THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Euler simulation of interacting particle systems and McKean-Vlasov SDEs with fully super-linear growth drifts in space and interaction

Citation for published version:
Chen, X & Dos Reis, G 2023, 'Euler simulation of interacting particle systems and McKean-Vlasov SDEs with fully super-linear growth drifts in space and interaction', *IMA Journal of Numerical Analysis*. https://doi.org/10.1093/imanum/drad022

Digital Object Identifier (DOI):
10.1093/imanum/drad022

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
IMA Journal of Numerical Analysis

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 16. Aug. 2023

EULER SIMULATION OF INTERACTING PARTICLE SYSTEMS AND MCKEAN-VLASOV SDES WITH FULLY SUPER-LINEAR GROWTH DRIFTS IN SPACE AND INTERACTION [*]

XINGYUAN CHEN [†] AND GONÇALO DOS REIS [†‡]

Abstract.
This work addresses the convergence of a split-step Euler type scheme (SSM) for the numerical simulation of interacting particle Stochastic Differential Equation (SDE) systems and McKean-Vlasov Stochastic Differential Equations (MV-SDEs) with full super-linear growth in the spatial and the interaction component in the drift, and non-constant Lipschitz diffusion coefficient. Super-linearity is understood in the sense that functions are assumed to behave polynomially but also satisfy a so-called one-sided Lipschitz condition.

The super-linear growth in the interaction (or measure) component stems from convolution operations with super-linear growth functions allowing in particular application to the granular media equation with multi-well confining potentials. From a methodological point of view, we avoid altogether functional inequality arguments (as we allow for non-constant non-bounded diffusion maps).

The scheme attains, in stepsize, a near-optimal classical (path-space) root mean-square error rate of $1/2-\varepsilon$ for $\varepsilon>0$ and an optimal rate $1/2$ in the non-path-space (pointwise) mean-square error metric. All findings are illustrated by numerical examples. In particular, the testing raises doubts if taming is a suitable methodology for this type of problem (with convolution terms and non-constant diffusion coefficients).

Key words. stochastic interacting particle systems, McKean-Vlasov equations, split-step Euler methods, super-linear growth in measure, super-linear growth in space

AMS subject classifications. 65C05, 65C30, 65C35

1. Introduction. Interactions of organisms, humans, and objects are common phenomena seen easily in collective behaviour within natural and social sciences. Models for interacting particle systems (IPS) and their mesoscopic limits, as the number of particles grows to infinity, receive presently enormous attention given their applicability in areas such as finance, mathematical neuroscience, biology, machine learning, and physics: animal swarming, cell movement induced by chemotaxis, opinion dynamics, particle movement in porous media, electrical battery modelling, self-assembly of particles (see for example [5, 10, 11, 13, 14, 24, 27, 29, 33, 37, 38, 43, 48, 51] and references). In this work, we address the numerical approximation of interacting particle systems given by stochastic differential equations (SDE) and their mesoscopic limit equations (or a class thereof) called McKean–Vlasov Stochastic Differential Equations (MV-SDE) that follow as the scaling limit of an infinite number of particles.

We understand the IPS as an N-dimensional system of $\mathbb{R}^d$-valued interacting particles where each particle is governed by a Stochastic Differential Equation (SDE). Let $i=1,\cdots,N$ and consider N particles $(X_t^{i,N})_{t\in[0,T]}$ with independent and identically distributed $X_0^{i,N}=X_0^i$ (the initial condition is random, but independent of other particles) and satisfying the $(\mathbb{R}^d)^N$-valued SDE (1.1)

$$\mathrm{d}X_t^{i,N}=\big(v(X_t^{i,N},\mu_t^{X,N})+b(t,X_t^{i,N},\mu_t^{X,N})\big)\mathrm{d}t+\sigma(t,X_t^{i,N},\mu_t^{X,N})\mathrm{d}W_t^i,\quad X_0^{i,N}=X_0^i\in L_0^m(\mathbb{R}^d), \tag{1.1}$$

$$\text{for } v\big(X_t^{i,N},\mu_t^{X,N}\big)=\Big(\frac{1}{N}\sum_{j=1}^N f(X_t^{i,N}-X_t^{j,N})\Big)+u\big(X_t^{i,N},\mu_t^{X,N}\big)\text{ with }\mu_t^{X,N}(\mathrm{d}x):=\frac{1}{N}\sum_{j=1}^N\delta_{X_t^{j,N}}(\mathrm{d}x), \tag{1.2}$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, $\{W^i\}_{i=1,\cdots,N\}}$ are independent Brownian motions and $L_0^m(\mathbb{R}^d)$ denotes the usual mth-moment integrable space of $\mathbb{R}^d$ random variables.

For the IPS class (1.1), the limiting class as $N\to\infty$ are called McKean-Vlasov SDEs and the passage to the limit operation is known as "Propagation of Chaos". This class was first described by McKean [50], where he

[*] Submitted to the editors on 2021/12/09; 1st revision submitted on 2022/08/24; 2nd revision submitted on 2023/01/19;

Funding: G.d.R. acknowledges support from the *Fundação para a Ciência e a Tecnologia* (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 and UIDP/00297/2020 (Centro de Matemática e Aplicações CMA/FCT/UNL).

[†] School of Mathematics, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK. (X.Chen-176@sms.ed.ac.uk, G.dosReis@ed.ac.uk).

[‡]Centro de Matemática e Aplicações (CMA), FCT, UNL, 2829-516 Caparica, Portugal

This manuscript is for review purposes only.

introduced the convolution type interaction (the v in (1.2)). This is a class of Markov processes associated with nonlinear parabolic equations where the map v in (1.2) is also called “self-stabilizing”. The IPS underpinning our work (1.1)-(1.2) has been studied widely, from a variety of points of view and as early as [55] (for a general survey under global Lipschitz conditions and boundedness).

McKean-Vlasov Stochastic Differential Equations (MV-SDEs) with convolution type drifts have general dynamics given by

$$\mathrm{d}X_t = \big(v(X_t,\mu_t^X) + b(t,X_t,\mu_t^X)\big)\mathrm{d}t + \sigma(t,X_t,\mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^m(\mathbb{R}^d), \tag{1.3}$$

$$\text{where } v(x,\mu) = \int_{\mathbb{R}^d} f(x-y)\mu(\mathrm{d}y) + u(x,\mu) \quad \text{with} \quad \mu_t^X = \mathrm{Law}(X_t), \tag{1.4}$$

where μ_t^X denotes the law of the solution process X at time t, W is a Brownian motion in $\mathbb{R}^d$, v, f, u, b, σ are measurable maps along with a sufficiently integrable initial condition X_0.

An embodiment (among many) for this typology of models is particle motion modelling that encapsulates three sources of forcing. Namely, the particle moves through a multi-well landscape potential gradient (the map u and b), the trajectories are affected by a Brownian motion (and associated diffusion coefficient σ), and the convolution self-stabilisation forcing characterises the influence of a large population of identical particles (under the same laws of motion v and f) on the particle. In effect, v acts on the particle as an average attractive/repulsive force exerted on the said particle by a population of similar particles (through the potential f), see [1, 57] and further examples in [37]. For instance, under certain constraints on f the map v adds inertia to the particle's motion, which in turn delays exit times from the domain of attraction and alters exit locations [1, 22, 31]. The self-stabilisation term in the system induces in the corresponding Fokker-Plank equation a nonlinear term of the form $\nabla[\rho.\nabla(f \star \rho)]$ (where ρ stands for the processes density while '$\star$' is the usual convolution operator) [13, 14, 37]. The granular media Fokker-Plank equation from biochemistry is a good example of an equation featuring this kind of structure [1, 15, 46]. The literature on MV-SDE is growing explosively with many contributions addressing well-posedness, regularity, ergodicity, nonlinear Fokker-Planck equations, large deviations [2, 3, 22, 34]. The convolution framework has been given particular attention as it underpins many settings of interest [15, 30, 46, 57]. The literature is even richer under the restriction to a constant diffusion term, $\sigma = \text{const}$, as it gives access to methodologies based on Langevin-type dynamics but also to the machinery of Functional inequalities (e.g., log-Sobolev and Poincare inequalities). We point to [30] for a nice overview on several *open* problems of interest where f is a singular kernel (and σ is a constant): including Coulomb interaction $f(x) = x/|x|^d$, Bio-Savart law $f(x) = x^\perp/|x|^d$; Cucker-Smale models $f(x) = (1+|x|^2)^{-\alpha}$ for $\alpha > 0$; crystallisation $f(x) = |x|^{-2p} - 2|x|^{-p}$ and take $p \to \infty$; 2D viscous vortex model with $f(x) = x/|x|^2$ [25].

Super-linear interaction forces. For the IPS (1.1)-(1.2) or the MV-SDE (1.3)-(1.4), we focus on the class where the involved functions are not (necessarily) globally Lipschitz functions. Concretely, the map v is a super-linear growth function in both space and measure component — we assume that f and u in (1.4) behave like a general polynomial but also satisfy a one-sided Lipschitz condition to control for radial growth (the specific details are given in Assumption 2.1 below); the maps b and σ are assumed globally Lipschitz functions.

From the theoretical point of view, this class is presently well understood. Well-posedness was generally established in [1]; [32] investigate different properties of the invariant measures for particles in double-well confining potential and later [57] investigate the convergence to stationary states. Large deviations and exit times for such self-stabilising diffusions are established in [1, 31]. The study of probabilistic properties and parametric inference (under constant diffusion) for this class is given in [26]. Two recent studies on parametric inference [7, 18] include numerical studies for the particle interaction ([26] does not) but do not tackle super-linear growth in the interaction component ([26] does).

To the best of our knowledge and except for [45], no numerical methods exists for this class as no general method allows for super-linear growth interaction kernels. For emphasis, standard SDE results for super-linear growth drifts do not yield convergence results independent of the number of particles N. In other words, by treating the interacting particle system (1.1) as an $(\mathbb{R}^d)^N$-dimensional SDE known results from SDE numerics with coefficients with super-linear growth can be applied directly. *However*, all estimates would depend on the

This manuscript is for review purposes only.

system's dimension, Nd, and hence "explode" as N tends to infinity. In this work, we introduce new technical elements to overcome this difficulty, which, to the best of our knowledge, are new. It's noteworthy to observe that the direct numerical discretization of the IPS system (1.1)-(1.2) leads to a costly computational cost of $\mathcal{O}(N^2)$ and hence care is needed.

Many of the current numerical methods in the literature of MV-SDEs rely on the particle approximation given by the IPS, and the known quantified rate for the propagation of chaos [1, 16, 40, 41]: taming [21, 39], time-adaptive [52], early Split-Step Methods (SSM) methods [17] – all these contributions allow for super-linear growth in space only. Further noteworthy contributions include [4, 6, 8, 12, 19, 23, 28, 36, 56]. Within the existing literature, no method can deal with a super-linear growth f component; all cited works make the assumption of a Lipschitz behaviour in $\mu \mapsto v(\cdot, \mu)$ (which, in essence, entail that ∇f is bounded).

Our contribution. *The results of this manuscript provide for both the numerical approximation of interacting particle SDE systems* (1.1)-(1.2)*, and McKean–Vlasov SDEs* (1.3)-(1.4).

The main contribution of this work is the numerical scheme and its convergence analysis. We present a particle approximation SSM algorithm inspired in [17] for the numerical approximation of MV-SDEs and associated particle systems with drifts featuring super-linear growth in space and measure, and where the diffusion coefficient satisfies a general Lipschitz condition. The well-posedness result (Theorem 2.3 below) and Propagation of Chaos (Proposition 2.5 below) follow from known literature [1] – in fact, our Proposition 2.5 establishes the well-posedness of the particle system hence closing the small gap present in [1, Theorem 3.14]. The only existing work tackling this involved setting via a fully implicit scheme is [45]. They rely on (Bakry-Emery) functional inequalities methodologies under specific structural assumptions (constant elliptic diffusion, $u = b = 0$ and differentiability) that we do not make.

The scheme we propose is a split-step scheme inspired in [17] (see Definition 2.6 below) that first solves an implicit equation given by the SDE's drift component only then takes that outcome and feeds it to the remaining dynamics of the SDE via a standard Euler step. The idea is that the implicit step deals with the problematic super-linear growth part, and the elements passed to the Euler step are better behaved. In [17], there is only super-linear growth in the space variables, and the measure component is assumed Lipschitz; here both space and measure component have super-linear growth. From a practical point of view, the implicit step in [17] for a particle i only depended on the elements of particle i (the measure being fixed to the previous time step); hence one solves N decoupled equations in $\mathbb{R}^d$. In this manuscript, the implicit step for particle i involves the whole system of particles entailing that one needs to solve one-single system but in $(\mathbb{R}^d)^N$ and the solution depends on all terms. This change in the scheme makes it much harder to obtain moment estimates for the scheme. For the setting of [17] there were already several competitive schemes present in the literature, e.g., taming [21, 39] and time-adaptive [52] and the numerical study there was comparative. For this work, no alternative numerical scheme exists – see below for further discussion regarding the implementation of taming for this class.

Results-wise, we provide two convergence results in the strong-error[1] sense. For the classical (path-space) root mean-square error, see Theorem 2.11, we achieve a nearly-optimal convergence rate of $1/2 - \varepsilon$ with $\varepsilon > 0$. The main difficulty, also where one of our main contributions lie, is in establishing higher-order moment bounds for the numerical scheme in a way that is compatible with the convolution component in (1.2) or (1.4) and Itô-type arguments – see Theorem 2.10. We provide a second strong (non-path-space) mean-square error criteria, see Theorem 2.10, that attains the optimal rate $1/2$. This 2nd result requires only the higher moments of the IPS' solution process and the 2nd-moments of the numerical approximation [9] (which are easier to obtain). We emphasise that this 2nd notion of strong convergence (see Theorem 2.10) is also standard (albeit less) within Monte Carlo literature. It also controls the variance of the approximation error (simply not in path-space). Hence, it is sufficient for the many uses one can give to the simulation output – as one would do given any other Monte Carlo estimators (e.g., confidence intervals). Lastly, we show that with a constant diffusion coefficient, one attains the higher convergence rate of 1.0 (see Theorem 2.13).

[1]We understand a "strong" error metric as a metric that depends on the joint distribution of the true solution and the numerical approximation. In contrast to the weak error where one needs only the marginals separately. Theorem 2.9 and 2.11 showcase two "strong" but different error metrics.

This manuscript is for review purposes only.

We illustrate our findings with extended numerical tests showing agreement with the theoretical results and discussing other properties for schemes: periodicity in phase-space, the impact of the number of particles and numerical rate of Propagation of Chaos, and complexity versus runtime. For comparison, we implement the taming algorithm [21] for the setting (without proof) and find that in the example with constant diffusion, taming performs similarly to the SSM. In the non-constant diffusion example, it performs very poorly. This latter finding raises questions (for future research) if taming is a suitable methodology for this class.

Organisation of the paper. In Section 2 we set the notation and framework. In Section 2.3, we state the SSM scheme and the two main convergence results. Section 3 provides numerical illustrations (for the granular media model and a double-well model with non-constant diffusion). All proofs are given in Section 4.

2. The split-step method for MV-SDEs and interacting particle systems.

We follow the notation and framework set in [1, 17].

2.1. Notation and Spaces.

Let $\mathbb{N}$ be the set of natural numbers starting at 0, $\mathbb{R}$ denotes the real numbers. For $a,b\in\mathbb{N}$ with $a\le b$, define $[\![a,b]\!]:=[a,b]\cap\mathbb{N}=\{a,\cdots,b\}$. For $x,y\in\mathbb{R}^d$ denote the scalar product of vectors by $x\cdot y$; and $|x|=(\sum_{j=1}^d x_j^2)^{1/2}$ the Euclidean distance. The $\mathbf{0}$ denotes the origin in $\mathbb{R}^d$. Let $\mathbb{1}_A$ be the indicator function of set $A\subset\mathbb{R}^d$. For a matrix $A\in\mathbb{R}^{d\times n}$ we denote by $A^\intercal$ its transpose and its Frobenius norm by $|A|=\text{Trace}\{AA^\intercal\}^{1/2}$. Let $I_d:\mathbb{R}^d\to\mathbb{R}^d$ be the identity map. For collections of vectors, let the upper indices denote the distinct vectors, whereas the lower index is a vector component, i.e., x_j^l denote the j-th component of l-th vector. ∇ denotes the vector differential operator, ∂ denotes the partial differential operator.

We introduce over $\mathbb{R}^d$ the space of probability measures $\mathcal{P}(\mathbb{R}^d)$ and its subset $\mathcal{P}_2(\mathbb{R}^d)$ of those with finite second moment. The space $\mathcal{P}_2(\mathbb{R}^d)$ is Polish under the Wasserstein distance

$$W^{(2)}(\mu,\nu)=\inf_{\pi\in\Pi(\mu,\nu)}\Big(\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\pi(\mathrm{d}x,\mathrm{d}y)\Big)^{\frac{1}{2}},\quad \mu,\nu\in\mathcal{P}_2(\mathbb{R}^d). \tag{2.1}$$

where $\Pi(\mu,\nu)$ is the set of couplings for μ and ν such that $\pi\in\Pi(\mu,\nu)$ is a probability measure on $\mathbb{R}^d\times\mathbb{R}^d$ such that $\pi(\cdot\times\mathbb{R}^d)=\mu$ and $\pi(\mathbb{R}^d\times\cdot)=\nu$.

Let our probability space be a completion of $(\Omega,\mathbb{F},\mathcal{F},\mathbb{P})$ with $\mathbb{F}=(\mathcal{F}_t)_{t\ge0}$ carrying an $\mathbb{R}^l$-valued Brownian motion $W=(W^1,\cdots,W^l)$ and generating the probability space's filtration, augmented by all $\mathbb{P}$-null sets, and with an additionally sufficiently rich sub σ-algebra $\mathcal{F}_0$ independent of W. We denote by $\mathbb{E}[\cdot]=\mathbb{E}^{\mathbb{P}}[\cdot]$ the usual expectation operator with respect to $\mathbb{P}$.

We consider some finite terminal time $T<\infty$ and use the following notation for spaces (standard in the (McKean-Vlasov) SDE literature [17, 21]). For $0\le t\le T$, let $L_t^p\left(\mathbb{R}^d\right)$ define the space of $\mathbb{R}^d$-valued, $\mathcal{F}_t$-measurable random variables X, that satisfy $\mathbb{E}\left[|X|^p\right]^{1/p}<\infty$. Define $\mathbb{S}^m([0,T])$ to be, for $m\geqslant 1$, the space of $\mathbb{R}^d$ -valued, $\mathcal{F}_\cdot$-adapted processes Z, that satisfy $\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}|Z_t|^m\right]^{1/m}<\infty$.

Throughout the text, C denotes a generic constant positive real number that may depend on the problem's data, may change from line to line but is always independent of the constants h,M,N (associated with the numerical scheme and specified below) but possibly depend on the terminal time T (and other fixed problem data).

2.2. Framework.

Let W be an l-dimensional Brownian motion and take the measurable maps $v:\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}^d$, $f:\mathbb{R}^d\to\mathbb{R}^d$, $b:[0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}^d$ and $\sigma:[0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}^{d\times l}$. The MV-SDE of interest for this work is Equation (1.3) (for some $m\ge1$), where μ_t^X denotes the law of the process X at time t, i.e., $\mu_t^X=\mathbb{P}\circ X_t^{-1}$. We make the following assumptions on the coefficients.

ASSUMPTION 2.1. *Let b and σ $1/2$-Hölder continuous in time, uniformly in $x\in\mathbb{R}^d$ and $\mu\in\mathcal{P}_2(\mathbb{R}^d)$. Assume that b,σ are uniformly Lipschitz in the sense that there exists $L_b,L_\sigma\ge0$ such that for all $t\in[0,T]$ and all $x,x'\in\mathbb{R}^d$ and $\forall\mu,\mu'\in\mathcal{P}_2(\mathbb{R}^d)$ we have that*

$$(\mathbf{A}^b)\qquad |b(t,x,\mu)-b(t,x',\mu')|^2\le L_b\big(|x-x'|^2+W^{(2)}(\mu,\mu')^2\big),$$

$$(\mathbf{A}^\sigma)\qquad |\sigma(t,x,\mu)-\sigma(t,x',\mu')|^2\le L_\sigma\big(|x-x'|^2+W^{(2)}(\mu,\mu')^2\big).$$

This manuscript is for review purposes only.

$(\mathbf{A}^u)$ *Let u satisfy: there exist $L_u \in \mathbb{R}$, $L_{\hat{u}} > 0$, $L_{\tilde{u}} \geq 0$, $q_1 > 0$ such that for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$ and $\forall \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, it holds that*

$$\begin{aligned} \langle x - x', u(x,\mu) - u(x',\mu)\rangle &\leq L_u |x-x'|^2 && \textit{(One-sided Lipschitz in space)},\\ |u(x,\mu) - u(x',\mu)| &\leq L_{\hat{u}}(1 + |x|^{q_1} + |x'|^{q_1})|x-x'| && \textit{(Locally Lipschitz in space)},\\ |u(x,\mu) - u(x,\mu')|^2 &\leq L_{\tilde{u}} W^{(2)}(\mu,\mu')^2 && \textit{(Lipschitz in measure)}. \end{aligned}$$

$(\mathbf{A}^f)$ *Let f satisfy: there exist $L_f \in \mathbb{R}$, $L_{\hat{f}} > 0$, $q_2 > 0$ such that for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$, it holds that*

$$\begin{aligned} \langle x - x', f(x) - f(x')\rangle &\leq L_f |x-x'|^2 && \textit{(One-sided Lipschitz)},\\ |f(x) - f(x')| &\leq L_{\hat{f}}(1 + |x|^{q_2} + |x'|^{q_2})|x-x'| && \textit{(Locally Lipschitz)},\\ f(x) &= -f(-x), && \textit{(Odd function)}. \end{aligned}$$

Assume the normalisation[2] *$f(\mathbf{0}) = \mathbf{0}$. Lastly, and for convenience, we set $q = \max\{q_1, q_2\}$ (and we have $q > 0$).*

The benefits of choosing drift$=v+b$ with b being uniformly Lipschitz are discussed below in Remark 2.7 (see also [17]). Certain useful properties can be derived from these assumptions.

REMARK 2.2 (Implied properties). *Under Assumption 2.1, take some $C > 0$. Then for all $t \in [0,T]$, $x, x', z \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, since f is a normalised odd function (i.e., $f(\mathbf{0}) = \mathbf{0}$), we have*

$$\langle x, f(x)\rangle = \langle x - \mathbf{0}, f(x) - f(\mathbf{0})\rangle + \langle x, f(\mathbf{0})\rangle \leq L_f|x|^2 + |x||f(\mathbf{0})| = L_f|x|^2.$$

Also, for the function u, define $\widehat{L}_u = L_u + 1/2,\ C_u = |u(0,\delta_0)|^2$, and thus by Young's inequality

$$\langle x, u(x,\mu)\rangle \leq C_u + \widehat{L}_u|x|^2 + L_{\tilde{u}} W^{(2)}(\mu,\delta_0)^2, \quad \langle x - x', u(x,\mu) - u(x',\mu')\rangle \leq \widehat{L}_u|x-x'|^2 + \frac{L_{\tilde{u}}}{2} W^{(2)}(\mu,\delta_0)^2.$$

Using the properties of the convolution, v of (1.3) also satisfies a one-sided Lipschitz condition in space

$$\langle x - x', v(x,\mu) - v(x',\mu)\rangle \leq \int_{\mathbb{R}^d} L_f|x-x'|^2 \mu(dz) + L_u|x-x'|^2 \ = \ (L_f + L_u)|x-x'|^2.$$

Moreover, for $\psi \in \{b, \sigma\}$, by Young's inequality, we have

$$\langle x, \psi(t,x,\mu)\rangle \leq C(1 + |x|^2 + W^{(2)}(\mu,\delta_0)^2) \quad \textit{and} \quad |\psi(t,x,\mu)|^2 \leq C(1 + |x|^2 + W^{(2)}(\mu,\delta_0)^2).$$

We first recall a result from [1] establishing well-posedness of the MV-SDE (1.3)-(1.4).

THEOREM 2.3 (Theorem 3.5 in [1]). *Let Assumption 2.1 hold and assume for some $m > 2(q+1)$, $X_0 \in L_0^m(\mathbb{R}^d)$. Then, there exists a unique solution X to MV-SDE (1.3) in $\mathbb{S}^m([0,T])$. For some constant $C > 0$ (depending on T and m) we have*

$$\mathbb{E}\big[\sup_{t\in[0,T]} |X_t|^{\widehat{m}}\big] \leq C\big(1 + \mathbb{E}\big[|X_0|^{\widehat{m}}\big]\big)e^{CT}, \qquad \textit{for any } \ \widehat{m} \in [2, m].$$

Proof. Our Assumption 2.1 is a particularisation of [1, Assumption 3.4] and hence our theorem follows directly from [1, Theorem 3.5]. ∎

The interacting particle system (1.1). As mentioned earlier, the numerical approximation results of this work apply directly if either one's starting point is the interacting particle system (1.1) or if one's starting point is the MV-SDE (1.3). On the latter, one can approximate the MV-SDE (1.3) (driven by the Brownian motion W) by

[2] This constraint is a soft as the framework allows to easily redefine f as $\hat{f}(x) := f(x) - f(\mathbf{0})$ with $f(\mathbf{0})$ merged into b.

This manuscript is for review purposes only.

the N-dimensional system $\mathbb{R}^d$-valued interacting particle system given in (1.1) and approximate it numerically with the gap closed by the Propagation of Chaos [17, 21, 52].

For completeness we recall the setup of (1.1). Let $i \in [\![1,N]\!]$ and consider N particles $(X^{i,N})_{t\in[0,T]}$ with independent and identically distributed (i.i.d.) initial conditions $X_0^{i,N} = X_0^i$ and satisfying the $(\mathbb{R}^d)^N$-valued SDE (1.1) (with v given in (1.4))

$$\mathrm{d}X_t^{i,N} = \big(v(X_t^{i,N}, \mu_t^{X,N}) + b(t, X_t^{i,N}, \mu_t^{X,N})\big)\mathrm{d}t + \sigma(t, X_t^{i,N}, \mu_t^{X,N})\mathrm{d}W_t^i, \quad X_0^{i,N} = X_0^i,$$

where $\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}(\mathrm{d}x)$ with $\delta_{X_t^{j,N}}$ being the Dirac measure at point $X_t^{j,N}$, and $W^i, i \in [\![1,N]\!]$ being independent Brownian motions (also independent of the BM W appearing in (1.3); with a slight abuse of notation to avoid re-defining the probability space's filtration).

REMARK 2.4 (The system through the lens of $\mathbb{R}^{Nd}$). *We introduce the map V to interpret* (1.1) *as one system of equations in $\mathbb{R}^{Nd}$ instead of N dependent equations each in $\mathbb{R}^d$. Namely, we define for v given by* (1.4)*,*

$$V = (V_1, \cdots, V_N) : (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N \ \text{ where for } \ i \in [\![1,N]\!] \ \ V_i : (\mathbb{R}^d)^N \to \mathbb{R}^d, \quad V_i(X^N) = v(X^{i,N}, \mu^{X,N}), \tag{2.2}$$

and $X^N = (X^{1,N}, \cdots, X^{N,N}) \in \mathbb{R}^{Nd}$ where each $X^{i,N}$ solves (1.1)*, $\mu^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X^{j,N}}(\mathrm{d}x)$.*

For $X^N, Y^N \in \mathbb{R}^{Nd}$ with corresponding measure $\mu^{X,N}, \mu^{Y,N}$ and letting Assumption 2.1 hold, the function V also satisfies a one-sided Lipschitz condition

$$\begin{aligned}
&\langle X^N - Y^N, V(X^N) - V(Y^N)\rangle \\
&= \frac{1}{2N}\sum_{i=1}^N\sum_{j=1}^N \Big\langle (X^{i,N} - X^{j,N}) - (Y^{i,N} - Y^{j,N}), f(X^{i,N} - X^{j,N}) - f(Y^{i,N} - Y^{j,N})\Big\rangle \\
&\quad + \sum_{i=1}^N \Big\langle X^{i,N} - Y^{i,N}, u(X^{i,N}, \mu^{X,N}) - u(Y^{i,N}, \mu^{X,N}) + u(Y^{i,N}, \mu^{X,N}) - u(Y^{i,N}, \mu^{Y,N})\Big\rangle \\
&\quad \le (2L_f^+ + L_u + \frac{1}{2} + \frac{L_{\tilde{u}}}{2})|X^N - Y^N|^2, \qquad L_f^+ = \max\{0, L_f\}.
\end{aligned}$$

In the last second step we changed the order of summation and used that f is odd.

Propagation of chaos (PoC). In order to show that the particle approximation (1.1) is of effective use to approximate the MV-SDE (1.3), we provide a pathwise propagation of chaos result (convergence as the number of particles increases and with rate). We introduce the auxiliary system of non interacting particles

$$\mathrm{d}X_t^i = \big(v(X_t^i, \mu_t^{X^i}) + b(t, X_t^i, \mu_t^{X^i})\big)\mathrm{d}t + \sigma(t, X_t^i, \mu_t^{X^i})\mathrm{d}W_t^i, \quad X_0^i = X_0^i\,, \quad t \in [0,T]\,, \tag{2.3}$$

which are just (decoupled) MV-SDEs with i.i.d. initial conditions X_0^i. Since the X^is are independent, $\mu_t^{X^i} = \mu_t^X$ for all i (and μ_t^X the law of the solution to (1.3) with v given as (1.4)).

The Propagation of chaos result (2.5) follows from [1, Theorem 3.14] under the assumption that the interacting particle system (1.1) is well-posed. The first statement of Proposition 2.5 establishes the well-posedness of the particle system hence closing the small gap left in [1, Theorem 3.14].

PROPOSITION 2.5. *Let the assumptions of Theorem 2.3 hold for some $m > 2(q+1)$. Then, for all $i \in [\![1,N]\!]$ there exists a unique solution $X^{i,N}$ to* (1.1) *in $\mathbb{S}^m([0,T])$ and for any $1 \le p \le m$ there exists $C > 0$ independent of N (but depending on T and m) such that*

$$\sup_{t\in[0,T]}\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|X_t^{i,N}|^p\big] \le C\Big(1 + \mathbb{E}\big[\,|X_0^{\cdot}|^p\big]\Big). \tag{2.4}$$

This manuscript is for review purposes only.

For $i \in [\![1,N]\!]$, let $X^i \in \mathbb{S}^m([0,T])$ be the solution to (2.3), ensured by Theorem 2.3. Suppose additionally that $m > \max\{2(q+1),4\}$. Then, there exists a constant $C>0$ independent of N (but depending on T and m) such that

$$\sup_{i\in[\![1,N]\!]}\sup_{0\le t\le T}\mathbb{E}\big[|X_t^i - X_t^{i,N}|^2\big] \le C\begin{cases} N^{-1/2}, & d<4\\ N^{-1/2}\log N, & d=4\\ N^{\frac{-2}{d+4}}, & d>4\end{cases}. \tag{2.5}$$

The proof and further details are presented in Appendix A. This result shows that the particle scheme will converge to the MV-SDE with a given quantified rate. Therefore, to show convergence between our numerical scheme and the MV-SDE, we only need to show that the numerical version of the particle scheme converges to the "true" particle scheme in a way that is independent of N. We note that the PoC rate can be optimised for the case of constant diffusion [17, Remark 2.5].

2.3. The scheme for the interacting particle system and main results. The split-step method (SSM) here is inspired by that of [17] and re-cast accordingly to the setup here. The critical difficulty arises from the convolution component in v (1.3). This term is the main hindrance in proving moment bounds. Before continuing recall the definition of V in Remark 2.4. We now introduce the SSM numerical scheme.

DEFINITION 2.6 (Definition of the SSM). *Let Assumption 2.1 hold. Define the uniform partition of $[0,T]$ as $\pi := \{t_n := nh : n \in [\![0,M]\!], h := T/M\}$ for a prescribed $M \in \mathbb{N}\setminus\{0\}$. Define recursively the SSM approximating (1.1) as: set $\hat{X}_0^{i,N} = X_0^i$ for $i \in [\![1,N]\!]$; iteratively over $n \in [\![0,M-1]\!]$ for all $i \in [\![1,N]\!]$ (recall Remark 2.4 and the definition of the map V in (2.2))*

$$Y_n^{\star,N} = \hat{X}_n^N + hV(Y_n^{\star,N}),\quad \hat{X}_n^N = (\cdots,\hat{X}_n^{i,N},\cdots),\quad Y_n^{\star,N} = (\cdots, Y_n^{i,\star,N},\cdots), \tag{2.6}$$

$$\text{where } Y_n^{i,\star,N} = \hat{X}_n^{i,N} + hv(Y_n^{i,\star,N},\hat{\mu}_n^{Y,N}),\qquad \hat{\mu}_n^{Y,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{Y_n^{j,\star,N}}(\mathrm{d}x), \tag{2.7}$$

$$\hat{X}_{n+1}^{i,N} = Y_n^{i,\star,N} + b(t_n, Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})h + \sigma(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^i,\qquad \Delta W_n^i = W_{t_{n+1}}^i - W_{t_n}^i. \tag{2.8}$$

The stepsize h is chosen as to belong to the interval (this constraint is soft in the sense of Remark 2.7)

$$h \in \Big(0,\min\big\{1,\frac{1}{\zeta}\big\}\Big) \quad \text{for } \zeta \text{ defined as} \quad \zeta = \max\Big\{2(L_f+L_u)\ ,\ 4L_f^+ + 2L_u + 2L_{\tilde{u}} + 1\ ,\ 0\Big\}. \tag{2.9}$$

In some cases where the original functions f,u might cause trouble to find a suitable choice of h, and by the Remark below, we can use the addition and subtraction trick to bypass the constraint, see Remark 4.1 and [17, Section 3.4] for more discussion.

REMARK 2.7 (The constraint on h in (2.9) is soft). *Our framework allows to change f,u,b in such a way as to have $\zeta = 0$ in (2.9) via addition and subtraction of linear terms to f,u and b. Concretely, take $\theta,\gamma \in \mathbb{R}$ and redefine f,u,b into $\widehat{f},\widehat{u},\widehat{b}$ as follows: for any $t\in[0,\infty), x\in\mathbb{R}^d, \mu\in\mathcal{P}_2(\mathbb{R}^d)$*

$$\widehat{f}(x) = f(x) - \theta x,\qquad \widehat{u}(x,\mu) = u(x,\mu) - \gamma x - \theta\int_{\mathbb{R}^d} z\mu(\mathrm{d}z),\quad \text{and}\quad \widehat{b}(t,x,\mu) = b(t,x,\mu) + (\gamma+\theta)x.$$

For judicious choices of θ,γ it is easy to see that ζ can be set to be zero (we invite the reader to carry out the calculations). We remark that this operation increases the Lipschitz constant of $\widehat{b}$.

Recall that the function V satisfies a one-sided Lipschitz condition in $X\in\mathbb{R}^{Nd}$ (Remark 2.4), and hence (under (2.9)) a unique solution $Y_n^{\star,N}$ to (2.6) as a function of $\hat{X}_n^N$ exists (details in Lemma 4.2). After introducing the discrete scheme, we define its continuous extension and provide the main convergence results.

This manuscript is for review purposes only.

Definition 2.8 (Continuous extension of the SSM). *Under the same choice of h and assumptions in Definition 2.6, for all $t \in [t_n, t_{n+1}]$, $n \in [\![0, M-1]\!]$, $i \in [\![1, N]\!]$, $\hat{X}_0^{i,N} = X_0^i \in L_0^m(\mathbb{R}^d)$, the continuous extension of the SSM is*

$$\mathrm{d}\hat{X}_t^{i,N} = \big(v(Y_{\kappa(t)}^{i,\star,N}, \hat{\mu}_{\kappa(t)}^{Y,N}) + b(\kappa(t), Y_{\kappa(t)}^{i,\star,N}, \hat{\mu}_{\kappa(t)}^{Y,N})\big)\mathrm{d}t + \sigma(\kappa(t), Y_{\kappa(t)}^{i,\star,N}, \hat{\mu}_{\kappa(t)}^{Y,N})\mathrm{d}W_t^i, \tag{2.10}$$

$$\hat{\mu}_n^{Y,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{Y_n^{j,\star,N}}(\mathrm{d}x), \qquad \kappa(t) = \sup\big\{t_n : t_n \le t,\ n \in [\![0, M-1]\!]\big\}, \qquad \hat{\mu}_{t_n}^{Y,N} = \hat{\mu}_n^{Y,N}.$$

The next result states our first strong convergence finding. It is a "strong" pointwise (non-path-space) convergence result that is not in the classical mean-square error form.

Theorem 2.9 (Non-path-space mean-square convergence). *Let Assumption 2.1 hold and choose h as in (2.9). Let $i \in [\![1, N]\!]$, take $X^{i,N}$ as the solution to (1.1) and let $\hat{X}^{i,N}$ be the continuous-time extension of the SSM given by (2.10). If $m \ge 4q + 4 > \max\{2(q+1), 4\}$, where $X_0^i \in L_0^m(\mathbb{R}^d)$ and q is as defined in Assumption 2.1, then there exists a constant $C > 0$ independent of h, N, M (but depending on T and m) such that*

$$\sup_{i \in [\![1,N]\!]} \sup_{0 \le t \le T} \mathbb{E}\big[\,|X_t^{i,N} - \hat{X}_t^{i,N}|^2\big] \le Ch. \tag{2.11}$$

The proof is presented in Section 4.2. This result does not need L^p-moment bounds of the scheme for $p > 2$. It needs *only* L^p-moments of the solution process of (1.1) and L^2-moments for the scheme [9]. The proof takes advantage of the elegant structure induced by the SSM where Proposition 4.3 and 4.4 are the crucial intermediate results to deal with the convolution term.

The next moment bound result is necessary for the subsequent uniform convergence result.

Theorem 2.10 (Moment bounds). *Let the setting of Theorem 2.9 hold. Let $m \ge 2$ where $X_0^i \in L_0^m(\mathbb{R}^d)$ for all $i \in [\![1, N]\!]$ and let $\hat{X}^{i,N}$ be the continuous-time extension of the SSM given by (2.10). Let $2p \in [2, m]$, then there exists a constant $C > 0$ independent of h, N, M (but depending on T and m) such that*

$$\sup_{i \in [\![1,N]\!]} \sup_{0 \le t \le T} \mathbb{E}\big[\,|\hat{X}_t^{i,N}|^{2p}\big] \le C\big(1 + \mathbb{E}\big[\,|\hat{X}_0^{\cdot}|^{2p}\big]\big) < \infty. \tag{2.12}$$

The proof is presented in Section 4.3 and builds around auxiliary Theorem 4.7. There, we expand (4.35) and (4.36), and leverage the properties of the SSM scheme stated in Proposition 4.3 and 4.4 to deal with the difficult convolution terms.

Next we state the classic mean-square error convergence result.

Theorem 2.11 (Classical path-space mean-square convergence). *Let the setting of Theorem 2.9 hold. Assume there exists some $\varepsilon \in (0, 1)$ such that $m \ge \max\{4q + 4, 2 + q + q/\varepsilon\} > \max\{2(q+1), 4\}$ with $X_0^i \in L_0^m(\mathbb{R}^d)$ for $i \in [\![1, N]\!]$ and q given as in Assumption 2.1. Then there exists a constant $C > 0$ independent of h, N, M (but depending on T and m) such that*

$$\sup_{i \in [\![1,N]\!]} \mathbb{E}\big[\sup_{0 \le t \le T} |X_t^{i,N} - \hat{X}_t^{i,N}|^2\big] \le Ch^{1-\varepsilon}. \tag{2.13}$$

The proof is presented in Section 4.4. For this result we need both the L^p-moments of the scheme and solution process. This in contrast to the proof methodology of Theorem 2.9 and the reason we introduce Theorem 2.10 as a main result. The nearly optimal error rate of $(1 - \varepsilon)$ is a consequence of the estimation of (4.46) (product of three unbounded random variables). The expectation is taken after the supremum and then we use Theorem 2.9 and 2.10 – this forces an ε sacrifice of the rate. The nearly optimal error rate of $(1 - \varepsilon)$ is also the present best one available even for higher-order differences $p > 2$ (although we do not present these calculations). It is still open how to prove (2.12) with the $\sup_t$ inside the expectation — the difficulty to be overcome relates to establishing (4.3) of Proposition 4.4 under higher moments $p > 2$ in a way that aligns with *carré-du-champs* type arguments and the convolution term (within the style of proof we provide, otherwise new arguments need be found). It remains an open problem to show (2.13) when $\varepsilon = 0$.

This manuscript is for review purposes only.

A particular result for granular media equation type models. We recast the earlier results to granular media type models where the diffusion coefficient is constant and higher convergence rates can be established.

ASSUMPTION 2.12. *Consider the following MV-SDE*

$$\mathrm{d}X_t = v(X_t, \mu_t^X)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0 \in L_0^m(\mathbb{R}^d), \quad v(x,\mu) = \int_{\mathbb{R}^d} f(x-y)\mu(\mathrm{d}y). \tag{2.14}$$

Let $f:\mathbb{R}^d \to \mathbb{R}^d$ be continuously differentiable satisfying ($\mathbf{A}^f$) of Assumption 2.1. There exist $L_{f'},\ L_{f''} > 0$, $q \in \mathbb{N}$ and $q > 1$, with q the same as in ($\mathbf{A}^f$), such that for all $x,\ x' \in \mathbb{R}^d$

$$|\nabla f(x)| \leq L_{f'}(1+|x|^q), \quad |\nabla f(x) - \nabla f(x')| \leq L_{f''}(1+|x|^{q-1}+|x'|^{q-1})|x-x'|. \tag{2.15}$$

The function $\sigma : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times l}$ is a constant matrix.

In the language of the granular media equation, MV-SDE (2.15) corresponds to the Fokker-Plank PDE $\partial_t \rho = \nabla \cdot [\nabla \rho + \rho \nabla W * \rho]$ where $\nabla W = f$ and ρ is the probability measure [45]. We have the following results.

THEOREM 2.13. *Let Assumption 2.12 hold and choose h as in (2.9). Let $i \in [\![1,N]\!]$, take $X^{i,N}$ to be the solution to (1.1), let $\hat{X}^{i,N}$ be the continuous-time extension of the SSM given by (2.10) and $X_0^i \in L_0^m(\mathbb{R}^d)$. Let $m \geq \max\{8q,\ 4q+4\} > \max\{2(q+1), 4\}$ with q as defined in Assumption 2.12. Then there exist a constant $C>0$ independent of h, N, M (but depending on T and m) such that*

$$\sup_{i\in[\![1,N]\!]} \sup_{0\leq t\leq T} \mathbb{E}\big[|X_t^{i,N} - \hat{X}_t^{i,N}|^2\big] \leq Ch^2. \tag{2.16}$$

This result is proved in Section 4.5. Supporting simulation results are presented in Section 3.1 and confirm the strong root mean square error rate of 1.0.

We note that one can use a proof methodology similar to that used for Theorem 2.11 to obtain (2.16) with the $\sup_t$ inside the expectation. This would deliver a rate of $h^{2-\varepsilon}$, the key steps are similar to (4.47)-(4.48).

3. Examples of interest.

We illustrate the SSM on three numerical examples.[3] The "true" solution in each case is unknown and the convergence rates for these examples are calculated in reference to a proxy solution given by the approximating scheme at a smaller timestep h and higher number of particles N (particular details are given below). The strong error between the proxy-true solution X_T and approximation $\hat{X}_T$ is as follows

$$\text{root Mean-square error (Strong error)} = \Big(\mathbb{E}\big[\,|X_T - \hat{X}_T|^2\big]\Big)^{\frac12} \approx \Big(\frac{1}{N}\sum_{j=1}^N |X_T^j - \hat{X}_T^j|^2\Big)^{\frac12}.$$

We also consider the path strong error define as follows, for $Mh = T$, $t_n = nh$,

$$\text{Strong error (Path)} \ = \Big(\mathbb{E}\big[\sup_t |X_t - \hat{X}_t|^2\big]\Big)^{\frac12} \approx \Big(\frac{1}{N}\sum_{j=1}^N \sup_{n\in[\![0,M]\!]} |X_{t_n}^j - \hat{X}_{t_n}^j|^2\Big)^{\frac12}.$$

The propagation of chaos (PoC) rate between different particle systems $\{\hat{X}_T^{i,N_l}\}_{i,l}$ where i denotes the i-th particle and N_l denotes the size of the system,

$$\text{Propagation of chaos error (PoC error)} \approx \Big(\frac{1}{N_l}\sum_{j=1}^{N_l} |\hat{X}_T^{j,N_l} - X_T^j|^2\Big)^{\frac12}.$$

[3] Implementation code in Python is available in https://github.com/AnandaChen/Simulation-of-super-measure

This manuscript is for review purposes only.

REMARK 3.1 ('Taming' algorithm). *For comparative purposes we implement the 'Taming' algorithm [17, 21] – any convergence analysis of the taming algorithm to the framework of this manuscript is an open question. Of the many variants of Taming possible, set the terminal time T with $Mh = T$, we implement as follows: $\int_{\mathbb{R}^d} f(\cdot - y)\mu(\mathrm{d}y)$ is replaced by $\int_{\mathbb{R}^d} f(\cdot - y)\mu(\mathrm{d}y)/(1 + M^{-\alpha}|\int_{\mathbb{R}^d} f(\cdot - y)\mu(\mathrm{d}y)|)$, and u is replaced by $u/(1+M^{-\alpha}|u|)$ with the choice of $\alpha = 1/2$ for non-constant diffusion and $\alpha = 1$ for constant diffusion.*

Within each example, the error rates of Taming and SSM are computed using the same Brownian motion paths.

Moreover, for the simulation study below, we fix the algorithmic parameters as follows:

1. For the strong error, the proxy-true solution is calculated with $h = 10^{-4}$ and the approximations are calculated with $h \in \{10^{-3}, 2 \times 10^{-3}, \dots, 10^{-1}\}$ with $N = 1000$ at $T = 1$ and using the same Brownian motion paths. We compare SSM and Taming with the proxy-true solutions provided by the same algorithm (SSM and Taming) respectively.
2. For the PoC error, the proxy-true solution is calculated with $N = 2560$ and the approximations are calculated with $N \in \{40, 80, \dots, 1280\}$, with $h = 0.001$ at $T = 1$ and using the same Brownian motion paths.
3. The implicit step (2.6) of the SSM algorithm is solved, in our examples, via a Newton method iteration. We point the reader to Appendix B for a full discussion. In practice, 2 to 4 Newton iterations are sufficient to ensure that the difference between two consecutive Newton iterates are not larger than $\sqrt{h}$ in $\|\cdot\|_\infty$-norm (in $\mathbb{R}^{Nd}$).

Lastly, the symbols $\mathcal{N}(\alpha, \beta)$ denote the normal distribution with mean $\alpha \in \mathbb{R}$ and variance $\beta \in (0, \infty)$.

3.1. Example: the granular media equation. The first example is the granular media Fokker-Plank equation taking the form $\partial_t \rho = \nabla \cdot [\nabla \rho + \rho \nabla W * \rho]$ with $W(x) = \frac{1}{3}|x|^3$ and ρ is the correspondent probability density [15, 45]. In MV-SDE form we have

$$\mathrm{d}X_t = v(X_t, \mu_t^X)\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \quad X_0 \in L_0^m(\mathbb{R}^d), \quad v(x,\mu) = \int_{\mathbb{R}^d} \Big(-\operatorname{sign}(x-y)|x-y|^2 \Big)\mu(\mathrm{d}y), \tag{3.1}$$

where $\operatorname{sign}(\cdot)$ is the standard sign function, μ_t^X is the law of the solution process X at time t. This granular media model has been well studied in [15, 45] and is a reference model to showcase the numerical approximation. For this specific case, starting from a normal distribution, the particles concentrate and move around its initial mean value (also its steady state). In Figure 3.1 (a) and (b) one sees the evolution of the density map across time $T \in \{1, 3, 10\}$ for two initial initial distributions $\mathcal{N}(0,1)$ and $\mathcal{N}(2,4)$ respectively, and $h = 0.01$. For this case, both methods approximate well the solution without any apparent leading difference between Taming and SSM.

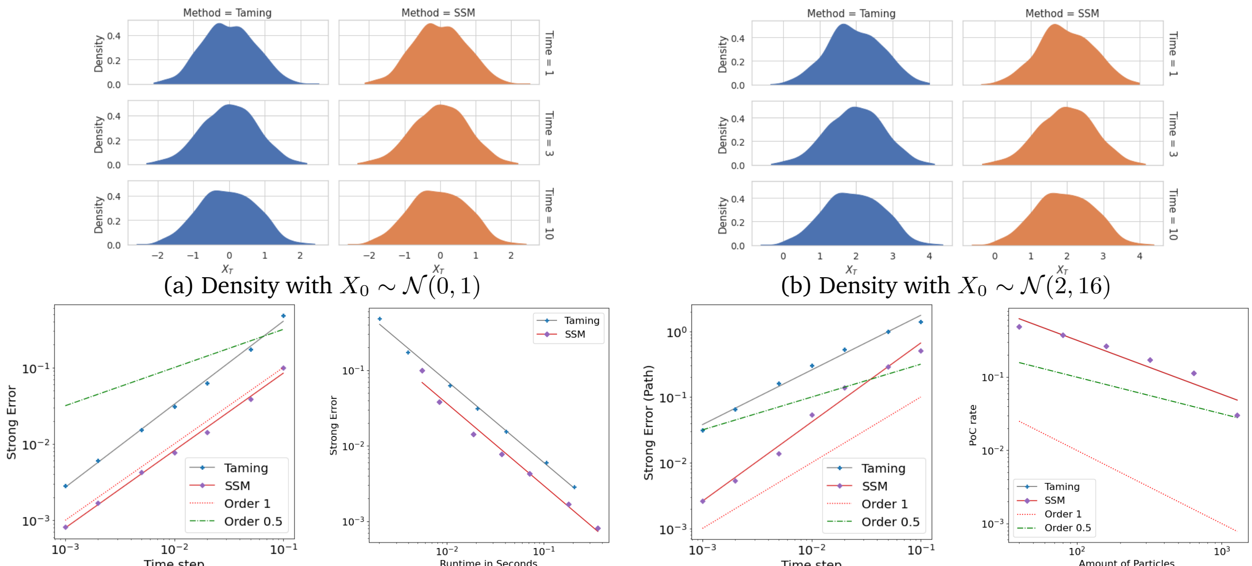
Figure 3.1 (c) shows strong error of both methods, computed at $T = 1$ across $h \in \{10^{-3}, 2\times 10^{-3}, \dots, 10^{-1}\}$. The proxy-true solution for each method is taken at $h = 10^{-4}$ and the baseline slopes for the "order 1" and "order 0.5" convergence rate are provided for comparison. The estimated rate of both method is 1.0 in accordance to Theorem 2.13 (under constant diffusion coefficient). Figure 3.1 (d) shows strong error v.s algorithm runtime of both methods under the same set up as in (c). The SSM perform slightly better than the Taming method.

Figure 3.1 (e) shows the path type strong error of both method, compare to the results in (c), the SSM preserve the error rate of near 1.0 and perform better than the Taming method. Figure 3.1 (f) shows the PoC error of both methods. The two results coincide since the differences between two methods are within 0.001. The PoC rates are near 0.5 which is better than the theoretical result of $1/4$ after we take square root in Proposition 2.5. This result is similar to [52, Example 4.1], and is explained theoretically by [20, Lemma 5.1] but under stronger assumptions than ours.

3.2. Example: Double-well model. We consider a limit model of particles under a symmetric double-well confinement. We test a variant of the model studied in [57] but change its diffusion coefficient to a non-constant one (in opposition to the previous example). Concretely, we study the following McKean-Vlasov equation

$$\mathrm{d}X_t = \big(v(X_t, \mu_t^X) + X_t\big)\mathrm{d}t + X_t \mathrm{d}W_t, \quad v(x,\mu) = -\frac{1}{4}x^3 + \int_{\mathbb{R}^d} -(x-y)^3 \mu(\mathrm{d}y). \tag{3.2}$$

This manuscript is for review purposes only.



(a) Density with $X_0 \sim \mathcal{N}(0, 1)$
(b) Density with $X_0 \sim \mathcal{N}(2, 16)$
(c) Strong Error
(d) Strong error v.s Runtime
(e) Strong error (Path)
(f) PoC Error

Figure 3.1: Simulation of the granular media equation (3.1) with $N = 1000$ particles. (a) and (b) show the density map for Taming (blue, left) and SSM (orange, right) with $h = 0.01$ at times $T = 1, 3, 10$ seen top-to-bottom and with different initial distribution. (c) Strong error (rMSE) of SSM and Taming with $X_0 \sim \mathcal{N}(2, 16)$. (d) Strong error (rMSE) of SSM and Taming w.r.t algorithm with $X_0 \sim \mathcal{N}(2, 16)$.(e) Strong error (Path) of SSM and Taming with $X_0 \sim \mathcal{N}(2, 16)$. (f) PoC error rate in N of SSM and Taming with $X_0 \sim \mathcal{N}(2, 9)$ with perfect overlap of errors.

The corresponding Fokker-Plank equation is $\partial_t \rho = \nabla \cdot [\ \nabla(\frac{\rho |x|^2}{2}) + \rho \nabla V + \rho \nabla W * \rho]$ with $W = \frac{1}{4}|x|^4$, $V = \frac{1}{16}|x|^4 - \frac{1}{2}|x|^2$, ρ is the corresponding density map. There are three stable states $\{-2, 0, 2\}$ for this model [57].

The example of Section 3.1 was a relatively mild with additive noise and where both methods performed well. For this double-well model of (3.2), the drift includes super-linear growth components in both space and measure and a non-constant unbounded diffusion coefficient.

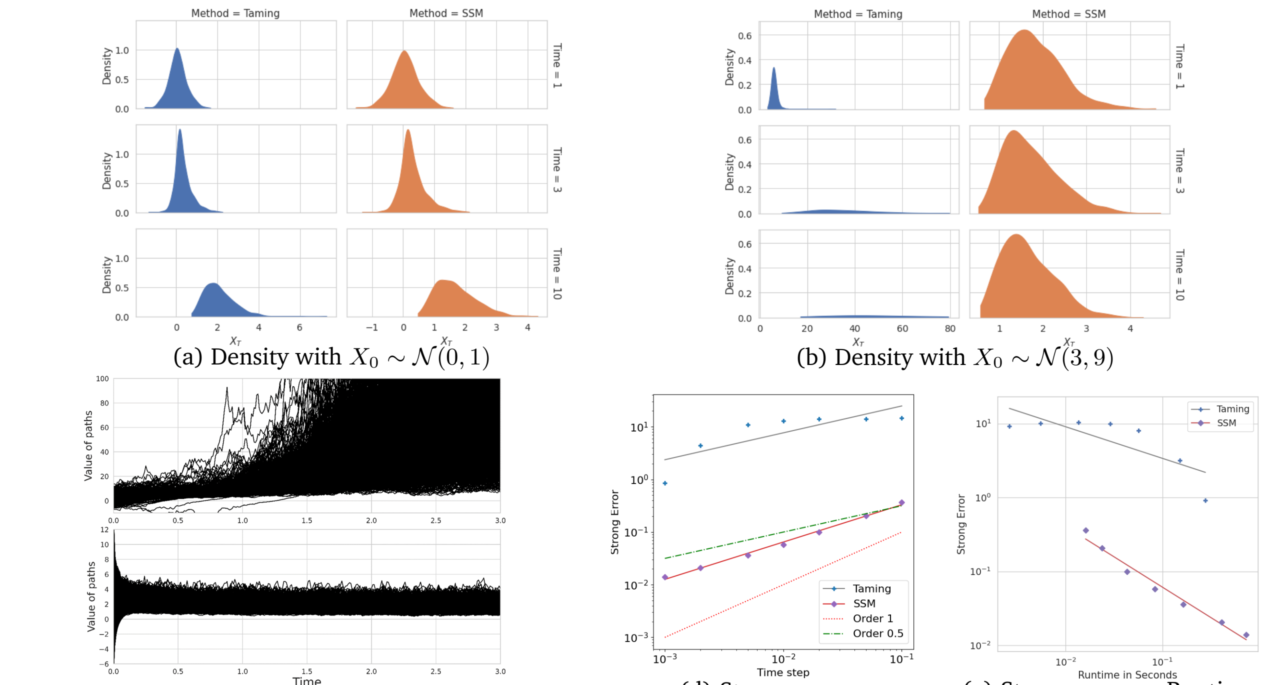
In Figure 3.2 (a) and (b), Taming (blue, left) fails to produce acceptable results of any type – Figure 3.2 (c) shows the simulated paths of both methods where it is noteworthy to see that Taming become unstable while the SSM paths remain stable. In respect to Figure 3.2 (a) and (b), the SSM (orange, right) depicts the distribution's evolution to one of the expected stable states ($x = 2$) as time evolves. It is interesting to find out that for the SSM in (a), where $X_0 \sim \mathcal{N}(0, 1)$, the particles shift from the zero (unstable) steady state to the positive stable steady state $x = 2$. However, in (b) with $X_0 \sim \mathcal{N}(3, 9)$, we find that the particles remain within the basin of attraction of the stable state $x = 2$. Figure 3.2 (d) displays under the same parameter choice for $h,\ T$ as for the granular media example of Section 3.1 with $X_0 \sim \mathcal{N}(2, 4)$ the estimated rate of convergence for the schemes. It shows the taming method fails to converge (but does not explode). The strong error rate of the SSM is the expected $1/2$ in-line with Theorem 2.9 (and Theorem 2.11).

The "order 1" and "order 0.5" lines are baselines corresponding to the slope of 1 and 0.5 rate of convergence.

Figure 3.2 (e) shows that, to reach the same strong error level Taming shall takes far more (over 100 times) runtime than the SSM.

3.3. Example: 2d Van der Pol (VdP) oscillator. We consider the Van der Pol (VdP) model described in [35, Section 4.2 and 4.3], with added super-linearity in measure and non-constant unbounded diffusion. We study

This manuscript is for review purposes only.



(a) Density with $X_0 \sim \mathcal{N}(0,1)$ (b) Density with $X_0 \sim \mathcal{N}(3,9)$

(c) Simulated paths of Taming (top) and SSM (bottom) (d) Strong error (e) Strong error v.s Runtime

Figure 3.2: Simulation of the Double-Well model (3.2) with $N = 1000$ particles. (a) and (b) show the density map for Taming (blue, left) and SSM (orange, right) with $h = 0.01$ at times $T = 1, 3, 10$ seen top-to-bottom and with different initial distribution. (c) simulated paths by Taming (top) and SSM (bottom) with $h = 0.01$ over $t \in [0, 3]$ and with $X_0 \sim \mathcal{N}(3, 9)$. (d) Strong error (rMSE) of SSM and Taming with $X_0 \sim \mathcal{N}(2, 4)$. (e) Strong error (rMSE) of SSM and Taming w.r.t algorithm Runtime with $X_0 \sim \mathcal{N}(2, 4)$.

the following MV-SDE dynamics: set $x = (x_1, x_2) \in \mathbb{R}^2$, for (1.3) define the functions f, u, b, σ as

$$(3.3) \qquad f(x) = -x|x|^2, \quad u(x) = \begin{bmatrix} -\frac{4}{3}x_1^3 \\ 0 \end{bmatrix}, \quad b(x) = \begin{bmatrix} 4(x_1 - x_2) \\ \frac{1}{4}x_1 \end{bmatrix}, \quad \sigma(x) = \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix},$$

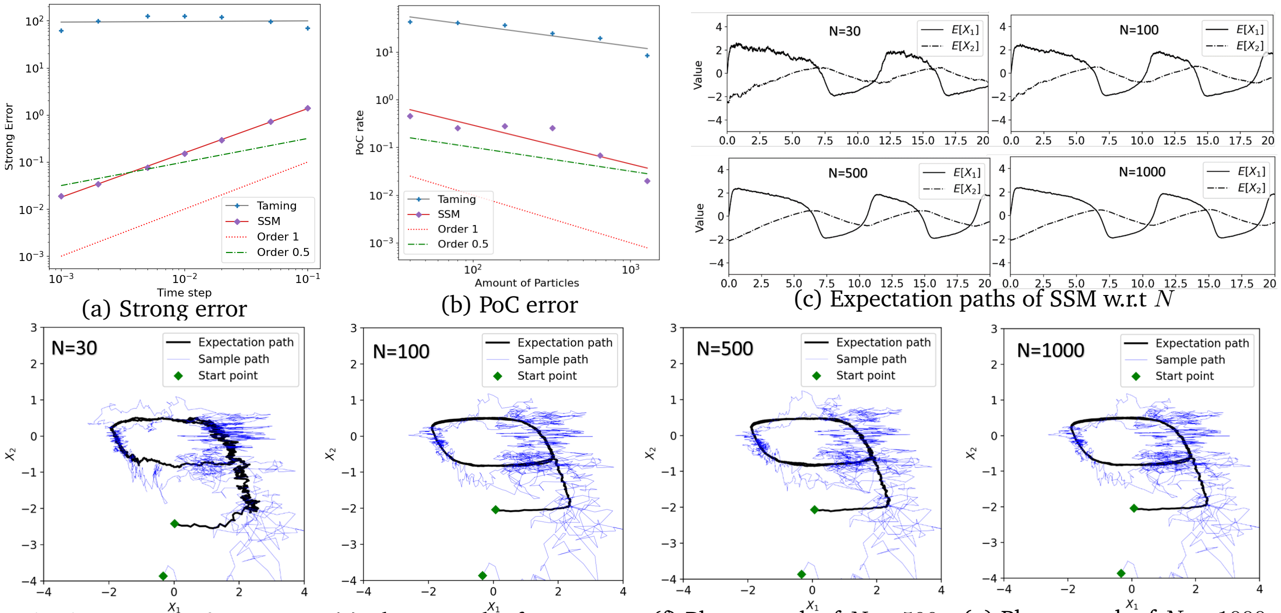
which satisfy the assumptions of this work.

Figure 3.3 (a) shows the strong error of both methods, the "order 1" and "order 0.5" lines are baselines with the slope of 1 and 0.5 for comparison. The estimated rate of the SSM is near 0.5 while Taming failed to converge. Figure 3.3 (b) shows the PoC error of both methods, Taming failed to converge while the estimated rate of the SSM is near 0.5 (see discussion of previous Section 3.1).

Figure 3.3 (c) shows the system's phase-space portraits (i.e., the parametric plot of $t \mapsto (X_{1,t}, X_{2,t})$ and $t \mapsto (\mathbb{E}[X_{1,t}], \mathbb{E}[X_{2,t}])$ over $t \in [0, 20]$) of the SSM with respect to different choices of $N \in \{30, 100, 500, 1000\}$. The impact of N on the quality of simulation is apparent as is the ability of the SSM to capture the periodic behaviour of the true dynamics. Figure 3.3 (d)-(e)-(f)-(g) shows the expectation's fluctuation (of Figure 3.3 (c)) and the system's phase-space path portraits of the SSM for different choices of N. The trajectory becomes smoother as N becomes larger and the paths are similar for $N \geq 500$.

3.4. Numerical complexity, discussion and various opens questions. Across the three examples the SSM converged and all examples recovered the theoretical convergence rate (of $1/2$ in general, and 1 for the additive noise case). In the latter two examples, Taming failed to converge while on the first example the SSM and taming are mostly similar. The main difference between examples is the diffusion coefficient.

This manuscript is for review purposes only.



(a) Strong error (b) PoC error (c) Expectation paths of SSM w.r.t N

(d) Phase graph of $N = 30$ (e) Phase graph of $N = 100$ (f) Phase graph of $N = 500$ (g) Phase graph of $N = 1000$

Figure 3.3: Simulation of the Vdp model (3.3) with $X_1 \sim \mathcal{N}(0, 4), X_2 \sim \mathcal{N}(-2, 4)$. (a) Strong error (rMSE) of the SSM and Taming with $T = 1,\ N = 1000$. (b) PoC error of the SSM and Taming with $T = 1,\ h = 0.001$. (c) the expectation overlays paths for the SSM with $T = 20, h = 0.01$ w.r.t different N. (d)-(e)-(f)-(g) the corresponding phase-space portraits in (c) with $N \in \{30, 100, 500, 1000\}$.

The SSM is robust in respect to small choices of h and N. In all three examples, the SSM remains convergent for all choices of h (even for $h = 0.1$) while taming fails to converge at all. In the Van der Pol (VdP) oscillator example of Section 3.3, when comparing across different particle sizes N, the SSM provides a good approximation for all choices of N (even for $N = 30$) and the PoC result is as expected. In general, we found that the runtime of the SSM is nearly the double of Taming for the same choices of h, but on the other hand, Taming takes over 100-times more runtime to reach the same accuracy as the SSM (if one considers the strong error against runtime).

Computational costs and open questions for future research. In the context of (1.1), assume one wants to simulate an N-particle system over a discretised finite time-domain with M time points. Since we deal with convolution type operator, the interaction term need to be computed for every single particle and thus, a standard explicit Euler scheme incurs a computational cost of $\mathcal{O}(N^2M)$. Without the convolution component, the cost is simply $\mathcal{O}(NM)$. For the SSM scheme in Definition 2.6, since it is has an implicit component there is an additional cost attached to it (more below).

At this level, two strategies can be thought to reduce the complexity. The first is by controlling the cost of computing the interaction itself, these have been proposed for example in the projected particle method [8] or the Random Batch Method (RBM) [37]. To date there is no general proof of these outside Lipschitz conditions (and constant diffusion coefficient in the RBM case) for the efficacy of the method, also, it is not clear how to use these methods in combination with Newton to solve the SSM's implicit equation (more below). The second is to better address the competition between the number of particles N, as dictated by the PoC result Proposition 2.5, and the time-step parameter M (or $1/h$). Our experimental work estimating the Propagation of chaos rate points to a convergence rate of order $1/2$ instead of the upper bound rate $1/4$ guaranteed by (2.5) in Theorem 2.5. This result is not surprising in view of the theoretical result [20, Lemma 5.1]; and numerically in [52, Example 4.1]. To the best of our knowledge, no known PoC rate result covers the examples presented here and Theorem 2.5 is

This manuscript is for review purposes only.

presently the best known general result.

Solving the implicit step in SSM - Newton's method. The SSM scheme contains an implicit Equation (2.6) that needs be solved at each timestep. It is left to the user to choose the most suitable method for given data and, in all generality, one needs an approximation scheme to solve (2.6). Proposition B.2 below shows that as long as said approximation is uniformly controlled within a ball of radius Ch of the true solution, then the SSM's convergence rate of Theorem 2.9 is preserved.

As mentioned in the initial part of Section 3, we use Newton's method (assuming extra differentiability of the involved maps) – see Appendix B for details where [54, Section 4.3] is used to guarantee convergence. The computation cost raises from $\mathcal{O}(N^2M)$ to $\mathcal{O}(\kappa N^2 M)$, where κ denotes the leading term cost of Newton after κ iterations. In practice, we found that within 2 to 4 iterations (i.e., $\kappa \leq 4$) two consecutive Newton iteration are sufficiently close for the purposes of the scheme's accuracy: denoting Newton's j^{th}-iteration by $y^j \in \mathbb{R}^{Nd}$, then $\|y^\kappa - y^{\kappa-1}\|_\infty < \sqrt{h}$ (which is the stop criteria used, see Appendix B).

Interacting particle systems like (1.1) induce a certain structure to the associated Jacobian matrix when seen through the lens of $(\mathbb{R}^d)^N$. The closed form expressions provided in Appendix B.2 point to a very sparse Jacobian matrix with a very specific block structure. For instance, the Γ matrix (see Appendix B.2) is a symmetric one and is multiplied by h/N making its entries very small: it stands to reason that Γ can be removed from the Jacobian matrix as one solves the system (provided its entries can be controlled) and thus suggests that an inexact or quasi-Newton method might be computationally more efficient. In [42, Section 3] the authors review [53] who address the case of using inexact Newton methods when the equation of interest (2.6) is a monotone map, which is indeed our case. The usage of Newton method is not a primary element of discussion and, as does [42], we point the reader to the comprehensive review [49] on practical quasi-Newton methods for nonlinear equations. In conclusion, it remains to explore how different versions of Newton method for sparse systems can be used as way to reduce its computational cost but, in light of our study, we found Newton method very fast and efficient even comparatively with the Explicit Euler taming method in Section 3.1.

4. Proof of split-step method (SSM) for MV-SDEs and interacting particle systems: convergence and stability.

The proof appearing in Section 4.2 depends in no way on Theorem 2.10 or its proof (in Section 4.3). Nonetheless, Section 4.3 has a strong complementary effect to fully understanding the proof in Section 4.2.

4.1. Some properties of the scheme.

Recall the SSM scheme of Definition 2.6. In this section we clarify further the choice of h and then introduce two critical results arising from the SSM's structure. Note that throughout $C > 0$ is a constant always independent of h, N, M.

REMARK 4.1 (Choice of h). *Let Assumption 2.1 hold, the constraint on h in (2.9) comes from (4.2), (4.3) and (4.19) below, where $L_f, L_u \in \mathbb{R}$ and $L_{\tilde{u}} \geq 0$. Following the notation of those inequalities, under (2.9) for $\zeta > 0$, there exists $\xi \in (0,1)$ such that $h < \xi/\zeta$ and*

$$\max\left\{\frac{1}{1-2(L_f+L_u)h}, \frac{1}{1-(4L_f^+ + 2L_u + 2L_{\tilde{u}} + 1)h}, \frac{1}{1-(4L_f^+ + 2L_u + L_{\tilde{u}} + 1)h}\right\} < \frac{1}{1-\xi}.$$

For $\zeta = 0$, the result is trivial and we conclude that there exist constants C_1, C_2 independent of h

$$\max\left\{\frac{1}{1-2(L_f+L_u)h}, \frac{1}{1-(4L_f^+ + 2L_u + 2L_{\tilde{u}} + 1)h}, \frac{1}{1-(4L_f^+ + 2L_u + L_{\tilde{u}} + 1)h}\right\} \leq C_1 \leq 1 + C_2 h.$$

As argued in Remark 2.7 the constraint on h can be lifted.

LEMMA 4.2. *Choose h as in (2.9). Then, given any $X \in \mathbb{R}^{Nd}$ there exists a unique solution $Y \in \mathbb{R}^{Nd}$ to*

$$Y = X + hV(Y). \tag{4.1}$$

The solution Y is a measurable map of X.

This manuscript is for review purposes only.

Proof. Recall Remark 2.4. The proof is an adaptation of the proof [17, Lemma 4.1] to the $\mathbb{R}^{Nd}$ case.

PROPOSITION 4.3 (Differences relationship). *Let Assumption 2.1 hold and choose h as in (2.9). For any $n \in [\![0, M]\!]$ and $Y_n^{*,N}$ in (2.6), there exists some constant $C > 0$ such that for all $i,\ j \in [\![1, N]\!]$,*

$$|Y_n^{i,\star,N} - Y_n^{j,\star,N}|^2 \le |\hat{X}_n^{i,N} - \hat{X}_n^{j,N}|^2 \frac{1}{1 - 2(L_f + L_u)h} \le (1 + Ch)|\hat{X}_n^{i,N} - \hat{X}_n^{j,N}|^2. \tag{4.2}$$

Proof. Take $n \in [\![0, M]\!]$, $i,\ j \in [\![1, N]\!]$. Using Remark 2.2 and Young's inequality we have

$$\begin{aligned}
&|Y_n^{i,\star,N} - Y_n^{j,\star,N}|^2 \\
&= \left\langle Y_n^{i,\star,N} - Y_n^{j,\star,N}, \hat{X}_n^{i,N} - \hat{X}_n^{j,N} \right\rangle + \left\langle Y_n^{i,\star,N} - Y_n^{j,\star,N}, v\left(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}\right) - v\left(Y_n^{j,\star,N}, \hat{\mu}_n^{Y,N}\right) \right\rangle h \\
&\le \frac{1}{2}|Y_n^{i,\star,N} - Y_n^{j,\star,N}|^2 + \frac{1}{2}|\hat{X}_n^{i,N} - \hat{X}_n^{j,N}|^2 + (L_f + L_u)|Y_n^{i,\star,N} - Y_n^{j,\star,N}|^2 h.
\end{aligned}$$

The argument regarding the uniformity of the constant C in regards to the parameters h, N, M follows from Remark 4.1. ∎

PROPOSITION 4.4 (Summation relationship). *Let Assumption 2.1 hold. Choose h as in (2.9). For the process in (2.7) there exists a constant $C > 0$ (independent of h, N, M) such that, for all $i \in [\![1, N]\!],\ n \in [\![0, M]\!]$,*

$$\frac{1}{N}\sum_{i=1}^{N}|Y_n^{i,\star,N}|^2 \le Ch + (1 + Ch)\frac{1}{N}\sum_{i=1}^{N}|\hat{X}_n^{i,N}|^2. \tag{4.3}$$

Proof. From (2.8) we have

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}|Y_n^{i,\star,N}|^2 &= \frac{1}{N}\sum_{i=1}^{N}\left\{\left\langle Y_n^{i,\star,N}, \hat{X}_n^{i,N}\right\rangle + \left\langle Y_n^{i,\star,N}, v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\right\rangle h\right\} \\
&\le \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{2}|Y_n^{i,\star,N}|^2 + \frac{1}{2}|\hat{X}_n^{i,N}|^2 + \left\langle Y_n^{i,\star,N}, u(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\right\rangle h + \frac{h}{N}\sum_{j=1}^{N}\left\langle Y_n^{i,\star,N}, f(Y_n^{i,\star,N} - Y_n^{j,\star,N})\right\rangle\right\}.
\end{aligned} \tag{4.4}$$

By Assumption 2.1 and Young's inequality, we have

$$\begin{aligned}
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\langle Y_n^{i,\star,N}, f(Y_n^{i,\star,N} - Y_n^{j,\star,N})\right\rangle &= \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\langle Y_n^{i,\star,N} - Y_n^{j,\star,N}, f(Y_n^{i,\star,N} - Y_n^{j,\star,N})\right\rangle \\
&\le \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} L_f|Y_n^{i,\star,N} - Y_n^{j,\star,N}|^2 \le \frac{2L_f^+}{N}\sum_{i=1}^{N}|Y_n^{i,\star,N}|^2, \quad L_f^+ = \max\{L_f, 0\}.
\end{aligned}$$

Plugging this into (4.4) and using Remark 2.2 with $\Lambda = 4L_f^+ + 2L_u + 2L_{\tilde{u}} + 1$, we have

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}|Y_n^{i,\star,N}|^2 &\le \frac{1}{N}\sum_{i=1}^{N}\left\{|\hat{X}_n^{i,N}|^2 + 2h\big(2L_f^+|Y_n^{i,\star,N}|^2 + C_u + \widehat{L}_u|Y_n^{i,\star,N}|^2 + L_{\tilde{u}}W^{(2)}(\hat{\mu}_n^{Y,N}, \delta_0)^2\big)\right\} \\
&\le \frac{1}{N}\sum_{i=1}^{N}\left\{|\hat{X}_n^{i,N}|^2 + 2h\big(2L_f^+|Y_n^{i,\star,N}|^2 + C_u + \widehat{L}_u|Y_n^{i,\star,N}|^2 + \frac{L_{\tilde{u}}}{N}\sum_{j=1}^{N}|Y_n^{j,\star,N}|^2\big)\right\} \\
&\le \frac{1}{1 - \Lambda h}\frac{1}{N}\sum_{i=1}^{N}\left\{|\hat{X}_n^{i,N}|^2 + 2C_u h\right\} = \frac{1}{N}\sum_{i=1}^{N}\left\{|\hat{X}_n^{i,N}|^2(1 + h\frac{\Lambda}{1 - \Lambda h}) + \frac{2C_u h}{1 - \Lambda h}\right\}.
\end{aligned}$$

Remark 4.1 yields the argument. ∎

This manuscript is for review purposes only.

From Lemma 4.2 we know a unique solution, $Y_n^{\star,N}$, to (2.6) as a function of $\hat{X}_n^N$ exists. We next show that the scheme we proposed in (2.6)-(2.8) is square integrable.

PROPOSITION 4.5 (Second moment bounds of SSM). *Let the setting of Theorem 2.9 hold. Let $m \geq 2$ where $\hat{X}_0^{i,N} \in L_0^m(\mathbb{R}^d)$ for all $i \in [\![1,N]\!]$, then there exists a constant $C>0$ independent of h,N,M (but depending on T) such that*

$$\sup_{i\in[\![1,N]\!]}\sup_{n\in[\![0,M]\!]}\mathbb{E}\big[|\hat{X}_n^{i,N}|^2\big]+\sup_{i\in[\![1,N]\!]}\sup_{n\in[\![0,M-1]\!]}\mathbb{E}\big[|Y_n^{i,\star,N}|^2\big]\leq C\big(1+\mathbb{E}\big[\,|\hat{X}_0^{\cdot,N}|^2\big]\big)<\infty.$$

Proof. Let $i \in [\![1,N]\!]$, $n \in [\![0,M-1]\!]$, by Assumption 2.1, from (2.6)-(2.8) and Proposition 4.4, since the particles are identically distributed, we have

$$\mathbb{E}\big[1+|Y_n^{i,\star,N}|^2|\big]\leq\mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big](1+Ch).$$

Similar to [17, Proposition 4.5], we have

$$|\hat{X}_{n+1}^{i,N}|^2\leq|\hat{X}_n^{i,N}|^2+C\Big(1+|Y_n^{i,\star,N}|^2+\frac{1}{N}\sum_{j=1}^N|Y_n^{j,\star,N}|^2\Big)(h+|\Delta W_n^i|^2)+2\Big\langle Y_n^{i,\star,N},\sigma(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^i\Big\rangle.$$

Taking expectations and summing 1 to both sides, Young's inequality yields

$$\mathbb{E}\big[1+|\hat{X}_{n+1}^{i,N}|^2\big]\leq\mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big](1+Ch).$$

By induction and using that the particles are identically distributed, we conclude that

$$\sup_{i\in[\![1,N]\!]}\sup_{n\in[\![0,M]\!]}\mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big]\leq\sup_{i\in[\![1,N]\!]}\mathbb{E}\big[1+|\hat{X}_0^{i,N}|^2\big](1+Ch)^M\leq(1+\mathbb{E}\big[|\hat{X}_0^{\cdot,N}|^2\big])e^{CT}<\infty, \tag{4.5}$$

where we used $Mh=T$ and that the $\{\hat{X}_0^{i,N}\}_i$ are i.i.d. The inequality for $\sup_{i\in[\![1,N]\!]}\sup_{n\in[\![0,M-1]\!]}\mathbb{E}\big[|Y_n^{i,\star,N}|^2\big]$ follows using similar argument. ∎

We provide the following auxiliary proposition to deal with the cross products terms in the later proofs.

PROPOSITION 4.6. *Take $N\in\mathbb{N}$, for all $i\in[\![1,N]\!]$, for any given $p\in\mathbb{N}$, sequences $\big\{\{a_i\}_i:\sum_{i=1}^N a_i=p,\ a_i\in\mathbb{N}\big\}$ and any collection of identically distributed L^p-integrable random variables $\{X_i\}_i$ we have*

$$\mathbb{E}\big[\prod_{i=1}^N|X_i|^{a_i}\big]\leq\mathbb{E}\big[|X_1|^p\big].$$

Proof. Using the notation above, by Young's inequality, for any $i,j\in[\![1,N]\!]$ we have

$$|X_i|^{a_i}|X_j|^{a_j}\leq\frac{a_i}{a_i+a_j}|X_i|^{a_i+a_j}+\frac{a_j}{a_i+a_j}|X_j|^{a_i+a_j}.$$

Thus, by induction and using that the $\{X_i\}_i$ are identically distributed, the result follows. ∎

4.2. Proof of Theorem 2.9: the pointwise mean-square convergence result. We provide here the proof of Theorem 2.9. Throughout this section, we follow the notation introduced in Theorem 2.9 and let Assumption 2.1 hold, h is chosen as in (2.9), $m\geq 4q+4$, where m is defined in (1.3) and q is defined in Assumption 2.1. Note that throughout $C>0$ is a constant always independent of h,N,M but possibly depending on T and m.

This manuscript is for review purposes only.

Proof. Let $i \in [\![1, N]\!]$, $n \in [\![0, M-1]\!]$, $s \in [0,h]$, $t_n = nh$ and $p \geq 2$ with $m \geq 4q+4$, using same notation as in (1.1), define the following auxiliary process

$$X_n^{i,N} = X_{t_n}^{i,N}, \quad \Delta X_{t_n+s}^i = X_{t_n+s}^{i,N} - \hat{X}_{t_n+s}^{i,N}, \quad t_n = nh, \quad \Delta W_{n,s}^i = W_{t_n+s}^i - W_{t_n}^i,$$

$$Y_n^{i,X,N} = X_n^{i,N} + hv(Y_n^{i,X,N}, \mu_n^{Y,X,N}), \qquad \mu_n^{Y,X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{Y_n^{j,X,N}}(\mathrm{d}x).$$

For all $n \in [\![0, M-1]\!],\ i \in [\![1, N]\!],\ r \in [0,h]$, from (2.10), we have

$$\begin{aligned}
|\Delta X_{t_n+r}^i|^2 = \Big|\Delta X_{t_n}^i &+ \int_{t_n}^{t_n+r} \big(v(X_s^{i,N}, \mu_s^{X,N}) - v(Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}s \\
&+ \int_{t_n}^{t_n+r} \big(v(Y_n^{i,X,N}, \mu_n^{Y,X,N}) - v\left(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}\right)\big)\mathrm{d}s + \int_{t_n}^{t_n+r} \big(b(s, X_s^{i,N}, \mu_s^{X,N}) - b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}s \\
&+ \int_{t_n}^{t_n+r} \big(b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big)\mathrm{d}s \\
&+ \int_{t_n}^{t_n+r} \big(\sigma(s, X_s^{i,N}, \mu_s^{X,N}) - \sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}W_s^i \\
&+ \int_{t_n}^{t_n+r} \big(\sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big)\mathrm{d}W_s^i\Big|^2.
\end{aligned}$$

Taking expectations on both side, using Jensen's inequality and Itô's isometry, we have

$$\mathbb{E}\big[|\Delta X_{t_n+r}^i|^2\big] \leq (1+h)I_1 + (1+\frac{1}{h})I_2 + 2I_3 + 2I_4, \tag{4.6}$$

where the terms I_1, I_2, I_3, I_4 are defines as follows

$$I_1 = \mathbb{E}\Big[\Big|\Delta X_{t_n}^i + \int_{t_n}^{t_n+r} \big(v(Y_n^{i,X,N}, \mu_n^{Y,X,N}) - v\left(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}\right)\big)\mathrm{d}s \tag{4.7}$$

$$+ \int_{t_n}^{t_n+r} \big(b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big)\mathrm{d}s\Big|^2\Big], \tag{4.8}$$

$$I_2 = \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r} \big(v(X_s^{i,N}, \mu_s^{X,N}) - v(Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}s \tag{4.9}$$

$$+ \int_{t_n}^{t_n+r} \big(b(s, X_s^{i,N}, \mu_s^{X,N}) - b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}s\Big|^2\Big], \tag{4.10}$$

$$I_3 = \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r} \big(\sigma(s, X_s^{i,N}, \mu_s^{X,N}) - \sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}W_s^i\Big|^2\Big], \tag{4.11}$$

$$I_4 = \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r} \big(\sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big)\mathrm{d}W_s^i\Big|^2\Big]. \tag{4.12}$$

For I_1, Young's inequality yields

$$\begin{aligned}
I_1 &= \mathbb{E}\Big[\Big|X_n^{i,N} + \big(V_n^{Y,i} + b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)r - \hat{X}_n^{i,N} - \big(V_n^{*,i} + b(t_n, Y_n^{i,\star,,N}, \hat{\mu}_n^{Y,N})\big)r\Big|^2\Big] \\
&\leq \mathbb{E}\Big[\Big|X_n^{i,N} - \hat{X}_n^{i,N} + \big(V_n^{Y,i} - V_n^{*,i}\big)r\Big|^2\Big](1+\frac{h}{2}) + \mathbb{E}\Big[\Big|b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - b(t_n, Y_n^{i,\star,,N}, \hat{\mu}_n^{Y,N})\Big|^2\Big](\frac{h}{2}+h),
\end{aligned} \tag{4.13}$$

This manuscript is for review purposes only.

where $V_n^{Y,i}$ and $V_n^{*,i}$ stand for $V_n^{Y,i} = v(Y_n^{i,X,N}, \mu_n^{Y,X,N})$ and $V_n^{*,i} = v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})$ respectively.

For the first term of (4.13), recall the SSM defined in (2.7). We have

$$
\begin{aligned}
&\mathbb{E}\Big[\Big|X_n^{i,N} - \hat{X}_n^{i,N} + \big(V_n^{Y,i} - V_n^{*,i}\big)r\Big|^2\Big] \\
&\quad= \mathbb{E}\Big[\Big\langle X_n^{i,N} - \hat{X}_n^{i,N} + \big(V_n^{Y,i} - V_n^{*,i}\big)r, Y_n^{i,X,N} - Y_n^{i,\star,N} + \big(V_n^{Y,i} - V_n^{*,i}\big)(r-h)\Big\rangle\Big] \\
&\quad= \mathbb{E}\Big[\Big\langle X_n^{i,N} - \hat{X}_n^{i,N}, Y_n^{i,X,N} - Y_n^{i,\star,N}\Big\rangle\Big] + \mathbb{E}\Big[\Big\langle X_n^{i,N} - \hat{X}_n^{i,N}, \big(V_n^{Y,i} - V_n^{*,i}\big)\Big\rangle\Big](r-h) \\
&\qquad + \mathbb{E}\Big[\Big\langle Y_n^{i,X,N} - Y_n^{i,\star,N}, \big(V_n^{Y,i} - V_n^{*,i}\big)\Big\rangle\Big]r - r(h-r)\mathbb{E}\Big[\Big|V_n^{Y,i} - V_n^{*,i}\Big|^2\Big].
\end{aligned}
$$

Using the relationship that (2.7) induces, we have

$$
V_n^{Y,i} - V_n^{*,i} = \frac{Y_n^{i,X,N} - X_n^{i,N} + Y_n^{i,\star,N} - \hat{X}_n^{i,N}}{h}.
$$

We first deduce that

$$
\begin{aligned}
\mathbb{E}\Big[\Big|X_n^{i,N} - \hat{X}_n^{i,N} + \big(V_n^{Y,i} - V_n^{*,i}\big)r\Big|^2\Big] &= \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big] + \mathbb{E}\Big[\Big\langle X_n^{i,N} - \hat{X}_n^{i,N}, V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]2r \\
&\quad + \mathbb{E}\Big[\Big\langle (Y_n^{i,X,N} - Y_n^{i,\star,N}) - (X_n^{i,N} - \hat{X}_n^{i,N}), V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]\frac{r^2}{h} \\
&= \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big](1 - C_{h,r}) + \mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big]C_{h,r} + \mathbb{E}\Big[\Big\langle Y_n^{i,X,N} - Y_n^{i,\star,N}, V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]\frac{r^2}{h}. \qquad (4.14)
\end{aligned}
$$

Where $C_{h,r} = (2hr - r^2)/2h$. Also, for the second term of (4.13), using Assumption 2.1 and that the particles are identically distributed

$$
\begin{aligned}
&\mathbb{E}\Big[\Big|b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Big|^2\Big] \\
&\quad\le C\mathbb{E}\Big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2 + W^{(2)}(\mu_n^{Y,X,N}, \hat{\mu}_n^{Y,N})\Big] \\
&\quad\le C\mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big] + C\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N |Y_n^{j,X,N} - Y_n^{j,\star,N}|^2\Big] \le C\mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big]. \qquad (4.15)
\end{aligned}
$$

By Assumption 2.1 and using Young's inequality once again

$$
\begin{aligned}
\mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big] &\le \mathbb{E}\Big[\Big\langle Y_n^{i,X,N} - Y_n^{i,\star,N}, X_n^{i,N} - \hat{X}_n^{i,N} + V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]h \qquad (4.16) \\
&\le \mathbb{E}\Big[\frac{1}{2}|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2 + \frac{1}{2}|X_n^{i,N} - \hat{X}_n^{i,N}|^2\Big] + \mathbb{E}\Big[\Big\langle Y_n^{i,X,N} - Y_n^{i,\star,N}, V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]h. \qquad (4.17)
\end{aligned}
$$

For the last term (4.17), since the particles are identically distributed, Assumption 2.1 and Remark 2.4 yield

$$
\begin{aligned}
\mathbb{E}\Big[\Big\langle Y_n^{i,X,N} - Y_n^{i,\star,N}, V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big] &\le \mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N \Big\langle Y_n^{j,X,N} - Y_n^{j,\star,N}, V_n^{Y,j} - V_n^{*,j}\Big\rangle\Big] \\
&\le \Big(2L_f^+ + L_u + \frac{1}{2} + \frac{L_{\tilde{u}}}{2}\Big)\mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big]. \qquad (4.18)
\end{aligned}
$$

Thus, injecting (4.18) back into (4.17) and (4.16), set $\Gamma_2 = 4L_f^+ + 2L_u + L_{\tilde{u}} + 1$, then by Remark 4.1,

$$
\mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big] \le \frac{1}{1 - \Gamma_2 h}\mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big] \le \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big](1 + Ch). \qquad (4.19)
$$

This manuscript is for review purposes only.

Plug (4.19) and (4.18) back into (4.14), (4.15) and (4.13). We then conclude that

$$I_1 \le \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big](1+Ch). \tag{4.20}$$

For I_2, by Young's and Jensen's inequality, we have

$$I_2 \le h\,\mathbb{E}\Big[\int_{t_n}^{t_n+h}\Big|v(X_s^{i,N},\mu_s^{X,N}) - v(Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2 \mathrm{d}s \tag{4.21}$$

$$+\int_{t_n}^{t_n+h}\Big|b(s,X_s^{i,N},\mu_s^{X,N}) - b(t_n,Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2 \mathrm{d}s\Big]. \tag{4.22}$$

For (4.21), from Assumption 2.1, using Young's, Jensen's, and Cauchy-Schwarz inequality

$$\mathbb{E}\Big[\Big|v(X_s^{i,N},\mu_s^{X,N}) - v(Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2\Big]$$

$$\le C\mathbb{E}\Big[\Big|u(X_s^{i,N},\mu_s^{X,N}) - u(Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2 + \frac{1}{N}\sum_{j=1}^N\Big|f(X_s^{i,N}-X_s^{j,N}) - f(Y_n^{i,X,N}-Y_n^{j,X,N})\Big|^2\Big] \tag{4.23}$$

$$\le \frac{C}{N}\sum_{j=1}^N\mathbb{E}\Big[\Big|\Big(1+|X_s^{i,N}-X_s^{j,N}|^q + |Y_n^{i,X,N}-Y_n^{j,X,N}|^q\Big)|X_s^{i,N}-Y_n^{i,X,N}-(X_s^{j,N}-Y_n^{j,X,N})|\Big|^2\Big]$$

$$+C\mathbb{E}\Big[(1+|X_s^{i,N}|^{2q}+|Y_n^{i,X,N}|^{2q})(|X_s^{i,N}-Y_n^{i,X,N}|^2) + \frac{1}{N}\sum_{j=1}^N|X_s^{j,N}-Y_n^{j,X,N}|^2\Big]$$

$$\le C\sqrt{\mathbb{E}\Big[1+|X_s^{i,N}|^{4q}+|Y_n^{i,X,N}|^{4q}\Big]\mathbb{E}\Big[|X_s^{i,N}-Y_n^{i,X,N}|^4\Big]} + \mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|X_s^{j,N}-Y_n^{j,X,N}|^2\Big] \tag{4.24}$$

$$+\frac{C}{N}\sum_{j=1}^N\sqrt{\mathbb{E}\Big[1+|X_s^{i,N}-X_s^{j,N}|^{4q}+|Y_n^{i,X,N}-Y_n^{j,X,N}|^{4q}\Big]\mathbb{E}\Big[|X_s^{i,N}-Y_n^{i,X,N}|^4+|X_s^{j,N}-Y_n^{j,X,N}|^4\Big]}. \tag{4.25}$$

Using the structure of the SSM, Young's and Jensen's inequality, and Proposition 4.3 we have

$$|X_s^{i,N}-Y_n^{i,X,N}|^2 \le 2|X_s^{i,N}-X_n^{i,N}|^2 + 2|X_n^{i,N}-Y_n^{i,X,N}|^2, \tag{4.26}$$

$$|X_n^{i,N}-Y_n^{i,X,N}|^2 = \Big|v(Y_n^{i,X,N},\mu_n^{Y,X,N})h\Big|^2 \le 2\Big|u(Y_n^{i,X,N},\mu_n^{Y,X,N})h\Big|^2 + \frac{2h^2}{N}\sum_{j=1}^N\Big|f(Y_n^{i,X,N}-Y_n^{j,X,N})\Big|^2$$

$$\le C\Big(1+|Y_n^{i,X,N}|^{2q+2}+\frac{1}{N}\sum_{j=1}^N|Y_n^{j,X,N}|^2\Big)h^2 + \frac{Ch^2}{N}\sum_{j=1}^N\Big(1+|Y_n^{i,X,N}-Y_n^{j,X,N}|^{2q+2}\Big)$$

$$\le C\Big(1+|Y_n^{i,X,N}|^{2q+2}+\frac{1}{N}\sum_{j=1}^N|Y_n^{j,X,N}|^2\Big)h^2 + \frac{Ch^2}{N}\sum_{j=1}^N\Big(1+|X_n^{i,N}-X_n^{j,N}|^{2q+2}\Big).$$

Similarly, we have

$$|X_s^{i,N}-Y_n^{i,X,N}|^4 \le 16|X_s^{i,N}-X_n^{i,N}|^4 + 16|X_n^{i,N}-Y_n^{i,X,N}|^4, \tag{4.27}$$

$$|X_n^{i,N}-Y_n^{i,X,N}|^4 \le C\Big(1+|Y_n^{i,X,N}|^{4q+4}+\frac{1}{N}\sum_{j=1}^N|Y_n^{j,X,N}|^4\Big)h^4 + \frac{Ch^4}{N}\sum_{j=1}^N\Big(1+|X_n^{i,N}-X_n^{j,N}|^{4q+4}\Big).$$

This manuscript is for review purposes only.

From (1.1) and using (2.4) (since $m \geq 4q+4$) alongside Young's inequality and Itô's isometry, we have

$$\mathbb{E}\big[|X_s^{i,N} - X_n^{i,N}|^2\big] \leq \mathbb{E}\Big[\Big|\int_{t_n}^s v(X_u^{i,N}, \mu_u^{X,N}) + b(u, X_u^{i,N}, \mu_u^{X,N})\mathrm{d}u + \int_{t_n}^s \sigma(u, X_u^{i,N}, \mu_u^{X,N})\mathrm{d}W_u^i\Big|^2\Big] \leq Ch,$$

$$\mathbb{E}\big[|X_s^{i,N} - X_n^{i,N}|^4\big] \leq \mathbb{E}\Big[\Big|\int_{t_n}^s v(X_u^{i,N}, \mu_u^{X,N}) + b(u, X_u^{i,N}, \mu_u^{X,N})\mathrm{d}u + \int_{t_n}^s \sigma(u, X_u^{i,N}, \mu_u^{X,N})\mathrm{d}W_u^i\Big|^4\Big] \leq Ch^2.$$

Also, using (2.4), Jensen's and Young's inequality (since $m \geq 4q+4$) we have

$$\mathbb{E}\Big[\frac{Ch^2}{N}\sum_{j=1}^N\Big(1 + |X_t^{i,N} - X_t^{j,N}|^{2q+2}\Big)\Big] \leq Ch^2 \quad \text{and} \quad \mathbb{E}\Big[\Big|\frac{Ch^2}{N}\sum_{j=1}^N\Big(1 + |X_t^{i,N} - X_t^{j,N}|^{2q+2}\Big)\Big|^2\Big] \leq Ch^4.$$

This next argument uses steps similar to those used in (4.35) and (4.36) (appearing in the proof of Theorem 4.7). Since $X^{\cdot,N}$ has bounded moments via (2.4) (this refers to the true interacting particle system), we have for any $m \geq p \geq 2$ that

$$\mathbb{E}\big[|Y_n^{i,X,N}|^p\big] \leq \Big(4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N |X_n^{i,N} - X_n^{j,N}|^p\Big] + 4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N (1 + |X_n^{j,N}|^2)\Big|^{p/2}\Big] + 1\Big)(1 + Ch) \leq C.$$

Collecting all the terms above, using that the particles are identically distributed, we have

$$\mathbb{E}\big[|X_s^{i,N} - Y_n^{i,X,N}|^2\big] \leq Ch, \qquad \mathbb{E}\big[|X_s^{i,N} - Y_n^{i,X,N}|^4\big] \leq Ch^2, \qquad \mathbb{E}\big[|Y_n^{i,X,N}|^p\big] \leq C, \tag{4.28}$$

$$\mathbb{E}\Big[\Big|W^{(2)}(\mu_s^{X,N}, \mu_n^{Y,X,N})\Big|^2\Big] \leq \mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N |X_s^{j,N} - Y_n^{j,X,N}|^2\Big] \leq Ch. \tag{4.29}$$

Plugging all the above inequalities back into (4.24) and (4.25), we conclude that

$$\mathbb{E}\Big[\Big|v(X_s^{i,N}, \mu_s^{X,N}) - v(Y_n^{i,X,N}, \mu_n^{Y,X,N})\Big|^2\Big] \leq Ch. \tag{4.30}$$

We now consider term (4.22) of I_2. By Assumption 2.1, using (4.28) and (4.29)

$$\mathbb{E}\Big[\Big|b(s, X_s^{i,N}, \mu_s^{X,N}) - b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\Big|^2\Big] \leq C\mathbb{E}\Big[h + |X_s^{i,N} - Y_n^{i,X,N}|^2 + \Big|W^{(2)}(\mu_s^{X,N}, \mu_n^{Y,X,N})\Big|^2\Big] \leq Ch. \tag{4.31}$$

Thus, plugging (4.30), (4.31) back into (4.21) and (4.22), we have

$$I_2 \leq Ch^3. \tag{4.32}$$

For I_3, by Itô's isometry, the results in (4.28) and (4.29), and using similar argument as in (4.31) we have

$$\begin{aligned} I_3 =& \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r}\Big(\sigma(s, X_s^{i,N}, \mu_s^{X,N}) - \sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\Big)\mathrm{d}W_s^i\Big|^2\Big] \\ \leq& \mathbb{E}\Big[\int_{t_n}^{t_n+h}\Big|\Big(\sigma(s, X_s^{i,N}, \mu_s^{X,N}) - \sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\Big)\Big|^2 \mathrm{d}s\Big] \leq Ch^2. \end{aligned} \tag{4.33}$$

Similarly for I_4, by Itô's isometry, Proposition 4.5, Equation (4.19) and using similar argument in (4.15)

$$\begin{aligned} I_4 =& \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r}\Big(\sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Big)\mathrm{d}W_s^i\Big|^2\Big] \\ \leq& \mathbb{E}\Big[\int_{t_n}^{t_n+h}\Big|\Big(\sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Big)\Big|^2 \mathrm{d}s\Big] \leq \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big]Ch. \end{aligned} \tag{4.34}$$

This manuscript is for review purposes only.

Plugging (4.20), (4.32) (4.33) and (4.34) back to (4.6), we have, for all $n \in [\![0, M-1]\!]$, $i \in [\![1, N]\!]$ and $r \in [0, h]$

$$\mathbb{E}\big[|\Delta X^i_{t_n+r}|^2\big] \leq (1+h)\mathbb{E}\big[|X^{i,N}_n - \hat{X}^{i,N}_n|^2\big](1+Ch) + (1+\frac{1}{h})Ch^3 + Ch^2 + \mathbb{E}\big[|X^{i,N}_n - \hat{X}^{i,N}_n|^2\big]Ch$$
$$\leq \mathbb{E}\big[|X^{i,N}_n - \hat{X}^{i,N}_n|^2\big](1+Ch) + Ch^2.$$

By backward induction, the discrete Grönwall's lemma delivers the result of (2.11). □

4.3. Proof of Theorem 2.10: the moment bound result. In this section prove Theorem 2.10. Throughout this section we follow the notation introduced in Theorem 2.10 and let: Assumption 2.1 hold, h is chosen as in (2.9) and $m \geq 2p$ with m as defined in (1.3).

We first prove a moment bounds result across the timegrid then extend it to the continues process as stated in Theorem 2.10.

THEOREM 4.7 (Moment bounds of SSM). *Let the setting of Theorem 2.9 hold. Let $m \geq 2$ where $\hat{X}^{i,N}_0 \in L^m_0(\mathbb{R}^d)$ for all $i \in [\![1, N]\!]$ and let $\hat{X}^{i,N}$ be the continuous-time extension of the SSM given by (2.10). Let $2p \in [2, m]$, then there exists a constant $C > 0$ independent of h, N, M (but depending on T and m) such that*

$$\sup_{i\in[\![1,N]\!]}\sup_{n\in[\![0,M]\!]}\mathbb{E}\big[|\hat{X}^{i,N}_n|^{2p}\big] + \sup_{i\in[\![1,N]\!]}\sup_{n\in[\![0,M-1]\!]}\mathbb{E}\big[|Y^{i,\star,N}_n|^{2p}\big] \leq C\big(1 + \sup_{i\in[\![1,N]\!]}\mathbb{E}\big[\,|\hat{X}^{i,N}_0|^{2p}\big]\big) < \infty.$$

Proof. The next inequality introduces the quantities $H^{X,p}_n$ and $H^{Y,p}_n$. For any $i \in [\![1, N]\!]$, $n \in [\![0, M]\!]$, by Young's and Jensen's inequality

$$\mathbb{E}\big[|\hat{X}^{i,N}_n|^{2p}\big] = \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N(\hat{X}^{i,N}_n - \hat{X}^{j,N}_n) + \frac{1}{N}\sum_{j=1}^N \hat{X}^{j,N}_n\Big|^{2p}\Big]$$
$$\leq 4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|\hat{X}^{i,N}_n - \hat{X}^{j,N}_n|^{2p}\Big] + 4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N(1+|\hat{X}^{j,N}_n|^2)\Big|^p\Big] + 1 = H^{X,p}_n, \tag{4.35}$$

$$\mathbb{E}\big[|Y^{i,\star,N}_n|^{2p}\big] \leq 4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|Y^{i,\star,N}_n - Y^{j,\star,N}_n|^{2p}\Big] + 4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N(1+|Y^{j,\star,N}_n|^2)\Big|^p\Big] + 1 = H^{Y,p}_n. \tag{4.36}$$

Using the following inequalities from Proposition 4.3 and 4.4, we have $H^{Y,p}_n \leq H^{X,p}_n(1+Ch)$,

$$|Y^{i,\star,N}_n - Y^{j,\star,N}_n|^2 \leq |\hat{X}^{i,N}_n - \hat{X}^{j,N}_n|^2(1+Ch) \quad\text{and}\quad \frac{1}{N}\sum_{j=1}^N(1+|Y^{j,\star,N}_n|^2) \leq \Big[\frac{1}{N}\sum_{j=1}^N(1+|\hat{X}^{j,N}_n|^2)\Big](1+Ch).$$

We now prove that $H^{X,p}_{n+1} \leq H^{Y,p}_n(1+Ch)$. For the first element composing $H^{X,p}_{n+1}$ we have

$$\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|\hat{X}^{i,N}_{n+1} - \hat{X}^{j,N}_{n+1}|^{2p}\Big] = \mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N\Big|\Big(Y^{i,\star,N}_n + b(t_n, Y^{i,\star,N}_n, \hat{\mu}^{Y,N}_n)h + \sigma(t_n, Y^{i,\star,N}_n, \hat{\mu}^{Y,N}_n)\Delta W^i_n\Big)$$
$$- \Big(Y^{j,\star,N}_n + b(t_n, Y^{j,\star,N}_n, \hat{\mu}^{Y,N}_n)h + \sigma(t_n, Y^{j,\star,N}_n, \hat{\mu}^{Y,N}_n)\Delta W^j_n\Big)\Big|^{2p}\Big]. \tag{4.37}$$

Introduce the extra (local) notation for $G^{i,j,n}_1$, $G^{i,j,n}_2$ and $G^{i,j,n}_3$ as

$$G^{i,j,n}_1 = Y^{i,\star,N}_n - Y^{j,\star,N}_n, \quad G^{i,j,n}_2 = b(t_n, Y^{i,\star,N}_n, \hat{\mu}^{Y,N}_n)h - b(t_n, Y^{j,\star,N}_n, \hat{\mu}^{Y,N}_n)h,$$
$$G^{i,j,n}_3 = \sigma(t_n, Y^{i,\star,N}_n, \hat{\mu}^{Y,N}_n)\Delta W^i_n - \sigma(t_n, Y^{j,\star,N}_n, \hat{\mu}^{Y,N}_n)\Delta W^j_n.$$

This manuscript is for review purposes only.

For $a+b+c=2p$, $a<2p$, $a,b,c\in\mathbb{N}$, by Assumption 2.1, Young's inequality, Jensen's inequality, Proposition 4.6 and the fact that the Brownian increments are independent, the particles are conditionally independent and identically distributed, for (4.37), we have

$$\mathbb{E}\Big[\frac{C}{N}\sum_{j=1}^{N}|G_1^{i,j,n}|^a|G_2^{i,j,n}|^b|G_3^{i,j,n}|^c\Big]\leq \mathbb{E}\big[|Y_n^{i,\star,N}|^{2p}\big]Ch\leq H_n^{Y,p}Ch.$$

Thus, for the first term of $H_{n+1}^{X,p}$, we conclude that

$$4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_{n+1}^{i,N}-\hat{X}_{n+1}^{j,N}|^{2p}\Big]\leq 4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^{N}|Y_n^{i,\star,N}-Y_n^{j,\star,N}|^{2p}\Big]+H_n^{Y,p}Ch. \tag{4.38}$$

For the second term of $H_{n+1}^{X,p}$ we have

$$\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}(1+|\hat{X}_{n+1}^{j,N}|^2)\Big|^p\Big]=\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}\Big[1+\Big(Y_n^{j,\star,N}+b(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})h+\sigma(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^j\Big)^2\Big]\Big|^p\Big].$$

Set the following (extra local) notation

$$G_4^n=\frac{1}{N}\sum_{j=1}^{N}(1+|Y_n^{j,\star,N}|^2),\quad G_5^n=\frac{1}{N}\sum_{j=1}^{N}\Big\langle 2Y_n^{j,\star,N}+\sigma(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^j,\sigma(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^j\Big\rangle,$$

$$G_6^n=\frac{1}{N}\sum_{j=1}^{N}\Big\langle 2Y_n^{j,\star,N}+b(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})h+2\sigma(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^j,b(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})h\Big\rangle.$$

We have once again using similar arguments as before, by Young's inequality, Jensen's inequality, Proposition 4.6, that the particles are conditionally independent and identically distributed, the independence property of the Brownian increments, the Lipschitz property for b and σ, and using the fact that for $l_1>l_2>1$, $|x|^{l_2}\leq 1+|x|^{l_1}$ we have

$$\mathbb{E}\big[|G_4^n|^a|G_5^n|^b|G_6^n|^c\big]\leq\mathbb{E}\big[|Y_n^{i,\star,N}|^{2p}+1\big]Ch\leq H_n^{Y,p}Ch.$$

Thus, for the second term of $H_{n+1}^{X,p}$, we conclude that

$$4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}(1+|\hat{X}_{n+1}^{j,N}|^2)\Big|^p\Big]\leq 4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}(1+|Y_n^{j,\star,N}|^2)\Big|^p\Big]+H_n^{Y,p}Ch. \tag{4.39}$$

Plug (4.38) and (4.39) into $H_{n+1}^{X,p}$ we have

$$\begin{aligned}H_{n+1}^{X,p}&=4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_{n+1}^{i,N}-\hat{X}_{n+1}^{j,N}|^{2p}\Big]+4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}(1+|\hat{X}_{n+1}^{j,N}|^2)\Big|^p\Big]+1\\&\leq 4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^{N}|Y_n^{i,\star,N}-Y_n^{j,\star,N}|^{2p}\Big]+4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}(1+|Y_n^{j,\star,N}|^2)\Big|^p\Big]+1+H_n^{Y,p}Ch\leq H_n^{Y,p}(1+Ch).\end{aligned}$$

Thus finally, for all $n\in[\![0,M-1]\!]$, $i\in[\![1,N]\!]$, by backward induction collecting all the results above, since $m\geq 2p$, where m is defined in (1.3), we have (for some $C>0$ independent of h,N,M)

$$\mathbb{E}\big[|\hat{X}_{n+1}^{i,N}|^{2p}\big]\leq H_{n+1}^{X,p}\leq H_n^{Y,p}(1+Ch)\leq H_n^{X,p}(1+Ch)^2\leq\cdots\leq H_0^{X,p}e^{CT}\leq C\mathbb{E}\big[|\hat{X}_0^{i,N}|^{2p}\big]+C<\infty.$$

Similar argument yields the result for $\mathbb{E}\big[|Y_n^{i,\star,N}|^{2p}\big]$. ∎

This manuscript is for review purposes only.

Proof of the Theorem 2.10.

Proof of the Theorem 2.10. Under the same assumptions and notations of Theorem 4.7, one can apply arguments similar to those used in [17, Proposition 4.6] to obtain the result. ∎

The final result of this section concerns the incremental (in time) moment bounds of $\hat{X}^{i,N}$. This result is in preparation for the next section.

PROPOSITION 4.8. *Under same assumptions and notations of Theorem 2.10, there exists a constant $C > 0$ independent of h, N, M (but depending on T and m) such that for any $p \geq 2$ satisfy $m \geq (q+1)p$, where m is defined in (1.3), q is defined in Assumption 2.1, we have*

$$\sup_{i\in[\![1,N]\!]} \sup_{0\leq t\leq T} \mathbb{E}\big[|\hat{X}^{i,N}_t - \hat{X}^{i,N}_{\kappa(t)}|^p\big] \leq Ch^{\frac{p}{2}}. \tag{4.40}$$

Proof. Under Assumption 2.1, and carefully applying Young's and Jensen's inequality, one can argue similarly as to [17, Proposition 4.7] and obtain the result (we omit further details). ∎

4.4. Proof of Theorem 2.11, the uniform convergence result in path-space. We now prove Theorem 2.11.

Proof of Theorem 2.11. Let Assumption 2.1 hold. Let $i \in [\![1,N]\!]$, $t \in [0,T]$, suppose $m \geq \max\{4q+4, 2+q+q/\varepsilon\}$, where $X^i_0 \in L^m_0(\mathbb{R}^d)$, q is as given in Assumption 2.1. From (2.4) and (2.12), both process $X^{i,N}$ and $\hat{X}^{i,N}$ have sufficient bounded moments for the following proof. Define $\Delta X^i := X^{i,N} - \hat{X}^{i,N}$. Itô's formula applied to $|X^{i,N}_t - \hat{X}^{i,N}_t|^2 = |\Delta X^i_t|^2$ yields

$$|\Delta X^i_t|^2 = 2\int_0^t \Big\langle v(X^{i,N}_s, \mu^{X,N}_s) - v(Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)}), \Delta X^i_s \Big\rangle \mathrm{d}s \tag{4.41}$$

$$+ 2\int_0^t \Big\langle b(s, X^{i,N}_s, \mu^{X,N}_s) - b(\kappa(s), Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)}), \Delta X^i_s \Big\rangle \mathrm{d}s \tag{4.42}$$

$$+ \int_0^t \Big|\sigma(s, X^{i,N}_s, \mu^{X,N}_s) - \sigma(\kappa(s), Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)})\Big|^2 \mathrm{d}s \tag{4.43}$$

$$+ 2\int_0^t \Big\langle \Delta X^i_s, \Big(\sigma(s, X^{i,N}_s, \mu^{X,N}_s) - \sigma(\kappa(s), Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)})\Big)\mathrm{d}W^i_s \Big\rangle. \tag{4.44}$$

We analyse the above terms one by one and will take supremum over time with expectation. For (4.41),

$$\begin{aligned} &\big\langle v(X^{i,N}_s, \mu^{X,N}_s) - v(Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)}), \Delta X^i_s \big\rangle \\ &= \big\langle v(X^{i,N}_s, \mu^{X,N}_s) - v(\hat{X}^{i,N}_s, \hat{\mu}^{X,N}_s), \Delta X^i_s \big\rangle + \big\langle v(\hat{X}^{i,N}_s, \hat{\mu}^{X,N}_s) - v(Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)}), \Delta X^i_s \big\rangle. \end{aligned} \tag{4.45}$$

This manuscript is for review purposes only.

For the first term above, by Assumption 2.1 and using Remark 2.2

$$\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle v(X_s^{i,N},\mu_s^{X,N})-v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle \mathrm{d}s\Big]$$

$$\le \mathbb{E}\Big[\int_0^T \frac{C}{N}\sum_{j=1}^N \Big|f(X_s^{i,N}-X_s^{j,N})-f(\hat{X}_s^{i,N}-\hat{X}_s^{j,N})\Big||\Delta X_s^i|\mathrm{d}s\Big]$$

$$+\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle u(X_s^{i,N},\mu_s^{X,N})-u(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle \mathrm{d}s\Big]$$

$$\le \mathbb{E}\Big[\int_0^T \frac{C}{N}\sum_{j=1}^N \Big\{\Big(1+|X_s^{i,N}-X_s^{j,N}|^q+|\hat{X}_s^{i,N}-\hat{X}_s^{j,N}|^q\Big)|\Delta X_s^i-\Delta X_s^j||\Delta X_s^i|\Big\}\mathrm{d}s\Big] \tag{4.46}$$

$$+\mathbb{E}\Big[\int_0^T \Big(\widehat{L}_u|\Delta X_s^i|^2+\frac{L_{\tilde{u}}}{2N}\sum_{j=1}^N|\Delta X_s^j|^2\Big)\mathrm{d}s\Big].$$

To deal with (4.46), using the following notations, for all $i,j\in[\![1,N]\!]$,

$$G_7^{i,j,s}=\Big(1+|X_s^{i,N}-X_s^{j,N}|^q+|\hat{X}_s^{i,N}-\hat{X}_s^{j,N}|^q\Big)\quad\text{and}\quad G_8^{i,j,s}=|\Delta X_s^i-\Delta X_s^j||\Delta X_s^i|.$$

The combination of $G_7^{i,j,s}$ and $G_8^{i,j,s}$ makes it difficult to obtain a domination via $|\Delta X_s^i|^2$, we overcome this by applying Chebyshev's inequality as follows. The indicator function is denoted as $\mathbb{1}_{\{\Omega\}}$. Recall the moment bound results on $X,\hat{X}$ in (2.4) and (2.12) respectively. Now, using Theorem 2.9, Proposition 4.6 and Young's inequality, we have

$$\mathbb{E}\big[G_7^{i,j,s}G_8^{i,j,s}\big]=\mathbb{E}\Big[G_7^{i,j,s}G_8^{i,j,s}(\mathbb{1}_{\{G_7^{i,j,s}<M^\varepsilon\}})\Big]+\mathbb{E}\Big[G_7^{i,j,s}G_8^{i,j,s}(\mathbb{1}_{\{G_7^{i,j,s}\ge M^\varepsilon\}})\Big] \tag{4.47}$$

$$\le \mathbb{E}\big[M^\varepsilon G_8^{i,j,s}\big]+\mathbb{E}\Big[\frac{|G_7^{i,j,s}|^{1/\varepsilon}}{M}G_7^{i,j,s}G_8^{i,j,s}\Big]\le 2\mathbb{E}\big[M^\varepsilon|\Delta X_s^i|^2\big]+h\mathbb{E}\big[|G_7^{i,j,s}|^{1/\varepsilon}G_7^{i,j,s}G_8^{i,j,s}\big]$$

$$\le Ch^{1-\varepsilon}+hC\Big(1+\mathbb{E}\Big[|X_s^{i,N}|^{2+q+q/\varepsilon}+|\hat{X}_s^{i,N}|^{2+q+q/\varepsilon}\Big]\Big)\le Ch^{1-\varepsilon}, \tag{4.48}$$

where for the last inequality, we used that the particles are identically distributed and there are sufficiently high bounded moments available for the process since $m\ge 2+q+q/\varepsilon$.

Thus, for the first term in (4.45) and using that the particles are identically distributed, we conclude that

$$\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle v(X_s^{i,N},\mu_s^{X,N})-v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle\mathrm{d}s\Big]\le C\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2\mathrm{d}s\Big]+Ch^{1-\varepsilon}. \tag{4.49}$$

For the second term in (4.45), under Assumption 2.1, using Young's inequality, Jensen's inequality, and Proposition 4.8 we have

$$\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N})-v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\mathrm{d}s\Big] \tag{4.50}$$

$$=\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle u(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N})-u(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\mathrm{d}s\Big] \tag{4.51}$$

$$+\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \frac{1}{N}\sum_{j=1}^N\langle f(\hat{X}_s^{i,N}-\hat{X}_s^{j,N})-f(Y_{\kappa(s)}^{i,\star,N}-Y_{\kappa(s)}^{j,\star,N}),\Delta X_s^i\Big\rangle\mathrm{d}s\Big] \tag{4.52}$$

$$\le \mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2\mathrm{d}s\Big]+I_2+I_3.$$

This manuscript is for review purposes only.

For I_2 (given by the domination of (4.51)), by Assumption 2.1, Young's inequality and Cauchy-Schwarz inequality

$$I_2 = L_{\hat{u}}\mathbb{E}\Big[\int_0^T \Big(1+|\hat{X}_s^{i,N}|^q+|Y_{\kappa(s)}^{i,\star,N}|^q\Big)^2|\hat{X}_s^{i,N}-Y_{\kappa(s)}^{i,\star,N}|^2\Big]\mathrm{d}s$$
$$\leq C\int_0^T \sqrt{\mathbb{E}\Big[\Big(1+|\hat{X}_s^{i,N}|^{2q}+|Y_{\kappa(s)}^{i,\star,N}|^{2q}\Big)^2\Big]\mathbb{E}\big[|\hat{X}_s^{j,N}-Y_{\kappa(s)}^{j,\star,N}|^4\big]}\mathrm{d}s.$$

For I_3 (given by the domination of (4.52) after extracting the $|\Delta X^i|$ term), by Assumption 2.1, Young's inequality and Cauchy-Schwarz inequality

$$I_3 = \frac{CL_{\hat{f}}}{N}\sum_{j=1}^N \mathbb{E}\Big[\int_0^T \Big(1+|\hat{X}_s^{i,N}-\hat{X}_s^{j,N}|^q+|Y_{\kappa(s)}^{i,\star,N}-Y_{\kappa(s)}^{j,\star,N}|^q\Big)^2\ \big|(\hat{X}_s^{i,N}-\hat{X}_s^{j,N})-(Y_{\kappa(s)}^{i,\star,N}-Y_{\kappa(s)}^{j,\star,N})\big|^2\Big]\mathrm{d}s$$
$$\leq \frac{C}{N}\sum_{j=1}^N\int_0^T \sqrt{\mathbb{E}\Big[\Big(1+|\hat{X}_s^{j,N}|^{2q}+|\hat{X}_s^{i,N}|^{2q}+|Y_{\kappa(s)}^{i,\star,N}|^{2q}+|Y_{\kappa(s)}^{j,\star,N}|^{2q}\Big)^2\Big]\mathbb{E}\big[|\hat{X}_s^{j,N}-Y_{\kappa(s)}^{j,\star,N}|^4\big]}\mathrm{d}s.$$

By (2.7), Assumption 2.1, Young's inequality, Jensen's inequality, since $m\geq 4q+4$, and by Theorem 4.7, we have

$$\mathbb{E}\big[|\hat{X}_{\kappa(s)}^{i,N}-Y_{\kappa(s)}^{i,\star,N}|^4\big]=\mathbb{E}\big[|hv(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_\kappa^{Y,N}(s))|^4\big]$$
$$\leq Ch^4\mathbb{E}\big[|u(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_\kappa^{Y,N}(s))|^4\big]+\frac{Ch^4}{N}\sum_{j=1}^N\mathbb{E}\big[|f(Y_{\kappa(s)}^{i,\star,N}-Y_{\kappa(s)}^{j,\star,N})|^4\big]$$
$$\leq Ch^4\mathbb{E}\Big[1+|Y_{\kappa(s)}^{i,\star,N}|^{4q+4}+\frac{1}{N}\sum_{j=1}^N|Y_{\kappa(s)}^{j,\star,N}|^4\Big]+\frac{Ch^4}{N}\sum_{j=1}^N\mathbb{E}\Big[(1+|Y_{\kappa(s)}^{i,\star,N}-Y_{\kappa(s)}^{j,\star,N}|^{4q})|Y_{\kappa(s)}^{i,\star,N}-Y_{\kappa(s)}^{j,\star,N}|^4\Big]$$
$$\leq\frac{Ch^4}{N}\sum_{j=1}^N\mathbb{E}\big[1+|Y_{\kappa(s)}^{j,\star,N}|^{4q+4}\big]\leq Ch^4.$$

Using this inequality in combination with Proposition 4.8 allows us to conclude that

$$\mathbb{E}\big[|\hat{X}_s^{j,N}-Y_{\kappa(s)}^{j,\star,N}|^4\big]\leq C\mathbb{E}\Big[|\hat{X}_s^{j,N}-\hat{X}_{\kappa(s)}^{j,N}|^4+|\hat{X}_{\kappa(s)}^{j,N}-Y_{\kappa(s)}^{j,\star,N}|^4\Big]\leq Ch^2. \tag{4.53}$$

Thus, for (4.45) injected back in (4.41), take supremum and expectation, and collecting all the necessary results above, we reach

$$\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t\Big\langle v(X_s^{i,N},\mu_s^N)-v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\mathrm{d}s\Big]\leq C\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2\mathrm{d}s\Big]+Ch^{1-\varepsilon}. \tag{4.54}$$

For the second term (4.42), the calculation is similar as in [17, Proof of Proposition 4.9], we conclude that

$$\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t\Big\langle b(s,X_s^{i,N},\mu_s^{X,N})-b(\kappa(s),Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\mathrm{d}s\Big]\leq Ch+C\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2\mathrm{d}s\Big]. \tag{4.55}$$

Similarly, for the third term (4.43) (these are just Lipschitz terms), we have

$$\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t\Big|\sigma(s,X_s^{i,N},\mu_s^{X,N})-\sigma(\kappa(s),Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\Big|^2\mathrm{d}s\Big]\leq Ch+C\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2\mathrm{d}s\Big]. \tag{4.56}$$

This manuscript is for review purposes only.

Consider the last term (4.44) – this is a Lipschitz term and dealt with similarly to [17, Proof of Proposition 4.9]. Using the Burkholder-Davis-Gundy's, Jensen's and Cauchy-Schwarz inequality, and the above results,

$$\begin{aligned}
(4.57)\qquad &\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle \Delta X_s^i, \Big(\sigma(s,X_s^{i,N},\mu_s^{X,N})-\sigma(\kappa(s),Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\Big)\mathrm{d}W_s^i\Big\rangle\Big]\\
&\le \frac{1}{4}\mathbb{E}\big[\sup_{0\le t\le T}|\Delta X_t^i|^2\big]+\mathbb{E}\Big[\int_0^T\Big|\sigma(s,X_s^{i,N},\mu_s^{X,N})-\sigma(\kappa(s),Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\Big|^2\mathrm{d}s\Big].
\end{aligned}$$

Again, gathering all the above results (4.54), (4.55), (4.56), and (4.57), plugging them back into (4.41), after taking supremum over $t\in[0,T]$ and expectation, for all $i\in[\![1,N]\!]$ we have

$$\begin{aligned}
\mathbb{E}\big[\sup_{0\le t\le T}|\Delta X_t^i|^2\big]&\le Ch^{1-\varepsilon}+C\mathbb{E}\Big[\int_0^T\sup_{0\le u\le s}|\Delta X_u^i|^2\mathrm{d}s\Big]+\frac{1}{2}\mathbb{E}\big[\sup_{0\le t\le T}|\Delta X_t^i|^2\big]\\
&\le Ch^{1-\varepsilon}+C\int_0^T\mathbb{E}\big[\sup_{0\le u\le s}|\Delta X_u^i|^2\big]\mathrm{d}s.
\end{aligned}$$

Grönwall's lemma delivers the final result after taking supremum over $i\in[\![1,N]\!]$. ◻

4.5. Discussion on the granular media type equation.

Throughout $C>0$ denotes a constant always independent of h,N,M but possibly depending on T and m.

Proof of Proposition 2.5. Recall the proof of (4.41) in Section 4.4. Under Assumption 2.12, for all $i\in[\![1,N]\!]$, $t\in[0,T]$, and using arguments similar to those of (4.45) we have

$$\begin{aligned}
\Delta X_t^i = X_t^{i,N}-\hat{X}_t^{i,N}&=\int_0^t v(X_s^{i,N},\mu_s^{X,N})-v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\,\mathrm{d}s,\\
(4.58)\qquad \Rightarrow \mathbb{E}\big[|\Delta X_t^i|^2\big]&\le 2\int_0^t\mathbb{E}\Big[\Big\langle v(X_s^{i,N},\mu_s^{X,N})-v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle\Big]\mathrm{d}s\\
(4.59)\qquad &\quad+2\int_0^t\mathbb{E}\Big[\Big\langle v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N})-v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\Big]\mathrm{d}s.
\end{aligned}$$

For (4.58), arguing as in (4.18), Remark 2.4 and using that the particles are identically distributed, we have

$$(4.60)\qquad \mathbb{E}\Big[\Big\langle v(X_s^{i,N},\mu_s^{X,N})-v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle\Big]\le 2L_f^{+}\mathbb{E}\big[|\Delta X_s^i|^2\big].$$

For (4.59), it is similar to the above, we have

$$(4.61)\quad 2\int_0^t\mathbb{E}\Big[\Big\langle v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N})-v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\Big]\mathrm{d}s=\frac{2}{N}\sum_{j=1}^N\int_0^t\mathbb{E}\Big[\Big\langle f(\Delta_s^{X,i,j})-f(\Delta_{\kappa(s)}^{Y,i,j}),\Delta X_s^i\Big\rangle\Big]\mathrm{d}s,$$

where we introduce the following handy notation (recall (2.7) and (2.10))

$$\begin{aligned}
&\Delta_t^{X,i,j}=\hat{X}_s^{i,N}-\hat{X}_s^{j,N},\qquad \Delta_{\kappa(s)}^{Y,i,j}=Y_{\kappa(s)}^{i,\star,N}-Y_{\kappa(s)}^{j,\star,N},\\
&\Delta_s^{X,i,j}=\Delta_{\kappa(s)}^{X,i,j}+G_9^{i,j,s}(s-\kappa(s))+G_{10}^{i,j,s},\qquad \Delta_{\kappa(s)}^{Y,i,j}=\Delta_{\kappa(s)}^{X,i,j}+G_9^{i,j,s}h,\\
(4.62)\qquad &G_9^{i,j,s}=\Big(v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})-v(Y_{\kappa(s)}^{j,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\Big)\quad\text{and}\quad G_{10}^{i,j,s}=\sigma\Big((W_s^i-W_{\kappa(s)}^i)-(W_s^j-W_{\kappa(s)}^j)\Big).
\end{aligned}$$

This manuscript is for review purposes only.

We now proceed to estimate (4.61). By the mean value theorem under Assumption 2.12, for (4.61), there exist $\rho_1, \rho_2 \in [0,1]$ such that

$$f(\Delta_s^{X,i,j}) = f(\Delta_{\kappa(s)}^{X,i,j}) + \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s}\Big) + \int_{\Delta_{\kappa(s)}^{X,i,j}}^{\Delta_s^{X,i,j}} \Big(\nabla f(u) - \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big) du$$

$$= f(\Delta_{\kappa(s)}^{X,i,j}) + \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s}\Big)$$

$$+ \Big(\nabla f(\Delta_{\kappa(s)}^{X,i,j} + \rho_1(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s})) - \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big)\Big(\Delta_s^{X,i,j} - \Delta_{\kappa(s)}^{X,i,j}\Big),$$

$$f(\Delta_{\kappa(s)}^{Y,i,j}) = f(\Delta_{\kappa(s)}^{X,i,j}) + \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}h\Big) + \Big(\nabla f(\Delta_{\kappa(s)}^{X,i,j} + \rho_2(G_{10}^{i,j,s}h)) - \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big)\Big(\Delta_{\kappa(s)}^{Y,i,j} - \Delta_{\kappa(s)}^{X,i,j}\Big).$$

Note that only G_{10} contains the Brownian increments. From the above, there exists $\rho_{1,s},\ \rho_{2,s} \in [0,1]$ for all $s \in [0,T]$, and by Young's inequality, we have

$$\text{(4.63)} \quad \int_0^t \mathbb{E}\Big[\Big\langle f(\Delta_s^{X,i,j}) - f(\Delta_{\kappa(s)}^{Y,i,j}), \Delta X_s^i\Big\rangle\Big] \mathrm{d}s$$

$$\text{(4.64)} \quad \leq \int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-h-\kappa(s)) + G_{10}^{i,j,s}\Big), \Delta X_s^i\Big\rangle\Big] \mathrm{d}s + C\int_0^t \mathbb{E}\Big[|\Delta X_s^i|^2\Big] \mathrm{d}s$$

$$\text{(4.65)} \quad + C\int_0^t \mathbb{E}\Big[\Big|\nabla f\Big(\Delta_{\kappa(s)}^{X,i,j} + \rho_{1,s}(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s})\Big) - \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big|^2 \Big|\Delta_s^{X,i,j} - \Delta_{\kappa(s)}^{X,i,j}\Big|^2\Big] \mathrm{d}s$$

$$\text{(4.66)} \quad + C\int_0^t \mathbb{E}\Big[\Big|\nabla f\big(\Delta_{\kappa(s)}^{X,i,j} + \rho_{2,s}(G_9^{i,j,s}h)\big) - \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big|^2 \Big|\Delta_{\kappa(s)}^{Y,i,j} - \Delta_{\kappa(s)}^{X,i,j}\Big|^2\Big] \mathrm{d}s.$$

For the first term of (4.64), by Young's inequality

$$\text{(4.67)} \quad \int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-h-\kappa(s)) + G_{10}^{i,j,s}\Big), \Delta X_s^i\Big\rangle\Big] \mathrm{d}s$$

$$\text{(4.68)} \quad \leq C\int_0^t \mathbb{E}\Big[|\Delta X_s^i|^2\Big] \mathrm{d}s + C\int_0^t \mathbb{E}\Big[\Big|\nabla f(\Delta_{\kappa(s)}^{X,i,j}) G_9^{i,j,s}(s-h-\kappa(s))\Big|^2\Big] \mathrm{d}s$$

$$\text{(4.69)} \quad + \int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\ G_{10}^{i,j,s}, \Delta X_s^i - \Delta X_{\kappa(s)}^i\Big\rangle\Big] \mathrm{d}s + \int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\ G_{10}^{i,j,s}, \Delta X_{\kappa(s)}^i\Big\rangle\Big] \mathrm{d}s.$$

For the second term of (4.68), since $m \geq 4q+2$, by Assumption 2.12 and Theorem 2.10, using calculations similar to those in (4.23) and Proposition 4.6, we have

$$C\int_0^t \mathbb{E}\Big[\Big|\nabla f(\Delta_{\kappa(s)}^{X,i,j}) G_9^{i,j,s}(s-h-\kappa(s))\Big|^2\Big] \mathrm{d}s \leq Ch^2 \int_0^t \mathbb{E}\Big[1 + |\hat{X}_{\kappa(s)}^{i,N}|^{4q+2} + |Y_{\kappa(s)}^{i,\star,N}|^{4q+2}\Big] \mathrm{d}s \leq Ch^2.$$

By Jensen's inequality and calculations close to those for I_3 in (4.52), since $m \geq 4q+2$, we have

$$\text{(4.70)} \quad \mathbb{E}\big[|\Delta X_t^i - \Delta X_{\kappa(t)}^i|^2\big] = \mathbb{E}\Big[\Big|\int_{\kappa(t)}^t \Big(v(X_s^{i,N}, \mu_s^{X,N}) - v(Y_{\kappa(s)}^{i,\star,N}, \hat{\mu}_{\kappa(s)}^{Y,N})\Big)\ \mathrm{d}s\Big|^2\Big]$$

$$\text{(4.71)} \quad \leq h\int_{\kappa(t)}^t \frac{1}{N}\sum_{j=1}^N \mathbb{E}\Big[\Big|f(X_s^{i,N} - X_s^{i,N}) - f(Y_{\kappa(s)}^{i,\star,N} - Y_{\kappa(s)}^{i,\star,N})\Big|^2\Big]\ \mathrm{d}s \leq Ch^3.$$

Thus, for the first term of (4.69), by Cauchy-Schwarz inequality and the properties of the Brownian increment

$$\int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\ G_{10}^{i,j,s}, \Delta X_s^i - \Delta X_{\kappa(s)}^i\Big\rangle\Big] \mathrm{d}s \leq \int_0^t \sqrt{\mathbb{E}\big[\big|\nabla f(\Delta_{\kappa(s)}^{X,i,j})\ G_{10}^{i,j,s}\big|^2\big]}\sqrt{\mathbb{E}\big[\big|\Delta X_s^i - \Delta X_{\kappa(s)}^i\big|^2\big]} \mathrm{d}s \leq Ch^2.$$

This manuscript is for review purposes only.

For the second term of (4.69), since $G_{10}^{i,j,s}$ of (4.62) is conditionally independent of $\Delta_{\kappa(s)}^{X,i,j}$ and $\Delta X_{\kappa(s)}^i$ (and contains the Brownian increments), the tower property yields

$$\int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\, G_{10}^{i,j,s}, \Delta X_{\kappa(s)}^i \Big\rangle\Big] \mathrm{d}s = 0. \tag{4.72}$$

Thus, plugging the above results back into (4.64), we conclude that

$$\int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-h-\kappa(s)) + G_{10}^{i,j,s}\Big), \Delta X_s^i \Big\rangle\Big] \mathrm{d}s \le Ch^2. \tag{4.73}$$

For (4.65), by Assumption 2.12, Cauchy-Schwarz inequality and the properties of the Brownian increment, and the condition $m \ge \max\{8q,\ 4q+4\}$

$$\begin{aligned}
&\mathbb{E}\Big[\big|\nabla f\big(\Delta_{\kappa(s)}^{X,i,j} + \rho_{1,s}(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s})\big) - \nabla f\big(\Delta_{\kappa(s)}^{X,i,j}\big)\big|^4\Big] \\
&\le C\mathbb{E}\Big[\Big|\Big(1 + \big|\Delta_{\kappa(s)}^{X,i,j} + \rho_{1,s}\big(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s}\big)\big|^{q-1} + \big|\Delta_{\kappa(s)}^{X,i,j}\big|^{q-1}\Big)\big|\rho_{1,s}\big(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s}\big)\big|^4\Big] \le Ch^2,
\end{aligned}$$

and

$$\mathbb{E}\big[\big|\Delta_s^{X,i,j} - \Delta_{\kappa(s)}^{X,i,j}\big|^4\big] \le C\mathbb{E}\Big[\big|\big(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s}\big)\big|^4\Big] \le Ch^2.$$

Thus, using Cauchy-Schwarz inequality again and the results above we conclude that

$$\int_0^t \mathbb{E}\Big[\big|\nabla f\Big(\Delta_{\kappa(s)}^{X,i,j} + \rho_{1,s}\big(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s}\big)\Big) - \nabla f(\Delta_{\kappa(s)}^{X,i,j})\big|^2 \big|\Delta_s^{X,i,j} - \Delta_{\kappa(s)}^{X,i,j}\big|^2\Big] \mathrm{d}s \le Ch^2. \tag{4.74}$$

For (4.66), recall (4.62). Similarly to above, by assumption $m \ge 4q+2$ and hence

$$\int_0^t \mathbb{E}\Big[\ \big|\nabla f\big(\Delta_{\kappa(s)}^{X,i,j} + \rho_{2,s} G_9^{i,j,s} h\big) - \nabla f\big(\Delta_{\kappa(s)}^{X,i,j}\big)\big|^2\ \big|G_9^{i,j,s} h\big|^2\ \Big] \mathrm{d}s \le Ch^2. \tag{4.75}$$

Thus, plugging (4.73), (4.74) and (4.75) back into (4.63), yields

$$\int_0^t \mathbb{E}\Big[\Big\langle f(\Delta_s^{X,i,j}) - f(\Delta_{\kappa(s)}^{Y,i,j}), \Delta X_s^i \Big\rangle\Big] \mathrm{d}s \le Ch^2 + C\int_0^t \mathbb{E}\big[|\Delta X_s^i|^2\big] \mathrm{d}s. \tag{4.76}$$

Plug the above result and (4.60) back to (4.58), we conclude that, for all $i \in [\![1, N]\!]$, $t \in [0, T]$

$$\mathbb{E}\big[|\Delta X_t^i|^2\big] \le C\int_0^t \mathbb{E}\big[|\Delta X_s^i|^2\big] \mathrm{d}s + Ch^2. \tag{4.77}$$

Grönwall's lemma delivers the final result after taking supremum over $i \in [\![1, N]\!]$. ∎

Appendix A. Well-posedness of the particle system and the PoC – Proposition 2.5 .

The Propagation of chaos result (2.5) follows directly from [1, Theorem 3.14]. The gap we close is the well-posedness result for the interacting particle system and the moment bound result. Note that throughout $C > 0$ is a constant always independent of h, N, M but possibly depending on T and m.

Proof of Proposition 2.5. We start by interpreting the interacting particle system (1.1) as a single SDE in $\mathbb{R}^{Nd}$. In Remark 2.4 we show that, as a system in $\mathbb{R}^{Nd}$, the function V (see (2.2) and (1.4)) satisfies a one-sided Lipschitz condition (as a map in $\mathbb{R}^{Nd}$). Thus: (i) the drift term of the whole system also satisfies one-sided Lipschitz condition as b satisfies a uniformly Lipschitz condition by $(\mathbf{A}^b)$; (ii) the diffusion coefficient satisfies a

This manuscript is for review purposes only.

Lipschitz condition (by $(\mathbf{A}^\sigma)$). In conclusion, the well-posedness of the interacting particle SDE $\mathbb{R}^{Nd}$-system is ensured by standard SDE results [47, Theorem 3.5 (p.58)].

The moment bound result of the $\mathbb{R}^{Nd}$-system that follows from [47, Theorem 3.5 (p.58)] does not lead to (2.4) as the constant appearing on the right-hand side *depends on* N and explode as $N \nearrow \infty$. Nonetheless, with well-posedness at hand, we are able to improve the bound and show (2.4).

The strategy of the proof is the same as that in Section 4.3. For all $m \geq 2p \geq 2$, $i \in [\![1, N]\!]$, $t \in [0, T]$, we have

$$
\begin{aligned}
\mathbb{E}\big[|X_t^{i,N}|^{2p}\big] &= \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N \big(X_t^{i,N} - X_t^{j,N}\big) + \frac{1}{N}\sum_{j=1}^N X_t^{j,N}\Big|^{2p}\Big] \\
&\leq 4^p \mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N |X_t^{i,N} - X_t^{j,N}|^{2p}\Big] + 4^p \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N |X_t^{j,N}|^2\Big|^p\Big] \\
&\leq 4^p \mathbb{E}\Big[|X_t^{i,N} - X_t^{j,N}|^{2p}\Big]_{i\neq j} + 4^p \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N |X_t^{j,N}|^2\Big|^p\Big].
\end{aligned}
\tag{A.1}
$$

For the first term in (A.1), by Itô's formula, for $i, j \in [\![1, N]\!], i \neq j$,

$$
\begin{aligned}
|X_t^{i,N} - X_t^{j,N}|^{2p} &= |X_0^{i,N} - X_0^{j,N}|^{2p} \\
&\quad + 2p \int_0^t |X_s^{i,N} - X_s^{j,N}|^{2p-2} \Big\langle X_s^{i,N} - X_s^{j,N}, v(X_s^{i,N}, \mu_s^{X,N}) - v(X_s^{j,N}, \mu_s^{X,N}) \Big\rangle \mathrm{d}s \\
&\quad + 2p \int_0^t |X_s^{i,N} - X_s^{j,N}|^{2p-2} \Big\langle X_s^{i,N} - X_s^{j,N}, b(s, X_s^{i,N}, \mu_s^{X,N}) - b(s, X_s^{j,N}, \mu_s^{X,N}) \Big\rangle \mathrm{d}s \\
&\quad + 2p \int_0^t |X_s^{i,N} - X_s^{j,N}|^{2p-2} \Big\langle X_s^{i,N} - X_s^{j,N}, \sigma(s, X_s^{i,N}, \mu_s^{X,N}) \mathrm{d}W_s^i - \sigma(s, X_s^{j,N}, \mu_s^{X,N}) \mathrm{d}W_s^j \Big\rangle \\
&\quad + \frac{2p(2p-1)}{2} \int_0^t |X_s^{i,N} - X_s^{j,N}|^{2p-2} \Big(|\sigma(s, X_s^{i,N}, \mu_s^{X,N})|^2 + |\sigma(s, X_s^{j,N}, \mu_s^{X,N})|^2\Big) \mathrm{d}s.
\end{aligned}
$$

By Assumption 2.1, Remark 2.2, Jensen's inequality, Proposition 4.6, take expectation on both side, by the particles are identically distributed and Burkholder-Davis-Gundy (BDG) inequality, we have

$$
\mathbb{E}\big[|X_t^{i,N} - X_t^{j,N}|^{2p}\big] \leq \mathbb{E}\big[|X_0^{i,N} - X_0^{j,N}|^{2p}\big] + C \int_0^t \mathbb{E}\big[|X_s^{i,N} - X_s^{j,N}|^{2p}\big] \mathrm{d}s + C \int_0^t \mathbb{E}\big[|X_s^{i,N}|^{2p}\big] \mathrm{d}s.
$$

For the second term in (A.1), similarly, and notice that,

$$
\begin{aligned}
\frac{1}{N}\sum_{j=1}^N |X_t^{j,N}|^2 &= \frac{1}{N}\sum_{j=1}^N |X_0^{j,N}|^2 + \frac{1}{N}\sum_{j=1}^N \int_0^t \Big\langle X_s^{j,N}, v(X_s^{j,N}, \mu_s^{X,N}) \Big\rangle \mathrm{d}s + \frac{1}{2N}\sum_{j=1}^N \int_0^t |\sigma(s, X_s^{j,N}, \mu_s^{X,N})|^2 \mathrm{d}s \\
&\quad + \frac{1}{N}\sum_{j=1}^N \int_0^t \Big\langle X_s^{j,N}, b(s, X_s^{j,N}, \mu_s^{X,N}) \Big\rangle \mathrm{d}s + \frac{1}{N}\sum_{j=1}^N \int_0^t \Big\langle X_s^{j,N}, \sigma(s, X_s^{j,N}, \mu_s^{X,N}) \mathrm{d}W_s^j \Big\rangle \\
&\leq \frac{1}{N}\sum_{j=1}^N \Big(|X_0^{j,N}|^2 + \int_0^t |X_s^{j,N}|^2 \mathrm{d}s + \int_0^t \Big\langle X_s^{j,N}, \sigma(s, X_s^{j,N}, \mu_s^{X,N}) \mathrm{d}W_s^j \Big\rangle\Big) + \frac{C}{N^2}\sum_{i=1}^N \sum_{j=1}^N \int_0^t |X_s^{i,N} - X_s^{j,N}|^2 \mathrm{d}s.
\end{aligned}
$$

Take power of p on both side and expectations. By Jensen's inequality, BDG inequality, Proposition 4.6, Assumption 2.1, the Lipschitz properties on σ, we can conclude with the highest order up to $2p$, we have

$$
\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N |X_t^{j,N}|^2\Big|^p\Big] \leq C + C\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N |X_0^{j,N}|^{2p}\Big] + C \int_0^t \mathbb{E}[|X_s^{i,N}|^{2p}] \mathrm{d}s + C \int_0^t \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N |X_s^{j,N}|^2\Big|^p\Big] \mathrm{d}s,
$$

This manuscript is for review purposes only.

where we used that the particles are identically distributed to deal with the third term on the righ-hand side.

Collecting all the above results and using (A.1) again, we have

$$\mathbb{E}\big[|X_t^{i,N}|^{2p}\big] \le \mathbb{E}\big[|X_t^{i,N}-X_t^{j,N}|^{2p}\big] + \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N |X_t^{j,N}|^2\Big|^p\Big]$$
$$\le \mathbb{E}\big[|X_0^{i,N}-X_0^{j,N}|^{2p}\big] + C\mathbb{E}\big[|X_0^{i,N}|^{2p}\big] + C\int_0^t \Big(\mathbb{E}\Big[|X_s^{i,N}-X_s^{j,N}|^{2p}\Big]_{i\neq j} + \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N |X_s^{j,N}|^2\Big|^p\Big]\Big)\mathrm{d}s.$$

Grönwall's lemma delivers the final result after taking supremum over $i \in [\![1,N]\!]$ and $t\in[0,T]$. □

Appendix B. Solving the implicit equation of the SSM and a deployment of Newton's method. In this section we address solving the implicit Equation (2.6) in the SSM. We first present a general result stating the level of precision on needs to solve (2.6) such that the final convergence rate of the SSM method is preserved (e.g., Theorem 2.9 and 2.11). Proposition B.2 is understood as a requirement of an adequate approximation method. In the subsequent section, we describe a deployment of Newton's method as one such method (among many) with the simulation results in Section 3 showing its efficiency.

B.1. Approximation scheme to the SSM. Recall the SSM from Definition 2.6. For any timestep $n\in[\![0,M-1]\!]$, for any particle $i\in[\![1,N]\!]$, define $\hat\Psi_i:\mathbb{R}^d\times\mathbb{R}^{Nd}\times[0,T]\to\mathbb{R}^d$ be the measurable map associating the unique solution $Y_n^{i,\star,N}$ of (2.6) to its data $\hat X_n^{i,N}$, $\hat X_n^N$ and h, i.e.,

$$\hat\Psi_i(\hat X_n^{i,N},\hat X_n^N,h)=Y_n^{i,\star,N},\qquad \hat\Psi=(\hat\Psi_1,\dots,\hat\Psi_N). \tag{B.1}$$

The existence of such a map $\hat\Psi$ is guaranteed by Lemma 4.2 (see also Proposition 4.3 and 4.4 for some of its good properties). We next introduce a version SSM of Definition 2.6 where the implicit equation is solved approximately only.

Definition B.1 (Approximation scheme to the SSM). *We follow the notation of Definition 2.6 hold. Denote the approximation mapping at each SSM step (2.6) as a measurable map $\overline\Psi_i:\mathbb{R}^d\times\mathbb{R}^{Nd}\times[0,T]\to\mathbb{R}^d$. The SSM variant is then, corresponding to (2.6)-(2.7): set $\overline X_0^{i,N}=X_0^i$ for $i\in[\![1,N]\!]$; then for all $i\in[\![1,N]\!]$ and $n\in[\![0,M-1]\!]$*

$$\overline Y_n^{i,\star,N}=\overline\Psi_i(\overline X_n^{i,N},\overline X_n^N,h),\quad \overline X_n^N=(\overline X_n^{1,N},\dots,\overline X_n^{N,N}),\quad \overline\mu_n^{Y,N}(\mathrm{d}x):=\frac{1}{N}\sum_{j=1}^N\delta_{\overline Y_n^{j,\star,N}}(\mathrm{d}x), \tag{B.2}$$

$$\overline X_{n+1}^{i,N}=\overline Y_n^{i,\star,N}+b(t_n,\overline Y_n^{i,\star,N},\overline\mu_n^{Y,N})h+\sigma(t_n,\overline Y_n^{i,\star,N},\overline\mu_n^{Y,N})\Delta W_n^i,\qquad \Delta W_n^i=W_{t_{n+1}}^i-W_{t_n}^i, \tag{B.3}$$

where for any i the map $\overline\Psi_i$ is an approximation to $\hat\Psi_i$ solving (B.1).

We emphasise that at this point, our assumption is that the maps $\overline\Psi_i$ can be found. We discuss how to find them in the next section.

Proposition B.2. *Let the assumptions of Theorem 2.10 hold. Recall the notation of Definition 2.6 and (B.1). For the $\hat\Psi_i$ and $\overline\Psi_i$ defined in (B.1) and (B.2) respectively, if $\sup_i\mathbb{E}[|\hat\Psi_i(x_i,x,h)-\overline\Psi_i(x_i,x,h)|^2]\le Ch$ for all $x=(x_1,\dots,x_N)\in L_0^2(\mathbb{R}^{Nd})$ and some constant C (independent of h,N,M but depending on T) , then*

$$\sup_{n\in[\![1,M]\!]}\ \sup_{i\in[\![1,N]\!]}\mathbb{E}\big[|\hat X_n^{i,N}-\overline X_n^{i,N}|^2\big]\le Ch. \tag{B.4}$$

The main interpretation is that as long as the implicit Equation (2.6) is solved approximately up to an accuracy of size h (the time-step increment) in L^2-norm, then the final order of convergence of the numerical scheme is preserved.

This manuscript is for review purposes only.

Proof. We proceed by induction since for all $i \in [\![1, N]\!]$, by definition, we have $\hat{X}_0^{i,N} = \overline{X}_0^{i,N} = X_0^i$.

Step: The initial case. We prove that $\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{X}_1^{i,N} - \overline{X}_1^{i,N}|^2\big] \le Ch$. By the assumptions of Proposition B.2 we have

$$\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|Y_0^{i,\star,N} - \overline{Y}_1^{i,\star,N}|^2\big] \le \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{\Psi}_i(X_0^i, X_0, h) - \overline{\Psi}_i(X_0^i, X_0, h)|^2\big] \le Ch.$$

For all $i \in [\![1, N]\!]$, since function b and σ are Lipschitz, by similar arguments in (4.31),

$$\begin{aligned}\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{X}_1^{i,N} - \overline{X}_1^{i,N}|^2\big] &\le C \sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[|Y_0^{i,\star,N} - \overline{Y}_1^{i,\star,N}|^2 + \big|W^{(2)}(\overline{\mu}_0^{Y,N}, \hat{\mu}_0^{Y,N})\big|^2 h\Big] \\ &\le \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|Y_0^{i,\star,N} - \overline{Y}_1^{i,\star,N}|^2\big] \le Ch. \end{aligned} \tag{B.5}$$

Step: The inductive case. For $n \in [\![1, M-1]\!]$, given $\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{X}_n^{i,N} - \overline{X}_n^{i,N}|^2\big] \le Ch$, we need to proof $\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{X}_{n+1}^{i,N} - \overline{X}_{n+1}^{i,N}|^2\big] \le Ch$, similarly, we first proof the result for the first step, from the assumption of Proposition B.2,

$$\begin{aligned}\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|Y_n^{i,\star,N} - \overline{Y}_n^{i,\star,N}|^2\big] &= \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{\Psi}_i(\hat{X}_n^i, \hat{X}_n, h) - \overline{\Psi}(\overline{X}_n^i, \overline{X}_n, h)|^2\big] \\ &\le 2 \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{\Psi}_i(\hat{X}_n^i, \hat{X}_n, h) - \hat{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h)|^2\big] + 2 \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h) - \overline{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h)|^2\big] \\ &\le 2 \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{\Psi}_i(\hat{X}_n^i, \hat{X}_n, h) - \hat{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h)|^2\big] + 2h. \end{aligned} \tag{B.6}$$

Recall the results in Section 4.2, the arguments in (4.19) are satisfied for all $i \in [\![1, N]\!]$, thus,

$$\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{\Psi}_i(\hat{X}_n^i, \hat{X}_n, h) - \hat{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h)|^2 \le \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{X}_n^i - \overline{X}_n^i|^2(1 + Ch) \le Ch.$$

Plug the result above into (B.6) to conclude

$$\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|Y_n^{i,\star,N} - \overline{Y}_n^{i,\star,N}|^2\big] \le Ch.$$

And, by similar argument in (B.5), we have

$$\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{X}_{n+1}^{i,N} - \overline{X}_{n+1}^{i,N}|^2\big] \le Ch. \qquad \square$$

B.2. Deploying Newton's method. We now provide a discussion on using Newton's method to solve (2.6) in the scope of the SSM. We first introduce Newton's method for high dimensions. Recall the functions V, u, f in (1.4), (2.2), and the SSM in Definition 2.6.

For simplicity of presentation, we assume that the function u only depends on the space-components (this is inline with the numerical examples section) and f has continuous second order derivative. Fix $x \in \mathbb{R}^{Nd}$, for $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$, for the functions $V, F : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$ and $u, f : \mathbb{R}^d \to \mathbb{R}^d$, we want to find a solution of $y \mapsto F(y)$ (given by (2.6)) defined as

$$\mathbb{R}^{Nd} \ni y \mapsto F(y) = y - x - hV(y) = 0, \quad V = (V_1, V_2, \ldots, V_N) \quad \text{and} \quad V_i(y) = u(y_i) + \frac{1}{N}\sum_{j=1}^{N} f(y_i - y_j).$$

This manuscript is for review purposes only.

For a fixed $x \in \mathbb{R}^{Nd}$, Lemma 4.2 ensures that a unique $y^\star$ exists satisfying $F(y^\star) = 0$. Setting as initial guess of $y^0 = x$, we denote the κ^{th}-iteration of the Newton method by y^κ and define it as

$$y^0 = x, \quad y^{\kappa+1} = y^\kappa - [\nabla F]^{-1}(y^\kappa) F(y^\kappa),$$

where ∇F stands for the Jacobian matrix of F.

Denoting I_{Nd} as the identity matrix in Nd-dimensions, we express the Jacobian of F in closed form as

$$[\nabla F](y) = I_{Nd} - hA(y) + \frac{h}{N}\Gamma(y) \quad \text{where for } y = (y_1, y_2, \dots, y_N) \in (\mathbb{R}^d)^N \text{ we have}$$

$$A(y) = \begin{bmatrix} \nabla u(y_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla u(y_N) \end{bmatrix} + \begin{bmatrix} \frac{1}{N}\sum_{j=1}^N \nabla f(y_1 - y_j) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{N}\sum_{j=1}^N \nabla f(y_N - y_j) \end{bmatrix}$$

$$\Gamma(y) = \begin{bmatrix} \nabla f(y_1 - y_1) & \cdots & \nabla f(y_1 - y_n) \\ \vdots & \ddots & \vdots \\ \nabla f(y_n - y_1) & \cdots & \nabla f(y_n - y_n) \end{bmatrix}.$$

The matrix $A(y)$ is a block diagonal matrix, and Γ is a symmetric matrix since f is odd and its main diagonal is equal to $\nabla f(\mathbf{0})$. We stop the Newton's iteration at step κ when the error tolerance rule $\|y^\kappa - y^{\kappa-1}\|_\infty < \sqrt{h}$ is satisfied. We note that since $\Gamma(\cdot)$ is a symmetric matrix weighted by $\frac{h}{N}$ which is an order $1/N$ smaller that I_{Nd} and $hA(\cdot)$ one can think of ignoring it in favour of an approximate Newton's method.

Theoretical foundation for methodological choices. As mentioned, Lemma 4.2 ensures a unique $y^\star$ exists solving $F(y^\star) = 0$. Proposition 4.3 and 4.4 ensure continuous dependence of $y^\star$ on x, and hence assuming h small enough the choice of $y^0 = x$ as the initial guess for $y^\star$ in the Newton method is justified. From [54, Theorem 4.4], under the extra assumption that F is twice differentiable with continuous derivatives, we have that the Newton iteration converges quadratically to the unique solution $y^\star$. In fact, given h small enough and complementing with the trick highlighted in Remark 2.7 one can show that V in (2.2) has a strictly negative one-sided Lipschitz constant and hence ∇V is strict negative definite matrix (see [44]) and hence so is ∇F – this ensures that ∇F is nonsingular (also at $y^\star$) and thus [54, Theorem 4.4] applies guaranteeing convergence.

In the scope of the examples presented in Section 3, with the choices above, we found that the condition $\|y^\kappa - y^{\kappa-1}\|_\infty < \sqrt{h}$ is attained within two to four Newton method iterations, i.e., with $\kappa \leq 4$.

This manuscript is for review purposes only.

References.

[1] D. Adams, G. dos Reis, R. Ravaille, W. Salkeld, and J. Tugaut. Large deviations and exit-times for reflected McKean-Vlasov equations with self-stabilising terms and superlinear drifts. *Stochastic Process. Appl.*, 146:264–310, 2022.

[2] D. Adams, M. H. Duong, and G. dos Reis. Entropic regularisation of non-gradient systems. *SIAM Journal on Mathematical Analysis*, 2022.

[3] D. Adams, M. H. Duong, and G. dos Reis. Operator-splitting schemes for degenerate, non-local, conservative-dissipative systems. *Discrete and Continuous Dynamical Systems*, 0:–, 2022.

[4] A. Agarwal and S. Pagliarani. A Fourier-based Picard-iteration approach for a class of McKean-Vlasov SDEs with Lévy jumps. *Stochastics*, 93(4):592–624, 2021.

[5] J. Baladron, D. Fasoli, O. Faugeras, and J. Touboul. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *J. Math. Neurosci.*, 2:Art. 10, 50, 2012.

[6] C. Beck, M. Hutzenthaler, A. Jentzen, and E. Magnani. Full history recursive multilevel Picard approximations for ordinary differential equations with expectations. *arXiv e-prints*, pages arXiv–2103, 2021.

[7] D. Belomestny, V. Pilipauskaitė, and M. Podolskij. Semiparametric estimation of McKean-Vlasov SDEs. *arXiv preprint arXiv:2107.00539*, 2021.

[8] D. Belomestny and J. Schoenmakers. Projected particle methods for solving McKean-Vlasov stochastic differential equations. *SIAM J. Numer. Anal.*, 56(6):3169–3195, 2018.

[9] W.-J. Beyn, E. Isaak, and R. Kruse. Stochastic C-stability and B-consistency of explicit and implicit Euler-type schemes. *J. Sci. Comput.*, 67(3):955–987, 2016.

[10] F. Bolley, J. A. Cañizo, and J. A. Carrillo. Stochastic mean-field limit: non-Lipschitz forces and swarming. *Math. Models Methods Appl. Sci.*, 21(11):2179–2210, 2011.

[11] M. Bossy, O. Faugeras, and D. Talay. Clarification and complement to “Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons”. *J. Math. Neurosci.*, 5:Art. 19, 23, 2015.

[12] M. Bossy and D. Talay. A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Math. Comp.*, 66(217):157–192, 1997.

[13] J. A. Carrillo, Y.-P. Choi, and O. Tse. Convergence to equilibrium in Wasserstein distance for damped Euler equations with interaction forces. *Comm. Math. Phys.*, 365(1):329–361, 2019.

[14] J. A. Carrillo, M. Fornasier, G. Toscani, and F. Vecil. Particle, kinetic, and hydrodynamic models of swarming. In *Mathematical modeling of collective behavior in socio-economic and life sciences*, Model. Simul. Sci. Eng. Technol., pages 297–336. Birkhäuser Boston, Boston, MA, 2010.

[15] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields*, 140(1-2):19–40, 2008.

[16] L.-P. Chaintron and A. Diez. Propagation of chaos: a review of models, methods and applications. *arXiv preprint arXiv:2106.14812*, 2021.

[17] X. Chen and G. dos Reis. A flexible split-step scheme for solving McKean-Vlasov stochastic differential equations. *Appl. Math. Comput.*, 427:Paper No. 127180, 23, 2022.

[18] F. Comte and V. Genon-Catalot. Nonparametric adaptive estimation for interacting particle systems. *HAL Id: hal-03696877*, 2022.

[19] D. Crisan and E. McMurray. Cubature on Wiener space for McKean-Vlasov SDEs with smooth scalar interaction. *Ann. Appl. Probab.*, 29(1):130–177, 2019.

[20] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: a central limit theorem. *Electron. J. Probab.*, 24:Paper No. 51, 54, 2019.

[21] G. dos Reis, S. Engelhardt, and G. Smith. Simulation of McKean-Vlasov SDEs with super-linear growth. *IMA J. Numer. Anal.*, 42(1):874–922, 2022.

[22] G. dos Reis, W. Salkeld, and J. Tugaut. Freidlin-Wentzell LDP in path space for McKean-Vlasov equations and the functional iterated logarithm law. *Ann. Appl. Probab.*, 29(3):1487–1540, 2019.

[23] G. dos Reis, G. Smith, and P. Tankov. Importance sampling for McKean-Vlasov SDEs. *arXiv preprint*

This manuscript is for review purposes only.

arXiv:1803.09320, 2018.
[24] W. Dreyer, M. Gaberšček, C. Guhlke, R. Huth, and J. Jamnik. Phase transition in a rechargeable lithium battery. *European J. Appl. Math.*, 22(3):267–290, 2011.
[25] N. Fournier, M. Hauray, and S. Mischler. Propagation of chaos for the 2D viscous vortex model. *J. Eur. Math. Soc. (JEMS)*, 16(7):1423–1466, 2014.
[26] V. Genon-Catalot and C. Larédo. Probabilistic properties and parametric inference of small variance nonlinear self-stabilizing stochastic differential equations. *Stochastic Process. Appl.*, 142:513–548, 2021.
[27] K. Giesecke, G. Schwenkler, and J. A. Sirignano. Inference for large financial systems. *Mathematical Finance*, 30(1):3–46, 2020.
[28] E. Gobet and S. Pagliarani. Analytical approximations of non-linear SDEs of McKean-Vlasov type. *J. Math. Anal. Appl.*, 466(1):71–106, 2018.
[29] C. Guhlke, P. Gajewski, M. Maurelli, P. K. Friz, and W. Dreyer. Stochastic many-particle model for LFP electrodes. *Contin. Mech. Thermodyn.*, 30(3):593–628, 2018.
[30] F. Harang and A. Mayorcas. Pathwise regularisation of singular interacting particle systems and their mean field limits. *arXiv preprint arXiv:2010.15517*, 2020.
[31] S. Herrmann, P. Imkeller, and D. Peithmann. Large deviations and a Kramers' type law for self-stabilizing diffusions. *Ann. Appl. Probab.*, 18(4):1379–1423, 2008.
[32] S. Herrmann and J. Tugaut. Self-stabilizing processes: uniqueness problem for stationary measures and convergence rate in the small-noise limit. *ESAIM Probab. Stat.*, 16:277–305, 2012.
[33] D. D. Holm and V. Putkaradze. Formation of clumps and patches in self-aggregation of finite-size particles. *Phys. D*, 220(2):183–196, 2006.
[34] X. Huang, P. Ren, and F.-Y. Wang. Distribution dependent stochastic differential equations. *Front. Math. China*, 16(2):257–301, 2021.
[35] M. Hutzenthaler and A. Jentzen. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *Mem. Amer. Math. Soc.*, 236(1112):v+99, 2015.
[36] M. Hutzenthaler, T. Kruse, and T. A. Nguyen. Multilevel Picard approximations for McKean-Vlasov stochastic differential equations. *J. Math. Anal. Appl.*, 507(1):Paper No. 125761, 14, 2022.
[37] S. Jin, L. Li, and J.-G. Liu. Random batch methods (RBM) for interacting particle systems. *J. Comput. Phys.*, 400:108877, 30, 2020.
[38] T. Kolokolnikov, J. A. Carrillo, A. Bertozzi, R. Fetecau, and M. Lewis. Emergent behaviour in multi-particle systems with non-local interactions [Editorial]. *Phys. D*, 260:1–4, 2013.
[39] C. Kumar and Neelima. On explicit Milstein-type scheme for McKean-Vlasov stochastic differential equations with super-linear drift coefficient. *Electron. J. Probab.*, 26:Paper No. 111, 32, 2021.
[40] D. Lacker. On a strong form of propagation of chaos for McKean-Vlasov equations. *Electron. Commun. Probab.*, 23:Paper No. 45, 11, 2018.
[41] D. Lacker. Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions. *arXiv e-prints*, pages arXiv–2105, 2021.
[42] D.-H. Li and W. Cheng. Recent progress in the global convergence of quasi-Newton methods for nonlinear equations. *Hokkaido Math. J.*, 36(4):729–743, 2007.
[43] L. Li, J.-G. Liu, and P. Yu. On the mean field limit for Brownian particles with Coulomb interaction in 3D. *J. Math. Phys.*, 60(11):111501, 34, 2019.
[44] A. Lionnet, G. dos Reis, and L. Szpruch. Time discretization of FBSDE with polynomial growth drivers and reaction-diffusion PDEs. *Ann. Appl. Probab.*, 25(5):2563–2625, 2015.
[45] F. Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, 13(2):540–560, 2003.
[46] F. Malrieu and D. Talay. Concentration inequalities for Euler schemes. In *Monte Carlo and quasi-Monte Carlo methods 2004*, pages 355–371. Springer, Berlin, 2006.
[47] X. Mao. *Stochastic differential equations and applications*. Horwood Publishing Limited, Chichester, second edition, 2008.

This manuscript is for review purposes only.

[48] C. Marchioro and M. Pulvirenti. *Mathematical theory of incompressible nonviscous fluids*, volume 96 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1994.

[49] J. M. Martınez. Practical quasi-Newton methods for solving nonlinear systems. *Journal of computational and Applied Mathematics*, 124(1-2):97–121, 2000.

[50] H. P. McKean, Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. U.S.A.*, 56:1907–1911, 1966.

[51] H. G. Othmer and A. Stevens. Aggregation, blowup, and collapse: the ABCs of taxis in reinforced random walks. *SIAM J. Appl. Math.*, 57(4):1044–1081, 1997.

[52] C. Reisinger and W. Stockinger. An adaptive Euler-Maruyama scheme for McKean-Vlasov SDEs with super-linear growth and application to the mean-field FitzHugh-Nagumo model. *J. Comput. Appl. Math.*, 400:Paper No. 113725, 23, 2022.

[53] M. V. Solodov and B. F. Svaiter. A globally convergent inexact Newton method for systems of monotone equations. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*, volume 22 of *Appl. Optim.*, pages 355–369. Kluwer Acad. Publ., Dordrecht, 1999.

[54] E. Süli and D. F. Mayers. *An introduction to numerical analysis*. Cambridge University Press, Cambridge, 2003.

[55] A.-S. Sznitman. Topics in propagation of chaos. *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989*, pages 165–251, 1991.

[56] D. Talay and O. Vaillant. A stochastic particle method with random weights for the computation of statistical solutions of McKean-Vlasov equations. *Ann. Appl. Probab.*, 13(1):140–180, 2003.

[57] J. Tugaut. Convergence to the equilibria for self-stabilizing processes in double-well landscape. *Ann. Probab.*, 41(3A):1427–1460, 2013.

This manuscript is for review purposes only.