(U_1^i(s\wedge\tau_R)\cdot U_1^j(s\wedge\tau_R)\big)=0$ for $i\neq j$ because of Doob's optional stopping theorem (see Theorem 2.2.1 in Ref. 3). Using Theorem 3.3, we deduce

$$\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N}U_1^i(s\wedge\tau_R)\right|^2\leq\frac{C}{N},\tag{4.22}$$

where C is independent of N. In the similar manner, we can obtain

$$\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N}U_2^i(s\wedge\tau_R)\right|^2\leq\frac{C}{N},\tag{4.23}$$

where C is independent of N. Using (4.22) and (4.23), we get the following estimate:

$$\mathbb{E}\int_0^{t\wedge\tau_R}|\bar{X}^{\mathcal{E}_s}(s)-\bar{X}(s)|^2ds\leq C\frac{e^{2\alpha K_u\sqrt{R}}}{N},$$

where C is independent of N and R. □

Theorem 4.2. *Let Assumptions* 3.1 *and* 3.3–3.5 *be satisfied. Let* X_N^i, $i=1,\ldots,N$, *solve* (2.11) *and* X^i, $i=1,\ldots,N$, *represent independent processes each solving* (2.14). *Assume that* X_N^i *and* X^i *are driven by the same Wiener processes and compound Poisson processes. Also assume that* $X_N^i(0)=X^i(0)$, *a.s.* $i=1,\ldots,N$, $\mathbb{E}|Z|^4\leq C$ *and* $\sup_{i=1,\ldots,N}\mathbb{E}|X^i(0)|^4=\sup_{i=1,\ldots,N}\mathbb{E}|X_N^i(0)|^4\leq C$. *Then,* X_N^i *converges to* X^i *in the mean-square sense when* $N\to\infty$, *i.e. for all* $t\in[0,T]$:

$$\lim_{N\to\infty}\sup_{i=1,\ldots,N}\mathbb{E}|X_N^i(t)-X^i(t)|^2=0.\tag{4.24}$$

Proof. Let $t\in(0,T]$. We can write

$$\begin{aligned}\mathbb{E}|X_N^i(t)-X^i(t)|^2 &= \mathbb{E}\big(|X_N^i(t)-X^i(t)|^2I_{\Omega_1(t)}\big)+\mathbb{E}\big(|X_N^i(t)-X^i(t)|^2I_{\Omega_2(t)}\big)\\ &=: E_1(t)+E_2(t),\end{aligned}$$

where $\Omega_1(t)$ and $\Omega_2(t)$ are from (4.15) and (4.16), respectively. Using the Cauchy–Bunyakovsky–Schwartz inequality and Markov's inequality, we obtain

$$\begin{aligned}E_1(t) &:= \mathbb{E}\big(|X_N^i(t)-X^i(t)|^2I_{\Omega_1(t)}\big)\\ &\leq \big(\mathbb{E}|X_N^i(t)-X^i(t)|^4\big)^{1/2}\big(\mathbb{E}I_{\Omega_1(t)}\big)^{1/2}\end{aligned}$$

$$\leq C\big(\mathbb{E}|X_N^i(t)|^4 + \mathbb{E}|X^i(t)|^4\big)^{1/2}\Bigg(\frac{1}{RN}\sum_{i=1}^{N}\mathbb{E}\sup_{0\leq s\leq t}|X_N^i(s)|^4 + \frac{1}{RN}\sum_{i=1}^{N}\mathbb{E}\sup_{0\leq s\leq t}|X^i(s)|^4\Bigg)^{1/2}.$$

We get the following estimate for $E_1(t)$ by applying Lemma 3.2 and Theorem 3.3:

$$E_1(t) \leq \frac{C}{\sqrt{R}}, \tag{4.25}$$

where C is a positive constant independent of N and R.

Now, we estimate $E_2(t)$. We have $\mathbb{E}(|X_N^i(t) - X^i(t)|^2 I_{\Omega_2(t)}) \leq \mathbb{E}(|X_N^i(t\wedge\tau_R) - X^i(t\wedge\tau_R)|^2)$. Using Ito's formula, we have

$$\begin{aligned}
&|X_N^i(t\wedge\tau_R) - X^i(t\wedge\tau_R)|^2\\
&\quad= |X_N^i(0) - X^i(0)|^2 - 2\mathbb{E}\int_0^{t\wedge\tau_R}\beta(s)(X_N^i(s) - X^i(s))\\
&\qquad\cdot(X_N^i(s) - \bar{X}_N(s) - X^i(s) + \bar{X}(s))ds\\
&\qquad + 2\int_0^{t\wedge\tau_R}\sigma^2(s)|\mathrm{Diag}(X_N^i(s) - \bar{X}_N(s) - X^i(s) + \bar{X}(s))|^2 ds\\
&\qquad + 2\sqrt{2}\int_0^{t\wedge\tau_R}\sigma(s)(X_N^i(s) - X^i(s))\cdot\mathrm{Diag}(X_N^i(s) - \bar{X}_N(s)\\
&\qquad - X^i(s) + \bar{X}(s))dW^i(s) + \int_0^{t\wedge\tau_R}\int_{\mathbb{R}^d}\Big(|X_N^i(s^-) - X^i(s^-)\\
&\qquad + \gamma(s)\,\mathrm{Diag}(X_N^i(s^-) - \bar{X}_N(s^-))z - \gamma(s)\,\mathrm{Diag}(X^i(s^-) - \bar{X}(s^-))z|^2\\
&\qquad - |X_N^i(s^-) - X^i(s^-)|^2\Big)\mathcal{N}^i(ds,dz).
\end{aligned} \tag{4.26}$$

Taking expectations on both sides of (4.26), using the Cauchy–Bunyakovsky–Schwartz inequality and Young's inequality, and applying Doob's optional stopping theorem (see Theorem 2.2.1 in Ref. 3), we get

$$\begin{aligned}
&\mathbb{E}|X_N^i(t\wedge\tau_R) - X^i(t\wedge\tau_R)|^2\\
&\quad\leq \mathbb{E}|X_N^i(0) - X^i(0)|^2 + C\mathbb{E}\int_0^{t\wedge\tau_R}\big(|X_N^i(s) - X^i(s)|^2 + |\bar{X}_N(s) - \bar{X}(s)|^2\big)ds\\
&\qquad + C\mathbb{E}\int_0^{t\wedge\tau_R}\int_{\mathbb{R}^d}(|X_N^i(s) - X^i(s)|^2 + |\bar{X}_N(s) - \bar{X}(s)|^2)|z|^2\rho_z(z)dzds\\
&\quad\leq \mathbb{E}|X_N^i(0) - X^i(0)|^2 + C\mathbb{E}\int_0^{t\wedge\tau_R}|X_N^i(s) - X^i(s)|^2 ds\\
&\qquad + C\mathbb{E}\int_0^{t\wedge\tau_R}|\bar{X}_N(s) - \bar{X}^{\mathcal{E}_s}(s)|^2 ds + C\mathbb{E}\int_0^{t\wedge\tau_R}|\bar{X}^{\mathcal{E}_s}(s) - \bar{X}(s)|^2 ds.
\end{aligned} \tag{4.27}$$

Substituting (4.17) and (4.21) in (4.27), we obtain

$$\begin{aligned}\mathbb{E}\big(|X_N^i(t\wedge\tau_R)-X^i(t\wedge\tau_R)|^2\big)&\\ \leq \mathbb{E}|X_N^i(0)-X^i(0)|^2+CRe^{4\alpha K_u\sqrt{R}}\int_0^t\frac{1}{N}\sum_{i=1}^N\mathbb{E}\big(|X_N^i(s\wedge\tau_R)&\\ -X^i(s\wedge\tau_R)|^2\big)ds+C\frac{e^{2\alpha K_u\sqrt{R}}}{N},&\end{aligned}$$

where $C>0$ is independent of N and R. Taking supremum over $i=1,\dots,N$, we get

$$\begin{aligned}\sup_{i=1,\dots,N}\mathbb{E}\big(|X_N^i(t\wedge\tau_R)-X^i(t\wedge\tau_R)|^2\big)&\\ \leq \sup_{i=1,\dots,N}\mathbb{E}|X_N^i(0)-X^i(0)|^2+CRe^{4\alpha K_u\sqrt{R}}\int_0^t\sup_{i=1,\dots,N}\mathbb{E}\big(|X_N^i(s\wedge\tau_R)&\\ -X^i(s\wedge\tau_R)|^2\big)ds+C\frac{e^{2\alpha K_u\sqrt{R}}}{N}.&\end{aligned}$$

Using Grönwall's inequality, we have

$$\begin{aligned}\sup_{i=1,\dots,N}\mathbb{E}\big(|X_N^i(t\wedge\tau_R)-X^i(t\wedge\tau_R)|^2\big)&\\ \leq\frac{C}{N}e^{CRe^{4\alpha K_u\sqrt{R}}}e^{2\alpha K_uR}\leq\frac{C}{N}e^{e^{C_u\sqrt{R}}},&\end{aligned}\tag{4.28}$$

where $C>0$ and $C_u>0$ are constants independent of N and R. In the above calculations, we have used the facts that $R<e^{2\alpha K_u\sqrt{R}}$ and $2\alpha K_u\sqrt{R}<e^{2\alpha K_u\sqrt{R}}$ for sufficiently large R.

We choose $R=\frac{1}{C_u^2}(\ln(\ln(N^{1/2})))^2$. Therefore,

$$\begin{aligned}\sup_{i=1,\dots,N}\mathbb{E}(|X_N^i(t)-X^i(t)|^2I_{\Omega_2(t)})&\\ \leq\sup_{i=1,\dots,N}\mathbb{E}\big(|X_N^i(t\wedge\tau_R)-X^i(t\wedge\tau_R)|^2\big)\leq\frac{C}{N^{1/2}},&\end{aligned}$$

which implies

$$\begin{aligned}\lim_{N\to\infty}\sup_{i=1,\dots,N}\mathbb{E}(|X_N^i(t)-X^i(t)|^2I_{\Omega_2(t)})&\\ =\lim_{N\to\infty}\sup_{i=1,\dots,N}\mathbb{E}\big(|X_N^i(t\wedge\tau_R)-X^i(t\wedge\tau_R)|^2\big)=0.&\end{aligned}\tag{4.29}$$

The term (4.25) and the choice of R provide the following estimate:

$$\mathbb{E}\big(|X_N^i(t)-X^i(t)|^2I_{\Omega_1}(t)\big)\leq\frac{C}{\sqrt{R}}\leq\frac{C}{\ln(\ln(N^{1/2}))},$$

where $C>0$ is independent of N and R. This yields

$$\lim_{N\to\infty}\sup_{i=1,\dots,N}\mathbb{E}\big(|X_N^i(t)-X^i(t)|^2I_{\Omega_1(t)}\big)=0.\tag{4.30}$$

As a consequence of (4.29) and (4.30), we get

$$\lim_{N\to\infty} \sup_{i=1,\dots,N} \mathbb{E}|X_N^i(t) - X^i(t)|^2 = 0$$

for all $t \in [0,T]$. □

Remark 4.1. It is not difficult to see from the above theorem that the empirical measure of the particle system (2.11) converges to the law of the mean-field SDEs (2.14) in 2-Wasserstein metric, i.e. for all $t \in [0,T]$:

$$\lim_{N\to\infty} \mathcal{W}_2^2(\mathcal{E}_t^N, \mathcal{L}_{X(t)}) = 0, \tag{4.31}$$

where $\mathcal{E}_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_N^i(t)}$.

Remark 4.2. Theorem 4.2 implies weak convergence of the empirical measure, $\mathcal{E}_t^N$ of interacting particle system towards $\mathcal{L}_{X(t)}$ which is the law of the mean-field limit process $X(t)$ (cf. Refs. 48 and 49).

4.3. *Convergence of the numerical scheme*

To implement the particle system (2.7), we have proposed to utilize the Euler scheme introduced in Sec. 2.2.3. The jump-diffusion SDEs (2.7), governing interacting particle system, have locally Lipschitz and linearly growing coefficients. Due to non-global Lipschitzness of the coefficients, it is not straightforward to deduce convergence of the Euler scheme to (2.7). In this subsection, we prove this convergence result uniform in N.

Introduce the function $\kappa_h(t) := t_k$, $t_k \le t < t_{k+1}$, where $0 = t_0 < \cdots < t_n = T$ is a uniform partition of $[0,T]$, i.e. $t_{k+1} - t_k = h$ for all $k = 0,\dots,n-1$. We write the continuous version of the numerical scheme (2.19) as follows:

$$\begin{aligned} dY_N^i(t) &= -\beta(t)(Y_N^i(\kappa_h(t)) - \bar{Y}_N(\kappa_h(t)))dt + \sqrt{2}\sigma(t)\,\mathrm{Diag}\left(Y_N^i(\kappa_h(t))\right. \\ &\quad \left. - \bar{Y}_N(\kappa_h(t))\right)dW^i(t) + \int_{\mathbb{R}^d} \mathrm{Diag}(Y_N^i(\kappa_h(t)) - \bar{Y}_N(\kappa_h(t)))z\mathcal{N}^i(dt,dz). \end{aligned} \tag{4.32}$$

In this subsection, our aim is to show mean-square convergence of $Y_N^i(t)$ to $X_N^i(t)$ uniformly in N.

Let Assumptions 3.1 and 3.2 hold. Let $\mathbb{E}|Y_N^i(0)|^2 < \infty$ and $\mathbb{E}|Z|^2 < \infty$, then the particle system (4.32) is well-posed (cf. Theorem 3.1). Moreover, if $\mathbb{E}|Y_N^i(0)|^{2p} < \infty$ and $\mathbb{E}|Z|^{2p} < \infty$ for some $p \ge 1$, then, due to Lemma 3.2, the following holds:

$$\mathbb{E}\sup_{0\le t\le T} |Y_N^i(t)|^{2p} \le K, \tag{4.33}$$

where we cannot say that K is independent of h. However, to prove the convergence of numerical scheme, we need the uniform in h and N moment bound, which we prove in the next lemma.

Lemma 4.5. *Let Assumptions* 3.1 *and* 3.3–3.5 *hold. Let* $p \geq 1$, $\mathbb{E}|Y_N^i(0)|^{2p} < \infty$ *and* $\mathbb{E}|Z|^{2p} < \infty$. *Then*

$$\sup_{i=1,\ldots,N} \mathbb{E} \sup_{0\leq t\leq T} |Y_N^i(t)|^{2p} \leq K_d, \tag{4.34}$$

where K_d *is a positive constant independent of* h *and* N.

Proof. Let p be a positive integer. Using Ito's formula, the Cauchy–Bunyakovsky–Schwartz inequality and Young's inequality, we have

$$\begin{aligned}
|Y_N^i(t)|^{2p} &\leq |Y_N^i(0)|^{2p} + C\int_0^t (|Y_N^i(s)|^{2p} + |Y_N^i(\kappa_h(s))|^{2p} + |\bar{Y}_N(\kappa_h(s))|^{2p})ds \\
&\quad + 2\sqrt{2}p\int_0^t \sigma(s)|Y_N^i(s)|^{2p-2}(Y_N^i(s)\cdot \mathrm{Diag}(Y_N^i(\kappa_h(s)) \\
&\quad - \bar{Y}_N(\kappa_h(s)))dW^i(s)) + C\int_0^t\int_{\mathbb{R}^d}\Big(|Y_N^i(s^-)|^{2p} + (|Y_N^i(\kappa_h(s))|^{2p} \\
&\quad + |\bar{Y}_N(\kappa_h(s))|^{2p})(1+|z|^{2p})\Big)\mathcal{N}^i(ds,dz).
\end{aligned}$$

First taking supremum over $0 \leq t \leq T$ and then expectation, we obtain

$$\begin{aligned}
&\mathbb{E}\sup_{0\leq t\leq T}|Y_N^i(t)|^{2p} \\
&\quad \leq \mathbb{E}|Y_N^i(0)|^{2p} + C\mathbb{E}\int_0^T \Big(|Y_N^i(s)|^{2p} + |Y_N^i(\kappa_h(s))|^{2p} + |\bar{Y}_N(\kappa_h(s))|^{2p}\Big)ds \\
&\qquad + 2\sqrt{2}p\mathbb{E}\sup_{0\leq t\leq T}\bigg|\int_0^t \sigma(s)|Y_N^i(s)|^{2p-2}(Y_N^i(s)\cdot \mathrm{Diag}(Y_N^i(\kappa_h(s)) \\
&\qquad - \bar{Y}_N(\kappa_h(s)))dW^i(s))\bigg| + C\mathbb{E}\int_0^T\int_{\mathbb{R}^d}\Big(|Y_N^i(s^-)|^{2p} + (|Y_N^i(\kappa_h(s))|^{2p} \\
&\qquad + |\bar{Y}_N(\kappa_h(s))|^{2p})(1+|z|^{2p})\Big)\mathcal{N}^i(ds,dz),
\end{aligned}$$

where C is independent of h and N. Using the Burkholder–Davis–Gundy inequality (note that we can apply this inequality due to (4.33)) and the fact that $\mathbb{E}|Z|^{2p} < \infty$, we get

$$\begin{aligned}
&\mathbb{E}\sup_{0\leq t\leq T}|Y_N^i(t)|^{2p} \\
&\quad \leq \mathbb{E}|Y_N^i(0)|^{2p} + C\mathbb{E}\int_0^T \Big(|Y_N^i(s)|^{2p} + |Y_N^i(\kappa_h(s))|^{2p} + |\bar{Y}_N(\kappa_h(s))|^{2p}\Big)ds
\end{aligned}$$

$$+ C\mathbb{E} \sup_{0\le t\le T} |Y_N^i(t)|^{2p-1} \left(\int_0^T |Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s))|^2 ds \right)^{1/2}.$$

Applying Young's inequality and Hölder's inequality, we ascertain (note that this is the same set of arguments used to obtain (3.28))

$$\mathbb{E} \sup_{0\le t\le T} |Y_N^i(t)|^{2p} \le \mathbb{E}|Y_N^i(0)|^{2p} + C \int_0^t (|Y_N^i(s)|^{2p} + |Y_N^i(\kappa_h(s))|^{2p}$$
$$+ |\bar{Y}_N(\kappa_h(s))|^{2p}) ds + \frac{1}{2}\mathbb{E} \sup_{0\le t\le T} |Y_N^i(t)|^{2p}$$
$$+ C\mathbb{E} \int_0^T |Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s))|^{2p} ds. \tag{4.35}$$

Using Jensen's inequality and (3.5), we have

$$|\bar{Y}_N(\kappa_h(s))|^2 \le L_1 + \frac{L_2}{N} \sum_{i=1}^N |Y_N^i(\kappa_h(s))|^2. \tag{4.36}$$

Therefore, substituting (4.36) in (4.35) yields

$$\mathbb{E} \sup_{0\le t\le T} |Y_N^i(t)|^{2p} \le 2\mathbb{E}|Y_N^i(0)|^{2p} + C + C\mathbb{E} \int_0^T \Big(|Y_N^i(s)|^{2p} + |Y_N^i(\kappa_h(s))|^{2p}$$
$$+ \frac{1}{N} \sum_{i=1}^N |Y_N^i(\kappa_h(s))|^{2p} \Big) ds$$
$$\le 2\mathbb{E}|Y_N^i(0)|^{2p} + C + C \int_0^T \Bigg(\mathbb{E} \sup_{0\le u\le s} |Y_N^i(u)|^{2p}$$
$$+ \frac{1}{N} \sum_{i=1}^N \mathbb{E} \sup_{0\le u\le s} |Y_N^i(u)|^{2p} \Bigg) ds,$$

where $C > 0$ is independent of h and N. Taking supremum over $i = 1, \ldots, N$, we get

$$\sup_{i=1,\ldots,N} \mathbb{E} \sup_{0\le t\le T} |Y_N^i(t)|^{2p}$$
$$\le 2\mathbb{E}|Y_N^i(0)|^{2p} + C + C \int_0^T \sup_{i=1,\ldots,N} \mathbb{E} \sup_{0\le u\le s} |Y_N^i(u)|^{2p} ds,$$

where $C > 0$ is independent of h and N. Using Grönwall's lemma, we have the desired result. ◻

Lemma 4.6. *Let Assumptions* 3.1 *and* 3.3–3.5 *hold. Let* $\sup_{i=1,\ldots,N} \mathbb{E}|X_N^i(0)|^4 < \infty$, $\sup_{i=1,\ldots,N} \mathbb{E}|Y_N^i(0)|^4 < \infty$, $\mathbb{E}|Z|^4 < \infty$. *Then*

$$\sup_{i=1,\ldots,N} \mathbb{E}|Y_N^i(t) - Y_N^i(\kappa_h(t))|^2 \le Ch,$$

where C is a positive constant independent of N and h.

Proof. We have

$$\begin{aligned}
&|Y_N^i(t) - Y_N^i(\kappa_h(t))|^2 \\
&\quad \le C\Bigg(\Bigg|\int_{\kappa_h(t)}^t (Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s)))ds\Bigg|^2 \\
&\qquad + \Bigg|\int_{\kappa_h(t)}^t \sigma(s)\,\mathrm{Diag}(Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s)))dW^i(s)\Bigg|^2 \\
&\qquad + \Bigg|\int_{\kappa_h(t)}^t \int_{\mathbb{R}^d} \gamma(s)\,\mathrm{Diag}(Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s)))z\mathcal{N}^i(ds,dz)\Bigg|^2\Bigg),
\end{aligned}$$

where C is independent of h and N. Taking expectation and using Ito's isometry (note that we can apply Ito's isometry due to Lemma 4.5), we get

$$\mathbb{E}|Y_N^i(t) - Y_N^i(\kappa_h(t))|^2 \le C(1 + \mathbb{E}|Z|^2)\Bigg(\int_{\kappa_h(t)}^t \mathbb{E}|Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s))|^2 ds\Bigg).$$

Therefore, use of (4.36) gives

$$\begin{aligned}
&\sup_{i=1,\ldots,N} \mathbb{E}|Y_N^i(t) - Y_N^i(\kappa_h(t))|^2 \\
&\quad \le C(1 + \mathbb{E}|Z|^2)\Bigg(\int_{\kappa_h(t)}^t \sup_{i=1,\ldots,N} \mathbb{E}|Y_N^i(\kappa_h(s))|^2 \\
&\qquad + L_1 + \frac{L_2}{N}\sum_{i=1}^N \sup_{i=1,\ldots,N} \big(\mathbb{E}|Y_N^i(\kappa_h(s))|^2\big)ds\Bigg).
\end{aligned}$$

Using Lemma 4.5, we get

$$\sup_{i=1,\ldots,N} \mathbb{E}|Y_N^i(t) - Y_N^i(\kappa_h(t))|^2 \le C(t - \kappa_h(t)) \le Ch,$$

where C is independent of N and h. □

Theorem 4.3. *Let Assumptions* 3.1 *and* 3.3–3.5 *hold. Assume that* $Y_N^i(0) = X_N^i(0)$, *a.s.*, $i = 1, \ldots, N$, $\sup_{i=1,\ldots,N} \mathbb{E}|X_N^i(0)|^4 = \sup_{i=1,\ldots,N} \mathbb{E}|Y_N^i(0)|^4 < \infty$ *and* $\mathbb{E}|Z|^4 < \infty$. *Then*

$$\begin{aligned}
&\lim_{h\to 0}\lim_{N\to\infty}\sup_{i=1,\ldots,N} \mathbb{E}|Y_N^i(t) - X_N^i(t)|^2 \\
&\quad = \lim_{N\to\infty}\lim_{h\to 0}\sup_{i=1,\ldots,N} \mathbb{E}|Y_N^i(t) - X_N^i(t)|^2 = 0 \qquad (4.37)
\end{aligned}$$

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

Proof. Introduce the stopping times

$$\tau_{1,R} = \inf\left\{ s \geq 0 : \frac{1}{N}\sum_{i=1}^{N} |X_N^i(s)|^4 \geq R \right\},$$

$$\tau_{3,R}^h = \inf\left\{ s \geq 0 : \frac{1}{N}\sum_{i=1}^{N} |Y_N^i(s)|^4 \geq R \right\},$$

$$\tau_R^h = \tau_{1,R} \wedge \tau_{3,R}^h$$

and the events $\Omega_3(t) = \{\tau_{1,R} \leq t\} \cup \{\tau_{3,R}^h \leq t\}$, $\Omega_4(t) = \Omega \backslash \Omega_3(t) = \{\tau_{1,R} > t\} \cap \{\tau_{3,R}^h > t\}$.

We have

$$\begin{aligned} \mathbb{E}|Y_N^i(t) - X_N^i(t)|^2 &= \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_3(t)}\big) + \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_4(t)}\big) \\ &=: E_3(t) + E_4(t). \end{aligned}$$

Let us first estimate the term $E_3(t)$. Using the Cauchy–Bunyakovsky–Schwartz inequality, Markov's inequality, Lemmas 3.2 and 4.5, we get

$$\begin{aligned} &\mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_3(t)}\big) \\ &\quad\leq \big(\mathbb{E}|Y_N^i(t) - X_N^i(t)|^4\big)^{1/2}\big(\mathbb{E} I_{\Omega_3(t)}\big)^{1/2} \\ &\quad\leq C\left(\frac{1}{RN}\sum_{i=1}^{N}\mathbb{E}\sup_{0\leq s\leq t}|Y_N^i(s)|^4 + \frac{1}{RN}\sum_{i=1}^{N}\mathbb{E}\sup_{0\leq s\leq t}|X_N^i(s)|^4\right)^{1/2} \leq \frac{C}{\sqrt{R}}, \end{aligned} \tag{4.38}$$

where C is independent of h, N and R.

Note that $\mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_4(t)}\big) \leq \mathbb{E}|Y_N^i(t\wedge\tau_R^h) - X_N^i(t\wedge\tau_R^h)|^2$. Using Ito's formula, we obtain

$$\begin{aligned} &|Y_N^i(t\wedge\tau_R^h) - X_N^i(t\wedge\tau_R^h)|^2 \\ &\quad= |Y_N^i(0) - X_N^i(0)|^2 - 2\int_0^{t\wedge\tau_R^h} \beta(s)\big((Y_N^i(s) - X_N^i(s)) \\ &\qquad\cdot (Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s)) - X_N^i(s) + \bar{X}_N(s))\big)ds \\ &\qquad+ 2\sqrt{2}\int_0^{t\wedge\tau_R^h} \sigma(s)\big((Y_N^i(s) - X_N^i(s)) \cdot \mathrm{Diag}(Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s)) \\ &\qquad-X_N^i(s) + \bar{X}_N(s))dW^i(s)\big) + 2\int_0^{t\wedge\tau_R^h} \sigma^2(s)|Y_N^i(\kappa_h(s)) \\ &\qquad- \bar{Y}_N(\kappa_h(s)) - X_N^i(s) + \bar{X}_N(s)|^2 ds + \int_0^{t\wedge\tau_R^h}\int_{\mathbb{R}^d} \big(|Y_N^i(s^-) - X_N^i(s^-) \\ &\qquad+ \gamma(s)\,\mathrm{Diag}(Y_N^i(\kappa_h(s)) - \bar{Y}_N(\kappa_h(s)))z - \gamma(s)\,\mathrm{Diag}(X_N^i(s) \end{aligned}$$

$$- \bar{X}_N(s))z|^2 - |Y_N^i(s^-) - X_N^i(s^-)|^2\Big)\mathcal{N}^i(ds, dz).$$

Taking expectation on both sides, and using the Cauchy–Bunyakovsky–Schwartz inequality, Young's inequality, Ito's isometry (note that we can apply Ito's isometry due to Lemma 4.5) and Doob's optional stopping theorem, we get

$$\begin{aligned}
&\mathbb{E}\big(|Y_N^i(t\wedge\tau_R^h) - X_N^i(t\wedge\tau_R^h)|^2\big) \\
&\quad \le C(1+\mathbb{E}|Z|^2)\mathbb{E}\int_0^{t\wedge\tau_R^h}\big(|Y_N^i(\kappa_h(s)) - X_N^i(s)|^2 \\
&\qquad + |Y_N^i(s) - X_N^i(s)|^2 + |\bar{Y}_N(\kappa_h(s)) - \bar{X}_N(s)|^2\Big)ds \\
&\quad \le C\mathbb{E}\int_0^{t\wedge\tau_R^h}\Big(|Y_N^i(\kappa_h(s)) - Y_N^i(s)|^2 + |Y_N^i(s) - X_N^i(s)|^2 \\
&\qquad + |\bar{Y}_N(\kappa_h(s)) - \bar{Y}_N(s)|^2 + |\bar{Y}_N(s) - \bar{X}_N(s)|^2\Big)ds. \qquad (4.39)
\end{aligned}$$

Due to Lemma 4.6, we have

$$\sup_{i=1,\ldots,N}\mathbb{E}|Y_N^i(\kappa_h(s)) - Y_N^i(s)|^2 \le Ch, \qquad (4.40)$$

where C is independent of h, N and R.

Now, we will estimate the term $|\bar{Y}_N(s) - \bar{Y}_N(\kappa_h(s))|$. Recall that we used discrete Jensen's inequality, Assumptions 3.3 and 3.4 and the Cauchy–Bunyakovsky–Schwartz inequality to obtain (4.20). We apply the same set of arguments as before to get

$$\begin{aligned}
&|\bar{Y}_N(s) - \bar{Y}_N(\kappa_h(s))|^2 \\
&\quad \le C\Bigg(e^{\frac{2\alpha K_u}{N}\sum_{j=1}^N |Y_N^j(s)|^2}\frac{1}{N}\sum_{i=1}^N|Y_N^i(s) - Y_N^i(\kappa_h(s))|^2 \\
&\qquad + e^{\frac{2\alpha K_u}{N}\sum_{j=1}^N(|Y_N^j(s)|^2 + |Y_N^j(\kappa_h(s))|^2)} \\
&\qquad \times\left(\frac{1}{N}\sum_{i=1}^N(1 + |Y_N^i(s)|^2 + |Y_N^i(\kappa_h(s))|^2)^2\right) \\
&\qquad \times\left(\frac{1}{N}\sum_{i=1}^N|Y_N^i(s) - Y_N^i(\kappa_h(s))|^2\right)\Bigg). \qquad (4.41)
\end{aligned}$$

In the similar manner, we can obtain the following bound:

$$\begin{aligned}
&|\bar{X}_N(s) - \bar{Y}_N(s)|^2 \\
&\quad \le C\Bigg(e^{\frac{2\alpha K_u}{N}\sum_{j=1}^N |X_N^j(s)|^2}\frac{1}{N}\sum_{i=1}^N|X_N^i(s) - Y_N^i(s)|^2
\end{aligned}$$

$$+ e^{\frac{2\alpha K_u}{N}\sum_{j=1}^{N}(|X_N^j(s)|^2+|Y_N^j(s)|^2)} \times \left(\frac{1}{N}\sum_{i=1}^{N}(1+|X_N^i(s)|^2+|Y_N^i(s)|^2)^2\right)$$
$$\times \left(\frac{1}{N}\sum_{i=1}^{N}|X_N^i(s)-Y_N^i(s)|^2\right)\Bigg), \tag{4.42}$$

where $C>0$ is independent of h, N and R. We substitute (4.40)–(4.42) in (4.39) to get

$$\begin{aligned}
&\mathbb{E}\big(|Y_N^i(t\wedge\tau_R^h)-X_N^i(t\wedge\tau_R^h)|^2\big)\\
&\quad\le C\mathbb{E}\int_0^{t\wedge\tau_R^h}\big(|X_N^i(s)-Y_N^i(s)|^2\big)ds + Ch\\
&\qquad + CRe^{4\alpha K_u\sqrt{R}}\left(\mathbb{E}\int_0^{t\wedge\tau_R^h}\frac{1}{N}\sum_{i=1}^{N}\big(|Y_N^i(s)-Y_N^i(\kappa_h(s))|^2\big)ds\right.\\
&\qquad \left.+\mathbb{E}\int_0^{t\wedge\tau_R^h}\frac{1}{N}\sum_{i=1}^{N}\big(|X_N^i(s)-Y_N^i(s)|^2\big)ds\right)\\
&\quad\le C\int_0^t\mathbb{E}\big(|X_N^i(s\wedge\tau_R^h)-Y_N^i(s\wedge\tau_R^h)|^2\big)ds+Ch\\
&\qquad + CRe^{4\alpha K_u\sqrt{R}}\int_0^t\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\big(|Y_N^i(s\wedge\tau_R^h)-Y_N^i(\kappa_h(s\wedge\tau_R^h))|^2\big)ds\\
&\qquad + CRe^{4\alpha K_u\sqrt{R}}\int_0^t\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\big(|X_N^i(s\wedge\tau_R^h)-Y_N^i(s\wedge\tau_R^h)|^2\big)ds,
\end{aligned}$$

where $C>0$ is independent of h, N and R. Taking supremum over $i=1,\dots,N$ and using Lemma 4.6, we obtain

$$\begin{aligned}
&\sup_{i=1,\dots,N}\mathbb{E}\big(|Y_N^i(t\wedge\tau_R^h)-X_N^i(t\wedge\tau_R^h)|^2\big)\\
&\quad\le CRe^{4\alpha K_u\sqrt{R}}h + CRe^{4\alpha K_u\sqrt{R}}\int_0^t\sup_{i=1,\dots,N}\mathbb{E}\big(|Y_N^i(s\wedge\tau_R^h)\\
&\qquad - X_N^i(s\wedge\tau_R^h)|^2\big)ds,
\end{aligned}$$

where C is independent of h, N and R. Using Grönwall's lemma, we get

$$\begin{aligned}
\sup_{i=1,\dots,N}\mathbb{E}\big(|Y_N^i(t\wedge\tau_R^h)-X_N^i(t\wedge\tau_R^h)|^2\big) &\le CRe^{4\alpha K_u\sqrt{R}}e^{CRe^{4\alpha K_u\sqrt{R}}}h\\
&\le Ce^{e^{C_u\sqrt{R}}}h,
\end{aligned}$$

where $C>0$ and $C_u>0$ are constants independent of h, N and R.

We choose $R = \frac{1}{C_u^2}(\ln(\ln(h^{-1/2})))^2$. Consequently, we have

$$\begin{aligned}\sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_4(t)}\big) &\leq \sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t\wedge\tau_R^h) - X_N^i(t\wedge\tau_R^h)|^2\big)\\ &\leq Ch^{1/2},\end{aligned}$$

where $C > 0$ is independent of h and N. This implies

$$\begin{aligned}&\lim_{h\to 0}\lim_{N\to\infty}\sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_4(t)}\big)\\ &\quad= \lim_{N\to\infty}\lim_{h\to 0}\sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_4(t)}\big) = 0. \qquad (4.43)\end{aligned}$$

The term (4.38) and the choice of R provide the following estimate:

$$\sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_3(t)}\big) \leq \frac{C}{\ln(\ln(h^{-1/2}))},$$

where C is independent of h and N. This gives

$$\begin{aligned}&\lim_{h\to 0}\lim_{N\to\infty}\sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_3(t)}\big)\\ &\quad= \lim_{N\to\infty}\lim_{h\to 0}\sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2 I_{\Omega_3(t)}\big) = 0. \qquad (4.44)\end{aligned}$$

As a consequence of (4.43) and (4.44), we get

$$\begin{aligned}&\lim_{h\to 0}\lim_{N\to\infty}\sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2\big)\\ &\quad= \lim_{N\to\infty}\lim_{h\to 0}\sup_{i=1,\dots,N} \mathbb{E}\big(|Y_N^i(t) - X_N^i(t)|^2\big) = 0.\end{aligned}$$

□

5. Numerical Examples

In this section, we conduct numerical experiments on the Rastrigin and Rosenbrock functions by implementing the models (2.5), (2.7), (2.17) and the model with common noise introduced in Refs. 33 and 34. We use the Euler scheme for implementation with $h = 0.01$. We run 100 simulations and quote the success rates. We call a run of N particles a success if $|\bar{Y}_N(T) - x_{\min}| \leq 0.25$. Defining success rate in this manner is consistent with earlier CBO papers.

Experiment 5.1. We perform the experiment with the CBO model (2.5), JumpCBO model (2.7), JumpCBOwCPN model (jump-diffuison CBO model with common Poisson noise from (2.17)), CBOwCWN model (CBO model with common Wiener noise of Refs. 33 and 34) for the Rastrigin function

$$f(x) = 10 + \sum_{i=1}^{d}\big((x_i - B)^2 - 10\cos(2\pi(x_i - B))\big)/d, \qquad (5.1)$$

where we take the dimension $d = 20$ and $B = 0$. The minimum is located at $(0,\dots,0) \in \mathbb{R}^{20}$. In this experiment for the Rastrigin function, the initial search

Table 1. Success rate for $\alpha = 20$.

N	CBO	CBOwCWN	JumpCBO	JumpCBOwCPN
20	53	1	61	65
50	62	0	69	72
80	22	2	41	40
100	1	2	29	25

Table 2. Success rate for $\alpha = 30$.

N	CBO	CBOwCWN	JumpCBO	JumpCBOwCPN
20	87	0	90	94
50	99	0	100	100
80	100	0	100	100
100	100	0	100	100

space is $[-6, 6]^{20}$ and final time, $T = 100$. We take $\beta = 1$, $\sigma = 5.1$ for CBO, CBOwCWN, JumpCBO and JumpCBOwCPN models. We take $\gamma(t) = 1$ when $t \leq 20$ and $\gamma(t) = e^{1-t/20}$ when $t > 20$ for JumpCBO and JumpCBOwCPN models. Also, $\mathcal{Z}$ is distributed as standard Gaussian random variable and we choose the jump intensity, λ, of the Poisson process equal to 20.

The results are presented in Tables 1 and 2. In the case of the Rastrigin function, the performance of JumpCBO model (2.7), JumpCBOwCPN model (2.17) and CBO model (2.5) is comparable. However, CBOwCWN of Refs. 33 and 34 does not perform well. As the alpha is increased from 20 to 30, the success rates increase to a very high level (except for CBOwCWN). We have taken constant β and σ, and decaying γ for the jump-diffusion CBO models. As one can see, the jump CBO models perform better than the earlier CBO models when $\alpha = 20$. Another fact to be noticed is that performance of the jump-diffusion models with common or independent Poisson processes is very similar. It is also clear from the experiment that CBOwCWN model of Refs. 33 and 34 does not induce enough noise in the dynamics of the particle system sufficient for effective space exploration.

Experiment 5.2. We perform the experiment with the CBO model (2.5), JumpCBO model (2.7), CBOwCN model (CBO model with common noise of Refs. 33 and 34) for the Rosenbrock function

$$\sum_{i=1}^{d-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]/d, \tag{5.2}$$

where we take $d = 5$. The minimum is located at $(1, \dots, 1) \in \mathbb{R}^5$. In this experiment for the Rosenbrock function, the initial search space is $[-1, 3]^5$ and final time $T = 120$. We take $\beta = 1$, $\sigma = 5$ for CBO as well as CBOwCN models. We take $\beta(t) = 2 - e^{-t/100}$, $\sigma(t) = 4 + e^{-t/90}$ and $\gamma(t) = 1$ for $t \leq 90$ and $\gamma(t) = e^{1-t/90}$ for $t > 90$. Note that $\beta(0) = 1$ and $\sigma(0) = 5$ which are the same as the parameters β and σ for the CBO and CBOwCN models. Also, $\mathcal{Z}$ is distributed as standard Gaussian

Table 3. Success rate for $\alpha = 20$.

N	CBO	CBOwCWN	JumpCBO	JumpCBOwCPN
20	2	1	35	37
50	3	1	75	76
80	3	0	96	89
100	4	4	85	94

Table 4. Success rate for $\alpha = 30$.

N	CBO	CBOwCWN	JumpCBO	JumpCBOwCPN
20	6	2	20	25
50	3	0	49	45
80	5	2	69	64
100	4	1	74	70

random variable and we choose the jump intensity, λ, of the Poisson process equal to 90.

The results are presented in Tables 3 and 4. In the case of the Rosenbrock function, there is a significant improvement in finding global minimum when using the jump-diffusion models (2.7) and (2.17) in comparison with (2.5) and CBOwCWN of Refs. 33 and 34. As is the case with the Rastrigin function, for the Rosenbrock function, both jump-diffusion models have similar performance. We note that the Rosenbrock function has quartic growth. We take time-dependent $\beta(t)$, $\sigma(t)$ and $\gamma(t)$ for the jump diffusion models so that $\beta(t)$ is increasing function, $\sigma(t)$ is a decreasing function, and $\gamma(t)$ is constant for some period of time and then starts decreasing exponentially. This experiment illustrates a good balance of *exploration* and *exploitation* delivered by the proposed jump-diffusion models. The particles explore the space until $t = 90$ and after that particles start exploiting the searched space.

6. Concluding Remarks

We have developed a new CBO algorithm based on jump-diffusion SDEs, for which we have studied its well-posedness both at the particle level and its mean-field approximation. In particular, we proved mean-square convergence of the interacting particle system to the mean-field limit and of a discretized particle system to the continuous-time dynamics. The key feature of the jump-diffusion CBO is a more effective energy landscape exploration driven by the randomness introduced by both Wiener and Poisson processes. In practice, this translates into better success rates in finding the global minimizer and a more robust initialization, which can be located far away from the global minimizer.

One of possible extensions of the proposed CBO model is to develop a particle system driven by path-dependent jump-diffusion SDEs. This model will have a non-Markovian structure which can potentially be utilized to force particles to search unexplored space. Another natural extension of the current work is a systematic study of CBO with constraints in the search space as recently discussed in Refs. 5, 14, 23 and 31. This is particularly challenging because of the need to accurately treat boundary conditions for the SDEs (see e.g. Ref. 43). One more interesting research direction is the exploration of jump-diffusion processes in the framework of kinetic-type CBO models.[8, 41]

Acknowledgments

AS was supported by EPSRC grant no. EP/W52251X/1. DK was supported by EPSRC grants EP/T024429/1 and EP/V04771X/1.

References

1. N. Agram and B. Oksendal, Stochastic Fokker-Planck PIDE for conditional McKean-Vlasov jump diffusions and applications to optimal control, preprint (2021), arXiv:2110.02193.
2. G. Albi, N. Bellomo, L. Fermo, S. Y. Ha, J. Kim, L. Pareschi, D. Poyato and J. Soler, Vehicular traffic, crowds, and swarms: From kinetic theory and multiscale methods to applications and research perspectives, *Math. Models Methods Appl. Sci.* **29** (2019) 1901–2005.
3. D. Applebaum, *Lévy Processes and Stochastic Calculus* (Cambridge Univ. Press, 2004).
4. T. Back, D. B. Fogel and Z. Michalewicz, *Handbook of Evolutionary Computation* (IOP, 1997).
5. H.-O. Bae, S.-Y. Ha, M. Kang, H. Lim, C. Min and J. Yoo, A constrained consensus based optimization algorithm and its application to finance, *Appl. Math. Comput.* **416** (2022) 126726.
6. N. Bellomo, D. Burini, G. Dosi, L. Gibelli, D. Knopoff, N. Outada, P. Terna and M. E. Virgillito, What is life? A perspective of the mathematical kinetic theory of active particles, *Math. Models Methods Appl. Sci.* **31** (2021) 1821–1866.
7. N. Bellomo, L. Gibelli, A. Quaini and A. Reali, Towards a mathematical theory of behavioral human crowds, *Math. Models Methods Appl. Sci.* **32** (2022) 321–358.
8. A. Benfenati, G. Borghi and L. Pareschi, Binary interaction methods for high dimensional global optimization and machine learning, *Appl. Math. Optim.* **86** (2022) 9.
9. A. Bertozzi, J. Rosado, M. Short and L. Wang, Contagion shocks in one dimension, *J. Stat. Phys.* **158** (2014) 647–664.
10. P. Bonacich and P. Lu, *Introduction to Mathematical Sociology* (Princeton Univ. Press, 2012).
11. J. A. Carrillo, Y. Choi, C. Totzeck and O. Tse, An analytical framework for consensus-based global optimization method, *Math. Models Methods Appl. Sci.* **28** (2018) 1037–1066.
12. J. A. Carrillo, M. Fornasier, J. Rosado and G. Toscani, Asymptotic flocking dynamics for the kinetic Cucker–Smale model, *SIAM J. Math. Anal.* **42** (2010) 218–236.
13. J. A. Carrillo, S. Jin, L. Li and Y. Zhu, A consensus-based global optimization method for high dimensional machine learning problems, *ESAIM Control Optim. Calc. Var.* **27** (2021) S5.

14. J. A. Carrillo, C. Totzeck and U. Vaes, Consensus-based optimization and ensemble Kalman inversion for global optimization problems with constraints, preprint (2021), arXiv:2111.02970.
15. M. Chak, N. Kantas and G. A. Pavliotis, On the generalised Langevin equation for simulated annealing, preprint (2020), arXiv:2003.06448.
16. A. R. Conn, K. Scheinberg and L. N. Vicente, *Introduction to Derivative-Free Optimization* (SIAM, 2009).
17. F. Cucker and S. Smale, On the mathematics of emergence, *Jpn. J. Math.* **2** (2007) 197–227.
18. K. Dareiotis, C. Kumar and S. Sabanis, On tamed Euler approximations of SDEs driven by Lévy noise with applications to delay equations, *SIAM J. Numer. Anal.* **54** (2016) 1840–1872.
19. G. Deligiannidis, S. Maurer and M. Tretyakov, Random walk algorithm for the Dirichlet problem for parabolic integro-differential equation, *BIT Numer. Math.* **61** (2021) 1223–1269.
20. M. Dorigo and C. Blum, Ant colony optimization theory: A survey, *Theor. Comput. Sci.* **344** (2005) 243–278.
21. X. Erny, Well-posedness and propagation of chaos for McKean–Vlasov equations with jumps and locally Lipschitz coefficients, *Stochastic Process. Appl.* **150** (2022) 192–214.
22. M. Fornasier, H. Huang, L. Pareschi and P. Sünnen, Consensus-based optimization on hypersurfaces: Well-posedness and mean-field limit, *Math. Models Methods Appl. Sci.* **30** (2020) 2725–2751.
23. M. Fornasier, H. Huang, L. Pareschi and P. Sünnen, Anisotropic diffusion in consensus-based optimization on the sphere, *SIAM J. Optim.* **32** (2022) 1984–2012.
24. M. Fornasier, T. Klock and K. Riedl, Consensus-based optimization methods converge globally, preprint (2021), arXiv:2103.15130.
25. M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, 2012).
26. S. Galam and S. Moscovici, Towards a theory of collective phenomena: Consensus and attitude changes in groups, *Eur. J. Soc. Psychol.* **21** (1991) 49–74.
27. D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of Second Order* (Springer, 1983).
28. C. Graham, McKean–Vlasov Ito–Skorohod equations, and nonlinear diffusions with discrete jump sets, *Stochastic Process. Appl.* **40** (1992) 69–82.
29. C. Graham, Nonlinear diffusion with jumps, *Ann. Inst. Henri Poincaré Probab. Stat.* **28** (1992) 393–402.
30. S. Grassi, H. Huang, L. Pareschi and J. Qiu, Mean-field particle swarm optimization, preprint (2021), arXiv:2108.00393.
31. S. Grassi and L. Pareschi, From particle swarm optimization to consensus based optimization: Stochastic modeling and mean-field limit, *Math. Models Methods Appl. Sci.* **31** (2021) 1625–1657.
32. I. Gyöngy and N. V. Krylov, On stochastic equations with respect to semimartingales I, *Stochastics* **4** (1980) 1–21.
33. S.-Y. Ha, S. Jin and D. Kim, Convergence of a first-order consensus-based global optimization algorithm, *Math. Models Methods Appl. Sci.* **30** (2020) 2417–2444.
34. S.-Y. Ha, S. Jin and D. Kim, Convergence and error estimates for time-discrete consensus-based optimization algorithms, *Numer. Math.* **147** (2021) 255–282.
35. S.-Y. Ha, M. Kang, D. Kim, J. Kim and I. Yang, Stochastic consensus dynamics for nonconvex optimization on the Stiefel manifold: Mean-field limit and convergence, *Math. Models Methods Appl. Sci.* **32** (2022) 533–617.

36. S.-Y. Ha and E. Tadmor, From particle to kinetic and hydrodynamic descriptions of flocking, *Kinet. Relat. Models* **1** (2008) 415–435.
37. R. A. Holley, S. Kusuoka and D. W. Stroock, Asymptotics of the spectral gap with applications to the theory of simulated annealing, *J. Funct. Anal.* **83** (1989) 333–347.
38. H. Huang and J. Qiu, On the mean-field limit for the consensus-based optimization, preprint (2021), arXiv:2105.12919.
39. S. Ken-Iti, *Lévy Processes and Infinitely Divisible Distributions* (Cambridge Univ. Press, 1999).
40. J. Kennedy, Particle swarm optimization, in *Encyclopedia of Machine Learning*, eds. C. Sammut and G. I. Webb (Springer, 2010), pp. 760–766.
41. D. Ko, S. Y. Ha, S. Jin and D. Kim, Convergence analysis of the discrete consensus-based optimization algorithm with random batch interactions and heterogeneous noises, *Math. Models Methods Appl. Sci.* **32** (2022) 1071–1107.
42. G. N. Milstein and M. V. Tretyakov, Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients, *SIAM J. Numer. Anal.* **43** (2005) 1139–1154.
43. G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, 2nd edn. (Springer, 2021).
44. S. Motsch and E. Tadmor, Heterophilious dynamics enhances consensus, *SIAM Rev.* **56** (2014) 577–621.
45. Neelima, S. Biswas, C. Kumar, G. dos Reis and C. Reisinger, Well-posedness and tamed Euler schemes for McKean–Vlasov equations driven by Lévy noise, preprint (2020), arXiv:2010.08585.
46. R. Pinnau, C. Totzeck, O. Tse and S. Martin, A consensus-based model for global optimization and its mean-field limit, *Math. Models Methods Appl. Sci.* **27** (2017) 183–204.
47. E. Platen and N. Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, Stochastic Modelling and Applied Probability (Springer, 2010).
48. A. Shiryaev, *Probability* (Springer, 2013).
49. A. S. Sznitman, Topics in propagation of chaos, in *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989*, ed. P.-L. Hennequin (Springer, 1991), pp. 165–251.
50. G. Toscani, Kinetic models of opinion formation, *Commun. Math. Sci.* **4** (2006) 481–496.
51. C. Totzeck, Trends in consensus-based optimization, in *Active Particles, Volume 3: Advances in Theory, Models, and Applications*, eds. N. Bellomo, J. A. Carrillo and E. Tadmor (Springer, 2022), pp. 201–226.
52. C. Villani, *Topics in Optimal Transportation* (Amer. Math. Soc., 2003).