Theory and simulation of Interacting particle systems and McKean-Vlasov processes: the super measure class, ergodicity and weak error

Xingyuan Chen

Doctor of Philosophy
University of Edinburgh
May 2024

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text, and that the work has not been submitted for any other degree or professional qualification except as specified.

(*Xingyuan Chen*)

To my parents and my friends.

人生天地之間，若白駒之過郤，忽然而已。

注然勃然，莫不出焉；油然漻然，莫不入焉。

——《庄子·知北游》

Abstract

This thesis is divided neatly into four collections of results.

In the first and second parts, we present an implicit Split-Step explicit Euler-type Method (SSM) for the simulation of McKean-Vlasov Stochastic Differential Equations (MV-SDEs) with drifts of super-linear growth in space and then super-linear growth in measure and non-constant Lipschitz diffusion coefficient. In the case that super-linear only happens in space, we prove a classical $1/2$ root mean square error (rMSE) convergence rate, for which other schemes have competitive results, but in the case that super-linear happens in space and measure (in convolution form), the SSM has a near-optimal classical (path-space) rMSE rate of $1/2 - \epsilon$ for $\epsilon > 0$ and an optimal rate $1/2$ in the non-path-space MSE, for which the result is new and no other scheme has been proven to work yet. The results are published in [47] and [48].

In the third part, we study a class of MV-SDEs with drifts and diffusions having super-linear growth in measure and space – the maps have general polynomial form but also satisfy a certain monotonicity condition. The combination of the drift's super-linear growth in measure (by way of a convolution) and the super-linear growth in space and measure of the diffusion coefficient requires novel technical elements in order to obtain the main results. We establish wellposedness, propagation of chaos (PoC). Further, we prove the SSM work in this class of MV-SDEs and attain a rate of $1/2$ in the non-path-space MSE, and under further assumptions on the model parameters, we show an exponential ergodicity property for the numerical scheme. The result is published in arXiv [49] and submitted to EJP.

In the fourth part, We study a non-Markovian Euler-type scheme with the same computational cost as the Euler scheme, for the approximation of the ergodic distribution of a one-dimensional McKean–Vlasov Stochastic Differential Equation (MV-SDE) under an assumption of strong convexity (finite-time results are also established). Based on a careful analysis of the variational processes and the backward Kolmogorov equation for the particle system associated to the MV-SDE, we show that the method attains a higher-order approximation accuracy in the long-time limit (weak convergence rate is $3/2$) than the standard Euler method (of weak order 1). While we use an interacting particle system (IPS) to approximate the MV-SDE, we show the convergence rate is independent of the dimension (N) of the IPS and this includes establishing uniform in time-decay estimates for moments of the IPS, the Kolmogorov equation and their derivatives. We establish several interesting results on the higher-order variation processes of the IPS which are of independent interest. The result will be published. The result submitted to EJP.

We provide different numerical tests of the theory and results for each part.

Lay summary

This thesis focuses on Mckean-Vlasov Stochastic Differential Equations (MV-SDEs), which have become increasingly prevalent due to advancements in computational power in the 21st century. One of the common ways to study MV-SDEs is to develop the corresponding interacting particle system and apply Monte Carlo simulation methods.

There are different challenges encountered during simulation such as the divergence of the traditional Euler method when dealing with MV-SDEs featuring a super-linear growth component. Another interesting challenge is to develop a more efficient (better convergence rate under the same computation) numerical method.

To investigate these challenges, we analyze the MV-SDES with different simulation methods and demonstrate theoretical results with different numerical examples.

Acknowledgements

First of all, I would like to thank my advisor, Gonçalo dos Reis, for his unwavering support, infinite patience, and inspirational guidance throughout my years of pursuing a Ph.D. It is through his tremendous help and encouragement that make me survived my PhD and finished this thesis.

Moreover, I would like to express my sincere gratitude to Wolfgang Stockinger, for his valuable comments and suggestions during my PhD journey. I am truly grateful for the opportunity to learn from him.

I extend my sincere gratitude to Francisco Bernal for his patience, trust, and invaluable contributions.

I am grateful to all academic and administration staff in the School of Mathematics who have kindly provided help and guidance during my PhD. Without them, it would be impossible for me to do my PhD in these fantastic three and half years.

On a more personal note, I would like to thank my parents (Kunlun Chen and Xiaoxia Huang), for their love and support. I am really fortunate to have parents like them. Their love and understanding without words encourage me a lot. I shall express my gratitude in Chinese:

我在新冠疫情间离家多年，在爱丁堡完成学业，经历有苦有乐皆人之常情，确实是有趣的经历，其中父母对我的信任和支持超越了语言可描述的范围，感谢他们，祝福他们，我将尽力分享传播最宝贵的知识以报答他们的恩情。

I also thank my friends who have given me unforgettable help, support, and spiritual encouragement during my PhD. Thanks to all sentient beings.

Contents

Chapter 1

Introduction

This thesis study on the numerical analysis and theory of the McKean-Vlasov Stochastic Differential Equations (MV-SDEs) mainly in two parts, the first part is on different types of super-linearity, where we mainly introduce a new favorable numerical method and shows the convergence in a strong sense. The second part focuses on the weak convergence analysis of a non-Markovian type numerical scheme for MV-SDEs with strong ergodicity under Lipschitz parameters, we prove some interesting results by studying the corresponding Fokker–Planck operators and high-order variation processes.

MV-SDEs differ from standard SDEs by means of the presence of the law of the solution process in the coefficients and their dynamics are of the following type

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

where μ_t^X denotes the law of the solution process X at time t, W is a multidimensional Brownian motion, b, σ are measurable maps and X_0 is a sufficiently integrable initial condition (in L_0^m for $m \geq 2$). These equations were introduced by McKean in the sixties [115] and have been the target of much research since. There is a rich literature on well-posedness [15, 121, 127, 133] and we point to the recent summary work [87] highlighting recent developments in regularity estimates, exponential ergodicity, long time large deviations (also [2, 61]), comparison Theorems and the Vlasov-Fokker-Plank equations associated to MV-SDEs. One particular element of interest is the so-called propagation of chaos (PoC) introduced by Kac [93] and further studied in the MV-SDE literature [100, 101, 118, 133]. The PoC phenomena states that an MV-SDE is the limit of certain weakly interacting particle systems (of standard SDEs) as the system's size increases to infinity. Namely, (2.1) is the limit, as $N \to \infty$, of the N-dimensional system of $\mathbb{R}^d$-valued interacting particles $X^{i,N}$,

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

with $\mu_t^{X,N}$ being the empirical measure given as $\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}(\mathrm{d}x)$ and $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, W^i are independent Brownian motions and with independent and identically distributed initial conditions X_0^i across $i = 1, \cdots, N$. In essence, the law μ_t^X in $\mathbb{R}^d$ of (2.1) is approximated by the empirical average $\mu_t^{X,N}$ generated by the $(\mathbb{R}^d)^N$-system (a high-dimensional system). This methodology, appealing to the interacting particle system, includes a well-known quantified speed of convergence result (summarised in Proposition 2.4 below) providing a path for a numerical approximation: from X to $\{X^{i,N}\}_i$ to $\{X^{i,N,\pi}\}_i$ with $X^{\cdot,N,\pi}$ the numerical approximation of each particle in the SDE system.

It is essential to highlight that although $\{X^{i,N}\}_i$ is a high-dimensional SDE and any existing numerical method a priori applies straightforwardly, all rates and results obtained in this straightforwardly way depend on the total system's dimension Nd. The constants then explode as N increases. Part of the difficulty of the method is to show that such constants, rates, and results are independent of N albeit depending on d. This step has its non-trivial difficulties which we discuss further below by drawing on prior work [60, 103].

1.1 Structure of the thesis

In the following chapters, we will study MV-SDEs 1.1 under different assumptions, we focus on numerical study and some intersting properties. We present brief information below, detailed introductions are provided in each corresponding chapter.

In Chapter 2, we study the MV-SDEs with drifts of superlinear growth in their spatial components with other terms that satisfy Lipschitz constraints, this is our starting point of studying superlinearity, we introduce and analyze the split-step method (SSM) for simulation and show several examples to verify our theoretical results. The result of this chapter is published in [47].

In Chapter 3, we continue to study the MV-SDEs with drifts of superlinear growth in their spatial and measure components(of convolution type), the diffusion term satisfies Lipschitz constraints. This is our second study on superlinearity and we again show the SSM converges which is the only proven-to-work scheme for this type of MV-SDEs until now. We have tested several examples in some extreme setups to verify our theoretical results. The result of this chapter is published in [48].

In Chapter 4, we provide the study the MV-SDEs with drifts and measure of superlinear growth in their spatial and measure components (of convolution type), this is the fully extension of superlinearity based on previous chapter. In this chapter, we first show the well-posedness results, then we provide the corresponding numerical simulation method using an B&C-method approach and show a ergodicity result for the numerical method. The result of this chapter is published in arXiv [49] and submitted to EJP.

In Chapter 5, comparing to the previous study on superlinearity and the strong error of the numerical method, we study the weak convergence to one-dimensional Lipschitz type MV-SDE of under an assumption of strong convexity. Based on the Kolmogorov backward equation, we analyze different types of variation processes and provide interesting techniques on the summation over different combination of stochastic processes. The result of this chapter is published in arXiv [50] and submitted to EJP.

1.2 Notation and spaces

Let $\mathbb{N} = \{1, 2, \cdots\}$ be the set of natural numbers starting at 1 and $\mathbb{R}$ denotes the real numbers where $\mathbb{R}^+ = [0, \infty)$. Also, we denote $[\![a, b]\!] := [a, b] \cap \mathbb{N} = \{a, \cdots, b\}$, for any $a, b \in \mathbb{N}$ with $a \le b$. For $x, y \in \mathbb{R}^d$ denote the scalar product of vectors by $\langle x, y\rangle$; and the Euclidean distance of x is $|x| = (\sum_{j=1}^d x_j^2)^{1/2}$. The indicator function of a set Ω is denoted as $\mathbb{1}_\Omega$. 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. For a twice continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, we denote by $\partial_{x_i} f : \mathbb{R}^d \to \mathbb{R}$ the partial derivative with respect to the i-th component, and by $\nabla f : \mathbb{R}^d \to \mathbb{R}^d$ its gradient $\nabla f = (\partial_{x_1} f, \dots, \partial_{x_d} f)$, and by $\nabla^2 f : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ its Hessian. For a multi-index $\ell = (\ell_1, \dots, \ell_d) \in \mathbb{N}^d$, we denote higher-order derivatives as

$$\partial^d_{x_{\ell_1},\dots,x_{\ell_d}} f.$$

The supnorm of f will be denoted by $|f|_\infty := \sup_{\boldsymbol{x}\in\mathbb{R}^d} |f(\boldsymbol{x})|$.

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

$$W^{(2)}(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \Big(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^2 \pi(\mathrm{d}x, \mathrm{d}y)\Big)^{\frac{1}{2}}, \quad \mu, \nu \in \mathcal{P}_2(\mathbb{R}^d). \tag{1.2}$$

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 [47, 60]). 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, constants C, K denote 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 C may depend on the terminal time T (and other fixed problem data) while K is independent of T.

1.3 Useful inequalities

Lemma 1.3.1. *(Hölder's inequality). For any real-valued random variables X and Y defined on the probability space $(\Omega,\mathbb{F},\mathbb{P})$, we have*

$$\mathbb{E}[XY]\le\Big(\mathbb{E}[|X|^p]\Big)^{\frac{1}{p}}\Big(\mathbb{E}[|Y|^q]\Big)^{\frac{1}{q}},$$

where $1/p+1/q=1,\ p,q>0$.

Lemma 1.3.2. *(Jensen's inequality). For any real-valued integrable random variables $X\in\mathbb{R}$ defined on the probability space $(\Omega,\mathbb{F},\mathbb{P})$ and real-valued convex function $f:\mathbb{R}\to\mathbb{R}$, we have*

$$f(\mathbb{E}[X])\le\mathbb{E}[f(X)].$$

Lemma 1.3.3. *(Cauchy–Schwarz inequality). For any real-valued random variables X and Y defined on the probability space $(\Omega,\mathbb{F},\mathbb{P})$, we have*

$$|\mathbb{E}[XY]|\le\sqrt{\mathbb{E}[|X|^2]\ \mathbb{E}[|Y|^2]}\ .$$

Lemma 1.3.4. *(Young's inequality). For any $a,b>0$ with $p>1$, we have*

$$ab\le\frac{p-1}{p}a^{\frac{p}{p-1}}+\frac{1}{p}b^p.$$

Lemma 1.3.5. *(Burkholder-Davis-Gundy inequality). For any $1\le p\le 1$ there exists some positive constants $C_{1,p}, C_{2,p}$ such that for all local martingales $\{M\}_t$ with $M_0=0$, we have*

$$C_{1,p}\mathbb{E}\big[|\langle M_t\rangle|^{p/2}\big]\le\mathbb{E}\big[\sup_{s\in[0,t]}|M_s|^{p/2}\big]\le C_{2,p}\mathbb{E}\big[|\langle M_t\rangle|^{p/2}\big],$$

where $\langle M_t\rangle$ denotes the quadratic variation of M_t.

Lemma 1.3.6. *(Gronwall's inequality). Let $T>0$ and let α,β and u be real-valued functions defined on $[0,T]$. Assume that α and u are continuous and that the negative part of β is integrable on every closed and bounded subinterval of $[0,T]$. If α is non-negative and if u satisfies the integral inequality*

$$u(t)\le\beta(t)+\int_0^t\alpha(s)u(s)\mathrm{d}s,\quad\forall t\in[0,T],$$

then

$$u(t)\le\beta(t)+\int_0^t\alpha(s)\beta(s)\exp\left(\int_s^t\alpha(r)\mathrm{d}r\right)\mathrm{d}s,\quad\forall t\in[0,T].$$

If we further have that β is non-decreasing, then

$$u(t)\le\beta(t)\exp\left(\int_0^t\alpha(s)\mathrm{d}s\right),\quad\forall t\in[0,T].$$

Chapter 2

Numerical analysis for McKean-Vlasov SDEs with super-linear growth drifts in space

2.1 Introduction

In this chapter, we study MV-SDEs of the following type

$$\mathrm{d}X_t = \Big(v(t, X_t, \mu_t^X) + b(t, X_t, \mu_t^X)\Big)\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^m(\mathbb{R}^d). \tag{2.1}$$

where μ_t^X denotes the law of the solution process X at time t, W is a multidimensional Brownian motion, v, b, σ are $\mathbb{F}$-adapted maps and X_0 is a sufficiently integrable initial condition (in L_0^m for $m \geq 2$).

The corresponding N-dimensional system of $\mathbb{R}^d$-valued interacting particles $X^{i,N}$ satisfy

$$\mathrm{d}X_t^{i,N} = \Big(v(t, X_t^{i,N}, \mu_t^{X,N}) + b(t, X_t^{i,N}, \mu_t^{X,N})\Big)\mathrm{d}t + \sigma(t, X_t^{i,N}, \mu_t^{X,N})\mathrm{d}W_t^i, \qquad X_0^{i,N} = X_0^i. \tag{2.2}$$

with $\mu_t^{X,N}$ being the empirical measure given as $\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}(\mathrm{d}x)$ and $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, W^i are independent Brownian motions and with independent and identically distributed initial conditions X_0^i across $i = 1, \cdots, N$.

In this chapter, we focus on the class of MV-SDE with drifts of superlinear growth in their spatial components [14, 60, 61, 98, 125], encapsulated in the function v, where b, σ are uniformly Lipschitz (in space and measure), and σ is non-constant – this structural Assumption on b, v is trivial from the theoretical perspective but plays an important role in the numerics. This class of MV-SDEs appears in several practical models in science, for example, in neuroscience [9, 27] introduce the mean-field FitzHugh-Nagumo model for a neuron networks in the brain; [25] discuss individual-based and swarming Cucker-Smale interaction models; and models of battery electrodes [63, 77]. These equations do not have explicit or closed-form solutions and numerical approximations are needed. Moreover, standard explicit numerical methods suitable for the Lipschitz case fail to converge on the superlinear growth setting. This is exemplified by the 'particle corruption' phenomena [60, Section 4.1] for MV-SDE particle system numerics. This phenomenon is akin to the divergence of SDE schemes in the superlinear growth setting as highlighted in the seminal work [89].

The numerical approximation of McKean–Vlasov equations in the continuous case was initiated in [29] and has been investigated further in several recent works. We briefly mention a few on numerical schemes for MV-SDEs under the superlinear setting and in the Brownian framework. Tamed Euler schemes appeared first [60], shortly followed by tamed Milstein schemes [13, 98] (appeared simultaneously) and Milstein schemes for delay MV-SDEs [14]. In [99] a tamed scheme is proposed (and a new wellposedness result) for MV-SDE featuring common noise in the particle system (which (2.2) does not) and where the diffusion is also allowed to grow superlinearly. Adaptive time-stepping methods come as an alternative to taming and in the context of MV-SDE they are proposed in [125] — these two methods will henceforth be referred to as the '*Taming*' and the '*Adaptive*' algorithm, respectively.

Outside the superlinear setting, [90] discusses computational complexity of MV-SDE algorithms (uniformly Lipschitz drift and constant diffusion coefficient) and we point the reader there for an in-depth overview on the state of the art in that regard. Numerical approximations for MV-SDEs with non-Lipschitz conditions in measure and space exist [59] but impose a linear growth condition on the coefficients. The case of simulating MV-SDEs with discontinuous coefficients has been addressed [104]. An alternative to the empirical measure approximation of (2.2) is to use a projection-type estimation of the marginal densities [18] where the error analysis requires differentiability of the coefficients. Variance reduction technique have been analysed for the class of MV-SDE, namely, importance sampling [62], antithetic multilevel Monte Carlo sampling [14] and antithetic sampling [20]. There also recent progress in the jump-diffusion setting [6, 123].

For the superlinear growth case described above, we are motivated by an open question left in [60] regarding implicit-type numerical methods for MV-SDEs. All methods described above are of explicit time-stepping nature which are known to lose some of the geometric properties of the original system. For instance, taming destroys the strict dissipativity of the drift map which then raises questions regarding the stability of the scheme's output across long time horizons (see [144]). Implicit methods for MV-SDEs are largely unexplored. The notable exceptions are [60, 110] which are also starting point for this chapter. In [110] the authors study MV-SDEs and associated particle systems with drifts of (symmetric) convolution kernel-type (as in [2]) and constant diffusion coefficient. The theoretical (Bacry-Emery) machinery employed there yields a critical logarithmic Sobolev inequality estimate that is used to establish a concentration inequality for their implicit Euler scheme. Assumption-wise, their setting and the setting of this chapter do not cover each other but agree over a small class. Further, it is unclear how to extend their methodology to non-constant possibly degenerate diffusion coefficients.

More recently, an implicit method to deal with the superlinear growth was proposed [60] (for general diffusion coefficients and without a concavity assumption) but convergence was shown under stronger restrictions than expected: the measure component of the drift is Lipschitz in Wasserstein-1 metric not just in Wasserstein-2 (plus uniformly bounded measure dependency). At the core of these difficulties was the use of stopping times arguments which revealed themselves difficult to handle with the measure dependency. The critical point is the proof of Lemma 5.12 (p.41) in [60] and the calculations executed in (p.48-49). Lastly, upon inspection of that proof, we emphasise that general θ-methods for SDEs [85] will face the same difficulties.

Our contributions. In this chapter, we revisit the framework of [60] and propose a split-step numerical scheme inspired by the earlier work [85]. Our contributions can be summarised as follows,

(I) *Main results.* We proposed a split-step method (SSM), see (2.5)-(2.6) below, for this class of MV-SDEs. We prove its convergence and recover the 1/2-convergence rate in root Mean Square Error (rMSE) under the same general assumptions as Taming [60] or Adaptive [125]. No differentability or non-degeneracy assumptions are imposed and stopping-time arguments are fully avoided.

We provide the stability analysis of Mean-square contractivity for the SSM (non-constant diffusion coefficient). To the best of our knowledge this has not yet been discussed for MV-SDE schemes in general. The stability of the SSM provides a theoretical foundation for carrying out simulation with larger timestep and we point to positive results by way of numerical simulation with the Cucker-Smale flocking model [67] where the SSM outperforms both Taming [60] and Adaptive [125] algorithms.

In regards to known findings, the SSM here overcomes the limitations of the implicit method proposed in [60, Section 3.2] and extends its scope of application. In the context of standard SDE simulation (where the coefficients are independent of the measure), our results lift the differentiability restriction of [85, Assumption 3.1], allow for time-dependence and can take advantage of (the possible) concavity of the map v (in (2.1)) – this is a mild improvement of known SDE results.

(II) *Computational gains.* The scheme is designed so that the superlinear term can be split from the main equation in a way that optimises/minimises the computational cost of the inversion method (from the implicit component). This flexibility is understood in the following way: given an MV-SDE it is left to the user to choose which terms form v and which terms form b in (2.1) (within restrictions). Moreover, since there is a split v Vs b the user may decide to add & subtract convenient terms to the drift (see Section 2.3.4 below; also [32, Eqs. (37) and (38)]) – this trick allows one to transform a non-dissipative term $x \mapsto v(\cdot, x, \cdot)$ into a dissipative one at the expense of an increase of the Lipschitz constant of b.

The SSM proposed allows to decouple the measure component making it amenable to a parallel implementation (e.g., [112]). Namely, one parallelizes the task of solving N-times an $\mathbb{R}^d$-system in opposition to the non-parallelizable task of solving once the $\mathbb{R}^{dN}$-system. The computational gains of the parallel implementations for the SMM are shown to be on par with Taming [60] and Adaptive [125] algorithms.

(III) *Comparative study against known literature.* We provide a comparative analysis against the Taming [60] and Adaptive [125] methods, across parallel and non-parallel implementations. The numerical study covers four examples of interest, highlighting a different flavour of these 3 algorithms including stability experiments.

We close this introduction with two comments. The general setting of [14, 60, 61, 98, 125] is based on drifts maps $(t, x, \mu) \mapsto \widehat{b}(t, x, \mu)$ and it is this map that satisfies a one-sided Lipschitz condition in space and a uniform Lipschitz condition in measure. In (2.1), we specify $\widehat{b}$ to be $(t, x, \mu) \mapsto \widehat{b}(t, x, \mu) = v(t, x, \mu) + b(t, x, \mu)$ where the polynomial growth is fully captured by v while b remains uniformly Lipschitz in its variables. Separating $\widehat{b}$ into v and b is natural non-limiting assumption.

Settings outside the scope of this chapter are: non-Lipschitz measure dependency [2, 110], the superlinear diffusion case of [99] (also [125, Section 3.1]) or jumps [123]. Weak error analysis is left unaddressed.

Organisation of this chapter In section 2, notations and necessary concepts for this chapter are given. In Section 2.2.2 we state the SSM, the main Theorem of convergence and the stability results for the scheme. Section 3 provides several numerical examples for comparison with other methods. Examples covering the stability analysis in the superlinear case, notably the Cucker-Smale model, are also given in Section 3. All proofs are postponed to Section 4. For convenience, 6.1 contains a short description of the Taming and Adaptive algorithms.

2.2 Main results

2.2.1 Framework

We study MV-SDE (2.1) for $m \geq 1$ and we denote the law of X at time t as μ_t^X. Take b, v, σ as measurable functions, $v : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d \times l}$. Throughout the text, we make the following assumption.

Assumption 2.2.1. *Assume that v, b and σ are 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 $L_b, L_\sigma \geq 0$ such that for all $t \in [0, T]$ all $x, x' \in \mathbb{R}^d$ and all $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ we have*

$$|b(t, x, \mu) - b(t, x', \mu')|^2 \leq L_b(|x - x'|^2 + W^{(2)}(\mu, \mu')^2),$$
$$|\sigma(t, x, \mu) - \sigma(t, x', \mu')|^2 \leq L_\sigma(|x - x'|^2 + W^{(2)}(\mu, \mu')^2).$$

For v, there exist $L_v \in \mathbb{R}$, $L_{\hat{v}} > 0$, $q \in \mathbb{N}$ and $q > 1$ such that for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$, all $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ we have

$$\langle x - x', v(t,x,\mu) - v(t,x',\mu)\rangle \le L_v |x-x'|^2, \qquad \textit{(One-sided Lipschitz in space)},$$
$$|v(t,x,\mu) - v(t,x',\mu)| \le L_{\hat{v}}(1 + |x|^q + |x'|^q)|x - x'|, \qquad \textit{(Locally Lipschitz in space)},$$
$$|v(t,x,\mu) - v(t,x,\mu')|^2 \le L_{\tilde{v}} W^{(2)}(\mu,\mu')^2, \qquad \textit{(Uniformly Lipschitz in measure)}.$$

The structural choice of having a drift $\widehat{b} = v + b$ with only v containing the superlinear growth component, as opposed to a single drift map $\widehat{b}$ in the style of [60, 125], is negligible and its use is discussed in Remark 2.2.7.

Immediate well-known properties can be derived from this assumption.

Remark 2.2.2 (Implied properties). *Under Assumption 2.2.1, define $\widehat{L}_v = 1/2 + L_v$ and $C_T = \sup_{t\in[0,T]} |v(t,0,\delta_0)|^2/2$. Let $C > 0$, then for all $t \in [0,T]$, $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ one has*

$$\begin{aligned}
&\langle x, v(t,x,\mu)\rangle = \langle x - 0, v(t,x,\mu) - v(t,0,\mu)\rangle + \langle x, v(t,0,\mu)\rangle \\
&\le L_v|x|^2 + |x||v(t,0,\mu)| \le (L_v + \frac{1}{2})|x|^2 + \frac{1}{2}|v(t,0,\mu) - v(t,0,\delta_0)|^2 + \frac{1}{2}|v(t,0,\delta_0)|^2 \\
&\le C_T + \widehat{L}_v|x|^2 + \frac{L_{\tilde{v}}}{2} W^{(2)}(\mu,\delta_0)^2.
\end{aligned}$$

where the last step follows using Young's inequality. Additionally, for $\psi \in \{b,\sigma\}$ one has

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

The above assumptions cover a larger range of models as highlighted in Section 2.3 below. This setting subsumes the standard globally Lipschitz assumptions. For further examples we point to [25, 61, 76, 111].

Theorem 2.2.3 (Theorem 3.3 in [61]). *Let Assumption 2.2.1 hold and suppose we have $X_0 \in L_0^m(\mathbb{R}^d)$ for some fixed $m \ge 2$. Then, there exists a unique solution X to the MV-SDE (2.1) and it satisfies $X \in \mathbb{S}^m([0,T])$.*

There exists a constant $C \in \mathbb{R}^+$ such that

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |X_t|^m\Big] \le C\left(1 + \mathbb{E}[\,|X_0|^m]\right) e^{CT}.$$

The interacting particle system approximation. In order to approximate of MV-SDE (2.1), we build an interacting particle system as follows:

1. For $i \in [\![1,N]\!]$, take $X_0^{i,N} = X_0^i$ as independent and identically distributed (i.i.d.) random initial condition for each particle.
2. Each particle is driven by its own independent Brownian motion W^i (all are i.i.d.)
3. We set $\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}(\mathrm{d}x)$ where δ_x is the Dirac measure at point $x \in \mathbb{R}^d$ and the dynamics of the particle system is given by (2.2).

Propagation of chaos (PoC).

Below we show a pathwise PoC result to control the difference between the original MV-SDE and the interacting particle system. For that, we introduce the auxiliary equation system of *non interacting particles*

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

This system is just N independent MV-SDEs (each in $\mathbb{R}^d$). The X^is are independent, we have $\mu_t^{X^i} = \mu_t^X$ (for all $i \in [\![1,N]\!]$) where μ_t^X is the law of X solution to (2.1)). Under Assumption

2.2.1, we have (see [33, 100, 118, 133])

$$\lim_{N\to\infty} \sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[\sup_{0\le t\le T} |X_t^{i,N} - X_t^i|^2 \Big] = 0\,.$$

By showing the following *propagation of chaos* result, we can connect in a quantifiable manner the MV-SDE and the interacting particle system.

The proof can be found in [60, Theorem 3.1] (and [60, Proposition 5.1] for the well-posedness of the particle system (2.2)).

Proposition 2.2.4 (Propagation of chaos)**.** *Let Assumption 2.2.1 hold and suppose we have for some $m \ge 2$ for all $i \in [\![1,N]\!]$ that $X_0^i \in L_0^m(\mathbb{R}^d)$. Then there exists a unique solution $\{X^{i,N}\}_i$ to* (2.2) *in $\mathbb{S}^m([0,T])$ and for any $1 \le p \le m$ there exists $C_p \in \mathbb{R}^+$ such that*

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

Moreover, let $X^i \in \mathbb{S}^m$ satisfy (2.3) *and assume $m > 4$. Then, the convergence rate between MV-SDE* (2.3) *and the interacting particle system* (2.2) *is given by*

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

Under this result, one can approximate MV-SDEs through particle scheme. Thus, by showing the convergence of the numerical methods to the interacting particle scheme, we can obtain the convergence rate between a numerical method and the MV-SDE.

Remark 2.2.5 (Optimising the PoC convergence rate)**.** *We highlight a result from [57] and later reviewed in [125] in the context of numerical methods for MV-SDEs. The PoC rate can be improved for the case d = 4, namely the $\log(N)$ term can be omitted (under restrictions). This holds under the additional constraint of a constant diffusion coefficient σ, and a bounded drift with bounded derivatives. It does not cover the superlinear growth case here, nonetheless, it is reasonable to expect that the result can be lifted to match the drift condition in this chapter and a diffusion coefficient that is bounded (and sufficiently smooth).*

2.2.2 The split-step method (SSM) for MV-SDEs: convergence and stability

The numerical scheme proposed in this chapter, and dubbed *Split-Step Method* (SSM), improves strongly on the implicit numerical scheme proposed in [60]. It is an enhanced variant of the split-step backward Euler scheme for standard SDEs [85, Eq. (3.8)-(3.9)] and here further optimised for the MV-SDE setting.

Define the uniform partition of $[0,T]$ as $\pi := \{t_n := nh : n \in [\![0,M]\!], h := T/M\}$ for a prescribed $M \in \mathbb{N}$. Define recursively the split-step method to approximate (2.2) as follows: for $i \in [\![1,N]\!]$ set $\hat{X}_0^{i,N} = X_0^i$, then, iteratively over $n \in [\![0,M-1]\!]$ for all $i \in [\![1,N]\!]$,

$$Y_n^{i,\star,N} = \hat{X}_n^{i,N} + hv\left(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{X,N}\right), \tag{2.5}$$

$$\hat{X}_{n+1}^{i,N} = Y_n^{i,\star,N} + b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})h + \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Delta W_n^i, \tag{2.6}$$

$$\text{where} \quad \hat{\mu}_n^{X,N}(dx) := \frac{1}{N}\sum_{j=1}^N \delta_{\hat{X}_n^{j,N}}(dx), \quad \hat{\mu}_n^{Y,N}(dx) := \frac{1}{N}\sum_{j=1}^N \delta_{Y_n^{j,\star,N}}(dx).$$

We state immediately the main convergence result between the continuous time extension of the scheme (2.5)-(2.6) and the solution to the MV-SDE (2.1). The proof is postponed to Section 2.4.1.

Theorem 2.2.6. *Let the assumptions of Proposition 2.2.4 hold. Let $m > 2(q+1)^2$, where q is the polynomial growth parameter of Assumption 2.2.1. Take the collection $\{\hat{X}_n^{i,N}\}_{i,n}$ for $n \in [\![0, M]\!]$, $i \in [\![1, N]\!]$ generated through the scheme* (2.5)-(2.6) *under the timestep constraint expressed through the one-sided Lipschitz constant L_v of v (Assumption 2.2.1, see Remark 2.2.7)*

$$\begin{cases} h > 0 \quad \text{and} \quad h \le \frac{1}{1+2L_v} & , \text{if } L_v > -\frac{1}{2}, \\ h > 0 & , \text{if } L_v \le -\frac{1}{2}. \end{cases}$$

Then, the following assertions hold. There exists a continuous-time extension, $(\hat{X}_t^{i,N})_{t\in[0,T]}$ to the SSM (2.5)-(2.6) *(and given in* (2.25) *below), satisfying ($C > 0$ is a constant independent of N, M but may depend on T, d):*

1. *Uniformly bounded p moments. Given $m \ge 2p \ge 1$ there exist constant $C > 0$ such that*

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

2. *Take $(X^{i,N})_i$ as the solution of the interacting particle system* (2.2)*. Then, scheme* (2.5)-(2.6) *converges to $X^{i,N}$ with a strong global convergence rate of $1/2$ in root mean square error (rMSE) over $[0, T]$, namely,*

$$\mathit{rMSE} = \sqrt{\sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[\sup_{0\le t\le T} |X_t^{i,N} - \hat{X}_t^{i,N}|^2\Big]} \le Ch^{\frac{1}{2}}. \tag{2.7}$$

3. *Let $(X^i)_i$ be the solution of the non-interacting particle system* (2.3)*. We have*

$$\sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[\sup_{0\le t\le T} |X_t^i - \hat{X}_t^{i,N}|^2\Big] \le C \begin{cases} h + N^{-1/2} & \text{if } d < 4, \\ h + N^{-1/2}\log(N) & \text{if } d = 4, \\ h + N^{-2/d} & \text{if } d > 4. \end{cases}$$

We remind the reader about Remark 2.2.5 concerning the PoC's convergence rate.

Remark 2.2.7 (The constraint on L_v is soft and removable.)**.** *The choice of what v and b are is left to the user. More precisely, to the drift map $\widehat{b}(t, x, \mu) = v(t, x, \mu) + b(t, x, \mu)$ one can always add and subtract a linear term γx ($\gamma \in \mathbb{R}$) and set $\widehat{b}(t, x, \mu) = \Big(v(t, x, \mu) - \gamma x\Big) + \Big(b(t, x, \mu) + \gamma x\Big)$. This means that the one-sided Lipschitz constant L_v becomes $L_v - \gamma$ and hence negative if γ is sufficiently large.*

We show below that this operation is not free of cost. Concretely, there is an implication in terms of the scheme's stability since for L_v to become negative the Lipschitz constant L_b increases proportionally. In Section 2.3.4 we discuss this in view of a numerical example and via a mean-square stability result we provide in Theorem 2.2.9 for the SSM (2.5)-(2.6)*.*

Lastly, to the best of our knowledge, the SSM scheme (2.5)-(2.6) *is not of the usual form SSM schemes are presented in the literature. Usually there is no structural separation of $v + b$ and one sets $b = 0$. Consequently there is no drift component in* (2.6) *(only a diffusion part), thus a discussion on the benefits/drawbacks of adding/subtracting of a γx-term seems generally absent.*

Remark 2.4.3 provides more details on the choice of h. The constraint of $L_v < -1/2$ is not sharp. In fact, it can be replaced by $L_v < -\varepsilon$ for some $\varepsilon \in (0, 1)$ at the expense of another constant growing proportionally to $1/\varepsilon$. We choose for simplicity $\varepsilon = 1/2$, see Remark 2.4.3 and the definition of $\widehat{L}_v$ in Remark 2.2.2 for further details. This issue is relevant in case one sets $b = 0$ as is usual in the SSM literature (and the trick of Remark 2.2.7 cannot be applied).

Theorem 2.2.6 shows the strong convergence rate of the SSM is 1/2 (rMSE) which is the same as Taming [60] and Adaptive [125]. Also, the complexity of the particle system is of order

N^2 in the worst situation, however, in several examples of section 2.3 the complexity for the calculation of the interaction term is of order N.

After the convergence study of Theorem 2.2.6 we introduce the notion of mean-square contractivity and study the stability of the SSM (2.5)-(2.6).

Definition 2.2.8 (Mean-square contractivity). *Suppose that $X_0 \in L_0^m(\mathbb{R}^d)$ and $Z_0 \in L_0^m(\mathbb{R}^d)$ for some sufficiently large $m \in \mathbb{N}$. Take two numerical solutions of the same numerical scheme $\hat{X}_n^{i,N}$ and $\hat{Z}_n^{i,N}$ of (2.1) with $\hat{X}_0^i$ and $\hat{Z}_0^i$ being i.i.d. copies of X_0 and Z_0 respectively. The scheme is called mean-square contractive if we have*

$$\lim_{n\to\infty} \sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[\,|\hat{X}_n^{i,N} - \hat{Z}_n^{i,N}|^2\Big] = 0.$$

The next result shows when the SSM is mean-square contractive.

Theorem 2.2.9. *Let the assumptions of Theorem 2.2.6 hold.*

Assume a further form of the Lipschitz condition of b, σ. Namely let $L_b, L_{\tilde{b}}, L_\sigma, L_{\tilde{\sigma}} \geq 0$ be such that

$$|b(t,x,\mu) - b(t,x',\mu')|^2 \leq L_b|x-x'|^2 + L_{\tilde{b}} W^{(2)}(\mu,\mu')^2,$$
$$|\sigma(t,x,\mu) - \sigma(t,x',\mu')|^2 \leq L_\sigma|x-x'|^2 + L_{\tilde{\sigma}} W^{(2)}(\mu,\mu')^2.$$

for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$. Suppose that $X_0 \in L_0^m(\mathbb{R}^d)$ and $Z_0 \in L_0^m(\mathbb{R}^d)$ for a sufficiently large $m \in \mathbb{N}$, and let $\hat{X}_0^i$ and $\hat{Z}_0^i$ be i.i.d. copies of X_0 and Z_0 respectively.

Set $h > 0$. Define two families $\{(X_n^{i,N}, Y_n^{i,\star,N})\}_{i,n}$ and $\{(Z_n^{i,N}, G_n^{i,\star,N})\}_{i,n}$ as the output of the SSM (2.5)-(2.6) w.r.t correspond empirical measure pairs $\{\hat{\mu}_n^{X,N}, \hat{\mu}_n^{Y,N}\}_n$ and $\{\hat{\mu}_n^{Z,N}, \hat{\mu}_n^{G,N}\}_n$ with input initial conditions $\{X_0^{i,N}\}_i$ and $\{Z_0^{i,N}\}_i$ respectively.

Then, for any $n \in \mathbb{N}$,

$$\sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[\,|\hat{X}_n^{i,N} - \hat{Z}_n^{i,N}|^2\Big] \leq (1+\beta h)^n \sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[|\hat{X}_0^{i,N} - \hat{Z}_0^{i,N}|^2\Big], \tag{2.8}$$

where

$$\beta = \frac{(2L_v + A + 1 + L_{\tilde{v}}) + h(L_{\tilde{v}}A + B) + h^2 B L_{\tilde{v}}}{1 - h(2L_v + 1)} \tag{2.9}$$

$$\text{and} \quad A = (2\sqrt{L_b} + 2\sqrt{L_{\tilde{b}}} + L_\sigma + L_{\tilde{\sigma}}) \quad B = L_b + L_{\tilde{b}}. \tag{2.10}$$

Under the choice of h stated in Theorem 2.2.6, the quantity $1 + \beta h$ is always positive. If $L_v < -(1 + L_{\tilde{v}} + A)/2 \leq -\frac{1}{2}$ and for a sufficient small h then $\beta < 0$ and thus the SSM is Mean-square contractive.

The proof of this result is postponed to Section 2.4.2. We illustrate immediately the scope of our findings with several numerical results. The reasoning behind the specification of the Lipschitz constants of b and σ will become apparent in the examples Section 2.3.4. The main motivation is to account for the contribution of the spatial and measure components separately as a way to explore the flexibility allowed by the scheme in choosing the drift coefficients v and b.

The equivalent definition of *Mean-square contractivity* (Definition 2.2.8) for the initial MV-SDE (2.1) is the so-called *Exponential mean-square stability inequality* defined next. We will make use of this definition in Section 2.3.4 below.

Definition 2.2.10 (Exponential mean-square contractive solutions). *Let X, Y be two solutions to (2.1) with initial conditions $X_0, Y_0 \in L_0^m(\mathbb{R}^d)$ respectively. If X, Y satisfy*

$$\mathbb{E}\Big[\,|X_t - Y_t|^2\Big] \leq e^{\eta t}\mathbb{E}\Big[|X_0 - Y_0|^2\Big]$$

for some real number $\eta < 0$, then the MV-SDE (2.1) is said to generate exponential mean-square contractive solutions.

2.3 Examples of interest

We now illustrate our numerical scheme (2.5),(2.6) through several examples of interest. Moreover, alongside the SSM simulations we also provide a comparative analysis with two other numerical schemes: Taming [60] and Adaptive timestepping [125] (the detailed forms are included in Section 6.1.1 and 6.1.2) . For convenience, Section 6.1 contains a brief description of the algorithms including convergence results and conditions.

Since the exact solution to each example is unknown we use a proxy for the true solution in order to compute approximation errors. Concretely, we compare SSM/Taming/Adaptive results with SSM/Taming/Adaptive results at a lower value of timestep h as a proxy (each method is compared with its own approximation of the true solution but at a much smaller timestep). Within each example, all the methods will use same Brownian motion paths.

We consider the weak error ($k = 1$) and the strong error ($k = 2$) between the true solution X_T and the approximation $\hat{X}_T$ as follow

$$\epsilon_k = \begin{cases} \left|\mathbb{E}\Big[\, X_T - \hat{X}_T \Big]\right|_2 \approx \left|\frac{1}{N}\sum_{j=1}^N X_T^j - \hat{X}_T^j|\right|_2, & k = 1, \\ \Big(\mathbb{E}\Big[\, |X_T - \hat{X}_T|^k \Big]\Big)^{1/k} \approx \Big(\frac{1}{N}\sum_{j=1}^N |X_T^j - \hat{X}_T^j|^k\Big)^{1/k}, & k = 2. \end{cases}$$

Our main theorem covers only the strong convergence result, nonetheless, we also present the weak convergence rate estimation.

We study several examples. The stochastic Ginzburg Landau example is a well-studied one and it provides a comparison example to other methods. The second example is a multi-dimensional FitzHugh-Nagumo model of McKean-Vlasov type from neuroscience which shows that the split-step method can properly deal with the superlinear term in a complex system. For the first example, we discuss the parallel implementation, for the second example, we discuss the accuracy w.r.t runtime.

The third example lies outside the scope of our assumptions by featuring a non-Lipschitz measure dependency, but all the methods still work. Proving the convergence of the method under this setting is left for future research.

In the last part, we discuss the stability of the SSM as understood by Theorem 2.2.9. We first look at a linear case to compare the conditions for mean-square contractivity between MV-SDEs and the numerical scheme. Then, the non-linear Ginzburg Landau type equation is used to illustrate the mean square contractivity for the split-step method. At last, the Cucker-Smale flocking model (a degenerate MV-SDE) shows the split-step method has better properties compared to the other methods under larger choices of timestep h.

Remark 2.3.1 (Parallel implementation)**.** *To implement the SSM* (2.5)-(2.6) *in parallel, at each timestep, we first solve step* (2.5) *of the SSM distributed by the cores, then calculate the empirical measure of the particle system in the first core, and finally the second step* (2.6) *of the SSM is executed also in parallel. To implement Taming* (6.1) *and Adaptive* (6.2) *in parallel, at each timestep h one needs to first communicate the empirical measure of the particle system between different cores and then each core can calculate its particles dynamics independently.*

A priori one should expect the taming to be the fastest of the three algorithms. We highlight that due to the nature of the empirical measure, after each timestep the processors need to communicate to update the calculation of the empirical measure, this leads to a well-known loss of parallelization power [21].

2.3.1 Example: the stochastic Ginzburg Landau equation

The first example we consider is the mean-field perturbation of the classic one-dimensional stochastic Ginzburg Landau equation, namely, for all $t \in [0, T]$,

$$dX_t = \Big(\frac{(\sigma')^2}{2}X_t - X_t^3 + c\mathbb{E}[X_t]\Big)dt + \sigma' X_t dW_t, \quad X_0 = x_0 \in \mathbb{R}. \tag{2.11}$$

where in SSM version 1:

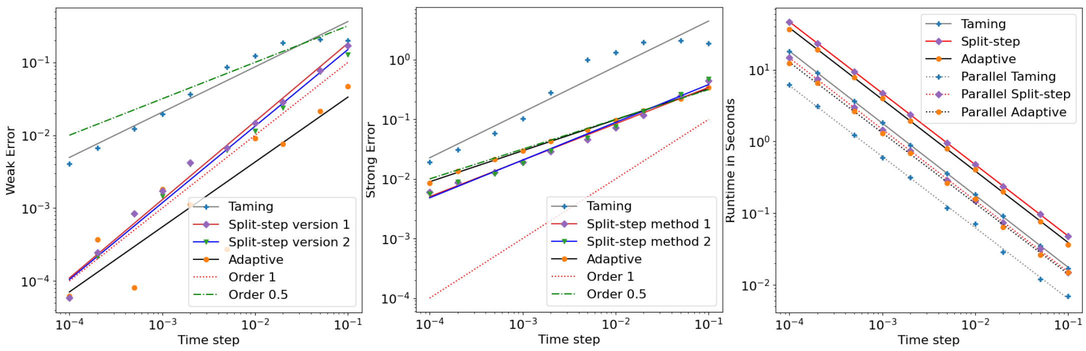
$$v(t,x,\mu) = -x^3, \quad b(t,x,\mu) = \frac{(\sigma')^2}{2}x + c\int_{\mathbb{R}} x\mu(dx), \quad \sigma(t,x,\mu) = \sigma' x, \tag{2.12}$$

in SSM version 2:

$$v(t,x,\mu) = -x^3 + c\int_{\mathbb{R}} x\mu(dx), \quad b(t,x,\mu) = \frac{(\sigma')^2}{2}x, \quad \sigma(t,x,\mu) = \sigma' x. \tag{2.13}$$

where σ', c are constants.

This equation is a toy one-dimensional MV-SDE that we use for a methodological comparison analysis. It features in [60, Section 4.1] and [125, Section 4.2], thus a comparison is insightful under the same choice of coefficients. Namely, we take $\sigma' = 1.5$, $x_0 = 1$, $c = 0.5$, $T = 1$ and $N = 1000$ particles. The results are shown in Figure 2.1. The timestep are $h \in \{10^{-4}, 2\times 10^{-4}, 5\times 10^{-4}, 10^{-3}, \ldots, 10^{-1}\}$. The true solution is calculated with $h = 10^{-5}$ (for each scheme). We have tested both versions of SSM, (2.12) and (2.13), and both show similar results (we present only one).



(a) Weak Error (b) Strong Error (rMSE) (c) Algorithm runtime

Figure 2.1: Simulations of the stochastic Ginzburg Landau equation with $N = 1000$ particles. (a) Weak error of different methods. (b) Strong error of different methods. (c) Runtime of different methods in serial and in parallel.

Taming is implemented with $\alpha = 0.5$ while Adaptive under the choice $\mathbf{h}^\delta(x) = h\min(1, |x|^{-2})$. Fig 2.1(a) shows the Weak error rate of the Taming to roughly be 1/2 where for other methods it is 1.0. Fig 2.1(b) shows the rMSE rate of all the methods to be around 1/2 with Taming's error being about one-order of magnitude higher than the other two errors (also observed in [125, Section 4]). The two versions of the SSM have similar behaviour for this model. Fig 2.1(c) depicts running times of 3 methods (version 1 of the SSM), the top 3 lines are the standard implementations (non-parallel) and the bottom 3 lines are the parallel implementation with 4 cores.

In both the parallel and non-parallel implementation, Taming is the fastest while the SSM takes a slightly longer time than the other methods but with a performance comparable to Adaptive. In a parallel implementation with 4 cores we reach a reduction to nearly 27% in relation to the non-parallel implementation's computational time. In this example, to reach the same strong error level Taming takes nearly 7-times more time than SSM; Adaptive is similar to SSM.

2.3.2 Example: the FitzHugh-Nagumo model

This is a three-dimensional ($d = 3$) MV-SDE (2.1) defined with $v : [0,T] \times \mathbb{R}^3 \to \mathbb{R}^3$, $b : [0,T] \times \mathbb{R}^3 \times \mathcal{P}_2(\mathbb{R}^3) \to \mathbb{R}^3$, $\sigma : [0,T] \times \mathbb{R}^3 \times \mathcal{P}_2(\mathbb{R}^3) \to \mathbb{R}^{3\times 3}$ for $x = (x_1, x_2, x_3) \in \mathbb{R}^3$, $z \in \mathbb{R}$ and μ^{x_3} is the marginal measure in x_3,

$$v(t,x) = \begin{pmatrix} -(x_1)^3/3 \\ 0 \\ 0 \end{pmatrix}, \quad b(t,x,\mu) = \begin{pmatrix} x_1 - x_2 + I - \int_{\mathbb{R}^3} J(x_1 - V_{rev}) z d\mu^{x_3}(z) \\ c(x_1 + a - bx_2) \\ a_r \frac{T_{max}(1-x_3)}{1+\exp(-\lambda(x_1 - V_T))} - a_d x_3 \end{pmatrix},$$

$$\sigma(t,x,\mu) = \begin{pmatrix} \sigma_{ext} & 0 & -\int_{\mathbb{R}^3} \sigma_J (x_1 - V_{rev}) z d\mu^{x_3}(z) \\ 0 & 0 & 0 \\ 0 & \sigma_{32}(x) & 0 \end{pmatrix},$$

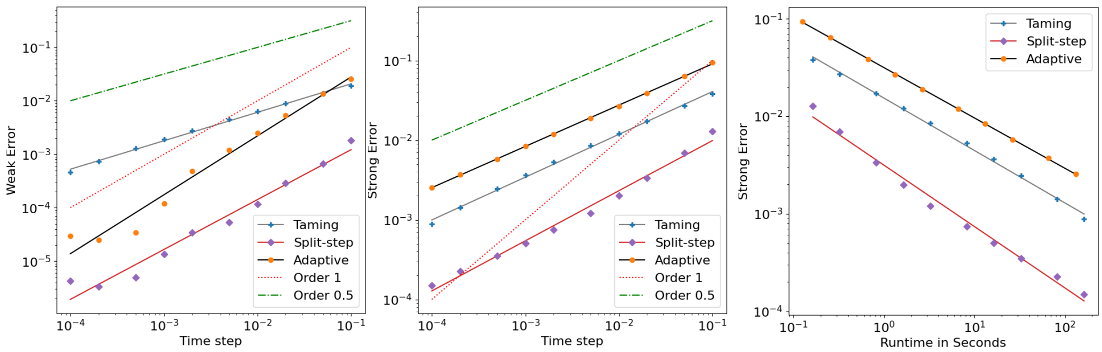
$$x_0 \sim \mathcal{N}\left(\begin{pmatrix} V_0 \\ w_0 \\ y_0 \end{pmatrix}, \begin{pmatrix} \sigma_{V_0} & 0 & 0 \\ 0 & \sigma_{w_0} & 0 \\ 0 & 0 & \sigma_{y_0} \end{pmatrix} \right),$$

where

$$\sigma_{32}(x) := \mathbb{1}_{\{x_3 \in (0,1)\}} \sqrt{a_r \frac{T_{max}(1-x_3)}{1+\exp(-\lambda(x_1 - V_T))} + a_d x_3} \times \Gamma \exp\left[-\frac{\Lambda}{1-(2x_3-1)^2} \right].$$

and $I, J, V_{rev}, V_T, T_{\max}, \Gamma, \Lambda, \sigma_{ext}, \sigma_J, a, b, c, a_r, a_d, \lambda$ are constants.

All the parameters are the same as in [60, Section 4.3] ($V_0 = 0, w_0 = 1/2, \dots$) – see also [125, Section 4.4]. The split-structure of the SSM enable us to flexibly deal with the superlinear term and the Lipschitz terms separately – we make judicious choices regarding v and b and thus we can optimise the solver of the implicit part. Otherwise, we would have been forced to use a general purpose solver which is costlier.



(a) Weak Error (b) Strong Error (rMSE) (c) rMSE w.r.t runtime

Figure 2.2: Simulations of the FitzHugh-Nagumo model with $N = 1000$ particles. (a) Weak error of different methods. (b) Strong error of different methods. (c) Strong error compare to different Algorithm runtime.

For the simulation, we take $N = 1000$, $T = 2$, the time step is taken from $h \in \{10^{-4}, 2 \times 10^{-4}, 5 \times 10^{-4}, 10^{-3}, \dots, 10^{-1}\}$ and the true solution is calculated with $h = 10^{-5}$. Taming is implemented with $\alpha = 1/2$ and Adaptive with $\mathbf{h}^\delta(x) = h \min(1, |x|^2/|b(t,x,\mu)|^2)$. Fig 2.2(a) shows the weak error rate of Taming to roughly be 1/2 with other methods being 1.0 (implementing Taming with $\alpha = 1$ yields a weak convergence rate of order 1.0, we do not present the result). Fig 2.2(b) shows the strong error rate of all the methods to roughly be 1/2, the error of SSM is an order of magnitude lower than the others. Fig 2.2(c) shows that, to reach the same strong error level Taming takes nearly 80-times more time than SSM; Adaptive takes nearly 10-times more time than SSM.

2.3.3 Example: Polynomial drift (non-Lipschitz measure dependency but still of one-sided Lipschitz type)

We present an example from [20, Section 3.3] that falls outside the theoretic framework of this chapter. Take the one-dimensional MV-SDE for $t \in [0, T]$ and $\gamma \in \mathbb{R}$

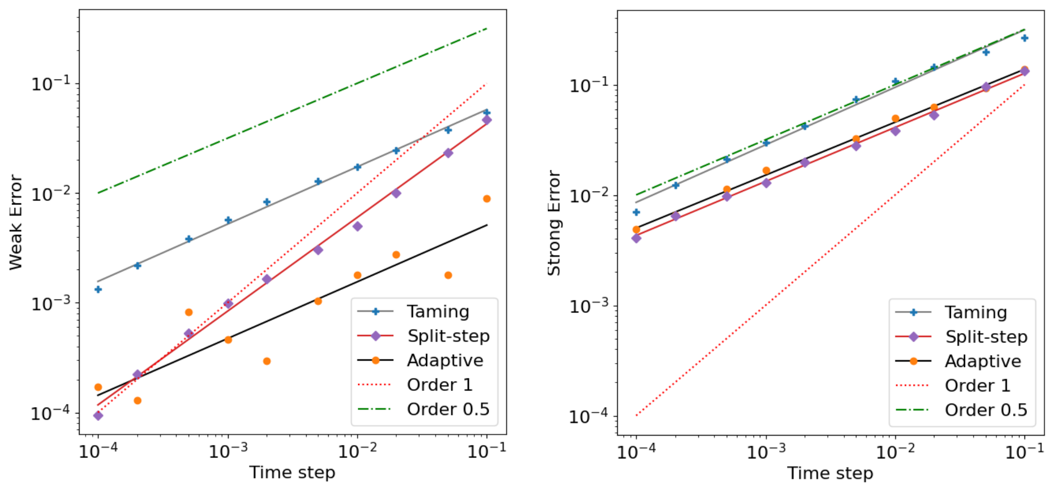
$$dX_t = \Big(\gamma X_t + \mathbb{E}[X_t] - X_t \mathbb{E}\big[|X_t|^2\big]\Big)dt + X_t dW_t, \quad \text{with } X(0) = x_0 \in \mathbb{R}, \quad (2.14)$$

$$\text{and set:} \quad v(t,x) = \gamma x - x \int_{\mathbb{R}} |x|^2 \mu(dx), \quad b(t,x,\mu) = \int_{\mathbb{R}} x\mu(dx), \quad \sigma(t,x,\mu) = x.$$

The dynamics of the interacting particle system (2.2), for $i \in [\![1, N]\!]$, $X^{i,N} \in \mathbb{R}$, is

$$dX_t^{i,N} = \Big[\gamma X_t^{i,N} + \frac{1}{N}\sum_{j=1}^{N} X_t^{j,N} - X_t^{i,N}\frac{1}{N}\sum_{j=1}^{N}|X_t^{j,N}|^2\Big]dt + X_t^{i,N} dW_t^i.$$

where $(W^i)_i$ are independent Brownian motions. Take $\gamma = -1$, $T = 1$. We have a superlinear measure component in (2.14) and none of the schemes is applicable (insofar as existing theoretical results allow). If the measure component was fixed (with finite 2nd moment), then, the drift would satisfy a one-sided Lipschitz condition – the convergence of this type of schemes is left for future work.



(a) Weak Error w.r.t Time step (b) Strong Error (rMSE) w.r.t Time step

Figure 2.3: Simulations of the polynomial drift model with $N = 1000$ particles. (a) Weak error of different methods. (b) Strong error of different methods.

The results are shown in Figure 2.3. We take $N = 1000$, $T = 2$ and the timestep is taken from $h \in \{10^{-4}, 2 \times 10^{-4}, 5 \times 10^{-4}, 10^{-3}, \dots, 10^{-1}\}$. The true solution is calculated under $h = 10^{-6}$. Taming is applied with $\alpha = 0.5$ while Adaptive under the choice $\mathbf{h}^\delta(x) = h \min(1, |x|^{-2})$. Figure 2.3(a) shows the weak error rate of Taming with $\alpha = 0.5$ to roughly be 1/2 with other methods being 1.0. Figure 2.3(b) shows strong error rate of all the methods to roughly be 1/2 (as expected).

2.3.4 Stability of the SSM: linear, non-linear and the Cucker-Smale model case

Recall from Theorem 2.2.9 the expression (2.9) for β. For the remainder of this section, let $t \in [0, T]$, we define $X_0, Z_0 \in L_0^m(\mathbb{R}^d)$ and $\hat{X}_0^i, \hat{Z}_0^i$, $i \in [\![1, N]\!]$ as i.i.d. samples from X_0 and Z_0 respectively, $X_n^{i,N}$ and $Z_n^{i,N}$ are defined as in Theorem 2.2.9 as outputs of our SSM scheme (2.5) and (2.6) (with the corresponding initial conditions). If $\beta < 0$ we then have $\mathbb{E}[|X_n^{i,N}|^2] = 0$ as $n \to \infty$.

Linear case: an Ornstein-Uhlenbeck McKean-Vlasov SDE

For the MV-SDE (see e.g. [20, Section 2.1]), for all $t \in [0,T]$ and $x_0 \in \mathbb{R}$

$$dX_t = \Big(\rho X_t + \lambda \mathbb{E}[X_t]\Big)dt + \eta\, dW_t, \quad X_0 = x_0, \tag{2.15}$$

$$\text{set:} \quad v(t,x,\mu) = \rho x, \quad b(t,x,\mu) = \lambda \int_{\mathbb{R}} x\mu(dx), \quad \sigma(t,x,\mu) = \eta. \tag{2.16}$$

where ρ, λ, η are constants. The first and second moments of X are respectively given by $\mathbb{E}\left[X_t\right] = x_0 \exp((\rho+\lambda)t)$ and $\mathbb{E}\left[X_t^2\right] = x_0^2 \exp(2(\rho+\lambda)t) + \frac{\eta^2}{2\rho}(\exp(2\rho t) - 1)$.

Let X, Z be two solution of (2.15) with X_0 and Z_0 as initial condition respectively, then by direct calculation

$$\mathbb{E}\Big[|X_t - Z_t|^2\Big] = \frac{1}{2\lambda} e^{2(\rho+\lambda)t} + \mathbb{E}\Big[|X_0 - Z_0|^2\Big] e^{2\rho t}.$$

Let $\rho \le 0$ and $\rho + \lambda < 0$ then from Definition 2.2.10, (2.15) generates exponential mean-square contractive solutions. The parameters of this example are $L_v = \rho$, $L_{\tilde{b}} = \lambda^2$, $L_{\tilde{v}} = L_b = L_\sigma = L_{\tilde{\sigma}} = 0$. Plugging these into (2.9) and in order to make $\beta < 0$, we need to choose h satisfying

$$h^{ssm} < -\frac{2\rho + 2\lambda + 1}{\lambda^2},$$

while in the standard Euler method, we need the following condition on the choice of h^{euler} to reach a contraction

$$h^{euler} < -\frac{\rho + |\lambda|}{\rho^2 + \lambda^2} \tag{2.17}$$

From Definition 2.2.8, to let the split-step method (2.5)-(2.6) admits mean-square contractive, h^{ssm} requires $\rho + \lambda < -1/2$, while the Euler method h^{euler} requires $\rho + |\lambda| < 0$. Recall the SDE to generate exponential mean-square contractive solutions need the constraint $\rho + \lambda < 0$. Thus, the condition for a mean-square stable for both numerical solution is slightly stronger than the condition for the SDE to generate exponential mean-square contractive solutions. Moreover, for $\lambda > -1/4$, the Euler method has less restriction on ρ than the SSM, and for negative $\lambda < -1/4$, the SSM has less restriction on ρ than the Euler method, and both method has same restriction on ρ when $\lambda = -1/4$.

Nonlinear case I: a stochastic Ginzburg Landau type equation

We illustrate the stability of the SSM scheme via the stochastic Ginzburg Landau type equation (in the style of that in Section 2.3.1), we consider the following one-dimensional MV-SDE for all $t \in [0,T]$

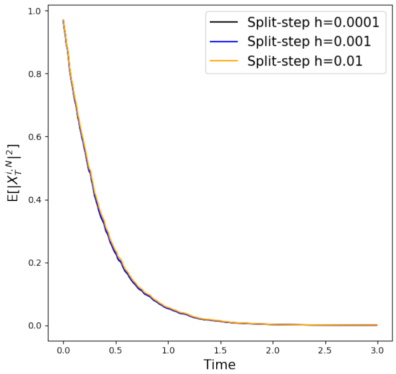
$$dX_t = \Big(-\frac{5}{2}X_t - \frac{1}{4}X_t^3 + \mathbb{E}[X_t]\Big)dt + X_t dW_t, \qquad X_0 = 1, \tag{2.18}$$

$$\text{set:} \quad v(t,x,\mu) = -\frac{5}{2}x - \frac{1}{4}x^3 - \gamma x, \quad b(t,x,\mu) = \int_{\mathbb{R}} x\mu(dx) + \gamma x \;\text{ for }\; \gamma \in \mathbb{R}, \quad \sigma(t,x,\mu) = x.$$

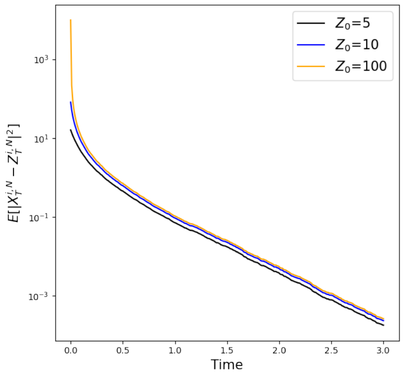
The parameters of this example are $L_v = -5/2 - \gamma$, $L_{\tilde{b}} = L_\sigma = 1$, and $L_b = \gamma^2$, $L_{\tilde{v}} = L_{\tilde{\sigma}} = 0$ – with these parameters it is known [143] that the system is conservative and the solution satisfies $X_t \to 0$ a.s. as $t \to \infty$. Plugging these into the mean-square stability β constant (2.9) and, when $\gamma = 0$, one must have small h in order to make $\beta < 0$. We now employ the split-step method, under different choices of h and initial values. Set the number of particles to be $N = 1000$.

Set $\gamma = 0$, $T = 3$ and $X_0 = 1$. Figure 2.4(a) shows that $\mathbb{E}[|X_T^{i,N}|^2]$ decreases to zero under different values of h (but small). Figure 2.4(b) shows mean square contraction property between X and Z highlights an exponential decay (where Z solves (2.18) for $Z_0 = 5, 10, 100$.)

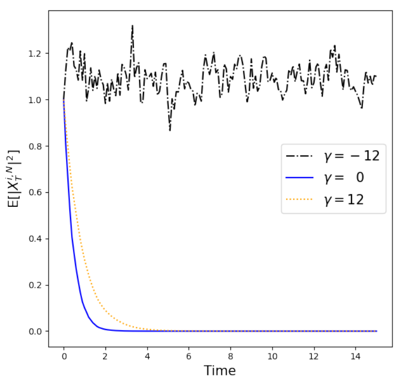
Fix $h = 0.1$. Figure 2.4(c) shows that when $\gamma = -12$ the scheme performs poorly, and this follows from the conditions of Theorem 2.2.9 not being satisfied. For $\gamma \in \{0, 12\}$, the scheme



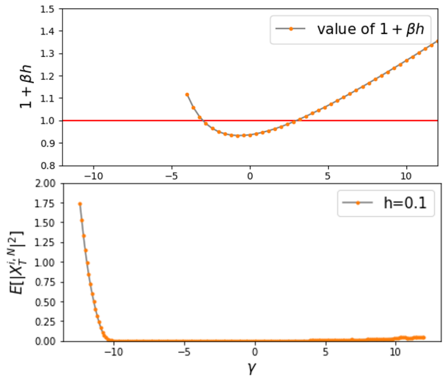
(a) $(h, t) \mapsto \mathbb{E}[|X_t^{i,N}|^2]$



(b) $(Z_0, t) \mapsto \mathbb{E}[|X_t^{i,N} - Z_t^{i,N}|^2]$



(c) $(\gamma, t) \mapsto \mathbb{E}[|X_t^{i,N}|^2]$



(d) top: $\gamma \mapsto 1 + \beta(\gamma)h$, bottom: $\gamma \mapsto \mathbb{E}[|X_T^{i,N}|^2]$

Figure 2.4: Simulations of the stochastic Ginzburg Landau type equation (2.18) with $N = 1000$ particles and $T = 3.0$. (a) shows $[0, 3] \ni t \mapsto \mathbb{E}[|X_t^{i,N}|^2]$ under three different timesteps $h \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ highlighting mean-square stability. (b) shows the mean square differences between X, Z when $h = 0.01$ for fixed $X_0 = 1$, $Z_0 \in \{5, 10, 100\}$ and $t \in [0, T]$ highlighting mean-square contractivity.
Highlighting mean-square stability/instability of approximation as map of γ under fixed $h = 0.1$. (c) this shows $[0, T] \ni t \mapsto \mathbb{E}[|X_t^{i,N}|^2]$ (for $T = 15$) under three different $\gamma \in \{-12, 0, 12\}$. (d) (top) shows $\gamma \mapsto 1 + \beta(\gamma)h$ where β is given in Theorem 2.2.9 and (bottom) $[-12, 12] \ni \gamma \mapsto \mathbb{E}[|X_T^{i,N}|^2]$, $T = 3$. As $\gamma \geq 5$ the method starts showing an error increase (bottom) which can be matched to $\beta(\gamma) > 0$ (top) and hence loss of stability.

shows contraction as $t \to \infty$, but it is much slower for $\gamma = 12$. Figure 2.4(d, lower graph) shows what happens when one shifts "slope from v to b" via the linear term γx (see Remark 2.2.7). We have now $L_v = -\gamma - 5/2$, thus, when $\gamma \in (-5, 5)$ the figure shows *contraction*, with $X_T^{i,N} \approx 0$ as expected. There is a significant change for $\gamma < -10$ where the approximation is not converging to the correct value. For $\gamma \geq 5$ and higher (recall that $h = 0.1$ is fixed) it seems the contraction is happening (although at a slower pace) but in Figure 2.4(d, upper graph) one sees that $\gamma \to 1 + \beta(\gamma)h$ is now above 1.0 which does not guarantee contraction (in the sense of Theorem 2.2.9).

Figure 2.4(c) and (d) highlight the trade off and care needed between: *(i)* making L_v negative via γ and thus removing the constraint on h imposed in Theorem 2.2.6, and, *(ii)* ensuring the stability of the scheme as imposed by Theorem 2.2.9.

Nonlinear case II: the two-dimensional Cucker-Smale flocking model

This example (see [67, Section 2]) highlights the stability of the split-step method. It is stable under larger timestep h by using the implicit step for the superlinear part. The explicit methods (Taming and Adaptive) fail to have acceptable results at this level of h.

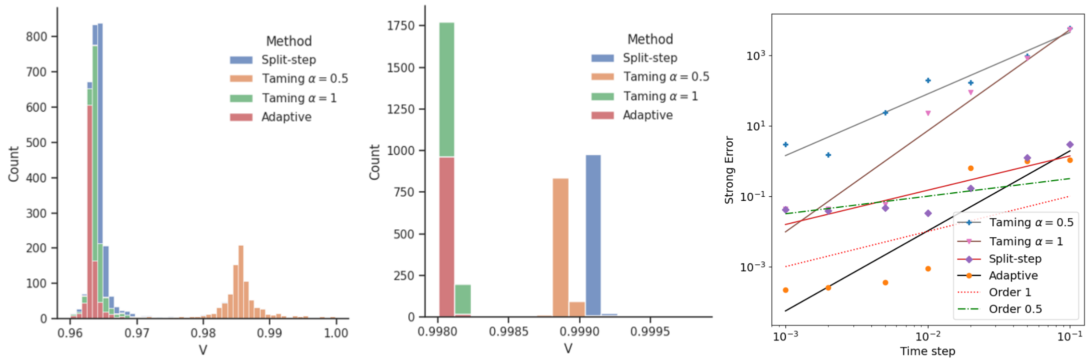
Applied our settings, this is a two-dimensional MV-SDE define under $v, b, \sigma : [0,T] \times \mathbb{R}^2 \times \mathcal{P}_2(\mathbb{R}^2) \to \mathbb{R}^2$ for $x = (V, X) \in \mathbb{R}^2$, $z \in \mathbb{R}$, μ^V is the measure of V as:

$$v(t,x,\mu) = \begin{pmatrix} -V^3 \\ 0 \end{pmatrix}, b(t,x,\mu) = \begin{pmatrix} 1 + \lambda \int_{\mathbb{R}} (V-z) d\mu^V(z) \\ V \end{pmatrix}, \sigma(t,x,\mu) = \begin{pmatrix} \sigma' \int_{\mathbb{R}} (V-z) d\mu^V(z) \\ 0 \end{pmatrix}.$$

where λ, σ' are constants. The dynamics of the particle system follows easily

$$dV_t^{i,N} = \Big(1 - (V_t^{i,N})^3 + \frac{\lambda}{N} \sum_{j=1}^{N} (V_t^{j,N} - V_t^{i,N})\Big) dt + \frac{\sigma'}{N} \sum_{j=1}^{N} (V_t^{j,N} - V_t^{i,N}) dW_t^i, \; dX_t^{i,N} = V_t^{i,N} dt.$$

where $i \in [\![1, N]\!]$, $V^{i,N}, X^{i,N} \in \mathbb{R}$, $(W^i)_i$ are independent Brownian motions.



(a) Histogram of V at $T = 1$ (b) Histogram of V at $T = 2$ (c) Strong Error of V

Figure 2.5: Simulations of the Cucker-Smale type flocking model. (a,b,c) Histogram of V at different time $T = 1, 2$ of different methods with $h = 10^{-3}$. (d) Strong error of different methods at $T = 2$.

Take $\lambda = 2$, $\sigma' = 4$, $T = 2$. With this choice of parameters the solution process V_t converges to 1 as $t \to \infty$ [67]. $V_0^{i,N} \sim \mathcal{N}(1, 2)$ are i.i.d. samples from standard normal distribution, the timesteps h are $h \in \{10^{-3}, 2 \times 10^{-3}, \dots, 10^{-1}\}$, particles $N = 1000$. The true solution is calculated under $h = 10^{-4}$. Taming is run with $\alpha = 0.5$ and 1 and Adaptive with $\mathbf{h}^\delta(x) = h \min(1, |x|^{-2})$. Figure 2.5 (a,b) show the distribution of V at time $T = 1, 2$. All four methods have same initial distribution (and same filtration) nonetheless there is a slight skew between the final results. Taming with $\alpha = 0.5$ has a different distribution than hte other three methods at $T = 1$, later, the SSM clusters at a different point than Adaptive and Taming method with $\alpha = 1$, but the deviation is very small ($< 10^{-2}$). Consider the strong error graph (c), the two Taming methods fail to have acceptable result with larger timestep; while SSM and Adaptive are at a similar position. For Adaptive, the error rate is nicely behaved but there is a jump at $h = 0.02$. The split-step method error rate decrease is stable as h decreases.

2.3.5 Discussion

We discuss some comparative advantages between the methods starting with generalist comments. All schemes have the same convergence error rate rMSE $\approx Ch^{\frac{1}{2}}$. Taming is by far the easiest to implement, with Adaptive the most complex requiring tuning the $\mathbf{h}^\delta$ map for each case (see 6.1). The SSM requires an implicit solver and ad-hoc choices of v and b for efficiency. Taming is the fastest algorithm with SSM and Adaptive running times being comparable with each other. In the way we presented the SSM: all methods are amenable to an efficient parallel implementation (under the caveat of processor communication [21]); moreover, in view of Remark 2.2.7, the SSM does not have any (real) restriction on the time-stepping although one solves an implicit method.

From the numerical examples, we see that

1. the strong error of the SSM is consistently one order of magnitude smaller than that of Taming. Under same choice of timestep, Taming is the fastest with SSM comparable to Adaptive. However, to reach the same strong error level, SSM takes less computational time than Adaptive and significantly less than Taming.

2. Compared to Adaptive, the SSM has in general no worse convergence than it and no clear domination of one over the other emerged. Implementation wise (at the level of computing the rMSE), to keep the same filtration for different timestep choices, the Brownian motion paths for Adaptive with function $\mathbf{h}^{\delta}(x)$ is much harder to generate (requiring sub-simulation from Brownian bridges) than the SSM with a fixed timestep. As a rule of thumb, Adaptive does on average a double amount of timesteps than SSM or Taming [125].

3. From the numerical examples and at the level of the strong error, the SSM performs better than Taming and Adaptive at larger time steps h (via comparative lower errors).

We have not investigated the effect of dimensionality, and we suspect that the running time gap of Taming between SSM or Adaptive will widen. The SSM we present has the extra advantage of flexibility in the way of how v and b are chosen. This means that a layer of optimisation can be added to the implicit solver. Lastly, and partially addressed here with the stability analysis, do the schemes preserve the finer properties of the underlying dynamics? Are they geometrically ergodic? Do they preserve oscillatory dynamics, such as amplitudes, frequencies and phases of oscillations? Even for large time steps? It is known that explicit Euler type schemes face difficulties in regards to this, with implicit or splitting methods being more stable [32].

2.4 Proof of the convergence result for the split-step method (SSM)

Throughout this section Assumption 2.2.1 is assumed to hold for all results.

2.4.1 Proof of the main convergence result, Theorem 2.2.6

For all auxiliary results next, we assume the conditions of Theorem 2.2.6 are in force and we thus do not state them.

Preliminary results

As a first step, we state a result that allows us to re-write (2.5) and (2.6) as a map of $\hat{X}^{i,N}$ without the presence of the $Y^{i,\star,N}$. We present first a new general version of [85, Lemma 3.4] where the differentiability Assumption is lifted and the maps are allowed to depend on time (and the measure component).

Lemma 2.4.1. *Let v be as in Assumption 2.2.1. Choose $h > 0$ satisfying $1 - h(2L_v + 1) > 0$. Then for $t \in [0,T]$, $c, d \in \mathbb{R}^d$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ the implicit equation,*
with d, μ, t fixed and c unknown,

$$c = d + hv(t, c, \mu). \tag{2.19}$$

has a unique solution in c. Define the functions v_h and F_h as

$$v_h(t,d,\mu) = v\Big(t, F_h(t,d,\mu), \mu\Big) \quad \textit{with} \quad [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \ni (t,d,\mu) \mapsto F_h(t,d,\mu) = c \in \mathbb{R}^d. \tag{2.20}$$

We then have for all $t \in [0,T]$, $x, y \in \mathbb{R}^d$, $\mu, \mu^x, \mu^y \in \mathcal{P}_2(\mathbb{R}^d)$ the following four inequalities,

$$
\begin{aligned}
|v_h(t,x,\mu)| &\leq \frac{|v(t,x,\mu)|}{1-hL_v}, & (2.21)\\
|F_h(t,x,\mu) - F_h(t,y,\mu)|^2 &\leq \frac{|x-y|^2}{1-2hL_v}, & (2.22)\\
|F_h(t,x,\mu^x) - F_h(t,y,\mu^y)|^2 &\leq \frac{1}{1-h(2L_v+1)}\Big(|x-y|^2 + L_{\tilde{v}} h \big(W^{(2)}(\mu^x,\mu^y)\big)^2\Big), & (2.23)\\
\langle x-y, v_h(t,x,\mu) - v_h(t,y,\mu)\rangle &\leq \frac{L_v}{1-2hL_v}|x-y|^2. & (2.24)
\end{aligned}
$$

For $x_i, y_i \in \mathbb{R}^d$, $i \in [\![1,N]\!]$ and $\mu^x, \mu^y \in \mathcal{P}_2(\mathbb{R}^d)$ being the empirical measures associated with the collections $\{x_i\}_i, \{y_i\}_i$, define the maps

$$
b_h(t,x_i,\mu^{F_{h,x,\mu^x}}) = b(t, F_h(t,x_i,\mu^x), \mu^{F_{h,x,\mu^x}}),\ \ \sigma_h(t,x_i,\mu^{F_{h,x,\mu^x}}) = \sigma(t, F_h(t,x_i,\mu^x), \mu^{F_{h,x,\mu^x}}),
$$

$$
\text{where}\ \ \mu^{F_{h,x,\mu^x}}(dx) = \frac{1}{N}\sum_{j=1}^N \delta_{F_h(t,x_j,\mu^y)}(dx),\ \ \mu^{F_{h,y,\mu^y}}(dx) = \frac{1}{N}\sum_{j=1}^N \delta_{F_h(t,y_j,\mu^y)}(dx).
$$

then, b_h and σ_h are satisfy

$$
\begin{aligned}
&\Big|b_h(t,x_i,\mu^{F_{h,x,\mu^x}}) - b_h(t,y_i,\mu^{F_{h,y,\mu^y}})\Big|^2\\
&\qquad \leq \frac{L_b}{1-h(2L_v+1)}\bigg(|x_i-y_i|^2 + \frac{2L_{\tilde{v}}h+1}{N}\sum_{j=1}^N |x_j-y_j|^2\bigg),\\
&\Big|\sigma_h(t,x_i,\mu^{F_{h,x,\mu^x}}) - \sigma_h(t,y_i,\mu^{F_{h,y,\mu^y}})\Big|^2\\
&\qquad \leq \frac{L_\sigma}{1-h(2L_v+1)}\bigg(|x_i-y_i|^2 + \frac{2L_{\tilde{v}}h+1}{N}\sum_{j=1}^N |x_j-y_j|^2\bigg).
\end{aligned}
$$

Lastly, $v_h \to v$, $b_h \to b$, and $\sigma_h \to \sigma$ uniformly over the compacts of $[0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ as $h \to 0^+$.

We observe that to establish (2.21) one only needs $1-hL_v > 0$, in other words: the condition $1-2hL_v > 0$ is not sharp for that result. Nonetheless, inequalities (2.22) and (2.24) are critical for our work and hence we write one single constraint.

Proof. Existence and uniqueness for (2.19) can be proved via a strict monotonicity contraction argument. Namely, fix some $t \in [0,T]$, from (2.19) one defines the operator $A : \mathbb{R}^d \to \mathbb{R}^d$ as $A(u) = u - hv(t,u,\mu)$ for $u \in \mathbb{R}^d$. Following [142, Definition 25.2 (p.500)], the operator A is continuous and strongly monotone (uniformly in t) under the restrictions $h > 0$ and $1-hL_v > 0$. This follows by directly injecting the one-sided condition of v (from Assumption 2.2.1) in the definition of *strongly monotone operator*. Finally, from [142, Theorem 26.A (p.557)] we conclude that the operator A is invertible and the inverse map is Lipschitz continuous. Thus, (2.19) has a unique measurable inverse given by F_h from (2.20). See also [107, (p.2596)].

We now determine the Lipschitz constant of v_h and F_h of (2.20). For (2.21), suppose $c, d \in \mathbb{R}^d$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ satisfy $c = d + hv(t,c,\mu)$ then, from Assumption 2.2.1:

$$
\begin{aligned}
c &= d + h\Big(v(t,c,\mu) - v(t,d,\mu)\Big) + hv(t,d,\mu),\\
\Rightarrow\ |c-d|^2 &= \Big\langle c-d, v(t,c,\mu) - v(t,d,\mu)\Big\rangle h + \langle c-d, v(t,d,\mu)\rangle h\\
&\leq hL_v|c-d|^2 + \langle c-d, v(t,d,\mu)\rangle h\\
&\Leftrightarrow (1-hL_v)\ |c-d|^2 \leq |c-d|\ |v(t,d,\mu)|\ h.
\end{aligned}
$$

Since $c = d + hv_h(t,d,\mu)$, we have by re-arranging the terms and plugging the inequality above

$$|v_h(t,d,\mu)| = \frac{1}{h}|c-d| \le \frac{|v(t,d,\mu)|}{1-hL_v}.$$

For (2.22), suppose $c_1, c_2, d_1, d_2 \in \mathbb{R}^d,\ \mu \in \mathcal{P}_2(\mathbb{R}^d)$ satisfy $c_1 = d_1 + hv(t,c_1,\mu)$ and $c_2 = d_2 + hv(t,c_2,\mu)$, we have

$$\begin{aligned}\Big|F_h(t,d_1,\mu) - F_h(t,d_2,\mu)\Big|^2 &= |c_1-c_2|^2 \\ &= \langle c_1-c_2, d_1-d_2\rangle + \Big\langle c_1-c_2, v(t,c_1,\mu)-v(t,c_2,\mu)\Big\rangle h \\ &\le \frac{1}{2}|c_1-c_2|^2 + \frac{1}{2}|d_1-d_2|^2 + hL_v|c_1-c_2|^2.\end{aligned}$$

For (2.23), suppose $x, y, \hat{x}, \hat{y} \in \mathbb{R}^d,\ \mu^x, \mu^y \in \mathcal{P}(\mathbb{R}^d)$ satisfy $\hat{x} = x + hv(t,\hat{x},\mu^x),\ \hat{y} = y + hv(t,\hat{y},\mu^y)$. We then have

$$\begin{aligned}&\Big|F_h(t,x,\mu^x) - F_h(t,y,\mu^y)\Big|^2 = |\hat{x}-\hat{y}|^2 = \langle \hat{x}-\hat{y}, x-y\rangle + \Big\langle \hat{x}-\hat{y}, v(t,\hat{x},\mu^x) - v(t,\hat{y},\mu^y)\Big\rangle h \\ &\le \frac{1}{2}|\hat{x}-\hat{y}|^2 + \frac{1}{2}|x-y|^2 + \Big\langle \hat{x}-\hat{y}, v(t,\hat{x},\mu^x) - v(t,\hat{y},\mu^x)\Big\rangle h \\ &\quad + \Big\langle \hat{x}-\hat{y}, v(t,\hat{y},\mu^x) - v(t,\hat{y},\mu^y)\Big\rangle h \\ &\le \frac{1}{2}|\hat{x}-\hat{y}|^2 + \frac{1}{2}|x-y|^2 + (L_v + \frac{1}{2})|\hat{x}-\hat{y}|^2 h + \frac{1}{2}L_{\tilde{v}}\big(W^{(2)}(\mu^x,\mu^y)\big)^2.\end{aligned}$$

To prove (2.24) we use the same notation/identities used to prove Inequality (2.22) above. We have that

$$\begin{aligned}&\Big\langle d_1-d_2, (d_1-d_2) + h\Big(v_h(t,d_1,\mu) - v_h(t,d_2,\mu)\Big)\Big\rangle \\ &\quad = \langle d_1-d_2, c_1-c_1\rangle \le \frac{1}{2}|d_1-d_2|^2 + \frac{1}{2}\frac{|d_1-d_2|^2}{1-2hL_v},\end{aligned}$$

and thus

$$\Big\langle d_1-d_2, v_h(t,d_1,\mu) - v_h(t,d_2,\mu)\Big\rangle \le \frac{L_v}{1-2hL_v}|d_1-d_2|^2.$$

We now address the Lipschitz property of b_h, σ_h. Since they are of the same nature, we provide only the proof for b_h as that for σ_h is identical. Let $i \in [\![1,N]\!]$, using the definition of b_h, then Assumption 2.2.1 followed by (2.23) we have

$$\begin{aligned}&\Big|b_h(t,x_i,\mu^{F_{h,x}}) - b_h(t,y_i,\mu^{F_{h,y}})\Big|^2 \le L_b\Big(|F_h(t,x_i,\mu^x) - F_h(t,y_i,\mu^y)|^2 + \big(W^{(2)}(\mu^{F_{h,x}},\mu^{F_{h,y}})\big)^2\Big) \\ &\le L_b\Big(|F_h(t,x_i,\mu^x) - F_h(t,y_i,\mu^y)|^2 + \frac{1}{N}\sum_{j=1}^N |F_h(t,x_j,\mu^x) - F_h(t,y_j,\mu^y)|^2\Big) \\ &\le \frac{L_b}{1-h(2L_v+1)}\Big(|x_i-y_i|^2 + \frac{2L_{\tilde{v}}h+1}{N}\sum_{j=1}^N |x_j-y_j|^2\Big).\end{aligned}$$

The convergence result in the final statement follows straightforwardly from [85, Lemma 3.4]. This convergence result is applied with fixed N and the parameter of the convergence is h (not N). One only needs to apply their arguments over $[0,T] \times \mathbb{R}^{dN}$ where the measures $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ are taken to be compactly supported on the compact where the family of points $\{x_i\}_i$ is contained.

□

After having introduced Lemma 2.4.1, we can finally address the continuous-time extension of the SSM (2.5)-(2.6) as referenced in Theorem 2.2.6. The SSM can be written as a continuous time SDE via linear interpolation of the iterates, namely, for $t \in [t_n, t_{n+1}]$, $i \in [\![1, N]\!]$, $\hat{X}_0^i \in L_0^m(\mathbb{R}^d)$:

$$d\hat{X}_t^{i,N} = \Big(v_h\big(\kappa(t), \hat{X}_{\kappa(t)}^{i,N}, \hat{\mu}_{\kappa(t)}^N\big) + b_h\big(\kappa(t), \hat{X}_{\kappa(t)}^{i,N}, \widetilde{\mu}_{\kappa(t)}^N\big)\Big)dt + \sigma_h\big(\kappa(t), \hat{X}_{\kappa(t)}^{i,N}, \widetilde{\mu}_{\kappa(t)}^N\big)dW_t^i, \quad (2.25)$$

$$\text{where} \quad \hat{\mu}_{\kappa(t)}^N(dx) := \frac{1}{N}\sum_{j=1}^N \delta_{\hat{X}_{\kappa(t)}^{j,N}}(dx) \quad \text{and} \quad \widetilde{\mu}_{\kappa(t)}^N(dx) := \frac{1}{N}\sum_{j=1}^N \delta_{F_h(\kappa(t), \hat{X}_{\kappa(t)}^{j,N}, \hat{\mu}_{\kappa(t)}^N)}(dx).$$

where $\kappa(t) = \sup\Big\{t_n : t_n \le t,\ n \in [\![0, M]\!]\Big\}$ and $\widetilde{\mu}_{t_n}^N = \widetilde{\mu}_n^N$.

Moment bounds

We now employ the results of Lemma 2.4.1 to establishing a domination of $|Y_n^{i,\star,N}|$ by $|\hat{X}_n^{i,N}|$.

Lemma 2.4.2. *Choose h as in Theorem 2.2.6 and recall $C_T, \widehat{L}_v$ as defined in Remark 2.2.2. Then, $|Y_n^{i,\star,N}|$ of (2.5) satisfies for any $i \in [\![1, N]\!]$, $n \in [\![0, M-1]\!]$,*

$$|Y_n^{i,\star,N}|^2 \le |\hat{X}_n^{i,N}|^2\Big(1 + \frac{2\widehat{L}_v}{1 - 2\widehat{L}_v h}h\Big) + \frac{L_{\tilde{v}}h}{1 - 2\widehat{L}_v h}\bigg(\frac{1}{N}\sum_{j=1}^N |\hat{X}_n^{j,N}|^2\bigg) + \frac{2C_T}{1 - 2\widehat{L}_v h}h. \quad (2.26)$$

Remark 2.4.3 (More on the stepsize h)**.** *The bound for h is necessary, for example, if $v(t, x, \mu) = 10x$, then the implicit solution gives $y = x/(1 - 10h)$ and one sees that the $h < 0.1$ condition is critical.*

Recall the stepsize constraint on h in Theorem 2.2.6. By inspection of the proof of this Lemma and the definition of $\widehat{L}_v$ in Remark 2.2.2, the reader will find that the constraint $L_v < -1/2$ needed to ensure $1 - h\widehat{L}_v > 0$ is not sharp and can be replaced by some number $\varepsilon \in (0, 1)$, i.e., $L_v < -\varepsilon$. The lack of sharpness arises from the choice of $\widehat{L}_v$ in Remark 2.2.2. There we used the Cauchy-Schwarz inequality where we could have used a Young type inequality from which the parameter ε would have arisen. We choose $1/2$ for ease of presentation.

However, through the split-step structure, one can use the "add and subtract a linear component" in the drift before the split-step is executed to make $L_v < 0$ and thus remove the constraint on h – see Section 2.3.4.

Proof. From (2.5) and Remark 2.2.2, for any i, n and any $t_n \in \pi$ we have using Cauchy-Schwarz and the properties of v that

$$\begin{aligned}|Y_n^{i,\star,N}|^2 &= \langle Y_n^{i,\star,N}, \hat{X}_n^{i,N}\rangle + \langle Y_n^{i,\star,N}, v(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^N)\rangle h \\ &\le \frac{1}{2}|Y_n^{i,\star,N}|^2 + \frac{1}{2}|\hat{X}_n^{i,N}|^2 + h\bigg(C_T + \widehat{L}_v|Y_n^{i,\star,N}|^2 + \frac{L_{\tilde{v}}}{2}\Big(\frac{1}{N}\sum_{j=1}^N |\hat{X}_n^{j,N}|^2\Big)\bigg).\end{aligned}$$

□

Remark 2.4.4. *Under Assumption 2.2.1 and under the choice of h in Theorem 2.2.6, $1/(1 - 2\widehat{L}_v h)$ is bounded above by some constant independent of h. We can thus claim that there exist constants $C > 0$, $\widetilde{C} \in \mathbb{R}$ depending on $L_v, C_T, \widehat{L}_v$ but independent of h such that from Lemma 2.4.1 and Lemma 2.4.2 we have (for any t, x, y, μ)*

$$|v_h(t, x, \mu)| \le C|v(t, x, \mu)|, \quad |Y_k^{i,\star,N}|^2 \le |\hat{X}_k^{i,N}|^2(1 + Ch) + h\frac{C}{N}\sum_{j=1}^N |\hat{X}_k^{j,N}|^2 + Ch,$$

$$\big\langle x - y, v_h(t, x, \mu) - v_h(t, y, \mu)\big\rangle \le \widetilde{C}|x - y|^2, \quad |F_h(t, x, \mu) - F_h(t, y, \mu)|^2 \le C|x - y|^2.$$

After this remark emphasising the independence of the constants in h, we are in a position to prove the moment bounds for the output of the SSM.

Proposition 2.4.5 (Moment bounds of SSM)**.** *Choose h as in Theorem 2.2.6. Then, there exists a constant $C \in \mathbb{R}^+$ such that for any $i \in [\![1, N]\!]$, $n \in [\![0, M]\!]$, for $\frac{m}{2} \geq p \geq 1$, the output $\hat{X}_n^{i,N}$ of the scheme* (2.5)-(2.6) *satisfies,*

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

Proof. Let $i \in [\![1, N]\!]$, $n \in [\![0, M-1]\!]$ and recall (2.6). Using Assumption 2.2.1, Remark 2.4.4, we have by taking squares and applying Young's inequality:

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

where we used $W^{(2)}(\hat{\mu}_n^{Y,N}, \delta_0)^2 \leq \frac{1}{N}\sum_{j=1}^N |Y_n^{j,\star,N}|^2$. By backward induction from $n+1$ to zero, we have (after some simplification)

$$\begin{aligned}
|\hat{X}_{n+1}^{i,N}|^2 &\leq |\hat{X}_0^{i,N}|^2 + 2\sum_{k=0}^n \Big(\Big\langle Y_k^{i,\star,N}, \sigma(t_k, Y_k^{i,\star,N}, \hat{\mu}_k^{Y,N})\Delta W_k^i\Big\rangle\Big) + C\sum_{k=0}^n \Big(|\Delta W_k^i|^2\Big) \\
&\quad + C\sum_{k=0}^n \Big(|\hat{X}_k^{i,N}|^2|\Delta W_k^i|^2\Big) + C\sum_{k=0}^n \Big(\frac{1}{N}\sum_{j=1}^N |\hat{X}_k^{j,N}|^2|\Delta W_k^i|^2\Big) + C\sum_{k=0}^n \Big(h\Big) \\
&\quad + C\sum_{k=0}^n \Big(|\hat{X}_k^{i,N}|^2h\Big) + C\sum_{k=0}^n \Big(\frac{1}{N}\sum_{j=1}^N |\hat{X}_k^{j,N}|^2h\Big).
\end{aligned}$$

Taking power $p \geq 2$ on both sides, expectations and re-organising the terms, we have (with C_p independent of h, N)

$$\begin{aligned}
&\mathbb{E}\Big[|\hat{X}_{n+1}^{i,N}|^{2p}\Big] \leq C_p\Big(\mathbb{E}\Big[|\hat{X}_0^{i,N}|^{2p}\Big] + \mathbb{E}\Big[\Big(\sum_{k=0}^n \Big(\langle Y_k^{i,\star,N}, \sigma(t_k, Y_k^{i,\star,N}, \hat{\mu}_k^{Y,N})\Delta W_k^i\rangle\Big)\Big)^p\Big] \\
&\quad + \mathbb{E}\Big[\Big(\sum_{k=0}^n \Big(|\Delta W_k^i|^2\Big)\Big)^p\Big] + \mathbb{E}\Big[\Big(\sum_{k=0}^n \Big(|\hat{X}_k^{i,N}|^2|\Delta W_k^i|^2\Big)\Big)^p\Big] + \mathbb{E}\Big[\Big(\sum_{k=0}^n \Big(|\hat{X}_k^{i,N}|^2h\Big)\Big)^p\Big] \\
&\quad + \mathbb{E}\Big[\Big(p\sum_{k=0}^n \Big(h\Big)\Big)^p\Big] + \mathbb{E}\Big[\Big(\sum_{k=0}^n \Big(\frac{1}{N}\sum_{j=1}^N |\hat{X}_k^{j,N}|^2|\Delta W_k^i|^2\Big)\Big)^p\Big] + \mathbb{E}\Big[\Big(\sum_{k=0}^n \Big(\frac{1}{N}\sum_{j=1}^N |\hat{X}_k^{j,N}|^2h\Big)\Big)^p\Big]\Big).
\end{aligned}$$

There are 8 terms to be estimated, but in essence only 3 arguments are needed. We present them only for the most complex terms since for the remaining ones it is just a simplification of the arguments presented. We present them in the form of supremum over $n \in [\![0, M-1]\!]$. We start with the last term of the 2nd line: apply Jensen's inequality twice after scaling the outer summation and then tower property to take advantage of the conditional independence between $\hat{X}_k^{j,N}$ and ΔW_k^i, namely, (recall $h = T/M$)

$$\mathbb{E}\Big[\sup_{n\in[\![0,M-1]\!]}\Big(\sum_{k=0}^{n}\Big(\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_k^{j,N}|^2|\Delta W_k^i|^2\Big)\Big)^p\Big] \le \mathbb{E}\Big[\frac{1}{M}\sum_{k=0}^{M-1}\Big(\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_k^{j,N}|^{2p}|\Delta W_k^i|^{2p}\Big)M^p\Big]$$
$$= h\mathbb{E}\Big[\sum_{k=0}^{M-1}\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_k^{j,N}|^{2p}\Big]CT^p.$$

We now address the second term in the 1st line. Using Burkholder–Davis–Gundy (BDG) inequality and Jensen's inequality as above, the Lipschitz property of σ, Jensen's inequality and the domination of $Y_k^{i,\star,N}$ by $\hat{X}_k^{i,N}$ in Lemma 2.4.2 gives

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{n\in[\![0,M-1]\!]}\Big(\sum_{k=0}^{n}\Big(\langle Y_k^{i,\star,N},\sigma(t_k,Y_k^{i,\star,N},\hat{\mu}_k^{Y,N})\Delta W_k^i\rangle\Big)\Big)^p\Big]\\
\le& C\mathbb{E}\Big[\Big(\sum_{k=0}^{M-1}\Big(|\langle Y_k^{i,\star,N},\sigma(t_k,Y_k^{i,\star,N},\hat{\mu}_k^{Y,N})\rangle|^2 h\Big)\Big)^{\frac{p}{2}}\Big]\\
\le& C\mathbb{E}\Big[\Big(\frac{1}{M}\sum_{k=0}^{M-1}\Big(1+|Y_k^{i,\star,N}|^4+\frac{1}{N}\sum_{j=1}^{N}|Y_k^{j,\star,N}|^4\Big)h\Big)^{\frac{p}{2}}\Big]M^{\frac{p}{2}}\\
\le& C\mathbb{E}\Big[1+\Big[\sum_{k=0}^{M-1}|\hat{X}_k^{i,N}|^{2p}\Big]h+\Big[\sum_{k=0}^{M-1}\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_k^{j,N}|^{2p}\Big]h\Big].
\end{aligned}$$

Finally, we address the last term in the 3rd line of the inequality. The result follows by applying Jensen's inequality

$$\mathbb{E}\Big[\sup_{n\in[\![0,M-1]\!]}\Big(\sum_{k=0}^{n}\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_k^{j,N}|^2 h\Big)^p\Big] = \mathbb{E}\Big[\Big(\frac{1}{M}\sum_{k=0}^{M-1}\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_k^{j,N}|^2 h\Big)^p M^p\Big]$$
$$\le h\mathbb{E}\Big[\sum_{k=0}^{M-1}\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_k^{j,N}|^{2p}\Big]T^p.$$

Collecting the several inequalities and injecting them back in the initial inequality, we conclude that

$$\mathbb{E}\Big[\sup_{n\in[\![0,M-1]\!]}|\hat{X}_{n+1}^{i,N}|^{2p}\Big] \le C\Big(1+\mathbb{E}\Big[|\hat{X}_0^{i,N}|^{2p}\Big]+\mathbb{E}\Big[\sum_{k=0}^{M-1}|\hat{X}_k^{i,N}|^{2p}h\Big]+\mathbb{E}\Big[\sum_{k=0}^{M-1}\frac{1}{N}\sum_{j=1}^{N}|\hat{X}_k^{j,N}|^{2p}h\Big]\Big).$$

Taking supremum and using that the particles are conditional i.d.d. (for fixed k)

$$\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[\sup_{n\in[\![0,M]\!]}|\hat{X}_n^{i,N}|^{2p}\Big] \le C\Big(1+\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}_0^{i,N}|^{2p}\Big]+\sum_{k=0}^{M-1}\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[\sup_{0\le n\le k}|\hat{X}_n^{i,N}|^{2p}\Big]h\Big).$$

The proof finishes after applying the discrete Grönwall's inequality to the inequality, and using that the $\hat{X}_0^{i,N}$ are independent and identically distributed (i.i.d). □

We now provide (moment) estimates for the continuous-time extension of the SSM.

Proposition 2.4.6. *Choose h as in Theorem 2.2.6. Take $(\hat{X}_t^{i,N})_{t\in[0,T]}$ as the map satisfying (2.25), i.e., the continuous time extension of the SSM. Then, for any for $\frac{m}{2} \ge p \ge 1$, there exist $C \in \mathbb{R}^+$:*

$$\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[\sup_{0\le t\le T}|\hat{X}_t^{i,N}|^{2p}\Big] < C\Big(1+\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}_0^{i,N}|^{2p}\Big]\Big) < \infty. \tag{2.27}$$

Proof. Let $i \in [\![1, N]\!]$ and $n \in [\![0, M-1]\!]$. From (2.25), set $t_n + s = t$, for all $t \in [0, T], n \in [\![0, M-1]\!]$, then

$$\hat{X}_t^{i,N} = \hat{X}_n^{i,N} + v_h(t_n, \hat{X}_n^{i,N}, \hat{\mu}_n^N)s + b_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)s + \sigma_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)(W_{t_n+s}^i - W_{t_n}^i). \tag{2.28}$$

Plugging (2.5) in (2.28) gives

$$\hat{X}_t^{i,N} = \hat{X}_n^{i,N}(1 - \frac{s}{h}) + \frac{s}{h} Y_n^{i,\star,N} + b_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)s + \sigma_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)(W_{t_n+s}^i - W_{t_n}^i).$$

Since $s \leq h$, by the Lipschitz conditions on b_h and σ_h, and Lemma 2.4.2, we have

$$\begin{aligned}
|\hat{X}_t^{i,N}|^2 \leq C\Big(1 &+ |\hat{X}_n^{i,N}|^2 + \frac{1}{N}\sum_{j=1}^N |\hat{X}_n^{j,N}|^2 + |b_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)|^2 \\
&+ \Big|\sigma_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)(W_{t_n+s}^i - W_{t_n}^i)\Big|^2\Big) \\
&\leq C\Big(1 + |\hat{X}_n^{i,N}|^2 + \frac{1}{N}\sum_{j=1}^N |\hat{X}_n^{j,N}|^2 + \Big|\sigma_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)(W_{t_n+s}^i - W_{t_n}^i)\Big|^2\Big).
\end{aligned}$$

Taking supremum over time and expectations on both sides yields

$$\begin{aligned}
\mathbb{E}\Big[\sup_{0\leq t\leq T} |\hat{X}_t^{i,N}|^{2p}\Big] &= \mathbb{E}\Big[\sup_{n\in[\![0,M-1]\!]}\sup_{0\leq s\leq h} |\hat{X}_{t_n+s}^{i,N}|^{2p}\Big] \\
&\leq C\mathbb{E}\Big[1 + \sup_{n\in[\![0,M-1]\!]}\Big\{\Big(|\hat{X}_n^{i,N}|^{2p} + \frac{1}{N}\sum_{j=1}^N |\hat{X}_n^{j,N}|^{2p}\Big) + I_h^{i,n}\Big\}\Big].
\end{aligned}$$

where $I_h^{i,n}$ is given by $I_h^{i,n} := \sup_{0\leq s\leq h}\Big|\sigma_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)(W_{t_n+s}^i - W_{t_n}^i)\Big|^{2p}$. Using the BDG inequality, Jensen's inequality, the Lipschitz condition on σ_h and Proposition 2.4.5, gives

$$\mathbb{E}[I_h^{i,n}] \leq C\mathbb{E}\Big[\Big(|\sigma_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)|^2 h\Big)^p\Big] \leq Ch^p\mathbb{E}\Big[1 + |\hat{X}_n^{i,N}|^{2p} + \frac{1}{N}\sum_{j=1}^N |\hat{X}_n^{j,N}|^{2p}\Big] \leq Ch^p.$$

Take supremum over $i \in [\![1, N]\!]$, then by Proposition 2.4.5 it follows that

$$\begin{aligned}
\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[\sup_{0\leq t\leq T} |\hat{X}_t^{i,N}|^{2p}\Big] &\leq \sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[C(1 + \sup_{0\leq n\leq M} |\hat{X}_n^{i,N}|^{2p} + Ch^p)\Big] \\
&\leq C\Big(1 + \sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}_0^{i,N}|^{2p}\Big]\Big) < \infty.
\end{aligned}$$

□

The last result in this block concerns incremental (in time) moment bounds of $\hat{X}^{i,N}$.

Proposition 2.4.7. *There exists $C \in \mathbb{R}^+$ such that for any $p \geq 2$, with $m \geq (q+1)p$,*

$$\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[\sup_{0\leq t\leq T} |\hat{X}_t^{i,N} - \hat{X}_{\kappa(t)}^{i,N}|^p\Big] \leq Ch^{\frac{p}{2}}. \tag{2.29}$$

Proof. From (2.25), for all $t \in [0, T]$ such that $t \in [t_n, t_{n+1}]$ set $s \in [0, h]$ such that $t_n + s = t$. Then

$$\hat{X}_t^{i,N} = \hat{X}_n^{i,N} + v_h(t_n, \hat{X}_n^{i,N}, \hat{\mu}_n^N)s + b_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)s + \sigma_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)(W_t^i - W_{t_n}^i).$$

Thus, we have a constant C_p only depending on p such that

$$|\hat{X}_t^{i,N} - \hat{X}_n^{i,N}|^p \leq C_p\Big(|v_h(t_n, \hat{X}_n^{i,N}, \hat{\mu}_n^N)|^p h^p + |b_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)|^p h^p + |\sigma_h(t_n, \hat{X}_n^{i,N}, \widetilde{\mu}_n^N)|^p |(W_t^i - W_{t_n}^i)|^p\Big).$$

Take sup over time and expectations on both sides. Using Assumption 2.2.1, one deals with the last term with the BDG inequality (using conditional expectations via $\mathbb{E}[\cdot|\mathcal{F}_{t_n}]$) and Jensen's inequality. From Lemma 2.4.1 and Propositions 2.4.5 and 2.4.6, there exists a positive C independent of h, N, M such that

$$\sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[\sup_{0\leq t\leq T} |\hat{X}_t^{i,N} - \hat{X}_{\kappa(t)}^{i,N}|^p\Big] \leq C(h^p + h^{\frac{p}{2}})\Big(1 + \sup_{i\in[\![1,N]\!]} \mathbb{E}\Big[\sup_{0\leq n\leq M} |\hat{X}_n^{i,N}|^{(q+1)p}\Big]\Big) \leq Ch^{\frac{p}{2}},$$

where q follows from the polynomial growth property of v in Assumption 2.2.1. □

Local errors

After having discussed moment bounds, we now discuss the local error.

Proposition 2.4.8. *Let the assumptions of Theorem 2.2.6 hold. Take the functions v, b, σ and the corresponding functions v_h, b_h, σ_h as defined in Lemma 2.4.1.*

Then, there exist positive constants C_1, C_2, C_3 and $q' = 2(q+1)^2$, such that for all $t \in [0, T]$, $i \in [\![1, N]\!]$, $z_i \in \mathbb{R}^d$, and the collection $\{z_i\}_i$, we have

$$|v_h(t, z_i, \mu^z) - v(t, z_i, \mu^z)|^2 \leq C_1\Big(1 + |z_i|^{q'} + \frac{1}{N}\sum_{j=1}^N |z_j|^{q'}\Big)h^2, \tag{2.30}$$

$$|b_h(t, z_i, \mu^{F_{h,z,\mu^z}}) - b(t, z_i, \mu^z)|^2 \leq C_2\Big(1 + |z_i|^{q'} + \frac{1}{N}\sum_{j=1}^N |z_j|^{q'}\Big)h^2, \tag{2.31}$$

$$|\sigma_h(t, z_i, \mu^{F_{h,z,\mu^z}}) - \sigma(t, z_i, \mu^z)|^2 \leq C_3\Big(1 + |z_i|^{q'} + \frac{1}{N}\sum_{j=1}^N |z_j|^{q'}\Big)h^2. \tag{2.32}$$

where μ^z and $\mu^{F_{h,z,\mu^z}}$ are the two empirical measures associated with $\{z_i\}_i$ and $\{F_h(t, z_i, \mu^z)\}_i$ respectively, i.e.,

$$\mu^z(dz) = \frac{1}{N}\sum_{j=1}^N \delta_{z_j}(dz), \quad \text{and} \quad \mu^{F_{h,z,\mu^z}}(dz) = \frac{1}{N}\sum_{j=1}^N \delta_{\{F_h(t,z_j,\mu^z)\}}(dz).$$

Proof. Recall the estimates given in Lemma 2.4.1. Using the identity $F_h(t, z_i, \mu^z) = z_i + hv_h(t, z_i, \mu^z)$, (2.21), Assumption 2.2.1, Young's inequality and Jensen's inequality, we have

$$\begin{aligned}
\Big|v_h(t, z_i, \mu^z) - v(t, z_i, \mu^z)\Big|^2 &= \Big|v\Big(t, z_i + hv_h(t, z_i, \mu^z), \mu^z\Big) - v(t, z_i, \mu^z)\Big|^2 \\
&\leq C\Big(1 + |z_i + hv_h(t, z_i, \mu^z)|^q + |z_i|^q\Big)^2 h^2 |v_h(t, z_i, \mu^z)|^2 \\
&\leq C\Big(1 + |z_i|^{2q} + |z_i|^{2q(q+1)}h^{2q} + h^{2q}\frac{1}{N}\sum_{j=1}^N |z_j|^{2q}\Big)\frac{h^2}{(1-hL_v)^2}\Big(1 + |z_i|^{2q+2} + \frac{1}{N}\sum_{j=1}^N |z_j|^2\Big) \\
&\leq C\Big(1 + |z_i|^{2(q+1)(q+1)} + \frac{1}{N}\sum_{j=1}^N |z_j|^{4q+2}\Big)h^2.
\end{aligned}$$

As in Lemma 2.4.1 we show only the result for b_h as the computation is the same for σ_h (and overall very close to that for v_h). Using the definition of b_h, the Lipschitz property of b and the definition of $\mu^{F_{h,z,\mu^z}}, \mu^z$, using similar calculations as above, by Young's inequality and Jensen's inequality, we have

$$
\begin{aligned}
\Big|b_h(t,z_i,\mu^{F_{h,z,\mu^z}})-b(t,z_i,\mu^z)\Big|^2 &= \Big|b\Big(t,F_h(t,z_i,\mu^z),\mu^{F_{h,z,\mu^z}}\Big)-b(t,z_i,\mu^z)\Big|^2\\
&\le L_b\Big(h^2|v_h(t,z_i,\mu^z)|^2+\big(W^{(2)}(\mu^{F_{h,z,\mu^z}},\mu^z)\big)^2\Big)\\
&\le L_b\Big(h^2|v_h(t,z_i,\mu^z)|^2+\frac{1}{N}\sum_{j=1}^N|F_h(t,z_j,\mu^z)-z_j|^2\Big)\\
&\le L_b\Big(h^2|v_h(t,z_i,\mu^z)|^2+\frac{1}{N}\sum_{j=1}^N h^2|v_h(t,z_j,\mu^z)|^2\Big).
\end{aligned}
$$

Applying Inequality (2.21) and the growth in v from Assumption 2.2.1 (as in the previous proof), we have the claim.

□

Proposition 2.4.9. *Let the assumptions of Theorem 2.2.6 holds with $m \ge 2(q+1)^2$. Let $i \in [\![1,N]\!]$ and take $X^{i,N}$ to be the solution to the interacting particle system* (2.2) *and let $\hat{X}^{i,N}$ be the continuous-time extension of the SSM given by* (2.25)*. We then have*

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

Proof. Take $i\in[\![1,N]\!]$, $t\in[0,T]$. From (2.4) and (2.27), both $X^{i,N}$ and $\hat{X}^{i,N}$ have bounded $2p$-moments ($p\ge 2$). Define the auxiliary quantity $\Delta X^i := X^{i,N}-\hat{X}^{i,N}$. Itô's formula applied to $|X_t^{i,N}-\hat{X}_t^{i,N}|^2=|\Delta X_t^i|^2$ yields

$$
\begin{aligned}
|\Delta X_t^i|^2 =&2\int_0^t\Big\langle v(s,X_s^{i,N},\mu_s^N)-v_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N),\Delta X_s^i\Big\rangle ds &(2.33)\\
&+2\int_0^t\Big\langle b(s,X_s^{i,N},\mu_s^N)-b_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\widetilde{\mu}_{\kappa(s)}^N),\Delta X_s^i\Big\rangle ds &(2.34)\\
&+\int_0^t\Big|\sigma(s,X_s^{i,N},\mu_s^N)-\sigma_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\widetilde{\mu}_{\kappa(s)}^N)\Big|^2 ds &(2.35)\\
&+2\int_0^t\Big\langle \Delta X_s^i,\Big(\sigma(s,X_s^{i,N},\mu_s^N)-\sigma_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\widetilde{\mu}_{\kappa(s)}^N)\Big)dW_s^i\Big\rangle. &(2.36)
\end{aligned}
$$

We analyse the components term by term. Namely, for (2.33)

$$
\begin{aligned}
&\Big\langle v(s,X_s^{i,N},\mu_s^N)-v_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N),\Delta X_s^i\Big\rangle\\
&\quad=\Big\langle v(s,X_s^{i,N},\mu_s^N)-v(s,\hat{X}_s^{i,N},\hat{\mu}_s^N),\Delta X_s^i\Big\rangle+\Big\langle v(s,\hat{X}_s^{i,N},\hat{\mu}_s^N)-v(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N),\Delta X_s^i\Big\rangle\\
&\qquad+\Big\langle v(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N)-v_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N),\Delta X_s^i\Big\rangle.
\end{aligned}
$$

From the Assumption 2.2.1, take supremum over $[0,T]$ and expectations, by the Young's inequality, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle v(s,X_s^{i,N},\mu_s^N)-v(s,\hat{X}_s^{i,N},\hat{\mu}_s^N),\Delta X_s^i\Big\rangle ds\Big]\\
&\le \mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle v(s,X_s^{i,N},\mu_s^N)-v(s,\hat{X}_s^{i,N},\mu_s^N)+v(s,\hat{X}_s^{i,N},\mu_s^N)-v(s,\hat{X}_s^{i,N},\hat{\mu}_s^N),\Delta X_s^i\Big\rangle ds\Big]\\
&\le \mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big[L_v|\Delta X_s^i|^2+\frac{1}{2}\Big|v(s,\hat{X}_s^{i,N},\mu_s^N)-v(s,\hat{X}_s^{i,N},\hat{\mu}_s^N)\Big|^2+\frac{1}{2}|\Delta X_s^i|^2\Big]ds\Big]\\
&\le \mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big[(L_v+\frac{1}{2})|\Delta X_s^i|^2+\frac{L_{\tilde{v}}}{2}W^{(2)}(\mu_s^N,\hat{\mu}_s^N)\Big]ds\Big]\\
&\le C\mathbb{E}\Big[\int_0^T \Big(|\Delta X_s^i|^2+\frac{1}{N}\sum_{j=1}^N|\Delta X_s^j|^2\Big)ds\Big]. \qquad (2.37)
\end{aligned}
$$

By the 1/2-Hölder regularity in time and the assumption on v, the particles being i.i.d. and the Cauchy-Schwarz inequality we have

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle v(s,\hat{X}_s^{i,N},\hat{\mu}_s^N)-v(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N),\Delta X_s^i\Big\rangle ds\Big]\\
&\le \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big|v(s,\hat{X}_s^{i,N},\hat{\mu}_s^N)-v(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N)\Big|^2ds\Big]+\frac{1}{2}\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2ds\Big]\\
&\le Ch+C\mathbb{E}\Big[\int_0^T\Big(\Big(1+|\hat{X}_s^{i,N}|^{2q}+|\hat{X}_{\kappa(s)}^{i,N}|^{2q}\Big)|\hat{X}_s^{i,N}-\hat{X}_{\kappa(s)}^{i,N}|^2+\frac{1}{N}\sum_{j=1}^N|\hat{X}_s^{j,N}-\hat{X}_{\kappa(s)}^{j,N}|^2\Big)ds\Big]\\
&\quad+\frac{1}{2}\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2ds\Big]\\
&\le Ch+\frac{1}{2}\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2ds\Big]+\frac{C}{N}\sum_{j=1}^N\int_0^T\mathbb{E}\Big[|\hat{X}_s^{j,N}-\hat{X}_{\kappa(s)}^{j,N}|^2\Big]\,ds\\
&\quad+C\int_0^T\sqrt{\mathbb{E}\Big[\Big(1+|\hat{X}_s^{i,N}|^{2q}+|\hat{X}_{\kappa(s)}^{i,N}|^{2q}\Big)^2\Big]\mathbb{E}\Big[|\hat{X}_s^{i,N}-\hat{X}_{\kappa(s)}^{i,N}|^4\Big]}\,ds\\
&\le Ch+\frac{1}{2}\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2ds\Big]. \qquad (2.38)
\end{aligned}
$$

where in the last inequality we used Hölder's inequality on the product term in combination with Proposition 2.4.6 and 2.4.7 with $m\ge 2(q+1)^2$ to guarantee the error satisfies $\mathbb{E}\big[|\hat{X}_s^{i,N}-\hat{X}_{\kappa(s)}^{i,N}|^4\big]\le Ch^2$. We now make use of Proposition 2.4.8 and arguments similar to those above to deal with the last term of the initial inequality

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le t\le T}\int_0^t \Big\langle v(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N)-v_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N),\Delta X_s^i\Big\rangle ds\Big]\\
&\le\frac{1}{2}\mathbb{E}\Big[\int_0^T\Big[\Big|v(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N)-v_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\hat{\mu}_{\kappa(s)}^N)\Big|^2+|\Delta X_s^i|^2\Big]ds\Big]\\
&\le\mathbb{E}\Big[\int_0^T C\Big[1+|\hat{X}_{\kappa(s)}^{i,N}|^{q'}+\frac{1}{N}\sum_{j=1}^N|\hat{X}_{\kappa(s)}^{j,N}|^{q'}\Big]h^2ds\Big]+\frac{1}{2}\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2ds\Big]\\
&\le Ch^2+\frac{1}{2}\mathbb{E}\Big[\int_0^T|\Delta X_s^i|^2ds\Big]. \qquad (2.39)
\end{aligned}
$$

where q' defined in Proposition 2.4.8 such that $m\ge 2(q+1)^2=q'$ as to guarantee $\mathbb{E}\big[|\hat{X}_{\kappa(s)}^{i,N}|^{q'}\big]\le$

C. We now proceed to estimate the b components. Using Young's inequality, for (2.34)

$$\begin{aligned}
&\Big\langle b(s, X_s^{i,N}, \mu_s^N) - b_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N), \Delta X_s^i \Big\rangle \\
&\leq \frac{1}{2}\Big| b(s, X_s^{i,N}, \mu_s^N) - b(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big|^2 + \frac{1}{2}|\Delta X_s^i|^2 \\
&\leq \Big| b(s, X_s^{i,N}, \mu_s^N) - b(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \hat{\mu}_{\kappa(s)}^N)\Big|^2 \\
&\quad + \Big| b(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \hat{\mu}_{\kappa(s)}^N) - b_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big|^2 + |\Delta X_s^i|^2. \qquad (2.40)
\end{aligned}$$

For the first term above, the Lipschitz condition on b, (2.22) and Proposition 2.4.7, yield

$$\begin{aligned}
&\Big| b(s, X_s^{i,N}, \mu_s^N) - b(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \hat{\mu}_{\kappa(s)}^N)\Big|^2 \leq C\Big(h + |X_s^{i,N} - \hat{X}_{\kappa(s)}^{i,N}|^2 + \frac{1}{N}\sum_{j=1}^N |X_s^{j,N} - \hat{X}_{\kappa(s)}^{j,N}|^2 \Big) \\
&\leq C\Big(h + |\Delta X_s^i|^2 + |\hat{X}_s^{i,N} - \hat{X}_{\kappa(s)}^{i,N}|^2 + \frac{1}{N}\sum_{j=1}^N |\Delta X_s^j|^2 + \frac{1}{N}\sum_{j=1}^N |\hat{X}_s^{j,N} - \hat{X}_{\kappa(s)}^{j,N}|^2 \Big).
\end{aligned}$$

and similarly we obtain:

$$\begin{aligned}
&\Big| \sigma(s, X_s^{i,N}, \mu_s^N) - \sigma_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big|^2 \\
&\leq C\Big[h + |\Delta X_s^i|^2 + |\hat{X}_s^{i,N} - \hat{X}_{\kappa(s)}^{i,N}|^2 + \frac{1}{N}\sum_{j=1}^N |\Delta X_s^j|^2 + \frac{1}{N}\sum_{j=1}^N |\hat{X}_s^{j,N} - \hat{X}_{\kappa(s)}^{j,N}|^2 \Big] \\
&\quad + 2\Big| \sigma(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \hat{\mu}_{\kappa(s)}^N) - \sigma_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big|^2. \qquad (2.41)
\end{aligned}$$

Consider the last term (2.36), take expectations, using the BDG inequality, Cauchy-Schwarz inequality and Proposition 2.4.8,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{0\leq t\leq T} \int_0^t \Big\langle \Delta X_s^i, \Big(\sigma(s, X_s^{i,N}, \mu_s^N) - \sigma_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big) dW_s^i \Big\rangle \Big] \\
&\leq \mathbb{E}\Big[C\Big(\int_0^T |\Delta X_s^i|^2 \Big| \sigma(s, X_s^{i,N}, \mu_s^N) - \sigma_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big|^2 ds\Big)^{\frac{1}{2}} \Big] \\
&\leq \mathbb{E}\Big[\Big(\frac{1}{4}\sup_{0\leq t\leq T} |\Delta X_t^i|^2 \; C\int_0^T \Big| \sigma(s, X_s^{i,N}, \mu_s^N) - \sigma_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big|^2 ds\Big)^{\frac{1}{2}} \Big] \\
&\leq \frac{1}{4}\mathbb{E}\Big[\sup_{0\leq t\leq T} |\Delta X_t^i|^2 \Big] C\mathbb{E}\Big[\int_0^T \Big| \sigma(s, X_s^{i,N}, \mu_s^N) - \sigma_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big|^2 ds \Big] \\
&\leq \frac{1}{4}\mathbb{E}\Big[\sup_{0\leq t\leq T} |\Delta X_t^i|^2 \Big] + \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big(\Big| \sigma(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \hat{\mu}_{\kappa(s)}^N) - \sigma_h(\kappa(s), \hat{X}_{\kappa(s)}^{i,N}, \widetilde{\mu}_{\kappa(s)}^N)\Big|^2 ds\Big) \Big] \\
&\quad + C\mathbb{E}\Big[\int_0^T \Big(h + |\Delta X_s^i|^2 + |\hat{X}_s^{i,N} - \hat{X}_{\kappa(s)}^{i,N}|^2 + \frac{1}{N}\sum_{j=1}^N |\Delta X_s^j|^2 + \frac{1}{N}\sum_{j=1}^N |\hat{X}_s^{j,N} - \hat{X}_{\kappa(s)}^{j,N}|^2 ds\Big) \Big] \\
&\qquad (2.42) \\
&\leq \frac{1}{4}\mathbb{E}\Big[\sup_{0\leq t\leq T} |\Delta X_t^i|^2 \Big] + C\mathbb{E}\Big[\int_0^T \Big(h + |\Delta X_s^i|^2 + \frac{1}{N}\sum_{j=1}^N |\Delta X_s^j|^2 ds\Big) \Big]. \qquad (2.43)
\end{aligned}$$

where we used (2.41) to reach (2.42), the last inequality (2.43) follows by Proposition 2.4.6, 2.4.7 and 2.4.8. Similarly, for the terms (2.40) and (2.41), by Proposition 2.4.8 and similar arguments as in (2.39), we conclude that

$$\mathbb{E}\Big[\sup_{0\le t\le T}\Big(2\int_0^t \Big\langle b(s,X_s^{i,N},\mu_s^N)-b_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\widetilde{\mu}_{\kappa(s)}^N),\Delta X_s^i\Big\rangle ds$$

$$+\int_0^t \Big|\sigma(s,X_s^{i,N},\mu_s^N)-\sigma_h(\kappa(s),\hat{X}_{\kappa(s)}^{i,N},\widetilde{\mu}_{\kappa(s)}^N)\Big|^2 ds\Big)\Big]$$

$$\le C\mathbb{E}\Big[\int_0^T \Big(h+|\Delta X_s^i|^2+\frac{1}{N}\sum_{j=1}^N|\Delta X_s^j|^2 ds\Big)\Big]. \tag{2.44}$$

Gathering all inequalities (2.37), (2.38), (2.39), (2.43) and (2.44) together, taking supremum on i, since the particles are i.i.d., we conclude (where h is the leading term)

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

$$\le \sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[C\int_0^T\Big(h+|X_s^{i,N}-\hat{X}_s^{i,N}|^2+\frac{1}{N}\sum_{j=1}^N|X_s^{j,N}-\hat{X}_s^{j,N}|^2\Big)ds\Big]$$

$$\le Ch+C\int_0^T\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[\sup_{0\le u\le s}|X_u^{i,N}-\hat{X}_u^{i,N}|^2\Big]ds.$$

Grönwall's lemma delivers the final result. □

Now, the proof of the main theorem is concluded as follow.

Proof of Theorem 2.2.6. In relation to Points 1, 2 and 3 in the theorem's statement: Point 1 follows from Proposition 2.4.6; Point 2 follows from Proposition 2.4.9; the last point follows by a straightforward combination of Proposition 2.2.4 and Proposition 2.4.9. □

2.4.2 Proof of the stability Theorem, Theorem 2.2.9

Proof of Theorem 2.2.9. Let $i\in[\![1,N]\!]$ and $n\in\mathbb{N}$. From (2.23) in Lemma 2.4.1, since the particles are identically distributed, we have

$$\mathbb{E}\Big[|Y_n^{i,\star,N}-G_n^{i,\star,N}|^2\Big]\le\mathbb{E}\Big[\frac{1}{1-h(2L_v+1)}\Big(|\hat{X}_n^{i,N}-\hat{Z}_n^{i,N}|^2+L_{\tilde{v}}hW^{(2)}(\hat{\mu}_n^{X,N},\hat{\mu}_n^{Z,N})\Big)\Big]$$

$$\le\frac{1+L_{\tilde{v}}h}{1-h(2L_v+1)}\mathbb{E}\Big[|\hat{X}_n^{i,N}-\hat{Z}_n^{i,N}|^2\Big].$$

By definition of the SSM, one also has $(Y_n^{i,\star,N}-\hat{X}_n^{i,N})=hv\big(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{X,N}\big)$ and $(G_n^{i,\star,N}-\hat{Z}_n^{i,N})=hv\big(t_n,G_n^{i,\star,N},\hat{\mu}_n^{Z,N}\big)$. Thus, from Assumption 2.2.1, for any n, the Cauchy-Schwarz inequality yields

$$
\begin{aligned}
&\mathbb{E}\Big[|\hat{X}^{i,N}_{n+1}-\hat{Z}^{i,N}_{n+1}|^2\Big]\\
&\leq \mathbb{E}\Big[|Y^{i,\star,N}_n-G^{i,\star,N}_n|^2+2\Big\langle Y^{i,\star,N}_n-G^{i,\star,N}_n, b(t_n,Y^{i,\star,N}_n,\hat{\mu}^{Y,N}_n)-b(t_n,G^{i,\star,N}_n,\hat{\mu}^{G,N}_n)\Big\rangle h\\
&\quad+\big|b(t_n,Y^{i,\star,N}_n,\hat{\mu}^{Y,N}_n)-b(t_n,G^{i,\star,N}_n,\hat{\mu}^{G,N}_n)\big|^2h^2\\
&\quad+\big|\sigma(t_n,Y^{i,\star,N}_n,\hat{\mu}^{Y,N}_n)-\sigma(t_n,G^{i,\star,N}_n,\hat{\mu}^{G,N}_n)\big|^2(\Delta W^i_n)^2\Big]\\
&\leq (1+L_\sigma h+L_b h^2)\mathbb{E}\Big[|Y^{i,\star,N}_n-G^{i,\star,N}_n|^2\Big]+(L_{\tilde{\sigma}}h+L_{\tilde{b}}h^2)\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|Y^{j,\star,N}_n-G^{j,\star,N}_n|^2\Big]\Big]\\
&\quad+2h\sqrt{\mathbb{E}\Big[|Y^{i,\star,N}_n-G^{i,\star,N}_n|^2\Big]}\sqrt{\mathbb{E}\Big[\big|b(t_n,Y^{i,\star,N}_n,\hat{\mu}^{Y,N}_n)-b(t_n,G^{i,\star,N}_n,\hat{\mu}^{Y,N}_n)\big|^2\Big]}\\
&\quad+2h\sqrt{\mathbb{E}\Big[|Y^{i,\star,N}_n-G^{i,\star,N}_n|^2\Big]}\sqrt{\mathbb{E}\Big[\big|b(t_n,G^{i,\star,N}_n,\hat{\mu}^{Y,N}_n)-b(t_n,G^{i,\star,N}_n,\hat{\mu}^{G,N}_n)\big|^2\Big]}\\
&\leq \Big(1+(2\sqrt{L_b}+2\sqrt{L_{\tilde{b}}}+L_\sigma+L_{\tilde{\sigma}})h+(L_b+L_{\tilde{b}})h^2\Big)\mathbb{E}\Big[|Y^{i,\star,N}_n-G^{i,\star,N}_n|^2\Big].
\end{aligned}
$$

where we used the tower property of the expectation with $\mathcal{F}_{t_n}$-conditional expectations to deal with the Brownian increment term (it holds that $\mathbb{E}[\,|\Delta W^i_n|^2|\mathcal{F}_{t_n}]=h$ after using that all $Y^{j,\star,N}_n$, $G^{j,\star,N}_n$ are $\mathcal{F}_{t_n}$-adapted), the Cauchy-Schwarz inequality and that the particles are identically distributed.

Taking supremum over i and using (2.23) yields

$$
\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}^{i,N}_{n+1}-\hat{Z}^{i,N}_{n+1}|^2\Big]\leq(1+\beta h)\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}^{i,N}_n-\hat{Z}^{i,N}_n|^2\Big].
$$

where β and α are exactly given by (2.9). A straightforward induction argument leads to

$$
\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}^{i,N}_n-\hat{Z}^{i,N}_n|^2\Big]\leq\big(1+\beta h\big)^n\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}^{i,N}_0-\hat{Z}^{i,N}_0|^2\Big].
$$

Recall (2.9) for the expression for β. If $L_v\geq-(1+L_{\tilde{v}}+A)/2$ then $1+\beta h>1$ and hence $\lim_{n\to\infty}\big(1+\beta h\big)^n\neq 0$, this implies that the SSM is not Mean-square contractive. On the other hand, since $(1+\beta h)$ is always positive then when $\beta<0$ and $(1+\beta h)\in(0,1)$ $\Leftrightarrow L_v<-(1+L_{\tilde{v}}+A)/2<-\frac{1}{2}$ with sufficient small h and consequently the SSM is guaranteed to be Mean-square contractive

$$
\lim_{n\to\infty}\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}^{i,N}_n-\hat{Z}^{i,N}_n|^2\Big]\leq\lim_{n\to\infty}(1+\beta h)^n\sup_{i\in[\![1,N]\!]}\mathbb{E}\Big[|\hat{X}^{i,N}_0-\hat{Z}^{i,N}_0|^2\Big]=0.
$$

□

Chapter 3

Numerical analysis for McKean-Vlasov SDEs with super-linear growth drifts in space and interaction

3.1 Introduction

In this chapter, we study the McKean-Vlasov Stochastic Differential Equations (MV-SDEs) with convolution-type drifts have general dynamics given by

$$\mathrm{d}X_t = \big(v(X_t,\mu_t^X) + b(t,X_t,\mu_t^X)\big)\mathrm{d}t + \sigma(t,X_t,\mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^m(\mathbb{R}^d), \tag{3.1}$$

$$\text{where } \ v(x,\mu) = \int_{\mathbb{R}^d} f(x-y)\mu(\mathrm{d}y) + u(x,\mu) \quad \text{with} \quad \mu_t^X = \mathrm{Law}(X_t), \tag{3.2}$$

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.

The corresponding IPS as an N-dimensional system of $\mathbb{R}^d$-valued interacting particles where each particle is governed by a Stochastic Differential Equation (SDE) is given below.

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 (3.3)

$$\mathrm{d}X_t^{i,N} = \big(v(X_t^{i,N},\mu_t^{X,N}) + b(t,X_t^{i,N},\mu_t^{X,N})\big)\mathrm{d}t + \sigma(t,X_t^{i,N},\mu_t^{X,N})\mathrm{d}W_t^i, \quad X_0^{i,N} = X_0^i \in L_0^m(\mathbb{R}^d), \tag{3.3}$$

$$\text{where } \begin{cases} v(X_t^{i,N},\mu_t^{X,N}) = \Big(\frac{1}{N}\sum_{j=1}^N f(X_t^{i,N} - X_t^{j,N})\Big) + u(X_t^{i,N},\mu_t^{X,N}) \\ \mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}(\mathrm{d}x), \end{cases} \tag{3.4}$$

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 (3.3), 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 [115], where he introduced the convolution type interaction (the v in (3.4)). This is a class of Markov processes associated with nonlinear parabolic equations where the map v in (3.4) is also called "self-stabilizing". The IPS underpinning our work (3.3)-(3.4) has been studied widely, from a variety of points of view and as early as [133] (for a general survey under global Lipschitz conditions and boundedness).

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 [2, 138] and further examples in [92]. 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 [2, 61, 83]. 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) [37, 38, 92]. The granular media Fokker-Plank equation from biochemistry is a good example of an equation featuring this kind of structure [2, 40, 111]. The literature on MV-SDE is growing explosively with many contributions addressing well-posedness, regularity, ergodicity, nonlinear Fokker-Planck equations, large deviations [3, 4, 61, 87]. The convolution framework has been given particular attention as it underpins many settings of interest [40, 82, 111, 138]. 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 [82] 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 $f(x) = x/|x|^2$ [72].

Super-linear interaction forces. For the IPS (3.3)-(3.4) or the MV-SDE (3.1)-(3.2), 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 (3.2) 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 3.2.1 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 [2]; [84] investigate different properties of the invariant measures for particles in double-well confining potential and later [138] investigate the convergence to stationary states. Large deviations and exit times for such self-stabilising diffusions are established in [2, 83]. The study of probabilistic properties and parametric inference (under constant diffusion) for this class is given in [74]. Two recent studies on parametric inference [17, 53] include numerical studies for the particle interaction ([74] does not) but do not tackle super-linear growth in the interaction component ([74] does).

To the best of our knowledge and except for [110], 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 (3.3) 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 system's dimension, Nd, and hence "explode" as N tends to infinity. In this chapter, 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 (3.3)-(3.4) 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 [2, 41, 101, 100]: taming [60, 98], time-adaptive [125], Split-Step Methods (SSM) methods in Chapter 2 – all these contributions allow for super-linear growth in space only. Further noteworthy contributions include [6, 16, 18, 29, 54, 62, 75, 91, 137]. 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 approxi-*

mation of interacting particle SDE systems (3.3)-(3.4)*, and McKean–Vlasov SDEs* (3.1)-(3.2).

The main contribution of this chapter is the numerical scheme and its convergence analysis. We extend the SSM algorithm in Chapter 2 to the 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 3.2.3 below) and Propagation of Chaos (Proposition 3.2.5 below) follow from known literature [2] – in fact, our Proposition 3.2.5 establishes the well-posedness of the particle system hence closing the small gap present in [2, Theorem 3.14]. The only existing work tackling this involved setting via a fully implicit scheme is [110]. 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 idea of the SSM 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 Chapter 2, 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. The SSM in Chapter 2 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 Chapter 2 there were already several competitive schemes present in the literature, e.g., taming [60, 98] and time-adaptive [125] 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 3.2.11, 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 (3.4) or (3.2) and Itô-type arguments – see Theorem 3.2.10. We provide a second strong (non-path-space) mean-square error criteria, see Theorem 3.13, 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 [22] (which are easier to obtain). We emphasise that this 2nd notion of strong convergence (see Theorem 3.13) 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 3.2.13).

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 [60] 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 this chapter. In Section 3.2 we set the notation and framework. In Section 3.2.2, we state the SSM scheme and the two main convergence results. Section 3.3 provides numerical illustrations (for the granular media model and a double-well model with non-constant diffusion). All proofs are given in Section 3.4.

[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 3.2.9 and 3.2.11 showcase two "strong" but different error metrics.

3.2 Main results

3.2.1 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 chapter is Equation (3.1) (for some $m \geq 1$), 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 3.2.1. *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 $L_b, L_\sigma \geq 0$ 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*

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

$(\mathbf{A}^u)$ *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*

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

$(\mathbf{A}^f)$ *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*

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

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 3.2.7 (see also Chapter 2). Certain useful properties can be derived from these assumptions.

Remark 3.2.2 (Implied properties)**.** *Under Assumption 3.2.1, 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

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

Using the properties of the convolution, v of (3.1) also satisfies a one-sided Lipschitz condition in space

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

[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.

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

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

We first recall a result from [2] establishing well-posedness of the MV-SDE (3.1)-(3.2).

Theorem 3.2.3 (Theorem 3.5 in [2])**.** *Let Assumption 3.2.1 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* (3.1) *in $\mathbb{S}^m([0,T])$. For some constant $C > 0$ (depending on T and m) we have*

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

Proof. Our Assumption 3.2.1 is a particularisation of [2, Assumption 3.4] and hence our theorem follows directly from [2, Theorem 3.5]. □

The interacting particle system (3.3)**.** As mentioned earlier, the numerical approximation results in this chapter apply directly if either one's starting point is the interacting particle system (3.3) or if one's starting point is the MV-SDE (3.1). On the latter, one can approximate the MV-SDE (3.1) (driven by the Brownian motion W) by the N-dimensional system $\mathbb{R}^d$-valued interacting particle system given in (3.3) and approximate it numerically with the gap closed by the Propagation of Chaos [47, 60, 125].

For completeness we recall the setup of (3.3). 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 (3.3) (with v given in (3.2))

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

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 (3.1); with a slight abuse of notation to avoid re-defining the probability space's filtration).

Remark 3.2.4 (The system through the lens of $\mathbb{R}^{Nd}$)**.** *We introduce the map V to interpret* (3.3) *as one system of equations in $\mathbb{R}^{Nd}$ instead of N dependent equations each in $\mathbb{R}^d$. Namely, we define v for $i \in [\![1, N]\!]$ given by* (3.2)*,*

$$V = (V_1, \cdots, V_N) : (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N, V_i : (\mathbb{R}^d)^N \to \mathbb{R}^d, \quad V_i(X^N) = v(X^{i,N}, \mu^{X,N}), \tag{3.5}$$

and $X^N = (X^{1,N}, \cdots, X^{N,N}) \in \mathbb{R}^{Nd}$ where each $X^{i,N}$ solves (3.3)*, with the Dirac measure $\mu^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X^{j,N}}(\mathrm{d}x)$.*

For $X^N, Y^N \in \mathbb{R}^{Nd}$ with corresponding measure $\mu^{X,N}, \mu^{Y,N}$ and letting Assumption 3.2.1 hold, the function V also satisfies a one-sided Lipschitz condition

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

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 (3.3) is of effective use to approximate the MV-SDE (3.1), 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

$$\mathrm{d}X_t^i = \big(v(X_t^i, \mu_t^{X^i}) + b(t, X_t^i, \mu_t^{X^i})\big)\mathrm{d}t + \sigma(t, X_t^i, \mu_t^{X^i})\mathrm{d}W_t^i, \quad X_0^i = X_0^i\,, \quad t \in [0,T]\,, \tag{3.6}$$

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 (3.1) with v given as (3.2)).

The Propagation of chaos result (3.8) follows from [2, Theorem 3.14] under the assumption that the interacting particle system (3.3) is well-posed. The first statement of Proposition 3.2.5 establishes the well-posedness of the particle system hence closing the small gap left in [2, Theorem 3.14].

Proposition 3.2.5. *Let the assumptions of Theorem 3.2.3 hold for some $m > 2(q+1)$. Then, for all $i \in [\![1,N]\!]$ there exists a unique solution $X^{i,N}$ to (3.3) 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*

$$\sup_{t\in[0,T]} \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|X_t^{i,N}|^p\big] \le C\Big(1 + \mathbb{E}\big[\,|X_0^{\cdot}|^p\big]\Big). \tag{3.7}$$

For $i \in [\![1,N]\!]$, let $X^i \in \mathbb{S}^m([0,T])$ be the solution to (3.6), ensured by Theorem 3.2.3. 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

$$\sup_{i\in[\![1,N]\!]} \sup_{0\le t\le T} \mathbb{E}\big[|X_t^i - X_t^{i,N}|^2\big] \le C \begin{cases} N^{-1/2}, & d<4 \\ N^{-1/2}\log N, & d=4 \\ N^{\frac{-2}{d+4}}, & d>4 \end{cases}. \tag{3.8}$$

The proof and further details are presented in Appendix 6.2.1. 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 Remark 2.2.5.

3.2.2 The scheme for the interacting particle system and main results

The split-step method (SSM) is an extension of Chapter 2 and re-cast accordingly to the setup here. The critical difficulty arises from the convolution component in v (3.1). This term is the main hindrance in proving moment bounds. Before continuing recall the definition of V in Remark 3.2.4. We now introduce the SSM numerical scheme.

Definition 3.2.6 (Definition of the SSM)**.** *Let Assumption 3.2.1 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 (3.3) 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 3.2.4 and the definition of the map V in (3.5))*

$$Y_n^{\star,N} = \hat{X}_n^N + hV(Y_n^{\star,N}),\ \hat{X}_n^N = (\cdots, \hat{X}_n^{i,N}, \cdots), \quad Y_n^{\star,N} = (\cdots, Y_n^{i,\star,N}, \cdots), \tag{3.9}$$

$$\textit{where}\ \ Y_n^{i,\star,N} = \hat{X}_n^{i,N} + hv(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}),\ \hat{\mu}_n^{Y,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{Y_n^{j,\star,N}}(\mathrm{d}x), \tag{3.10}$$

$$\hat{X}_{n+1}^{i,N} = Y_n^{i,\star,N} + b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})h + \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Delta W_n^i,\ \Delta W_n^i = W_{t_{n+1}}^i - W_{t_n}^i. \tag{3.11}$$

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

Remark 3.2.7)

$$h \in \Big(0, \min\big\{1, \frac{1}{\zeta}\big\}\Big) \quad \textit{for } \zeta \textit{ defined as} \quad \zeta = \max\Big\{2(L_f + L_u)\ ,\ 4L_f^+ + 2L_u + 2L_{\tilde{u}} + 1\ ,\ 0\Big\}. \tag{3.12}$$

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 3.4.1 and Remark 2.4.3 for more discussion.

Remark 3.2.7 (The constraint on h in (3.12) is soft)**.** *Our framework allows to change f, u, b in such a way as to have $\zeta = 0$ in* (3.12) *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 $t \in [0, \infty), x \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)$*

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

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 3.2.4), and hence (under (3.12)) a unique solution $Y_n^{\star,N}$ to (3.9) as a function of $\hat{X}_n^N$ exists (details in Lemma 3.4.2). After introducing the discrete scheme, we define its continuous extension and provide the main convergence results.

Definition 3.2.8 (Continuous extension of the SSM)**.** *Under the same choice of h and assumptions in Definition 3.2.6, 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*

$$\mathrm{d}\hat{X}_t^{i,N} = \big(v(Y_{\kappa(t)}^{i,\star,N}, \hat{\mu}_{\kappa(t)}^{Y,N}) + b(\kappa(t), Y_{\kappa(t)}^{i,\star,N}, \hat{\mu}_{\kappa(t)}^{Y,N})\big)\mathrm{d}t + \sigma(\kappa(t), Y_{\kappa(t)}^{i,\star,N}, \hat{\mu}_{\kappa(t)}^{Y,N})\mathrm{d}W_t^i, \tag{3.13}$$

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

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 3.2.9 (Non-path-space mean-square convergence)**.** *Let Assumption 3.2.1 hold and choose h as in* (3.12)*. Let $i \in [\![1, N]\!]$, take $X^{i,N}$ as the solution to* (3.3) *and let $\hat{X}^{i,N}$ be the continuous-time extension of the SSM given by* (3.13)*. 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 3.2.1, then there exists a constant $C > 0$ independent of h, N, M (but depending on T and m) such that*

$$\sup_{i \in [\![1,N]\!]} \sup_{0 \le t \le T} \mathbb{E}\big[\,|X_t^{i,N} - \hat{X}_t^{i,N}|^2\big] \le Ch. \tag{3.14}$$

The proof is presented in Section 3.4.2. 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 (3.3) and L^2-moments for the scheme [22]. The proof takes advantage of the elegant structure induced by the SSM where Proposition 3.4.3 and 3.4.4 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 3.2.10 (Moment bounds)**.** *Let the setting of Theorem 3.2.9 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* (3.13)*. 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*

$$\sup_{i\in[\![1,N]\!]}\sup_{0\le t\le T}\mathbb{E}\big[\,|\hat{X}_t^{i,N}|^{2p}\big]\le C\big(1+\mathbb{E}\big[\,|\hat{X}_0^{\cdot}|^{2p}\big]\big)<\infty. \tag{3.15}$$

The proof is presented in Section 3.4.3 and builds around auxiliary Theorem 3.4.7. There, we expand (3.59) and (3.60), and leverage the properties of the SSM scheme stated in Proposition 3.4.3 and 3.4.4 to deal with the difficult convolution terms.

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

Theorem 3.2.11 (Classical path-space mean-square convergence)**.** *Let the setting of Theorem 3.2.9 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 3.2.1. Then there exists a constant $C>0$ independent of h,N,M (but depending on T and m) such that*

$$\sup_{i\in[\![1,N]\!]}\mathbb{E}\big[\sup_{0\le t\le T}|X_t^{i,N}-\hat{X}_t^{i,N}|^2\big]\le Ch^{1-\varepsilon}. \tag{3.16}$$

The proof is presented in Section 3.4.4. 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 3.2.9 and the reason we introduce Theorem 3.2.10 as a main result. The nearly optimal error rate of $(1-\varepsilon)$ is a consequence of the estimation of (3.70) (product of three unbounded random variables). The expectation is taken after the supremum and then we use Theorem 3.2.9 and 3.2.10 – 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 (3.15) with the $\sup_t$ inside the expectation — the difficulty to be overcome relates to establishing (3.25) of Proposition 3.4.4 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 (3.16) when $\varepsilon=0$.

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 3.2.12. *Consider the following MV-SDE*

$$\mathrm{d}X_t=v(X_t,\mu_t^X)\mathrm{d}t+\sigma\mathrm{d}W_t,\quad X_0\in L_0^m(\mathbb{R}^d),\quad v(x,\mu)=\int_{\mathbb{R}^d}f(x-y)\mu(\mathrm{d}y). \tag{3.17}$$

Let $f:\mathbb{R}^d\to\mathbb{R}^d$ be continuously differentiable satisfying $(\mathbf{A}^f)$ of Assumption 3.2.1. 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 $x,\ x'\in\mathbb{R}^d$

$$|\nabla f(x)|\le L_{f'}(1+|x|^q),\quad |\nabla f(x)-\nabla f(x')|\le L_{f''}(1+|x|^{q-1}+|x'|^{q-1})|x-x'|. \tag{3.18}$$

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 (3.18) 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 [110]. We have the following results.

Theorem 3.2.13. *Let Assumption 3.2.12 hold and choose h as in* (3.12)*. Let $i\in[\![1,N]\!]$, take $X^{i,N}$ to be the solution to* (3.3)*, let $\hat{X}^{i,N}$ be the continuous-time extension of the SSM given by* (3.13) *and $X_0^i\in L_0^m(\mathbb{R}^d)$. Let $m\ge\max\{8q,\ 4q+4\}>\max\{2(q+1),4\}$ with q as defined in Assumption 3.2.12. Then there exist a constant $C>0$ independent of h,N,M (but depending on T and m) such that*

$$\sup_{i\in[\![1,N]\!]}\sup_{0\le t\le T}\mathbb{E}\big[|X_t^{i,N}-\hat{X}_t^{i,N}|^2\big]\le Ch^2. \tag{3.19}$$

This result is proved in Section 3.4.5. Supporting simulation results are presented in Section 3.3.1 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 3.2.11 to obtain (3.19) with the $\sup_t$ inside the expectation. This would deliver a rate of $h^{2-\varepsilon}$, the key steps are similar to (3.71)-(3.72).

3.3 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

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

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

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

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,

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

Remark 3.3.1 ('Taming' algorithm)**.** *For comparative purposes we implement the 'Taming' algorithm [47, 60] – 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 (3.9) of the SSM algorithm is solved, in our examples, via a Newton method iteration. We point the reader to Appendix 6.2.2 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 $\beta \in (0, \infty)$.

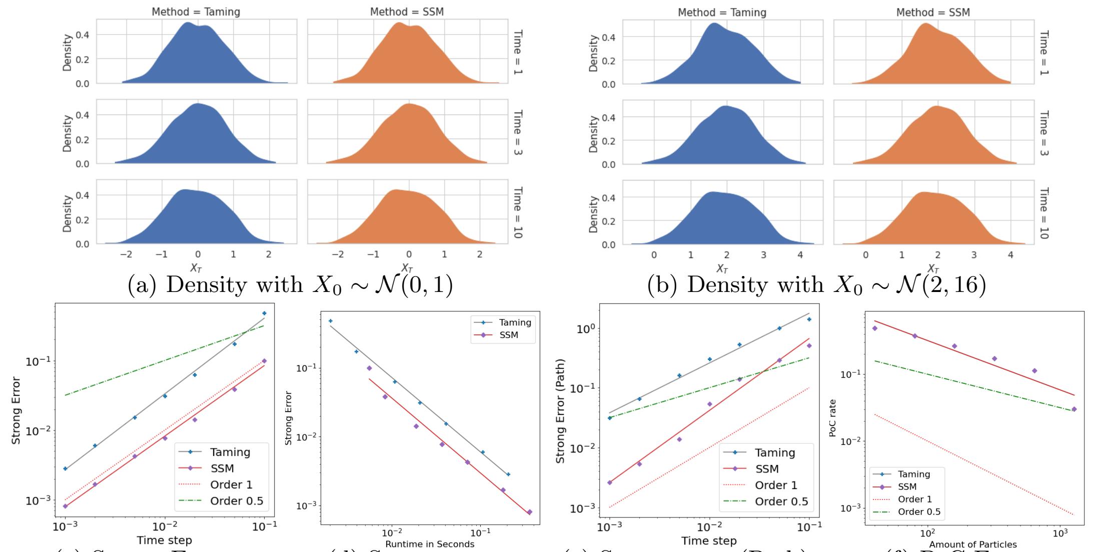
[3]Implementation code in Python is available in https://github.com/AnandaChen/Simulation-of-super-measure

3.3.1 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 [40, 110]. In MV-SDE form we have

$$\mathrm{d}X_t = v(X_t, \mu_t^X)\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \quad X_0 \in L_0^m(\mathbb{R}^d), \quad v(x,\mu) = \int_{\mathbb{R}^d} \Big(-\operatorname{sign}(x-y)|x-y|^2 \Big)\mu(\mathrm{d}y), \tag{3.20}$$

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 [40, 110] and is a reference model to showcase



(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 (e) Strong error (Path) (f) PoC Error

Figure 3.1: Simulation of the granular media equation (3.20) 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 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 $h \in \{10^{-3}, 2 \times 10^{-3}, \ldots, 10^{-1}\}$. 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 3.2.13 (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.

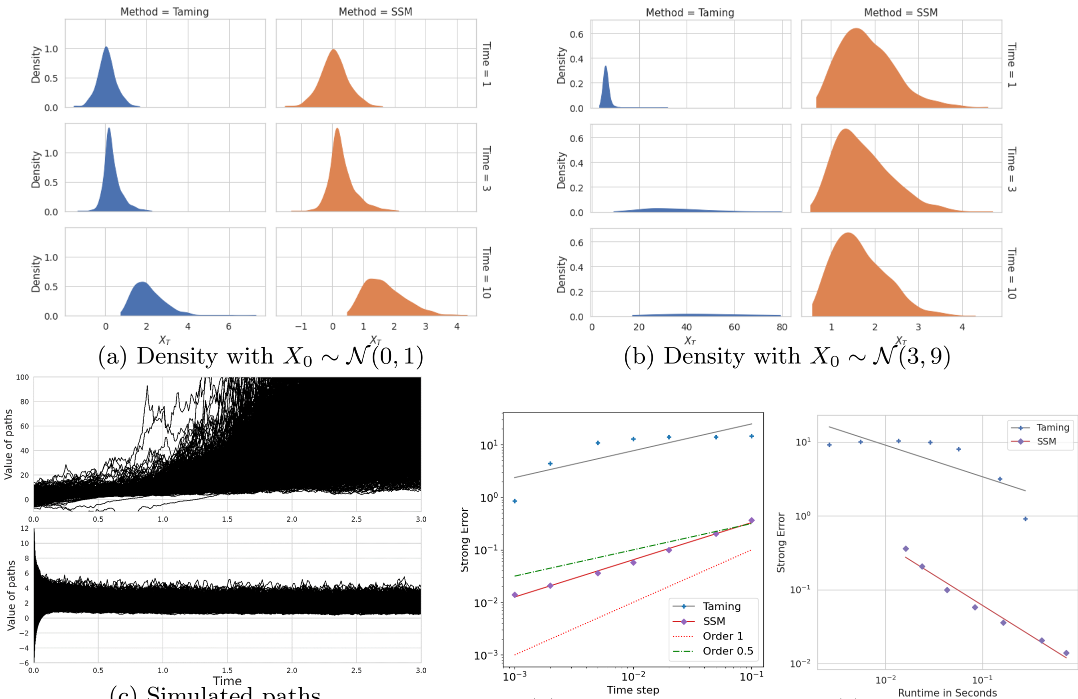
Figure 3.1 (e) shows the path type strong error of both method, compare to the results in (c), the SSM preserve the error rate of near 1.0 and perform better than the Taming method. Figure 3.1 (f) shows the PoC error of both methods. The two results coincide since the differences between two methods 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 3.2.5. This result is similar to [125, Example 4.1], and is explained theoretically by [57, Lemma 5.1] but under stronger assumptions than ours.

3.3.2 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 [138] but change its diffusion coefficient to a non-constant one (in opposition to the previous example). Concretely, we study the following McKean-Vlasov equation

$$\mathrm{d}X_t = \big(v(X_t, \mu_t^X) + X_t\big)\mathrm{d}t + X_t \mathrm{d}W_t, \quad v(x, \mu) = -\frac{1}{4}x^3 + \int_{\mathbb{R}^d} -\big(x - y\big)^3 \mu(\mathrm{d}y). \qquad (3.21)$$

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 [138].



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

(c) Simulated paths (d) Strong error (e) Strong error

Figure 3.2: Simulation of the Double-Well model (3.21) 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 example of Section 3.3.1 was a relatively mild with additive noise and where both methods performed well. For this double-well model of (3.21), 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 3.3.1 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 3.2.9 (and Theorem 3.2.11).

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.

3.3.3 Example: 2d Van der Pol (VdP) oscillator

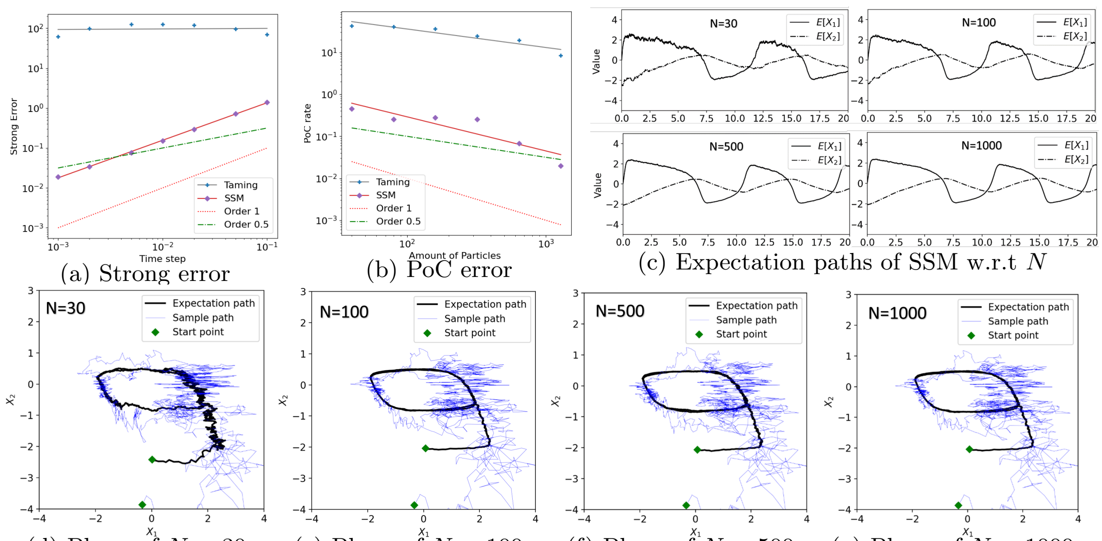
We consider the Van der Pol (VdP) model described in [88, Section 4.2 and 4.3], with added super-linearity in measure and non-constant unbounded diffusion. We study the following MV-SDE dynamics: set $x = (x_1, x_2) \in \mathbb{R}^2$, for (3.1) define the functions f, u, b, σ as

$$f(x) = -x|x|^2, \quad u(x) = \begin{bmatrix} -\frac{4}{3}x_1^3 \\ 0 \end{bmatrix}, \quad b(x) = \begin{bmatrix} 4(x_1 - x_2) \\ \frac{1}{4}x_1 \end{bmatrix}, \quad \sigma(x) = \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix}, \quad (3.22)$$

which satisfy the assumptions of this chapter.

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 3.3.1).

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 $N \geq 500$.



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

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

Figure 3.3: Simulation of the Vdp model (3.22) 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 $N \in \{30, 100, 500, 1000\}$.

3.3.4 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.

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 3.3.3, 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 (3.3), 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 3.2.6, 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 [18] or the Random Batch Method (RBM) [92]. 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 3.2.5, 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 (3.8) in Theorem 3.2.5. This result is not surprising in view of the theoretical result [57, Lemma 5.1]; and numerically in [125, Example 4.1]. To the best of our knowledge, no known PoC rate result covers the examples presented here and Theorem 3.2.5 is presently the best known general result.

Solving the implicit step in SSM - Newton's method

The SSM scheme contains an implicit Equation (3.9) 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 (3.9). Proposition 6.2.2 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 3.2.9 is preserved.

As mentioned in the initial part of Section 3.3, we use Newton's method (assuming extra differentiability of the involved maps) – see Appendix 6.2.2 for details where [131, Section 4.3] is used to guarantee convergence. The computation cost raises from $\mathcal{O}(N^2M)$ to $\mathcal{O}(\kappa N^2M)$, 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 6.2.2).

Interacting particle systems like (3.3) 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 6.2.2 point to a very sparse Jacobian matrix with a very specific block structure. For instance, the Γ matrix (see Appendix 6.2.2) 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 [105, Section 3] the authors review [130] who address the case of using inexact Newton methods when the equation of interest (3.9) is a monotone map, which is indeed our case. The usage of Newton method is not a primary element of discussion and, as does [105], we point the reader to the comprehensive review [114] 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 3.3.1.

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

The proof appearing in Section 3.4.2 depends in no way on Theorem 3.2.10 or its proof (in Section 3.4.3). Nonetheless, Section 3.4.3 has a strong complementary effect to fully understanding the proof in Section 3.4.2.

3.4.1 Some properties of the scheme

Recall the SSM scheme of Definition 3.2.6. 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 3.4.1 (Choice of h)**.** *Let Assumption 3.2.1 hold, the constraint on h in* (3.12) *comes from* (3.24), (3.25) *and* (3.41) *below, where $L_f, L_u \in \mathbb{R}$ and $L_{\tilde{u}} \geq 0$. Following the notation of those inequalities, under* (3.12) *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

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

As argued in Remark 3.2.7 the constraint on h can be lifted.

Lemma 3.4.2. *Choose h as in* (3.12)*. Then, given any $X \in \mathbb{R}^{Nd}$ there exists a unique solution $Y \in \mathbb{R}^{Nd}$ to*

$$Y = X + hV(Y). \tag{3.23}$$

The solution Y is a measurable map of X.

Proof. Recall Remark 3.2.4. The proof is an adaptation of the proof Lemma 2.4.1 to the $\mathbb{R}^{Nd}$ case.

□

Proposition 3.4.3 (Differences relationship)**.** *Let Assumption 3.2.1 hold and choose h as in* (3.12)*. For any $n \in [\![0, M]\!]$ and $Y_n^{*,N}$ in* (3.9)*, there exists some constant $C > 0$ such that for all $i,\ j \in [\![1, N]\!]$,*

$$|Y_n^{i,\star,N} - Y_n^{j,\star,N}|^2 \leq |\hat{X}_n^{i,N} - \hat{X}_n^{j,N}|^2 \frac{1}{1-2(L_f+L_u)h} \leq (1+Ch)|\hat{X}_n^{i,N} - \hat{X}_n^{j,N}|^2. \tag{3.24}$$

Proof. Take $n \in [\![0, M]\!]$, $i,\ j \in [\![1, N]\!]$. Using Remark 3.2.2 and Young's inequality we have

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

The argument regarding the uniformity of the constant C in regards to the parameters h, N, M follows from Remark 3.4.1. □

Proposition 3.4.4 (Summation relationship)**.** *Let Assumption 3.2.1 hold. Choose h as in* (3.12)*. For the process in* (3.10) *there exists a constant $C > 0$ (independent of h, N, M) such that, for all $i \in [\![1, N]\!],\ n \in [\![0, M]\!]$,*

$$
\frac{1}{N}\sum_{i=1}^{N} |Y_n^{i,\star,N}|^2 \leq Ch + (1 + Ch)\frac{1}{N}\sum_{i=1}^{N} |\hat{X}_n^{i,N}|^2. \tag{3.25}
$$

Proof. From (3.11) we have

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} |Y_n^{i,\star,N}|^2 &= \frac{1}{N}\sum_{i=1}^{N} \Big(\Big\langle Y_n^{i,\star,N}, \hat{X}_n^{i,N} \Big\rangle + \Big\langle Y_n^{i,\star,N}, v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}) \Big\rangle h \Big) \\
&\leq \frac{1}{N}\sum_{i=1}^{N} \Big(\frac{1}{2}|Y_n^{i,\star,N}|^2 + \frac{1}{2}|\hat{X}_n^{i,N}|^2 + \Big\langle Y_n^{i,\star,N}, u(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}) \Big\rangle h \\
&\qquad + \frac{h}{N}\sum_{j=1}^{N} \Big\langle Y_n^{i,\star,N}, f(Y_n^{i,\star,N} - Y_n^{j,\star,N}) \Big\rangle \Big).
\end{aligned} \tag{3.26}
$$

By Assumption 3.2.1 and Young's inequality, we have

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

Plugging this into (3.26) and using Remark 3.2.2 with $\Lambda = 4L_f^+ + 2L_u + 2L_{\tilde{u}} + 1$, we have

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

Remark 3.4.1 yields the argument. □

From Lemma 3.4.2 we know a unique solution, $Y_n^{\star,N}$, to (3.9) as a function of $\hat{X}_n^N$ exists.

We next show that the scheme we proposed in (3.9)-(3.11) is square integrable.

Proposition 3.4.5 (Second moment bounds of SSM)**.** *Let the setting of Theorem 3.2.9 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 3.2.1, from (3.9)-(3.11) and Proposition 3.4.4, 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 Proposition 2.4.5, we have

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

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

$$\sup_{i\in[\![1,N]\!]} \sup_{n\in[\![0,M]\!]} \mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big] \leq \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[1+|\hat{X}_0^{i,N}|^2\big](1+Ch)^M \leq (1+\mathbb{E}\big[|\hat{X}_0^{\cdot,N}|^2\big])e^{CT} < \infty, \tag{3.27}$$

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 3.4.6. *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. □

3.4.2 Proof of Theorem 3.2.9: the pointwise mean-square convergence result

We provide here the proof of Theorem 3.2.9. Throughout this section, we follow the notation introduced in Theorem 3.2.9 and let Assumption 3.2.1 hold, h is chosen as in (3.12), $m \geq 4q+4$, where m is defined in (3.1) and q is defined in Assumption 3.2.1. Note that throughout $C > 0$ is a constant always independent of h, N, M but possibly depending on T and m.

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 (3.3), 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 (3.13), 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

$$\mathbb{E}\big[|\Delta X_{t_n+r}^i|^2\big] \leq (1+h)I_1 + (1+\frac{1}{h})I_2 + 2I_3 + 2I_4, \tag{3.28}$$

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

$$I_1 = \mathbb{E}\Big[\Big|\Delta X_{t_n}^i + \int_{t_n}^{t_n+r} \big(v(Y_n^{i,X,N}, \mu_n^{Y,X,N}) - v\left(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}\right)\big)\mathrm{d}s \tag{3.29}$$

$$+ \int_{t_n}^{t_n+r} \big(b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big)\mathrm{d}s\Big|^2\Big], \tag{3.30}$$

$$I_2 = \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r} \big(v(X_s^{i,N}, \mu_s^{X,N}) - v(Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}s \tag{3.31}$$

$$+ \int_{t_n}^{t_n+r} \big(b(s, X_s^{i,N}, \mu_s^{X,N}) - b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}s\Big|^2\Big], \tag{3.32}$$

$$I_3 = \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r} \big(\sigma(s, X_s^{i,N}, \mu_s^{X,N}) - \sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)\mathrm{d}W_s^i\Big|^2\Big], \tag{3.33}$$

$$I_4 = \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r} \big(\sigma(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big)\mathrm{d}W_s^i\Big|^2\Big]. \tag{3.34}$$

For I_1, Young's inequality yields

$$I_1 = \mathbb{E}\Big[\Big|X_n^{i,N} + \big(V_n^{Y,i} + b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N})\big)r - \hat{X}_n^{i,N} - \big(V_n^{*,i} + b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big)r\Big|^2\Big]$$
$$\leq \mathbb{E}\Big[\Big|X_n^{i,N} - \hat{X}_n^{i,N} + \big(V_n^{Y,i} - V_n^{*,i}\big)r\Big|^2\Big](1+\frac{h}{2})$$
$$+ \mathbb{E}\Big[\Big|b(t_n, Y_n^{i,X,N}, \mu_n^{Y,X,N}) - b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Big|^2\Big](\frac{h}{2}+h), \tag{3.35}$$

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 (3.35), recall the SSM defined in (3.10). We have

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

where we write $r(r-h) = -r(h-r)$ to more clearly illustrate the results in the later content. Using the relationship that (3.10) 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

$$\mathbb{E}\Big[\Big|X_n^{i,N} - \hat{X}_n^{i,N} + \big(V_n^{Y,i} - V_n^{*,i}\big)r\Big|^2\Big] = \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big] + \mathbb{E}\Big[\Big\langle X_n^{i,N} - \hat{X}_n^{i,N}, V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]2r$$
$$+ \mathbb{E}\Big[\Big\langle (Y_n^{i,X,N} - Y_n^{i,\star,N}) - (X_n^{i,N} - \hat{X}_n^{i,N}), V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]\frac{r^2}{h}$$
$$\leq \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big](1 - C_{h,r}) + \mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big]C_{h,r}$$
$$+ \mathbb{E}\Big[\Big\langle Y_n^{i,X,N} - Y_n^{i,\star,N}, V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]\frac{r^2}{h}, \tag{3.36}$$

where $C_{h,r} = (2hr - r^2)/2h^2$. Also, for the second term of (3.35), using Assumption 3.2.1 and that the particles are identically distributed

$$\mathbb{E}\Big[\Big|b(t_n, Y_n^{i,X,N},\ \mu_n^{Y,X,N}) - b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Big|^2\Big]$$
$$\leq C\mathbb{E}\Big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2 + W^{(2)}(\mu_n^{Y,X,N}, \hat{\mu}_n^{Y,N})\Big]$$
$$\leq C\mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big] + C\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^{N}|Y_n^{j,X,N} - Y_n^{j,\star,N}|^2\Big] \leq C\mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big]. \tag{3.37}$$

By Assumption 3.2.1 and using Young's inequality once again

$$\mathbb{E}\big[|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2\big] \leq \mathbb{E}\Big[\Big\langle Y_n^{i,X,N} - Y_n^{i,\star,N}, X_n^{i,N} - \hat{X}_n^{i,N} + V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]h \tag{3.38}$$
$$=\mathbb{E}\Big[\frac{1}{2}|Y_n^{i,X,N} - Y_n^{i,\star,N}|^2 + \frac{1}{2}|X_n^{i,N} - \hat{X}_n^{i,N}|^2\Big] + \mathbb{E}\Big[\Big\langle Y_n^{i,X,N} - Y_n^{i,\star,N}, V_n^{Y,i} - V_n^{*,i}\Big\rangle\Big]h. \tag{3.39}$$

For the last term (3.39), since the particles are identically distributed, Assumption 3.2.1 and

Remark 3.2.4 yield

$$\mathbb{E}\Big[\Big\langle Y_n^{i,X,N}-Y_n^{i,\star,N},V_n^{Y,i}-V_n^{*,i}\Big\rangle\Big]\le\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N\Big\langle Y_n^{j,X,N}-Y_n^{j,\star,N},V_n^{Y,j}-V_n^{*,j}\Big\rangle\Big]$$
$$\le\Big(2L_f^++L_u+\frac{1}{2}+\frac{L_{\tilde{u}}}{2}\Big)\mathbb{E}\big[|Y_n^{i,X,N}-Y_n^{i,\star,N}|^2\big]. \quad (3.40)$$

Thus, injecting (3.40) back into (3.39) and (3.38), set $\Gamma_2=4L_f^++2L_u+L_{\tilde{u}}+1$, then by Remark 3.4.1,

$$\mathbb{E}\big[|Y_n^{i,X,N}-Y_n^{i,\star,N}|^2\big]\le\frac{1}{1-\Gamma_2h}\mathbb{E}\big[|X_n^{i,N}-\hat{X}_n^{i,N}|^2\big]\le\mathbb{E}\big[|X_n^{i,N}-\hat{X}_n^{i,N}|^2\big](1+Ch). \quad (3.41)$$

Plug (3.41) and (3.40) back into (3.36), (3.37) and (3.35). We then conclude that

$$I_1\le\mathbb{E}\big[|X_n^{i,N}-\hat{X}_n^{i,N}|^2\big](1+Ch). \quad (3.42)$$

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

$$I_2\le h\ \mathbb{E}\Big[\int_{t_n}^{t_n+h}\Big|v(X_s^{i,N},\mu_s^{X,N})-v(Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2\mathrm{d}s \quad (3.43)$$
$$+\int_{t_n}^{t_n+h}\Big|b(s,X_s^{i,N},\mu_s^{X,N})-b(t_n,Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2\mathrm{d}s\Big]. \quad (3.44)$$

For (3.43), from Assumption 3.2.1, using Young's, Jensen's, and Cauchy-Schwarz inequality

$$\mathbb{E}\Big[\Big|v(X_s^{i,N},\mu_s^{X,N})-v(Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2\Big]$$
$$\le C\mathbb{E}\Big[\Big|u(X_s^{i,N},\mu_s^{X,N})-u(Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2$$
$$+\frac{1}{N}\sum_{j=1}^N\Big|f(X_s^{i,N}-X_s^{j,N})-f(Y_n^{i,X,N}-Y_n^{j,X,N})\Big|^2\Big] \quad (3.45)$$
$$\le\frac{C}{N}\sum_{j=1}^N\mathbb{E}\Big[\Big|\Big(1+|X_s^{i,N}-X_s^{j,N}|^q+|Y_n^{i,X,N}-Y_n^{j,X,N}|^q\Big)$$
$$\cdot|X_s^{i,N}-Y_n^{i,X,N}-(X_s^{j,N}-Y_n^{j,X,N})|\Big|^2\Big]$$
$$+C\mathbb{E}\Big[(1+|X_s^{i,N}|^{2q}+|Y_n^{i,X,N}|^{2q})(|X_s^{i,N}-Y_n^{i,X,N}|^2)+\frac{1}{N}\sum_{j=1}^N|X_s^{j,N}-Y_n^{j,X,N}|^2\Big]$$
$$\le C\sqrt{\mathbb{E}\Big[1+|X_s^{i,N}|^{4q}+|Y_n^{i,X,N}|^{4q}\Big]\mathbb{E}\Big[|X_s^{i,N}-Y_n^{i,X,N}|^4\Big]}+\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|X_s^{j,N}-Y_n^{j,X,N}|^2\Big] \quad (3.46)$$
$$+\frac{C}{N}\sum_{j=1}^N\sqrt{\mathbb{E}\Big[1+|X_s^{i,N}-X_s^{j,N}|^{4q}+|Y_n^{i,X,N}-Y_n^{j,X,N}|^{4q}\Big]}$$
$$\times\sqrt{\mathbb{E}\Big[|X_s^{i,N}-Y_n^{i,X,N}|^4+|X_s^{j,N}-Y_n^{j,X,N}|^4\Big]}. \quad (3.47)$$

Using the structure of the SSM, Young's and Jensen's inequality, and Proposition 3.4.3 we have

$$
\begin{aligned}
|X_s^{i,N} - Y_n^{i,X,N}|^2 &\leq 2|X_s^{i,N} - X_n^{i,N}|^2 + 2|X_n^{i,N} - Y_n^{i,X,N}|^2, &(3.48)\\
|X_n^{i,N} - Y_n^{i,X,N}|^2 &= \Big|v(Y_n^{i,X,N}, \mu_n^{Y,X,N})h\Big|^2\\
&\leq 2\Big|u(Y_n^{i,X,N}, \mu_n^{Y,X,N})h\Big|^2 + \frac{2h^2}{N}\sum_{j=1}^N \Big|f(Y_n^{i,X,N} - Y_n^{j,X,N})\Big|^2 &(3.49)\\
&\leq C\Big(1 + |Y_n^{i,X,N}|^{2q+2} + \frac{1}{N}\sum_{j=1}^N |Y_n^{j,X,N}|^2\Big)h^2 + \frac{Ch^2}{N}\sum_{j=1}^N\Big(1 + |Y_n^{i,X,N} - Y_n^{j,X,N}|^{2q+2}\Big)\\
&\leq C\Big(1 + |Y_n^{i,X,N}|^{2q+2} + \frac{1}{N}\sum_{j=1}^N |Y_n^{j,X,N}|^2\Big)h^2 + \frac{Ch^2}{N}\sum_{j=1}^N\Big(1 + |X_n^{i,N} - X_n^{j,N}|^{2q+2}\Big).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&|X_s^{i,N} - Y_n^{i,X,N}|^4 \leq 16|X_s^{i,N} - X_n^{i,N}|^4 + 16|X_n^{i,N} - Y_n^{i,X,N}|^4, &(3.50)\\
&|X_n^{i,N} - Y_n^{i,X,N}|^4\\
&\qquad\leq C\Big(1 + |Y_n^{i,X,N}|^{4q+4} + \frac{1}{N}\sum_{j=1}^N |Y_n^{j,X,N}|^4\Big)h^4 + \frac{Ch^4}{N}\sum_{j=1}^N\Big(1 + |X_n^{i,N} - X_n^{j,N}|^{4q+4}\Big).
\end{aligned}
$$

From (3.3) and using (3.7) (since $m \geq 4q+4$) alongside Young's inequality and Itô's isometry, we have

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

Also, using (3.7), Jensen's and Young's inequality (since $m \geq 4q+4$) we have

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

This next argument uses steps similar to those used in (3.59) and (3.60) (appearing in the proof of Theorem 3.4.7). Since $X^{\cdot,N}$ has bounded moments via (3.7) (this refers to the true interacting particle system), we have for any $m \geq p \geq 2$ that

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

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

$$
\mathbb{E}\big[|X_s^{i,N} - Y_n^{i,X,N}|^2\big] \leq Ch, \quad \mathbb{E}\big[|X_s^{i,N} - Y_n^{i,X,N}|^4\big] \leq Ch^2, \quad \mathbb{E}\big[|Y_n^{i,X,N}|^p\big] \leq C, \quad (3.51)
$$

$$
\mathbb{E}\Big[\Big|W^{(2)}(\mu_s^{X,N}, \mu_n^{Y,X,N})\Big|^2\Big] \leq \mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N |X_s^{j,N} - Y_n^{j,X,N}|^2\Big] \leq Ch. \quad (3.52)
$$

Plugging all the above inequalities back into (3.46) and (3.47), we conclude that

$$\mathbb{E}\Big[\Big|v(X_s^{i,N},\mu_s^{X,N})-v(Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2\Big]\le Ch. \tag{3.53}$$

We now consider term (3.44) of I_2. By Assumption 3.2.1, using (3.51) and (3.52)

$$\begin{aligned}&\mathbb{E}\Big[\Big|b(s,X_s^{i,N},\mu_s^{X,N})-b(t_n,Y_n^{i,X,N},\mu_n^{Y,X,N})\Big|^2\Big]\\&\quad\le C\mathbb{E}\Big[h+|X_s^{i,N}-Y_n^{i,X,N}|^2+\Big|W^{(2)}(\mu_s^{X,N},\mu_n^{Y,X,N})\Big|^2\Big]\le Ch.\end{aligned} \tag{3.54}$$

Thus, plugging (3.53), (3.54) back into (3.43) and (3.44), we have

$$I_2\le Ch^3. \tag{3.55}$$

For I_3, by Itô's isometry, the results in (3.51) and (3.52), and using similar argument as in (3.54) we have

$$\begin{aligned}I_3=&\mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r}\Big(\sigma(s,X_s^{i,N},\mu_s^{X,N})-\sigma(t_n,Y_n^{i,X,N},\mu_n^{Y,X,N})\Big)\mathrm{d}W_s^i\Big|^2\Big]\\\le&\mathbb{E}\Big[\int_{t_n}^{t_n+h}\Big|\Big(\sigma(s,X_s^{i,N},\mu_s^{X,N})-\sigma(t_n,Y_n^{i,X,N},\mu_n^{Y,X,N})\Big)\Big|^2\,\mathrm{d}s\Big]\le Ch^2.\end{aligned} \tag{3.56}$$

Similarly for I_4, by Itô's isometry, Proposition 3.4.5, Equation (3.41) and using similar argument in (3.37)

$$\begin{aligned}I_4=&\mathbb{E}\Big[\Big|\int_{t_n}^{t_n+r}\Big(\sigma(t_n,Y_n^{i,X,N},\mu_n^{Y,X,N})-\sigma(t_n,Y_n^{i,\star,N},\hat\mu_n^{Y,N})\Big)\mathrm{d}W_s^i\Big|^2\Big]\\\le&\mathbb{E}\Big[\int_{t_n}^{t_n+h}\Big|\Big(\sigma(t_n,Y_n^{i,X,N},\mu_n^{Y,X,N})-\sigma(t_n,Y_n^{i,\star,N},\hat\mu_n^{Y,N})\Big)\Big|^2\,\mathrm{d}s\Big] \qquad (3.57)\\\le&\int_{t_n}^{t_n+h}\mathbb{E}\big[|Y_n^{i,X,N}-Y_n^{i,\star,N}|^2\big]\mathrm{d}s\le(1+Ch)\int_{t_n}^{t_n+h}\mathbb{E}\big[|X_n^{i,N}-\hat X_n^{i,N}|^2\big]\mathrm{d}s\\\le&\mathbb{E}\big[|X_n^{i,N}-\hat X_n^{i,N}|^2\big]Ch.\end{aligned} \tag{3.58}$$

Plugging (3.42), (3.55) (3.56) and (3.58) back to (3.28), we have, for all $n\in[\![0,M-1]\!]$, $i\in[\![1,N]\!]$ and $r\in[0,h]$

$$\begin{aligned}&\mathbb{E}\big[|\Delta X_{t_n+r}^i|^2\big]\\&\le(1+h)\mathbb{E}\big[|X_n^{i,N}-\hat X_n^{i,N}|^2\big](1+Ch)+(1+\frac{1}{h})Ch^3+Ch^2+\mathbb{E}\big[|X_n^{i,N}-\hat X_n^{i,N}|^2\big]Ch\\&\le\mathbb{E}\big[|X_n^{i,N}-\hat X_n^{i,N}|^2\big](1+Ch)+Ch^2.\end{aligned}$$

By backward induction, the discrete Grönwall's lemma delivers the result of (3.14). □

3.4.3 Proof of Theorem 3.2.10: the moment bound result

In this section prove Theorem 3.2.10, and we we follow the notation and assumptions introduced in Theorem 3.2.10 throughout this section.

We first prove a moment bounds result across the timegrid then extend it to the continues process as stated in Theorem 3.2.10.

Theorem 3.4.7 (Moment bounds of SSM)**.** *Let the setting of Theorem 3.2.9 hold. Let $m\ge2$ where $\hat X_0^{i,N}\in L_0^m(\mathbb{R}^d)$ for all $i\in[\![1,N]\!]$ and let $\hat X^{i,N}$ be the continuous-time extension of the SSM given by* (3.13)*. 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

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

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

$$\begin{aligned}\mathbb{E}\big[|\hat{X}_n^{i,N}|^{2p}\big]&=\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N(\hat{X}_n^{i,N}-\hat{X}_n^{j,N})+\frac{1}{N}\sum_{j=1}^N\hat{X}_n^{j,N}\Big|^{2p}\Big]\\&\le 4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|\hat{X}_n^{i,N}-\hat{X}_n^{j,N}|^{2p}\Big]+4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N(1+|\hat{X}_n^{j,N}|^2)\Big|^p\Big]+1=H_n^{X,p},\end{aligned}\tag{3.59}$$

$$\mathbb{E}\big[|Y_n^{i,\star,N}|^{2p}\big]\le 4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|Y_n^{i,\star,N}-Y_n^{j,\star,N}|^{2p}\Big]+4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N(1+|Y_n^{j,\star,N}|^2)\Big|^p\Big]+1=H_n^{Y,p}.\tag{3.60}$$

Using the following results from Proposition 3.4.3 and 3.4.4, we have $H_n^{Y,p}\le H_n^{X,p}(1+Ch)$,

$$\begin{aligned}|Y_n^{i,\star,N}-Y_n^{j,\star,N}|^2&\le|\hat{X}_n^{i,N}-\hat{X}_n^{j,N}|^2(1+Ch)\\\frac{1}{N}\sum_{j=1}^N(1+|Y_n^{j,\star,N}|^2)&\le\Big[\frac{1}{N}\sum_{j=1}^N(1+|\hat{X}_n^{j,N}|^2)\Big](1+Ch).\end{aligned}$$

We now prove that $H_{n+1}^{X,p}\le H_n^{Y,p}(1+Ch)$. For the first element composing $H_{n+1}^{X,p}$ we have

$$\begin{aligned}\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|\hat{X}_{n+1}^{i,N}-\hat{X}_{n+1}^{j,N}|^{2p}\Big]&=\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N\Big|\Big(Y_n^{i,\star,N}+b(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})h+\sigma(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})\\&\quad\cdot\Delta W_n^i\Big)-\Big(Y_n^{j,\star,N}+b(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})h+\sigma(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^j\Big)\Big|^{2p}\Big].\end{aligned}\tag{3.61}$$

Introduce the extra (local) notation for $G_1^{i,j,n}$, $G_2^{i,j,n}$ and $G_3^{i,j,n}$ as

$$\begin{aligned}G_1^{i,j,n}&=Y_n^{i,\star,N}-Y_n^{j,\star,N},\quad G_2^{i,j,n}=b(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})h-b(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})h,\\G_3^{i,j,n}&=\sigma(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^i-\sigma(t_n,Y_n^{j,\star,N},\hat{\mu}_n^{Y,N})\Delta W_n^j.\end{aligned}$$

For $a+b+c=2p,\ a<2p,\ a,b,c\in\mathbb{N}$, by Assumption 3.2.1, Young's inequality, Jensen's inequality, Proposition 3.4.6 and the fact that the Brownian increments are independent, the particles are conditionally independent and identically distributed, for (3.61), we have

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

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

$$4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|\hat{X}_{n+1}^{i,N}-\hat{X}_{n+1}^{j,N}|^{2p}\Big]\le 4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N|Y_n^{i,\star,N}-Y_n^{j,\star,N}|^{2p}\Big]+H_n^{Y,p}Ch.\tag{3.62}$$

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

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

Set the following (extra local) notation

$$
\begin{aligned}
G^n_4&=\frac{1}{N}\sum_{j=1}^{N}(1+|Y^{j,\star,N}_n|^2),\\
G^n_5&=\frac{1}{N}\sum_{j=1}^{N}\Big\langle 2Y^{j,\star,N}_n+\sigma(t_n,Y^{j,\star,N}_n,\hat{\mu}^{Y,N}_n)\Delta W^j_n,\sigma(t_n,Y^{j,\star,N}_n,\hat{\mu}^{Y,N}_n)\Delta W^j_n\Big\rangle,\\
G^n_6&=\frac{1}{N}\sum_{j=1}^{N}\Big\langle 2Y^{j,\star,N}_n+b(t_n,Y^{j,\star,N}_n,\hat{\mu}^{Y,N}_n)h+2\sigma(t_n,Y^{j,\star,N}_n,\hat{\mu}^{Y,N}_n)\Delta W^j_n,b(t_n,Y^{j,\star,N}_n,\hat{\mu}^{Y,N}_n)h\Big\rangle.
\end{aligned}
$$

We have once again using similar arguments as before, by Young's inequality, Jensen's inequality, Proposition 3.4.6, 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 $l_1>l_2>1,\ |x|^{l_2}\le 1+|x|^{l_1}$ we have

$$
\mathbb{E}\big[|G^n_4|^a|G^n_5|^b|G^n_6|^c\big]\le\mathbb{E}\big[|Y^{i,\star,N}_n|^{2p}+1\big]Ch\le H^{Y,p}_nCh.
$$

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

$$
4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}(1+|\hat{X}^{j,N}_{n+1}|^2)\Big|^p\Big]\le 4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}(1+|Y^{j,\star,N}_n|^2)\Big|^p\Big]+H^{Y,p}_nCh. \tag{3.63}
$$

Plug (3.62) and (3.63) into $H^{X,p}_{n+1}$ we have

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

Thus finally, for all $n\in[\![0,M-1]\!],\ i\in[\![1,N]\!]$, by backward induction collecting all the results above, since $m\ge 2p$, where m is defined in (3.1), we have (for some $C>0$ independent of h,N,M)

$$
\begin{aligned}
\mathbb{E}\big[|\hat{X}^{i,N}_{n+1}|^{2p}\big]&\le H^{X,p}_{n+1}\le H^{Y,p}_n(1+Ch)\le H^{X,p}_n(1+Ch)^2\le\cdots\le H^{X,p}_0e^{CT}\\
&\le C\mathbb{E}\big[|\hat{X}^{i,N}_0|^{2p}\big]+C<\infty.
\end{aligned}
$$

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

Proof of the Theorem 3.2.10

Proof of the Theorem 3.2.10. Under the same assumptions and notations of Theorem 3.4.7, one can apply arguments similar to those used in Proposition 2.4.6 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 3.4.8. *Under same assumptions and notations of Theorem 3.2.10, 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* (3.1)*, q is defined in Assumption 3.2.1, we have*

$$\sup_{i\in[\![1,N]\!]}\sup_{0\leq t\leq T}\mathbb{E}\big[|\hat{X}_t^{i,N}-\hat{X}_{\kappa(t)}^{i,N}|^p\big]\leq Ch^{\frac{p}{2}}. \tag{3.64}$$

Proof. Under Assumption 3.2.1, and carefully applying Young's and Jensen's inequality, one can argue similarly as to Proposition 2.4.7 and obtain the result (we omit further details). □

3.4.4 Proof of Theorem 3.2.11, the uniform convergence result in path-space

We now prove Theorem 3.2.11.

Proof of Theorem 3.2.11. Let Assumption 3.2.1 hold. Let $i\in[\![1,N]\!]$, $t\in[0,T]$, suppose $m\geq \max\{4q+4,2+q+q/\varepsilon\}$, where $X_0^i\in L_0^m(\mathbb{R}^d)$, q is as given in Assumption 3.2.1. From (3.7) and (3.15), 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 $|X_t^{i,N}-\hat{X}_t^{i,N}|^2=|\Delta X_t^i|^2$ yields

$$|\Delta X_t^i|^2=2\int_0^t\Big\langle v(X_s^{i,N},\mu_s^{X,N})-v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\mathrm{d}s \tag{3.65}$$

$$+2\int_0^t\Big\langle b(s,X_s^{i,N},\mu_s^{X,N})-b(\kappa(s),Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\mathrm{d}s \tag{3.66}$$

$$+\int_0^t\Big|\sigma(s,X_s^{i,N},\mu_s^{X,N})-\sigma(\kappa(s),Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\Big|^2\mathrm{d}s \tag{3.67}$$

$$+2\int_0^t\Big\langle\Delta X_s^i,\Big(\sigma(s,X_s^{i,N},\mu_s^{X,N})-\sigma(\kappa(s),Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\Big)\mathrm{d}W_s^i\Big\rangle. \tag{3.68}$$

We analyse the above terms one by one and will take supremum over time with expectation. For (3.65),

$$\begin{aligned}&\big\langle v(X_s^{i,N},\mu_s^{X,N})-v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\big\rangle\\&=\big\langle v(X_s^{i,N},\mu_s^{X,N})-v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\big\rangle+\big\langle v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N})-v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\big\rangle.\end{aligned} \tag{3.69}$$

For the first term above, by Assumption 3.2.1 and using Remark 3.2.2

$$\begin{aligned}&\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t\Big\langle v(X_s^{i,N},\mu_s^{X,N})-v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle\mathrm{d}s\Big]\\
\leq&\mathbb{E}\Big[\int_0^T\frac{C}{N}\sum_{j=1}^N\Big|f(X_s^{i,N}-X_s^{j,N})-f(\hat{X}_s^{i,N}-\hat{X}_s^{j,N})\Big||\Delta X_s^i|\mathrm{d}s\Big]\\
&+\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t\Big\langle u(X_s^{i,N},\mu_s^{X,N})-u(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle\mathrm{d}s\Big]\\
\leq&\mathbb{E}\Big[\int_0^T\frac{C}{N}\sum_{j=1}^N\Big\{\Big(1+|X_s^{i,N}-X_s^{j,N}|^q+|\hat{X}_s^{i,N}-\hat{X}_s^{j,N}|^q\Big)|\Delta X_s^i-\Delta X_s^j||\Delta X_s^i|\Big\}\mathrm{d}s\Big] \qquad (3.70)\\
&+\mathbb{E}\Big[\int_0^T\Big(\widehat{L}_u|\Delta X_s^i|^2+\frac{L_{\tilde{u}}}{2N}\sum_{j=1}^N|\Delta X_s^j|^2\Big)\mathrm{d}s\Big].\end{aligned}$$

To deal with (3.70), 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 (3.7) and (3.15) respectively. Now, using Theorem 3.2.9, Proposition 3.4.6 and Young's inequality, we have

$$\mathbb{E}\big[G_7^{i,j,s}G_8^{i,j,s}\big] = \mathbb{E}\Big[G_7^{i,j,s}G_8^{i,j,s}(\mathbb{1}_{\{G_7^{i,j,s}<M^\varepsilon\}})\Big] + \mathbb{E}\Big[G_7^{i,j,s}G_8^{i,j,s}(\mathbb{1}_{\{G_7^{i,j,s}\geq M^\varepsilon\}})\Big] \tag{3.71}$$

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

$$\leq Ch^{1-\varepsilon} + hC\Big(1 + \mathbb{E}\Big[|X_s^{i,N}|^{2+q+q/\varepsilon} + |\hat{X}_s^{i,N}|^{2+q+q/\varepsilon}\Big]\Big) \leq Ch^{1-\varepsilon}, \tag{3.72}$$

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 $m \geq 2 + q + q/\varepsilon$.

Thus, for the first term in (3.69) and using that the particles are identically distributed, we conclude that

$$\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \Big\langle v(X_s^{i,N}, \mu_s^{X,N}) - v(\hat{X}_s^{i,N}, \hat{\mu}_s^{X,N}), \Delta X_s^i\Big\rangle \mathrm{d}s\Big] \leq C\mathbb{E}\Big[\int_0^T |\Delta X_s^i|^2 \mathrm{d}s\Big] + Ch^{1-\varepsilon}. \tag{3.73}$$

For the second term in (3.69), under Assumption 3.2.1, using Young's inequality, Jensen's inequality, and Proposition 3.4.8 we have

$$\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \Big\langle v(\hat{X}_s^{i,N}, \hat{\mu}_s^{X,N}) - v(Y_{\kappa(s)}^{i,\star,N}, \hat{\mu}_{\kappa(s)}^{Y,N}), \Delta X_s^i\Big\rangle \mathrm{d}s\Big] \tag{3.74}$$

$$=\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \Big\langle u(\hat{X}_s^{i,N}, \hat{\mu}_s^{X,N}) - u(Y_{\kappa(s)}^{i,\star,N}, \hat{\mu}_{\kappa(s)}^{Y,N}), \Delta X_s^i\Big\rangle \mathrm{d}s\Big] \tag{3.75}$$

$$+ \mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \frac{1}{N}\sum_{j=1}^N \langle f(\hat{X}_s^{i,N} - \hat{X}_s^{j,N}) - f(Y_{\kappa(s)}^{i,\star,N} - Y_{\kappa(s)}^{j,\star,N}), \Delta X_s^i\Big\rangle \mathrm{d}s\Big] \tag{3.76}$$

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

For I_2 (given by the domination of (3.75)), by Assumption 3.2.1, 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 (3.76) after extracting the $|\Delta X^i|$ term), by Assumption 3.2.1, Young's inequality and Cauchy-Schwarz inequality

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

$$\times \big|(\hat{X}_s^{i,N} - \hat{X}_s^{j,N}) - (Y_{\kappa(s)}^{i,\star,N} - Y_{\kappa(s)}^{j,\star,N})\big|^2\Big]\mathrm{d}s$$

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

By (3.10), Assumption 3.2.1, Young's inequality, Jensen's inequality, since $m \geq 4q + 4$, and by

Theorem 3.4.7, we have

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

Using this inequality in combination with Proposition 3.4.8 allows us to conclude that

$$
\mathbb{E}\big[|\hat{X}^{j,N}_{s} - Y^{j,\star,N}_{\kappa(s)}|^4\big] \leq C\mathbb{E}\Big[|\hat{X}^{j,N}_{s} - \hat{X}^{j,N}_{\kappa(s)}|^4 + |\hat{X}^{j,N}_{\kappa(s)} - Y^{j,\star,N}_{\kappa(s)}|^4\Big] \leq Ch^2. \tag{3.77}
$$

Thus, for (3.69) injected back in (3.65), take supremum and expectation, and collecting all the necessary results above, we reach

$$
\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \Big\langle v(X^{i,N}_s, \mu^N_s) - v(Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)}), \Delta X^i_s\Big\rangle \mathrm{d}s\Big] \leq C\mathbb{E}\Big[\int_0^T |\Delta X^i_s|^2\mathrm{d}s\Big] + Ch^{1-\varepsilon}. \tag{3.78}
$$

For the second term (3.66), the calculation is similar as in Proposition 2.4.9, we conclude that

$$
\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \Big\langle b(s, X^{i,N}_s, \mu^{X,N}_s) - b(\kappa(s), Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)}), \Delta X^i_s\Big\rangle \mathrm{d}s\Big] \leq Ch + C\mathbb{E}\Big[\int_0^T |\Delta X^i_s|^2\mathrm{d}s\Big]. \tag{3.79}
$$

Similarly, for the third term (3.67) (these are just Lipschitz terms), we have

$$
\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \Big|\sigma(s, X^{i,N}_s, \mu^{X,N}_s) - \sigma(\kappa(s), Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)})\Big|^2 \mathrm{d}s\Big] \leq Ch + C\mathbb{E}\Big[\int_0^T |\Delta X^i_s|^2\mathrm{d}s\Big]. \tag{3.80}
$$

Consider the last term (3.68) – this is a Lipschitz term and dealt with similarly to Proposition 2.4.9. Using the Burkholder-Davis-Gundy's, Jensen's and Cauchy-Schwarz inequality

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \Big\langle \Delta X^i_s, \Big(\sigma(s, X^{i,N}_s, \mu^{X,N}_s) - \sigma(\kappa(s), Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)})\Big)\mathrm{d}W^i_s\Big\rangle\ \Big] \qquad (3.81) \\
&\leq \frac{1}{4}\mathbb{E}\big[\sup_{0\leq t\leq T}|\Delta X^i_t|^2\big] + \mathbb{E}\Big[\int_0^T \Big|\sigma(s, X^{i,N}_s, \mu^{X,N}_s) - \sigma(\kappa(s), Y^{i,\star,N}_{\kappa(s)}, \hat{\mu}^{Y,N}_{\kappa(s)})\Big|^2 \mathrm{d}s\ \Big].
\end{aligned}
$$

Again, gathering all the above results (3.78), (3.79), (3.80), and (3.81), plugging them back into (3.65), after taking supremum over $t \in [0, T]$ and expectation, for all $i \in [\![1, N]\!]$ we have

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

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

3.4.5 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 Theorem 3.2.13. Recall the proof of (3.65) in Section 3.4.4. Under Assumption 3.2.12, for all $i\in[\![1,N]\!]$, $t\in[0,T]$, and using arguments similar to those of (3.69) we have

$$\Delta X_s^i = X_t^{i,N}-\hat{X}_t^{i,N} = \int_0^t v(X_s^{i,N},\mu_s^{X,N}) - v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\,\mathrm{d}s,$$

$$\Rightarrow \mathbb{E}\big[|\Delta X_t^i|^2\big] \le 2\int_0^t \mathbb{E}\Big[\Big\langle v(X_s^{i,N},\mu_s^{X,N}) - v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle\Big]\mathrm{d}s \tag{3.82}$$

$$+2\int_0^t \mathbb{E}\Big[\Big\langle v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}) - v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\Big]\mathrm{d}s. \tag{3.83}$$

For (3.82), arguing as in (3.40), Remark 3.2.4 and using that the particles are identically distributed, we have

$$\mathbb{E}\Big[\Big\langle v(X_s^{i,N},\mu_s^{X,N}) - v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}),\Delta X_s^i\Big\rangle\Big] \le 2L_f^{+}\mathbb{E}\big[|\Delta X_s^i|^2\big]. \tag{3.84}$$

For (3.83), it is similar to the above, we have

$$2\int_0^t \mathbb{E}\Big[\Big\langle v(\hat{X}_s^{i,N},\hat{\mu}_s^{X,N}) - v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}),\Delta X_s^i\Big\rangle\Big]\mathrm{d}s$$

$$=\frac{2}{N}\sum_{j=1}^N\int_0^t \mathbb{E}\Big[\Big\langle f(\Delta_s^{X,i,j}) - f(\Delta_{\kappa(s)}^{Y,i,j}),\Delta X_s^i\Big\rangle\Big]\mathrm{d}s, \tag{3.85}$$

where we introduce the following handy notation (recall (3.10) and (3.13))

$$\begin{aligned}
\Delta_t^{X,i,j} &= \hat{X}_s^{i,N}-\hat{X}_s^{j,N}, \qquad \Delta_{\kappa(s)}^{Y,i,j} = Y_{\kappa(s)}^{i,\star,N}-Y_{\kappa(s)}^{j,\star,N},\\
\Delta_s^{X,i,j} &= \Delta_{\kappa(s)}^{X,i,j}+G_9^{i,j,s}(s-\kappa(s))+G_{10}^{i,j,s}, \qquad \Delta_{\kappa(s)}^{Y,i,j} = \Delta_{\kappa(s)}^{X,i,j}+G_9^{i,j,s}h,\\
G_9^{i,j,s} &= \Big(v(Y_{\kappa(s)}^{i,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N}) - v(Y_{\kappa(s)}^{j,\star,N},\hat{\mu}_{\kappa(s)}^{Y,N})\Big) \text{ and } G_{10}^{i,j,s} = \sigma\Big((W_s^i-W_{\kappa(s)}^i)-(W_s^j-W_{\kappa(s)}^j)\Big).
\end{aligned} \tag{3.86}$$

We now proceed to estimate (3.85). By the mean value theorem under Assumption 3.2.12, for (3.85), there exist $\rho_1,\rho_2\in[0,1]$ such that

$$\begin{aligned}
&f(\Delta_s^{X,i,j})\\
&= f(\Delta_{\kappa(s)}^{X,i,j}) + \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-\kappa(s))+G_{10}^{i,j,s}\Big) + \int_{\Delta_{\kappa(s)}^{X,i,j}}^{\Delta_s^{X,i,j}}\Big(\nabla f(u)-\nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big)du\\
&= f(\Delta_{\kappa(s)}^{X,i,j}) + \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-\kappa(s))+G_{10}^{i,j,s}\Big)\\
&\quad+\Big(\nabla f\big(\Delta_{\kappa(s)}^{X,i,j}+\rho_1(G_9^{i,j,s}(s-\kappa(s))+G_{10}^{i,j,s})\big)-\nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big)\Big(\Delta_s^{X,i,j}-\Delta_{\kappa(s)}^{X,i,j}\Big),\\
&f(\Delta_{\kappa(s)}^{Y,i,j}) = f(\Delta_{\kappa(s)}^{X,i,j}) + \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}h\Big)\\
&\quad+\Big(\nabla f\big(\Delta_{\kappa(s)}^{X,i,j}+\rho_2(G_{10}^{i,j,s}h)\big)-\nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big)\Big(\Delta_{\kappa(s)}^{Y,i,j}-\Delta_{\kappa(s)}^{X,i,j}\Big).
\end{aligned}$$

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

$$\int_0^t \mathbb{E}\Big[\Big\langle f(\Delta_s^{X,i,j}) - f(\Delta_{\kappa(s)}^{Y,i,j}), \Delta X_s^i\Big\rangle\Big]\mathrm{d}s \tag{3.87}$$

$$\leq \int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-h-\kappa(s)) + G_{10}^{i,j,s}\Big), \Delta X_s^i\Big\rangle\Big]\mathrm{d}s + C\int_0^t \mathbb{E}\Big[|\Delta X_s^i|^2\Big]\mathrm{d}s \tag{3.88}$$

$$+ C\int_0^t \mathbb{E}\Big[\Big|\nabla f\Big(\Delta_{\kappa(s)}^{X,i,j} + \rho_{1,s}(G_9^{i,j,s}(s-\kappa(s)) + G_{10}^{i,j,s})\Big) - \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big|^2\Big|\Delta_s^{X,i,j} - \Delta_{\kappa(s)}^{X,i,j}\Big|^2\Big]\mathrm{d}s \tag{3.89}$$

$$+ C\int_0^t \mathbb{E}\Big[\Big|\nabla f\big(\Delta_{\kappa(s)}^{X,i,j} + \rho_{2,s}(G_9^{i,j,s}h)\big) - \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big|^2\Big|\Delta_{\kappa(s)}^{Y,i,j} - \Delta_{\kappa(s)}^{X,i,j}\Big|^2\Big]\mathrm{d}s. \tag{3.90}$$

For the first term of (3.88), by Young's inequality

$$\int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\Big(G_9^{i,j,s}(s-h-\kappa(s)) + G_{10}^{i,j,s}\Big), \Delta X_s^i\Big\rangle\Big]\mathrm{d}s \tag{3.91}$$

$$\leq C\int_0^t \mathbb{E}\Big[|\Delta X_s^i|^2\Big]\mathrm{d}s + C\int_0^t \mathbb{E}\Big[\Big|\nabla f(\Delta_{\kappa(s)}^{X,i,j})G_9^{i,j,s}(s-h-\kappa(s))\Big|^2\Big]\mathrm{d}s \tag{3.92}$$

$$+ \int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\ G_{10}^{i,j,s}, \Delta X_s^i - \Delta X_{\kappa(s)}^i\Big\rangle\Big]\mathrm{d}s + \int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\ G_{10}^{i,j,s}, \Delta X_{\kappa(s)}^i\Big\rangle\Big]\mathrm{d}s. \tag{3.93}$$

For the second term of (3.92), since $m \geq 4q+2$, by Assumption 3.2.12 and Theorem 3.2.10, using calculations similar to those in (3.45) and Proposition 3.4.6, 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 (3.76), since $m \geq 4q+2$, we have

$$\mathbb{E}\big[|\Delta X_t^i - \Delta X_{\kappa(t)}^i|^2\big] = \mathbb{E}\Big[\Big|\int_{\kappa(t)}^t \Big(v(X_s^{i,N}, \mu_s^{X,N}) - v(Y_{\kappa(s)}^{i,\star,N}, \hat{\mu}_{\kappa(s)}^{Y,N})\Big)\ \mathrm{d}s\Big|^2\Big] \tag{3.94}$$

$$\leq h\int_{\kappa(t)}^t \frac{1}{N}\sum_{j=1}^N \mathbb{E}\Big[\Big|f(X_s^{i,N} - X_s^{j,N}) - f(Y_{\kappa(s)}^{i,\star,N} - Y_{\kappa(s)}^{j,\star,N})\Big|^2\Big]\ \mathrm{d}s \leq Ch^3. \tag{3.95}$$

Thus, for the first term of (3.93), by Cauchy-Schwarz inequality and the properties of the Brownian increment

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

For the second term of (3.93), since $G_{10}^{i,j,s}$ of (3.86) 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

$$\int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta_{\kappa(s)}^{X,i,j})\ G_{10}^{i,j,s}, \Delta X_{\kappa(s)}^i\Big\rangle\Big]\mathrm{d}s = 0. \tag{3.96}$$

Thus, plugging the above results back into (3.88), we conclude that

$$\int_0^t \mathbb{E}\Big[\Big\langle \nabla f(\Delta^{X,i,j}_{\kappa(s)})\Big(G_9^{i,j,s}(s-h-\kappa(s))+G_{10}^{i,j,s}\Big), \Delta X^i_s\Big\rangle\Big]\mathrm{d}s \le Ch^2. \tag{3.97}$$

For (3.89), by Assumption 3.2.12, Cauchy-Schwarz inequality and the properties of the Brownian increment, and the condition $m \ge \max\{8q,\ 4q+4\}$

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

and

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

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

$$\int_0^t \mathbb{E}\Big[\big|\nabla f\big(\Delta^{X,i,j}_{\kappa(s)}+\rho_{1,s}\big(G_9^{i,j,s}(s-\kappa(s))+G_{10}^{i,j,s}\big)\big)-\nabla f(\Delta^{X,i,j}_{\kappa(s)})\big|^2\big|\Delta^{X,i,j}_s-\Delta^{X,i,j}_{\kappa(s)}\big|^2\Big]\mathrm{d}s \le Ch^2. \tag{3.98}$$

For (3.90), recall (3.86). Similarly to above, by assumption $m \ge 4q+2$ and hence

$$\int_0^t \mathbb{E}\Big[\ \big|\nabla f\big(\Delta^{X,i,j}_{\kappa(s)}+\rho_{2,s}G_9^{i,j,s}h\big)-\nabla f\big(\Delta^{X,i,j}_{\kappa(s)}\big)\big|^2\ \big|G_9^{i,j,s}h\big|^2\ \Big]\mathrm{d}s \le Ch^2. \tag{3.99}$$

Thus, plugging (3.97), (3.98) and (3.99) back into (3.87), yields

$$\int_0^t \mathbb{E}\Big[\Big\langle f(\Delta^{X,i,j}_s)-f(\Delta^{Y,i,j}_{\kappa(s)}), \Delta X^i_s\Big\rangle\Big]\mathrm{d}s \le Ch^2 + C\int_0^t \mathbb{E}\big[|\Delta X^i_s|^2\big]\mathrm{d}s. \tag{3.100}$$

Plug the above result and (3.84) back to (3.82), we conclude that, for all $i \in [\![1,N]\!]$, $t\in[0,T]$

$$\mathbb{E}\big[|\Delta X^i_t|^2\big] \le C\int_0^t \mathbb{E}\big[|\Delta X^i_s|^2\big]\mathrm{d}s + Ch^2. \tag{3.101}$$

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

Chapter 4

Well-posedness, ergodicity and numerical analysis for McKean-Vlasov SDEs with fully super-linear growth drifts and diffusion in space and interaction

4.1 Introduction

In this chapter, we analyze a class of McKean–Vlasov Stochastic Differential Equations (MV-SDEs) having drift and diffusion components of convolution type, akin to the porous media equation or interaction kernel modelling, which allows for *super-linear growth in measure and space* in both coefficients – one may think of the super-linearity as higher-order polynomials under some additional conditions regulating spatial and measure radial growth. We work with MV-SDE dynamics of the form

$$\mathrm{d}X_t = \big(v(X_t,\mu_t^X) + b(t,X_t,\mu_t^X)\big)\mathrm{d}t + \overline{\sigma}(t,X_t,\mu_t^X)\mathrm{d}W_t, \quad X_0 \in L^m(\mathbb{R}^d),\ m>2 \tag{4.1}$$

$$\text{where} \quad \begin{cases} v(x,\mu) &= \int_{\mathbb{R}^d} f(x-y)\mu(\mathrm{d}y) + u(x,\mu), \\ \overline{\sigma}(t,x,\mu) &= \sigma(t,x,\mu) + \int_{\mathbb{R}^d} f_\sigma(x-y)\mu(\mathrm{d}y). \end{cases} \tag{4.2}$$

Above, μ_t^X denotes the law of the solution process X at time t, $L^m(\mathbb{R}^d)$ is the space of $\mathcal{F}_0$-measurable random variables with finite m-th moments, W is a multidimensional Brownian motion, u, b, σ and f, f_σ are measurable maps. Critically, f, f_σ, u, σ are maps of super-linear growth but not assumed to be differentiable and $\overline{\sigma}$ may degenerate. The function u in v allows to incorporate measure dependencies other than convolution type.

In terms of a modelling perspective in the context of particle dynamics, (4.1)-(4.2) model the dynamics of particle motion where the particle is affected by different sources of forcing. The map u represents a multi-well gradient potential confining the particle (and the source of super-linear growth in the spatial component) and the convolution map f contains information on the forces affecting the particles (attractive, repulsive), see [2, 83, 138]. As argued in [83], under certain assumptions, v (and f) adds inertia to the particle's dynamic in turn affecting its exit time from a domain of attraction (by accelerating or delaying it) and alters exit locations [2, 61, 83] (also [56]). To motivate the study of equations with a (nonlinear) convolution term $(f_\sigma * \mu)(\cdot) := \int_{\mathbb{R}^d} f_\sigma(\cdot - y)\mu(\mathrm{d}y)$ in the diffusion component, which is the main feature of our work, we first mention [80]. There, a Cucker–Smale model incorporating random communication is rewritten as a Cucker–Smale model with multiplicative noise (the diffusion coefficient has the form $(\mathbb{E}[X_t] - X_t) = \int_{\mathbb{R}^d}(X_t - y)\mu_t(\mathrm{d}y)$), which helps to stabilize flocking states as the effect of the noise diminishes the closer the particles concentrate around their mean; see also [7].

These works give a clear motivation to analyze convolution type diffusion maps (diffusions whose strength depends on the density) – also [64] studies a kinetic flocking model with a more general distance potential (communication rate) function than [80]. In addition, [25] considers general stochastic systems of interacting particles with Brownian noise to study models for the collective behavior (swarming) – more particularly, [25, Section 1.2.2] highlights several open-question model extensions to nonlinear diffusion coefficients (though beyond the scope of this chapter). The recent works [26, 36] investigate consensus-based optimization (CBO) methods for solving high-dimensional nonlinear unconstrained minimization problems. A CBO scheme updates the particle's position in an iterative manner to explore the optimization landscape. There, particles far away from the equilibrium state are expected to exhibit more exploration (i.e., the noise level should be larger) compared to particles close to it. Inspired by the above discussed works, we offer a new class of MV-SDEs adding a new element in the diffusion coefficient by means of a reversion to the population mean expressed through a *fully non-Lipschitz* $f_\sigma * \mu$ significantly beyond the linear interaction diffusion coefficients studied in the mentioned works.

More generally, the motivation to study this class of MV-SDEs and associated interacting particle systems is to present a unified framework to address wellposedness and establish properties useful for downstream applications. For instance, from emerging models of mean-field type in neuroscience [68], understating particle motion and exit times [2, 83, 84, 138, 141], parametric inference [17] (also [53, 74]) is an important consideration. We also point to Section 4 and 5 of [92] for a variety of general interacting systems that are subsumed by our class. Our results can also be viewed as an addition to the literature on granular media type equations as studied in [39, 78, 126].

The existence and uniqueness of solutions to MV-SDEs, in a strong and weak sense, has been extensively studied, see e.g., [35, 43, 87, 94, 100, 110, 116, 121, 127, 133] and references therein, but *none* cover the setting presented here. To the best of our knowledge, the existence and uniqueness of strong solutions to equations with super-linear growth in the measure component of the drift and the diffusion has not been addressed in general. There exist various works considering super-linearly growing coefficients (in state) but do not incorporate f or f_σ, see e.g., [99, 116, 140] and its references. In [2] the authors deal with a super-linear f, $f_\sigma \equiv 0$, and a (unbounded) uniformly Lipschitz continuous σ, and derive wellposedness (i.e., existence and uniqueness of a strong solution) and large deviation results. Further, [143] allows for a setting similar to ours but requires upfront strong dissipativity and non-degeneracy. Their aim was to study ergodicity, nonetheless, it is unclear how to adapt their methodology if working with the goal of proving wellposedness over $[0,T]$ under milder conditions. From the initial work [2], our goal is to develop a general framework to study (4.1)-(4.2) in terms of wellposedness (over $[0,T]$), with a super-linearly growing σ and f_σ, ergodicity and approximation schemes. **Our contributions.**

Our *first main contribution* concerns wellposedness and propagation of chaos (PoC) results for the finite time horizon $[0,T]$ case. The critical nontrivial hurdle of this setting is in establishing L^p-moment bounds for $p > 2$ under the presence of the super-linear growths of f, f_σ and σ – *this issue appears solely due* to the simultaneous presence of nonlinearities (in space and measure) in the drift and diffusion, otherwise techniques like those of [2] or [99] would suffice. To overcome this hurdle, we introduce a new condition dubbed 'additional symmetry', which is new in the literature (to the best of our knowledge). For a quick perspective, we suggest the reader glance at Lemma 6.3.1 and 6.3.2 (in Appendix) and the proof of Theorem 4.2.5 to see how one deals with the convolution terms and to note the importance of the 'additional symmetry' condition – a discussion on this latter condition is presented in Remark 4.2.6. We also address a propagation of chaos result [57, 61, 71, 118, 133] for this class. We show the interacting particle system, obtained by replacing μ^X by the system's N-particle empirical distribution, recovers the original MV-SDE in the particle limit $N \to \infty$. Under a very mild higher-integrability assumption, a convergence rate is obtained.

The *second main contribution* of this chapter is a numerical method to approximate (4.1)-(4.2) over $[0,T]$ via its interacting particle system. Most of our theoretical results are only proven for the finite time case, but we successfully apply the scheme for the long-term simulation of a particle system as well. There are presently many studies for numerical methods allowing super-linear spatial growth of drifts (and diffusions): Euler type methods, e.g., taming [60], time-adaptive [125], semi-implicit methods (Chapter 2, 3), projection methods [18]; Milstein type methods e.g., [14, 13, 98, 123] with some allowing super-linear σ in space. There are variations on the assumptions, but *all* these contributions require drifts and diffusions to be globally Lipschitz continuous in measure (with respect to the Wasserstein distance with quadratic cost). Two recent contributions [106, 119] allow for weaker continuity conditions than Lipschitz for the coefficients but require a linear growth in space and measure. Only the results in Chapter 3 allows for general super-linear growth in f, u but still limits $\overline{\sigma}$ to satisfy Lipschitz assumptions – we detail below the differences between Chapter 3 and this manuscript in more detail.

The scheme we propose here belongs to the split-step method (SSM) class as in previous chapters. We follow the strategy of approximating (in time) the interacting particle system associated with the MV-SDE and using a quantitative propagation of chaos (PoC) convergence result, see [29] and [60, 123, 125] for earlier uses of this strategy. From a methodological point of view, the convergence proof of our numerical scheme is different from any used to study MV-SDE numerical schemes in the literature; our highly-non-linear setting forces us to draw on the stochastic C-stability and B-consistency mechanics proposed in [22]. Its use in the context of numerical schemes for MV-SDEs and interacting particle systems is novel in the literature – except for the very recent [24] that studies higher order strong scheme for MV-SDE under non-differentiability conditions using a randomisation method. In [24], the authors work with generic Lipschitz assumptions and need to change the underpinning error norms to cope with the complexity arising from the randomization step due to an explicit non-differentiability assumption on the drift coefficient. Our approach and requirements differ, and so does the analysis (albeit similar at points). We show that it is possible to work directly with the concepts of [22] to deal with the interacting particle system, see Section 4.2.4 – we emphasize that the main goal of the analysis is to guarantee that core moment estimates are uniformly independent of the number of particles N of the interacting system (but may depend on the initial system's underlying dimension d). As is common in the MV-SDE literature, results from the SDE numerics literature on super-linear growth do not carry over to the MV-SDE one.

Closest to our work with regards to the SSM in Chapter 3, where the authors propose an SSM scheme similar to the one here for interacting particle systems that have (4.1) (with $f_\sigma \equiv 0$ and σ globally Lipschitz continuous in space and measure) as limit. There they overcome the barrier of super-linear growth in space and measure for the drift (that [47, 60, 125] do not), but work with a diffusion component of Lipschitz type; the focus of Chapter 3 is solely the analysis of the numerical scheme and not wellposedness or ergodicity. Our setting is more involved those of [47, 48] and requires novel proof techniques – here, drawing from [22], specifically due to the simultaneous super-linearity in drift and diffusion. In Chapter 3, bounds for higher order moments of the discrete process obtained by the time-stepping scheme can be developed by commonly used assumptions like ($\mathbf{A}^u$, $\mathbf{A}^\sigma$) below. For our situation, the super-linearity in the diffusion coefficient (in space and measure) is controlled by a drift satisfying a suitable one-sided Lipschitz condition. However, the simultaneous appearance of nonlinearities in the diffusion and the nonlinear convolution $f * \mu$ in the drift causes difficulties. This is the reason why, for the scheme in this manuscript, only L^2-moment bounds were established, and the proof methodology takes recourse in the stochastic C-stability and B-consistency mechanics of [22] (which does not require to establish bounds for higher order moments of the SSM). It remains unclear how to obtain higher moments (even with the help of the new additional symmetry assumption).

We present several numerical examples of interest and for comparison we implement, without proof, two intuitive versions of taming methods [60, 123]. Our examples show the SSM to perform very well for the approximation of a solution to (4.1) on $[0, T]$ in an L^2-sense or the approximation for the ergodic distribution. The numerical results using taming are mixed but suffice to highlight which version can be expected to converge (theoretically). *We show a surprising numerical divergence finding for taming given the choice of initial condition and which does not appear when using the SSM* (see our Section 4.3.1) – this confirms the SSM as a stable/robust choice of scheme for this class. The SSM is shown to preserve periodicity of phase-space. Lastly, we provide a numerical example with the aim of estimating the PoC rate across dimensions and which highlights a gap in the literature: we observe the rate of [57, 118] (not applicable to our setting) instead of those in [34, 71] (which are used to prove our PoC result). Future research focuses on the study of uniform in-time PoC results and strong convergence rates for the SSM on $[0, \infty)$.

We also establish ergodicity result for the SSM, we demonstrate with an example, the study on invariant distribution of fully super-linear growth type MV-SDEs remains an open challenge and is left for future investigation.

Organization of this chapter. Section 4.2 contains: framework, wellposedness results, the particle approximation and the propagation of chaos statement, and the numerical scheme alongside associated convergence results. Several numerical examples are provided in Section 4.3. They cover the non-dissipative case in short and long time horizons; approximation of the invariant distribution; preservation of periodicity in phase-space and numerical estimation of PoC rates across dimension. All proofs are postponed to Section 4.4. Generic auxiliary results are given in the Appendix.

4.2 Main results

4.2.1 Framework

Let $v : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, $b : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and $\overline{\sigma} : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times l}$ be measurable maps. The MV-SDE of interest of this chapter is Equation (4.1) (for some $m > 2$), 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 4.2.1. *The functions b and σ are 1/2-Hölder continuous in time, uniformly in $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\sup_{t\in[0,\infty)} \big(|b(t, 0, \delta_0)| + |\sigma(t, 0, \delta_0)|\big) \le L$, for some constant $L \ge 0$.*

($\mathbf{A}^b$) *Let b be uniformly Lipschitz continuous in the sense that there exists $L^{(1)}_{(b)}$, $L^{(3)}_{(b)} \ge 0$ and $L^{(2)}_{(b)} \in \mathbb{R}$ such that for all $t \in [0, \infty)$, $x, x' \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ we have that*

$$|b(t, x, \mu) - b(t, x', \mu')|^2 \le L^{(1)}_{(b)}\big(|x - x'|^2 + \big(W^{(2)}(\mu, \mu')\big)^2\big),$$
$$\langle x - x', b(t, x, \mu) - b(t, x', \mu')\rangle \le L^{(2)}_{(b)}|x - x'|^2 + L^{(3)}_{(b)}\big(W^{(2)}(\mu, \mu')\big)^2,$$

where the first Lipschitz condition leads to the second one-sided Lipschitz condition, we separate these conditions with specific constants for clearer demonstration of the later results (e.g Theorem 4.2.19).

($\mathbf{A}^u$, $\mathbf{A}^\sigma$) *Let u, σ satisfy: there exist $L^{(1)}_{(u\sigma)} \in \mathbb{R}$, and $L^{(2)}_{(u\sigma)}, L^{(3)}_{(u\sigma)}, L^{(4)}_{(u\sigma)}, q_1 \ge 0$ such that for all $t \in [0, \infty)$, $x, x' \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, with $m > 2$ in (4.1), we have that*

$$\begin{aligned}&\langle x - x', u(x, \mu) - u(x', \mu')\rangle + 2(m - 1)|\sigma(t, x, \mu) - \sigma(t, x', \mu')|^2 \\ &\qquad\le L^{(1)}_{(u\sigma)}|x - x'|^2 + L^{(2)}_{(u\sigma)}\big(W^{(2)}(\mu, \mu')\big)^2, \end{aligned} \tag{4.3}$$

$$\begin{aligned}&|u(x, \mu) - u(x', \mu)| + |\sigma(t, x, \mu) - \sigma(t, x', \mu)| \\ &\qquad\le L^{(3)}_{(u\sigma)}(1 + |x|^{q_1} + |x'|^{q_1})|x - x'|,\end{aligned} \tag{4.4}$$

$$|u(x, \mu) - u(x, \mu')|^2 + |\sigma(t, x, \mu) - \sigma(t, x, \mu')|^2 \le L^{(4)}_{(u\sigma)}\big(W^{(2)}(\mu, \mu')\big)^2.$$

($\mathbf{A}^f$, $\mathbf{A}^{f_\sigma}$) *Let f, f_σ satisfy: there exist $L^{(1)}_{(f)}, L^{(3)}_{(f)} \in \mathbb{R}$, and $L^{(2)}_{(f)}, q_2 \ge 0$, such that for all $x, x' \in \mathbb{R}^d$, $2 < p \le m$, we have that*

$$\begin{aligned}&\langle x - x', f(x) - f(x')\rangle + 2(m - 1)|f_\sigma(x) - f_\sigma(x')|^2 \\ &\qquad\le L^{(1)}_{(f)}|x - x'|^2, \text{(One-sided Lipschitz, monotonicity condition)},\end{aligned}$$
$$|f(x) - f(x')| + |f_\sigma(x) - f_\sigma(x') \le L^{(2)}_{(f)}(1 + |x|^{q_2} + |x'|^{q_2})|x - x'|, \quad \text{(Locally Lipschitz)},$$
$$f(x) = -f(-x), \qquad \text{(Odd function)},$$
$$(|x|^{p-2} - |x'|^{p-2})\langle x + x', f(x - x')\rangle \le L^{(3)}_{(f)}(|x|^p + |x'|^p), \qquad \text{(Additional symmetry)}.$$

Assume the normalization[1] $f(0) = f_\sigma(0) = 0$. *Lastly, and for convenience, we set* $q = \max\{q_1, q_2\} \geq 0$.

Remark 4.2.2 (Time dependency for u)**.** *To avoid added complexity to an already complex work, we do not address time-dependence on u. A close inspection of the proof for wellposedness and convergence of the numerical scheme shows that as long as the time dependence does not interfere with constraints imposed by Assumption 4.2.1 the results will hold. Additionally one would require a 1/2-Hölder continuity property for the function.*

All elements in the above assumption are standard, except the 'additional symmetry' restriction. The 'additional symmetry' is a new type of restriction which we have not found previously in the literature and we discuss it in more detail at several points in the text, in particular, in Remark 4.2.6.

This condition is trivially satisfied when $d = 1$ (see (4.5)) or when the function is Lipschitz. We next provide a non-trivial example in $d > 1$ for f satisfying the 'extra symmetry' condition.

Example 4.2.3. *For $x \in \mathbb{R}^d$ define $f(x) = -x|x|^2$. Then, for any $p > 2$, $x, y \in \mathbb{R}^d$ it holds that*

$$(|x|^{p-2} - |y|^{p-2})\langle x + y, -(x - y)|x - y|^2\rangle = -(|x|^{p-2} - |y|^{p-2})(|x|^2 - |y|^2)|x - y|^2 \leq 0,$$

and the conclusion follows from the monotonicity of the polynomial function.

Remark 4.2.4 (Implied properties)**.** *Let Assumption 4.2.1 hold with $m > 2$. We provide the following estimates for some positive constant C which may change line by line, which are derived using the one-sided Lipschitz condition and Young's inequality (see Remark 3.2.2 for details). For all $t \in [0,T]$, $x, x', z \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, we have*

$$\begin{aligned}
\langle x, f(x)\rangle + 2(m-1)|f_\sigma(x)|^2 &\leq L^{(1)}_{(f)}|x|^2, \ |b(t,x,\mu)|^2 \leq C\big(1 + |x|^2 + (W^{(2)}(\mu,\delta_0))^2\big),\\
\langle x, u(x,\mu)\rangle + (m-1)|\sigma(t,x,\mu)|^2 &\leq C\big(1 + |x|^2 + (W^{(2)}(\mu,\delta_0))^2\big),\\
\langle x - x', u(x,\mu) - u(x',\mu')\rangle &\leq C\big(|x - x'|^2 + (W^{(2)}(\mu,\mu'))^2\big),\\
\langle x, b(t,x,\mu)\rangle &\leq C(1 + |x|^2 + (W^{(2)}(\mu,\delta_0))^2),\\
\langle x - x', v(x,\mu) - v(x',\mu)\rangle &\leq (L^{(1)}_{(u\sigma)} + L^{(1)}_{(f)})|x - x'|^2,\\
|\overline{\sigma}(t,x,\mu)|^2 &\leq 2|\sigma(t,x,\mu)|^2 + 2\Big|\int_{\mathbb{R}^d} f_\sigma(x-y)\mu(\mathrm{d}y)\Big|^2\\
&\leq 2\Big(|\sigma(t,x,\mu)|^2 + \int_{\mathbb{R}^d} |f_\sigma(x-y)|^2\mu(\mathrm{d}y)\Big).
\end{aligned}$$

Let $f : \mathbb{R} \to \mathbb{R}$ be an odd-function satisfying the one-sided Lipschitz condition, then f satisfies the additional symmetry condition, i.e., for $x, y \in \mathbb{R}, x \neq y, p \geq 2$, we have

$$\begin{aligned}
(|x|^{p-2} - |y|^{p-2})\langle x + y, f(x-y)\rangle &= \frac{(|x|^{p-2} - |y|^{p-2})(x+y)}{x-y}\langle x - y, f(x-y)\rangle\\
&\leq C(|x|^p + |y|^p). \qquad (4.5)
\end{aligned}$$

The following decomposition is crucial for the remaining parts of this chapter

[1] This constraint is not restrictive since the framework allows to easily redefine f as $\hat{f}(x) := f(x) - f(0)$ with $f(0)$ merged into b.

for $x\in\mathbb{R}^d$, $m\geq p>2$, $\mu\in\mathcal{P}_m(\mathbb{R}^d)$ it holds that

$$\begin{aligned}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|x|^{p-2}\langle x,f(x-y)\rangle\mu(\mathrm{d}y)\mu(\mathrm{d}x)&=\frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\langle|x|^{p-2}x-|y|^{p-2}y,f(x-y)\rangle\mu(\mathrm{d}y)\mu(\mathrm{d}x)\\ &=\frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\langle(|x|^{p-2}x-|x|^{p-2}y)+(|x|^{p-2}y-|y|^{p-2}y),f(x-y)\rangle\mu(\mathrm{d}y)\mu(\mathrm{d}x)\\ &=\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\big(\tfrac{1}{2}|x|^{p-2}\langle x-y,f(x-y)\rangle+\tfrac{1}{4}(|x|^{p-2}-|y|^{p-2})\langle x+y,f(x-y)\rangle\big)\mu(\mathrm{d}y)\mu(\mathrm{d}x).\end{aligned}\tag{4.6}$$

The decomposition in (4.6) *along with* ($\mathbf{A}^f$, $\mathbf{A}^{f_\sigma}$) *will be used to incorporate the nonlinearity of f_σ.*

4.2.2 Existence and uniqueness of the MV-SDE

Let us start by stating the wellposedness result of MV-SDE (4.1).

Theorem 4.2.5 (Wellposedness)**.** *Let Assumption 4.2.1 hold with $m>2q+2$, then there exists a unique strong solution X to MV-SDE* (4.1) *satisfying the following estimates: For some constant $C>0$, we have a pointwise estimate*

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

The proof of the wellposedness theorem is postponed to Section 4.4.1.

Remark 4.2.6 (On the 'additional symmetry' restriction)**.** *The critical element of the proof for this result, is the difficulty in establishing (finite) bounds for higher order moments of the solution process. The 'additional symmetry' assumption is a technical condition without which we were not able to establish L^p-moment bounds for $p>2$ (and $d>1$) – proving L^2-moment bounds or uniqueness of the solution is straightforward and the condition is not needed. The requirement of 'additional symmetry' stems solely from having a super-linearly growing σ, f_σ and a super-linear growth of the convolution term appearing in the drift. If either of them is of linear growth (or $d=1$), then the 'additional symmetry' condition can be removed and the results hold.*

The strategy used in [2] to establish L^p-moment bounds, working with Assumption 4.2.1 but with a linearly growing σ, is to bound $\mathbb{E}\big[|X_t|^{2p}\big]$ via

$$\mathbb{E}\big[|X_t|^{2p}\big]\leq C\big(\mathbb{E}\big[\big|X_t-\mathbb{E}[X_t]\big|^{2p}\big]+\mathbb{E}\big[|X_t|^2\big]^p\big),$$

and then noticing that

$$\begin{aligned}\mathbb{E}\big[\big|X_t-\mathbb{E}[X_t]\big|^{2p}\big]&=\int_{\mathbb{R}^d}\Big|x-\int_{\mathbb{R}^d}y\mu_t(\mathrm{d}y)\Big|^{2p}\mu_t(\mathrm{d}x)\\ &\leq\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|x-y|^{2p}\mu_t(\mathrm{d}y)\mu_t(\mathrm{d}x)=\mathbb{E}\big[|X_t-\tilde{X}_t|^{2p}\big],\end{aligned}$$

with $\tilde{X}$ an independent copy of X driven by its independent Brownian motion, see Lemma 6.3.1 and Lemma 6.3.2 for extra details. This trick allows to deal with the convolution term, employing its symmetry, (see Lemma 6.3.2), but does not give control of the super-linear diffusion. To be precise, Itô's formula applied to $|X-\tilde{X}|^{2p}$ forces one to use the polynomial growth condition on σ (4.4)*, which involves higher moments, instead of* (4.3)*.*

Without the trick described above, and following more classical approaches [123, Theorem 2.1], it is possible to control the super-linear growth of σ in space (via (4.3)*) but it is unclear how to simultaneously control the super-linear growth of the convolution terms in a tractable way (the tricks of Lemma 6.3.1 and Lemma 6.3.2 do not carry over).*

All in all, there is competition between the growths of f and σ, f_σ, and neither just described technique is adequate to establish L^p-moment estimates. The 'additional symmetry' condition *offsets this difficulty. See details in the proof in Section 4.4.1. Lifting this restriction is left as an open question.*

4.2.3 Particle approximation of the MV-SDE

We now turn to the particle approximation of the MV-SDE with the ultimate goal of establishing a working numerical scheme for the equation. All results here are only concerned with the finite-time case.

As in [29, 47, 125], we approximate the MV-SDE (4.1) (driven by the Brownian motion W) by an interacting particle system, i.e., an N-dimensional system of $\mathbb{R}^d$-valued interacting particles. Let $i \in [\![1, N]\!]$ and consider N particles $(X_t^{i,N})_{t\in[0,T]}$ with independent and identically distributed (i.i.d.) initial data $X_0^{i,N} = X_0^i$ (an independent copy of X_0) satisfying the $(\mathbb{R}^d)^N$-valued SDE

$$\mathrm{d}X_t^{i,N} = \big(v(X_t^{i,N}, \mu_t^{X,N}) + b(t, X_t^{i,N}, \mu_t^{X,N})\big)\mathrm{d}t + \overline{\sigma}(t, X_t^{i,N}, \mu_t^{X,N})\mathrm{d}W_t^i, \quad X_0^{i,N} = X_0^i, \quad (4.7)$$

where $\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}(\mathrm{d}x)$ with δ_x being the Dirac measure at $x \in \mathbb{R}^d$, and W^i being independent Brownian motions (also independent of the Brownian motion appearing in (4.1)).

We introduce similarly to Remark 3.2.4 the auxiliary maps V, and $\hat{\Sigma}$ to view (4.7) as a system in $\mathbb{R}^{Nd}$.

Lemma 4.2.7 (Properties of the particle system as a system in $\mathbb{R}^{Nd}$)**.** *Define $V : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$, $\hat{\Sigma} : [0,T] \times \mathbb{R}^{Nd} \to \mathbb{R}^{N(d\times l)}$ by $V(x^N) = \big(\dots, v(x^{i,N}, \mu^{x,N}), \dots\big)$, and $\hat{\Sigma}(t, x^N) = \big(\dots, \overline{\sigma}(t, x^{i,N}, \mu^{x,N}), \dots\big)$ with $x^N = (x^{1,N}, \dots, x^{N,N}) \in \mathbb{R}^{Nd}$, $t \in [0,T]$.*

Then, under Assumption 4.2.1 with $m > 2$, for any $x^N, y^N \in \mathbb{R}^{Nd}$ with corresponding empirical measures $\mu^{x,N} = \frac{1}{N}\sum_{j=1}^N \delta_{x^{j,N}}$, and $\mu^{y,N} = \frac{1}{N}\sum_{j=1}^N \delta_{y^{j,N}}$, the functions V, $\hat{\Sigma}$ also satisfy a monotonicity condition (see first item of ($\mathbf{A}^f$, $\mathbf{A}^{f_\sigma}$) in Assumption 4.2.1) in $\mathbb{R}^{Nd}$ (with constants independent of N).

Proof. From Assumption 4.2.1, Remark 4.2.4 and Jensen's inequality, we deduce, for all $x^N, y^N \in \mathbb{R}^{Nd}$, $t \in [0,T]$,

$$\begin{aligned}
&\langle x^N - y^N, V(x^N) - V(y^N)\rangle + \frac{(m-1)}{2}|\hat{\Sigma}(t,x^N) - \hat{\Sigma}(t,y^N)|^2 \\
&\leq \frac{1}{2N}\sum_{i=1}^N\sum_{j=1}^N \big\langle (x^{i,N} - x^{j,N}) - (y^{i,N} - y^{j,N}), f(x^{i,N} - x^{j,N}) - f(y^{i,N} - y^{j,N})\big\rangle \\
&\quad + \sum_{i=1}^N \Big(\big\langle x^{i,N} - y^{i,N}, u(x^{i,N}, \mu^{x,N}) - u(y^{i,N}, \mu^{y,N})\big\rangle \\
&\qquad\qquad + (m-1)|\sigma(t, x^{i,N}, \mu^{x,N}) - \sigma(t, y^{i,N}, \mu^{y,N})|^2\Big) \\
&\quad + \frac{m-1}{N}\sum_{i=1}^N\sum_{j=1}^N |f_\sigma(x^{i,N} - x^{j,N}) - f_\sigma(y^{i,N} - y^{j,N})|^2 \leq C|x^N - x^N|^2,
\end{aligned}$$

where $C > 0$ is independent of N.

□

Propagation of chaos (PoC). In order to show that the particle approximation (4.7) is effective to approximate the underlying MV-SDE, we present a pathwise propagation of chaos result (convergence as the number of particles increases). To do so, we introduce the system of non interacting particles

$$\mathrm{d}X_t^i = \big(v(X_t^i, \mu_t^{X^i}) + b(t, X_t^i, \mu_t^{X^i})\big)\mathrm{d}t + \overline{\sigma}(t, X_t^i, \mu_t^{X^i})\mathrm{d}W_t^i, \quad t \in [0,T], \quad (4.8)$$

which are (decoupled) MV-SDEs with i.i.d. initial conditions X_0^i (an independent copy of X_0). Since the X^i's are independent, $\mu_t^{X^i} = \mu_t^X$ for all i (and μ_t^X the marginal law of the solution to (4.1)). We are interested in the strong error-type metrics for the numerical approximation and the relevant PoC result for our case is given in the next theorem, the proof is postponed to Section 4.4.

Theorem 4.2.8 (Propagation of Chaos)**.** *Let the Assumptions of Theorem 4.2.5 hold for some $m > 2(q+1)$. Let X^i be the solution to (4.8) in the sense of Theorem 4.2.5. Then, there exists a unique solution $X^{i,N}$ to (4.7) and for any $1 \leq p \leq m$ there exists $C > 0$ independent of N such that*

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

Moreover, suppose that $m > 2(q+1)$ and $m > 4$, then we have the following convergence result

$$\sup_{i\in[\![1,N]\!]} \sup_{t\in[0,T]} \mathbb{E}\big[|X_t^{i,N} - X_t^i|^2\big] \leq C \begin{cases} N^{-1/2}, & d < 4, \\ N^{-1/2}\log N, & d = 4, \\ N^{-\frac{2}{d+4}}, & d > 4. \end{cases} \tag{4.9}$$

This result shows that the particle approximation will converge to the MV-SDE with a given rate. Therefore, to establish convergence of our numerical scheme to the MV-SDE (in a strong sense), we only need to show that the discrete-time version of the particle system converges to the "true" particle system.

4.2.4 C-stability and B-consistency for the particle system

Before introducing our numerical scheme and the corresponding strong convergence result, we first present a definition of C-stability and B-consistency for the particle system. The following definitions and methodologies are modifications of the original work in [22, 24] tailored to the present particle system setting. The probability space in this section supports (at least) the N driving Brownian motions of the particle system and the filtration corresponds to the enlarged filtration generated by all Brownian motions augmented by a rich enough σ-algebra $\mathcal{F}_0$.

Definition 4.2.9. *Let $h \in (0,1)$ be the stepsize and $\Psi_i : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times [0,T] \times (0,1) \times \Omega \to \mathbb{R}^d$ for all $i \in [\![1,N]\!]$ be a mapping satisfying the following measurability and integrability condition: For every $t, t+h \in [0,T],\ h \in (0,1)$ and $X = (X^1,\ldots,X^N) \in L^2\left(\Omega,\mathcal{F}_t,\mathbb{P};\mathbb{R}^{Nd}\right)$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ it holds*

$$\Psi_i(X^i,\mu,t,h) \in L^2\big(\Omega,\mathcal{F}_{t+h},\mathbb{P};\mathbb{R}^d\big), \quad \Psi = (\Psi_1,\ldots,\Psi_N). \tag{4.10}$$

Then, for $M \in \mathbb{N}, Mh = T,\ k \in [\![0, M-1]\!],\ t_k = kh$, we say that a particle system $\hat{X}_k^N = (\hat{X}_k^{1,N},\ldots,\hat{X}_k^{N,N}) \in \mathbb{R}^{Nd}$ is generated by the stochastic one-step method (Ψ,h,ξ) with initial condition $\xi = (\xi^1,\ldots,\xi^N) \in L^2\left(\Omega,\mathcal{F}_0,\mathbb{P};\mathbb{R}^{Nd}\right)$, $\Psi = (\Psi_1,\ldots,\Psi_N)$, if

$$\begin{aligned} &\hat{X}_{k+1}^{i,N} = \Psi_i(\hat{X}_k^{i,N},\hat{\mu}_k^{X,N},t_k,h), \quad \hat{\mu}_k^{X,N}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^N \delta_{\hat{X}_k^{j,N}}(\mathrm{d}x), \\ &\hat{X}_0^{i,N} = \xi^i, \quad i \in [\![1,N]\!]. \end{aligned}$$

We call Ψ the one-step map of the method.

Definition 4.2.10. *A stochastic one-step method (Ψ,h,ξ) is called stochastically C-stable if there exists a constant $C > 0$ and a parameter $\eta \in (1,\infty)$ such that for all $t, t+h \in [0,T],\ h > 0$ and all random variables $X_t^{i,N}, Z_t^{i,N} \in L^2\left(\Omega,\mathcal{F}_t,\mathbb{P};\mathbb{R}^d\right),\ i \in [\![1,N]\!]$, from identically distributed* *particle systems with their empirical measures* $\mu_t^{X,N},\ \mu_t^{Z,N} \in \mathcal{P}_2(\mathbb{R}^d)$, *it holds*

$$\begin{aligned}\mathbb{E}\Big[\Big|\mathbb{E}\big[\Psi_i(X_t^{i,N},\mu_t^{X,N},t,h) - \Psi_i(Z_t^{i,N},\mu_t^{Z,N},t,h) \mid \mathcal{F}_t\big]\Big|^2\Big] \\ + \eta\mathbb{E}\Big[\Big|\big(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_t]\big)\left(\Psi_i(X_t^{i,N},\mu_t^{X,N},t,h) - \Psi_i(Z_t^{i,N},\mu_t^{Z,N},t,h)\right)\Big|^2\Big] \\ \leq (1+Ch)\,\mathbb{E}\big[|X_t^{i,N} - Z_t^{i,N}|^2\big] + Ch\big(W^{(2)}(\mu_t^{X,N},\mu_t^{Z,N})\big)^2.\end{aligned}$$

Here, and in what follows we denote by $(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_t])\,Y = Y - \mathbb{E}[Y \mid \mathcal{F}_t]$ the projection of an $\mathcal{F}_{t+h}$-measurable random variables Y orthogonal to the conditional expectation $\mathbb{E}[\cdot \mid \mathcal{F}_t]$.

Definition 4.2.11. *Let* $X^{i,N},\ i \in [\![1,N]\!]$, *be the unique strong solution to* (4.7), *with* $\mu^{X,N}$ *being the corresponding empirical measure. A stochastic one-step method* (Ψ,h,ξ) *is called stochastically* B*-consistent of order* $\gamma > 0$ *if there exists a constant* $C > 0$ *such that for all* $t, t+h \in [0,T],\ h \in (0,1)$, *it holds*

$$\begin{aligned}\mathbb{E}\Big[\Big|\mathbb{E}\big[X_{t+h}^{i,N} - \Psi_i(X_t^{i,N},\mu_t^{X,N},t,h) \mid \mathcal{F}_t\big]\Big|^2\Big] &\leq Ch^{2\gamma+2}, \\ \mathbb{E}\Big[\Big|\big(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_t]\big)\left(X_{t+h}^{i,N} - \Psi_i(X_t^{i,N},\mu_t^{X,N},t,h)\right)\Big|^2\Big] &\leq Ch^{2\gamma+1}.\end{aligned}$$

Next, we show the convergence results based on the definitions above.

Lemma 4.2.12. *Let* (Ψ,h,ξ) *be a stochastically* C*-stable one-step method with some* $\eta \in (1,\infty)$. *For the particle system* $X^{i,N}$, *given by* (4.7) *with its empirical distribution* $\mu^{X,N}$, *we have*

$$\begin{aligned}\sup_{n\in[\![0,M]\!]}\sup_{i\in[\![1,N]\!]}\mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big] \leq e^{CT}\Bigg[\mathbb{E}\big[|X_0^{i,N} - \xi^i|^2\big] \\ + \sum_{k=1}^{M}\sup_{i\in[\![1,N]\!]}\Bigg((1+h^{-1})\mathbb{E}\Big[\Big|\mathbb{E}\big[X_k^{i,N} - \Psi_i(X_{k-1}^{i,N},\mu_{k-1}^{X,N},t_{k-1},h) \mid \mathcal{F}_{t_{k-1}}\big]\Big|^2\Big] \\ + C_\eta\ \mathbb{E}\Big[\Big|\ \big(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_{t_{k-1}}]\big)\left(X_k^{i,N} - \Psi_i(X_{k-1}^{i,N},\mu_{k-1}^{X,N},t_{k-1},h)\right)\Big|^2\Big]\Bigg)\Bigg]\end{aligned}$$

where $C_\eta = 1 + (\eta-1)^{-1}$ *and* $\hat{X}_n^{i,N}$ *denotes the particles generated by* (Ψ,h,ξ), *with* $X_k^{i,N} = X_{t_k}^{i,N}$, $\mu_k^{X,N} = \mu_{t_k}^{X,N}$, $t_k = kh$ *for all* $k \in [\![0,M]\!]$.

Theorem 4.2.13. *Let the stochastic one-step method* (Ψ,h,ξ) *be stochastically* C*-stable and stochastically* B*-consistent of order* $\gamma > 0$. *If* $\xi^i = X_0^{i,N} = \hat{X}_0^{i,N}$, *then there exists a constant* C *independent of* N, h *such that*

$$\sup_{n\in[\![0,M]\!]}\sup_{i\in[\![1,N]\!]}\mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big] \leq Ch^{2\gamma},$$

where $X^{i,N}$ *denotes the exact solution to* (4.7) *and* $\hat{X}^{i,N}$ *is the particle generated by* (Ψ,h,ξ). *In particular,* (Ψ,h,ξ) *is strongly convergent of order* γ.

4.2.5 The numerical scheme

The split-step method (SSM) proposed here follows the steps in Chapter 3 and will be re-casted accordingly. The critical difficulty arises from the simultaneous appearance of the convolution component in v (4.1) and the super-linear diffusion coefficient. The presence of both nonlinearities is the main hindrance to proving moment bounds of order $p > 2$ for the numerical scheme. Therefore, we rely on the C-stability and B-consistency methodology, as this approach does not require proving moment stability of higher order for the numerical scheme. This is in stark contrast to the techniques used in Chapter 3, where the time-stepping scheme has stable moments of higher order (depending on the regularity of the initial data) and strong convergence rates are proven without employing the C-stability and B-consistency procedure.

Here, we wish to emphasize that even with the symmetry condition it is unclear how to prove L^p-moment bounds of the numerical scheme for $p > 2$.

Definition 4.2.14 (Definition of the SSM)**.** *Let Assumption 4.2.1 hold, let h satisfy* (4.14) *and let $M \in \mathbb{N}$ such that $Mh = T$. Define recursively the SSM approximating of* (4.7) *as: set $\hat{X}_0^{i,N} = X_0^i$, for $i \in [\![1, N]\!]$; for $n \in [\![0, M-1]\!]$ and $i \in [\![1, N]\!]$ (recall Lemma 4.2.7), $t_n = nh$, we have with $\Delta W_n^i = W_{t_{n+1}}^i - W_{t_n}^i$*

$$Y_n^{\star,N} = \hat{X}_n^N + hV(Y_n^{\star,N}), \quad \hat{X}_n^N = (\ldots, \hat{X}_n^{i,N}, \ldots), \quad Y_n^{\star,N} = (\ldots, Y_n^{i,\star,N}, \ldots), \tag{4.11}$$

$$\text{where } \ Y_n^{i,\star,N} = \hat{X}_n^{i,N} + hv(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}), \qquad \hat{\mu}_n^{Y,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{Y_n^{j,\star,N}}(\mathrm{d}x), \tag{4.12}$$

$$\hat{X}_{n+1}^{i,N} = Y_n^{i,\star,N} + b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})h + \overline{\sigma}(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Delta W_n^i. \tag{4.13}$$

The stepsize h satisfies (this constraint is soft, see Remark 3.2.7 for details)

$$h \in \Big(0, \min\big\{1, \tfrac{1}{\zeta}\big\}\Big) \text{ where } \zeta = \max\Big\{2(L_{(f)}^{(1)} + L_{(u\sigma)}^{(1)}),\ 2(2L_{(f)}^{(1),+} + L_{(u\sigma)}^{(1)} + L_{(u\sigma)}^{(2)}),\ 0\Big\}. \tag{4.14}$$

It is immediate to see that (4.11) or (4.12) are implicit equations (given $\hat{X}_n^N$). The solvability of $Y_n^{\star,N}$ as a unique implicit map of the input $\hat{X}_n^N$ is addressed in Remark 4.2.16 below. The choice of h is discussed next.

Remark 4.2.15 (Choice of h)**.** *Let Assumption 4.2.1 hold (the constraint on h in* (4.14) *comes from* (4.38), (4.42), (4.43) *and* (4.34) *below) and following the notation of these inequalities, under* (4.14) *with $\zeta > 0$, there exists $\lambda \in (0, 1)$ such that $h < \lambda/\zeta$ and*

$$\max\left\{\frac{1}{1 - 2(L_{(f)}^{(1)} + L_{(u\sigma)}^{(1)})h},\ \frac{1}{1 - 2(2L_{(f)}^{(1),+} + L_{(u\sigma)}^{(1)} + L_{(u\sigma)}^{(2)})h}\right\} < \frac{1}{1-\lambda}.$$

For $\zeta = 0$, the result is trivial and we conclude that there exists a constant C independent of h such that

$$\max\left\{\frac{1}{1 - 2(L_{(f)}^{(1)} + L_{(u\sigma)}^{(1)})h},\ \frac{1}{1 - 2(2L_{(f)}^{(1),+} + L_{(u\sigma)}^{(1)} + L_{(u\sigma)}^{(2)})h}\right\} \le 1 + Ch.$$

As argued in Remark 3.2.7, the constraint on h can be lifted.

Remark 4.2.16 (Solvability of the implicit equation (4.12))**.** *Recall that the function V (defined in Lemma 4.2.7) satisfies a one-sided Lipschitz condition in $\mathbb{R}^{Nd}$, and hence (under* (4.14)*) a unique solution $Y_n^{\star,N}$ to* (4.11) *as a function of $\hat{X}_n^N$ exists.*

This result follows a well-known argument using results on strongly monotone operators [142, Theorem 26.A (p.557)] and is fully detailed in Lemma 2.4.1. The result is a straightforward adaptation to the $\mathbb{R}^{Nd}$ case (see proof section).

After introducing the discrete scheme, we introduce its continuous extension and the main convergence results.

Definition 4.2.17 (Continuous extension of the SSM)**.** *Under the same choice of h and assumptions in Definition 4.2.14, for all $t \in [t_n, t_{n+1}]$, $n \in [\![0, M-1]\!]$, $t_n = nh$, $i \in [\![1, N]\!]$, $\hat{X}_0^{i,N} = X_0^i$, for X_0^i in* (4.7)*, the continuous extension of the SSM is*

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

$$\text{where} \qquad \hat{\mu}_n^{Y,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{Y_n^{j,\star,N}}(\mathrm{d}x), \qquad \hat{\mu}_{t_n}^{Y,N} = \hat{\mu}_n^{Y,N},$$

and $\kappa(t) = \sup\big\{t_n : t_n \le t,\ n \in [\![0, M-1]\!]\big\}$.

Theorem 4.2.18 (Convergence of the SSM). *Let the assumptions of Theorem 4.2.8 hold. Assume additionally that $\eta > 1$ and $m > 4q+4 > \max\{2(q+1),4\}$, with $X_0^i \in L^m(\mathbb{R}^d)$ and q as defined in Assumption 4.2.1. Choose h as in* (4.14)*. Then for the SSM scheme defined in* (4.11)-(4.13)*, we have the following properties.*

1. *The SSM is C-stable;*
2. *The SSM is B-consistent with $\gamma = 1/2$ in Definition 4.2.11;*
3. *For $i \in [\![1,N]\!]$, let $X^{i,N}$ be the solution to* (4.7)*, then there exists a constant $C > 0$ (independent of N and h) such that*

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

Lastly, we present a result about long time stability of the numerical scheme proposed as means to access the invariant distribution of the original MV-SDE by way of simulation. In other words, we provide sufficient conditions for our scheme to be ergodic as $T \to \infty$.

Theorem 4.2.19 (Ergodicity of the SSM). *Let the Assumptions of Theorem 4.2.18 holds. Suppose that $X_0 \in L^m(\mathbb{R}^d)$ and $Z_0 \in L^m(\mathbb{R}^d)$ for $m > 4q+4$ as in Theorem 4.2.18, and let $\hat{X}_0^{i,N}$ and $\hat{Z}_0^{i,N}$ be i.i.d. copies of X_0 and Z_0 respectively, for all $i \in [\![1,N]\!]$.*

Set $h > 0$. For $i \in [\![1,N]\!]$ and $n \in [\![1,M]\!]$, define $(\hat{X}_n^{i,N}, Y_n^{i,X,N})$ and $(\hat{Z}_n^{i,N}, Y_n^{i,Z,N})$ as the output of the SSM (4.12)-(4.13) *(i.e., $\star = X, Z$) corresponding to the empirical measure pairs $(\hat{\mu}_n^{X,N}, \hat{\mu}_n^{Y,X,N})$ and $(\hat{\mu}_n^{Z,N}, \hat{\mu}_n^{Y,Z,N})$ with initial conditions $X_0^{i,N}$ and $Z_0^{i,N}$ respectively. Then, for any $n \in [\![1,M]\!]$,*

$$\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[\,|\hat{X}_n^{i,N} - \hat{Z}_n^{i,N}|^2\big] \leq (1+\beta h)^n \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{X}_0^{i,N} - \hat{Z}_0^{i,N}|^2\big],$$

where we recall the parameters in Assumption 4.2.1

$$\beta = \frac{\rho_1 + 2L_{(b)}^{(1)}h}{1 - h(4L_{(f)}^{(1),+} + 2L_{(u\sigma)}^{(1)} + 2L_{(u\sigma)}^{(2)})}, \quad \rho_1 = 4L_{(f)}^{(1),+} + 2L_{(u\sigma)}^{(1)} + 2L_{(u\sigma)}^{(2)} + 2L_{(b)}^{(2)} + 2L_{(b)}^{(3)}.$$

Under the choice of h stated in Theorem 4.2.18, the quantity $1+\beta h$ is always positive. If $\rho_1 < 0$ and h sufficiently small then $\beta < 0$.

4.3 Examples of interest

We illustrate the performance of the SSM on several numerical examples. As the "true" solution of the considered models is unknown, the convergence rates for these examples are calculated in reference to a proxy solution given by an approximation at a smaller timestep h. The strong error between the proxy-true solution X_T and approximation $\hat{X}_T$ is as follows

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

We also consider the path type strong error as follows

$$\text{Strong error (path)} = \Big(\mathbb{E}\big[\sup_{t\in[0,T]} |X_t - \hat{X}_t|^2\big]\Big)^{\frac{1}{2}} \approx \Big(\frac{1}{N}\sum_{j=1}^{N} \sup_{n\in[\![0,M]\!]} |X_n^j - \hat{X}_n^j|^2\Big)^{\frac{1}{2}}.$$

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, is measured by

$$\text{Propagation of chaos Error (PoC-Error)} \approx \Big(\frac{1}{N_l}\sum_{j=1}^{N_l}|\hat{X}_T^{j,N_l} - \hat{X}_T^{j,N_{l+1}}|^2\Big)^{\frac{1}{2}}. \tag{4.15}$$

Above $N_{l+1} = 2N_l$ and the first half of the N_{l+1} particles use the same Brownian motions as the whole N_l particle system. In this section, the rMSE takes $h \in \{10^{-1}, 5\times 10^{-2}, 2\times 10^{-2}, 10^{-2}, 5\times 10^{-3}, 2\times 10^{-3}, 10^{-3}\}$ with $N = 1000$, the proxy solution takes $h = 10^{-4}$. The PoC takes $N \in \{40, 80, 160, 320, 640, 1280\}$ with $h = 10^{-3}$, the proxy solution takes $N = 2560$.

Remark 4.3.1 ('Taming' algorithm)**.** *For comparative purposes, we implement the 'Taming' algorithm [47, 60] – any convergence analysis of the taming algorithm in the framework of this manuscript is an open question. Of the many possible taming variants, we implement the following two cases: taming f (and similarly f_σ) inside the convolution term ('Taming-in') and taming the convolution itself ('Taming-out'). Concretely, set $Mh = T$, then f is replaced by (for $\alpha \in (0,1]$)*

- *'Taming-out': $\int_{\mathbb{R}^d} f(\cdot - y)\mu(\mathrm{d}y)$ is replaced by $\int_{\mathbb{R}^d} f(\cdot - y)\mu(\mathrm{d}y)/\big(\ 1 + M^\alpha|\int_{\mathbb{R}^d} f(\cdot - y)\mu(\mathrm{d}y)|\ \big)$.*
- *'Taming-in': f is replaced by $f/\big(1 + M^\alpha|f|\big)$.*

Note that the proxy solution for the SSM is computed using the SSM and analogously for the taming schemes. For each example, the error rates of Taming and SSM are computed using the same Brownian motion paths and same initial data. To avoid confusion later in the numerical results, we clarify that due to the super-linear convolution kernel, we do not expect the Taming method to converge. However, under mild initial conditions, it is rare to observe the divergence, so we test high variance cases to show the Taming method does not work in general while the SSM works as expected. We remark that the first step (4.11) of the SSM requires to solve an implicit equation in $\mathbb{R}^{Nd}$, which is done employing Newton's method (see Section 6.2.2 for details).

Below, the symbols $\mathcal{N}(\alpha,\beta)$ denote the normal distribution with mean $\alpha \in \mathbb{R}$ and variance $\beta \in (0,\infty)$, the symbol $U(a,b)$ denotes the uniform distribution over $[a,b]$ for $-\infty < a < b < \infty$, the symbol $B(c,p)$ denotes the binomial distribution for random variables X such that $X = 0$ with probability p and $X = c$ with probability $1-p$.

4.3.1 Example: Symmetric double-well type model

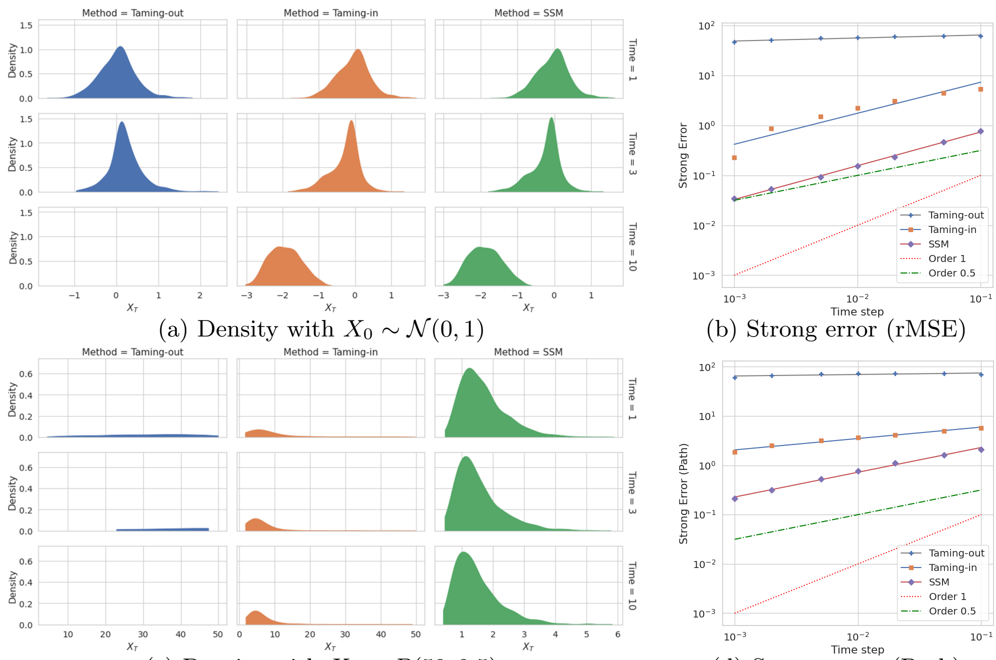
We consider an extension to the symmetric double-well model [138] of confinement type with extra super-linearity [128, Section 5] in the diffusion coefficient,

$$\mathrm{d}X_t = \big(v(X_t,\mu_t^X) + X_t\big)\mathrm{d}t + (X_t + \tfrac{1}{4}X_t^2)\mathrm{d}W_t,\ v(x,\mu) = -\tfrac{1}{4}x^3 + \int_{\mathbb{R}} -\big(x-y\big)^3\mu(\mathrm{d}y). \tag{4.16}$$

The corresponding Fokker-Planck equation is $\partial_t\rho = \nabla[\ \nabla\frac{\rho}{2}|x + \frac{1}{4}x^2|^2 + \rho\nabla V + \rho\nabla W * \rho]$ with $W = \frac{1}{4}|x|^4$, $V = \frac{1}{16}|x|^4 - \frac{1}{2}|x|^2$, and ρ is the corresponding density map. Due to the structure of the drift term, we expect three cluster states around $x \in \{-2, 0, 2\}$.

The goal of this example is to simulate the interacting particle system associated to (4.16) up to $T = 10$ using the three numerical methods available. Figure 4.1 (a) and (c) show the evolution of the density map at $T \in \{1, 3, 10\}$. In (a) with $X_0 \sim \mathcal{N}(0,1)$, all three methods yield similar results, but (c) shows that with $X_0 \sim B(50, 0.5)$, Taming-out (blue, left) and Taming-in fail to produce acceptable results, while the SSM produces the expected results.

Figure 4.1 (b) shows the strong convergence of the methods, Taming-out failed to converge. Taming-in and the SSM converge under all time step choices (all satisfying (4.14)) and nearly attain the 1/2 strong error rate, the error of SSM is one order of magnitude smaller than the error of Taming-in. Figure 4.1 (d) shows the path type strong convergence of both methods, and we observe that Taming-out and Taming-in failed to converge or at least converge with a



(a) Density with $X_0 \sim \mathcal{N}(0, 1)$ (b) Strong error (rMSE)

(c) Density with $X_0 \sim B(50, 0.5)$ (d) Strong error (Path)

Figure 4.1: Simulation of the double-well model (4.16) with $N = 1000$ particles. All schemes are initialized on the exact same samples. (a) and (c) show the density map for Taming-out (left), Taming-in (middle) and SSM (right) with $h = 0.01$ at times $T \in \{1, 3, 10\}$ seen top-to-bottom and with different initial distribution. (b) Strong error (rMSE) of SSM and Taming with $X_0 \sim \mathcal{N}(3, 9)$ in log-scale. (d) Strong error (Path) of SSM and Taming with $X_0 \sim \mathcal{N}(3, 9)$ in log-scale.

very low rate. The SSM converges under all time step choices but the errors are one order of magnitude greater than the standard strong error.

As mentioned earlier, we do not have any theoretical support for the convergence of the taming methods. This example shows that a convergence proof for Taming-in might be feasible, possibly, under the caveat of an additional condition on the distribution/support of the initial condition – this was fully unforeseen. These results for Taming-out are discouraging, nonetheless, under strong dissipativity Taming-out seems stable (see next example).

4.3.2 Example: Approximating the possible invariant distribution

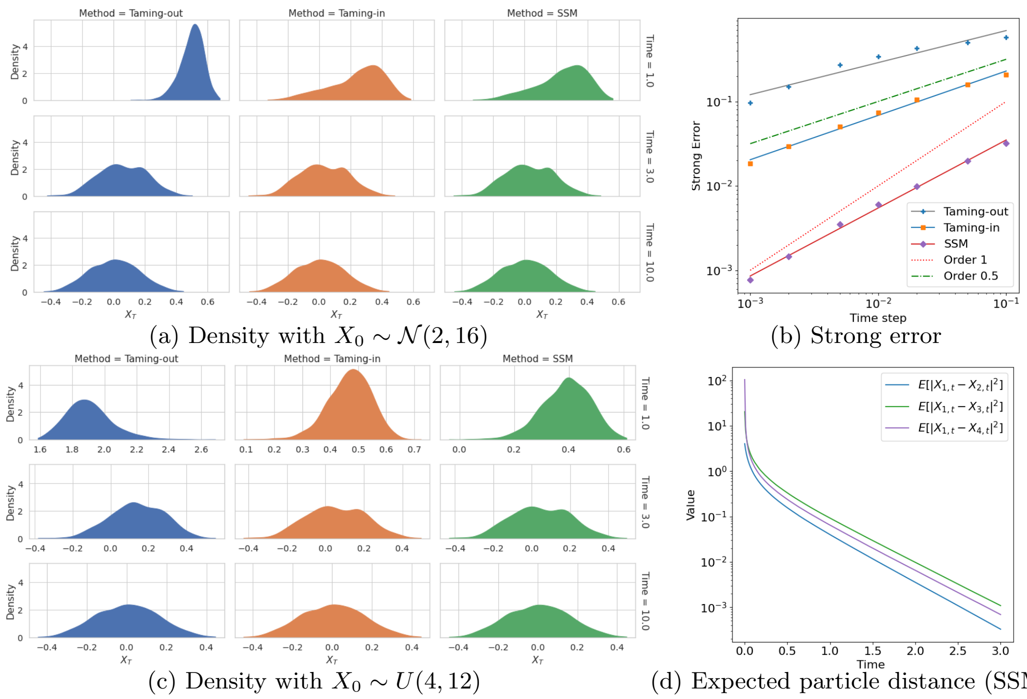
This example aims to illustrate the long-time simulation for the purpose of approximating the invariant distribution of the system

$$\mathrm{d}X_t = \big(v(X_t, \mu_t^X) - X_t\big)\mathrm{d}t + \tfrac{1}{4}(1 - X_t^2)\mathrm{d}W_t,\ v(x, \mu) = -x^3 + \int_{\mathbb{R}} -\big(x - y\big)^3 \mu(\mathrm{d}y). \tag{4.17}$$

The corresponding Fokker-Planck equation is $\partial_t \rho = \nabla[\ \nabla \frac{\rho}{32}|1 - x^2|^2 + \rho\nabla V + \rho\nabla W * \rho]$ with $W = \frac{1}{4}|x|^4$, $V = \frac{1}{4}|x|^4 + \frac{1}{2}|x|^2$, and ρ is the corresponding density map. Here, the cluster state is $x = 0$.

Figure 4.2 (a) and (c) show the evolution of the particle distribution under different initial conditions. All three methods produce similar outputs at $T \in \{3,\ 10\}$, with Taming-out taking longer to contract and to converge than the other methods under $X_0 \sim \mathcal{N}(2, 16)$ in (a) and $X_0 \sim U(4, 12)$ in (c). The similar results obtained at $T \in \{3, 10\}$ are due to the fact that the model (4.17) has an invariant distribution and the initial distribution is compactly supported around the cluster state $x = 0$.

Figure 4.2 (b) illustrates the strong convergence of the three methods: they all converge and the rates are of order close to 1/2, the SSM outperforms the other two methods by 1 to 2 orders



(a) Density with $X_0 \sim \mathcal{N}(2,16)$ (b) Strong error

(c) Density with $X_0 \sim U(4,12)$ (d) Expected particle distance (SSM)

Figure 4.2: Approximation of the invariant distribution of (4.17) with $N = 1000$ particles. The simulated Brownian motion paths and initial distribution are the same for all schemes. (a) and (c) show the distribution for Taming-out (left), Taming-in (middle) and SSM (right) with $h = 0.01$ at times $T \in \{1, 3, 10\}$ seen top-to-bottom and with different initial distribution; x- and y-scales are fixed. (b) Strong error (rMSE) of SSM and Taming with $X_0 \sim \mathcal{N}(2,16)$. (d) Expected distance (in log-scale) between particles under different initial distributions with $h = 10^{-3}$ for the SSM.

of magnitude. Figure 4.2 (d) depicts the expected exponential decay rate for the SSM under different initial conditions: $X_{1,0} \sim \mathcal{N}(0,1)$, $X_{2,0} \sim U(-3,3)$, $X_{3,0} \sim \mathcal{N}(2,16)$, $X_{4,0} \sim \mathcal{N}(2,100)$ (same Brownian motion samples).

4.3.3 Example: Kinetic 2d Van der Pol oscillator and periodic phase-space

We consider a two-dimensional Van der Pol (VdP) oscillator model with added super-linearity terms. The VdP model was proposed to describe stable oscillation [88, Section 4.2 and 4.3] and for a system of many coupled oscillators in the presence of noise the limit model is a MV-SDE [1]. Here, we build a two-dimensional VdP-type model with mean-field components and super-diffusivity that features a periodicity of phase-space to show that the SSM preserves the theoretical periodic behaviour in simulation scenarios – see [32, Section 7.3].

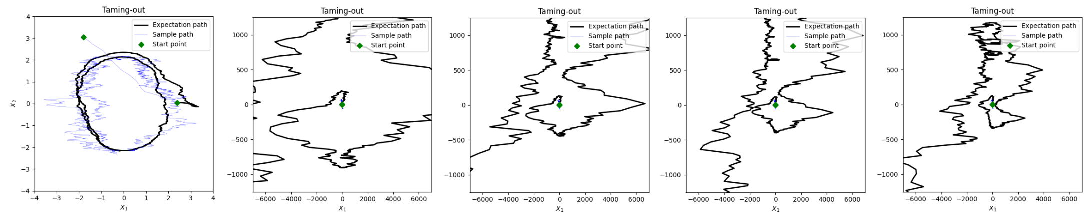
Set $x = (x_1, x_2) \in \mathbb{R}^2$ and define the functions f, u, b, σ as

$$f(x) = -x|x|^2, \ u(x) = \begin{bmatrix} -\frac{1}{3}x_1^3 \\ 0 \end{bmatrix}, \ b(x) = \begin{bmatrix} x_1 - x_2 \\ x_1 \end{bmatrix}, \ \sigma(x) = \begin{bmatrix} 1 + 1/4\ x_1^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad (4.18)$$

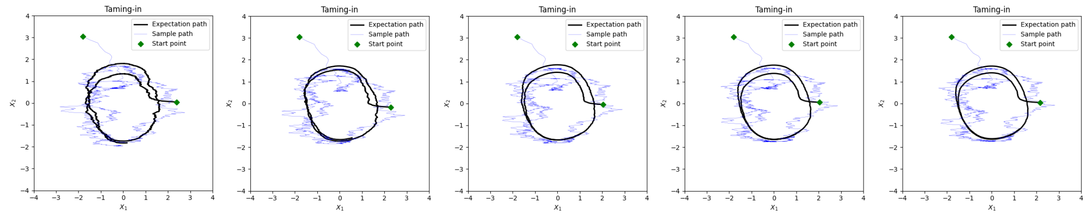
where f satisfies $(\mathbf{A}^f)$.

Figure 4.3 (a)-(o) show the system's phase-space portraits (i.e., the parametric plot of $t \mapsto (X_{1,t}, X_{2,t})$ and $t \mapsto (\mathbb{E}[X_{1,t}], \mathbb{E}[X_{2,t}])$) for the three methods with different choices of N.

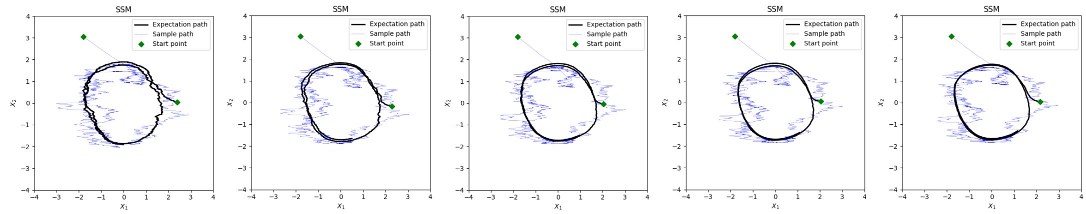
In the first row of Figure 4.3, (a)-(e) shows the result of the Taming-out method, the system fails to converge for $N > 50$. The second row and third row of Figure 4.3 show the result of Taming-in and the SSM, both methods converge and the trajectory becomes smoother as more particles are taken. However, there is a big difference on the expectation trajectories of the SSM and Taming in, the expectation trajectories of the SSM do not cross themselves while the expectation trajectories of Taming-in always cross themselves, which is not expected since the



(a) $N = 50$ (b) $N = 200$ (c) $N = 500$ (d) $N = 1000$ (e) $N = 2000$



(f) $N = 50$ (g) $N = 200$ (h) $N = 500$ (i) $N = 1000$ (j) $N = 2000$



(k) $N = 50$ (l) $N = 200$ (m) $N = 500$ (n) $N = 1000$ (o) $N = 2000$

Figure 4.3: Simulation of the Vdp model (4.18) with a different number of particles and $h = 10^{-2}$, $T = 12$, $X_{1,0} \sim \mathcal{N}(2, 16), X_{2,0} \sim \mathcal{N}(0, 16)$. (a)(b)(c)(d)(e) are phase portraits of the Taming-out method with different choices of N. (f)(g)(h)(i)(j) are phase portraits of the Taming-in method with different choices of N. (k)(l)(m)(n)(o) are phase portraits of the SSM with different choices of N.

slope fields of the VdP model are smooth and do not admit the cross. Moreover, comparing the first few steps in the sample paths, the particles generated by the SSM concentrate to the expectation path within two steps while the one generated by Taming-in takes about 10 steps. This is because the SSM preserves the super-linear power from the convolution kernel while the Taming-in turns this power to an asymptotic linear one. Thus, the SSM preserves more geometric properties than the taming method even though the approximation obtained via taming may not blow up.

4.3.4 Example: Super-linear growth of measure components in diffusion

This example illustrates the effect of two additional types of measure-nonlinearities included in the diffusion term; Case 1 corresponds to a convolution term in the diffusion and Case 2 is a variance-type term (which is beyond the scope of the thesis). Note that the assumptions of the wellposedness result are not satisfied as the estimate (4.3) does not hold (but could readily be achieved by slightly modifying the constants of the coefficients), which indicates that this bound is not sharp. We consider

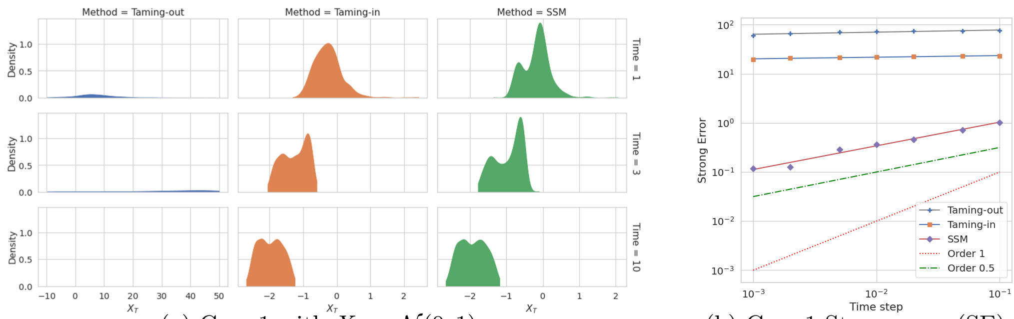
$$\mathrm{d}X_t = \big(v(X_t, \mu_t^X) + X_t\big)\mathrm{d}t + \big(X_t + \tfrac{1}{4}X_t^2 + f_\sigma(X_t, \mu_t^X)\big)\mathrm{d}W_t, \tag{4.19}$$

$$\textit{with } v(x, \mu) = -\tfrac{1}{4}x^3 + \int_{\mathbb{R}} -(x-y)^3 \mu(\mathrm{d}y),$$

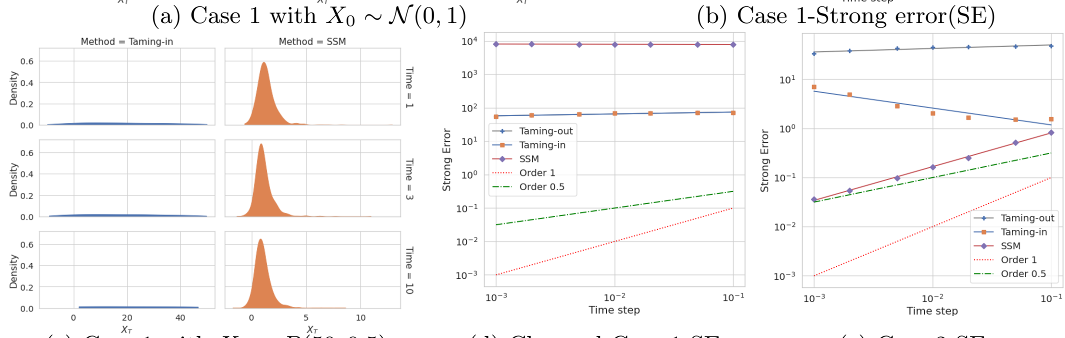
$$f_\sigma(x, \mu) = \begin{cases} \int_{\mathbb{R}} (x-y)^2 \mu(\mathrm{d}y), \textit{ Case 1}, \\ \int_{\mathbb{R}} \int_{\mathbb{R}} (y-z)^2 \mu(\mathrm{d}y)\mu(\mathrm{d}z), \textit{ Case 2}. \end{cases}$$

For Case 1, we have a nonlinear convolution kernel $f_\sigma(x) = x^2$ for all $x \in \mathbb{R}$. Figure 4.4, in particular, subplots (a)-(c), illustrates that the SSM converges, in a pointwise sense, with strong order 1/2 and recovers reasonable density estimates for different choices of the initial distribution. Similar behaviour is not observed for different taming approaches which fail to recover the anticipated strong convergence order of 1/2 and we observe that taming schemes do not capture the density of the solution well for high-variance initial data. We conducted an analogous test with $v(x, \mu) = -x^3/4$ in (d), i.e., we removed the convolution term in the drift, and our experiments failed, in the sense that the approximate solutions computed by the SSM did not converge. This supports our theoretical results that a suitable drift compensation for the nonlinear measure component appearing in the diffusion is indeed needed.

Case 2 corresponds to an example, where the convolution term is again integrated, i.e., resembles a variance-type term. We are not aware of an existing result that yields wellposedness of the underlying MV-SDE including such a term (even without the nonlinear convolution terms). Further, it is not clear which assumptions would be required for a numerical scheme to converge in a strong sense. The expected strong convergence order is observed for the SSM in (e), but no taming approach appears to be a reasonable alternative. We additionally conducted a numerical experiment for Case 2 with $v(x, \mu) = -x^3/4$, in order to investigate if the variance-type term requires a compensation term (similar to changed Case 1). We also observed that no time-stepping scheme (i.e., taming and SSM) seemed to converge (the result is similar to (d) and we do not present here), which again indicates that the drift's convolution term can also help to control variance-type terms in the diffusion.



(a) Case 1 with $X_0 \sim \mathcal{N}(0, 1)$ (b) Case 1-Strong error(SE)



(c) Case 1 with $X_0 \sim B(50, 0.5)$ (d) Changed Case 1-SE (e) Case 2-SE

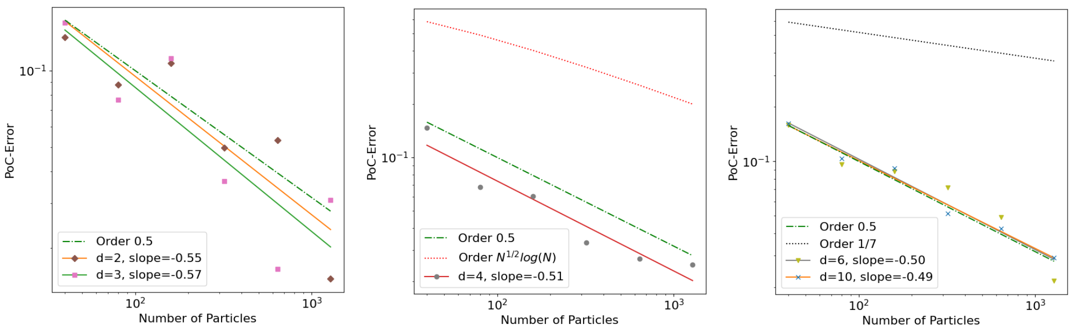
Figure 4.4: Approximation of (4.19) with $N = 1000$ particles. The simulated Brownian motion sample paths and initial distribution are the same for all schemes. (a) and (c) show the distribution for Taming-out (left), Taming-in (middle) and SSM (right) with $h = 0.01$ at times $T \in \{1, 3, 10\}$ seen top-to-bottom and with different initial distribution; x- and y-scales are fixed. (b), (d) and (e) show the strong error (rMSE) of SSM and Taming with $X_0 \sim \mathcal{N}(1, 1)$ for different cases. The changed Case 1 in (d) is Case 1 with $v(x, \mu) = -x^3/4$.

4.3.5 Example: Propagation of Chaos rate across dimensions

In this example, we estimate the PoC rate depending on the dimension and compare the findings to the theoretical upper bounds established in Theorem 4.2.8. For equation (4.1)-(4.2) we make the following choices: Let $d \geq 2$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, the initial condition X_0 is a vector distributed according to d-independent $\mathcal{N}(1,1)$-random variables, and

$$f(x) = -x|x|^2, \quad u(x) = -\frac{1}{3}\left[\ x_1^3, x_2^3, \cdots, x_d^3\ \right]^\intercal, \quad b(t,x,\mu) = x,$$
$$\overline{\sigma}(x) = \begin{bmatrix} x_1 + 1/4\ x_1^2 & x_2 & \cdots & x_d \\ x_1 & x_2 + 1/4\ x_2^2 & \cdots & x_d \\ \cdots & \cdots & \cdots & \cdots \\ x_1 & x_2 & \cdots & x_d + 1/4\ x_d^2 \end{bmatrix}. \tag{4.20}$$

This is a toy model with a high-dimensional fully coupled convolution kernel and super-linear diffusion term. We observe in Figure 4.5 a strong PoC rate, estimated via (4.15), of order of roughly $1/2$ across dimension d. By the ordinary least squares linear regression, for dimension $d \in \{2,3,4,6,10\}$, the corresponding slopes are $\{\text{slopes}_d\}_d = \{-0.55, -0.57, -0.5, -0.50, -0.49\}$ and the corresponding R-square measure is $\{R_d^2\}_d = \{0.81, 0.75, 0.92, 0.91, 0.98\}$.



(a) PoC rates for $d \in \{2,3\}$ (b) PoC rates for $d = 4$ (c) PoC rates for $d \in \{6,10\}$

Figure 4.5: Estimation of PoC rate for equation (4.1)-(4.2) under (4.20) using SSM (4.11)-(4.13) with fixed stepsize $h = 10^{-3}$, $T = 1$ and number of particles $N \in \{40, 80, 160, 320, 640, 1280, 2560\}$. In all figures the reference rate 0.5 and the upper bound rate from Theorem 4.2.8 are displayed.

These findings are inline with those obtained in the one-dimensional example of [125, Example 4.1]. Theorem 4.2.8 establishes a strong convergence rate (in terms of number of particles in a pathwise sense) of order $1/4$ for dimensions $d < 4$ only and these results are smaller than the upper bounds of PoC in Theorem 4.2.8 – this highlights a gap in the literature to be explored in future research. For perspective, at a theoretical level the rate $1/2$ in N is not new under stronger assumptions. This was obtained in [57, Lemma 5.1] or [134] when the drift and diffusion coefficients are assumed to satisfy strong regularity assumptions. Also in [118] for linear type MV-SDEs featuring diffusions $\mathbb{R}^d \ni x \mapsto \overline{\sigma}(x)$ and drifts with structure of the type $\mathbb{R}^d \ni x \mapsto \int_{\mathbb{R}^d} b(x,y)\mu(\mathrm{d}y)$, and requiring that $b, \overline{\sigma}$ are uniformly Lipschitz, the convergence rate $1/2$ in the number of particles is obtained; also in [58].

4.3.6 Discussion

We discuss the advantages of the SSM compared with the taming methods. The SSM converges under all cases, while the two types of taming failed to converge in some cases. The SSM requires an implicit solver for the convolution kernel but the running time of the SSM compared to the taming methods is only 2 to 3 times longer. From the numerical examples, we see that:

1. The two types of strong errors of the SSM are of order 0.5 and consistently outperform that of the proposed taming schemes. In fact, the taming methods are not even expected to converge, however, under a mild initial condition, it is hard to observe the divergence. In the tests with high variance initial distributions, the taming methods diverge while SSM converges consistently. The SSM preserves convergence for larger time steps h (via comparative lower errors) and is also suitable for long-time simulation.

2. The SSM preserves important geometric properties (the concentration speed of the particles is fast, the expected trajectory coincides with the vector field result), while the taming methods appear to fail to capture these crucial properties.

3. We applied the SSM to examples, where the diffusion also involves certain nonlinear measure terms. As long as a suitable monotonicity condition is satisfied the SSM yields promising results.

4. We perform a PoC rate test across dimensions with non-trivial convolution kernel. The rate which we observe numerically is better than the one suggested by the PoC results.

4.4 Proof of the main results

4.4.1 Proof of Theorem 4.2.5 : Wellposedness and moment stability

Proof of Theorem 4.2.5. The existence and uniqueness follow from modifications of the methodologies used in [2, Theorem 3.5].

Wellposedness. The proof for existence and uniqueness follows along the same lines as the arguments presented in [2, Theorem 3.5]. We repeat here the main steps for convenience. As opposed to more classical approaches, the fixed point argument is carried out over a suitable function space, see [19], instead of a measure space.

To be precise, one considers the function space $\Lambda_{[0,T],q}$, for q as in Assumption 4.2.1, defined as the space of continuous functions $\boldsymbol{g} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^{d\times l}$, $\boldsymbol{g}(t,x) = (g_1(t,x), g_2(t,x))$ with $g_1 : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $g_2 : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times l}$, satisfying

$$\|\boldsymbol{g}\|_{[0,T],q} := \sup_{t\in[0,T]} \left(\sup_{x\in\mathbb{R}^d} \frac{|\boldsymbol{g}(t,x)|}{1+|x|^{q+1}} \right) < \infty, \tag{4.21}$$

and there exists a constant $L_1 \geq 0$ such that (with m as in Assumption 4.2.1)

$$\langle x-y, g_1(t,x) - g_1(t,y)\rangle + 2(m-1)|g_2(t,x) - g_2(t,y)|^2 \leq L_1|x-y|^2, \tag{4.22}$$

for all $t \in [0,T]$, $x, y \in \mathbb{R}^d$. In particular, this implies that there exists a constant $L_2 \geq 0$ such that

$$\langle x, g_1(t,x)\rangle + 2(m-1)|g_2(t,x)|^2 \leq L_2(1 + |\boldsymbol{g}(t,0)|^2 + |x|^2).$$

For some $K > 0$ (chosen below), and a small enough terminal time T_0, we define now

$$E := \{\boldsymbol{g} \in \Lambda_{[0,T_0],q} : \|\boldsymbol{g}\|_{[0,T_0],q} \leq K\}.$$

We claim that there exist choices for T_0 and K such that $\Gamma : E \to E$ defined by

$$\Gamma[\boldsymbol{g}](t,x) = (\Gamma[\boldsymbol{g}]_1(t,x), \Gamma[\boldsymbol{g}]_2(t,x)) := (f * \mu_t^{\boldsymbol{g}}(x), f_\sigma * \mu_t^{\boldsymbol{g}}(x)),$$

forms a contraction. Here, $\mu^{\boldsymbol{g}}$ is the law of the solution to the MV-SDE

$$\mathrm{d}X_t^{\boldsymbol{g}} = \big(\overline{v}(t, X_t^{\boldsymbol{g}}, \mu_t^{\boldsymbol{g}}) + b(t, X_t^{\boldsymbol{g}}, \mu_t^{\boldsymbol{g}})\big)\mathrm{d}t + \overline{\sigma}(t, X_t^{\boldsymbol{g}}, \mu_t^{\boldsymbol{g}})\mathrm{d}W_t, \quad X_0^{\boldsymbol{g}} = X_0 \in L^m(\mathbb{R}^d) \tag{4.23}$$

$$\overline{v}(t,x,\mu) = g_1(t,x) + u(x,\mu), \quad \overline{\sigma}(t,x,\mu) = \sigma(t,x,\mu) + g_2(t,x). \tag{4.24}$$

The existence of a unique strong solution to $(X_t^{\boldsymbol{g}})_{t\in[0,T]}$ satisfying $\sup_{t\in[0,T]} \mathbb{E}[|X_t^{\boldsymbol{g}}|^m] \leq C$ for some constant $C > 0$, is shown in [99, 123].

We first show that there exist $0 < T_0 < T$ and K such that Γ indeed maps E onto itself. Let $\boldsymbol{g} \in E$. First, we observe that for all $x, y \in \mathbb{R}^d$, $t \in [0, T_0]$

$$\begin{aligned}
&\langle x-y, \Gamma[\boldsymbol{g}]_1(t,x) - \Gamma[\boldsymbol{g}]_1(t,y)\rangle + 2(m-1)|\Gamma[\boldsymbol{g}]_2(t,x) - \Gamma[\boldsymbol{g}]_2(t,y)|^2 \\
&\leq \int_{\mathbb{R}^d} \big(\langle x-y, f(x-u) - f(y-u)\rangle + 2(m-1)|f_\sigma(x-u) - f_\sigma(y-u)|^2\big)\, \mu_t^{\boldsymbol{g}}(\mathrm{d}u) \\
&\leq L_1|x-y|^2.
\end{aligned}$$

Further, we derive, using that $(X_t^{\boldsymbol{g}})_{t\in[0,T]}$ has finite moments of order $m > 2(q+1)$ that there exist constants $C > 0$ and $C(q, \mathbb{E}[|X_0|^{q+1}]) > 0$ (depending on the moment bounds of the initial data, q, and the model parameters) such that

$$\|\Gamma[\boldsymbol{g}]\|_{[0,T_0],q} \le \sup_{t\in[0,T_0]} \left(\sup_{x\in\mathbb{R}^d} \frac{|(f * \mu_t^{\boldsymbol{g}})(x)| + |(f_\sigma * \mu_t^{\boldsymbol{g}})(x)|}{1+|x|^{q+1}} \right) \le C\Big(1 + \sup_{t\in[0,T_0]} \mathbb{E}\big[|X_t^{\boldsymbol{g}}|^{q+1}\big]\Big) \tag{4.25}$$

$$\begin{aligned}
&\le C + Ce^{CT_0}\Bigg(\mathbb{E}[|X_0|^{q+1}] \\
&+ \int_0^{T_0} \big(|b(s,0,\delta_0)|^{q+1} + |g_1(s,0)|^{q+1} + |g_2(s,0)|^{q+1} + |u(0,\delta_0)|^{q+1} + |\sigma(s,0,\delta_0)|^{q+1}\big)\mathrm{d}s\Bigg) \\
&\le C + Ce^{CT_0}\Bigg(\mathbb{E}[|X_0|^{q+1}] + T_0\|\boldsymbol{g}\|_{[0,T_0],q}^{q+1} \\
&+ \int_0^{T_0} \big(|b(s,0,\delta_0)|^{q+1} + |u(0,\delta_0)|^{q+1} + |\sigma(s,0,\delta_0)|^{q+1}\big)\mathrm{d}s\Bigg) \\
&\le C + Ce^{CT_0}\Bigg(\mathbb{E}[|X_0|^{q+1}] + T_0K^{q+1} \\
&+ \int_0^{T_0} \big(|b(s,0,\delta_0)|^{q+1} + |u(0,\delta_0)|^{q+1} + |\sigma(s,0,\delta_0)|^{q+1}\big)\mathrm{d}s\Bigg) \le K,
\end{aligned}$$

for a sufficiently small $T_0 > 0$ and the choice $K = 2C(1 + e^{CT}\mathbb{E}[|X_0|^{q+1}])$. It remains to show that the map $\Gamma : E \to E$ forms a contraction, i.e., for any $\boldsymbol{g}_1 = (g_{1,1}, g_{1,2}), \boldsymbol{g}_2 = (g_{2,1}, g_{2,2}) \in E$, we have

$$\|\Gamma[\boldsymbol{g}_1] - \Gamma[\boldsymbol{g}_2]\|_{[0,T_0],q} \le c\|\boldsymbol{g}_1 - \boldsymbol{g}_2\|_{[0,T_0],q},$$

for $c \in (0,1)$ and a T_0 possibly even smaller than chosen above.

An application of Itô's formula shows for $t \in [0, T_0]$

$$\begin{aligned}
&\mathbb{E}[|X_t^{\boldsymbol{g}_1} - X_t^{\boldsymbol{g}_2}|^2] \le \mathbb{E}[|X_0^{\boldsymbol{g}_1} - X_0^{\boldsymbol{g}_2}|^2] + \int_0^t \mathbb{E}\Big[\big|\overline{\sigma}(s, X_s^{\boldsymbol{g}_1}, \mu_s^{\boldsymbol{g}_1}) - \overline{\sigma}(s, X_s^{\boldsymbol{g}_2}, \mu_s^{\boldsymbol{g}_2})\big|^2\Big]\mathrm{d}s \\
&+ 2\int_0^t \mathbb{E}\Big[\langle X_s^{\boldsymbol{g}_1} - X_s^{\boldsymbol{g}_2}, b(s, X_s^{\boldsymbol{g}_1}, \mu_s^{\boldsymbol{g}_1}) - b(s, X_s^{\boldsymbol{g}_2}, \mu_s^{\boldsymbol{g}_2})\rangle\Big]\mathrm{d}s \\
&+ 2\int_0^t \mathbb{E}\Big[\langle X_s^{\boldsymbol{g}_1} - X_s^{\boldsymbol{g}_2}, \overline{v}(s, X_s^{\boldsymbol{g}_1}, \mu_s^{\boldsymbol{g}_1}) - \overline{v}(s, X_s^{\boldsymbol{g}_2}, \mu_s^{\boldsymbol{g}_2})\rangle\Big]\mathrm{d}s \\
&\le \mathbb{E}[|X_0^{\boldsymbol{g}_1} - X_0^{\boldsymbol{g}_2}|^2] + \int_0^t C\mathbb{E}\Big[\big|X_s^{\boldsymbol{g}_1} - X_s^{\boldsymbol{g}_2}\big|^2\Big] + 2\mathbb{E}\Big[\big|g_{1,2}(s, X_s^{\boldsymbol{g}_1}) - g_{1,2}(s, X_s^{\boldsymbol{g}_2})\big|^2\Big]\mathrm{d}s \\
&+ 2\int_0^t \mathbb{E}\Big[\langle X_s^{\boldsymbol{g}_1} - X_s^{\boldsymbol{g}_2}, g_{1,1}(s, X_s^{\boldsymbol{g}_1}) - g_{1,1}(s, X_s^{\boldsymbol{g}_2})\rangle\Big]\mathrm{d}s \\
&+ 2\int_0^t \mathbb{E}\Big[\langle X_s^{\boldsymbol{g}_1} - X_s^{\boldsymbol{g}_2}, g_{1,1}(s, X_s^{\boldsymbol{g}_2}) - g_{2,1}(s, X_s^{\boldsymbol{g}_2})\rangle\Big]\mathrm{d}s \\
&+ 2\int_0^t \mathbb{E}\Big[\big|g_{1,2}(s, X_s^{\boldsymbol{g}_2}) - g_{2,2}(s, X_s^{\boldsymbol{g}_2})\big|^2\Big]\mathrm{d}s \\
&\le \mathbb{E}[|X_0^{\boldsymbol{g}_1} - X_0^{\boldsymbol{g}_2}|^2] + C\int_0^t \mathbb{E}\Big[\big|X_s^{\boldsymbol{g}_1} - X_s^{\boldsymbol{g}_2}\big|^2\Big]\mathrm{d}s \\
&+ 4\int_0^t \mathbb{E}\Big[\big|g_{1,1}(s, X_s^{\boldsymbol{g}_2}) - g_{2,1}(s, X_s^{\boldsymbol{g}_2})\big|^2 + \big|g_{1,2}(s, X_s^{\boldsymbol{g}_2}) - g_{2,2}(s, X_s^{\boldsymbol{g}_2})\big|^2\Big]\mathrm{d}s \\
&\le \mathbb{E}[|X_0^{\boldsymbol{g}_1} - X_0^{\boldsymbol{g}_2}|^2] + C\int_0^t \mathbb{E}\Big[\big|X_s^{\boldsymbol{g}_1} - X_s^{\boldsymbol{g}_2}\big|^2\Big]\mathrm{d}s \\
&+ C\int_0^t \|\boldsymbol{g}_1 - \boldsymbol{g}_2\|_{[0,T_0],q}^2 \mathbb{E}[1 + |X_s^{\boldsymbol{g}_2}|^{2q+2}]\mathrm{d}s,
\end{aligned}$$

where we used Young's inequality in the last display. By Grönwall's Lemma, we have

$$\sup_{t\in[0,T_0]}\mathbb{E}[|X_t^{\boldsymbol{g}_1}-X_t^{\boldsymbol{g}_2}|^2]\le CT_0e^{CT_0}\sup_{t\in[0,T_0]}\mathbb{E}[1+|X_t^{\boldsymbol{g}_2}|^{2q+2}]\|\boldsymbol{g}_1-\boldsymbol{g}_2\|^2_{[0,T_0],q}$$

From the result above, we have

$$\begin{aligned}
&\|\Gamma[\boldsymbol{g}_1]-\Gamma[\boldsymbol{g}_2]\|_{[0,T_0],q}\\
&\le \sup_{t\in[0,T_0]}\left(\sup_{x\in\mathbb{R}^d}\frac{|(f*\mu_t^{\boldsymbol{g}_1})(x)-(f*\mu_t^{\boldsymbol{g}_2})(x)|+|(f_\sigma*\mu_t^{\boldsymbol{g}_1})(x)-(f_\sigma*\mu_t^{\boldsymbol{g}_2})(x)|}{1+|x|^{q+1}}\right)\\
&\le C\sup_{t\in[0,T_0]}\left(\sup_{x\in\mathbb{R}^d}\frac{\mathbb{E}\left[|X_t^{\boldsymbol{g}_1}-X_t^{\boldsymbol{g}_2}|(1+|x|^{q+1})\big(1+|X_t^{\boldsymbol{g}_1}|^q+|X_t^{\boldsymbol{g}_2}|^q\big)\right]}{1+|x|^{q+1}}\right)\\
&\le C\sup_{t\in[0,T_0]}\mathbb{E}\left[|X_t^{\boldsymbol{g}_1}-X_t^{\boldsymbol{g}_2}|\big(1+|X_t^{\boldsymbol{g}_1}|^q+|X_t^{\boldsymbol{g}_2}|^q\big)\right]\\
&\le C\Big(\sup_{t\in[0,T_0]}\mathbb{E}[|X_t^{\boldsymbol{g}_1}-X_t^{\boldsymbol{g}_2}|^2]\Big)^{1/2}\Big(\sup_{t\in[0,T_0]}\mathbb{E}\left[\big(1+|X_t^{\boldsymbol{g}_1}|^q+|X_t^{\boldsymbol{g}_2}|^q\big)^2\right]\Big)^{1/2}\\
&\le C\Big(e^{CT_0}\sqrt{T_0}\Big)\Big(\sup_{t\in[0,T_0]}\mathbb{E}[1+|X_t^{\boldsymbol{g}_2}|^{2q+2}]\Big)^{1/2}\\
&\quad\times\Big(\sup_{t\in[0,T_0]}\mathbb{E}\left[\big(1+|X_t^{\boldsymbol{g}_1}|^q+|X_t^{\boldsymbol{g}_2}|^q\big)^2\right]\Big)^{1/2}\|\boldsymbol{g}_1-\boldsymbol{g}_2\|_{[0,T_0],q}\\
&\le C\Big(e^{CT_0}\sqrt{T_0}\Big)\Big(1+\sup_{t\in[0,T_0]}\mathbb{E}[|X_t^{\boldsymbol{g}_1}|^{2q+2}]+\sup_{t\in[0,T_0]}\mathbb{E}[|X_t^{\boldsymbol{g}_2}|^{2q+2}]\Big)\|\boldsymbol{g}_1-\boldsymbol{g}_2\|_{[0,T_0],q},
\end{aligned}$$

where we used Young's inequality in the last estimate. Performing similar calculations as above for the moments of $X_t^{\boldsymbol{g}_1}$ and $X_t^{\boldsymbol{g}_2}$, which by assumption exist up to order $m>2q+2$, allows to deduce that T_0 can indeed be chosen small enough such that Γ maps E onto E is a contraction operator. We conclude that the sequence $(\boldsymbol{g}^n)_{n\ge0}$ defined by $\boldsymbol{g}^{n+1}=\Gamma[\boldsymbol{g}^n]$, for $\boldsymbol{g}^0\in E$, is a Cauchy sequence belonging to E and converges with respect to the $\|\cdot\|_{[0,T_0],q}$-norm to $\boldsymbol{g}=\Gamma[\boldsymbol{g}]$ satisfying (4.22). Thus, for all $t\in[0,T_0]$, we have

$$\boldsymbol{g}(t,X_t^{\boldsymbol{g}})=(f*\mu_t^{\boldsymbol{g}}(X_t^{\boldsymbol{g}}),f_\sigma*\mu_t^{\boldsymbol{g}}(X_t^{\boldsymbol{g}})).$$

Substituting this into (4.23), yields (4.1) and thus $(X_t)_{t\in[0,T_0]}$ with $\sup_{t\in[0,T_0]}\mathbb{E}[|X_t|^m]<\infty$.

Our challenge now is to find a solution over the whole interval $[0,T]$. From the above analysis, we observe that the implied constants (and therefore the choice of T_0) depend on the moments of X_0. Therefore, we are not immediately able to deduce the existence of a solution on $[0,T]$. We need to ensure that these constants do not explode.

Below, we show pointwise p-th moment estimates for $m\ge p>2$ (the case $p=2$ follows in a straightforward manner from the below arguments where one would use Lemma 6.3.1 and Lemma 6.3.2 instead of the additional symmetry property. From Itô's formula, Assumption 4.2.1 and Remark 4.2.4, for all $t\in[0,T_0]$, we deduce

$$\begin{aligned}
|X_t|^p\le{}&|X_0|^p+p\int_0^t|X_s|^{p-2}\langle X_s,v(X_s,\mu_s^X)\rangle\mathrm{d}s+p\int_0^t|X_s|^{p-2}\langle X_s,\overline{\sigma}(s,X_s,\mu_s^X)\mathrm{d}W_s\rangle\\
&+p\int_0^t|X_s|^{p-2}\langle X_s,b(s,X_s,\mu_s^X)\rangle\mathrm{d}s \qquad (4.26)\\
&+p(p-1)\int_0^t|X_s|^{p-2}\Big(|\sigma(s,X_s,\mu_s^X)|^2+\int_{\mathbb{R}^d}|f_\sigma(X_s-y)|^2\mu_s^X(\mathrm{d}y)\Big)\mathrm{d}s\\
\le{}&|X_0|^p+C\int_0^t\Big(1+|X_s|^p+\big(W^{(2)}(\mu_s^X,\delta_0)\big)^p\Big)\mathrm{d}s+p\int_0^t|X_s|^{p-2}\langle X_s,\overline{\sigma}(s,X_s,\mu_s^X)\mathrm{d}W_s\rangle\\
&+p\int_0^t|X_s|^{p-2}\Big(\langle X_s,\int_{\mathbb{R}^d}f(X_s-y)\mu_s^X(\mathrm{d}y)\rangle+(p-1)\int_{\mathbb{R}^d}|f_\sigma(X_s-y)|^2\mu_s^X(\mathrm{d}y)\Big)\mathrm{d}s.
\end{aligned}$$

Taking expectation on both sides, using Assumption 4.2.1, in particular ($\mathbf{A}^f$, $\mathbf{A}^{f_\sigma}$), and Remark

4.2.4, we derive

$$
\begin{aligned}
\mathbb{E}\big[|X_t|^p\big] \leq & \mathbb{E}\big[|X_0|^p\big] + C\int_0^t (1+\mathbb{E}\big[|X_s|^p\big])\mathrm{d}s \\
& + \frac{p}{4}\int_0^t \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (|x|^{p-2}-|y|^{p-2})\langle x+y, f(x-y)\rangle \mu_s^X(\mathrm{d}x)\mu_s^X(\mathrm{d}y)\mathrm{d}s \\
& + \frac{p}{2}\int_0^t \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x|^{p-2}\big(\langle x-y, f(x-y)\rangle + 2(p-1)|f_\sigma(x-y)|^2\big)\mu_s^X(\mathrm{d}x)\mu_s^X(\mathrm{d}y)\mathrm{d}s \\
\leq & \mathbb{E}\big[|X_0|^p\big] + C\int_0^t \mathbb{E}\big[|X_s|^p\big]\mathrm{d}s + Ct.
\end{aligned}
$$

Gronwall's lemma yields the pointwise moment estimate,

$$
\sup_{t\in[0,T_0]} \mathbb{E}\big[|X_t|^{\widetilde{m}}\big] \leq C\left(1+\mathbb{E}[|X_0|^{\widetilde{m}}]\right)e^{CT_0}, \qquad \text{for any } \widetilde{m}\in[2,m]. \tag{4.27}
$$

Since, we have established a-priori L^p-moment bounds, for $p\in[2,m]$, (which substitutes for [2, Proposition 3.13]), we can repeat the arguments from above to establish the existence of a solution to an arbitrary time interval $[0,T]$. To be more precise, we first show that we can choose constants $K_1, T_1>0$ (independent of T_0) such that for $T_0+T_1\in[0,T]$ we have $\|\Gamma[\boldsymbol{g}]\|_{[T_0,T_0+T_1],q}\leq K_1$, for $\|\boldsymbol{g}\|_{[T_0,T_0+T_1],q}\leq K_1$. From Equation 4.25, we get

$$
\begin{aligned}
&\|\Gamma[\boldsymbol{g}]\|_{[T_0,T_0+T_1],q} \\
&\quad\leq C + Ce^{CT_1}\bigg(\sup_{t\in[0,T_0]}\mathbb{E}[|X_t|^{q+1}] + T_1\|\boldsymbol{g}\|^{q+1}_{[T_0,T_0+T_1],q} \\
&\qquad\quad + \int_{T_0}^{T_0+T_1}\Big(|b(s,0,\delta_0)|^{q+1} + |u(0,\delta_0)|^{q+1} + |\sigma(s,0,\delta_0)|^{q+1}\Big)\mathrm{d}s\bigg) \\
&\quad\leq C + Ce^{CT_1}\bigg(\sup_{t\in[0,T_0]}\mathbb{E}[|X_t|^{q+1}] + T_1K_1^{q+1} \\
&\qquad\quad + \int_{T_0}^{T_0+T_1}\Big(|b(s,0,\delta_0)|^{q+1} + |u(0,\delta_0)|^{q+1} + |\sigma(s,0,\delta_0)|^{q+1}\Big)\mathrm{d}s\bigg) \\
&\quad\leq C + Ce^{CT_1}\bigg(e^{CT_0}\Big(1+\mathbb{E}[|X_0|^{q+1}]\Big) + T_1K_1^{q+1} \\
&\qquad\quad + \int_{T_0}^{T_0+T_1}\Big(|b(s,0,\delta_0)|^{q+1} + |u(0,\delta_0)|^{q+1} + |\sigma(s,0,\delta_0)|^{q+1}\Big)\mathrm{d}s\bigg),
\end{aligned}
$$

where we used (4.27) in the last inequality.

Let now $K_1 = 2C(1+e^{CT}+e^{CT}\mathbb{E}[|X_0|^{q+1}])$. Then, we choose $T_1>0$ (independent of T_0) small enough such that for any $\|\boldsymbol{g}\|_{[T_0,T_0+T_1],q}\leq K_1$, we have $\|\Gamma[\boldsymbol{g}]\|_{[T_0,T_0+T_1],q}\leq K_1$. Similary as above, we can show that the map $\Gamma: E_1\to E_1$, where

$$
E_1 := \{\boldsymbol{g}\in\Lambda_{[T_0,T_0+T_1],q} : \|\boldsymbol{g}\|_{[T_0,T_0+T_1],q}\leq K_1\},
$$

forms a contraction (eventually choosing T_1 even smaller as above). The argument from above (choosing K_2 etc. as K_1) can be repeated to establish the existence of a solution on the time interval $[0,T]$.

□

4.4.2 Proof of Theorem 4.2.8: Propagation of chaos

Proof. Due to Lemma 4.2.7 and conditions ($\mathbf{A}^u$, $\mathbf{A}^\sigma$, $\mathbf{A}^f$, $\mathbf{A}^{f_\sigma}$), we observe that the drift and diffusion of the interacting particle system (viewed as an SDE in $\mathbb{R}^{Nd}$) satisfy a monotonicity condition as in [113, Section 2] which allow us to deduce that the interacting particle system has a unique strong solution. Critically, the wellposedness result therein does not yield moment estimates that are independent of N, as we interpreted the particle system as one single SDE in $\mathbb{R}^{Nd}$. In the next step, we prove moment bounds independent of N. By Itô's formula, Assumption 4.2.1, Remark 4.2.4 and Jensen's inequality, we have, for all $t \in [0,T]$, $i \in [\![1,N]\!]$, $2 \le p \le m$,

$$
\begin{aligned}
&\mathbb{E}\big[|X_t^{i,N}|^p\big] \le \mathbb{E}\big[|X_0^{i,N}|^p\big] + p\mathbb{E}\Big[\int_0^t |X_s^{i,N}|^{p-2}\langle X_s^{i,N}, v(X_s^{i,N},\mu_s^{X,N}) + b(s,X_s^{i,N},\mu_s^{X,N})\rangle\Big]\mathrm{d}s\\
&+ p\mathbb{E}\Big[\int_0^t |X_s^{i,N}|^{p-2}\Big(\langle X_s^{i,N}, \overline{\sigma}(s,X_s^{i,N},\mu_s^{X,N})\mathrm{d}W_s^i\rangle + \tfrac{(p-1)}{2}|\overline{\sigma}(s,X_s^{i,N},\mu_s^{X,N})|^2\mathrm{d}s\Big)\Big]\\
&\le \mathbb{E}\big[|X_0^{i,N}|^p\big] + C\int_0^t \mathbb{E}\big[|X_s^{i,N}|^p\big]\mathrm{d}s + CT\\
&+ p\int_0^t \mathbb{E}\Big[|X_s^{i,N}|^{p-2}\Big(\langle X_s^{i,N}, v(X_s^{i,N},\mu_s^{X,N})\rangle + (p-1)|\sigma(s,X_s^{i,N},\mu_s^{X,N})|^2\\
&\qquad + \langle X_s^{i,N}, \frac{1}{N}\sum_{j=1}^N f(X_s^{i,N}-X_s^{j,N})\rangle + (p-1)\frac{1}{N}\sum_{j=1}^N \big|f_\sigma(X_s^{i,N}-X_s^{j,N})\big|^2\Big)\Big]\mathrm{d}s\\
&\le \mathbb{E}\big[|X_0^{i,N}|^p\big] + CT + C\int_0^t \mathbb{E}\big[|X_s^{i,N}|^p\big]\mathrm{d}s\\
&+ \frac{p}{2N}\sum_{j=1}^N\int_0^t \mathbb{E}\Big[|X_s^{i,N}|^{p-2}\langle X_s^{i,N}-X_s^{j,N}, f(X_s^{i,N}-X_s^{j,N})\rangle\Big]\mathrm{d}s\\
&+ \frac{p(p-1)}{N}\sum_{j=1}^N\int_0^t \mathbb{E}\Big[|X_s^{i,N}|^{p-2}\ \big|f_\sigma(X_s^{i,N}-X_s^{j,N})\big|^2\Big]\mathrm{d}s\\
&+ \frac{p}{4N}\sum_{j=1}^N\int_0^t \mathbb{E}\Big[(|X_s^{i,N}|^{p-2}-|X_s^{j,N}|^{p-2})\langle X_s^{i,N}+X_s^{j,N}, f(X_s^{i,N}-X_s^{j,N})\rangle\Big]\mathrm{d}s\\
&\le \mathbb{E}\big[|X_0^{i,N}|^p\big] + C\int_0^T \mathbb{E}\big[|X_s^{i,N}|^p\big]\mathrm{d}s + CT.
\end{aligned}
$$

where we use the following estimate that

$$
\begin{aligned}
&\sum_{j=1}^N \mathbb{E}\Big[|X_s^{i,N}|^{p-2}\langle X_s^{i,N}, f(X_s^{i,N}-X_s^{j,N})\rangle\Big]\\
&= \frac{1}{2}\sum_{j=1}^N \mathbb{E}\Big[\langle |X_s^{i,N}|^{p-2}X_s^{i,N} - |X_s^{j,N}|^{p-2}X_s^{j,N}, f(X_s^{i,N}-X_s^{j,N})\rangle\Big]\\
&= \frac{1}{2}\sum_{j=1}^N \mathbb{E}\Big[\langle |X_s^{i,N}|^{p-2}X_s^{i,N} - |X_s^{i,N}|^{p-2}X_s^{j,N} + |X_s^{i,N}|^{p-2}X_s^{j,N}\\
&\qquad\qquad - |X_s^{j,N}|^{p-2}X_s^{j,N}, f(X_s^{i,N}-X_s^{j,N})\rangle\Big]\\
&= \frac{1}{2}\sum_{j=1}^N \mathbb{E}\Big[|X_s^{i,N}|^{p-2}\langle X_s^{i,N}-X_s^{j,N}, f(X_s^{i,N}-X_s^{j,N})\rangle\Big]\\
&\quad + \frac{1}{4}\sum_{j=1}^N \mathbb{E}\Big[\big\langle\big(|X_s^{i,N}|^{p-2}X_s^{j,N} - |X_s^{j,N}|^{p-2}X_s^{j,N}\big) - \big(|X_s^{j,N}|^{p-2}X_s^{i,N}\\
&\qquad\qquad - |X_s^{i,N}|^{p-2}X_s^{i,N}\big), f(X_s^{i,N}-X_s^{j,N})\big\rangle\Big]\\
&= \frac{1}{2}\sum_{j=1}^N \mathbb{E}\Big[|X_s^{i,N}|^{p-2}\langle X_s^{i,N}-X_s^{j,N}, f(X_s^{i,N}-X_s^{j,N})\rangle\Big]
\end{aligned}
$$

$$+\frac{1}{4}\sum_{j=1}^{N}\mathbb{E}\Big[(|X_s^{i,N}|^{p-2}-|X_s^{j,N}|^{p-2})\langle X_s^{i,N}+X_s^{j,N},f(X_s^{i,N}-X_s^{j,N})\rangle\Big].$$

Taking supremum over i and t, shows the claim using Gronwall's lemma; Jensen's inequality yields the estimate for $1\le p<2$.

The estimate (4.9) is then a consequence of [2, Theorem 3.14], we provides some key differences here

Using Itô's formula, we have

$$\begin{aligned}
\sum_{i=1}^{N}\mathbb{E}\big[|X_t^{i,N}-X_t^i|^2\big] &= 2\int_0^t \mathbb{E}\big[\langle X_s^{i,N}-X_s^i, v(X_s^{i,N},\mu_s^{X,N})-v(X_s^i,\mu_s^X)\rangle\big]\mathrm{d}s \qquad (4.28)\\
&+\int_0^t \mathbb{E}\big[2\langle X_s-Y_s, b(s,X_s^{i,N},\mu_s^{X,N})-b(s,X_s^i,\mu_s^X)\rangle\\
&\qquad +|\overline{\sigma}(s,X_s^{i,N},\mu_s^{X,N})-\overline{\sigma}(s,X_s^i,\mu_s^X)|^2\big]\mathrm{d}s\\
&\le \int_0^t\sum_{i=1}^{N}2\mathbb{E}\big[\langle X_t^{i,N}-X_t^i, b(s,X_s^{i,N},\mu_s^{X,N})-b(s,X_s^i,\mu_s^X)\rangle\big]\mathrm{d}s\\
&+2\sum_{i=1}^{N}\int_0^t\mathbb{E}\big[\langle X_t^{i,N}-X_t^i,\frac{1}{N}\sum_{j=1}^{N}f(X_t^{i,N}-X_t^{j,N})-\frac{1}{N}\sum_{j=1}^{N}f(X_t^i-X_t^j)\rangle\big]\mathrm{d}s\\
&+2\sum_{i=1}^{N}\int_0^t\mathbb{E}\big[\langle X_t^{i,N}-X_t^i,\frac{1}{N}\sum_{j=1}^{N}f(X_t^i-X_t^j)-\int_{\mathbb{R}^d}f(X_t^i-y)\mu_s^X(dy)\rangle\big]\mathrm{d}s\\
&+4\sum_{i=1}^{N}\int_0^t\mathbb{E}\big[\big|\frac{1}{N}\sum_{j=1}^{N}f_\sigma(X_t^{i,N}-X_t^{j,N})-\frac{1}{N}\sum_{j=1}^{N}f_\sigma(X_t^i-X_t^j)\big|^2\big]\mathrm{d}s\\
&+4\sum_{i=1}^{N}\int_0^t\mathbb{E}\big[\big|\frac{1}{N}\sum_{j=1}^{N}f_\sigma(X_t^i-X_t^j)-\int_{\mathbb{R}^d}f_\sigma(X_t^i-y)\mu_s^X(dy)\ \big|^2\big]\mathrm{d}s\\
&+2\sum_{i=1}^{N}\int_0^t\mathbb{E}\Big[\langle X_t^{i,N}-X_t^i, u(X_s^{i,N},\mu_s^{X,N})-u(X_s^i,\mu_s^X)\rangle\\
&\qquad +|\sigma(s,X_s^{i,N},\mu_s^{X,N})-\sigma(s,X_s^i,\mu_s^X)|^2\Big]\mathrm{d}s\\
&\le C\sum_{i=1}^{N}\int_0^t\mathbb{E}\big[|X_t^{i,N}-X_t^i|^2\big]+(W^{(2)}(\mu_s^X,\mu_s^{X,N}))^2\mathrm{d}s\\
&+\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_0^t\Big(\big\langle (X_t^{i,N}-X_t^i)-(X_t^{j,N}-X_t^j), f(X_t^{i,N}-X_t^{j,N})-f(X_t^i-X_t^j)\big\rangle\\
&\qquad +4|f_\sigma(X_t^{i,N}-X_t^{j,N})-f_\sigma(X_t^i-X_t^j)|^2\Big)\mathrm{d}s\\
&+2\sum_{i=1}^{N}\int_0^t\mathbb{E}\big[\langle X_t^{i,N}-X_t^i,\frac{1}{N}\sum_{j=1}^{N}f(X_t^i-X_t^j)-\int_{\mathbb{R}^d}f(X_t^i-y)\mu_s^X(dy)\rangle\big]\mathrm{d}s\\
&+4\sum_{i=1}^{N}\int_0^t\mathbb{E}\big[\big|\frac{1}{N}\sum_{j=1}^{N}f_\sigma(X_t^i-X_t^j)-\int_{\mathbb{R}^d}f_\sigma(X_t^i-y)\mu_s^X(dy)\ \big|^2\big]\mathrm{d}s\\
&\le C\sum_{i=1}^{N}\int_0^t\mathbb{E}\big[|X_t^{i,N}-X_t^i|^2\big]+(W^{(2)}(\mu_s^X,\mu_s^{X,N}))^2\mathrm{d}s\\
&+\frac{C}{N}\sum_{i=1}^{N}\int_0^t\Big(\mathbb{E}\big[\big|X_t^{i,N}-X_t^i\big|^2\big]\Big)^{1/2}
\end{aligned}$$

$$\cdot \Big(\mathbb{E}\big[\big|\sum_{j=1}^{N} f(X_t^i - X_t^j) - \int_{\mathbb{R}^d} f(X_t^i - y)\mu_s^X(dy)\big|^2\big]\Big)^{1/2} \mathrm{d}s \tag{4.29}$$

$$+ \frac{C}{N^2}\sum_{i=1}^{N}\int_0^t \mathbb{E}\big[\big|\sum_{j=1}^{N} f_\sigma(X_t^i - X_t^j) - \int_{\mathbb{R}^d} f_\sigma(X_t^i - y)\mu_s^X(dy)\big|^2\big]\ \mathrm{d}s. \tag{4.30}$$

Now, to deal with (4.29) and (4.30), we use similar arguments as in [2, Equation 3.25], consider (4.29), we have

$$\begin{aligned}
&\mathbb{E}\big[\big|\sum_{j=1}^{N} f(X_t^i - X_t^j) - \int_{\mathbb{R}^d} f(X_t^i - y)\mu_s^X(dy)\big|^2\big] \\
&= \sum_{j,k=1}^{N} \mathbb{E}\big[\big\langle f(X_t^i - X_t^j) - \int_{\mathbb{R}^d} f(X_t^i - y)\mu_s^X(dy), f(X_t^i - X_t^k) - \int_{\mathbb{R}^d} f(X_t^i - y)\mu_s^X(dy)\big\rangle\big],
\end{aligned}$$

where for $i \neq j \neq k$, we X_t^i, X_t^j, X_t^k are independent and identically distributed, we have

$$\begin{aligned}
&\mathbb{E}[\langle f(X_t^i - X_t^j) - \int_{\mathbb{R}^d} f(X_t^i - y)\mu_s^X(dy), f(X_t^i - X_t^k) - \int_{\mathbb{R}^d} f(X_t^i - y)\mu_s^X(dy)\rangle] \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \langle f(x-y_1) - f(x-z_1), \\
&\qquad f(x-y_2) - f(x-z_2)\rangle \mu_s^X(dx)\mu_s^X(dy_1)\mu_s^X(dy_2)\mu_s^X(dz_1)\mu_s^X(dz_2) \\
&= (1-1+1-1)\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \langle f(x-y_1), f(x-y_2)\rangle \mu_s^X(dx)\mu_s^X(dy_1)\mu_s^X(dy_2) = 0.
\end{aligned}$$

So that only the cases $j = k$ matters, we have

$$\begin{aligned}
&\mathbb{E}\big[\big|\sum_{j=1}^{N} f(X_t^i - X_t^j) - \int_{\mathbb{R}^d} f(X_t^i - y)\mu_s^X(dy)\big|^2\big] \\
&\qquad = \sum_{j=1}^{N} \mathbb{E}\big[\big| f(X_t^i - X_t^j) - \int_{\mathbb{R}^d} f(X_t^i - y)\mu_s^X(dy)\big|^2\big] \leq CN,
\end{aligned}$$

where we use the Locally Lipschitz Assumption 2.2.1 on f with the moment bound results in Theorem 2.2.3 and Theorem 4.2.8. Similar arguments works for f_σ in (4.30). Now, the other steps follow similarly as in [2, Theorem 3.14] , we have for (4.28)

$$\sum_{i=1}^{N} \mathbb{E}\big[|X_t^{i,N} - X_t^i|^2\big] \leq C\sum_{i=1}^{N}\int_0^t \Big(\mathbb{E}\big[|X_t^{i,N} - X_t^i|^2\big] + (W^{(2)}(\mu_s^X, \mu_s^{X,N}))^2 + \frac{1}{\sqrt{N}} + \frac{1}{N}\Big)\mathrm{d}s,$$

The estimate (4.9) follows as in [2, Theorem 3.14]. □

4.4.3 Proof of Lemma 4.2.12: Stochastic C-Stability

The proof shown in this section is an extension of the results for classical SDEs in [22] to the particle system considered in this chapter.

Proof. For every $n \in [\![0, M]\!]$, we denote the difference of the two particles by

$$e_n^{i,N} := X_n^{i,N} - \hat{X}_n^{i,N}.$$

By the orthogonality of the conditional expectation it holds

$$\mathbb{E}\big[|e_n^{i,N}|^2\big] = \mathbb{E}\Big[\big|\mathbb{E}\big[e_n^{i,N} \mid \mathcal{F}_{t_{n-1}}\big]\big|^2\Big] + \mathbb{E}\Big[\big|e_n^{i,N} - \mathbb{E}\big[e_n^{i,N} \mid \mathcal{F}_{t_{n-1}}\big]\big|^2\Big]. \tag{4.31}$$

The term $e_n^{i,N}$ can be expressed as follows

$$e_n^{i,N} = X_n^{i,N} + \Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h) - \Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h) - \hat{X}_n^{i,N}.$$

Thus, for the first term in (4.31), it follows from the inequality $(a+b)^2 = a^2 + 2ab + b^2 \leq \big(1+h^{-1}\big)\, a^2 + (1+h)\, b^2$ that, we have

$$\begin{aligned}\mathbb{E}\Big[\big|\mathbb{E}\big[e_n^{i,N} \mid \mathcal{F}_{t_{n-1}}\big]\big|^2\Big] &\leq (1+\tfrac{1}{h})\mathbb{E}\Big[\Big|\mathbb{E}\big[X_n^{i,N} - \Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h) \mid \mathcal{F}_{t_{n-1}}\big]\Big|^2\Big] \\ &\quad + (1+h)\mathbb{E}\Big[\Big|\mathbb{E}\big[\Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h) - \hat{X}_n^{i,N} \mid \mathcal{F}_{t_{n-1}}\big]\Big|^2\Big]. \end{aligned} \tag{4.32}$$

Similarly, for the second term in (4.31), choose η such that $1 < \eta \leq (m-1)$ in order to use (4.3) in Assumption 4.2.1 ($\mathbf{A}^u$, $\mathbf{A}^\sigma$), we have

$$\begin{aligned}&\mathbb{E}\Big[\Big|e_n^{i,N} - \mathbb{E}\big[e_n^{i,N}|\mathcal{F}_{t_{n-1}}\big]\Big|^2\Big] \\ &\quad\leq (1+\tfrac{1}{\eta-1})\mathbb{E}\Big[\Big|\big(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_{t_{n-1}}]\big)\Big(X_n^{i,N} - \Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h)\Big)\Big|^2\Big] \\ &\qquad + \eta\, \mathbb{E}\Big[\Big|\big(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_{t_{n-1}}]\big)\Big(\Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h) - \hat{X}_n^{i,N}\Big)\Big|^2\Big]. \end{aligned} \tag{4.33}$$

Using the fact $\hat{X}_n^{i,N} = \Psi_i(\hat{X}_{n-1}^{i,N}, \hat{\mu}_{n-1}^{X,N}, t_{n-1}, h)$, and the definition of C-stability for the terms (4.32), (4.33) (note that $h \in (0,1)$)

$$\begin{aligned}&(1+h)\mathbb{E}\Big[\Big|\mathbb{E}\big[\Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h) - \hat{X}_n^{i,N} \mid \mathcal{F}_{t_{n-1}}\big]\Big|^2\Big] \\ &\quad + \eta\, \mathbb{E}\Big[\Big|\big(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_{t_{n-1}}]\big)\Big(\Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h) - \hat{X}_n^{i,N}\Big)\Big|^2\Big] \\ &\quad \leq (1+h)\Big((1+Ch)\mathbb{E}\big[|e_{n-1}^{i,N}|^2\big] + Ch\mathbb{E}\big[|W^{(2)}(\hat{\mu}_{n-1}^{X,N}, \mu_{n-1}^{X,N})|^2\big]\Big). \end{aligned}$$

We then further estimate (4.31) by

$$\begin{aligned}\mathbb{E}\big[|e_n^{i,N}|^2\big] &\leq (1+\tfrac{1}{h})\mathbb{E}\Big[\Big|\mathbb{E}\big[X_n^{i,N} - \Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h) \mid \mathcal{F}_{t_{n-1}}\big]\Big|^2\Big] \\ &\quad + (1+\tfrac{1}{\eta-1})\mathbb{E}\Big[\Big|\big(\mathrm{id} - \mathbb{E}\left[\cdot \mid \mathcal{F}_{t_{n-1}}\right]\big)\left(X_n^{i,N} - \Psi_i(X_{n-1}^{i,N}, \mu_{n-1}^{X,N}, t_{n-1}, h)\right)\Big|^2\Big] \\ &\quad + (1+Ch)\mathbb{E}\big[|e_{n-1}^{i,N}|^2\big] + Ch\mathbb{E}\big[|W^{(2)}(\hat{\mu}_{n-1}^{X,N}, \mu_{n-1}^{X,N})|^2\big]. \end{aligned}$$

Using the fact that the particles are identically distributed

$$\mathbb{E}\big[|W^{(2)}(\hat{\mu}_{n-1}^{X,N}, \mu_{n-1}^{X,N})|^2\big] \leq \frac{1}{N}\sum_{j=1}^N \mathbb{E}[|e_{n-1}^{j,N}|^2] = \mathbb{E}\big[|e_{n-1}^{i,N}|^2\big].$$

By induction, with $C_\eta = 1 + (\eta-1)^{-1}$, we have

$$\begin{aligned}\sup_{n\in[\![0,M]\!]} \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big] &\leq \mathbb{E}\big[|\hat{X}_0^{i,N} - \xi^i|^2\big] \\ &\quad + \sum_{k=1}^M \big(1+h^{-1}\big)\, \mathbb{E}\Big[\Big|\mathbb{E}\big[X_k^{i,N} - \Psi_i(X_{k-1}^{i,N}, \mu_{k-1}^{X,N}, t_{k-1}, h) \mid \mathcal{F}_{t_{k-1}}\big]\Big|^2\Big] \end{aligned}$$

$$+ C_\eta \sum_{k=1}^{M} \mathbb{E}\Big[\Big| \big(\mathrm{id} - \mathbb{E}\left[\cdot \mid \mathcal{F}_{t_{k-1}}\right]\big) \left(X_k^{i,N} - \Psi_i(X_{k-1}^{i,N}, \mu_{k-1}^{X,N}, t_{k-1}, h)\right) \Big|^2 \Big]$$

$$+ Ch \sum_{k=1}^{M} \mathbb{E}\big[|X_k^{i,N} - \hat{X}_k^{i,N}|^2\big] + \frac{Ch}{N} \sum_{k=1}^{M} \sum_{j=1}^{N} \mathbb{E}\big[|X_k^{j,N} - \hat{X}_k^{j,N}|^2\big].$$

Taking supremum over $i \in [\![1, N]\!]$ and applying the discrete Gronwall's Lemma yields the result. □

4.4.4 Proof of Theorem 4.2.13

Proof. Using Definition 4.2.10, Definition 4.2.11 and the result in Lemma 4.2.12, we obtain

$$\sup_{n \in [\![0,M]\!]} \sup_{i \in [\![1,N]\!]} \mathbb{E}\big[|X_n^{i,N} - \hat{X}_n^{i,N}|^2\big] \leq \mathrm{e}^{CT} \Bigg[\mathbb{E}\big[|X_0^{i,N} - \hat{X}_0^{i,N}|^2\big]$$

$$+ \sum_{k=1}^{M} \sup_{i \in [\![1,N]\!]} \Bigg(\left(1 + h^{-1}\right) \mathbb{E}\Big[\Big| \mathbb{E}\big[X_k^{i,N} - \Psi_i\left(X_{k-1}^{i,N}, \mu_{k-1}^{X,N}, t_{k-1}, h\right) \mid \mathcal{F}_{t_{k-1}}\big] \Big|^2 \Big]$$

$$+ C_\eta\, \mathbb{E}\Big[\Big| \big(\mathrm{id} - \mathbb{E}\left[\cdot \mid \mathcal{F}_{t_{k-1}}\right]\big) \left(X_k^{i,N} - \Psi_i(X_{k-1}^{i,N}, \mu_{k-1}^{X,N}, t_{k-1}, h)\right) \Big|^2 \Big] \Bigg) \Bigg]$$

$$\leq C\mathrm{e}^{CT} \sum_{k=1}^{M} \left((1 + h^{-1}) h^{2+2\gamma} + C_\eta h^{1+2\gamma} \right) \leq C h^{2\gamma},$$

where in the second last estimate we used $Mh = T$. □

4.4.5 Proof of Theorem 4.2.18: Convergence of the SSM scheme

The SSM is C-stable

We first need to prove (4.10), i.e., $\hat{X}_{n+1}^{i,N} \in L^2\big(\Omega, \mathcal{F}_{t_n+h}, \mathbb{P}; \mathbb{R}^d\big)$ for all $n \in [\![0, M-1]\!]$ and $i \in [\![1, N]\!]$ given $\hat{X}_n^{i,N} \in L^2\big(\Omega, \mathcal{F}_{t_n}, \mathbb{P}; \mathbb{R}^d\big)$, where $\hat{X}^{i,N}$ is constructed by the SSM scheme defined in (4.12) and (4.13). We first provide the following useful result for the later proof.

Proposition 4.4.1 (Summation relationship)**.** *Let Assumption 4.2.1 hold and choose h as in* (4.14)*. Then there exists a constant $C > 0$ such that, for all $n \in [\![0, M-1]\!]$,*

$$\frac{1}{N} \sum_{j=1}^{N} |Y_n^{j,\star,N}|^2 \leq Ch + (1 + Ch)\ \frac{1}{N} \sum_{j=1}^{N} |\hat{X}_n^{j,N}|^2. \tag{4.34}$$

Proof. See Proposition 3.4.4. □

Proposition 4.4.2 (Second order moment bounds of SSM)**.** *Let the setting of Theorem 4.2.18 hold. Then there exists a constant $C > 0$ independent of h, N, M such that*

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

Proof. The proof is similar to Proposition 3.4.5. By Assumption 4.2.1, Proposition 4.4.1 , and the fact that the particles are identically distributed, we deduce that there exists a constant $C > 0$ such that for any $i \in [\![1, N]\!]$, $n \in [\![0, M-1]\!]$

$$\mathbb{E}\big[1 + |Y_n^{i,\star,N}|^2\big] = \frac{1}{N} \sum_{j=1}^{N} \mathbb{E}\big[1 + |Y_n^{j,\star,N}|^2\big]$$

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

From (4.12) and Jensen's inequality, we have

$$\begin{aligned}
|Y_n^{i,\star,N}|^2 &= \big\langle Y_n^{i,\star,N}, \hat{X}_n^{i,N} + hv(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big\rangle \\
&\leq \frac{1}{2}|Y_n^{i,\star,N}|^2 + \frac{1}{2}|\hat{X}_n^{i,N}|^2 + h\big\langle Y_n^{i,\star,N}, v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big\rangle,
\end{aligned}$$

and hence,

$$|Y_n^{i,\star,N}|^2 \leq |\hat{X}_n^{i,N}|^2 + 2h\big\langle Y_n^{i,\star,N}, v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big\rangle. \tag{4.35}$$

Also, from (4.13) and using the result above, we have

$$|\hat{X}_{n+1}^{i,N}|^2 = \big|Y_n^{i,\star,N} + b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})h + \overline{\sigma}(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Delta W_n^i\big|^2.$$

Taking expectation on both sides, by Jensen's inequality, (4.35), Assumption 4.2.1 and Remark 4.2.4, we have

$$\begin{aligned}
&\mathbb{E}\big[1+|\hat{X}_{n+1}^{i,N}|^2\big] \\
&\quad\leq (1+Ch)\mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big] + h\mathbb{E}\Big[2\big\langle Y_n^{i,\star,N}, v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big\rangle + |\overline{\sigma}(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})|^2\Big] \\
&\quad\leq (1+Ch)\mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big] + 2h\mathbb{E}\Big[\big\langle Y_n^{i,\star,N}, u(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big\rangle + |\sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})|^2\Big] \\
&\qquad + \frac{h}{N}\sum_{j=1}^{N}\mathbb{E}\Big[\big\langle Y_n^{i,\star,N} - Y_n^{j,\star,N}, f(Y_n^{i,\star,N} - Y_n^{j,\star,N})\big\rangle + 2|f_\sigma(Y_n^{i,\star,N} - Y_n^{j,\star,N})|^2\Big] \\
&\quad\leq (1+Ch)\mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big] + Ch\mathbb{E}\Big[1+|Y_n^{i,\star,N}|^2 + \frac{1}{N}\sum_{j=1}^{N}|Y_n^{j,\star,N}|^2\Big] \\
&\qquad + \frac{Ch}{N}\sum_{j=1}^{N}\mathbb{E}\Big[|Y_n^{i,\star,N} - Y_n^{j,\star,N}|^2\Big] \\
&\quad\leq (1+Ch)\mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big] + Ch\mathbb{E}\Big[1+|Y_n^{i,\star,N}|^2 + \frac{1}{N}\sum_{j=1}^{N}|Y_n^{j,\star,N}|^2\Big] \\
&\quad\leq (1+Ch)\mathbb{E}\big[1+|\hat{X}_n^{i,N}|^2\big].
\end{aligned}$$

□

Proposition 4.4.2 shows that the one-step map of the SSM, $\Psi = (\Psi_1, \ldots, \Psi_N)$ in Definition 4.2.9, $\Psi_i(\hat{X}_n^{i,N}, \hat{\mu}_n^{X,N}, t_n, h) = \hat{X}_{n+1}^{i,N}$ is indeed an L^2-operator. We now prove the SSM is C-stable.

Proof of statement 1 in Theorem 4.2.18. We use (4.12) and (4.13) to define the mapping $\Psi = (\Psi_1, \ldots, \Psi_N)$ and consequently to generate the following two processes $\hat{X}_n^{i,N}$ and $\hat{Z}_n^{i,N}$ for all $i \in [\![1, N]\!]$, $n \in [\![0, M-1]\!]$, with the corresponding empirical measures $\hat{\mu}_n^{X,N}$, $\hat{\mu}_n^{Z,N} \in \mathcal{P}_2(\mathbb{R}^d)$ and $\Delta W_n^i = W_{t_{n+1}}^i - W_{t_n}^i$

$$\begin{aligned}
Y_n^{i,X,N} &= \hat{X}_n^{i,N} + hv(Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N}), \qquad \hat{\mu}_n^{Y,X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N}\delta_{Y_n^{j,X,N}}(\mathrm{d}x), \\
\hat{X}_{n+1}^{i,N} &= Y_n^{i,X,N} + b(t_n, Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N})h + \overline{\sigma}(t_n, Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N})\Delta W_n^i, \\
Y_n^{i,Z,N} &= \hat{Z}_n^{i,N} + hv(Y_n^{i,Z,N}, \hat{\mu}_n^{Y,Z,N}), \qquad \hat{\mu}_n^{Y,Z,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N}\delta_{Y_n^{j,Z,N}}(\mathrm{d}x),
\end{aligned}$$

$$\hat{Z}^{i,N}_{n+1} = Y^{i,Z,N}_n + b(t_n, Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)h + \overline{\sigma}(t_n, Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)\Delta W^i_n.$$

Thus, $\hat{X}^{i,N}_{n+1} = \Psi_i(\hat{X}^{i,N}_n, \hat{\mu}^{X,N}_n, t_n, h)$ and $\hat{Z}^{i,N}_{n+1} = \Psi_i(\hat{Z}^{i,N}_n, \hat{\mu}^{Z,N}_n, t_n, h)$. We need to prove

$$\mathbb{E}\Big[\Big|\mathbb{E}\big[\Psi_i(\hat{X}^{i,N}_n, \hat{\mu}^{X,N}_n, t_n, h) - \Psi_i(\hat{Z}^{i,N}_n, \hat{\mu}^{Z,N}_n, t_n, h) \mid \mathcal{F}_{t_n}\big]\Big|^2\Big] \tag{4.36}$$

$$+\eta\, \mathbb{E}\Big[\Big|(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_t])\,(\Psi_i(\hat{X}^{i,N}_n, \hat{\mu}^{X,N}_n, t_n, h) - \Psi_i(\hat{Z}^{i,N}_n, \hat{\mu}^{Z,N}_n, t_n, h))\Big|^2\Big] \tag{4.37}$$

$$\le (1 + Ch)\,\mathbb{E}\big[|\hat{X}^{i,N}_n - \hat{Z}^{i,N}_n|^2\big] + Ch\mathbb{E}\big[|W^{(2)}(\hat{\mu}^{X,N}_n, \hat{\mu}^{Z,N}_n)|^2\big].$$

For the first component (4.36), using the Lipschitz continuity of b, we get

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\mathbb{E}\big[\Psi_i(\hat{X}^{i,N}_n, \hat{\mu}^{X,N}_n, t_n, h) - \Psi_i(\hat{Z}^{i,N}_n, \hat{\mu}^{Z,N}_n, t_n, h) \mid \mathcal{F}_{t_n}\big]\Big|^2\Big] \\
&\quad = \mathbb{E}\big[\big|Y^{i,X,N}_n + b(t_n, Y^{i,X,N}_n, \hat{\mu}^{Y,X,N}_n)h - Y^{i,Z,N}_n - b(t_n, Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)h\big|^2\big] \\
&\quad \le (1 + Ch)\mathbb{E}\big[|Y^{i,X,N}_n - Y^{i,Z,N}_n|^2\big] + Ch\mathbb{E}\big[|W^{(2)}(\hat{\mu}^{Y,X,N}_n, \hat{\mu}^{Y,Z,N}_n)|^2\big].
\end{aligned}$$

Due to Remark 4.2.4 and Lemma 4.2.7, we observe

$$\begin{aligned}
&|Y^{i,X,N}_n - Y^{i,Z,N}_n|^2 \\
&\quad = \langle Y^{i,X,N}_n - Y^{i,Z,N}_n, \hat{X}^{i,N}_n - \hat{Z}^{i,N}_n + v(Y^{i,X,N}_n, \hat{\mu}^{Y,X,N}_n)h - v(Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)h\rangle \\
&\quad \le \frac{1}{2}\big(|Y^{i,X,N}_n - Y^{i,Z,N}_n|^2 + |\hat{X}^{i,N}_n - \hat{Z}^{i,N}_n|^2\big) \\
&\qquad + h\langle Y^{i,X,N}_n - Y^{i,Z,N}_n, v(Y^{i,X,N}_n, \hat{\mu}^{Y,X,N}_n) - v(Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)\rangle,
\end{aligned}$$

and therefore

$$\begin{aligned}
&|Y^{i,X,N}_n - Y^{i,Z,N}_n|^2 \\
&\quad \le |\hat{X}^{i,N}_n - \hat{Z}^{i,N}_n|^2 + 2h\langle Y^{i,X,N}_n - Y^{i,Z,N}_n, v(Y^{i,X,N}_n, \hat{\mu}^{Y,X,N}_n) - v(Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)\rangle \\
&\quad \le |\hat{X}^{i,N}_n - \hat{Z}^{i,N}_n|^2 \\
&\quad + 2h\Big\langle Y^{i,X,N}_n - Y^{i,Z,N}_n, \frac{1}{N}\sum_{j=1}^N \big(f(Y^{i,X,N}_n - Y^{j,X,N}_n) - f(Y^{i,Z,N}_n - Y^{j,Z,N}_n)\big)\Big\rangle \\
&\quad + 2h\langle Y^{i,X,N}_n - Y^{i,Z,N}_n, u(Y^{i,X,N}_n, \hat{\mu}^{Y,X,N}_n) - u(Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)\rangle.
\end{aligned} \tag{4.38}$$

For the second component (4.37), by Jensen's inequality, we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big|(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_{t_n}])\,(\Psi_i(\hat{X}^{i,N}_n, \hat{\mu}^{X,N}_n, t_n, h) - \Psi_i(\hat{Z}^{i,N}_n, \hat{\mu}^{Z,N}_n, t_n, h))\Big|^2\Big] \\
&\quad = \mathbb{E}\big[|\overline{\sigma}(t_n, Y^{i,X,N}_n, \hat{\mu}^{Y,X,N}_n)\Delta W^i_n - \overline{\sigma}(t_n, Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)\Delta W^i_n|^2\big] \\
&\quad \le 2h\mathbb{E}\Big[|\sigma(t_n, Y^{i,X,N}_n, \hat{\mu}^{Y,X,N}_n) - \sigma(t_n, Y^{i,Z,N}_n, \hat{\mu}^{Y,Z,N}_n)|^2 \\
&\qquad + \frac{1}{N}\sum_{j=1}^N |f_\sigma(Y^{i,X,N}_n - Y^{j,X,N}_n) - f_\sigma(Y^{i,Z,N}_n - Y^{j,Z,N}_n)|^2\Big].
\end{aligned}$$

From Assumption 4.2.1 and Remark 4.2.4, we derive, for some $\eta > 1$,

$$\begin{aligned}
&\mathbb{E}\Big[\Big\langle Y^{i,X,N}_n - Y^{i,Z,N}_n, \frac{1}{N}\sum_{j=1}^N \big(f(Y^{i,X,N}_n - Y^{j,X,N}_n) - f(Y^{i,Z,N}_n - Y^{j,Z,N}_n)\big)\Big\rangle\Big] \\
&\quad + \eta\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N |f_\sigma(Y^{i,X,N}_n - Y^{j,X,N}_n) - f_\sigma(Y^{i,Z,N}_n - Y^{j,Z,N}_n)|^2\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\Big[\big\langle (Y_n^{i,X,N}-Y_n^{j,X,N})-(Y_n^{i,Z,N}-Y_n^{j,Z,N}),\\
&\qquad\qquad f(Y_n^{i,X,N}-Y_n^{j,X,N})-f(Y_n^{i,Z,N}-Y_n^{j,Z,N})\big\rangle\Big]\\
&\quad+\frac{\eta}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\Big[|f_\sigma(Y_n^{i,X,N}-Y_n^{j,X,N})-f_\sigma(Y_n^{i,Z,N}-Y_n^{j,Z,N})|^2\Big]\\
&\le \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\Big[L_{(f)}^{(1)}\big|(Y_n^{i,X,N}-Y_n^{j,X,N})-(Y_n^{i,Z,N}-Y_n^{j,Z,N})\big|^2\Big]\\
&\le 2L_{(f)}^{(1),+}\mathbb{E}\big[|Y_n^{i,X,N}-Y_n^{i,Z,N}|^2\big]. \qquad (4.39)
\end{aligned}
$$

Collecting the above estimates and using Remark 4.2.4, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\Big|\mathbb{E}\big[\Psi_i(\hat{X}_n^{i,N},\hat{\mu}_n^{X,N},t_n,h)-\Psi_i(\hat{Z}_n^{i,N},\hat{\mu}_n^{Z,N},t_n,h)\mid\mathcal{F}_t\big]\Big|^2\Big]\\
&\quad+\eta\mathbb{E}\Big[\Big|(\mathrm{id}-\mathbb{E}[\cdot\mid\mathcal{F}_t])\left(\Psi_i(\hat{X}_n^{i,N},\hat{\mu}_n^{X,N},t_n,h)-\Psi_i(\hat{Z}_n^{i,N},\hat{\mu}_n^{Z,N},t_n,h)\right)\Big|^2\Big]\\
\le&\mathbb{E}\Big[|\hat{X}_n^{i,N}-\hat{Z}_n^{i,N}|^2+4L_{(f)}^{(1),+}h|Y_n^{i,X,N}-Y_n^{i,Z,N}|^2 \qquad (4.40)\\
&\quad+2\eta h|\sigma(t_n,Y_n^{i,X,N},\hat{\mu}_n^{Y,X,N})-\sigma(t_n,Y_n^{i,Z,N},\hat{\mu}_n^{Y,Z,N})|^2\\
&\quad+2h\langle Y_n^{i,X,N}-Y_n^{i,Z,N},u(Y_n^{i,X,N},\hat{\mu}_n^{Y,X,N})-u(Y_n^{i,Z,N},\hat{\mu}_n^{Y,Z,N})\rangle\Big](1+Ch)\\
\le&(1+Ch)\mathbb{E}\big[|\hat{X}_n^{i,N}-\hat{Z}_n^{i,N}|^2\big]+Ch\left(\mathbb{E}\big[|Y_n^{i,X,N}-Y_n^{i,Z,N}|^2\big]+\mathbb{E}\big[|W^{(2)}(\hat{\mu}_n^{Y,X,N},\hat{\mu}_n^{Y,Z,N})|^2\big]\right), \qquad (4.41)
\end{aligned}
$$

where we use that the particles are identically distributed and the following inequality

$$
\begin{aligned}
&\mathbb{E}\Big[\langle Y_n^{i,X,N}-Y_n^{i,Z,N},u(Y_n^{i,X,N},\hat{\mu}_n^{Y,X,N})-u(Y_n^{i,Z,N},\hat{\mu}_n^{Y,Z,N})\rangle\\
&\qquad+\eta h|\sigma(t_n,Y_n^{i,X,N},\hat{\mu}_n^{Y,X,N})-\sigma(t_n,Y_n^{i,Z,N},\hat{\mu}_n^{Y,Z,N})|^2\Big]\\
&\le (L_{(u\sigma)}^{(1)}+L_{(u\sigma)}^{(2)})\mathbb{E}\big[|Y_n^{i,X,N}-Y_n^{i,Z,N}|^2\big]. \qquad (4.42)
\end{aligned}
$$

Substitute the estimates from above into (4.38), and take Remark 4.2.15 into account, we get

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

Further, we note that

$$
\mathbb{E}\big[|W^{(2)}(\hat{\mu}_n^{Y,X,N},\hat{\mu}_n^{Y,Z,N})|^2\big]\le\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}[|Y_n^{j,X,N}-Y_n^{j,Z,N}|^2]\le(1+Ch)\mathbb{E}\big[|\hat{X}_n^{i,N}-\hat{Z}_n^{i,N}|^2\big].
$$

Substitute these estimates in (4.41), allowing to deduce the claim. □

The SSM is B-consistent

We first state the following auxiliary results and recall that the constant C is positive and independent of h,N,M.

Proposition 4.4.3 (Difference relationship)**.** *Let Assumption 4.2.1 hold and choose h as in* (4.14)*. For any $n\in[\![0,M]\!]$, let $Y_n^{\star,N}$ defined as in* (4.11)-(4.12)*. Then, there exists a constant $C>0$ such that for all $i,\ j\in[\![1,N]\!]$,*

$$
|Y_n^{i,\star,N}-Y_n^{j,\star,N}|^2\le\frac{1}{1-2(L_{(f)}^{(1)}+L_{(u\sigma)}^{(1)})h}|\hat{X}_n^{i,N}-\hat{X}_n^{j,N}|^2\le(1+Ch)|\hat{X}_n^{i,N}-\hat{X}_n^{j,N}|^2. \qquad (4.43)
$$

Proof. See Proposition 3.4.3. ☐

Now, we state the following moment relationship for the first step of the SSM.

Proposition 4.4.4 (Moment relationship)**.** *Let Assumption 4.2.1 hold and choose h as in* (4.14)*, then there exist a constant $C > 0$ independent of N, such that for all $i \in [\![1, N]\!]$, $n \in [\![0, M]\!]$, $p \geq 1$ we have*

$$\mathbb{E}\big[|Y_n^{i,\star,N}|^{2p}\big] \leq C\Big(\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}[|X_n^{i,N} - X_n^{j,N}|^{2p}] + \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}(1+|X_n^{j,N}|^2)\Big|^p\Big] + 1\Big). \tag{4.44}$$

Proof. By Young's inequality and Jensen's inequality

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

Combining Propositions 4.4.3 and 4.4.1 allows to conclude the claim.

☐

The main goal of this section is to prove that $X_t^{i,N}$ defined by (4.7) satisfies for all $t \in [0, T]$, $i \in [\![1, N]\!]$ the following estimates with $\gamma = 1/2$.

$$\mathbb{E}\Big[\Big|\mathbb{E}\big[X_{t+h}^{i,N} - \Psi_i(X_t^{i,N}, \mu_t^{X,N}, t, h) \mid \mathcal{F}_t\big]\Big|^2\Big] \leq Ch^{2\gamma+2}, \tag{4.45}$$

$$\mathbb{E}\Big[\Big|(\mathrm{id} - \mathbb{E}\,[\cdot \mid \mathcal{F}_t])\,(X_{t+h}^{i,N} - \Psi_i(X_t^{i,N}, \mu_t^{X,N}, t, h))\Big|^2\Big] \leq Ch^{2\gamma+1}. \tag{4.46}$$

Proof of statement 2 in Theorem 4.2.18. Recall (4.7) and the SSM given in (4.11)-(4.13). Then, we introduce the following quantities, for all $t \in [0, T]$, $i \in [\![1, N]\!]$,

$$X_{t+h}^{i,N} = X_t^{i,N} + \int_t^{t+h}\Big(v(X_s^{i,N}, \mu_s^{X,N}) + b(s, X_s^{i,N}, \mu_s^{X,N})\Big)\mathrm{d}s + \int_t^{t+h}\overline{\sigma}(s, X_s^{i,N}, \mu_s^{X,N})\mathrm{d}W_s^i, \tag{4.47}$$

$$Y_t^{i,N} = X_t^{i,N} + v(Y_t^{i,N}, \mu_t^{Y,N})h, \qquad \mu_t^{Y,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N}\delta_{Y_t^{j,N}}(\mathrm{d}x), \tag{4.48}$$

$$\begin{aligned}&\Psi_i(X_t^{i,N}, \mu_t^{X,N}, t, h) \\ &\quad= X_t^{i,N} + \int_t^{t+h}\Big(v(Y_t^{i,N}, \mu_t^{Y,N}) + b(t, Y_t^{i,N}, \mu_t^{Y,N})\Big)\mathrm{d}s + \int_t^{t+h}\overline{\sigma}(t, Y_t^{i,N}, \mu_t^{Y,N})\mathrm{d}W_s^i,\end{aligned}$$

where the last equation is the integration form for the one-step map of SSM. Therefore, the first term (4.45) can be estimated by Jensen's inequality

$$\mathbb{E}\Big[\Big|\mathbb{E}\big[X_{t+h}^{i,N} - \Psi_i(X_t^{i,N}, \mu_t^{X,N}, t, h) \mid \mathcal{F}_t\big]\Big|^2\Big] \tag{4.49}$$

$$\leq 2h\int_t^{t+h}\mathbb{E}\big[|v(X_s^{i,N}, \mu_s^{X,N}) - v(Y_t^{i,N}, \mu_t^{Y,N})|^2\big]\,\mathrm{d}s \tag{4.50}$$

$$+\,2h\int_t^{t+h}\mathbb{E}\big[|b(s, X_s^{i,N}, \mu_s^{X,N}) - b(t, Y_t^{i,N}, \mu_t^{Y,N})|^2\big]\,\mathrm{d}s.$$

For the second term (4.46), we get

$$\mathbb{E}\Big[\Big|\big(\mathrm{id} - \mathbb{E}[\cdot \mid \mathcal{F}_t]\big)\,(X_{t+h}^{i,N} - \Psi_i(X_t^{i,N}, \mu_t^{X,N}, t, h))\Big|^2\Big] \tag{4.51}$$

$$\leq C \int_t^{t+h} \mathbb{E}\big[|\overline{\sigma}(s, X_s^{i,N}, \mu_s^{X,N}) - \overline{\sigma}(t, Y_t^{i,N}, \mu_t^{Y,N})|^2\big] \, \mathrm{d}s.$$

By Young's inequality and Jensen's inequality, Assumption 4.2.1 and Proposition 4.4.3, for $s \in [t, t+h]$, we have

$$\begin{aligned}
&|X_s^{i,N} - Y_t^{i,N}|^2 \leq 2|X_s^{i,N} - X_t^{i,N}|^2 + 2|X_t^{i,N} - Y_t^{i,N}|^2, \\
&|X_t^{i,N} - Y_t^{i,N}|^2 = |v(Y_t^{i,N}, \mu_t^{Y,N})h|^2 \leq \frac{2h^2}{N} \sum_{j=1}^N |f(Y_t^{i,N} - Y_t^{j,N})|^2 + 2h^2 |u(Y_t^{i,N}, \mu_t^{Y,N})|^2 \\
&\quad \leq \frac{Ch^2}{N} \sum_{j=1}^N \Big(1 + |Y_t^{i,N} - Y_t^{j,N}|^{2q+2}\Big) + Ch^2 \Big(1 + |Y_t^{i,N}|^{2q+2} + \frac{1}{N} \sum_{j=1}^N |Y_t^{j,N}|^2\Big) \\
&\quad \leq \frac{Ch^2}{N} \sum_{j=1}^N \Big(1 + |X_t^{i,N} - X_t^{j,N}|^{2q+2}\Big) + Ch^2 \Big(1 + |Y_t^{i,N}|^{2q+2} + \frac{1}{N} \sum_{j=1}^N |Y_t^{j,N}|^2\Big).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&|X_s^{i,N} - Y_t^{i,N}|^4 \leq 16|X_s^{i,N} - X_t^{i,N}|^4 + 16|X_t^{i,N} - Y_t^{i,N}|^4, \\
&|X_t^{i,N} - Y_t^{i,N}|^4 \\
&\quad \leq Ch^4 \Big(1 + |Y_t^{i,N}|^{4q+4} + \frac{1}{N} \sum_{j=1}^N |Y_t^{j,N}|^4\Big) + \frac{Ch^4}{N} \sum_{j=1}^N \Big(1 + |X_t^{i,N} - X_t^{j,N}|^{4q+4}\Big). \qquad (4.52)
\end{aligned}$$

Using the moment stability of $X^{i,N}$ (note $m > 4q + 4 > \max\{2(q+1), 4\}$) and Jensen's inequality, we get

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

By (4.47) and another application of Jensen's inequality

$$\begin{aligned}
\mathbb{E}\big[|X_s^{i,N} - X_t^{i,N}|^2\big] \leq & Ch \int_t^s \mathbb{E}\big[|v(X_u^{i,N}, \mu_u^{X,N}) + b(u, X_u^{i,N}, \mu_u^{X,N})|^2\big] \mathrm{d}u \\
& + C \int_t^s \mathbb{E}\big[|\overline{\sigma}(u, X_u^{i,N}, \mu_u^{X,N})|^2\big] \mathrm{d}u \ \leq Ch.
\end{aligned}$$

Similarly, we have

$$\mathbb{E}\big[|X_s^{i,N} - X_t^{i,N}|^4\big] \leq Ch^2.$$

Using the above results and we have sufficient moment bounds for $Y_t^{i,N}$ from Proposition 4.4.4, we conclude that

$$\begin{aligned}
&\mathbb{E}\big[|X_s^{i,N} - Y_t^{i,N}|^2\big] \leq Ch, \quad \mathbb{E}\big[|X_s^{i,N} - Y_t^{i,N}|^4\big] \leq Ch^2, \\
&\mathbb{E}\big[|W^{(2)}(\mu_s^{X,N}, \mu_t^{Y,N})|^2\big] \leq \frac{1}{N} \sum_{j=1}^N \mathbb{E}\big[|X_s^{j,N} - Y_t^{j,N}|^2\big] \leq Ch.
\end{aligned}$$

Thus, for the term (4.50), taking Assumption 4.2.1 into account, following the arguments in Section 3.4.2, Jensen's inequality, Cauchy-Schwarz inequality and Young's inequality yield

$$\mathbb{E}\big[|v(X_s^{i,N}, \mu_s^{X,N}) - v(Y_t^{i,N}, \mu_t^{Y,N})|^2\big]$$

$$\leq C\sqrt{\mathbb{E}\big[1+|X_s^{i,N}|^{4q}+|Y_t^{i,N}|^{4q}\big]\mathbb{E}\big[|X_s^{i,N}-Y_t^{i,N}|^4\big]}+C\mathbb{E}\big[|X_s^{i,N}-Y_t^{i,N}|^2\big]\leq Ch.$$

Also, from Assumption 4.2.1, we have

$$\begin{aligned}&\mathbb{E}\big[|b(s,X_s^{i,N},\mu_s^{X,N})-b(t,Y_t^{i,N},\mu_t^{Y,N})|^2\big]\\&\quad\leq C\big(h+\mathbb{E}\big[|X_s^{i,N}-Y_t^{i,N}|^2\big]+\mathbb{E}\big[|W^{(2)}(\mu_s^{X,N},\mu_t^{Y,N})|^2\big]\big)\leq Ch,\end{aligned}$$

and similarly, from Jensen's inequality and Remark 4.2.4, we have

$$\begin{aligned}&\mathbb{E}\big[|\overline{\sigma}(s,X_s^{i,N},\mu_s^{X,N})-\overline{\sigma}(t,Y_t^{i,N},\mu_t^{Y,N})|^2\big]\\&\quad\leq C\mathbb{E}\Big[h+\big(1+|X_s^{i,N}|^{2q}+|Y_t^{i,N}|^{2q}\big)|X_s^{i,N}-Y_t^{i,N}|^2+\frac{1}{N}\sum_{j=1}^N|X_s^{j,N}-Y_t^{j,N}|^{2q+2}\Big]\leq Ch.\end{aligned}$$

Substituting the results above back to (4.49) and (4.51), we have

$$\begin{aligned}\mathbb{E}\Big[\Big|\mathbb{E}\big[X_{t+h}^{i,N}-\Psi_i(X_t^{i,N},\mu_t^{X,N},t,h)\mid\mathcal{F}_t\big]\Big|^2\Big]&\leq Ch\int_t^{t+h}h\mathrm{d}s\leq Ch^3,\\ \mathbb{E}\Big[\Big|(\mathrm{id}-\mathbb{E}[\cdot\mid\mathcal{F}_t])\,(X_{t+h}^{i,N}-\Psi_i(X_t^{i,N},\mu_t^{X,N},t,h))\Big|^2\Big]&\leq C\int_t^{t+h}h\mathrm{d}s\leq Ch^2.\end{aligned}$$

□

Proof of convergence for the SSM scheme

Proof of statement 3 in Theorem 4.2.18. At last, we will prove the third statement in Theorem 4.2.18. By combining the first two statements and Theorem 4.2.13, we first have

$$\sup_{n\in[\![0,M]\!]}\sup_{i\in[\![1,N]\!]}\mathbb{E}\big[\,|X_n^{i,N}-\hat{X}_n^{i,N}|^2\big]\leq Ch. \tag{4.53}$$

Now, we extend the strong convergence rate to the continuous time version of the SSM, which has not been discussed in [22]. In order to extend the result above to the continuous extension of the SSM, we consider, for all $n\in[\![0,M-1]\!]$, $i\in[\![1,N]\!]$, $r\in[0,h]$,

$$\begin{aligned}|X_{t_n+r}^{i,N}-\hat{X}_{t_n+r}^{i,N}|^2=\Big|&X_{t_n}^{i,N}-\hat{X}_n^{i,N}+\int_{t_n}^{t_n+r}\big(v(X_s^{i,N},\mu_s^{X,N})-v(Y_n^{i,N},\mu_n^{Y,N})\big)\mathrm{d}s &(4.54)\\&+\int_{t_n}^{t_n+r}\big(b(s,X_s^{i,N},\mu_s^{X,N})-b(t_n,Y_n^{i,N},\mu_n^{Y,N})\big)\mathrm{d}s &(4.55)\\&+\int_{t_n}^{t_n+r}\big(\overline{\sigma}(s,X_s^{i,N},\mu_s^{X,N})-\overline{\sigma}(t_n,Y_n^{i,N},\mu_n^{Y,N})\big)\mathrm{d}W_s^i &(4.56)\\&+\int_{t_n}^{t_n+r}\big(v(Y_n^{i,N},\mu_n^{Y,N})-v(Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})\big)\mathrm{d}s\\&+\int_{t_n}^{t_n+r}\big(b(t_n,Y_n^{i,N},\mu_n^{Y,N})-b(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})\big)\mathrm{d}s\\&+\int_{t_n}^{t_n+r}\big(\overline{\sigma}(t_n,Y_n^{i,N},\mu_n^{Y,N})-\overline{\sigma}(t_n,Y_n^{i,\star,N},\hat{\mu}_n^{Y,N})\big)\mathrm{d}W_s^i\Big|^2,\end{aligned}$$

where $Y_n^{i,N}=Y_{t_n}^{i,N}$, $\mu_n^{Y,N}=\mu_{t_n}^{Y,N}$ are defined in (4.48). Taking expectation on both sides and using Jensen's inequality, we derive

$$\begin{aligned}&\mathbb{E}\big[|X_{t_n+r}^{i,N}-\hat{X}_{t_n+r}^{i,N}|^2\big]\\&\quad\leq C\mathbb{E}\Big[\Big|\big(X_{t_n}^{i,N}+v(Y_n^{i,N},\mu_n^{Y,N})r+b(t_n,Y_n^{i,N},\mu_n^{Y,N})r+\overline{\sigma}(t_n,Y_n^{i,N},\mu_n^{Y,N})\Delta W_{n,r}^i\big)\end{aligned}$$

$$- \big(\hat{X}_n^{i,N} + v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})r + b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})r + \overline{\sigma}(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Delta W_{n,r}^i\big)\Big|^2\Big] + Ch,$$

where $\Delta W_{n,r}^i = W_{t_n+r}^i - W_{t_n}^i$ and we remark that the integral terms in (4.54)-(4.56) can be analyzed using the results in Section 4.4.5. We now consider the following differences: From (4.12) and following similar calculations to Section 3.4.2, we have

$$\begin{aligned}
&\mathbb{E}\big[\big|\big(X_{t_n}^{i,N} + v(Y_n^{i,N}, \mu_n^{Y,N})r\big) - \big(\hat{X}_n^{i,N} + v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})r\big)\big|^2\big] \\
&\quad = \mathbb{E}\big[\big\langle\big(X_{t_n}^{i,N} - \hat{X}_n^{i,N}\big) + r\Delta V_n^Y, \big(Y_n^{i,N} - Y_n^{i,\star,N}\big) - (h-r)\Delta V_n^Y\big\rangle\big] \\
&\quad \leq \mathbb{E}\big[|X_{t_n}^{i,N} - \hat{X}_n^{i,N}|^2\big]\tfrac{2h-r}{2h} + \mathbb{E}\big[|Y_n^{i,N} - Y_n^{i,\star,N}|^2\big]\tfrac{r}{2h} + \mathbb{E}\big[\big\langle Y_n^{i,N} - Y_n^{i,\star,N}, \Delta V_n^Y\big\rangle\big]r,
\end{aligned}$$

where $\Delta V_n^Y = v(Y_n^{i,N}, \mu_n^{Y,N}) - v(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})$. By Jensen's inequality and the results in Section 4.4.5, we conclude that for all $n \in [\![0, M-1]\!]$, $i \in [\![1, N]\!]$, $r \in [0, h]$, we have

$$\begin{aligned}
\mathbb{E}\big[|X_{t_n+r}^{i,N} - \hat{X}_{t_n+r}^{i,N}|^2\big] &\leq Ch + C\mathbb{E}\big[|X_{t_n}^{i,N} - \hat{X}_n^{i,N}|^2\big] \\
&+ 2r\mathbb{E}\Big[\big\langle Y_n^{i,N} - Y_n^{i,\star,N}, u(Y_n^{i,N}, \mu_n^{Y,N}) - u(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big\rangle\Big] \\
&+ 2r\mathbb{E}\Big[\big|\sigma(t_n, Y_n^{i,N}, \mu_n^{Y,N}) - \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\big|^2\Big] \\
&+ \frac{2r}{N}\sum_{j=1}^N \mathbb{E}\Big[\big\langle Y_n^{i,N} - Y_n^{i,\star,N}, f(Y_n^{i,N} - Y_n^{j,N}) - f(Y_n^{i,\star,N} - Y_n^{j,\star,N})\big\rangle\Big] \\
&+ \frac{2r}{N}\sum_{j=1}^N \mathbb{E}\Big[|f_\sigma(Y_n^{i,N} - Y_n^{j,N}) - f_\sigma(Y_n^{i,\star,N} - Y_n^{j,\star,N})|^2\Big] \\
&\leq Ch + C\mathbb{E}\big[|X_{t_n}^{i,N} - \hat{X}_n^{i,N}|^2\big] + C\mathbb{E}\big[|Y_n^{i,N} - Y_n^{i,\star,N}|^2\big] \leq Ch,
\end{aligned}$$

where we used (4.53) and $\mathbb{E}\big[|Y_n^{i,N} - Y_n^{i,\star,N}|^2\big] \leq (1 + Ch)\mathbb{E}\big[|X_{t_n}^{i,N} - \hat{X}_n^{i,N}|^2\big]$.

□

4.4.6 Proof of Theorem 4.2.19: Ergodicity for the SSM

Proof of Theorem 4.2.19. Using the notations of Theorem 4.2.19 and Section 4.4.5, and recalling the results in (4.38) and (4.39), for all $i \in [\![1, N]\!]$, $n \in [\![0, M-1]\!]$, we have

$$\begin{aligned}
&\mathbb{E}\big[\,|Y_n^{i,X,N} - Y_n^{i,Z,N}|^2\big] \\
&\quad \leq \frac{2h}{N}\sum_{j=1}^N \mathbb{E}\big[\,\langle Y_n^{i,X,N} - Y_n^{i,Z,N}, f(Y_n^{i,X,N} - Y_n^{j,X,N}) - f(Y_n^{i,Z,N} - Y_n^{j,Z,N})\rangle\big] \\
&\quad + \mathbb{E}\big[|\hat{X}_n^{i,N} - \hat{Z}_n^{i,N}|^2\big] + 2h\mathbb{E}\big[\langle Y_n^{i,X,N} - Y_n^{i,Z,N}, u(Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N}) - u(Y_n^{i,Z,N}, \hat{\mu}_n^{Y,Z,N})\rangle\big] \\
&\quad \leq |\hat{X}_n^{i,N} - \hat{Z}_n^{i,N}|^2 + h(4L_{(f)}^{(1),+} + 2L_{(u\sigma)}^{(1)} + 2L_{(u\sigma)}^{(2)})\mathbb{E}\big[\,|Y_n^{i,X,N} - Y_n^{i,Z,N}|^2\big],
\end{aligned}$$

and therefore

$$\mathbb{E}\big[\,|Y_n^{i,X,N} - Y_n^{i,Z,N}|^2\big] \leq \mathbb{E}\big[\,|\hat{X}_n^{i,N} - \hat{Z}_n^{i,N}|^2\big]\frac{1}{1 - h(4L_{(f)}^{(1),+} + 2L_{(u\sigma)}^{(1)} + 2L_{(u\sigma)}^{(2)})}. \tag{4.57}$$

Next, we consider

$$\begin{aligned}
\mathbb{E}\big[\,|\hat{X}_{n+1}^{i,N} - \hat{Z}_{n+1}^{i,N}|^2\big] &= \mathbb{E}\Big[\,\big|Y_n^{i,X,N} + b(t_n, Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N})h + \overline{\sigma}(t_n, Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N})\Delta W_n^i \\
&\qquad - Y_n^{i,Z,N} - b(t_n, Y_n^{i,Z,N}, \hat{\mu}_n^{Y,Z,N})h - \overline{\sigma}(t_n, Y_n^{i,Z,N}, \hat{\mu}_n^{Y,Z,N})\Delta W_n^i\big|^2\Big] \\
&= \mathbb{E}\big[\,|Y_n^{i,X,N} - Y_n^{i,Z,N}\big|^2 + \big|\overline{\sigma}(t_n, Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N}) - \overline{\sigma}(t_n, Y_n^{i,Z,N}, \hat{\mu}_n^{Y,Z,N})\big|^2 h\big] \\
&\quad + h^2\mathbb{E}\big[\,|b(t_n, Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N}) - b(t_n, Y_n^{i,Z,N}, \hat{\mu}_n^{Y,Z,N})|^2\big]
\end{aligned}$$

$$
\begin{aligned}
&+ 2h\mathbb{E}\big[\,\langle Y_n^{i,X,N} - Y_n^{i,Z,N}, b(t_n, Y_n^{i,X,N}, \hat{\mu}_n^{Y,X,N}) - b(t_n, Y_n^{i,Z,N}, \hat{\mu}_n^{Y,Z,N})\rangle\big] \\
&\leq \mathbb{E}\big[\,|\hat{X}_n^{i,N} - \hat{Z}_n^{i,N}|^2\big] \\
&+ \mathbb{E}\big[\,|Y_n^{i,X,N} - Y_n^{i,Z,N}|^2\big]\Big(h\big(4L_{(f)}^{(1),+} + 2L_{(u\sigma)}^{(1)} + 2L_{(u\sigma)}^{(2)} + 2L_{(b)}^{(2)} + 2L_{(b)}^{(3)}\big) + 2L_{(b)}^{(1)}h^2\Big),
\end{aligned}
$$

where in the last inequality we used the results above, (4.39) and Cauchy–Schwarz inequality. Substituting (4.57) into the last inequality yields the result. □

Chapter 5

Numerical analysis for Langevin type McKean-Vlasov SDEs on variation processes and weak convergence

5.1 Introduction

In this chapter, for $t \geq 0$, we consider a class of McKean–Vlasov Stochastic Differential Equations (MV-SDE), specifically the one-dimensional mean-field Langevin (MFL) equation:

$$X_t = \xi - \int_0^t \Big(\nabla U(X_s) + \nabla V * \mu_s(X_s) \Big) \mathrm{d}s + \sigma W_t, \tag{5.1}$$

where μ_t is the law of X_t, $\sigma \in \mathbb{R}$ and $\xi \in L^p(\Omega, \mathbb{R})$ for some given $p \geq 2$ (i.e. the initial state is an $\mathcal{F}_0$-measurable random variable with finite p-th moments). Following a statistical physics interpretation [40], the map $U : \mathbb{R} \to \mathbb{R}$ is labelled the confining potential and $V : \mathbb{R} \to \mathbb{R}$ the interaction potential. $*$ denotes the usual convolution operator given by $(f * \mu)(\cdot) = \int_{\mathbb{R}} f(\cdot - y)\mu(\mathrm{d}y)$ for some given integrable function $f : \mathbb{R} \to \mathbb{R}$. Of particular interest is the equation's stationary distribution μ^* and specifically how to efficiently generate samples from it. The latter problem has garnered additional interest due to its role in the study of training neural networks (via stochastic gradient descent algorithms) in the mean-field regime [51, 52, 55, 86, 117, 129]. We consider the case where the functions U, V satisfy suitable regularity and convexity assumptions (Assumption 5.2.1), thus the process described in (5.1) admits a unique stationary distribution μ^* (e.g., [40, 49]) with a well-known implicit form satisfying

$$\mu^*(x) \propto \exp\Big(- \frac{2}{\sigma^2} U(x) - \frac{2}{\sigma^2} \int_{\mathbb{R}} V(x-y)\mu^*(\mathrm{d}y) \Big). \tag{5.2}$$

The mainstream method to sample from μ^* is to simulate the weakly-interacting N-particle SDE system (IPS) that approximates (5.1) in the mean-field limit. That is, the law of the solution to (5.1) is approximated, for some sufficiently large N, by the empirical distribution associated with the IPS $(\boldsymbol{X}_t^N)_{t \geq 0} \coloneqq (X_t^{1,N}, \dots, X_t^{N,N})_{t \geq 0}$ with components defined for $i \in \{1, \dots, N\}$ by

$$X_t^{i,N} = \xi^i - \int_0^t \Big(\nabla U(X_s^{i,N}) + \frac{1}{N} \sum_{j=1}^N \nabla V(X_s^{i,N} - X_s^{j,N}) \Big) \mathrm{d}s + \sigma W_t^i, \tag{5.3}$$

where $(\xi^i, W^i)_{i \in \{1,\dots,N\}}$ are a collection of i.i.d. copies of (ξ, W). We briefly review some important results providing quantitative convergence guarantees for the approximation of (5.1) through its IPS. Broadly speaking, the error of approximating (5.1) by an N-particle system (5.3) has been widely studied in the literature of Propagation of Chaos (PoC), stemming from the seminal works [118, 133]. In a nutshell, the accuracy of the N-particle approximation is known to behave like $\mathcal{O}(1/N)$ in the *squared L^2-norm* (under Lipschitz type assumptions on the interaction and confining potentials). These quantitative convergence results are often referred to as strong PoC [29, 118, 133]. More recently, [101, Theorem 2.2] shows that the PoC convergence rate for models of the form (5.1) (in relative entropy and finite time) can be improved from $\mathcal{O}(1/N)$ to $\mathcal{O}(1/N^2)$ (under certain smallness conditions), with [101, Example 2.8] showing that the rate is optimal. We refer the reader to the introductions of the papers [101, 102] and review articles [41, 41] for a holistic discussion on this topic.

When the PoC holds up to $T = \infty$, e.g., [40, 102], it is called uniform in time PoC – see our Proposition 5.2.2 for a formulation in the strong sense. Uniform in time PoC results for the model (5.1) have received significant attention in the last few years; see e.g. [40, 46, 65, 102] and references cited therein.

In the context of quantitative weak PoC results (i.e. the absolute difference between expectations), we refer to the recent work [42] establishing finite-time higher order weak PoC results via techniques from differential calculus on the space of measures along with the study of Kolmogorov backward PDEs written on the Wasserstein space. Also, [81, Theorem 3.1] establishes a finite-time weak PoC result of rate $\mathcal{O}(1/N)$ based on a more classical approach, via a Talay–Tubaro expansion and an analysis of a Kolmogorov backward PDE associated with the whole IPS.

Once (5.3) and a PoC rate is established, the sampling from μ^* (5.2) is obtained via discretization of (5.3) with a convenient numerical scheme (also called numerical integrator or sampler). There is a growing body of contributions on the topic of sampling from the (overdamped) MFL stationary distribution using this method [97, 132]. Although said works provide a variety of quantitative PoC rate type results (under a variety of conditions), *the time discretization schemes used are all of Euler type* – see [132, Theorem 3 and 4] or [97, Section 4] – the final error rates or sampling guarantees encase a leading order 1 dependence on the time-discretization stepsize. We also mention two recent contributions on unadjusted Hamiltonian Monte Carlo for the simulation of the (underdamped or kinetic) MFL [30, 31]. The main emphasis of the present work is to improve the weak convergence order (to the stationary distribution) of the standard Euler scheme using a non-Markovian version of it.

We briefly mention some (recent) contributions on numerical schemes for MV-SDEs. The seminal works [28, 29] investigate the convergence rates of the particle system's and Euler scheme's approximation accuracy of the cumulative distribution (in L^1-norm) for the Burger's type MV-SDE using density estimates or using a Malliavin calculus approach [8]. In the context of finite time-horizon simulation there are many recent contributions (e.g., Euler and Milstein schemes) focusing on the approximation error stemming from the time discretization of the IPS [13, 23, 48, 49, 60]. Cubature type algorithms, a class of weak approximation algorithms, for (Stratonovich) MV-SDEs have been proposed in [44, 54, 122]. Lastly, we mention [5] and its references for numerical methods to approximate MV-SDEs directly fully avoiding the IPS approach.

Motivation, weak convergence schemes and the non-Markovian Euler scheme for SDEs. The classical Euler scheme is an easy to implement and ubiquitous method for the numerical approximation of solutions to SDEs. In the classic overdamped Langevin context, i.e. if one sets $\nabla V = 0$ in (5.3), the Euler scheme attains a strong and weak rate $\mathcal{O}(h)$ (where h is the time-discretization stepsize) in either finite or infinite time horizon [120, 124]. Informally, under certain conditions the Euler scheme's weak error at time $T = Mh$ and stepsize $h > 0$ ($M \in \mathbb{N}$) can be expressed (see, Talay–Tubaru [120, 136]) in the form

$$\text{Weak Error}^{\text{Euler}}(h;T) = C_T h + \mathcal{O}(h^2) \qquad \text{where} \qquad \lim_{T\to\infty} C_T = \text{Const} > 0. \tag{5.4}$$

In [103] (for general dimensions), setting ∇V=0 in (5.1) and denoting by $X^h_{t_m}$ the numerical approximation to X_{t_m}, the following variant of the Euler scheme was analyzed (for any $m \in \mathbb{N}$,

$t_m = mh$):

$$X^h_{t_{m+1}} = X^h_{t_m} - \nabla U(X^h_{t_m})h + \frac{\sigma}{2}(\Delta W_m + \Delta W_{m+1}) \text{ with } \Delta W_{m+1} = W_{t_{m+1}} - W_{t_m}. \quad (5.5)$$

This scheme is called the *Leimkuhler–Matthews method* or the *non-Markovian Euler scheme* since $X^h_{t_{m+1}}$ is computed using the current and past Brownian increments, ΔW_{m+1} and ΔW_m. It is shown in [103] that (5.4) holds for $T < \infty$ (with a different C_T), but as $T \to \infty$ one has

$$\lim_{T\to\infty} C_T = 0 \quad \Rightarrow \quad \lim_{T\to\infty} \text{Weak Error}^{\text{non-Mark. Euler}}(h;T) = \mathcal{O}(h^2),$$

and thus the *non-Markovian Euler scheme* is a weak order-2 method as $T \to \infty$. An intuition behind the result is offered by [139] through the concept of *Postprocessed integrators*. There, (5.5) is re-written as a two-step method where the second step corrects the $\mathcal{O}(h)$ bias of the first step in such a way that in the long time limit the weak error is $\mathcal{O}(h^2)$; see [139, Equation (2.4)].

The focus of this work is to study the non-Markovian Euler scheme (5.5) in the context of the overdamped MFL dynamics (5.1) as way to simulate (5.2) via a higher-order weak scheme. As described, the MFL (5.1) is first approximated by the IPS (5.3) and then the IPS is time-discretized using the non-Markovian scheme (5.5).

In terms of proof methodologies for weak errors, for either SDEs or MV-SDEs, it is well known since the seminal work [136], that weak error analysis can be tackled via the Kolmogorov backward PDE [120, Chap. 2]. This approach for MV-SDEs and the IPS is well-reviewed in [81] and is the approach we take. An alternative method is the use of Malliavin Calculus [10, 96, 122] as it offers a path to completely bypass the analysis of the Kolmogorov backward equation. In addition, we highlight the classic backward error analysis approach drawing on Itô- or Stratonovich–Taylor expansions [70, 95, 120]. In a different spirit, results showing density approximations, via Fokker–Plank PDE analysis or Malliavin calculus have been obtained in [11, 12, 28, 29].

The scheme's convergence results for the MFL class. The main contribution of this paper is to establish the techniques needed to understand and quantify the weak errors for the non-Markovian Euler scheme applied to (5.3) *in a way such that the convergence rate is independent of the number of particles N in the IPS.* In terms of the convergence results, for any $T > 0$, the weak approximation error for smooth test functions $g : \mathbb{R}^N \to \mathbb{R}$ (satisfying Assumption 5.3.3) is defined as

$$\text{Weak Error} := \mathbb{E}\big[g\big(\boldsymbol{X}^N_T\big)\big] - \mathbb{E}\big[g\big(\boldsymbol{X}^{N,h}_T\big)\big], \qquad \text{with } T = Mh \text{ for } h > 0, M \in \mathbb{N}, \quad (5.6)$$

where $\boldsymbol{X}^N_T$ denotes the solution of (5.3) and $\boldsymbol{X}^{N,h}_T$ denotes the $\mathbb{R}^N$-valued output of M-steps ($T = Mh$) of the non-Markovian Euler scheme applied to (5.3) (and explicitly given in (5.13)). Informally, our main result (Theorem 5.3.8) states that

$$\begin{aligned} \big|\, \mathbb{E}[g(\boldsymbol{X}^N_T)] - \mathbb{E}[g(\boldsymbol{X}^{N,h}_T)] \,\big| &\le C_0(T)h + Kh^{3/2} \\ &\quad \text{where} \quad |C_0(T)| \le K\exp(-\lambda_0 T) + Kh^{1/2}, \end{aligned} \quad (5.7)$$

for some positive constants λ_0, K independent of h, T, M and N. In other words, the scheme is uniformly (in the number of particles) of weak order $\mathcal{O}(h^{3/2})$ as $T \to \infty$ and has standard weak order $\mathcal{O}(h)$ for $T < \infty$. We provide an in-depth technical discussion (Remarks 5.3.4 and 5.6.2) on the missing $h^{1/2}$ order in the convergence rate when comparing this to the second order weak convergence result obtained in [103, 139].

A secondary contribution of this work (Proposition 5.2.3 below), is the clarification of the nuance that the higher-order weak convergence of the non-Markovian Euler scheme comes at the cost of having a *uniform in time* strong L^2-convergence order of $\mathcal{O}(h^{1/2})$. It is lower than the $\mathcal{O}(h)$ strong L^2-convergence of the classical Euler scheme.

Methodology, contributions and existing literature. The main methodology we follow is an involved variant of the Talay–Tubaro approach to the study of weak convergence [120, 136], which is also the approach used by [103] (for SDEs) and [42, 81] (to study weak quantitative PoC over $T < \infty$). At its core, this method relates the expectations appearing in the definition of the weak error (5.6) to a Kolmogorov backward PDE with terminal condition given by the function g – see the PDE (5.17) linked to the driving SDE (5.3), which in flow form is given in (5.16) and written as $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\geq t\geq 0}$, for $\boldsymbol{x} \in \mathbb{R}^N$ denoting the starting point of the IPS at time $t \geq 0$. This analysis involves establishing certain bounds for the variation processes of the IPS (5.3), or more precisely for the flow process $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\geq t\geq 0}$. In this regard, our approach is closest in spirit to that of [81] as we work with Kolmogorov backward PDEs connected to the full particle system. However, our focus is on the *time-discretization* analysis *uniformly in N over infinite time* as opposed to the weak error analysis in the number of particles [42, 58, 81] (these works consider $T < \infty$ and deal only with the continuous time IPS equation). Our case has therefore fundamental added complexities in relation to the mentioned works, as we require estimates which are not only uniform in N but also in time.

Technical challenges. As mentioned, (5.7) is proved via a Talay–Tubaro type expansion which, for the case of the non-Markovian Euler scheme, is an involved collection of terms arising from Taylor expansions using the Kolmogorov backward equation associated with the flow equation for the IPS (5.3). This expansion has been given in [103, Equation (3.17)] for SDEs and we recast it to our setting (in Lemma 5.3.7 and also in Section 5.6 and in Appendix 6.4.4). In the following, we highlight several technical elements of our work and point out crucial differences to [81, 103]:

(i) The terms arising from said Taylor expansions involve up to 6-th order (cross)-derivatives in the spatial variable of the solution to the Kolmogorov backward PDE (see our Assumption 5.3.3 and Lemma 5.5.2). Critically, the usual pointwise estimates from PDE theory e.g. [103, Equation (3.3)] (or [135, 139]), do not directly apply to our case as those would not be independent of the number of particles. It is not clear how the right-hand side in [103, Equality (3.3)] depends on the problem's dimension. Therefore, we derive suitable new estimates in L^p-norm of the solution to the Kolmogorov backward equation that decay exponentially in time in a non-explosive way in N (see, Lemma 5.5.2 for an intermediate pointwise result and Lemma 5.5.3 for the final L^p-estimates used to show the main theorem) – this is in stark contrast to [81] (in particular their Appendix B) which establishes pointwise estimates. For clarity, the derivatives of the solution to the Kolmogorov backward equation are intrinsically linked to certain moment estimates for variation processes of the IPS' flow SDE $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\geq t}$ (see Lemma 5.5.2). In order to control the time dependence of the implied constants for the moment estimates of the variation processes, a careful analysis of the terms involving the convolution kernel is needed. Consequently, we are only able to establish the bounds in Lemma 5.5.3 in an L^p-sense. The estimates of Lemma 5.5.3 are obtained in [81] in a pointwise sense but crucially without the exponential time decay component (see RHS of (5.114) and (5.115)); their analysis is carried out in finite time for which this issue is not a concern.

   Further, our analysis requires to study the time regularity of the solution to the Kolmogorov backward PDE, which needs estimates for the differences of the IPS' flow SDE process, $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\geq t}$, concretely differences of the form $|\boldsymbol{X}_u^{t,\boldsymbol{x},N} - \boldsymbol{X}_u^{s,\boldsymbol{x},N}|$ for $0 \leq t \leq s \leq u$.

   Lastly, it is noteworthy to highlight that the weak error test function g in (5.6) depends on the whole IPS (as in [81] but not as in [20][1]) making the analysis much more involved.

(ii) Before addressing the estimates for the Kolmogorov backward equation, we derive L^p-norm estimates for the *variation processes* of the flow of the IPS (decaying over time uniformly over N) up to general n-order (although only 6-orders are needed). This is done in Section 5.4. Our approach shows a way to analyze the terms arising from the

[1] The analysis of the Kolmogorov backward PDE over a single-particle $X^{i,N}$ instead of $\boldsymbol{X}^N = (X^{1,N}, \ldots, X^{N,N})$ on the test function g enables an advantageous simplifying decoupling effect at a later point; such is not the case here.

interacting kernel and their recurring contributions across the different orders of the variation processes and *across different particle indices* – compare (5.32) and (5.33) for the first order case and check Lemma 5.4.7 for general cases. A further crucial component of the analysis is to establish the correct decay in terms of number of particles across different orders of variation processes. This, in particular, subsequently allows to control the growth of the derivatives of the solution to the Kolmogorov backward equation. The depth of the analysis is well beyond the results of [81] who carry out a related approach over finite time (or in [58] over the torus over infinite time horizon).

(iii) Regarding the strong convergence analysis of the non-Markovian Euler scheme, the one-timestep error propagation analysis requires analysing 3 sub-steps of the scheme which is in contrast to such analysis for the standard Euler scheme (loosely speaking, only 1 sub-step is analyzed) and thus, the analysis is lengthier than usual – see e.g. the proof of Proposition 5.2.3.

Gaps, conjectures and pathways for further study. Our analysis addresses MFL dynamics in $\mathbb{R}$ through $\mathbb{R}^N$-valued IPS. It is believed that our results could be established in the multi-dimensional case $d > 1$ if the measure dependence in (5.1) was of the form $\mathbb{E}[h(X_t)]$ instead of an interaction kernel. The tools and techniques we have employed to show our main result do not use measure derivatives (due to the simplicity of the underlying model) or concentration inequalities. It seems possible, although presently unclear, that drawing on Log-Sobolev inequalities and related machinery would provide means to lift the technical constraint in dimension arising from the convolution term. Overall, to establish higher L^p-moments for the variation processes in Section 5.4, we require the *symmetrization trick* (in Remark 5.4.3 to deal with (5.25)) and an inequality of the type $(|x|^{p-2}x - |y|^{p-2}y) \cdot (x - y) \geq 0$ to hold – this would not naturally hold in the $d > 1$ case (this issued is hinted at in Chapter 3 and appears explicitly in Chapter 4). A possible alternative methodology to establish our main result is the postprocessed integrators machinery presented in [139]. To use it, we benefit from all the results shown in this work. However, some others would still need to be established, e.g. one has to derive results that imply Assumption 2.3 or Theorem 4.1 of [139] in the IPS (5.3) setting that remain uniform in the particle number; this is yet to be explored and left for future research. In addition, it would be interesting to see if techniques from Malliavin calculus could be used [10, 96, 122], even in the standard SDE context, to establish the weak convergence results shown in this article. It is interesting to question if the weak results in our manuscript could be extended to the difference between the densities of (5.1) and (5.3) as in [12, Corollary 2.1] (as the diffusion of (5.3) is uniformly elliptic) – in fact, the question is also pertinent in the context of standard SDEs itself (do the results of [103, 139] also hold for densities as in [12]).

Further afield and more broadly is if these results could be established under the setting of common-noise MFL dynamics [108], or in the context of the kinetic/underdamped MFL [45, 79]. It appears to be possible to combine portions of the methodology here with that in [81] to extend their finite-time weak PoC result to the $T \to \infty$ setting. Our work also paves the way to study stochastic gradient descent convergence [97, 132] but using the non-Markovian Euler scheme as the update instead of the standard Euler one.

Paper organization. This paper is organized as follows. In Section 5.2, we state the main assumptions, introduce the non-Markovian scheme and state basic results regarding wellposedness of the underlying model. In Section 5.3, we present our weak error expansions based on the result in [103]. We state the main technical difficulties when applying the scheme to the IPS and explain why we cannot reach weak rate of order 2 in the case for classical SDEs. All proofs of Section 5.2 and 5.3 are postponed to the final part of the paper. Section 5.4 contains the results relating to the analysis of several variation processes, while Section 5.5 contains the decay estimates for the solution to the Kolmogorov backward equation. Section 5.6 contains the proof of the weak error result (of Section 5.3). An illustrative numerical example is provided in Section 5.7.

5.2 Main results

We consider the following one-dimensional MV-SDE, for $t \geq 0$,

$$X_t = \xi - \int_0^t \Big(\nabla U(X_s) + \nabla V * \mu_s(X_s)\Big)\mathrm{d}s + \sigma W_t, \tag{5.8}$$

where $\sigma \in \mathbb{R}$, and $\xi \in L^p(\Omega, \mathbb{R})$ for some given $p \geq 2$. $U : \mathbb{R} \to \mathbb{R}$ is the confining potential and $V : \mathbb{R} \to \mathbb{R}$ is the interaction potential, with $*$ denoting the usual convolution operator where $(f * \mu)(\cdot) = \int_{\mathbb{R}} f(\cdot - y)\mu(\mathrm{d}y)$. We impose the following standard assumptions on U and V.

Assumption 5.2.1. *Let $U : \mathbb{R} \to \mathbb{R}$ and $V : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable functions with globally Lipschitz continuous gradients. Further suppose that*

(1) U is uniformly convex in the sense that there exists $\lambda > 0$ such that for all $x, y \in \mathbb{R}$,

$$\big(\nabla U(x) - \nabla U(y)\big)\big(x - y\big) \geq \lambda |x - y|^2, \tag{5.9}$$

which implies $\nabla^2 U \geq \lambda$.

(2) V is even (thus ∇V is odd), and convex, i.e., for all $x, y \in \mathbb{R}$,

$$\big(\nabla V(x) - \nabla V(y)\big)(x - y) \geq 0,$$

and there exists $K_V > 0$ such that $|\nabla^2 V|_\infty \leq K_V$.

The Interacting particle system (IPS). Define the $\mathbb{R}^N$-valued map B as

$$\mathbb{R}^N \ni \boldsymbol{x} = (x_1, \ldots, x_N) \mapsto B(\boldsymbol{x}) := \big(B_1(x_1, \ldots, x_N), \ldots, B_N(x_1, \ldots, x_N)\big),$$

where

$$B_i(\boldsymbol{x}) = B_i(x_1, \ldots, x_N) := -\nabla U(x_i) - \frac{1}{N}\sum_{j=1}^N \nabla V(x_i - x_j). \tag{5.10}$$

Let (ξ^i, W^i) for $i \in \{1, \ldots, N\}$ be i.i.d. copies of (ξ, W) and define the IPS associated with (5.8) to be

$$X_t^{i,N} = \xi^i + \int_0^t B_i(X_s^{1,N}, \ldots, X_s^{N,N})\mathrm{d}s + \sigma W_t^i, \tag{5.11}$$

$$\boldsymbol{X}_t^N = \boldsymbol{\xi} + \int_0^t B(\boldsymbol{X}_s^N)\mathrm{d}s + \sigma \boldsymbol{W}_t, \tag{5.12}$$

with solution process $(\boldsymbol{X}_t^N)_{t\geq 0} := (X_t^{1,N}, \ldots, X_t^{N,N})_{t\geq 0}$, where we introduced $\boldsymbol{\xi} = (\xi^1, \ldots, \xi^N)$ and $(\boldsymbol{W}_t)_{t\geq 0} := (W_t^1, \ldots, W_t^N)_{t\geq 0}$.

Preliminary results.

The next proposition collects some basic properties of the MFL equation (5.8) and the IPS (5.11).

Proposition 5.2.2. *Let Assumption 5.2.1 hold and let $\xi \in L^p(\Omega, \mathbb{R})$ for some $p \geq 2$. Then the following hold:*

(1) The MV-SDE (5.8) and the IPS (5.11) each admit a unique strong solution. There exist constants $\kappa \in (0, \lambda)$ and $K \geq 0$ (K, κ are independent of t and N) such that for any $t \geq 0$

$$\mathbb{E}\big[\,|X_t|^p\big] \leq K\big(1 + \mathbb{E}[\,|\xi|^p]e^{-p\kappa t}\big) \quad \text{and} \quad \max_{i\in\{1,\ldots,N\}} \mathbb{E}\big[\,|X_t^{i,N}|^p\big] \leq K\big(1 + \mathbb{E}[\,|\xi|^p]e^{-p\kappa t}\big).$$

(2) Uniform propagation of chaos (PoC) holds, i.e., there exists $K \geq 0$ such that for every $N \geq 1$

$$\max_{i\in\{1,\dots,N\}} \sup_{t\geq 0} \mathbb{E}\big[\,|X_t^i - X_t^{i,N}|^2\big] \leq \frac{K}{N},$$

where X^i is the solution of (5.8) *with (ξ, W) replaced by (ξ^i, W^i) (i.e., the so-called non-interacting particle system).*

(3) There exists a unique stationary distribution for (5.8) *and* (5.11)*, denoted by μ^* and $\mu^{N,*}$, respectively. Moreover, $W^{(2)}(\mu_t, \mu^*) \to 0$ and $W^{(2)}(\mu_t^N, \mu^{N,*}) \to 0$ as $t \to \infty$.*

Proof. See Appendix 6.4.1. □

5.2.1 The non-Markovian Euler scheme

Let $h \in (0, 1)$ denote the timestep and take $m \in \{0, \dots, M-1\}$ for a given $M \in \mathbb{N}$. Inspired by [103, Equation (1.7)] (also [139]), we introduce the following non-Markovian Euler scheme

$$X_{t_{m+1}}^{i,N,h} = X_{t_m}^{i,N,h} - \Big(\nabla U(X_{t_m}^{i,N,h}) + \frac{1}{N}\sum_{j=1}^{N} \nabla V(X_{t_m}^{i,N,h} - X_{t_m}^{j,N,h})\Big)h + \frac{\sigma}{2}(\Delta W_m^i + \Delta W_{m+1}^i), \tag{5.13}$$

with $X_{t_0}^{i,N,h} = X_{t_0}^{i,N}$, to approximate the IPS (5.11), where we set the time grid points as $t_m := mh$ up to some time $T := t_M = Mh$, and the random increments as $\Delta W_{m+1}^i = W_{t_{m+1}}^i - W_{t_m}^i$ with $\Delta W_0^i = 0$. In analogy to the IPS, we write for the solution process $(\boldsymbol{X}_{t_m}^{N,h})_{m\in\{0,\dots,M\}} := (X_{t_m}^{1,N,h}, \dots, X_{t_m}^{N,N,h})_{m\in\{0,\dots,M\}}$. We aim to analyze the behaviour of this scheme as $T \to \infty$.

The following result establishes some fundamental properties for the non-Markovian Euler scheme (5.13): moment estimates and L^2-strong convergence (we were unable to find a proof in the literature regarding the L^2-strong convergence for this scheme (even for SDEs) and we thus provide it here). Critically, the moment estimates obtained are independent of the time horizon (i.e., the constant K appearing below is independent of T). Lastly, as in [103] or [139], the result holds for a sufficiently small timestep h.

Proposition 5.2.3. *Let Assumption 5.2.1 hold, let $\xi \in L^p(\Omega, \mathbb{R})$ for some $p \geq 2$, and let $N, M \in \mathbb{N}$. Then the following statements hold for the process defined in* (5.13).

(1) There exist $\kappa, K > 0$ (both are independent of h, T, M and N) such that for any sufficiently small timestep $0 < h \ll \min\{1/2\lambda, 1\}$ and $m \in \{0, \dots, M\}$,

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\big[\,|X_{t_m}^{i,N,h}|^p\big] \leq K\big(1 + \mathbb{E}\big[\,|\xi|^p\big]e^{-\kappa t_m}\big).$$

(2) L^2-strong error. There exists $K > 0$ independent of h, T, M and N such that for any sufficiently small timestep $0 < h \ll \min\{1/2\lambda, 1\}$,

$$\max_{i\in\{1,\dots,N\}} \max_{m\in\{0,\dots,M\}} \mathbb{E}\big[\,|X_{t_m}^{i,N} - X_{t_m}^{i,N,h}|^2\big] \leq Kh,$$

where $X^{i,N}$ and $X^{i,N,h}$ are the processes defined in (5.11) *and* (5.13) *respectively.*

Proof. See Appendix 6.4.2. □

5.3 The weak error expansion

Before presenting the framework for the weak error analysis and our main result, we require the following definition which will be helpful to characterize and analyze the higher-order variation processes.

Definition 5.3.1. *Let $n, m, N \in \mathbb{N}$ with $N \gg n, m$ be given integers. Define the set of multi-indices*

$$\Pi_n^N := \big\{\gamma = (\gamma_1, \ldots, \gamma_n) : \gamma_i \in \{1, \ldots, N\} \text{ for all } i \in \{1, \ldots, n\}\big\},$$

with $\Pi_0^N := \emptyset$ denoting the empty set. For a subset $\overline{\gamma} \subseteq \gamma$, let $|\overline{\gamma}|$ be its length.

For a given $\gamma \in \Pi_n^N$, let $\hat{\gamma}$ be a set of length N counting the frequency of each $j \in \{1, \ldots, N\}$ in γ, and define $\hat{\mathcal{O}}(\gamma) := \{\text{number of non-zero values in } \hat{\gamma}\}$. We also use the following two operations for the multi-indices: for $\gamma^{(1)} \in \Pi_n^N$ and $\gamma^{(2)} \in \Pi_m^N$, the difference $\gamma^{(1)} \setminus \gamma^{(2)} \in \Pi_k^N$ is specified through the counting set (as $|\hat{\gamma}^{(1)}| = |\hat{\gamma}^{(2)}| = N$)

$$\hat{\gamma}^{diff} := \{\max\{\hat{\gamma}_1^{(1)} - \hat{\gamma}_1^{(2)}, 0\}, \ldots, \max\{\hat{\gamma}_N^{(1)} - \hat{\gamma}_N^{(2)}, 0\}\},$$

where k is the number of non-zero elements in[2] *$\hat{\gamma}^{diff}$.*

The union is defined by

$$\gamma^{(1)} \bigcup \gamma^{(2)} = (\gamma_1^{(1)}, \ldots, \gamma_n^{(1)}, \gamma_1^{(2)}, \ldots, \gamma_m^{(2)}) \in \Pi_{n+m}^N.$$

For two sets of multi-indices, Π_n^N and Π_m^N, with $n \neq m$, the union is defined as

$$\Pi_n^N \bigcup \Pi_m^N = \{\gamma : \gamma \in \Pi_n^N \text{ or } \gamma \in \Pi_m^N\}.$$

The shuffle product (see, [66]) for two multi-indices $\gamma^{(1)} \in \Pi_n^N$ and $\gamma^{(2)} \in \Pi_m^N$ is denoted by $\gamma^{(1)} \sqcup\!\sqcup \gamma^{(2)}$. We write $\gamma^{(1)} \simeq \gamma^{(2)}$ if $m = n$ and there exists a permutation $\pi \in S_n$ such that $(\gamma_1^{(1)}, \ldots, \gamma_n^{(1)}) = (\gamma_{\pi(1)}^{(2)}, \ldots, \gamma_{\pi(n)}^{(2)})$, where S_n is the symmetric group on the set $\{1, \ldots, N\}$.

Example 5.3.2. *We present the following examples to make Definition 5.3.1 clearer: For $N = 3$, we have:*

$$\Pi_1^3 = \big\{(1), (2), (3)\big\}, \quad \Pi_2^3 = \big\{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\big\}.$$

For $N \geq 3$, let $(1,1,1)$ (i.e., $\hat{\gamma} = (3, 0, \ldots, 0)$, $|\hat{\gamma}| = N$), $(1,1,2)$ (i.e., $\hat{\gamma} = (2, 1, 0, \ldots, 0)$) , $(1,2,2) \in \Pi_3^N$ (i.e., $\hat{\gamma} = (1, 2, 0, \ldots, 0)$). Then we have

$$\hat{\mathcal{O}}\big((1,1,1)\big) = 1, \quad \hat{\mathcal{O}}\big((1,1,2)\big) = 2, \quad \hat{\mathcal{O}}\big((1,2,3)\big) = 3.$$

With regards to the set operations, we present an example for $\alpha = (1,2,3,3), \beta = (3,5)$:

$$\alpha \setminus \beta = (1,2,3), \quad \alpha \bigcup \beta = (1,2,3,3,3,5).$$

In regards to the shuffle product, take for example $\alpha = (1,2)$, $\beta = (3)$, then

$$\alpha \sqcup\!\sqcup \beta = \big\{(3,1,2), (1,3,2), (1,2,3)\big\},$$

where we understand $(1,2,3) \simeq (1,3,2) \simeq (2,1,3) \simeq (2,3,1) \simeq (3,1,2) \simeq (3,2,1)$.

The following summation of the elements in an $N \times N$ matrix with elements $a_{i,j}$ for $i, j \in \{1, \ldots, N\}$, demonstrates the meaning of $\hat{\mathcal{O}}(\cdot)$ in the definition for the case $n = 2$:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_{i,j} = \sum_{\gamma \in \Pi_2^N} a_{\gamma_1, \gamma_2} = \sum_{\gamma \in \Pi_2^N, \hat{\mathcal{O}}(\gamma) = 1} a_{\gamma_1, \gamma_2} + \sum_{\gamma \in \Pi_2^N, \hat{\mathcal{O}}(\gamma) = 2} a_{\gamma_1, \gamma_2},$$

where we partitioned the summation into diagonal and off-diagonal elements.

For our analysis, we impose the following assumption.

[2] In this work the difference $\gamma^{(1)} \setminus \gamma^{(2)}$ is *never* used directly, only the quantity $\hat{\mathcal{O}}(\gamma^{(1)} \setminus \gamma^{(2)})$ will be used and thus we require only $\hat{\gamma}^{\text{diff}}$. For practical purposes, one can think of the γ as being ordered vectors (in increasing order) – see further Example 5.3.2.

Assumption 5.3.3. *Assumption 5.2.1 holds. Further, suppose that:*

(1) The potentials $U, V \in \mathcal{C}^7(\mathbb{R})$, and all derivatives of $\nabla U, \nabla V$ are uniformly bounded. (This in particular implies that $\nabla U, \nabla V$ are Lipschitz continuous.)

(2) The convexity parameters λ, K_V satisfy $\lambda \geq 7K_V$.

(3) Let $N \in \mathbb{N}$ with $N \gg 6$. For any $n \in \{1, \ldots, 6\}$ and $(\gamma_1, \ldots, \gamma_{|\gamma|}) = \gamma \in \bigcup_{k=1}^{n} \Pi_k^N$, with integers $\gamma_j \in \{1, \ldots, N\}$, the function $g : \mathbb{R}^N \to \mathbb{R}$, satisfies $|\partial^{|\gamma|}_{x_{\gamma_1},\ldots,x_{\gamma_{|\gamma|}}} g|_\infty = \mathcal{O}(N^{-\hat{\mathcal{O}}(\gamma)})$, with an implied constant independent of N.

(4) The function g and its derivatives up to order n are Lipschitz. (Note that item (3) implies that the function g and its derivatives up to order $n-1$ are Lipschitz.)

Remark 5.3.4. *(On point (1) of Assumption 5.3.3): Our analysis has a small reduction regarding the order of regularity when compared to [103] who require the drift of the underlying model to be 8-times continuously differentiable. In [103], in an intermediate step, weak convergence of order $\mathcal{O}(h)$ is first established for the non-Markovian Euler scheme (which only requires the drift to be 6-times continuously differentiable).*

The intermediate result is then employed to show weak convergence of a certain term where the test function involves 2nd order derivatives of the drift and the solution to the Kolmogorov PDE. In our notation this test function is denoted by L and is precisely defined in (5.20) *below. Since L already involves second order derivatives (of the potentials), the higher regularity is needed. We are not able to show that L possesses sufficient regularity properties (i.e., item (3) in Assumption 5.3.3) to apply the intermediate result concerning the first order weak convergence and resort to a strong convergence result instead (as a consequence we only derive the weak convergence rate 1.5 instead of 2); see Remark 5.6.2 for full details.*

Remark 5.3.5 (On point (2) of Assumption 5.3.3)**.** *The constraint $\lambda \geq 7K_V$ is not sharp and relates to the need of sufficient convexity as we analyze the n-th order variation processes (for the process defined in* (5.16)*). For instance, Lemma 5.4.5 establishes moment bounds of the* 2*nd variation processes of* (5.16) *and for it we require $\lambda > (2 + 1/N)K_V$ in* (5.78)*. For the moment bounds of the* 6*-th variation processes (in Lemma 5.4.7), the requirement ends up being $\lambda > (6 + 1/N)K_V$ and we streamline it to $\lambda \geq 7K_V$. This is a technical constraint of the analysis stemming from the interplay between the confinement potential and the interaction kernel functions and will be made more precise in the proofs of the lemmas. Lastly, this assumption is not comparable to those in [103] or [139] as neither have interaction kernels; only confining potentials (see* (5.8)*).*

Remark 5.3.6 (On point (3) of Assumption 5.3.3)**.** *Typical examples for g satisfying the above assumptions would be $g(\boldsymbol{x}) = \tilde{g}\left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right)$, for some functions $f, \tilde{g} : \mathbb{R} \to \mathbb{R}$ that are sufficiently often differentiable with bounded derivatives. For instance, consider the case $n = 3$ and let $\gamma \in \Pi_1^N \bigcup \Pi_2^N \bigcup \Pi_3^N$ for which $\hat{\mathcal{O}}(\gamma) = 3$ (e.g., $\gamma = (1, 2, 3)$). Therefore, our assumption requires $|\partial^3_{x_1,x_2,x_3} g|_\infty = \mathcal{O}(N^{-3})$, which is satisfied for regular enough functions $\tilde{g}$ and f. As a further example, consider $\gamma = (1, 1, 3)$ for which $\hat{\mathcal{O}}(\gamma) = 2$, and hence, $|\partial^3_{x_1,x_1,x_3} g|_\infty = \mathcal{O}(N^{-2})$. As a last example, if $f = id$ then for any $|\gamma|$-order derivative, one has automatically $|\partial^{|\gamma|}_{x_{\gamma_1},\ldots,x_{\gamma_{|\gamma|}}} g|_\infty = \mathcal{O}(N^{-|\gamma|})$.*

Weak error. We define the weak error induced by the non-Markovian scheme (see, (5.13)) approximating the IPS $\boldsymbol{X}_T^N$ (5.11) as follows: Let the test function $g : \mathbb{R}^N \to \mathbb{R}$ satisfy Assumption 5.3.3. For any $T > 0$, the weak approximation error satisfies

$$\text{Weak Error} := \mathbb{E}\big[g\big(\boldsymbol{X}_T^N\big)\big] - \mathbb{E}\big[g\big(\boldsymbol{X}_T^{N,h}\big)\big]. \tag{5.14}$$

The Kolmogorov backward equation for the flow and the weak error expansion. We study the weak error (5.14) via an analysis of the Kolmogorov backward equation for the stochastic flow equation associated with the dynamics of (5.12). Concretely, let $N \in \mathbb{N}$, $\boldsymbol{x} \in \mathbb{R}^N$, and $0 \le t \le s$. Then we introduce $\boldsymbol{X}_s^{t,\boldsymbol{x},N} = (X_s^{t,x_1,1,N}, \dots, X_s^{t,x_N,N,N})$, where

$$X_s^{t,x_i,i,N} = x_i + \int_t^s B_i(X_u^{t,x_1,1,N}, \dots, X_u^{t,x_N,N,N})\mathrm{d}u + \sigma(W_s^i - W_t^i), \quad i \in \{1, \dots, N\}, \quad (5.15)$$

$$\boldsymbol{X}_s^{t,\boldsymbol{x},N} = \boldsymbol{x} + \int_t^s B(\boldsymbol{X}_r^{t,\boldsymbol{x},N})\mathrm{d}r + \sigma(\boldsymbol{W}_s - \boldsymbol{W}_t). \quad (5.16)$$

The wellposedness of (5.15) or (5.16) under our assumptions is clear. The generator for (5.16) is defined by

$$\mathcal{L}_N = \sum_{i=1}^N B_i \partial_{x_i} + \frac{1}{2}\sigma^2 \partial^2_{x_i,x_i},$$

where B_i is the drift term for the i-th particle as in (5.15). Now, for $u : [0,T] \times \mathbb{R}^N \to \mathbb{R}$, we introduce the Kolmogorov backward equation:

$$\partial_t u + \mathcal{L}_N u = 0, \quad t \in [0,T), \quad u(T, \boldsymbol{x}) = g(\boldsymbol{x}), \quad (5.17)$$

for the above test function $g : \mathbb{R}^N \to \mathbb{R}$. Under the above assumptions, the solution of the above PDE is given by the Feynman-Kac formula [73, 120]:

$$u(t, \boldsymbol{x}) = \mathbb{E}\left[g(\boldsymbol{X}_T^N) | X_t^{i,N} = x_i, i \in \{1, \dots, N\}\right]. \quad (5.18)$$

To analyze the weak error (5.14), we need to expand it akin to a Talay–Tubaru expansion [136] (see also [81, 120]), but with certain fundamental differences. The following expansion is shown in [103].

Lemma 5.3.7 (Weak error expansion, Equation (3.17) in [103])**.** *Let Assumption 5.3.3 hold. Then the following expansion of the weak error holds for the processes defined in* (5.11) *and* (5.13)*: for any sufficiently small timestep* $0 < h \ll \min\{1/2\lambda, 1\}$ *and* $m \in \{0, \dots, M-1\}$ *for a given* $M \in \mathbb{N}$ *(recall* $T = t_M = Mh$*), we have*

$$\mathbb{E}[g(\boldsymbol{X}_T^N)] - \mathbb{E}[g(\boldsymbol{X}_T^{N,h})] = h^2 \mathbb{E}\left[\sum_{m=0}^{M-1} L(t_m, \boldsymbol{X}_{t_m}^{N,h})\right] + \mathbb{E}\left[\sum_{m=0}^{M-1} R(t_m, \boldsymbol{X}_{t_m}^{N,h})\right], \quad (5.19)$$

where $L : \mathbb{R}_+ \times \mathbb{R}^N \to \mathbb{R}$ *is defined via the map* u *defined in* (5.18) *and the drifts* $(B_i)_{i\in\{1,\dots,N\}}$ *in* (5.11) *as*

$$L(t, \boldsymbol{x}) = \frac{1}{2}\Big[\sum_{i,j=1}^N B_j(\boldsymbol{x})\partial_{x_j} B_i(\boldsymbol{x})\partial_{x_i} u(t, \boldsymbol{x}) + \frac{\sigma^2}{2}\sum_{i,j=1}^N \partial_{x_j} B_i(\boldsymbol{x})\partial^2_{x_i,x_j} u(t, \boldsymbol{x}) + \frac{\sigma^2}{2}\sum_{i,j=1}^N \partial^2_{x_j,x_j} B_i(\boldsymbol{x})\partial_{x_i} u(t, \boldsymbol{x})\Big], \quad t \ge 0, \ \boldsymbol{x} \in \mathbb{R}^N, \quad (5.20)$$

and $R(\cdot,\cdot)$ *is a collection of remainder terms (discussed and analyzed in Section 5.6.2).*

Proof. This expansion is derived and presented in [103, Equation (3.17)] and we do not reproduce it here. Our function L given in (5.20) is denoted as B_0 in [103] (see their Theorem 3.4). The sum of remainders $R(t_m, \cdot)$ in (5.19) corresponds to the second sum of remainders $h^3 r(t_m, \cdot)$ in [103, Equation (3.17)] – where the $h^3 r(t_m, \cdot)$ itself is a linear combination of remainders $h^3 r_j$ for $j \in \{1, \dots, 8\}$ appearing in Equations (3.8), (3.10)–(3.16) in [103, p7-9]. This derivation is discussed in more depth in Section 5.6.2. □

We aim to control the growth of L and the remainder R in terms of quantities that do not grow in N – this is a central technical difference to [103]. A key point in the growth analysis for L (and R) is the suitable control of moment bounds for the variation processes of $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\ge t\ge 0}$, which will be discussed in the next section. We end this section with this manuscript's main result. Its proof is given in Section 5.6 after establishing a large collection of auxiliary results in Sections 5.4 and 5.5, and the appendix.

Theorem 5.3.8. *Let Assumption 5.3.3 hold, let $\xi \in L^{10}(\Omega, \mathbb{R})$ and let $0 < h \ll \min\{1/2\lambda, 1\}$. Then the following expansion for the weak error for the processes defined in* (5.11) *and* (5.13) *holds: for any $N, M \in \mathbb{N}$ (with $T = Mh$),*

$$\mathbb{E}[g(\boldsymbol{X}_T^N)] - \mathbb{E}[g(\boldsymbol{X}_T^{N,h})] = C_0(T)h + Kh^{3/2},$$

where

$$C_0(T) := \mathbb{E}\left[\int_0^T L(t, \boldsymbol{X}_t^N)\mathrm{d}t\right],$$

and

$$|C_0(T)| \leq K\exp(-\lambda_0 T) + Kh^{1/2},$$

for some positive constants λ_0, K independent of h, T, M and N.

Proof. This result follows as a consequence of Lemmas 5.6.1 and 5.6.3 in Section 5.6. □

It is clear from the main statement that, as $T \to \infty$, the weak error is of order $3/2$ uniformly in N, and at finite time $T < \infty$ it is of order 1 uniformly in N. We flag that for standard SDEs (without concern for uniformity in N), both [103, 139] obtain an error of order h^2 (as $T \to \infty$). This gap in our result is of technical nature (appearing in Section 5.6) and is detailed in Remark 5.6.2. The assumption that $\xi \in L^{10}(\Omega, \mathbb{R})$ stems from the analysis of the remainder terms R appearing in the weak error expansion (5.19) (see Section 5.6.2). The analysis of the L term only requires $\xi \in L^4(\Omega, \mathbb{R})$ (see Section 5.6.1).

5.4 Analysis of Variation processes

The results for the variation processes established below are key to studying the uniform-in-N and uniform-in-time decay of the solution to the Kolmogorov backward equation. For completeness, we state the following lemma regarding wellposedness of the multiple variation process used throughout this section and drawing from classical SDE theory. The subsequent results of the section are devoted to establishing L^p-estimates of these processes that decay exponentially in time in a non-explosive way in N.

Lemma 5.4.1. *Let Assumption 5.3.3 hold, and let $\mathbb{N} \ni n \leq 6$ and $T > 0$. For any $\boldsymbol{x} \in \mathbb{R}^N$, and $T \geq s \geq t \geq 0$, let $\boldsymbol{X}_s^{t,\boldsymbol{x},N}$ be defined by* (5.15)*. Then its first n-variation processes given by* (5.21)*,* (5.57)*, and* (5.79) *have unique solutions.*

Proof. For any fixed N and $T > 0$, Assumption 5.3.3 implies Assumption (A) (p108), condition (5.12) and condition (5.15) (at any higher-order; see pages 120 and 122, respectively) in [73]. This suffices to ensure the wellposedness of the first (5.21), second (5.57) and higher-order (5.79) variation processes via Theorem 5.3 and Theorem 5.4 in [73] (see the comment after the proof of [73, Theorem 5.5 (p123)] regarding the extension of Theorem 5.4 to higher order derivatives). Note that the analysis in the later sections of this article is carried out for an arbitrary $T > 0$, and $T \to \infty$ is only considered in the final step; in particular, a wellposedness result for the variation processes in finite time suffices for our purposes. □

5.4.1 First Variation process

Here and below let $T > 0$ be an arbitrary terminal time and let $T \geq s \geq t \geq 0$, $N \in \mathbb{N}$. The first variation process of $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s \geq t \geq 0}$ defined in (5.15), is given by

$$X_{s,x_j}^{t,x_i,i,N} = \delta_{i,j} + \int_t^s \sum_{l=1}^N \partial_{x_l} B_i(\boldsymbol{X}_u^{t,\boldsymbol{x},N}) X_{u,x_j}^{t,x_l,l,N}\mathrm{d}u, \tag{5.21}$$

where $\delta_{i,j}$ is the usual Kronecker symbol. The subindex x_j in $X^{t,x_i,i,N}_{s,x_j}$ indicates the perturbation with respect to the j-th component of the initial data $\boldsymbol{x} \in \mathbb{R}^N$ of the flow process $(\boldsymbol{X}^{t,\boldsymbol{x},N}_s)_{s\geq t\geq 0}$. Note that the processes $(X^{t,x_i,i,N}_{s,x_j})_{s\geq t\geq 0}$ (for different indices i,j) are, in general, not identically distributed. However, if the starting positions x_i, for $i \in \{1,\dots,N\}$, are all sampled from the same distribution, then the 'diagonal' elements of $(\boldsymbol{X}^{t,\boldsymbol{x},N}_s)_{s\geq t\geq 0}$ are identically distributed. (The same argument applies to the off-diagonal ones). This first lemma accounts for the different behaviours of L^p-moments for (5.21).

Lemma 5.4.2. *Let Assumption 5.3.3 hold and let $p \geq 2$. Consider the first variation process $(X^{t,x_i,i,N}_{s,x_j})_{i,j\in\{1,\dots,N\}}$ defined by (5.21) for $T \geq s \geq t \geq 0$. Then there exist constants $\lambda_1 \in (0,\lambda)$ and $K > 0$ (both are independent of s,t,T and N) such that for any $T \geq s \geq t \geq 0$*

$$\sum_{i=1}^N \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j}|^p\Big] \leq Ke^{-\lambda p(s-t)} \qquad \text{and} \qquad \sum_{i=1,i\neq j}^N \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j}|^p\Big] \leq \frac{K}{N^{p-1}}e^{-\lambda_1 p(s-t)}. \tag{5.22}$$

This lemma and the earlier Remark 5.3.6 highlight the main difficulty faced in this manuscript's analysis. The above inequalities suggest that the term $i = j$ delivers the $\mathcal{O}(1)$ behaviour while all other (cross-derivative $i \neq j$) elements decay proportionally to the number of particles. Our analysis throughout this section is involved, as this behaviour needs to be tracked across higher-order variation processes.

Lastly, since (5.21) is a linear ODE with random coefficients (bounded in $\boldsymbol{X}^{t,\boldsymbol{x},N}$) and initial condition $\delta_{i,j}$ we are able to obtain all L^p-moments without imposing further constraints on the integrability of $\boldsymbol{X}^{t,\boldsymbol{x},N}$.

Remark 5.4.3 (The 'symmetrization trick')**.** *We employ a recurring argument in this proof which we coin as the symmetrization trick. This trick exploits that $\nabla^2 V$ is an even function. That is, for any $x \in \mathbb{R}$ we have $\nabla^2 V(x) = \frac{1}{2}(\nabla^2 V(x) + \nabla^2 V(-x))$.*

Proof. Note that in the following proof, the positive constant K is independent of s,t,T,N and may change line by line. The essence of this proof is the application of Itô's formula followed by standard domination arguments.

Applying Itô's formula yields for any $i,j \in \{1,\dots,N\}$

$$\begin{aligned}
&e^{\lambda p(s-t)}\mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j}|^p\Big] \\
&\leq \delta_{i,j} + p\int_t^s e^{\lambda p(u-t)} \\
&\quad \Big(\mathbb{E}\Big[X^{t,x_i,i,N}_{u,x_j} \cdot \Big(\sum_{l=1}^N \partial_{x_l} B_i(\boldsymbol{X}^{t,\boldsymbol{x},N}_u) X^{t,x_l,l,N}_{u,x_j}\Big)|X^{t,x_i,i,N}_{u,x_j}|^{p-2}\Big] + \lambda\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j}|^p\Big]\Big)\mathrm{d}u \\
&= \delta_{i,j} - p\int_t^s e^{\lambda p(u-t)} \\
&\quad \Big(\mathbb{E}\Big[X^{t,x_i,i,N}_{u,x_j} \cdot \Big(\nabla^2 U(X^{t,x_i,i,N}_u) X^{t,x_i,i,N}_{u,x_j}\Big)|X^{t,x_i,i,N}_{u,x_j}|^{p-2}\Big] - \lambda\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j}|^p\Big]\Big)\mathrm{d}u \\
&\quad - p\int_t^s e^{\lambda p(u-t)} \\
&\quad \mathbb{E}\Big[X^{t,x_i,i,N}_{u,x_j} \cdot \Big(\sum_{l=1}^N \frac{1}{N}\sum_{k=1}^N \partial_{x_l}\nabla V(X^{t,x_i,i,N}_u - X^{t,x_k,k,N}_u) X^{t,x_l,l,N}_{u,x_j}\Big)|X^{t,x_i,i,N}_{u,x_j}|^{p-2}\Big]\mathrm{d}u \\
&\leq \delta_{i,j} - p\int_t^s e^{\lambda p(u-t)} \\
&\quad \mathbb{E}\Big[X^{t,x_i,i,N}_{u,x_j} \cdot \Big(\sum_{l=1}^N \frac{1}{N}\sum_{k=1}^N \partial_{x_l}\nabla V(X^{t,x_i,i,N}_u - X^{t,x_k,k,N}_u) X^{t,x_l,l,N}_{u,x_j}\Big)|X^{t,x_i,i,N}_{u,x_j}|^{p-2}\Big]\mathrm{d}u,
\end{aligned}$$

where we used (5.9) of Assumption 5.2.1 to obtain

$$-X_{u,x_j}^{t,x_i,i,N} \cdot \Big(\nabla^2 U(X_u^{t,x_i,i,N}) X_{u,x_j}^{t,x_i,i,N}\Big) \leq -\lambda |X_{u,x_j}^{t,x_i,i,N}|^2. \tag{5.23}$$

We further note that by the chain rule,

$$\begin{aligned}\sum_{l=1}^{N} \frac{1}{N} \sum_{k=1}^{N} &\partial_{x_l} \nabla V(X_u^{t,x_i,i,N} - X_u^{t,x_k,k,N}) X_{u,x_j}^{t,x_l,l,N} \\ &= \frac{1}{N} \sum_{l=1}^{N} \nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_l,l,N})(X_{u,x_j}^{t,x_i,i,N} - X_{u,x_j}^{t,x_l,l,N}). \end{aligned} \tag{5.24}$$

Hence taking summation over $i \in \{1, \ldots, N\}$ in (5.24), we have

$$\begin{aligned}
&\sum_{i=1}^{N} \mathbb{E}\Big[X_{u,x_j}^{t,x_i,i,N} \cdot \Big(\sum_{l=1}^{N} \frac{1}{N} \sum_{k=1}^{N} \partial_{x_l} \nabla V(X_u^{t,x_i,i,N} - X_u^{t,x_k,k,N}) X_{u,x_j}^{t,x_l,l,N}\Big) |X_{u,x_j}^{t,x_i,i,N}|^{p-2}\Big] \\
&= \mathbb{E}\Big[\frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \Big(|X_{u,x_j}^{t,x_i,i,N}|^{p-2} X_{u,x_j}^{t,x_i,i,N} \cdot \nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_l,l,N})(X_{u,x_j}^{t,x_i,i,N} - X_{u,x_j}^{t,x_l,l,N})\Big)\Big] \\
&= \mathbb{E}\Big[\frac{1}{2N} \sum_{i=1}^{N} \sum_{l=1}^{N} \Big(|X_{u,x_j}^{t,x_i,i,N}|^{p-2} X_{u,x_j}^{t,x_i,i,N} \cdot \nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_l,l,N})(X_{u,x_j}^{t,x_i,i,N} - X_{u,x_j}^{t,x_l,l,N})\Big)\Big] \\
&\quad + \mathbb{E}\Big[\frac{1}{2N} \sum_{l=1}^{N} \sum_{i=1}^{N} \Big(|X_{u,x_j}^{t,x_l,l,N}|^{p-2} X_{u,x_j}^{t,x_l,l,N} \cdot \nabla^2 V(X_u^{t,x_l,l,N} - X_u^{t,x_i,i,N})(X_{u,x_j}^{t,x_l,l,N} - X_{u,x_j}^{t,x_i,i,N})\Big)\Big]
\end{aligned} \tag{5.25}$$

$$\begin{aligned}
&= \mathbb{E}\Big[\frac{1}{2N} \sum_{i=1}^{N} \sum_{l=1}^{N} \Big(\big(|X_{u,x_j}^{t,x_i,i,N}|^{p-2} X_{u,x_j}^{t,x_i,i,N} - |X_{u,x_j}^{t,x_l,l,N}|^{p-2} X_{u,x_j}^{t,x_l,l,N}\big) \\
&\qquad \cdot \nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_l,l,N})(X_{u,x_j}^{t,x_i,i,N} - X_{u,x_j}^{t,x_l,l,N})\Big)\Big] \geq 0,
\end{aligned} \tag{5.26}$$

where in (5.26), we used the inequality $\big(|x|^{p-2}x - |y|^{p-2}y\big) \cdot (x - y) \geq 0$ for all $x, y \in \mathbb{R}$ and the fact that $\nabla^2 V(x) \geq 0$, for all $x \in \mathbb{R}$. In accordance to Remark 5.4.3, we used the *symmetrization trick* to derive (5.25). Hence, taking summation over $i \in \{1, \ldots, N\}$, we deduce that

$$e^{\lambda p(s-t)} \sum_{i=1}^{N} \mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p\Big] \leq \sum_{i=1}^{N} \delta_{i,j} = 1$$

and consequently for all $j \in \{1, \ldots, N\}$, we have

$$\sum_{i=1}^{N} \mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p\Big] \leq e^{-\lambda p(s-t)}. \tag{5.27}$$

To prove the claim of the second result in (5.21), we first derive for any $j \in \{1, \ldots, N\}$, $\lambda_1 \in (0, \lambda)$ (using Itô's formula and (5.23)),

$$\begin{aligned}
e^{\lambda_1 p(s-t)} \sum_{i=1, i\neq j}^{N} \mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p\Big] &\leq p \int_t^s (\lambda_1 - \lambda) e^{\lambda_1(u-t)} \sum_{i=1, i\neq j}^{N} \mathbb{E}\Big[|X_{u,x_j}^{t,x_i,i,N}|^p\Big] \mathrm{d}u \\
&\quad - p \int_t^s e^{\lambda_1(u-t)} \sum_{i=1, i\neq j}^{N} \mathbb{E}\Big[|X_{u,x_j}^{t,x_i,i,N}|^{p-2} X_{u,x_j}^{t,x_i,i,N} \\
&\quad \cdot \Big(\sum_{l=1}^{N} \frac{1}{N} \sum_{k=1}^{N} \partial_{x_l} \nabla V(X_u^{t,x_i,i,N} - X_u^{t,x_k,k,N}) X_{u,x_j}^{t,x_l,l,N}\Big)\Big] \mathrm{d}u.
\end{aligned}$$

For the last term above, we note that

$$
\begin{aligned}
&-\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X_{u,x_j}^{t,x_i,i,N}|^{p-2}X_{u,x_j}^{t,x_i,i,N}\cdot\Big(\sum_{l=1}^{N}\frac{1}{N}\sum_{k=1}^{N}\partial_{x_l}\nabla V(X_u^{t,x_i,i,N}-X_u^{t,x_k,k,N})X_{u,x_j}^{t,x_l,l,N}\Big)\Big]\\
&=-\mathbb{E}\Big[\frac{1}{N}\sum_{i=1,i\neq j}^{N}\sum_{l=1,l\neq j}^{N}\Big(|X_{u,x_j}^{t,x_i,i,N}|^{p-2}X_{u,x_j}^{t,x_i,i,N}\\
&\quad\cdot\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_l,l,N})(X_{u,x_j}^{t,x_i,i,N}-X_{u,x_j}^{t,x_l,l,N})\Big)\Big] \qquad (5.28)\\
&\quad-\mathbb{E}\Big[\frac{1}{N}\sum_{i=1,i\neq j}^{N}\Big(|X_{u,x_j}^{t,x_i,i,N}|^{p-2}\cdot X_{u,x_j}^{t,x_i,i,N},\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_j,j,N})(X_{u,x_j}^{t,x_i,i,N}-X_{u,x_j}^{t,x_j,j,N})\Big)\Big]\\
&\leq-\mathbb{E}\Big[\frac{1}{2N}\sum_{i=1,i\neq j}^{N}\sum_{l=1,l\neq j}^{N}\Big(\big(|X_{u,x_j}^{t,x_i,i,N}|^{p-2}X_{u,x_j}^{t,x_i,i,N}-|X_{u,x_j}^{t,x_l,l,N}|^{p-2}X_{u,x_j}^{t,x_l,l,N}\big)\\
&\quad\cdot\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_l,l,N})(X_{u,x_j}^{t,x_i,i,N}-X_{u,x_j}^{t,x_l,l,N})\Big)\Big] \qquad (5.29)\\
&\quad+\mathbb{E}\Big[\frac{1}{N}\sum_{i=1,i\neq j}^{N}|\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_j,j,N})|\,|X_{u,x_j}^{t,x_i,i,N}|^{p-1}\,|X_{u,x_j}^{t,x_j,j,N}|\Big]\\
&\leq K_V\mathbb{E}\Big[\sum_{i=1,i\neq j}^{N}|X_{u,x_j}^{t,x_i,i,N}|^{p-1}\,\frac{|X_{u,x_j}^{t,x_j,j,N}|}{N}\Big] \qquad (5.30)\\
&\leq\varepsilon\Big(\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X_{u,x_j}^{t,x_i,i,N}|^p\Big]\Big)+\frac{K}{N^{p-1}}\mathbb{E}\Big[|X_{u,x_j}^{t,x_j,j,N}|^p\Big], \qquad (5.31)
\end{aligned}
$$

where in (5.29), we used the *symmetrization trick* again. (5.30) follows from Assumption 5.2.1(2) and (5.31) is a consequence of Young's inequality, where ε is some positive constant which can be chosen to be arbitrarily small. Note that one can choose ε, K to be positive constants (both independent of s, t, N) satisfying $\varepsilon\in(0,\lambda-\lambda_1)$ for any $\lambda_1\in(0,\lambda)$. Thus, we conclude that

$$
\begin{aligned}
e^{\lambda_1 p(s-t)}&\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p\Big]\\
&\leq p\int_t^s e^{\lambda_1 p(u-t)}\Big((\lambda_1+\varepsilon-\lambda)\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X_{u,x_j}^{t,x_i,i,N}|^p\Big]+\frac{K}{N^{p-1}}\mathbb{E}\Big[|X_{u,x_j}^{t,x_j,j,N}|^p\Big]\Big)\mathrm{d}u\\
&\leq\frac{Kp}{N^{p-1}}\int_t^s e^{p(\lambda_1-\lambda)(u-t)}\mathrm{d}u\leq\frac{K}{N^{p-1}(\lambda-\lambda_1)},
\end{aligned}
$$

where we noted that ε can be chosen to be arbitrarily small, such that $\lambda_1+\varepsilon-\lambda$ remains negative (so the summation term can be upper bounded by zero). We used (5.27) to bound $\mathbb{E}\big[|X_{u,x_j}^{t,x_j,j,N}|^p\big]$ and then noted that $\int_t^s e^{p(\lambda_1-\lambda)(u-t)}\mathrm{d}u\leq 1/p(\lambda-\lambda_1)$ to conclude the result.

Therefore, for all $\lambda_1\in(0,\lambda)$, we have

$$
\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p\Big]\leq\frac{K}{N^{p-1}}e^{-\lambda_1 p(s-t)}.
$$

□

The following proposition provides L^2-estimates for the differences of the processes defined in (5.21) with the same initial points, but at different starting times. The results are used in Section 5.6.1 to establish time-regularity estimates for the derivatives of the function u.

Proposition 5.4.4. *Let Assumption 5.3.3 hold. Consider the first variation process with components $(X^{t,x_i,i,N}_{s,x_j})_{s\geq t\geq 0}$ defined by (5.21) for $i,j\in\{1,\dots,N\}$ and assume that the starting positions $x_i\in L^4(\Omega,\mathbb{R})$ are $\mathcal{F}_t$-measurable random variables that are identically distributed over all $i\in\{1,\dots,N\}$. Then there exist $\lambda_2\in(0,\min\{\lambda-2K_V,\lambda_1\})$, $\lambda_3\in(0,\min\{\lambda-2K_V,\lambda_2\})$, and $K>0$ (all independent of s,t,T,N) such that for all $T\geq s\geq t\geq 0$ with $s-t<1$,*

$$\sum_{i=1}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j}-X^{s,x_i,i,N}_{T,x_j}|^2\Big]\leq K(s-t)e^{-2\lambda_2(T-s)}, \tag{5.32}$$

$$\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j}-X^{s,x_i,i,N}_{T,x_j}|^2\Big]\leq \frac{K(s-t)}{N}e^{-2\lambda_3(T-s)}. \tag{5.33}$$

Proof. Note that in the following proof, the positive constant K is independent of s,t,T,N and may change line by line.

Part 1: Preliminary manipulations. Similar to the calculations in the proof of Lemma 5.4.2, we derive that (recalling that (5.21) is an ODE with random coefficients), for all $i,j\in\{1,\dots,N\}$, $T\geq s\geq t\geq 0$, $\lambda_2\in(0,\min\{\lambda-2K_V,\lambda_1\})$,

$$\begin{aligned}
&e^{2\lambda_2(T-s)}|X^{t,x_i,i,N}_{T,x_j}-X^{s,x_i,i,N}_{T,x_j}|^2\\
&=|X^{t,x_i,i,N}_{s,x_j}-X^{s,x_i,i,N}_{s,x_j}|^2+2\lambda_2\int_0^{T-s}e^{2\lambda_2u}|X^{t,x_i,i,N}_{s+u,x_j}-X^{s,x_i,i,N}_{s+u,x_j}|^2\mathrm{d}u\\
&\quad-2\int_0^{T-s}e^{2\lambda_2u}\Big(X^{t,x_i,i,N}_{s+u,x_j}-X^{s,x_i,i,N}_{s+u,x_j}\Big)\\
&\qquad\cdot\Big(\nabla^2U(X^{t,x_i,i,N}_{s+u})X^{t,x_i,i,N}_{s+u,x_j}-\nabla^2U(X^{s,x_i,i,N}_{s+u})X^{s,x_i,i,N}_{s+u,x_j}\Big)\mathrm{d}u\\
&\quad-2\int_0^{T-s}e^{2\lambda_2u}\Big(X^{t,x_i,i,N}_{s+u,x_j}-X^{s,x_i,i,N}_{s+u,x_j}\Big)\\
&\qquad\cdot\Big(\frac{1}{N}\sum_{l=1}^{N}\nabla^2V(X^{t,x_i,i,N}_{s+u}-X^{t,x_l,l,N}_{s+u})(X^{t,x_i,i,N}_{s+u,x_j}-X^{t,x_l,l,N}_{s+u,x_j})\\
&\qquad\qquad-\frac{1}{N}\sum_{l=1}^{N}\nabla^2V(X^{s,x_i,i,N}_{s+u}-X^{s,x_l,l,N}_{s+u})(X^{s,x_i,i,N}_{s+u,x_j}-X^{s,x_l,l,N}_{s+u,x_j})\Big)\mathrm{d}u.
\end{aligned}$$

Therefore, using (5.9), we have

$$\begin{aligned}
&e^{2\lambda_2(T-s)}|X^{t,x_i,i,N}_{T,x_j}-X^{s,x_i,i,N}_{T,x_j}|^2 &(5.34)\\
&\leq|X^{t,x_i,i,N}_{s,x_j}-X^{s,x_i,i,N}_{s,x_j}|^2+2(\lambda_2-\lambda)\int_0^{T-s}e^{2\lambda_2u}|X^{t,x_i,i,N}_{s+u,x_j}-X^{s,x_i,i,N}_{s+u,x_j}|^2\mathrm{d}u\\
&\quad-2\int_0^{T-s}e^{2\lambda_2u}\Big(X^{t,x_i,i,N}_{s+u,x_j}-X^{s,x_i,i,N}_{s+u,x_j}\Big) &(5.35)\\
&\qquad\cdot\Big(\nabla^2U(X^{t,x_i,i,N}_{s+u})X^{s,x_i,i,N}_{s+u,x_j}-\nabla^2U(X^{s,x_i,i,N}_{s+u})X^{s,x_i,i,N}_{s+u,x_j}\Big)\mathrm{d}u\\
&\quad-2\int_0^{T-s}e^{2\lambda_2u}\Big(X^{t,x_i,i,N}_{s+u,x_j}-X^{s,x_i,i,N}_{s+u,x_j}\Big)\\
&\qquad\cdot\Bigg(\Big(\frac{1}{N}\sum_{l=1}^{N}\nabla^2V(X^{t,x_i,i,N}_{s+u}-X^{t,x_l,l,N}_{s+u})(X^{t,x_i,i,N}_{s+u,x_j}-X^{t,x_l,l,N}_{s+u,x_j})\\
&\qquad-R^{i,t,s}_{s+u}\Big)+\Big(R^{i,t,s}_{s+u}-\frac{1}{N}\sum_{l=1}^{N}\nabla^2V(X^{s,x_i,i,N}_{s+u}-X^{s,x_l,l,N}_{s+u})(X^{s,x_i,i,N}_{s+u,x_j}-X^{s,x_l,l,N}_{s+u,x_j})\Big)\Bigg)\mathrm{d}u,\\
&&(5.36)
\end{aligned}$$

where we added and subtract the following auxiliary term:

$$R^{i,t,s}_{s+u} := \frac{1}{N}\sum_{l=1}^{N}\Big[\nabla^2 V(X^{t,x_i,i,N}_{s+u} - X^{t,x_l,l,N}_{s+u})(X^{s,x_i,i,N}_{s+u,x_j} - X^{s,x_l,l,N}_{s+u,x_j})\Big].$$

The result in (5.22) and the fact that the starting positions x_i are identically distributed, yield that for all $i,j \in \{1,\dots,N\}, i \neq j,\ \lambda_1 \in (0,\lambda)$

$$\mathbb{E}\Big[\big|X^{s,x_i,i,N}_{s+u,x_i}\big|^4\Big] \leq Ke^{-4\lambda_1 u}, \qquad \mathbb{E}\Big[\big|X^{s,x_i,i,N}_{s+u,x_j}\big|^4\Big] \leq \frac{K}{N^4}e^{-4\lambda_1 u}. \tag{5.37}$$

Part 2: Establishing (5.32). We further estimate the term involving $\nabla^2 U$, (5.35): under Assumption 5.2.1, we have

$$\begin{aligned}
\sum_{i=1}^{N}&\mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\Big)\cdot\Big(\nabla^2 U(X^{t,x_i,i,N}_{s+u})X^{s,x_i,i,N}_{s+u,x_j} - \nabla^2 U(X^{s,x_i,i,N}_{s+u})X^{s,x_i,i,N}_{s+u,x_j}\Big)\Big]\\
&\leq \sum_{i=1}^{N}\lambda\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|\,\big|X^{t,x_i,i,N}_{s+u} - X^{s,x_i,i,N}_{s+u}\big|\,\big|X^{s,x_i,i,N}_{s+u,x_j}\big|\Big]\\
&\leq \varepsilon\sum_{i=1}^{N}\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|^2\Big] + K\sum_{i=1}^{N}\sqrt{\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u} - X^{s,x_i,i,N}_{s+u}\big|^4\Big]\mathbb{E}\Big[\big|X^{s,x_i,i,N}_{s+u,x_j}\big|^4\Big]},
\end{aligned} \tag{5.38}$$

where we employed Young's inequality (with constants $\varepsilon, K > 0$ independent of s,t and N) and the Cauchy–Schwarz inequality. We further bound (5.38) by applying Lemma 6.4.2

$$\begin{aligned}
(5.38) &\leq \varepsilon\sum_{i=1}^{N}\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|^2\Big]\\
&\qquad + K(s-t)e^{-2\lambda_2 u}\bigg(\sqrt{\mathbb{E}\Big[\big|X^{s,x_j,j,N}_{s+u,x_j}\big|^4\Big]} + \sum_{i=1,i\neq j}^{N}\sqrt{\mathbb{E}\Big[\big|X^{s,x_i,i,N}_{s+u,x_j}\big|^4\Big]}\bigg)\\
&\leq \varepsilon\sum_{i=1}^{N}\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|^2\Big] + K(s-t)e^{-2\lambda_2 u}\Big(e^{-2\lambda_1 u} + \frac{1}{N^2}\sum_{i=1,i\neq j}^{N}e^{-2\lambda_1 u}\Big) \qquad (5.39)\\
&\leq \varepsilon\sum_{i=1}^{N}\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|^2\Big] + K(s-t)e^{-4\lambda_2 u}, \qquad (5.40)
\end{aligned}$$

for some $\lambda_2 \in (0, \min\{\lambda - 2K_V, \lambda_1\})$, where we injected the estimate (5.37) and used that the processes $(X^{t,x_i,i,N}_s)_{s\geq t\geq 0}$ for $i \in \{1,\dots,N\}$ are identically distributed (due to the assumption in 5.4.4 on the starting positions x_i being identically distributed over $i \in \{1,\dots,N\}$) in the second inequality.

As for the term involving $\nabla^2 V$, (5.36), after taking the expectation and summing over $i \in \{1,\dots,N\}$,

$$\begin{aligned}
&-\sum_{i=1}^{N}\mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\Big)\\
&\quad\cdot\Big(\frac{1}{N}\sum_{l=1}^{N}\nabla^2 V(X^{t,x_i,i,N}_{s+u} - X^{t,x_l,l,N}_{s+u})(X^{t,x_i,i,N}_{s+u,x_j} - X^{t,x_l,l,N}_{s+u,x_j}) - R^{i,t,s}_{s+u}\Big)\Big]\\
&= -\frac{1}{2N}\sum_{i=1}^{N}\sum_{l=1}^{N}\mathbb{E}\Big[\Big(\big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big) - \big(X^{t,x_l,l,N}_{s+u,x_j} - X^{s,x_l,l,N}_{s+u,x_j}\big)\Big)
\end{aligned}$$

$$\cdot \nabla^2 V(X^{t,x_i,i,N}_{s+u} - X^{t,x_l,l,N}_{s+u})\Big(\big(X^{t,x_i,i,N}_{s+u,x_j} - X^{t,x_l,l,N}_{s+u,x_j}\big) - \big(X^{s,x_i,i,N}_{s+u,x_j} - X^{s,x_l,l,N}_{s+u,x_j}\big)\Big)\Big] \le 0, \tag{5.41}$$

where we once again use the *symmetrization trick* and that $\nabla^2 V(x) \ge 0$ for all $x \in \mathbb{R}$. Similar to the analysis involving $\nabla^2 U$, we obtain

$$\begin{aligned}
&\sum_{i=1}^{N} \mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\Big) \\
&\quad \cdot \Big(R^{i,t,s}_{s+u} - \frac{1}{N}\sum_{l=1}^{N} \nabla^2 V(X^{s,x_i,i,N}_{s+u} - X^{s,x_l,l,N}_{s+u})(X^{s,x_i,i,N}_{s+u,x_j} - X^{s,x_l,l,N}_{s+u,x_j})\Big)\Big] \\
&\le \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N} K_V \mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big| \, \big|(X^{t,x_i,i,N}_{s+u} - X^{t,x_l,l,N}_{s+u}) - (X^{s,x_i,i,N}_{s+u} - X^{s,x_l,l,N}_{s+u})\big| \\
&\quad \cdot \ \big|X^{s,x_i,i,N}_{s+u,x_j} - X^{s,x_l,l,N}_{s+u,x_j}\big|\Big] \le \varepsilon \sum_{i=1}^{N} \mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|^2\Big] + K(s-t)e^{-4\lambda_2 u},
\end{aligned} \tag{5.42}$$

where we applied similar calculations as in (5.39) and (5.40). After taking the expectation, summing over $i \in \{1, \ldots, N\}$, and injecting our established estimates (5.40), (5.41) and (5.42), we have for an arbitrary small $\varepsilon > 0$,

$$\begin{aligned}
&e^{2\lambda_2(T-s)} \sum_{i=1}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j} - X^{s,x_i,i,N}_{T,x_j}|^2\Big] \\
&\qquad \le \sum_{i=1}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j} - X^{s,x_i,i,N}_{s,x_j}|^2\Big] + K(s-t)\int_0^{T-s} e^{(2\lambda_2 - 4\lambda_2)u}\mathrm{d}u \\
&\qquad\quad + 2(2\varepsilon + \lambda_2 - \lambda)\int_0^{T-s} e^{2\lambda_2 u}\sum_{i=1}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}|^2\Big]\mathrm{d}u.
\end{aligned} \tag{5.43}$$

Further notice that for the first summation term of (5.43), we obtain

$$\begin{aligned}
&\sum_{i=1}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j} - X^{s,x_i,i,N}_{s,x_j}|^2\Big] \\
&\quad = \sum_{i=1}^{N}\mathbb{E}\Big[\Big|\int_0^{s-t}\Big(\frac{1}{N}\sum_{l=1}^{N}\nabla^2 V(X^{t,x_i,i,N}_{t+u} - X^{t,x_l,l,N}_{t+u})(X^{t,x_i,i,N}_{t+u,x_j} - X^{t,x_l,l,N}_{t+u,x_j}) \\
&\qquad\qquad + \nabla^2 U(X^{t,x_i,i,N}_{t+u})X^{t,x_i,i,N}_{t+u,x_j}\Big)\mathrm{d}u\Big|^2\Big] \\
&\quad \le K(s-t)\int_0^{s-t}\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N}\mathbb{E}\Big[|\nabla^2 V(X^{t,x_i,i,N}_{t+u} - X^{t,x_l,l,N}_{t+u})|^2\,|X^{t,x_i,i,N}_{t+u,x_j} - X^{t,x_l,l,N}_{t+u,x_j}|^2\Big] \\
&\qquad\qquad + \sum_{i=1}^{N}\mathbb{E}\Big[|\nabla^2 U(X^{t,x_i,i,N}_{t+u})|^2\,|X^{t,x_i,i,N}_{t+u,x_j}|^2\Big]\Big)\,\mathrm{d}u
\end{aligned} \tag{5.44}$$

$$\le K\big(4K_V + \lambda\big)\int_0^{s-t}\sum_{i=1}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{t+u,x_j}|^2\Big]\mathrm{d}u \tag{5.45}$$

$$\le K(s-t)\int_0^{s-t} e^{-2\lambda_1 u}\mathrm{d}u \le K(s-t), \tag{5.46}$$

where we used Jensen's inequality in (5.44), Assumption 5.2.1 in (5.45) and Lemma 5.4.2 with $\lambda_1 \in (0, \lambda)$ to establish (5.46). Consequently, for all $j \in \{1, \ldots, N\}$, we have (by choosing ε

arbitrarily small)

$$
\begin{aligned}
e^{2\lambda_2(T-s)} \sum_{i=1}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j} - X^{s,x_i,i,N}_{T,x_j}|^2\Big] \\
&\le K(s-t) + K(s-t)\int_0^{T-s} e^{-2\lambda_2 u}\mathrm{d}u \\
&\quad + 2(2\varepsilon+\lambda_2-\lambda)\int_0^{T-s} e^{2\lambda_2 u}\sum_{i=1}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}|^2\Big].
\end{aligned}
$$

Using $\int_0^{T-s} e^{-2\lambda_2 u}\mathrm{d}u \le 1/(2\lambda_2)$ and $(2\varepsilon+\lambda_2-\lambda)<0$, we deduce that

$$
\sum_{i=1}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j} - X^{s,x_i,i,N}_{T,x_j}|^2\Big] \le K(s-t)e^{-2\lambda_2(T-s)}. \tag{5.47}
$$

This concludes the first part of the statement (5.32).

Part 3: Establishing (5.33)*.* Mimicking the estimates in (5.29)–(5.31), we first establish a result to deal with the term involving $\nabla^2 V$:

$$
\begin{aligned}
&-\sum_{i=1,i\neq j}^{N} \mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\Big) \\
&\quad\cdot\Big(\frac{1}{N}\sum_{l=1}^{N}\nabla^2 V(X^{t,x_i,i,N}_{s+u} - X^{t,x_l,l,N}_{s+u})(X^{t,x_i,i,N}_{s+u,x_j} - X^{t,x_l,l,N}_{s+u,x_j}) - R^{i,t,s}_{s+u}\Big)\Big] \\
&= -\frac{1}{2N}\sum_{i=1,i\neq j}^{N}\sum_{l=1,l\neq j}^{N}\mathbb{E}\Big[\Big(\big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big) - \big(X^{t,x_l,l,N}_{s+u,x_j} - X^{s,x_l,l,N}_{s+u,x_j}\big)\Big) \\
&\qquad\cdot\Big(\nabla^2 V(X^{t,x_i,i,N}_{s+u} - X^{t,x_l,l,N}_{s+u})\big(X^{t,x_i,i,N}_{s+u,x_j} - X^{t,x_l,l,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j} + X^{s,x_l,l,N}_{s+u,x_j}\big)\Big)\Big] \\
&\quad - \frac{1}{N}\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\Big)\cdot\nabla^2 V(X^{t,x_i,i,N}_{s+u} - X^{t,x_j,l,N}_{s+u}) \\
&\qquad\qquad\cdot\Big(\big(X^{t,x_i,i,N}_{s+u,x_j} - X^{t,x_j,j,N}_{s+u,x_j}\big) - \big(X^{s,x_i,i,N}_{s+u,x_j} - X^{s,x_j,j,N}_{s+u,x_j}\big)\Big)\Big]
\end{aligned} \tag{5.48}
$$

$$
\le K_V \sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[\Big(|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}|\Big)\cdot\Big(\frac{1}{N}|X^{t,x_j,j,N}_{s+u,x_j} - X^{s,x_j,j,N}_{s+u,x_j}|\Big)\Big] \tag{5.49}
$$

$$
\le \varepsilon \sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}|^2\Big] + \frac{K}{N^2}\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_j,j,N}_{s+u,x_j} - X^{s,x_j,j,N}_{s+u,x_j}|^2\Big], \tag{5.50}
$$

where we once again used the *symmetrization trick* in (5.48), 5.2.1(2) in (5.49) and Young's inequality with constants ε, K chosen such that $\varepsilon \ll \lambda$ in (5.50). Similarly, we have

$$
\begin{aligned}
&-\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\Big) \\
&\quad\cdot\Big(R^{i,t,s}_{s+u} - \frac{1}{N}\sum_{l=1}^{N}\nabla^2 V(X^{s,x_i,i,N}_{s+u} - X^{s,x_l,l,N}_{s+u})(X^{s,x_i,i,N}_{s+u,x_j} - X^{s,x_l,l,N}_{s+u,x_j})\Big)\Big] \\
&\le \frac{1}{N}\sum_{i=1,i\neq j}^{N}\sum_{l=1,l\neq j}^{N} K_V\mathbb{E}\Big[|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}|\ |(X^{t,x_i,i,N}_{s+u} - X^{t,x_l,l,N}_{s+u}) - (X^{s,x_i,i,N}_{s+u} - X^{s,x_l,l,N}_{s+u})|
\end{aligned}
$$

$$
\cdot \big|X^{s,x_i,i,N}_{s+u,x_j} - X^{s,x_l,l,N}_{s+u,x_j}\big|\Big]
$$

$$
+ \frac{1}{N}\sum_{i=1,i\neq j}^{N} K_V \mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|\ \big|(X^{t,x_i,i,N}_{s+u} - X^{t,x_j,j,N}_{s+u}) - (X^{s,x_i,i,N}_{s+u} - X^{s,x_j,j,N}_{s+u})\big|
$$

$$
\cdot \big|X^{s,x_i,i,N}_{s+u,x_j} - X^{s,x_j,j,N}_{s+u,x_j}\big|\Big]
$$

$$
\leq \frac{1}{N}\sum_{i=1,i\neq j}^{N}\sum_{l=1,l\neq j}^{N}\Big(\varepsilon\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|^2\Big] + \frac{K(s-t)}{N^2}e^{-4\lambda_2 u}\Big)
$$

$$
+ \sum_{i=1,i\neq j}^{N}\Big(\varepsilon\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|^2\Big] + \frac{K(s-t)}{N^2}e^{-4\lambda_2 u}\Big) \tag{5.51}
$$

$$
\leq 2\varepsilon\Big(\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[\big|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\big|^2\Big]\Big) + \frac{K(s-t)}{N}e^{-4\lambda_2 u}, \tag{5.52}
$$

where we used the moment bounds established in (5.37) to get (5.51) and applied similar calculations as in (5.40). Taking summation over $i, i \neq j$ and collecting the estimates in (5.50) and (5.52),

$$
e^{2\lambda_3(T-s)}\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j} - X^{s,x_i,i,N}_{T,x_j}|^2\Big] \leq \sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j} - X^{s,x_i,i,N}_{s,x_j}|^2\Big]
$$

$$
+ 2(\lambda_3 + 2\varepsilon - \lambda)\int_0^{T-s} e^{2\lambda_3 u}\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}|^2\Big]\mathrm{d}u \tag{5.53}
$$

$$
+ \frac{2K}{N^2}\int_0^{T-s} e^{2\lambda_3 u}\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_j,j,N}_{s+u,x_j} - X^{s,x_j,j,N}_{s+u,x_j}|^2\Big]\mathrm{d}u + \frac{K(s-t)}{N}\int_0^{T-s} e^{2\lambda_3 - 4\lambda_2 u}\mathrm{d}u. \tag{5.54}
$$

Note that (5.53)≤ 0 since $\lambda_3 < \lambda$ and ε can be chosen to be arbitrarily small, so that this term remains negative. Implementing a crude upper bound (5.47) and using that $\lambda_3 < \lambda_2$ (hence the integrals remain bounded as T gets large), we have

$$
(5.54) \leq \frac{K^2(s-t)}{N(\lambda_2 - \lambda_3)} + \frac{K(s-t)}{2N(2\lambda_2 - \lambda_3)}.
$$

Joining together the terms and estimates terms, we have

$$
e^{2\lambda_3(T-s)}\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j} - X^{s,x_i,i,N}_{T,x_j}|^2\Big]
$$

$$
\leq \sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j} - X^{s,x_i,i,N}_{s,x_j}|^2\Big] + \frac{K(s-t)}{N}. \tag{5.55}
$$

To analyze the summation term in (5.55), we first provide the following estimate: For all $u \geq 0$ and $i \in \{1, \dots, N\}$:

$$
\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{l=1}^{N}\nabla^2 V(X^{t,x_i,i,N}_{t+u} - X^{t,x_l,l,N}_{t+u})\ (X^{t,x_i,i,N}_{t+u,x_j} - X^{t,x_l,l,N}_{t+u,x_j})\Big|^2\Big]
$$

$$
\leq 2\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{l=1,l\neq j}^{N}\nabla^2 V(X^{t,x_i,i,N}_{t+u} - X^{t,x_l,l,N}_{t+u})\ (X^{t,x_i,i,N}_{t+u,x_j} - X^{t,x_l,l,N}_{t+u,x_j})\Big|^2\Big]
$$

$$
+ \frac{2}{N^2}\mathbb{E}\Big[\Big|\nabla^2 V(X^{t,x_i,i,N}_{t+u} - X^{t,x_j,j,N}_{t+u})\ (X^{t,x_i,i,N}_{t+u,x_j} - X^{t,x_j,j,N}_{t+u,x_j})\Big|^2\Big]
$$

$$\leq \frac{K}{N}\sum_{l=1,l\neq j}^{N}\mathbb{E}\Big[\Big|\nabla^2 V(X_{t+u}^{t,x_i,i,N}-X_{t+u}^{t,x_l,l,N})\;(X_{t+u,x_j}^{t,x_i,i,N}-X_{t+u,x_j}^{t,x_l,l,N})\Big|^2\Big]$$
$$+\frac{K}{N^2}\mathbb{E}\Big[\Big|X_{t+u,x_j}^{t,x_i,i,N}-X_{t+u,x_j}^{t,x_j,j,N}\Big|^2\Big]$$
$$\leq \frac{K}{N}\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X_{t+u,x_j}^{t,x_i,i,N}|^2\Big]+\frac{K}{N^2}\mathbb{E}\Big[|X_{t+u,x_j}^{t,x_j,j,N}|^2\Big]\leq \frac{K}{N^2}e^{-2\lambda_1 u},$$

where we isolated the $l=j$ term, applied $(a+b)^2\leq 2(a^2+b^2)$ for $a,b\in\mathbb{R}$, before applying Jensen's inequality and Assumption 5.2.1. Therefore, for the summation term in (5.55):

$$\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}-X_{s,x_j}^{s,x_i,i,N}|^2\Big]$$
$$\leq K(s-t)\int_0^{s-t}\Bigg(\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{l=1}^{N}\nabla^2 V(X_{t+u}^{t,x_i,i,N}-X_{t+u}^{t,x_l,l,N})\;(X_{t+u,x_j}^{t,x_i,i,N}-X_{t+u,x_j}^{t,x_l,l,N})\Big|^2\Big]$$
$$+\sum_{i=1,i\neq j}^{N}\mathbb{E}\Big[|\nabla^2 U(X_{t+u}^{t,x_i,i,N})|^2\;|X_{t+u,x_j}^{t,x_i,i,N}|^2\Big]\Bigg)\,\mathrm{d}u$$
$$\leq K(s-t)\int_0^{s-t}\Bigg(\Big(\sum_{i=1,i\neq j}^{N}\frac{e^{-2\lambda_1 u}}{N^2}\Big)+\frac{e^{-2\lambda_1 u}}{N}\Bigg)\,\mathrm{d}u\leq\frac{K(s-t)}{N}\int_0^{s-t}e^{-2\lambda_1 u}\mathrm{d}u\leq\frac{K(s-t)}{N},\tag{5.56}$$

where we used Lemma 5.4.2 in the last line. Consequently, substituting (5.56) into (5.55), we conclude

$$\sum_{i=1,i\neq j}^{N}\mathbb{E}\big[|X_{T,x_j}^{t,x_i,i,N}-X_{T,x_j}^{s,x_i,i,N}|^2\big]\leq\frac{K(s-t)}{N}e^{-2\lambda_3(T-s)}.$$

□

5.4.2 Second Variation process

Let $T\geq s\geq t\geq 0$, $N\in\mathbb{N}$. The second variation process of $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\geq t\geq 0}$ is defined, for $i,j\in\{1,\ldots,N\}$, as

$$X_{s,x_j,x_k}^{t,x_i,i,N}=\int_t^s\sum_{l=1}^{N}\partial_{x_l}B_i(\boldsymbol{X}_u^{t,\boldsymbol{x},N})X_{u,x_j,x_k}^{t,x_l,l,N}\mathrm{d}u$$
$$+\int_t^s\sum_{l=1}^{N}\sum_{l'=1}^{N}\partial^2_{x_l,x_{l'}}B_i(\boldsymbol{X}_u^{t,\boldsymbol{x},N})X_{u,x_j}^{t,x_l,l,N}X_{u,x_k}^{t,x_{l'},l',N}\mathrm{d}u.\tag{5.57}$$

The following lemma proceeds the results in Lemma 5.4.2 and accounts for the different behaviours of L^p-moments for the second order variation processes defined in (5.57), which is needed in Lemma 5.4.7 and contributes to the analysis in Section 5.6.

Lemma 5.4.5. *Let Assumption 5.3.3 hold and let $p\geq 2$. Consider the second variation process (5.57) and assume that the starting positions $x_i\in L^2(\Omega,\mathbb{R})$ are $\mathcal{F}_t$-measurable random variables that are identically distributed over all $i\in\{1,\ldots,N\}$. Then there exist $\lambda_4\in(0,\min\{\lambda-(2+1/N)K_V,\lambda_3\})$ and $K>0$ (both independent of s,t,T and N) such that for any $T\geq s\geq t\geq 0$ and $i\in\{1,\ldots,N\}$*

$$\mathbb{E}\Big[|X_{s,x_i,x_i}^{t,x_i,i,N}|^p\Big]\leq Ke^{-\lambda_4 p(s-t)},\qquad \sum_{i,j,k=1,\;i\neq j\neq k}^{N}\mathbb{E}\Big[|X_{s,x_j,x_k}^{t,x_i,i,N}|^p\Big]\leq\frac{K}{N^{2p-3}}e^{-\lambda_4 p(s-t)},$$

$$\textit{and} \quad \sum_{i,k=1,\ i\neq k}^{N} \Big(\mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_k,x_k}|^p\Big] + \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_i,x_k}|^p\Big] + \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_k,x_i}|^p\Big]\Big) \leq \frac{K}{N^{p-2}} e^{-\lambda_4 p(s-t)}.$$

Remark 5.4.6. *This lemma continues to highlight the main difficulty faced in this manuscript's analysis. The above inequalities suggest that the second-order variation process, where $i = j = k$ implies $\hat{\mathcal{O}}((i,j,k)) = 1$, yields the $\mathcal{O}(1)$ behaviour, while all other elements (i.e., the cross-derivatives with $\hat{\mathcal{O}}((i,j,k)) \geq 2$) decay differently with respect to the number of particles. We refer to Lemma 5.4.7 for a general result.*

Proof. Let $p \geq 2$ be a given integer. Note that in the following proof, the positive constant K is independent of s, t, T, N and may change line by line.

Part 1: Preliminary manipulations. For $\lambda_4 \in (0, \min\{\lambda - (2 + 1/N)K_V, \lambda_3\})$, we define $I^{2,p}_{t,s}$ for all $s \geq t \geq 0$ as

$$\begin{aligned} I^{2,p}_{t,s} := e^{p\lambda_4(s-t)} \Big(& \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_i,x_i}|^p\Big] \\ & + N^{p-2} \sum_{i,k=1,\ i\neq k}^{N} \Big(\mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_k,x_k}|^p\Big] + \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_i,x_k}|^p\Big] + \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_k,x_i}|^p\Big]\Big) \\ & + N^{2p-3} \sum_{i,j,k=1,\ i\neq j\neq k}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j,x_k}|^p\Big]\Big), \end{aligned} \tag{5.58}$$

for which we will aim to show that we can upper bound $I^{2,p}_{t,s} \leq K$. We start by analysing each of the second variation processes: For any $i, j, k \in \{1, \ldots, N\}$, $\lambda_4 \in (0, \min\{\lambda - (2+1/N)K_V, \lambda_3\})$, $s \geq t \geq 0$, we have that

$$\begin{aligned} & e^{p\lambda_4(s-t)} \mathbb{E}\Big[|X^{t,x_i,i,N}_{s,x_j,x_k}|^p\Big] \\ & = -p \int_t^s e^{p\lambda_4(u-t)} \mathbb{E}\Big[\big(X^{t,x_i,i,N}_{u,x_j,x_k}\big) \cdot \big(\nabla^2 U(X^{t,x_i,i,N}_u) X^{t,x_i,i,N}_{u,x_j,x_k}\big) |X^{t,x_i,i,N}_{u,x_j,x_k}|^{p-2}\Big] \mathrm{d}u \\ & \quad - p \int_t^s e^{p\lambda_4(u-t)} \mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{u,x_j,x_k}\Big) \\ & \quad \cdot \Big(\sum_{l=1}^{N} \partial_{x_l} \frac{1}{N} \sum_{q=1}^{N} \nabla V(X^{t,x_i,i,N}_u - X^{t,x_q,q,N}_u) X^{t,x_l,l,N}_{u,x_j,x_k}\Big) |X^{t,x_i,i,N}_{u,x_j,x_k}|^{p-2}\Big] \mathrm{d}u \\ & \quad + p \int_t^s \lambda_4 e^{p\lambda_4(u-t)} \mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\Big] \mathrm{d}u \\ & \quad + p \int_t^s e^{p\lambda_4(u-t)} \mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{u,x_j,x_k}\Big) \\ & \quad \cdot \Big(\sum_{l=1}^{N} \sum_{l'=1}^{N} \partial^2_{x_l,x_{l'}} B_i(\boldsymbol{X}^{t,\boldsymbol{x},N}_u) X^{t,x_l,l,N}_{u,x_j} X^{t,x_{l'},l',N}_{u,x_k}\Big) |X^{t,x_i,i,N}_{u,x_j,x_k}|^{p-2}\Big] \mathrm{d}u \\ & \leq p \int_t^s (-\lambda + \lambda_4) e^{p\lambda_4(u-t)} \mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\Big] \mathrm{d}u && (5.59) \\ & \quad - p \int_t^s e^{p\lambda_4(u-t)} \mathbb{E}\Big[\Big(X^{t,x_i,i,N}_{u,x_j,x_k}\Big) && (5.60) \\ & \quad \cdot \Big(\frac{1}{N} \sum_{l=1}^{N} \nabla V(X^{t,x_i,i,N}_u - X^{t,x_l,l,N}_u)(X^{t,x_i,i,N}_{u,x_j,x_k} - X^{t,x_l,l,N}_{u,x_j,x_k})\Big) |X^{t,x_i,i,N}_{u,x_j,x_k}|^{p-2}\Big] \mathrm{d}u \\ & \quad + p \int_t^s e^{p\lambda_4(u-t)} \mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^{p-1} \sum_{l=1}^{N} \sum_{l'=1}^{N} |\partial^2_{x_l,x_{l'}} B_i(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)|\ |X^{t,x_l,l,N}_{u,x_j}|\ |X^{t,x_{l'},l',N}_{u,x_k}|\Big] \mathrm{d}u, && (5.61) \end{aligned}$$

where we used Assumption 5.2.1 on the first term to obtain (5.59). In what follows, the last two terms, which we will refer to as the convolution term (5.60) and the lower order variation term (5.61), will be investigated in more detail.

Part 2: Analysis of the convolution term (5.60). We analyze the convolution term (5.60) by considering five different cases: $i=j=k$, $i\neq j=k$, $i=j\neq k$, $i=k\neq j$ and $i\neq j\neq k$, where we will see that different methodologies need to be implemented based on the values $\hat{\mathcal{O}}((i,j,k))$ of the second variation processes.

Case $i=j=k$: The convolution term, after summing over all $i\in\{1,\dots,N\}$, simplifies to

$$
\begin{aligned}
&-\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbb{E}\left[\left(|X_{u,x_i,x_i}^{t,x_i,i,N}|^{p-2}X_{u,x_i,x_i}^{t,x_i,i,N}\right)\cdot\left(\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_l,l,N})(X_{u,x_i,x_i}^{t,x_i,i,N}-X_{u,x_i,x_i}^{t,x_l,l,N})\right)\right]\\
&=-\frac{1}{N}\sum_{i,l=1,\ i\neq l}^{N}\mathbb{E}\Big[\left(|X_{u,x_i,x_i}^{t,x_i,i,N}|^{p-2}X_{u,x_i,x_i}^{t,x_i,i,N}\right)\cdot\left(\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_l,l,N})X_{u,x_i,x_i}^{t,x_i,i,N}\right)\Big]\\
&\quad+\frac{1}{N}\sum_{i,l=1,\ i\neq l}^{N}\mathbb{E}\Big[\left(|X_{u,x_i,x_i}^{t,x_i,i,N}|^{p-2}X_{u,x_i,x_i}^{t,x_i,i,N}\right)\cdot\left(\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_l,l,N})X_{u,x_i,x_i}^{t,x_l,l,N}\right)\Big]\\
&\leq\frac{K_V}{N}\sum_{i,l=1,\ i\neq l}^{N}\mathbb{E}\Big[|X_{u,x_i,x_i}^{t,x_i,i,N}|^{p-1}\cdot|X_{u,x_i,x_i}^{t,x_l,l,N}|\Big] &(5.62)\\
&\leq\frac{K_V(p-1)}{p}\sum_{i=1}^{N}\mathbb{E}\Big[|X_{u,x_i,x_i}^{t,x_i,i,N}|^{p}\Big]+\frac{K_V}{Np}\sum_{i,l=1,\ i\neq l}^{N}\mathbb{E}\Big[|X_{u,x_i,x_i}^{t,x_l,l,N}|^{p}\Big], &(5.63)
\end{aligned}
$$

where we used $\nabla^2V(x)\geq 0$ for all $x\in\mathbb{R}$ and Assumption 5.2.1 to derive (5.62) and (5.63) is a consequence of Young's inequality.

Case $i\neq j=k$: In this situation, after summing over all $i,k\in\{1,\dots,N\}$, $i\neq k$ and splitting the summation over l, we derive using the *symmetrization trick* that

$$
\begin{aligned}
&-\sum_{i,k=1,\ i\neq k}^{N}\mathbb{E}\Big[\frac{1}{N}\sum_{l=1}^{N}\left(|X_{u,x_k,x_k}^{t,x_i,i,N}|^{p-2}X_{u,x_k,x_k}^{t,x_i,i,N}\right)\\
&\qquad\cdot\left(\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_l,l,N})(X_{u,x_k,x_k}^{t,x_i,i,N}-X_{u,x_k,x_k}^{t,x_l,l,N})\right)\Big]\\
&=-\frac{1}{2N}\sum_{i,k,l=1,\ i,l\neq k}^{N}\mathbb{E}\Big[\left(|X_{u,x_k,x_k}^{t,x_i,i,N}|^{p-2}X_{u,x_k,x_k}^{t,x_i,i,N}-|X_{u,x_k,x_k}^{t,x_l,l,N}|^{p-2}X_{u,x_k,x_k}^{t,x_l,l,N}\right)\\
&\qquad\cdot\left(\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_l,l,N})(X_{u,x_k,x_k}^{t,x_i,i,N}-X_{u,x_k,x_k}^{t,x_l,l,N})\right)\Big]\\
&\quad-\mathbb{E}\Big[\frac{1}{N}\sum_{i,k=1,\ i\neq k}^{N}\left(|X_{u,x_k,x_k}^{t,x_i,i,N}|^{p-2}X_{u,x_k,x_k}^{t,x_i,i,N}\right)\\
&\qquad\cdot\left(\nabla^2V(X_u^{t,x_i,i,N}-X_u^{t,x_k,k,N})(X_{u,x_k,x_k}^{t,x_i,i,N}-X_{u,x_k,x_k}^{t,x_k,k,N})\right)\Big]\\
&\leq K_V\sum_{i,k=1,\ i\neq k}^{N}\mathbb{E}\Big[|X_{u,x_k,x_k}^{t,x_i,i,N}|^{p-1}\cdot\frac{|X_{u,x_k,x_k}^{t,x_k,k,N}|}{N}\Big] &(5.64)\\
&\leq\frac{K_V(p-1)}{p}\sum_{i,k=1,\ i\neq k}^{N}\mathbb{E}\Big[|X_{u,x_k,x_k}^{t,x_i,i,N}|^{p}\Big]+\frac{K_V}{pN^{p-1}}\sum_{k=1}^{N}\mathbb{E}\Big[|X_{u,x_k,x_k}^{t,x_k,k,N}|^{p}\Big], &(5.65)
\end{aligned}
$$

where we used $\nabla^2V(x)\geq 0$ for all $x\in\mathbb{R}$ and Assumption 5.2.1 to derive (5.64). (5.65) is a consequence of Young's inequality.

Cases $i = j \neq k$ and $i = k \neq j$: These two cases share similar calculations and we show the first of these. Summing over all $i, k \in \{1, \dots, N\},\ i \neq k$, we have

$$
\begin{aligned}
&- \sum_{i,k=1,\ i\neq k}^{N} \mathbb{E}\Big[\frac{1}{N} \sum_{l=1}^{N} \Big(|X_{u,x_i,x_k}^{t,x_i,i,N}|^{p-2} X_{u,x_i,x_k}^{t,x_i,i,N} \Big) \\
&\qquad\qquad\qquad \cdot \Big(\nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_l,l,N})(X_{u,x_i,x_k}^{t,x_i,i,N} - X_{u,x_i,x_k}^{t,x_l,l,N}) \Big) \Big] \\
&\le -\frac{1}{N} \sum_{i,k,l=1,\ i\neq l\neq k}^{N} \mathbb{E}\Big[\Big| \Big(|X_{u,x_i,x_k}^{t,x_i,i,N}|^{p-2} X_{u,x_i,x_k}^{t,x_i,i,N} \Big) \cdot \Big(\nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_l,l,N}) X_{u,x_i,x_k}^{t,x_l,l,N} \Big) \Big| \Big] \\
&\quad + \frac{1}{N} \sum_{i,k=1,\ i\neq k}^{N} \mathbb{E}\Big[\Big| \Big(|X_{u,x_i,x_k}^{t,x_i,i,N}|^{p-2} X_{u,x_i,x_k}^{t,x_i,i,N} \Big) \cdot \Big(\nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_k,k,N}) X_{u,x_i,x_k}^{t,x_k,k,N} \Big) \Big| \Big] \\
&\le \frac{K_V}{N} \sum_{i,k,l=1,\ i\neq l\neq k}^{N} \Big(\frac{p-1}{p} \mathbb{E}\Big[|X_{u,x_i,x_k}^{t,x_i,i,N}|^p \Big] + \frac{1}{p} \mathbb{E}\Big[|X_{u,x_i,x_k}^{t,x_l,l,N}|^p \Big] \Big) \\
&\qquad\qquad + \frac{K_V}{N} \sum_{i,k=1,\ i\neq k}^{N} \Big(\frac{p-1}{p} \mathbb{E}\Big[|X_{u,x_i,x_k}^{t,x_i,i,N}|^p \Big] + \frac{1}{p} \mathbb{E}\Big[|X_{u,x_i,x_k}^{t,x_k,k,N}|^p \Big] \Big)
\end{aligned} \tag{5.66}
$$

$$
\begin{aligned}
&\le \frac{K_V(p-1)}{p} \sum_{i,k=1,\ i\neq k}^{N} \mathbb{E}\Big[|X_{u,x_i,x_k}^{t,x_i,i,N}|^p \Big] + \frac{K_V}{Np} \sum_{i,k,l=1,\ i\neq l\neq k}^{N} \mathbb{E}\Big[|X_{u,x_i,x_k}^{t,x_l,l,N}|^p \Big] \\
&\quad + \frac{K_V}{N} \sum_{i,k=1,\ i\neq k}^{N} \Big(\frac{p-1}{p} \mathbb{E}\Big[|X_{u,x_i,x_k}^{t,x_i,i,N}|^p \Big] + \frac{1}{p} \mathbb{E}\Big[|X_{u,x_i,x_k}^{t,x_k,k,N}|^p \Big] \Big),
\end{aligned} \tag{5.67}
$$

where once again, we split up the summation over l and repeatedly apply Assumption 5.2.1 and Young's inequality to obtain (5.66) and merely rearranging terms yields (5.67).

Case $i \neq j \neq k$**:** Summing over all $i, j, k \in \{1, \dots, N\},\ i \neq j \neq k$, we obtain

$$
\begin{aligned}
&- \sum_{i,j,k=1,\ i\neq j\neq k}^{N} \mathbb{E}\Big[\frac{1}{N} \sum_{l=1}^{N} \Big(|X_{u,x_j,x_k}^{t,x_i,i,N}|^{p-2} X_{u,x_j,x_k}^{t,x_i,i,N} \Big) \\
&\qquad \cdot \Big(\nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_l,l,N})(X_{u,x_j,x_k}^{t,x_i,i,N} - X_{u,x_j,x_k}^{t,x_l,l,N}) \Big) \Big] \\
&= -\frac{1}{2N} \sum_{i,j,k,l=1,\ i\neq j\neq k,\ l\neq j\neq k}^{N} \mathbb{E}\Big[\Big(|X_{u,x_j,x_k}^{t,x_i,i,N}|^{p-2} X_{u,x_j,x_k}^{t,x_i,i,N} - |X_{u,x_j,x_k}^{t,x_l,l,N}|^{p-2} X_{u,x_j,x_k}^{t,x_l,l,N} \Big) \\
&\qquad \cdot \Big(\nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_l,l,N})(X_{u,x_j,x_k}^{t,x_i,i,N} - X_{u,x_j,x_k}^{t,x_l,l,N}) \Big) \Big] \\
&\quad - \mathbb{E}\Big[\frac{1}{N} \sum_{i,j,k=1,\ i\neq j\neq k}^{N} \Big(|X_{u,x_j,x_k}^{t,x_i,i,N}|^{p-2} X_{u,x_j,x_k}^{t,x_i,i,N} \Big) \\
&\qquad \cdot \Big(\nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_j,j,N})(X_{u,x_j,x_k}^{t,x_i,i,N} - X_{u,x_j,x_k}^{t,x_j,j,N}) \Big) \Big] \\
&\quad - \mathbb{E}\Big[\frac{1}{N} \sum_{i,j,k=1,\ i\neq j\neq k}^{N} \Big(|X_{u,x_j,x_k}^{t,x_i,i,N}|^{p-2} X_{u,x_j,x_k}^{t,x_i,i,N} \Big) \\
&\qquad \cdot \Big(\nabla^2 V(X_u^{t,x_i,i,N} - X_u^{t,x_k,k,N})(X_{u,x_j,x_k}^{t,x_i,i,N} - X_{u,x_j,x_k}^{t,x_k,k,N}) \Big) \Big]
\end{aligned} \tag{5.68}
$$

$$\leq K_V \sum_{i,j,k=1,\ i\neq j\neq k}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^{p-1}\cdot\Big(\frac{1}{N}|X^{t,x_j,j,N}_{u,x_j,x_k}|+\frac{1}{N}|X^{t,x_k,k,N}_{u,x_j,x_k}|\Big)\Big] \tag{5.69}$$

$$\leq K_V \sum_{i,j,k=1,\ i\neq j\neq k}^{N} \Big(\frac{2(p-1)}{p}\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\Big]+\frac{1}{pN^p}\Big(\mathbb{E}\Big[|X^{t,x_j,j,N}_{u,x_j,x_k}|^p\Big]+\mathbb{E}\Big[|X^{t,x_k,k,N}_{u,x_j,x_k}|^p\Big]\Big)\Big) \tag{5.70}$$

$$\leq \frac{2K_V(p-1)}{p}\sum_{i,j,k=1,\ i\neq j\neq k}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\Big] + \frac{K_V}{pN^{p-1}}\sum_{j,k=1,j\neq k}^{N}\Big(\mathbb{E}\Big[|X^{t,x_j,j,N}_{u,x_j,x_k}|^p\Big]+\mathbb{E}\Big[|X^{t,x_k,k,N}_{u,x_j,x_k}|^p\Big]\Big). \tag{5.71}$$

This is once again established via splitting up the sum over $l \in \{1,\dots,N\}$ and using the *symmetrization trick* to obtain (5.68) and Assumption 5.2.1 to obtain (5.69). Young's inequality yields (5.70) before a final rearrangement of terms provides the final inequality (5.71).

Part 3: Analysis of the lower order variation term (5.61). Lemma 5.4.2 with $\lambda_1 \in (\lambda_2,\lambda)$ implies that there exists $K>0$ such that for all $i,j \in \{1,\dots,N\}, i\neq j,\ s\geq t\geq 0$,

$$\begin{aligned}\mathbb{E}[|X^{t,x_i,i,N}_{s,x_i}|^p] &\leq \sum_{k=1}^{N}\mathbb{E}[|X^{t,x_k,k,N}_{s,x_i}|^p]\leq Ke^{-\lambda_1 p(s-t)},\\ \mathbb{E}[|X^{t,x_i,i,N}_{s,x_j}|^p] &= \frac{1}{N-1}\sum_{k=1,\ k\neq j}^{N}\mathbb{E}[|X^{t,x_k,k,N}_{s,x_j}|^p]\leq \frac{K}{N^p}e^{-\lambda_1 p(s-t)},\end{aligned} \tag{5.72}$$

where for the equality in the second line we used that the starting points x_i are identically distributed. For the lower-order variation terms (5.61), using Hölder's inequality, we have that for all $i,j,k \in \{1,\dots,N\}$ and $u\geq t\geq 0$,

$$\begin{aligned}&e^{p\lambda_4(u-t)}\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^{p-1}\sum_{l=1}^{N}\sum_{l'=1}^{N}|\partial^2_{x_l,x_{l'}}B_i(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)|\ |X^{t,x_l,l,N}_{u,x_j}|\ |X^{t,x_{l'},l',N}_{u,x_k}|\Big]\\ &\leq e^{p\lambda_4(u-t)}\sum_{l=1}^{N}\sum_{l'=1}^{N}|\partial^2_{x_l,x_{l'}}B_i|_\infty\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\right]\right)^{(p-1)/p}\cdot\left(\mathbb{E}\left[|X^{t,x_l,l,N}_{u,x_j}|^p\ |X^{t,x_{l'},l',N}_{u,x_k}|^p\right]\right)^{1/p}\\ &=: e^{p\lambda_4(u-t)}\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\right]\right)^{(p-1)/p}\cdot I^{1,i,j,k}_{t,u}.\end{aligned} \tag{5.73}$$

For all $i,j,k \in \{1,\dots,N\},\ u\geq t\geq 0$, we defined the following:

$$\begin{aligned}I^{1,i,j,k}_{t,u} &:= \sum_{l=1}^{N}\sum_{l'=1}^{N}|\partial^2_{x_l,x_{l'}}B_i|_\infty\left(\mathbb{E}\left[|X^{t,x_l,l,N}_{u,x_j}|^p\ |X^{t,x_{l'},l',N}_{u,x_k}|^p\right]\right)^{1/p}\\ &\leq \sum_{l=1}^{N}\sum_{l'=1}^{N}|\partial^2_{x_l,x_{l'}}B_i|_\infty\left(\mathbb{E}\left[|X^{t,x_l,l,N}_{u,x_j}|^{2p}\right]\ \mathbb{E}\left[|X^{t,x_{l'},l',N}_{u,x_k}|^{2p}\right]\right)^{1/2p}\\ &= |\partial^2_{x_i,x_i}B_i|_\infty\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_j}|^{2p}\right]\ \mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_k}|^{2p}\right]\right)^{1/2p}\\ &\quad+\sum_{l'=1,l'\neq i}^{N}|\partial^2_{x_i,x_{l'}}B_i|_\infty\left(\mathbb{E}\left[|X^{t,x_i,N}_{u,x_j}|^{2p}\right]\ \mathbb{E}\left[|X^{t,x_{l'},l',N}_{u,x_k}|^{2p}\right]\right)^{1/2p}\\ &\quad+\sum_{l=1,l\neq i}^{N}|\partial^2_{x_l,x_i}B_i|_\infty\left(\mathbb{E}\left[|X^{t,x_l,N}_{u,x_j}|^{2p}\right]\ \mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_k}|^{2p}\right]\right)^{1/2p}\end{aligned}$$

$$+\sum_{l=1,\ i\neq l}^{N}\sum_{l'=1,\ i\neq l'}^{N}|\partial^2_{x_l,x_{l'}}B_i|_\infty\left(\mathbb{E}\left[|X^{t,x_l,l,N}_{u,x_j}|^{2p}\right]\ \mathbb{E}\left[|X^{t,x_{l'},l',N}_{u,x_k}|^{2p}\right]\right)^{1/2p},$$

where we implemented the Cauchy–Schwarz inequality in the first inequality. Now using Assumption 5.3.3,

$$|\partial^2_{x_l,x_{l'}}B_i|_\infty=\begin{cases}\mathcal{O}(1), & i=l=l',\\ \mathcal{O}(N^{-1}), & i=l\neq l'\ \textit{or}\ i=l'\neq l\ \textit{or}\ i\neq l=l',\\ 0, & i\neq l\neq l',\end{cases}$$

and the results in (5.72) show that there exists $K>0$ such that

$$I^{1,i,j,k}_{t,u}\leq Ke^{-2\lambda_1(u-t)}\cdot\begin{cases}1, & i=j=k,\\ N^{-1}, & i=j\neq k\ \textit{or}\ i=k\neq j\ \textit{or}\ i\neq j=k,\\ N^{-2}, & i\neq j\neq k.\end{cases}\tag{5.74}$$

Having established this general estimate for $I^{1,i,j,k}_{t,u}$ in (5.73), we distinguish the following scenarios:

Case 1: $i=j=k$. We have

$$\begin{aligned}&\sum_{i=1}^{N}e^{p\lambda_4(u-t)}\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_i,x_i}|^p\right]\right)^{(p-1)/p}\cdot I^{1,i,i,i}_{t,u}\\
&\leq Ke^{p\lambda_4(u-t)}\sum_{i=1}^{N}\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_i,x_i}|^p\right]\right)^{(p-1)/p}\cdot e^{-2\lambda_1(u-t)}\\
&=Ke^{(\lambda_4-2\lambda_1)(u-t)}\cdot\sum_{i=1}^{N}\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_i,x_i}|^p\right]\right)^{(p-1)/p}e^{(p-1)\lambda_4(u-t)}\\
&\leq Ke^{(\lambda_4-2\lambda_1)(u-t)}\cdot\left(N+e^{p\lambda_4(u-t)}\sum_{i=1}^{N}\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_i,x_i}|^p\right]\right),\end{aligned}\tag{5.75}$$

where we deploy (5.74) to obtain the first inequality. Using Young's inequality $ab\leq a^{q_1}/q_1+b^{q_2}/q_2$ with $a=\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_i,x_i}|^p\right]\right)^{(p-1)/p}e^{(p-1)\lambda_4(u-t)}$, $b=1$, $q_1=p/(p-1)$, $q_2=p$ yields (5.75).

Case 2: $i\neq j=k$. We have

$$\begin{aligned}&\sum_{i,k=1,\ i\neq k}^{N}e^{p\lambda_4(u-t)}\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_k,x_k}|^p\right]\right)^{(p-1)/p}\cdot I^{1,i,k,k}_{t,u}\\
&\leq e^{p\lambda_4(u-t)}\sum_{i,k=1,\ i\neq k}^{N}\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_k,x_k}|^p\right]\right)^{(p-1)/p}\cdot\frac{K}{N}e^{-2\lambda_1(u-t)}\\
&\leq Ke^{(\lambda_4-2\lambda_1)(u-t)}\cdot\left(\frac{1}{N^{p-2}}+e^{p\lambda_4(u-t)}\sum_{i,k=1,\ i\neq k}^{N}\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_k,x_k}|^p\right]\right),\end{aligned}\tag{5.76}$$

where we used Young's inequality $ab\leq a^{q_1}/q_1+b^{q_2}/q_2$ with $a=\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_k,x_k}|^p\right]\right)^{(p-1)/p}e^{(p-1)\lambda_1(u-t)}$, $b=1/N$, $q_1=p/(p-1)$, $q_2=p$ to derive (5.76).

Similar calculations apply to the cases $i=j\neq k$ and $i=k\neq j$:

$$\sum_{i,k=1,\ i\neq k}^{N}e^{p\lambda_4(u-t)}\left(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_i,x_k}|^p\right]\right)^{(p-1)/p}\cdot I^{1,i,i,k}_{t,u}$$

$$\leq Ke^{(\lambda_4-2\lambda_1)(u-t)}\cdot\Big(\frac{1}{N^{p-2}}+e^{p\lambda_4(u-t)}\sum_{i,k=1,\ i\neq k}^{N}\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_i,x_k}|^p\right]\Big),$$

$$\sum_{i,j=1,\ i\neq j}^{N}e^{p\lambda_4(u-t)}\Big(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_j,x_j}|^p\right]\Big)^{(p-1)/p}\cdot I^{1,i,j,i}_{t,u}$$

$$\leq Ke^{(\lambda_4-2\lambda_1)(u-t)}\cdot\Big(\frac{1}{N^{p-2}}+e^{p\lambda_4(u-t)}\sum_{i,k=j,\ i\neq j}^{N}\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_j,x_i}|^p\right]\Big).$$

Case 3: $i\neq j\neq k$. We have

$$\sum_{i,j,k=1,\ i\neq j\neq k}^{N}e^{p\lambda_4(u-t)}\Big(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\right]\Big)^{(p-1)/p}\cdot I^{1,i,j,k}_{t,u}$$

$$\leq e^{p\lambda_4(u-t)}\sum_{i,j,k=1,\ i\neq j\neq k}^{N}\Big(\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\right]\Big)^{(p-1)/p}\cdot\frac{K}{N^2}e^{-2\lambda_1(u-t)}$$

$$\leq Ke^{(\lambda_4-2\lambda_1)(u-t)}\cdot\Big(\frac{1}{N^{2p-3}}+e^{p\lambda_4(u-t)}\sum_{i,j,k=1,\ i\neq j\neq k}^{N}\mathbb{E}\left[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\right]\Big),\tag{5.77}$$

where we used Young's inequality $ab\leq a^{q_1}/q_1+b^{q_2}/q_2$ with $a=(\mathbb{E}[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p])^{(p-1)/p}e^{(p-1)\lambda_4(u-t)}$, $b=1/N^2$, $q_1=p/(p-1)$, $q_2=p$ to derive (5.77).

Part 4: Collecting estimates and conclusion. Gathering up all the results in (5.63)–(5.77) and recalling the definition of $I^{2,p}_{t,s}$ in (5.58), we have for all $\lambda_4\in(0,\min\{\lambda-(2+1/N)K_V,\lambda_3\})$,

$$\begin{aligned}
I^{2,p}_{t,s}\leq{}& p\int_t^s\Big(-\lambda+\lambda_4+\frac{2K_V(p-1)}{p}+Ke^{(\lambda_4-2\lambda_1)(u-t)}\Big)I^p_{t,u}\mathrm{d}u+pK\int_t^s e^{(\lambda_4-2\lambda_1)(u-t)}\mathrm{d}u\\
&+p\int_t^s\frac{K_V}{N^2p}\sum_{i,k=1,\ i\neq k}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_k,x_k}|^p\Big]\mathrm{d}u+p\int_t^s\frac{K_V}{Np}\sum_{i=1}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_i,x_i}|^p\Big]\mathrm{d}u\\
&+p\int_t^s\Big(\frac{2K_VN^{p-3}}{p}\sum_{i,j,k=1,\ i\neq j\neq k}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\Big]\\
&\qquad+K_VN^{p-3}\sum_{i,k=1,\ i\neq k}^{N}\Big(\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_i,x_k}|^p\Big]+\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_k,x_i}|^p\Big]\Big)\Big)\mathrm{d}u\\
&+p\int_t^s\frac{K_V}{p}N^{p-2}\sum_{i,k=1,\ i\neq k}^{N}\Big(\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_i,x_k}|^p\Big]+\mathbb{E}\Big[|X^{t,x_k,k,N}_{u,x_i,x_k}|^p\Big]\Big)\mathrm{d}u.
\end{aligned}$$

Combining the terms above, we further have

$$\begin{aligned}
I^{2,p}_{t,s}\leq{}& p\int_t^s\Big(-\lambda+\lambda_4+\frac{2K_V(p-1)}{p}+Ke^{(\lambda_4-2\lambda_1)(u-t)}\Big)I^p_{t,u}\mathrm{d}u+pK\int_t^s e^{(\lambda_4-2\lambda_1)(u-t)}\mathrm{d}u\\
&+p\int_t^s\frac{K_V}{p}\Big(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_i,x_i}|^p\Big]\\
&\qquad+N^{p-2}\sum_{i,k=1,\ i\neq k}^{N}\Big(\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_i,x_k}|^p\Big]+\mathbb{E}\Big[|X^{t,x_k,k,N}_{u,x_i,x_k}|^p\Big]\Big)\Big)\mathrm{d}u\\
&+p\int_t^s\frac{K_V}{N}\Big(N^{p-2}\sum_{i,k=1,\ i\neq k}^{N}\Big(\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_k,x_k}|^p\Big]+\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_i,x_k}|^p\Big]+\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_k,x_i}|^p\Big]\Big)
\end{aligned}$$

$$+\frac{2}{pN^{p-1}}N^{2p-3}\sum_{i,j,k=1,\ i\neq j\neq k}^{N}\mathbb{E}\Big[|X^{t,x_i,i,N}_{u,x_j,x_k}|^p\Big]\Big)\mathrm{d}u.$$

Hence, using $I^{2,p}_{t,s}$ in (5.58) to dominate the sums of expectations (in each integral) we have

$$I^{2,p}_{t,s}\leq p\int_t^s\Big(-\lambda+\lambda_4+2K_V+\frac{K_V}{N}+Ke^{(\lambda_4-2\lambda_1)(u-t)}\Big)I^{2,p}_{t,u}\mathrm{d}u+pK\int_t^s e^{(\lambda_4-2\lambda_1)(u-t)}\mathrm{d}u. \tag{5.78}$$

Recalling that is λ_4 chosen such that $\lambda_4\in(0,\min\{\lambda-(2+1/N)K_V,\lambda_3\})$ and $0<\lambda_3<\lambda_1<\lambda$, which in turn implies $\lambda_4-2\lambda_1<0$, we have

$$I^{2,p}_{t,s}\leq pK\int_t^s e^{(\lambda_4-2\lambda_1)(u-t)}I^{2,p}_{t,u}\mathrm{d}u+pK\int_t^s e^{(\lambda_4-2\lambda_1)(u-t)}\mathrm{d}u\leq K,$$

where for the last inequality, we used Gronwall's inequality (see Lemma 6.4.6) with

$$\alpha(s)=pKe^{(\lambda_4-2\lambda_1)(s-t)},\quad \beta(s)=pK\int_t^s e^{(\lambda_4-2\lambda_1)(u-t)}\mathrm{d}u\leq pK/(2\lambda_1-\lambda_4)\quad\text{and}$$

$u(s)=I^{2,p}_{t,s}$ in combination with the estimate $\int_t^u\alpha(s)\mathrm{d}s\leq pK/(2\lambda_1-\lambda_4)$.

□

5.4.3 The n-Variation process

Let $T\geq s\geq t\geq 0$, $N\in\mathbb{N}$ and $1<n\leq 6$ be an integer. Recalling the quantity Π^N_n in Definition 5.3.1, the n-variation process of $(\boldsymbol{X}^{t,\boldsymbol{x},N}_s)_{s\geq t\geq 0}$ is given for $i\in\{1,\dots,N\}$, $s\geq t$, $\gamma\in\Pi^N_n$, by

$$\begin{aligned}X^{t,x_i,i,N}_{s,x_{\gamma_1},\dots,x_{\gamma_n}}&=\int_t^s\Big(\sum_{l=1}^N\partial_{x_l}B_i(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)X^{t,x_l,l,N}_{u,x_{\gamma_1}}\Big)_{x_{\gamma_2},\dots,x_{\gamma_n}}\mathrm{d}u\\&=\int_t^s\sum_{l=1}^N\partial_{x_l}B_i(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)X^{t,x_l,l,N}_{u,x_{\gamma_1},\dots,x_{\gamma_n}}\mathrm{d}u\\&\quad+\sum_{\substack{\alpha,\beta\in\bigcup_{k=0}^{n-1}\Pi^N_k,\\|\alpha|>0,\ \gamma\setminus(\gamma_1)\in\alpha\sqcup\beta}}\int_t^s\sum_{l=1}^N\Big(\partial_{x_l}B_i(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)\Big)_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}}\Big(X^{t,x_l,l,N}_{u,x_{\gamma_1}}\Big)_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}}\mathrm{d}u,\end{aligned}\tag{5.79}$$

where $\alpha=(\alpha_1,\dots,\alpha_{|\alpha|})$, $\beta=(\beta_1,\dots,\beta_{|\beta|})$ with $\alpha_i\in\{1,\dots,N\}$, and $\beta_i\in\{1,\dots,N\}$ for $i\in\{1,\dots,|\beta|\}$. To clarify the notation, we present the following examples for $n\in\{2,3\}$:

$$\Big\{\alpha,\beta\in\bigcup_{k=0}^{1}\Pi^N_k:|\alpha|>0,\ \gamma\setminus(\gamma_1)\in\alpha\sqcup\beta\Big\}=\Big\{\{\alpha=(\gamma_2),\beta=\emptyset\}\Big\},$$

when $\gamma=(\gamma_1,\gamma_2)$; for the case $n=3$ and setting $\gamma=(\gamma_1,\gamma_2,\gamma_3)$

$$\begin{aligned}&\Big\{\alpha,\beta\in\bigcup_{k=0}^{2}\Pi^N_k:|\alpha|>0,\ \gamma\setminus(\gamma_1)\in\alpha\sqcup\beta\Big\}\\&\quad=\Big\{\{\alpha=(\gamma_2,\gamma_3),\beta=\emptyset\},\{\alpha=(\gamma_2),\beta=(\gamma_3)\},\{\alpha=(\gamma_3),\beta=(\gamma_2)\}\Big\}.\end{aligned}$$

The following lemma is a generalization of Lemma 5.4.2 and Lemma 5.4.5, which describes the behaviour of the higher-order variation processes and is needed in the proofs in Section 5.5 and Section 5.6.

Lemma 5.4.7. *Let the assumptions of Lemma 5.4.5 hold and let $p\geq 2$ be given. Consider the n-variation process with components $(X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n+1}}})_{s\geq t\geq 0}$ defined by* (5.79) *for $T\geq s\geq t\geq 0$,*

$n, N \in \mathbb{N}$, $\gamma \in \Pi_{n+1}^N$, $1 \le n \le 6$. *Then for each* $1 \le n \le 6$, *there exist constants* $\lambda_0^{(n)} \in (0, \lambda)$ *and* $K > 0$ *(both independent of* s, t, T *and* N*) such that for any* $m \in \{1, \dots, n+1\}$, *we have*

$$\sum_{\gamma \in \Pi_{n+1}^N,\ \hat{\mathcal{O}}(\gamma)=m} \mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n+1}}}|^p\Big] \le \frac{K}{N^{p(m-1)-m}} e^{-\lambda_0^{(n)} p(s-t)}.$$

In particular, this implies that, for all $\gamma \in \Pi_{n+1}^N$, *such that* $\hat{\mathcal{O}}(\gamma) = m$, $m \in \{1, \dots, n+1\}$, *we have*

$$\mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n+1}}}|^p\Big] \le \frac{K}{N^{p(m-1)}} e^{-\lambda_0^{(n)} p(s-t)}.$$

Remark 5.4.8. *We take* $\gamma \in \Pi_{n+1}^N$ *here in Lemma 5.4.7: this still corresponds to the highest order of the derivatives remaining as 6. We add the extra index since* γ_1 *corresponds to the index of the starting position, not an index which corresponds to any derivatives being taken.*

Remark 5.4.9. *As in Lemma 5.4.2, where the constant appearing in the bound of the second variation process* λ_4 *was strictly less than* λ_1 *(the constant which arose bounding the first variation process), we choose* $\lambda_0^{(n)}$ *such that the sequence* $\lambda_0^{(i)}$, $i \in \{1, \dots, 6\}$ *is a strictly decreasing sequence. Note that here, our* $\lambda_0^{(1)}$ *and* $\lambda_0^{(2)}$ *subsumes our* λ_1 *and* λ_4 *in the previous proofs respectively.*

Proof. Note that in the following proof, the positive constant K is independent of s, t, T, N and may change line by line.

Part 1: Preliminary manipulations. We shall prove this by induction. The result follows for $n = 1$ and $n = 2$, from Lemma 5.4.2 and Lemma 5.4.5 respectively.

Now, suppose that the claim holds for some $n_1 \in \mathbb{N}$, $n_1 < 6$, i.e., there exists a sufficiently small constant $\lambda_0^{(n_1)} \in (0, \lambda)$, such that for $m \in \{1, \dots, n_1 + 1\}$,

$$\sum_{\gamma \in \Pi_{n_1+1}^N,\ \hat{\mathcal{O}}(\gamma)=m} \mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+1}}}|^p\Big] \le \frac{K}{N^{p(m-1)-m}} e^{-\lambda_0^{(n_1)} p(s-t)}. \tag{5.80}$$

We need to prove the statement for $m \in \{1, \dots, n_1 + 2\}$, i.e., there exists some constant $\lambda_0^{(n_1+1)} \in (0, \lambda^{(n_1)})$ such that

$$\sum_{\gamma \in \Pi_{n_1+2}^N,\ \hat{\mathcal{O}}(\gamma)=m} \mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^p\Big] \le \frac{K}{N^{p(m-1)-m}} e^{-\lambda_0^{(n_1+1)} p(s-t)}.$$

We shall use the induction hypothesis (5.80) in *Part 2.2*.

From (5.79), we have for all $\gamma \in \Pi_{n_1+2}^N$

$$\begin{aligned}
&X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \\
&\quad = \int_t^s \Big(\sum_{l=1}^N \partial_{x_l} B_{\gamma_1}(\boldsymbol{X}_u^{t,\boldsymbol{x},N}) X^{t,x_l,l,N}_{u,x_{\gamma_2}}\Big)_{x_{\gamma_3},\dots,x_{\gamma_{n_1+2}}} \mathrm{d}u \\
&\quad = \int_t^s \sum_{l=1}^N \partial_{x_l} B_{\gamma_1}(\boldsymbol{X}_u^{t,\boldsymbol{x},N}) X^{t,x_l,l,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \mathrm{d}u \\
&\qquad + \sum_{\substack{\alpha,\beta \in \bigcup_{k=0}^{n_1} \Pi_k^N, \\ |\alpha|>0,\ \gamma \setminus (\gamma_1,\gamma_2) \in \alpha \sqcup \beta}} \int_t^s \sum_{l=1}^N \Big(\partial_{x_l} B_{\gamma_1}(\boldsymbol{X}_u^{t,\boldsymbol{x},N})\Big)_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}} \Big(X^{t,x_l,l,N}_{u,x_{\gamma_2}}\Big)_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}} \mathrm{d}u.
\end{aligned}$$

An application of Itô's formula yields for all $s \ge t$, $\lambda_0^{(n_1+1)} > 0$,

$$e^{p\lambda_0^{(n_1+1)}(s-t)} \mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^p\Big]$$

$$
\begin{aligned}
&= p\int_t^s e^{p\lambda_0^{(n_1+1)}(u-t)}\mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^{p-2}\Big(X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}\Big)\\
&\quad\cdot\Big(\sum_{l=1}^N \partial_{x_l}B_{\gamma_1}(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)X^{t,x_l,l,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}\\
&\qquad+\sum_{\substack{\alpha,\beta\in\bigcup_{k=0}^{n_1}\Pi_k^N,\\ |\alpha|>0,\ \gamma\setminus(\gamma_1,\gamma_2)\in\alpha\sqcup\beta}}\sum_{l=1}^N\Big(\partial_{x_l}B_{\gamma_1}(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)\Big)_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}}\Big(X^{t,x_l,l,N}_{u,x_{\gamma_2}}\Big)_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}}\Big)\Big]\mathrm{d}u\\
&\quad+p\lambda_0^{(n_1+1)}\int_t^s e^{p\lambda_0^{(n_1+1)}(u-t)}\mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^p\Big]\mathrm{d}u\\
&\le p\int_t^s(\lambda_0^{(n_1+1)}-\lambda)e^{p\lambda_0^{(n_1+1)}(u-t)}\mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^p\Big]\mathrm{d}u &&(5.81)\\
&\quad-p\int_t^s e^{p\lambda_0^{(n_1+1)}(u-t)}\mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^{p-2}\Big(X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}\Big)\\
&\qquad\cdot\Big(\frac{1}{N}\sum_{l=1}^N\nabla^2V(X^{t,x_{\gamma_1},\gamma_1,N}_u-X^{t,x_l,l,N}_u)\ \big(X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}-X^{t,x_l,l,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}\big)\Big)\Big]\mathrm{d}u\\
& &&(5.82)\\
&\quad-p\int_t^s e^{p\lambda_0^{(n_1+1)}(u-t)}\mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^{p-2}\Big(X^{t,x_{\gamma_1},\gamma_1,N}_{u,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}\Big)\\
&\qquad\cdot\Big(\sum_{\substack{\alpha,\beta\in\bigcup_{k=0}^{n_1}\Pi_k^N,\\ |\alpha|>0,\ \gamma\setminus(\gamma_1,\gamma_2)\in\alpha\sqcup\beta}}\sum_{l=1}^N\Big(\partial_{x_l}B_{\gamma_1}(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)\Big)_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}}\Big(X^{t,x_l,l,N}_{u,x_{\gamma_2}}\Big)_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}}\Big)\Big]\mathrm{d}u.\\
& &&(5.83)
\end{aligned}
$$

The analysis of the convolution term (5.82) and the lower order variation term (5.83), mimic that of the proof of Lemma 5.4.5 - all arguments are complete generalizations. We present these for clarity.

For a given $p\ge 2$, we define the process $I^{n_1+1,p}_{t,s}$ (similar to the term $I^{2,p}_{t,s}$ from (5.58)):

$$
\begin{aligned}
I^{n_1+1,p}_{t,s} :=& e^{p\lambda_0^{(n_1+1)}(s-t)}\sum_{m=1}^{n_1+2}\Big(N^{(m-1)p-m}\sum_{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m}\mathbb{E}\Big[|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^p\Big]\Big) \qquad (5.84)\\
\le& \sum_{m=1}^{n_1+2}\Big(N^{(m-1)p-m}\sum_{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m}(5.81)+(5.82)+(5.83)\Big),
\end{aligned}
$$

which we now work towards bounding $I^{n_1+1,p}_{t,s}\le K$, by separating our analysis into that of the convolution term (5.82) and the lower order variation term (5.83).

Part 2.1: Analysis of the convolution term (5.82). We highlight the main steps that deal specifically with the convolution term (5.82) (similar to (5.63)–(5.71) in the proof steps for $n=2$). Summing over $\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m$, gives

$$
\begin{aligned}
&-\sum_{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m}\mathbb{E}\Big[\frac{1}{N}\sum_{l=1}^N\Big(|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^{p-2}X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}\Big)\\
&\qquad\cdot\Big(\nabla^2V(X^{t,x_{\gamma_1},\gamma_1,N}_s-X^{t,x_l,l,N}_s)(X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}-X^{t,x_l,l,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}})\Big)\Big]\\
&=-\Big(\sum_{\substack{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1))=m}}+\sum_{\substack{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1))=m-1}}\Big)\mathbb{E}\Big[\frac{1}{N}\sum_{l=1}^N\Big(|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^{p-2}X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}\Big)
\end{aligned}
$$

$$\cdot\Big(\nabla^2 V(X_s^{t,x_{\gamma_1},\gamma_1,N} - X_s^{t,x_l,l,N})(X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N} - X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_l,l,N})\Big)\Big]. \tag{5.85}$$

To continue from (5.85), we need to consider whether or not γ_1 is a unique index in γ, i.e., the two sums appearing in (5.85).

Case 1: γ_1 is not a unique index in γ (i.e., $\hat{\mathcal{O}}(\gamma) = \hat{\mathcal{O}}(\gamma\setminus(\gamma_1)) = m$). Note that this case is only meaningful for $m \in \{1,\dots,n_1+1\}$, as γ has at most n_1+1 different elements (note that γ_1 appears at least twice). Further, we remark that $\hat{\mathcal{O}}(\gamma) = m$ but $\hat{\mathcal{O}}(\gamma\setminus(\gamma_1)\bigcup(l)) \in \{m, m+1\}$, and therefore we need to consider $X_\cdot^{t,x_{\gamma_1},\gamma_1,N}$ and $X_\cdot^{t,x_l,l,N}$ separately (in a similar fashion to (5.63) and (5.67)). Hence, we write

$$\begin{aligned}
&-\sum_{\substack{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1))=m}} \mathbb{E}\Big[\frac{1}{N}\sum_{l=1}^N \Big(|X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N}|^{p-2} X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N}\Big)\\
&\qquad\cdot\Big(\nabla^2 V(X_s^{t,x_{\gamma_1},\gamma_1,N} - X_s^{t,x_l,l,N})(X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N} - X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_l,l,N})\Big)\Big]\\
&= -\frac{1}{N}\sum_{\substack{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1))=m}} \Big(\sum_{\substack{l=1,\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1)\bigcup(l))=m}}^N + \sum_{\substack{l=1,\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1)\bigcup(l))=m+1}}^N\Big)\\
&\qquad\cdot\mathbb{E}\Big[\Big(|X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N}|^{p-2} X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N}\Big)\\
&\qquad\cdot\Big(\nabla^2 V(X_s^{t,x_{\gamma_1},\gamma_1,N} - X_s^{t,x_l,l,N})(X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N} - X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_l,l,N})\Big)\Big].
\end{aligned} \tag{5.86}$$

We address each of the inner sums over l in (5.86) separately. Concretely: for the first inner sum, where the summation is over $\hat{\mathcal{O}}(\gamma\setminus(\gamma_1)\bigcup(l)) = m$ with l taking values in $\gamma\setminus(\gamma_1)$ (as per the outer summation), we use the *symmetrization trick* to get (5.87) and we substitute the upper indices $\gamma_1 \mapsto l_1$ and $l \mapsto l_2$ for added clarity. For the second inner sum, where the summation is over $\hat{\mathcal{O}}(\gamma\setminus(\gamma_1)\bigcup(l)) = m+1$ with $l \notin \gamma$, we apply Young's inequality to get (5.88). We point out that similar arguments were employed in the estimation of the second variation process – refer to (5.63) and (5.67) for example. We then have,

$$\begin{aligned}
&(5.86)\\
&\le -\frac{1}{2N}\sum_{\substack{\gamma\in\Pi^N_{n_1+1},\ \hat{\mathcal{O}}(\gamma)=m\\ l_1,l_2\in\gamma}} \mathbb{E}\Big[\Big(|X_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}^{t,x_{l_1},l_1,N}|^{p-2} X_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}^{t,x_{l_1},l_1,N}\\
&\qquad - |X_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}^{t,x_{l_2},l_2,N}|^{p-2} X_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}^{t,x_{l_2},l_2,N}\Big)\\
&\qquad\cdot\Big(\nabla^2 V(X_s^{t,x_{l_1},l_1,N} - X_s^{t,x_{l_2},l_2,N})(X_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}^{t,x_{l_1},l_1,N} - X_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}^{t,x_{l_2},l_2,N})\Big)\Big] \qquad (5.87)\\
&\quad + \frac{K_V}{N}\sum_{\substack{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1))=m}} \sum_{\substack{l=1,\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1)\bigcup(l))=m+1}}^N\\
&\qquad\Big(\frac{p-1}{p}\mathbb{E}\Big[\Big|X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N}\Big|^p\Big] + \frac{1}{p}\mathbb{E}\Big[\Big|X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_l,l,N}\Big|^p\Big]\Big) \qquad (5.88)\\
&\le \frac{K_V(p-1)}{p}\sum_{\substack{\gamma\in\Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m\\ \hat{\mathcal{O}}(\gamma\setminus(\gamma_1))=m}} \mathbb{E}\Big[\Big|X_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}^{t,x_{\gamma_1},\gamma_1,N}\Big|^p\Big]
\end{aligned}$$

$$+ \frac{K_V}{Np} \sum_{\substack{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m+1 \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1))=m}} \mathbb{E}\Big[\Big| X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big|^p \Big], \tag{5.89}$$

where (5.89) follows from the fact $\nabla^2 V(x) \geq 0$ for all $x \in \mathbb{R}$ (implying that (5.87) is bounded above by zero), and an application of Young's inequality is used to get (5.88). This will provide acceptable weights in (5.94) below.

Case 2: γ_1 is a unique index in γ (i.e., $\hat{\mathcal{O}}(\gamma) = m$, $\hat{\mathcal{O}}(\gamma \setminus (\gamma_1)) = m-1$). Note that this case is only meaningful for $m \in \{2, \dots, n_1+2\}$, as γ has at least 2 distinct elements (at least one of $\gamma_2, \dots, \gamma_{n_1+2}$ has to be different to γ_1). Further, we remark that $\hat{\mathcal{O}}(\gamma) = m$ but $\hat{\mathcal{O}}(\gamma \setminus (\gamma_1) \bigcup (l)) \in \{m, m-1\}$, and therefore we need to consider $X^{t,x_{\gamma_1},\gamma_1,N}_{\cdot}$ and $X^{t,x_l,l,N}_{\cdot}$ separately (in a similar fashion to (5.65) and (5.71)). Hence, we write

$$\begin{aligned}
&- \sum_{\substack{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1))=m-1}} \mathbb{E}\Big[\frac{1}{N} \sum_{l=1}^N \Big(|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^{p-2} X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big) \\
&\qquad \cdot \Big(\nabla^2 V(X^{t,x_{\gamma_1},\gamma_1,N}_s - X^{t,x_l,l,N}_s)(X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} - X^{t,x_l,l,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}) \Big) \Big] \\
&= -\frac{1}{N} \sum_{\substack{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1))=m-1}} \Bigg(\sum_{\substack{l=1, \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1) \bigcup (l))=m}}^N + \sum_{\substack{l=1, \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1) \bigcup (l))=m-1}}^N \Bigg) \mathbb{E}\Big[\Big(|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^{p-2} \\
&\quad \cdot X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big) \cdot \Big(\nabla^2 V(X^{t,x_{\gamma_1},\gamma_1,N}_s - X^{t,x_l,l,N}_s)(X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} - X^{t,x_l,l,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}) \Big) \Big].
\end{aligned} \tag{5.90}$$

We exploit the *symmetrization trick* for the summation over $\hat{\mathcal{O}}(\gamma \setminus (\gamma_1) \bigcup (l)) = m$ where l takes values out of $\gamma \setminus (\gamma_1)$, and Young's inequality for the summation over $\hat{\mathcal{O}}(\gamma \setminus (\gamma_1) \bigcup (l)) = m-1$ where $l \in (\gamma \setminus (\gamma_1))$ (check (5.65) and (5.71) for example). As before, substitute the new upper indices (5.90): $\gamma_1 \mapsto l_1$ and $l \mapsto l_2$ to obtain

$$\begin{aligned}
(5.90) \leq &-\frac{1}{2N} \sum_{\substack{\gamma \in \Pi^N_{n_1+1},\ \hat{\mathcal{O}}(\gamma)=m-1 \\ l_1, l_2 \notin \gamma}} \mathbb{E}\Big[\Big(|X^{t,x_{l_1},l_1,N}_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}|^{p-2} X^{t,x_{l_1},l_1,N}_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}} - |X^{t,x_{l_2},l_2,N}_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}|^{p-2} \\
&\cdot X^{t,x_{l_2},l_2,N}_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}} \Big) \cdot \Big(\nabla^2 V(X^{t,x_{l_1},l_1,N}_s - X^{t,x_{l_2},l_2,N}_s)(X^{t,x_{l_1},l_1,N}_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}} - X^{t,x_{l_2},l_2,N}_{s,x_{\gamma_1},\dots,x_{\gamma_{n_1+1}}}) \Big) \Big]
\end{aligned} \tag{5.91}$$

$$\begin{aligned}
&+ K_V \sum_{\substack{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1))=m-1}} \sum_{\substack{l=1, \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1) \bigcup (l))=m-1}}^N \qquad (5.92) \\
&\qquad \Big(\frac{p-1}{p} \mathbb{E}\Big[\Big| X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big|^p \Big] + \frac{1}{pN^p} \mathbb{E}\Big[\Big| X^{t,x_l,l,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big|^p \Big] \Big) \\
\leq\ & \frac{6K_V(p-1)}{p} \sum_{\substack{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1))=m-1}} \mathbb{E}\Big[\Big| X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big|^p \Big] \\
&+ \frac{6K_V}{pN^p} \sum_{\substack{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m-1 \\ \hat{\mathcal{O}}(\gamma \setminus (\gamma_1))=m-1}} \mathbb{E}\Big[\Big| X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big|^p \Big], \qquad (5.93)
\end{aligned}$$

where we used $\nabla^2 V(x) \geq 0$ to bound (5.91) above by 0, then apply Jensen's inequality to (5.92) to match the acceptable weights on N in (5.94). The constant 6 arises from the fact that there are at most 6 different values for $l \in \gamma \setminus (\gamma_1), m \le n_1 + 2 \le 7$ such that $\hat{\mathcal{O}}(\gamma \setminus (\gamma_1) \bigcup (l)) = \hat{\mathcal{O}}(\gamma \setminus (\gamma_1)) = m - 1$.

Combining case 1 and case 2: We inject the established estimates in (5.89) and (5.93) into (5.85) and consider the summation (5.84), (noting the inclusion of the $N^{(m-1)p-m}$ term in the summand) to obtain

$$\begin{aligned}
&\sum_{m=1}^{n_1+2} \Bigg(N^{(m-1)p-m} \sum_{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m} \mathbb{E}\Bigg[\frac{1}{N} \sum_{l=1}^{N} \Big(|X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}|^{p-2} X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big) \\
&\qquad \cdot \Big(\nabla^2 V(X^{t,x_{\gamma_1},\gamma_1,N}_s - X^{t,x_l,l,N}_s)(X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} - X^{t,x_l,l,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}}) \Big) \Bigg] \Bigg) \\
&\le \Big(\frac{6K_V(p-1)}{p} + \frac{K_V}{Np} + \frac{6K_V}{Np} \Big) \sum_{m=1}^{n_1+2} \Bigg(N^{(m-1)p-m} \sum_{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m} \mathbb{E}\Big[\Big| X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big|^p \Big] \Bigg) \\
&\le K_V \Big(6 + \frac{1}{Np} \Big) \sum_{m=1}^{n_1+2} \Bigg(N^{(m-1)p-m} \sum_{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m} \mathbb{E}\Big[\Big| X^{t,x_{\gamma_1},\gamma_1,N}_{s,x_{\gamma_2},\dots,x_{\gamma_{n_1+2}}} \Big|^p \Big] \Bigg) \qquad (5.94) \\
&= K_V \Big(6 + \frac{1}{Np} \Big) e^{-\lambda_0^{(n_1+1)}(s-t)} I^{n_1+1,p}_{s,t}. \qquad (5.95)
\end{aligned}$$

Part 2.2: Analysis of the lower order variation term (5.83). We provide the following details on the derivatives of the function B_{γ_1}: We notice that, from Assumption 5.3.3, for any $\boldsymbol{x} \in \mathbb{R}^N$,

$$\partial_{x_l} B_{\gamma_1} = \begin{cases} -\nabla^2 U(x_{\gamma_1}) - \frac{1}{N}\sum_{j=1}^N \nabla^2 V(x_{\gamma_1} - x_j), & \gamma_1 = l, \\ -\frac{1}{N}\partial_{x_l} \sum_{j=1}^N \nabla V(x_{\gamma_1} - x_j) = \frac{1}{N}\nabla^2 V(x_{\gamma_1} - x_l), & \gamma_1 \ne l. \end{cases}$$

Thus, we have

$$|\partial_{x_l} B_{\gamma_1}|_\infty \le \begin{cases} |\nabla^2 U|_\infty + \frac{1}{N}\sum_{j=1}^N |\nabla^2 V|_\infty = \mathcal{O}(1), & \gamma_1 = l, \\ \frac{1}{N}|\nabla^2 V|_\infty = \mathcal{O}(N^{-1}), & \gamma_1 \ne l, \end{cases} \qquad (5.96)$$

and

$$|\partial^2_{x_l,x_{l'}} B_{\gamma_1}|_\infty \le \begin{cases} |\nabla^3 U|_\infty + \frac{1}{N}\sum_{j=1}^N |\nabla^3 V|_\infty = \mathcal{O}(1), & \gamma_1 = l = l', \\ \frac{1}{N}|\nabla^3 V|_\infty = \mathcal{O}(N^{-1}), & \gamma_1 = l \ne l', \\ \frac{1}{N}|\nabla^3 V|_\infty = \mathcal{O}(N^{-1}), & \gamma_1 = l' \ne l, \\ \frac{1}{N}|\nabla^3 V|_\infty = \mathcal{O}(N^{-1}), & \gamma_1 \ne l = l', \\ 0, & \gamma_1 \ne l \ne l'. \end{cases}$$

Using this methodology, one can easily establish that for $n \ge 1,\ \eta \in \Pi^N_n$

$$|\partial^n_{x_{\eta_1},\dots,x_{\eta_n}} B_{\gamma_1}|_\infty = \begin{cases} \mathcal{O}(N^{1-\hat{\mathcal{O}}(\eta \bigcup (\gamma_1))}), & \hat{\mathcal{O}}(\eta \bigcup (\gamma_1)) \in \{1,2\}, \\ 0, & \hat{\mathcal{O}}(\eta \bigcup (\gamma_1)) \ge 3, \end{cases} \qquad (5.97)$$

with an implied constant independent of N, since only tuples of the form $(\gamma_1, \eta_1, \dots, \eta_n)$, where elements take at most two different values yield a non-zero contribution in (5.97). Proceeding exactly as in the manner of (5.74)–(5.77), applying our induction hypothesis (5.80), we can establish the existence of positive constants K and α such that

$$\sum_{m=1}^{n_1+2} \Bigg(N^{(m-1)p-m} \sum_{\gamma \in \Pi^N_{n_1+2},\ \hat{\mathcal{O}}(\gamma)=m} (5.83) \Bigg) \le K \int_t^s e^{-\alpha(u-t)} \mathrm{d}u + K \int_t^s e^{-\alpha(u-t)} I^{n_1+1,p}_{t,u} \mathrm{d}u. \qquad (5.98)$$

Part 3: Collecting terms and conclusion. By collecting (5.81), (5.95) and (5.98), we have that

$$\begin{aligned} I_{t,s}^{n_1+1,p} \leq & K\int_t^s e^{-\alpha(u-t)}\mathrm{d}u \\ & + p\Big(\lambda_0^{(n_1+1)} - \lambda + K_V(6+\frac{1}{Np})\Big)\int_t^s I_{t,u}^{n_1+1,p}\mathrm{d}u + K\int_t^s e^{-\alpha(u-t)} I_{t,u}^{n_1+1,p}\mathrm{d}u. \end{aligned}$$

Using the convexity assumption $\lambda \geq 7K_V$, we can conclude the existence of $\lambda_0^{(n_1+1)} \in (0,\lambda)$ such that the term in front of the second integral remains negative. The result follows from Gronwall's inequality (as was done in the case $n=2$). □

5.5 Decay estimates for the Kolmogorov backward equation

We establish the following estimates for the derivatives of the solution to the Kolmogorov backward equation in terms of moments of the variation processes. This will further help us to apply the results developed in Section 5.4 to the weak error analysis in Section 5.6.

Lemma 5.5.1. *Let u be the solution to the Kolmogorov backward equation* (5.17) *with g as in Assumption 5.3.3 and let $T \geq t \geq 0, N \in \mathbb{N}$. Then there exists a constant $K > 0$ (independent of t, T, N), such that for any $j \in \{1,\dots,N\}$ and $\boldsymbol{x} \in \mathbb{R}^N$*

$$|\partial_{x_j} u(t,\boldsymbol{x})|^2 \leq \frac{K}{N^2}\mathbb{E}\Big[|X_{T,x_j}^{t,x_j,j,N}|^2\Big] + \frac{K}{N}\sum_{i=1,\ i\neq j}^{N} \mathbb{E}\left[|X_{T,x_j}^{t,x_i,i,N}|^2\right].$$

Proof. From the definition of u, using [73, Theorem 5.5 in Chapter 5] as in [81, Proposition B.3], we deduce

$$\begin{aligned} & |\partial_{x_j} u(t,\boldsymbol{x})|^2 \\ & = \Big|\mathbb{E}\Big[\sum_{i=1}^{N}\big(\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})\big)\cdot\big(X_{T,x_j}^{t,x_i,i,N}\big)\Big]\Big|^2 \\ & \leq 2\Big|\mathbb{E}\Big[|\partial_{x_j} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})|\ |X_{T,x_j}^{t,x_j,j,N}|\Big]\Big|^2 + 2\Big|\mathbb{E}\Big[\sum_{i=1,\ i\neq j}^{N}\big(\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})\big)\cdot\big(X_{T,x_j}^{t,x_i,i,N}\big)\Big]\Big|^2 \\ & \leq \frac{K}{N^2}\mathbb{E}\Big[|X_{T,x_j}^{t,x_j,j,N}|^2\Big] + KN\sum_{i=1,\ i\neq j}^{N}\mathbb{E}\left[\Big||\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})|\ |X_{T,x_j}^{t,x_i,i,N}|\Big|^2\right] \\ & \leq \frac{K}{N^2}\mathbb{E}\Big[|X_{T,x_j}^{t,x_j,j,N}|^2\Big] + \frac{K}{N}\sum_{i=1,\ i\neq j}^{N}\mathbb{E}\left[|X_{T,x_j}^{t,x_i,i,N}|^2\right], \end{aligned}$$

where we used Jensen's inequality and the growth of derivatives of g in Assumption 5.3.3. □

The next result, Lemma 5.5.2, generalizes Lemma 5.5.1 to derivatives of order $1 < n \leq 6$, whose proof is involved and relies on carefully applying Jensen's inequality. The next calculations aim to clarify that partitioning summations in particular ways before the application of Jensen's inequality can yield sharper upper bounds.

For each $\gamma = (\gamma_1,\dots,\gamma_n) \in \Pi_n^N$, let $x_{\gamma_1,\dots,\gamma_n}$ be a real number. Then applying Jensen's inequality directly we have

$$\Big|\sum_{\gamma\in\Pi_1^N} x_{\gamma_1}\Big|^2 = \Big|\sum_{i=1}^{N} x_i\Big|^2 = N^2\Big|\frac{1}{N}\sum_{i=1}^{N} x_i\Big|^2 \leq N\sum_{i=1}^{N}|x_i|^2,$$

$$\Big|\sum_{\gamma\in\Pi_n^N} x_{\gamma_1,\dots,\gamma_n}\Big|^2 = N^{2n}\Big|\frac{1}{N^n}\sum_{\gamma\in\Pi_n^N} x_{\gamma_1,\dots,\gamma_n}\Big|^2 \le \frac{N^{2n}}{N^n}\sum_{\gamma\in\Pi_n^N}|x_{\gamma_1,\dots,\gamma_n}|^2 = N^n\sum_{\gamma\in\Pi_n^N}|x_{\gamma_1,\dots,\gamma_n}|^2. \tag{5.99}$$

Consider now the specific two-dimensional example of $x_{\gamma_1,\gamma_2} = N^{1-\hat{\mathcal{O}}(\gamma)}$ (corresponding to a 2×2 matrix with diagonal entries 1 and otherwise $1/N$). Using (5.99) we have that

$$\Big|\sum_{\gamma\in\Pi_2^N} x_{\gamma_1,\gamma_2}\Big|^2 \le N^2\sum_{i,j=1}^N |x_{i,j}|^2 = N^2\sum_{i=1}^N|x_{i,i}|^2 + N^2\sum_{i,j=1,i\neq j}^N |x_{i,j}|^2 = N^3+N^2 \le 2N^3.$$

This estimate is too naive and can be improved, as we can instead consider

$$\begin{aligned}\Big|\sum_{\gamma\in\Pi_2^N} x_{\gamma_1,\gamma_2}\Big|^2 &\le 2\Big|\sum_{i=1}^N x_{i,i}\Big|^2 + 2\Big|\sum_{i,j=1,i\neq j}^N x_{i,j}\Big|^2 \le 2N\sum_{i=1}^N|x_{i,i}|^2 + 2N^2\sum_{i,j=1,i\neq j}^N|x_{i,j}|^2\\ &= 2N^2 + \frac{2N^3(N-1)}{N^2} \le 4N^2,\end{aligned}$$

which is indeed a sharper upper bound.

This argument can be extended to the general n-dimensional case; that is, take $x_{\gamma_1,\dots,\gamma_n} = N^{1-\hat{\mathcal{O}}(\gamma)}$ as the entries of a $\bigotimes_{i=1}^n \mathbb{R}^N$-valued n-tensor. Observe that

$$|\{\gamma\in\Pi_n^N : \hat{\mathcal{O}}(\gamma)=m\}| = \mathcal{O}(N^m), \tag{5.100}$$

a result which follows from a simple combinatorial argument: regardless of the length n of γ, we have m distinct values chosen out of the range $\{1,\dots,N\}$; that is $\mathcal{O}(N^m)$ possibilities. We have

$$\begin{aligned}\Big|\sum_{\gamma\in\Pi_n^N} x_{\gamma_1,\dots,\gamma_n}\Big|^2 &= \Big|\sum_{m=1}^n\sum_{\substack{\gamma\in\Pi_n^N,\\ \hat{\mathcal{O}}(\gamma)=m}} x_{\gamma_1,\dots,\gamma_n}\Big|^2 \le K\sum_{m=1}^n\Big|\sum_{\substack{\gamma\in\Pi_n^N,\\ \hat{\mathcal{O}}(\gamma)=m}} x_{\gamma_1,\dots,\gamma_n}\Big|^2\\ &\le K\sum_{m=1}^n N^m\sum_{\substack{\gamma\in\Pi_n^N,\\ \hat{\mathcal{O}}(\gamma)=m}}|x_{\gamma_1,\dots,\gamma_n}|^2 \le K\sum_{m=1}^n N^{2m}N^{2-2m} \le KN^2,\end{aligned}$$

with K a constant, changing across the inequalities, independent of N but dependent on n where $n\le 6$. This is also a sharper upper bound than applying Jensen's inequality directly. We now state the generalization of Lemma 5.5.1 for $1\le n\le 6$.

Lemma 5.5.2. *Let u satisfy the Kolmogorov backward equation* (5.17) *with g as in Assumption 5.3.3 and let $T\ge t\ge 0, N\in\mathbb{N}$. Then there exists a constant $K>0$ (independent of t,T,N), such that for any $n\in\mathbb{N}, 1\le n\le 6$, $\gamma\in\Pi_n^N$, and $\boldsymbol{x}\in\mathbb{R}^N$*

$$\begin{aligned}&|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^2\\ &\le K\sum_{m=0}^n \sum_{\substack{\ell\in\bigcup_{k=1}^n\Pi_k^N,\\ \hat{\mathcal{O}}(\ell\bigcup\gamma)=\hat{\mathcal{O}}(\gamma)+m}} N^{m-2\hat{\mathcal{O}}(\ell)} \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|}\in\bigcup_{k=1}^n\Pi_k^N,\\ \bigcup_{i=1}^{|\ell|}\alpha_i\simeq\gamma}} \mathbb{E}\Big[\prod_{i=1}^{|\ell|}\Big|X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}}\Big|^2\Big],\end{aligned} \tag{5.101}$$

where $\alpha_i = (\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|})$ and $\alpha_{i,j}\in\{1,\dots,N\}$ for $j\in\{1,\dots,|\alpha_i|\}$. Further, we have

$$|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^4$$

$$\leq K\sum_{m=0}^{n}\sum_{\substack{\ell\in\bigcup_{k=1}^{n}\Pi_k^N,\\ \hat{\mathcal{O}}(\ell\bigcup\gamma)=\hat{\mathcal{O}}(\gamma)+m}} N^{3m-4\hat{\mathcal{O}}(\ell)}\sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|}\in\bigcup_{k=1}^{n}\Pi_k^N,\\ \bigcup_{i=1}^{|\ell|}\alpha_i\simeq\gamma}}\mathbb{E}\Big[\prod_{i=1}^{|\ell|}\Big|X_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}}^{t,x_{\ell_i},\ell_i,N}\Big|^4\Big]. \tag{5.102}$$

Proof. Note that in the following proof, the positive constant K is independent of t,T,N and may change line by line. Before establishing the main results (*Part 2* and *Part 3*), we present (*Part 1*) a study for the second derivatives of u (i.e., for the case $n=2$) to demonstrate the approach used and its nuances in regards to estimate (5.101) of the lemma.

Part 1: Comparing the direct calculation with (5.101) *for the example case* $n=2$. For all $j,k\in\{1,\dots,N\}$, $j\neq k$, using [73, Theorem 5.5 in Chapter 5], we have

$$\begin{aligned}
&|\partial^2_{x_j,x_k}u(t,\boldsymbol{x})|^2\\
&=\Big|\mathbb{E}\Big[\sum_{i=1}^{N}\partial_{x_i}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j,x_k}^{t,x_i,i,N}\Big]+\mathbb{E}\Big[\sum_{i=1}^{N}\sum_{i'=1}^{N}\partial^2_{x_i,x_{i'}}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_i,i,N}X_{T,x_k}^{t,x_{i'},i',N}\Big]\Big|^2\\
&\leq K\Bigg(\Big|\mathbb{E}\Big[\partial_{x_j}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j,x_k}^{t,x_j,j,N}+\partial_{x_k}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j,x_k}^{t,x_k,k,N}\Big]\Big|^2 &(5.103)\\
&\quad+\Big|\mathbb{E}\Big[\sum_{i=1,\ i\neq j\neq k}^{N}\partial_{x_i}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j,x_k}^{t,x_i,i,N}\Big]\Big|^2 &(5.104)\\
&\quad+\Big|\mathbb{E}\Big[\partial^2_{x_j,x_j}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_j,j,N}X_{T,x_k}^{t,x_j,j,N}+\partial^2_{x_k,x_k}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_k,k,N}X_{T,x_k}^{t,x_k,k,N}\Big]\Big|^2 &(5.105)\\
&\quad+\Big|\mathbb{E}\Big[\partial^2_{x_j,x_k}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_j,j,N}X_{T,x_k}^{t,x_k,k,N}+\partial^2_{x_k,x_j}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_k,k,N}X_{T,x_k}^{t,x_j,j,N}\Big]\Big|^2 &(5.106)\\
&\quad+\Big|\mathbb{E}\Big[\sum_{i=1,\ i\neq j\neq k}^{N}\partial^2_{x_i,x_i}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_i,i,N}X_{T,x_k}^{t,x_i,i,N}\Big]\Big|^2 &(5.107)\\
&\quad+\Big|\mathbb{E}\Big[\sum_{i=1,\ i\neq j\neq k}^{N}\Big(\partial^2_{x_j,x_i}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_j,j,N}X_{T,x_k}^{t,x_i,i,N}+\partial^2_{x_i,x_j}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_i,i,N}X_{T,x_k}^{t,x_j,j,N}\\
&\qquad+\partial^2_{x_k,x_i}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_k,k,N}X_{T,x_k}^{t,x_i,i,N}+\partial^2_{x_i,x_k}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_i,i,N}X_{T,x_k}^{t,x_k,k,N}\Big)\Big]\Big|^2 &(5.108)\\
&\quad+\Big|\mathbb{E}\Big[\sum_{i,i'=1,\ j\neq i\neq i'\neq k}^{N}\partial^2_{x_i,x_{i'}}g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})X_{T,x_j}^{t,x_i,i,N}X_{T,x_k}^{t,x_{i'},i',N}\Big]\Big|^2\Bigg), &(5.109)
\end{aligned}$$

where the inequality follows simply by separating the summation terms by number of distinct derivatives of g matching also to the order of decay in N, and then applying Jensen's inequality (but still leaving the square outside the expectations).

Under Assumption 5.3.3 on the derivatives of g and observing that (5.104), (5.107) and (5.108) have $\mathcal{O}(N)$ terms and that (5.109) has $\mathcal{O}(N^2)$ terms in the summand, we apply Jensen's inequality once more to get

$$\begin{aligned}
(5.103)+(5.105)+(5.109)&\leq\frac{K}{N^2}\mathbb{E}\Big[|X_{T,x_j,x_k}^{t,x_j,j,N}|^2+|X_{T,x_j,x_k}^{t,x_k,k,N}|^2+|X_{T,x_j}^{t,x_j,j,N}X_{T,x_k}^{t,x_j,j,N}|^2\Big]\\
&\quad+\frac{K}{N^2}\mathbb{E}\Big[|X_{T,x_j}^{t,x_k,k,N}X_{T,x_k}^{t,x_k,k,N}|^2\Big]+\sum_{i,i'=1,\ j\neq i\neq i'\neq k}^{N}\Big[|X_{T,x_j}^{t,x_i,i,N}X_{T,x_k}^{t,x_{i'},i',N}|^2\Big]\\
(5.104)+(5.107)&\leq\frac{K}{N}\sum_{i=1,\ i\neq j\neq k}^{N}\mathbb{E}\Big[|X_{T,x_j,x_k}^{t,x_i,i,N}|^2+|X_{T,x_j}^{t,x_i,i,N}X_{T,x_k}^{t,x_i,i,N}|^2\Big]\\
(5.106)&\leq\frac{K}{N^4}\mathbb{E}\Big[|X_{T,x_j}^{t,x_j,j,N}X_{T,x_k}^{t,x_k,k,N}|^2+|X_{T,x_j}^{t,x_k,k,N}X_{T,x_k}^{t,x_j,j,N}|^2\Big]
\end{aligned}$$

$$(5.108) \leq \frac{K}{N^3} \sum_{i=1,\ i\neq j\neq k}^{N} \mathbb{E}\Big[|X^{t,x_j,j,N}_{T,x_j} X^{t,x_i,i,N}_{T,x_k}|^2 + |X^{t,x_i,i,N}_{T,x_j} X^{t,x_j,j,N}_{T,x_k}|^2 + |X^{t,x_k,k,N}_{T,x_j} X^{t,x_i,i,N}_{T,x_k}|^2 + |X^{t,x_i,i,N}_{T,x_j} X^{t,x_k,k,N}_{T,x_k}|^2\Big].$$

We now demonstrate how these estimates (5.103)–(5.109), when combined together, can be upper bounded by the form given in (5.101) for $n = 2$, $\gamma = (j,k)$, $j \neq k$:

$$\begin{aligned}
&|\partial^2_{x_j,x_k} u(t,\boldsymbol{x})|^2 \\
&\leq K \sum_{m=0}^{2} \sum_{\substack{\ell\in\bigcup_{k=1}^2 \Pi_k^N,\\ \hat{\mathcal{O}}(\ell\bigcup\gamma)=\hat{\mathcal{O}}(\gamma)+m}} N^{m-2\hat{\mathcal{O}}(\ell)} \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|}\in\bigcup_{k=1}^2 \Pi_k^N,\\ \bigcup_{i=1}^{|\ell|}\alpha_i\simeq\gamma}} \mathbb{E}\Big[\prod_{i=1}^{|\ell|} \Big|X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}}\Big|^2\Big] \\
&= K\Bigg(\frac{1}{N^2}\mathbb{E}\Big[|X^{t,x_j,j,N}_{T,x_j,x_k}|^2 + |X^{t,x_k,k,N}_{T,x_j,x_k}|^2\Big] \\
&\quad + \frac{1}{N^2}\mathbb{E}\Big[|X^{t,x_j,j,N}_{T,x_j}|^2\ |X^{t,x_j,j,N}_{T,x_k}|^2 + |X^{t,x_j,j,N}_{T,x_k}|^2\ |X^{t,x_j,j,N}_{T,x_j}|^2 \\
&\quad + |X^{t,x_k,k,N}_{T,x_j}|^2\ |X^{t,x_k,k,N}_{T,x_k}|^2 + |X^{t,x_k,k,N}_{T,x_k}|^2\ |X^{t,x_k,k,N}_{T,x_j}|^2\Big] + \frac{1}{N^4}\mathbb{E}\Big[|X^{t,x_j,j,N}_{T,x_j}|^2\ |X^{t,x_k,k,N}_{T,x_k}|^2 \\
&\quad + |X^{t,x_j,j,N}_{T,x_k}|^2\ |X^{t,x_k,k,N}_{T,x_j}|^2 + |X^{t,x_k,k,N}_{T,x_j}|^2\ |X^{t,x_j,k,N}_{T,x_k}|^2 + |X^{t,x_k,k,N}_{T,x_k}|^2\ |X^{t,x_j,j,N}_{T,x_j}|^2\Big] \\
&\quad + \frac{1}{N}\sum_{i=1,i\neq j\neq k}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j,x_k}|^2 + |X^{t,x_i,i,N}_{T,x_k,x_j}|^2\Big] + \frac{1}{N^3}\sum_{i=1,i\neq j\neq k}^{N} \mathbb{E}\Big[|X^{t,x_j,j,N}_{T,x_j}|^2\ |X^{t,x_i,i,N}_{T,x_k}|^2 \\
&\quad + |X^{t,x_j,j,N}_{T,x_k}|^2\ |X^{t,x_i,i,N}_{T,x_j}|^2 + |X^{t,x_i,i,N}_{T,x_j}|^2\ |X^{t,x_j,j,N}_{T,x_k}|^2 \\
&\quad + |X^{t,x_i,i,N}_{T,x_k}|^2\ |X^{t,x_j,j,N}_{T,x_j}|^2 + |X^{t,x_k,k,N}_{T,x_j}|^2\ |X^{t,x_i,i,N}_{T,x_k}|^2 \\
&\quad + |X^{t,x_k,k,N}_{T,x_k}|^2\ |X^{t,x_i,i,N}_{T,x_j}|^2 + |X^{t,x_i,i,N}_{T,x_j}|^2\ |X^{t,x_k,k,N}_{T,x_k}|^2 + |X^{t,x_i,i,N}_{T,x_k}|^2\ |X^{t,x_k,k,N}_{T,x_j}|^2\Big] \\
&\quad + \frac{1}{N}\sum_{i=1,i\neq j\neq k}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j}|^2\ |X^{t,x_i,i,N}_{T,x_k}|^2 + |X^{t,x_i,i,N}_{T,x_k}|^2\ |X^{t,x_i,i,N}_{T,x_j}|^2\Big] \\
&\quad + \frac{1}{N^2}\sum_{i,i'=1,\ j\neq i\neq i'\neq k}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j}|^2\ |X^{t,x_{i'},i',N}_{T,x_k}|^2 + |X^{t,x_{i'},i',N}_{T,x_j}|^2\ |X^{t,x_i,i,N}_{T,x_k}|^2\Big]\Bigg). \qquad (5.110)
\end{aligned}$$

Comparing to the results of (5.103)–(5.109), we can see that (5.110) contains more terms. For example, consider the term

$$\frac{1}{N}\sum_{i=1,i\neq j\neq k}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j,x_k}|^2 + |X^{t,x_i,i,N}_{T,x_k,x_j}|^2\Big].$$

For the case $m = 1, \ell \in \Pi_1^N$, we not only take summation over $|X^{t,x_i,i,N}_{T,x_j,x_k}|^2$, but also consider $|X^{t,x_i,i,N}_{T,x_k,x_j}|^2$, so that the sum of the terms (5.103)–(5.109) is bounded above by (5.110), verifying our result in the case $n = 2$, $\gamma = (j,k)$, $j \neq k$.

Part 2. The bound (5.101). From the above estimates, one can see that the idea is to essentially partition the sums based on the relationship between indices i, i', j, k, to keep consistent orders of N. Having this separation trick in mind, for any $\gamma \in \Pi_n^N$, $1 \leq n \leq 6$, we consider (recall the notation for $\hat{\mathcal{O}}(\cdot)$ and $|\cdot|$ in Definition 5.3.1):

$$|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^2$$

$$
\begin{aligned}
&= \left| \mathbb{E}\left[\sum_{\substack{\alpha,\beta \in \bigcup_{k=0}^{n-1} \Pi_k^N, \\ \gamma \backslash (\gamma_1) \in \alpha \sqcup \beta}} \sum_{i=1}^{N} \Big(\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \Big)_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}} \big(X_{T,x_{\gamma_1}}^{t,x_i,i,N} \big)_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}} \right] \right|^2 \\
&\leq \left| \mathbb{E}\left[\sum_{m=1}^{n} \sum_{\ell \in \Pi_m^N} \left(\left| \partial^m_{x_{\ell_1},\dots,x_{\ell_m}} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \right| \sum_{\substack{\alpha_1,\dots,\alpha_m \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \bigcup_{i=1}^m \alpha_i \simeq \gamma}} \prod_{i=1}^{m} \left| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \right| \right) \right] \right|^2 \\
&= \left| \mathbb{E}\left[\sum_{m=1}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \hat{\mathcal{O}}(\ell) = m}} \left(\left| \partial^{|\ell|}_{x_{\ell_1},\dots,x_{\ell_{|\ell|}}} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \right| \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \left| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \right| \right) \right] \right|^2 \qquad (5.111) \\
&= \left| \mathbb{E}\left[\sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m}} \left(\left| \partial^{|\ell|}_{x_{\ell_1},\dots,x_{\ell_{|\ell|}}} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \right| \right.\right.\right. \\
&\qquad \left.\left.\left. \cdot \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \left| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \right| \right) \right] \right|^2 \qquad (5.112) \\
&\leq K \sum_{m=0}^{n} N^{2m}\, \mathbb{E}\left[\left| \frac{1}{N^m} \sum_{\substack{\ell \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m}} \left(\left| \partial^{|\ell|}_{x_{\ell_1},\dots,x_{\ell_{|\ell|}}} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \right| \right.\right.\right. \\
&\qquad \left.\left.\left. \cdot \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \left| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \right| \right) \right|^2 \right] \\
&\leq K \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m}} N^m\, \mathbb{E}\left[\left| \partial^{|\ell|}_{x_{\ell_1},\dots,x_{\ell_{|\ell|}}} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \right|^2 \right. \\
&\qquad \left. \cdot \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \left| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \right|^2 \right] \qquad (5.113) \\
&\leq K \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m}} N^{m - 2\hat{\mathcal{O}}(\ell)}\, \mathbb{E}\left[\sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^{n} \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \left| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \right|^2 \right],
\end{aligned}
$$

where in (5.111) and (5.112), we regroup the summation over all $\ell \in \bigcup_{k=1}^{n} \Pi_k^N$ based on the magnitude of $\hat{\mathcal{O}}(\ell)$ and $\hat{\mathcal{O}}(\ell \bigcup \gamma)$. In (5.113), we apply Jensen's inequality to the second summation where the set $\{\ell \in \bigcup_{k=1}^{n} \Pi_k^N : \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m\}$ has $\mathcal{O}(N^m)$ terms (ℓ has m degrees of freedom), thus we end up with a factor of N^m after calculation. For the last line, we used Assumption 5.3.3: $\left| \partial^{|\ell|}_{x_{\ell_1},\dots,x_{\ell_{|\ell|}}} g \right|_\infty = \mathcal{O}(N^{-\hat{\mathcal{O}}(\ell)})$.

Part 3: The bound (5.102). One proves (5.102) using the same arguments as in *Part 2*. We have

$$
|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^4
$$

$$\leq K \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^n \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma)+m}} N^{3m}\, \mathbb{E}\Bigg[\Big| \partial^{|\ell|}_{x_{\ell_1},\dots,x_{\ell_{|\ell|}}} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \Big|^4 \cdot \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^n \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \Big| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \Big|^4 \Bigg]$$

$$\leq K \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^n \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma)+m}} N^{3m-4\hat{\mathcal{O}}(\ell)}\, \mathbb{E}\Bigg[\sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^n \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \Big| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \Big|^4 \Bigg].$$

□

Lemma 5.5.3. *Let u satisfy the Kolmogorov backward equation (5.17) with g as in Assumption 5.3.3, let $T \geq t \geq 0, N \in \mathbb{N}$ and assume that the starting points x_i are $\mathcal{F}_t$-measurable random variables in $L^2(\Omega, \mathbb{R})$ sampled from the same distribution for all $i \in \{1, \dots, N\}$. Then there exist positive constants K, $\lambda_0 \in (0, \lambda)$ (both are independent of t, T, N) such that for any $1 \leq n \leq 6$, $\gamma \in \Pi_n^N$*

$$\mathbb{E}\Big[\big| \partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x}) \big|^2 \Big] \leq K e^{-\lambda_0 (T-t)} N^{-2\hat{\mathcal{O}}(\gamma)}, \tag{5.114}$$

$$\mathbb{E}\Big[\big| \partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x}) \big|^4 \Big] \leq K e^{-\lambda_0 (T-t)} N^{-4\hat{\mathcal{O}}(\gamma)}. \tag{5.115}$$

Proof. Note that in the following proof, the positive constants K, λ_0 are independent of t, T, N and may change line by line. Recall the results and notations in Lemma 5.5.2, after taking the expectation, we have

$$\mathbb{E}\Big[\big| \partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x}) \big|^2 \Big]$$

$$\leq K \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^n \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma)+m}} N^{m-2\hat{\mathcal{O}}(\ell)} \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^n \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \mathbb{E}\Big[\prod_{i=1}^{|\ell|} \Big| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \Big|^2 \Big]$$

$$\leq K \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^n \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma)+m}} N^{m-2\hat{\mathcal{O}}(\ell)} \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^n \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \Big(\mathbb{E}\Big[\Big| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \Big|^{2|\ell|} \Big] \Big)^{1/|\ell|},$$

where we employed Hölder's inequality. We have the following estimate for the variation processes in the above product:

$$\begin{aligned}
\Big(\mathbb{E}\Big[\Big| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \Big|^{2|\ell|} \Big] \Big)^{1/|\ell|} &\leq \Bigg(\frac{1}{N^{\hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i)}} \sum_{\substack{\beta \in \Pi^N_{|\alpha_i|+1}, \\ \hat{\mathcal{O}}(\beta) = \hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i)}} \mathbb{E}\Big[\Big| X^{t,x_{\beta_1},\beta_1,N}_{T,x_{\beta_2},\dots,x_{\beta_{|\beta|}}} \Big|^{2|\ell|} \Big] \Bigg)^{1/|\ell|} \\
&\leq \Big(\frac{K}{N^{2|\ell|(\hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i)-1)}} e^{-\lambda_0 2|\ell|(T-t)} \Big)^{1/|\ell|} \\
&\leq K e^{-2\lambda_0 (T-t)} N^{-2(\hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i)-1)},
\end{aligned}$$

where we used the second and first part of Lemma 5.4.7 with $p = 2|\ell|$, $m = \hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i)$ to obtain the first two inequalities. We now relate the orders $(\hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i) - 1)$ from the previous estimate and $\hat{\mathcal{O}}(\gamma)$ appearing in (5.114), by first showing that $\hat{\mathcal{O}}(\gamma \setminus \ell) \leq \sum_{i=1}^{|\ell|} \Big(\hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i) - 1 \Big)$.

Concretely, the constraint in the second summation, $\bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma$, implies

$$\begin{aligned}
\sum_{i=1}^{|\ell|} \Big(\hat{\mathcal{O}}((\ell_i)\bigcup \alpha_i) - 1\Big) &= \sum_{i=1}^{|\ell|} \Big(\hat{\mathcal{O}}\Big((\ell_i)\bigcup \Big(\alpha_i \bigcap (\gamma \setminus \ell)\Big) \bigcup \Big(\alpha_i \setminus (\gamma \setminus \ell)\Big)\Big) - 1\Big) \\
&= \sum_{i=1}^{|\ell|} \Big(\hat{\mathcal{O}}\Big(\alpha_i \bigcap (\gamma \setminus \ell)\Big) + \hat{\mathcal{O}}\Big((\ell_i)\bigcup \Big(\alpha_i \setminus (\gamma \setminus \ell)\Big)\Big) - 1\Big) \\
&\geq \sum_{i=1}^{|\ell|} \hat{\mathcal{O}}\Big(\alpha_i \bigcap (\gamma \setminus \ell)\Big) \geq \hat{\mathcal{O}}\Big((\bigcup_{i=1}^{|\ell|} \alpha_i) \bigcap (\gamma \setminus \ell)\Big) = \hat{\mathcal{O}}(\gamma \setminus \ell).
\end{aligned}$$

Then using the constraint $\hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m$ of the first summation we infer

$$\sum_{i=1}^{|\ell|} \Big(\hat{\mathcal{O}}((\ell_i)\bigcup \alpha_i) - 1\Big) \geq \hat{\mathcal{O}}(\gamma \setminus \ell) = \hat{\mathcal{O}}(\ell \bigcup \gamma) - \hat{\mathcal{O}}(\ell) = \hat{\mathcal{O}}(\gamma) + m - \hat{\mathcal{O}}(\ell).$$

Hence,

$$\begin{aligned}
&\mathbb{E}\Big[|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^2\Big] \\
&\quad \leq K e^{-2\lambda_0(T-t)} \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^n \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m}} N^{m - 2\hat{\mathcal{O}}(\ell)} \; N^{-2 \sum_{i=1}^{|\ell|} \left(\hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i) - 1\right)} \\
&\quad \leq K e^{-2\lambda_0(T-t)} \sum_{m=0}^{n} N^{2m - 2\hat{\mathcal{O}}(\ell) - 2\left(\hat{\mathcal{O}}(\gamma) - \hat{\mathcal{O}}(\ell) + m\right)} \leq K e^{-2\lambda_0(T-t)} N^{-2\hat{\mathcal{O}}(\gamma)},
\end{aligned}$$

which yields (5.114), as sought. Similar calculations deliver (5.115). That is, we have

$$\begin{aligned}
&\mathbb{E}\Big[|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^4\Big] \\
&\quad \leq K \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^n \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m}} N^{3m - 4\hat{\mathcal{O}}(\ell)} \\
&\qquad\qquad \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^n \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \prod_{i=1}^{|\ell|} \left(\mathbb{E}\left[\left|X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}}\right|^{4|\ell|}\right]\right)^{1/|\ell|} \\
&\quad \leq K e^{-2\lambda_0(T-t)} \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^n \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma) + m}} N^{3m - 4\hat{\mathcal{O}}(\ell)} \; N^{-4 \sum_{i=1}^{|\ell|} \left(\hat{\mathcal{O}}((\ell_i) \bigcup \alpha_i) - 1\right)} \\
&\quad \leq K e^{-2\lambda_0(T-t)} \sum_{m=0}^{n} N^{4m - 4\hat{\mathcal{O}}(\ell) - 4\left(\hat{\mathcal{O}}(\gamma) - \hat{\mathcal{O}}(\ell) + m\right)} \leq K e^{-2\lambda_0(T-t)} N^{-4\hat{\mathcal{O}}(\gamma)}.
\end{aligned}$$

□

5.6 Weak error expansion and its analysis

For the convenience of the reader, we recall Lemma 5.3.7 which provides an expansion for the global weak error (5.14) under Assumption 5.3.3, for the processes defined in (5.11) and (5.13) as follows:

$$\mathbb{E}\Big[g(\boldsymbol{X}_T^N)\Big] - \mathbb{E}\Big[g(\boldsymbol{X}_T^{N,h})\Big] = h^2\mathbb{E}\left[\sum_{m=0}^{M-1} L(t_m, \boldsymbol{X}_{t_m}^{N,h})\right] + \mathbb{E}\left[\sum_{m=0}^{M-1} R(t_m, \boldsymbol{X}_{t_m}^{N,h})\right], \tag{5.116}$$

where the map $L : \mathbb{R}_+ \times \mathbb{R}^N \to \mathbb{R}$ is defined via the maps u and $(B_i)_{i\in\{1,\dots,N\}}$:

$$\begin{aligned} L(t,\boldsymbol{x}) = \frac{1}{2}\Big[\sum_{i,j=1}^N B_j(\boldsymbol{x})\partial_{x_j}B_i(\boldsymbol{x})\partial_{x_i}u(t,\boldsymbol{x}) + \frac{\sigma^2}{2}\sum_{i,j=1}^N \partial_{x_j}B_i(\boldsymbol{x})\partial^2_{x_i,x_j}u(t,\boldsymbol{x}) \\ + \frac{\sigma^2}{2}\sum_{i,j=1}^N \partial^2_{x_j,x_j}B_i(\boldsymbol{x})\partial_{x_i}u(t,\boldsymbol{x})\Big]. \end{aligned} \tag{5.117}$$

The remainder term $R(\cdot,\cdot)$ will later be written as a linear combination of 8 remainder terms, which we will analyze in Section 5.6.2.

5.6.1 Estimates for the leading error term L

We consider the first term in (5.116) expressed as a telescoping sum:

$$h^2\mathbb{E}\Big[\sum_{m=0}^{M-1} L(t_m, \boldsymbol{X}_{t_m}^{N,h})\Big] = h\mathbb{E}\Big[\int_0^T L(s,\boldsymbol{X}_s^N)\mathrm{d}s\Big] \tag{5.118}$$

$$+ h\sum_{m=0}^{M-1}\mathbb{E}\Big[\int_{t_m}^{t_{m+1}} \Big(L(t_m,\boldsymbol{X}_{t_m}^{N,h}) - L(s,\boldsymbol{X}_s^N)\Big)\mathrm{d}s\Big], \tag{5.119}$$

and derive the following result.

Lemma 5.6.1. *Let Assumption 5.3.3 hold and let $\xi \in L^4(\Omega,\mathbb{R})$. Let L be defined in (5.117). Then there exists a positive constant $\lambda_0 \in (0,\lambda)$ such that*

$$h^2\mathbb{E}\Big[\sum_{m=0}^{M-1} L(t_m, \boldsymbol{X}_{t_m}^{N,h})\Big] \le Kh^{3/2} + Khe^{-\lambda_0 T}. \tag{5.120}$$

Proof. Note that in the following proof, the positive constants K, λ_0 are independent of t, T, N and may change line by line. The proof is carried out by analysing (5.118) and (5.119) separately to reach (5.120).

Part 1: Estimating (5.118). Let $(X_*^{1,N},\dots,X_*^{N,N}) = \boldsymbol{X}_*^N \sim \mu^{N,*}$, where $\mu^{N,*}$ is the stationary distribution of the IPS (viewed as a $\mathbb{R}^N$-valued SDE). Using integration by parts, we have

$$\mathbb{E}\Big[L(t,\boldsymbol{X}_*^N)\Big] = \int_{\mathbb{R}^N} L(t,\boldsymbol{x})\mu^{N,*}(\mathrm{d}\boldsymbol{x}) = 0.$$

Hence, we may write

$$\begin{aligned} \mathbb{E}\Big[\int_0^T L(s,\boldsymbol{X}_s^N)\mathrm{d}s\Big] &= \int_0^T \mathbb{E}\Big[L(s,\boldsymbol{X}_s^N) - L(s,\boldsymbol{X}_*^N)\Big]\mathrm{d}s \\ &= \int_0^T\int_0^1 \mathbb{E}\Big[\big\langle \partial_{\boldsymbol{x}}L(s,\rho\boldsymbol{X}_s^N + (1-\rho)\boldsymbol{X}_*^N,\ \boldsymbol{X}_s^N - \boldsymbol{X}_*^N\big\rangle\Big]\mathrm{d}\rho\ \mathrm{d}s. \end{aligned}$$

Let $\boldsymbol{X}_{s,\rho}^N := \rho\boldsymbol{X}_s^N + (1-\rho)\boldsymbol{X}_*^N$. Using the chain rule, we deduce the following:

$$\mathbb{E}\Big[\Big\langle \partial_{\boldsymbol{x}}L(s,\boldsymbol{X}_{s,\rho}^N),\ \boldsymbol{X}_s^N - \boldsymbol{X}_*^N\Big\rangle\Big] = \frac{1}{2}\mathbb{E}\Big[\sum_{i,j,k=1}^N \Big(\partial_{x_k}B_j(\boldsymbol{X}_{s,\rho}^N)\partial_{x_j}B_i(\boldsymbol{X}_{s,\rho}^N)\partial_{x_i}u(s,\boldsymbol{X}_{s,\rho}^N)$$

$$
\begin{aligned}
&+ B_j(\boldsymbol{X}^N_{s,\rho})\partial^2_{x_j,x_k} B_i(\boldsymbol{X}^N_{s,\rho})\partial_{x_i} u(s,\boldsymbol{X}^N_{s,\rho}) + B_j(\boldsymbol{X}^N_{s,\rho})\partial_{x_j} B_i(\boldsymbol{X}^N_{s,\rho})\partial^2_{x_i,x_k} u(s,\boldsymbol{X}^N_{s,\rho})\Big) \\
&\qquad\qquad \cdot \Big(X^{k,N}_s - X^{k,N}_*\Big) \\
&+ \frac{\sigma^2}{2}\sum_{i,j,k=1}^N \Big(\partial^2_{x_j,x_k} B_i(\boldsymbol{X}^N_{s,\rho})\partial^2_{x_i,x_j} u(s,\boldsymbol{X}^N_{s,\rho}) + \partial_{x_j} B_i(\boldsymbol{X}^N_{s,\rho})\partial^3_{x_i,x_j,x_k} u(s,\boldsymbol{X}^N_{s,\rho})\Big) \\
&\qquad\qquad \cdot \Big(X^{k,N}_s - X^{k,N}_*\Big) \\
&+ \frac{\sigma^2}{2}\sum_{i,j,k=1}^N \Big(\partial^3_{x_j,x_j,x_k} B_i(\boldsymbol{X}^N_{s,\rho})\partial_{x_i} u(s,\boldsymbol{X}^N_{s,\rho}) + \partial^2_{x_j,x_j} B_i(\boldsymbol{X}^N_{s,\rho})\partial^2_{x_i,x_k} u(s,\boldsymbol{X}^N_{s,\rho})\Big) \\
&\qquad\qquad \cdot \Big(X^{k,N}_s - X^{k,N}_*\Big)\Big] \\
\le{}& K\sum_{i,j,k=1}^N \sqrt{\mathbb{E}\Big[\big|X^{k,N}_s - X^{k,N}_*\big|^2\Big]} \cdot \Big\{ \sqrt{\mathbb{E}\Big[\big|\partial_{x_k} B_j(\boldsymbol{X}^N_{s,\rho})\partial_{x_j} B_i(\boldsymbol{X}^N_{s,\rho})\partial_{x_i} u(s,\boldsymbol{X}^N_{s,\rho})\big|^2\Big]} \\
&+ \sqrt{\mathbb{E}\Big[\big|B_j(\boldsymbol{X}^N_{s,\rho})\partial^2_{x_j,x_k} B_i(\boldsymbol{X}^N_{s,\rho})\partial_{x_i} u(s,\boldsymbol{X}^N_{s,\rho})\big|^2\Big]} \\
&+ \sqrt{\mathbb{E}\Big[\big|B_j(\boldsymbol{X}^N_{s,\rho})\partial_{x_j} B_i(\boldsymbol{X}^N_{s,\rho})\partial^2_{x_i,x_k} u(s,\boldsymbol{X}^N_{s,\rho})\big|^2\Big]} \\
&+ \sqrt{\mathbb{E}\Big[\big|\partial^2_{x_j,x_k} B_i(\boldsymbol{X}^N_{s,\rho})\partial^2_{x_i,x_j} u(s,\boldsymbol{X}^N_{s,\rho})\big|^2\Big]} + \sqrt{\mathbb{E}\Big[\big|\partial_{x_j} B_i(\boldsymbol{X}^N_{s,\rho})\partial^3_{x_i,x_j,x_k} u(s,\boldsymbol{X}^N_{s,\rho})\big|^2\Big]} \\
&+ \sqrt{\mathbb{E}\Big[\big|\partial^3_{x_j,x_j,x_k} B_i(\boldsymbol{X}^N_{s,\rho})\partial_{x_i} u(s,\boldsymbol{X}^N_{s,\rho})\big|^2\Big]} + \sqrt{\mathbb{E}\Big[\big|\partial^2_{x_j,x_j} B_i(\boldsymbol{X}^N_{s,\rho})\partial^2_{x_i,x_k} u(s,\boldsymbol{X}^N_{s,\rho})\big|^2\Big]}\ \Big\},
\end{aligned} \tag{5.121}
$$

where we used the Cauchy–Schwarz inequality. We now work through (5.121). As for the L^2-distance to the invariant distribution (the first expectation in the sum), note that (6.23) in Lemma 6.4.4 implies that for any $s \ge 0$ we have $\mathbb{E}[|X^{k,N}_s - X^{k,N}_*|^2] \le Ke^{-2\lambda s}$. The approach to deal with the remaining seven terms is more or less identical. We inject the estimate (5.114) of Lemma 5.5.3 for the derivatives of u and the bounds for the derivatives of B established in (5.97). The inequality below preserves the exact ordering of (5.121), and we highlight that obtaining (5.122) requires the additional use of the linear growth of B_j, the Cauchy–Schwarz inequality, the L^4-estimates of $\boldsymbol{X}^N$ (in Proposition 5.2.2) and the L^4-estimate (5.115) for the derivatives of u (recalling that $\xi \in L^4(\Omega,\mathbb{R})$); for some positive constant $\lambda_0 \in (0,\lambda)$ chosen small enough we have

$$
\begin{aligned}
&\mathbb{E}\Big[\Big\langle \partial_{\boldsymbol{x}} L(s,\boldsymbol{X}^N_{s,\rho}),\ \boldsymbol{X}^N_s - \boldsymbol{X}^N_*\Big\rangle\Big] \\
&\quad \le K\sqrt{e^{-2\lambda s}} \cdot \sqrt{e^{-\lambda_0(T-s)}} \sum_{\gamma\in\Pi^N_3} \Big(\frac{1}{N^{(\hat{\mathcal{O}}((\gamma_2,\gamma_3))-1)+(\hat{\mathcal{O}}((\gamma_1,\gamma_2))-1)+1}} \\
&\qquad + \frac{1}{N^{(\hat{\mathcal{O}}(\gamma)-1)+1}} + \frac{1}{N^{(\hat{\mathcal{O}}((\gamma_1,\gamma_2))-1)+\hat{\mathcal{O}}((\gamma_1,\gamma_3))}} \\
&\qquad + \frac{1}{N^{(\hat{\mathcal{O}}(\gamma)-1)+\hat{\mathcal{O}}((\gamma_1,\gamma_2))}} + \frac{1}{N^{(\hat{\mathcal{O}}((\gamma_1,\gamma_2))-1)+\hat{\mathcal{O}}(\gamma)}} \\
&\qquad + \frac{1}{N^{(\hat{\mathcal{O}}(\gamma)-1)+1}} + \frac{1}{N^{(\hat{\mathcal{O}}((\gamma_1,\gamma_2))-1)+\hat{\mathcal{O}}((\gamma_1,\gamma_3))}} \Big)
\end{aligned} \tag{5.122}
$$

$$
\le K\sqrt{e^{-2\lambda s}e^{-\lambda_0(T-s)}} \sum_{\gamma\in\Pi^N_3} \frac{1}{N^{\hat{\mathcal{O}}(\gamma)}} \tag{5.123}
$$

$$
\le K\sqrt{e^{-2\lambda s}e^{-\lambda_0(T-s)}}, \tag{5.124}
$$

where the inequality in (5.123) follows from the fact that $\hat{\mathcal{O}}((\gamma_1,\gamma_2)) + \hat{\mathcal{O}}((\gamma_1,\gamma_3)) - 1 \ge \hat{\mathcal{O}}(\gamma)$ for any $\gamma \in \Pi^N_3$ (seen by checking the cases). The final result (5.124) follows from recalling

(5.100), in turn implying that the summation term in (5.123) is indeed $\mathcal{O}(1)$.

To conclude this first part of the proof, we gather our estimates and obtain

$$\mathbb{E}\Big[\int_0^T L(s,\boldsymbol{X}_s^N)\mathrm{d}s\Big] \le K\int_0^T\int_0^1 \sqrt{e^{-\lambda_0(T-s)}e^{-2\lambda s}}\,\mathrm{d}\rho\mathrm{d}s$$
$$= Ke^{-\frac{\lambda_0}{2}T}\int_0^T e^{\frac{\lambda_0-2\lambda}{2}s}\mathrm{d}s \le Ke^{-\frac{\lambda_0}{2}T}.$$

Part 2: Estimating (5.119). For the term (5.119), we have

$$\sum_{m=0}^{M-1}\mathbb{E}\Big[\int_{t_m}^{t_{m+1}}\Big(L(t_m,\boldsymbol{X}_{t_m}^{N,h}) - L(s,\boldsymbol{X}_s^N)\Big)\mathrm{d}s\Big]$$
$$= \sum_{m=0}^{M-1}\int_{t_m}^{t_{m+1}}\mathbb{E}\Big[L(t_m,\boldsymbol{X}_{t_m}^{N,h}) - L(t_m,\boldsymbol{X}_{t_m}^{N})\Big]\mathrm{d}s \tag{5.125}$$
$$+ \sum_{m=0}^{M-1}\int_{t_m}^{t_{m+1}}\mathbb{E}\Big[L(t_m,\boldsymbol{X}_{t_m}^{N}) - L(s,\boldsymbol{X}_{t_m}^{N})\Big]\mathrm{d}s \tag{5.126}$$
$$+ \sum_{m=0}^{M-1}\int_{t_m}^{t_{m+1}}\mathbb{E}\Big[L(s,\boldsymbol{X}_{t_m}^{N}) - L(s,\boldsymbol{X}_{s}^{N})\Big]\mathrm{d}s. \tag{5.127}$$

Part 2.1: Estimating (5.125) *and* (5.127). For (5.125), similar to the calculations (5.121)–(5.124) (in the previous part of the proof), we derive

$$\mathbb{E}\Big[L(t_m,\boldsymbol{X}_{t_m}^{N,h}) - L(t_m,\boldsymbol{X}_{t_m}^{N})\Big]$$
$$= \int_0^1\mathbb{E}\Big[\Big\langle\partial_{\boldsymbol{x}}L\big(t_m,\rho\boldsymbol{X}_{t_m}^{N,h}+(1-\rho)\boldsymbol{X}_{t_m}^{N}\big),\ \boldsymbol{X}_{t_m}^{N,h}-\boldsymbol{X}_{t_m}^{N}\Big\rangle\Big]\mathrm{d}\rho$$
$$\le \frac{K}{N}\sum_{i=1}^N\sqrt{e^{-\lambda_0(T-t_m)}\mathbb{E}\Big[\big|X_{t_m}^{i,N,h}-X_{t_m}^{i,N}\big|^2\Big]} \le Kh^{1/2}e^{-\lambda_0(T-t_{m+1})/2}, \tag{5.128}$$

where we used Proposition 5.2.3 for the strong error rate. Similarly, for (5.127), we have for $s\in[t_m,t_{m+1}]$,

$$\mathbb{E}\Big[L(s,\boldsymbol{X}_{t_m}^{N}) - L(s,\boldsymbol{X}_{s}^{N})\Big] = \int_0^1\mathbb{E}\Big[\Big\langle\partial_{\boldsymbol{x}}L\big(s,\rho\boldsymbol{X}_{t_m}^{N}+(1-\rho)\boldsymbol{X}_{s}^{N}\big),\ \boldsymbol{X}_{s}^{N}-\boldsymbol{X}_{t_m}^{N}\Big\rangle\Big]\mathrm{d}\rho$$
$$\le \frac{K}{N}\sum_{i=1}^N\sqrt{e^{-\lambda_0(T-s)}\mathbb{E}\Big[\big|X_{s}^{i,N}-X_{t_m}^{i,N}\big|^2\Big]}$$
$$\le K\sqrt{s-t_m}e^{-\lambda_0(T-s)/2} \le Kh^{1/2}e^{-\lambda_0(T-t_{m+1})/2}, \tag{5.129}$$

where we used Proposition 6.4.1 and that $h\ge s-t_m$.

Part 2.2: Estimating (5.126). For the term (5.126), we get for all m and $s\in[t_m,t_{m+1}]$

$$\mathbb{E}\Big[L(t_m,\boldsymbol{X}_{t_m}^{N}) - L(s,\boldsymbol{X}_{t_m}^{N})\Big]$$
$$= \frac{1}{2}\mathbb{E}\Big[\sum_{i,j=1}^N\Big(B_j(\boldsymbol{X}_{t_m}^N)\partial_{x_j}B_i(\boldsymbol{X}_{t_m}^N)\Big(\partial_{x_i}u(t_m,\boldsymbol{X}_{t_m}^N)-\partial_{x_i}u(s,\boldsymbol{X}_{t_m}^N)\Big)$$
$$+\frac{\sigma^2}{2}\partial_{x_j}B_i(\boldsymbol{X}_{t_m}^N)\Big(\partial^2_{x_i,x_j}u(t_m,\boldsymbol{X}_{t_m}^N)-\partial^2_{x_i,x_j}u(s,\boldsymbol{X}_{t_m}^N)\Big)$$
$$+\frac{\sigma^2}{2}\partial^2_{x_j,x_j}B_i(\boldsymbol{X}_{t_m}^N)\Big(\partial_{x_i}u(t_m,\boldsymbol{X}_{t_m}^N)-\partial_{x_i}u(s,\boldsymbol{X}_{t_m}^N)\Big)\Big)\Big]$$

$$\leq K \sum_{i,j=1}^{N} \sqrt{\mathbb{E}\Big[\Big|B_j(\boldsymbol{X}_{t_m}^N)\partial_{x_j} B_i(\boldsymbol{X}_{t_m}^N)\Big|^2\Big]\mathbb{E}\Big[\Big|\partial_{x_i} u(t_m, \boldsymbol{X}_{t_m}^N) - \partial_{x_i} u(s, \boldsymbol{X}_{t_m}^N)\Big|^2\Big]} \tag{5.130}$$

$$+ K \sum_{i,j=1}^{N} \sqrt{\mathbb{E}\Big[\Big|\partial_{x_j} B_i(\boldsymbol{X}_{t_m}^N)\Big|^2\Big]\mathbb{E}\Big[\Big|\partial^2_{x_i,x_j} u(t_m, \boldsymbol{X}_{t_m}^N) - \partial^2_{x_i,x_j} u(s, \boldsymbol{X}_{t_m}^N)\Big|^2\Big]} \tag{5.131}$$

$$+ K \sum_{i,j=1}^{N} \sqrt{\mathbb{E}\Big[\Big|\partial^2_{x_j,x_j} B_i(\boldsymbol{X}_{t_m}^N)\Big|^2\Big]\mathbb{E}\Big[\Big|\partial_{x_i} u(t_m, \boldsymbol{X}_{t_m}^N) - \partial_{x_i} u(s, \boldsymbol{X}_{t_m}^N)\Big|^2\Big]}. \tag{5.132}$$

We first study the differences of first order derivatives of u in (5.130) and (5.132), and then study the difference of second order derivatives of u (5.131). Collecting these estimates and using the bounds on the derivatives of B yields the upper bound on (5.126). Then we will be in a position to conclude the main result.

Part 2.2.1: The first order derivative terms. We first derive the following estimate:

$$\begin{aligned}
&\mathbb{E}\Big[\big|\partial_{x_i} u(t_m, \boldsymbol{X}_{t_m}^N) - \partial_{x_i} u(s, \boldsymbol{X}_{t_m}^N)\big|^2\Big] \\
&= \mathbb{E}\Big[\Big| \sum_{j=1}^{N} \mathbb{E}\Big[\partial_{x_j} g(\boldsymbol{X}_T^{t_m,\boldsymbol{x},N}) \cdot X_{T,x_i}^{t_m,x_j,j,N} - \partial_{x_j} g(\boldsymbol{X}_T^{s,\boldsymbol{x},N}) \cdot X_{T,x_i}^{s,x_j,j,N}\Big] \Big|_{\boldsymbol{x}=\boldsymbol{X}_{t_m}^N} \Big|^2\Big] \\
&\leq K\mathbb{E}\Big[\Big| \mathbb{E}\Big[\sum_{j=1,j\neq i}^{N} \Big(\partial_{x_j} g(\boldsymbol{X}_T^{t_m,\boldsymbol{x},N}) \cdot X_{T,x_i}^{t_m,x_j,j,N} - \partial_{x_j} g(\boldsymbol{X}_T^{s,\boldsymbol{x},N}) \cdot X_{T,x_i}^{s,x_j,j,N}\Big)\Big] \Big|_{\boldsymbol{x}=\boldsymbol{X}_{t_m}^N} \Big|^2\Big] \\
&\quad + K\mathbb{E}\Big[\Big| \mathbb{E}\Big[\partial_{x_i} g(\boldsymbol{X}_T^{t_m,\boldsymbol{x},N}) \cdot X_{T,x_i}^{t_m,x_i,i,N} - \partial_{x_i} g(\boldsymbol{X}_T^{s,\boldsymbol{x},N}) \cdot X_{T,x_i}^{s,x_i,i,N}\Big] \Big|_{\boldsymbol{x}=\boldsymbol{X}_{t_m}^N} \Big|^2\Big] \\
&\leq K\mathbb{E}\Big[\Big| \sum_{j=1,j\neq i}^{N} \Big(\partial_{x_j} g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}_{t_m}^N,N}) \cdot X_{T,x_i}^{t_m,X_{t_m}^{j,N},j,N} - \partial_{x_j} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N}) \cdot X_{T,x_i}^{s,X_{t_m}^{j,N},j,N}\Big) \Big|^2\Big] \\
&\quad + K\mathbb{E}\Big[\Big| \partial_{x_i} g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}_{t_m}^N,N}) \cdot X_{T,x_i}^{t_m,X_{t_m}^{i,N},i,N} - \partial_{x_i} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N}) \cdot X_{T,x_i}^{s,X_{t_m}^{i,N},i,N} \Big|^2\Big]
\end{aligned} \tag{5.133}$$

$$\begin{aligned}
&\leq KN \sum_{j=1,j\neq i}^{N} \mathbb{E}\Big[\Big| \Big(\partial_{x_j} g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}_{t_m}^N,N}) - \partial_{x_j} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N})\Big) \cdot X_{T,x_i}^{t_m,X_{t_m}^{j,N},j,N} \Big|^2\Big] \\
&\quad + KN \sum_{j=1,j\neq i}^{N} \mathbb{E}\Big[\Big|\partial_{x_j} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N}) \cdot \Big(X_{T,x_i}^{t_m,X_{t_m}^{j,N},j,N} - X_{T,x_i}^{s,X_{t_m}^{j,N},j,N}\Big) \Big|^2\Big] \\
&\quad + K\mathbb{E}\Big[\Big| \Big(\partial_{x_i} g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}_{t_m}^N,N}) - \partial_{x_i} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N})\Big) \cdot X_{T,x_i}^{t_m,X_{t_m}^{i,N},i,N} \Big|^2\Big] \\
&\quad + K\mathbb{E}\Big[\Big| \partial_{x_i} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N}) \cdot \Big(X_{T,x_i}^{t_m,X_{t_m}^{i,N},i,N} - X_{T,x_i}^{s,X_{t_m}^{i,N},i,N}\Big) \Big|^2\Big],
\end{aligned} \tag{5.134}$$

where we employed the Cauchy–Schwarz and Jensen inequalities, as well as the tower property for conditional expectations to obtain (5.133). Inequality (5.134) follows from a standard rearrangement of (5.133).

An application of Hölder's inequality and Assumption 5.3.3 on the function g further yields

$$\begin{aligned}
(5.134) \leq \frac{K}{N} \sum_{j=1,j\neq i}^{N} &\sqrt{\mathbb{E}\Big[\big|X_T^{t_m,X_{t_m}^{j,N},j,N} - X_T^{s,X_{t_m}^{j,N},j,N}\big|^4\Big]\ \mathbb{E}\Big[\big|X_{T,x_i}^{t_m,X_{t_m}^{j,N},j,N}\big|^4\Big]} \\
&+ \frac{K}{N} \sum_{j=1,j\neq i}^{N} \mathbb{E}\Big[\big|X_{T,x_i}^{t_m,X_{t_m}^{j,N},j,N} - X_{T,x_i}^{s,X_{t_m}^{j,N},j,N}\big|^2\Big]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{K}{N^2}\sqrt{\mathbb{E}\Big[\big|X_T^{t_m,X_{t_m}^{i,N},i,N}-X_T^{s,X_{t_m}^{i,N},i,N}\big|^4\Big]\,\mathbb{E}\Big[\big|X_{T,x_i}^{t_m,X_{t_m}^{i,N},i,N}\big|^4\Big]}\\
&+\frac{K}{N^2}\mathbb{E}\Big[\big|X_{T,x_i}^{t_m,X_{t_m}^{i,N},i,N}-X_{T,x_i}^{s,X_{t_m}^{i,N},i,N}\big|^2\Big]\\
\le{}&\frac{K}{N^2}(s-t_m)e^{-2\lambda_2(T-s)}, \qquad (5.135)
\end{aligned}
$$

where we used Lemma 5.4.7 with $n=1$ (note that λ_0 can be replaced by λ_1 in this case), Lemma 6.4.2 and Proposition 5.4.4 with $\lambda_2\in(0,\min\{\lambda-2K_V,\lambda_1\})$ for the final inequality.

Part 2.2.2: The second order derivative terms. Similarly, for the differences of the second order derivatives of u in (5.131), applying the tower property of the conditional expectation once more we have

$$
\begin{aligned}
&\mathbb{E}\Big[\big|\partial^2_{x_i,x_j}u(t_m,\boldsymbol{X}^N_{t_m})-\partial^2_{x_i,x_j}u(s,\boldsymbol{X}^N_{t_m})\big|^2\Big]\\
&=\mathbb{E}\Bigg[\bigg|\ \sum_{k=1}^N\mathbb{E}\Big[\partial_{x_k}g(\boldsymbol{X}_T^{t_m,\boldsymbol{x},N})\cdot X_{T,x_i,x_j}^{t_m,x_k,k,N}-\partial_{x_k}g(\boldsymbol{X}_T^{s,\boldsymbol{x},N})\cdot X_{T,x_i,x_j}^{s,x_k,k,N}\Big]\\
&\quad+\sum_{k,k'=1}^N\mathbb{E}\Big[\partial_{x_k,x_{k'}}g(\boldsymbol{X}_T^{t_m,\boldsymbol{x},N})\cdot X_{T,x_i}^{t_m,x_k,k,N}X_{T,x_j}^{t_m,x_{k'},k',N}\\
&\qquad\qquad-\partial_{x_k,x_{k'}}g(\boldsymbol{X}_T^{s,\boldsymbol{x},N})\cdot X_{T,x_i}^{s,x_k,k,N}X_{T,x_j}^{s,x_{k'},k',N}\Big]\ \bigg|_{\boldsymbol{x}=\boldsymbol{X}^N_{t_m}}\ \bigg|^2\Bigg]\\
&\le K\mathbb{E}\Big[\Big|\ \sum_{k=1}^N\Big(\partial_{x_k}g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}^N_{t_m},N})-\partial_{x_k}g(\boldsymbol{X}_T^{s,\boldsymbol{X}^N_{t_m},N})\Big)\cdot X_{T,x_i,x_j}^{t_m,X_{t_m}^{k,N},k,N}\Big|^2\Big] \qquad (5.136)\\
&\quad+K\mathbb{E}\Big[\Big|\sum_{k=1}^N\Big(X_{T,x_i,x_j}^{t_m,X_{t_m}^{k,N},k,N}-X_{T,x_i,x_j}^{s,X_{t_m}^{k,N},k,N}\Big)\cdot\partial_{x_k}g(\boldsymbol{X}_T^{s,\boldsymbol{X}^N_{t_m},N})\Big|^2\Big] \qquad (5.137)\\
&\quad+K\mathbb{E}\Big[\Big|\ \sum_{k,k'=1}^N\Big(\partial_{x_k,x_{k'}}g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}^N_{t_m},N})-\partial_{x_k,x_{k'}}g(\boldsymbol{X}_T^{s,\boldsymbol{X}^N_{t_m},N})\Big)\\
&\qquad\qquad\cdot X_{T,x_i}^{t_m,X_{t_m}^{k,N},k,N}X_{T,x_j}^{t_m,X_{t_m}^{k',N},k',N}\Big|^2\Big] \qquad (5.138)\\
&\quad+K\mathbb{E}\Big[\Big|\ \sum_{k,k'=1}^N\Big(X_{T,x_i}^{t_m,X_{t_m}^{k,N},k,N}-X_{T,x_i}^{s,X_{t_m}^{k,N},k,N}\Big)\cdot\partial_{x_k,x_{k'}}g(\boldsymbol{X}_T^{s,\boldsymbol{X}^N_{t_m},N})\cdot X_{T,x_j}^{t_m,X_{t_m}^{k',N},k',N}\Big|^2\Big] \qquad (5.139)\\
&\quad+K\mathbb{E}\Big[\Big|\ \sum_{k,k'=1}^N\Big(X_{T,x_j}^{t_m,X_{t_m}^{k',N},k',N}-X_{T,x_j}^{s,X_{t_m}^{k',N},k',N}\Big)\cdot\partial_{x_k,x_{k'}}g(\boldsymbol{X}_T^{s,\boldsymbol{X}^N_{t_m},N})\cdot X_{T,x_i}^{s,X_{t_m}^{k,N},k,N}\Big|^2\Big]. \qquad (5.140)
\end{aligned}
$$

We are required to analyze each of the terms (5.136)–(5.140) separately. By further applying Jensen's inequality and Hölder's inequality, we derive the following estimates for the first two terms (5.136) and (5.137) based on the values of $\hat{\mathcal{O}}((i,j,k))$

$$
\begin{aligned}
&(5.136)\\
&\le K\sqrt{\mathbb{E}\Big[\Big|\ \partial_{x_i}g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}^N_{t_m},N})-\partial_{x_i}g(\boldsymbol{X}_T^{s,\boldsymbol{X}^N_{t_m},N})\Big|^4\Big]\mathbb{E}\Big[\Big|X_{T,x_i,x_j}^{t_m,X_{t_m}^{i,N},i,N}\Big|^4\ \Big]}\\
&\quad+K\sqrt{\mathbb{E}\Big[\Big|\ \partial_{x_j}g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}^N_{t_m},N})-\partial_{x_j}g(\boldsymbol{X}_T^{s,\boldsymbol{X}^N_{t_m},N})\Big|^4\Big]\mathbb{E}\Big[\Big|X_{T,x_i,x_j}^{t_m,X_{t_m}^{j,N},j,N}\Big|^4\ \Big]}\\
&\quad+NK\sum_{k=1,k\notin\{i,j\}}^N\sqrt{\mathbb{E}\Big[\Big|\ \partial_{x_k}g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}^N_{t_m},N})-\partial_{x_k}g(\boldsymbol{X}_T^{s,\boldsymbol{X}^N_{t_m},N})\Big|^4\Big]\mathbb{E}\Big[\Big|X_{T,x_i,x_j}^{t_m,X_{t_m}^{k,N},k,N}\Big|^4\ \Big]}. \qquad (5.141)
\end{aligned}
$$

Now, similar to (5.135), an application of Assumption 5.3.3 for the function g and then injecting the bounds from Lemma 5.4.7 with $n=2$ (note that λ_0 can be replaced by λ_4 in this case) and Lemma 6.4.2 yield

$$\begin{aligned}
(5.141) \leq \frac{K}{N^2}&\sqrt{\mathbb{E}\Big[\big|X_T^{t_m,X_{t_m}^{i,N},i,N} - X_T^{s,X_{t_m}^{i,N},i,N}\big|^4\Big]\ \mathbb{E}\Big[\big|X_{T,x_i,x_j}^{t_m,X_{t_m}^{i,N},i,N}\big|^4\Big]} \\
&+ \frac{K}{N^2}\sqrt{\mathbb{E}\Big[\big|X_T^{t_m,X_{t_m}^{j,N},j,N} - X_T^{s,X_{t_m}^{j,N},j,N}\big|^4\Big]\ \mathbb{E}\Big[\big|X_{T,x_i,x_j}^{t_m,X_{t_m}^{j,N},j,N}\big|^4\Big]} \\
&+ \frac{K}{N}\sum_{k=1,k\notin\{i,j\}}^{N}\sqrt{\mathbb{E}\Big[\big|X_T^{t_m,X_{t_m}^{k,N},k,N} - X_T^{s,X_{t_m}^{k,N},k,N}\big|^4\Big]\ \mathbb{E}\Big[\big|X_{T,x_i,x_j}^{t_m,X_{t_m}^{k,N},k,N}\big|^4\Big]} \\
\leq K(s-t_m)&e^{-2\lambda_2(T-s)}e^{-2\lambda_4(T-t_m)}\left(\frac{1}{N^2\cdot N^{2\hat{\mathcal{O}}((i,j))-2}} + \frac{1}{N}\sum_{k=1,k\notin\{i,j\}}^{N}\frac{1}{N^{2\hat{\mathcal{O}}((i,j,k))-2}}\right).
\end{aligned} \tag{5.142}$$

Similarly, for (5.137), using Proposition 6.4.3, we have

$$\begin{aligned}
(5.137) \leq K&\sqrt{\mathbb{E}\Big[\Big|\Big(X_{T,x_i,x_j}^{t_m,X_{t_m}^{i,N},i,N} - X_{T,x_i,x_j}^{s,X_{t_m}^{i,N},i,N}\Big)\Big|^4\Big]\mathbb{E}\Big[\Big|\partial_{x_i} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N})\Big|^4\Big]} \\
&+ K\sqrt{\mathbb{E}\Big[\Big|\Big(X_{T,x_i,x_j}^{t_m,X_{t_m}^{j,N},j,N} - X_{T,x_i,x_j}^{s,X_{t_m}^{j,N},j,N}\Big)\Big|^4\Big]\mathbb{E}\Big[\Big|\partial_{x_j} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N})\Big|^4\Big]} \\
&+ NK\sum_{k=1,k\notin\{i,j\}}^{N}\sqrt{\mathbb{E}\Big[\Big|\Big(X_{T,x_i,x_j}^{t_m,X_{t_m}^{k,N},k,N} - X_{T,x_i,x_j}^{s,X_{t_m}^{k,N},k,N}\Big)\Big|^4\Big]\mathbb{E}\Big[\Big|\partial_{x_k} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N})\Big|^4\Big]} \\
&\leq K(s-t_m)e^{-2\lambda_4(T-t_m)}\left(\frac{1}{N^2\cdot N^{2\hat{\mathcal{O}}((i,j))-2}} + \frac{1}{N}\sum_{k=1,k\notin\{i,j\}}^{N}\frac{1}{N^{2\hat{\mathcal{O}}((i,j,k))-2}}\right).
\end{aligned} \tag{5.143}$$

For (5.138)–(5.140), we will make use of the following bound, which follows from exploiting the Lipschitz property of $\partial_{x_k,x_k'} g$ from Assumption 5.3.3 and applying Lemma 6.4.2

$$\mathbb{E}\Big[\Big|\partial_{x_k,x_{k'}} g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}_{t_m}^N,N}) - \partial_{x_k,x_{k'}} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N})\Big|^4\Big] \leq \frac{K}{N^{4\hat{\mathcal{O}}((k,k'))}}e^{-4\lambda_2(T-s)}.$$

Hence, similarly to (5.142)–(5.143), partitioning the sum based on values taken by $\hat{\mathcal{O}}((i,j,k,k'))$

$$\begin{aligned}
(5.138) \leq K&\Bigg(\sum_{\substack{k,k'\in\{1,\dots,N\},\\ \hat{\mathcal{O}}((i,j,k,k'))=\hat{\mathcal{O}}((i,j))}} + N\sum_{\substack{k,k'\in\{1,\dots,N\},\\ \hat{\mathcal{O}}((i,j,k,k'))=\hat{\mathcal{O}}((i,j))+1}} + N^2\sum_{\substack{k,k'\in\{1,\dots,N\},\\ \hat{\mathcal{O}}((i,j,k,k'))=\hat{\mathcal{O}}((i,j))+2}}\Bigg) \\
&\cdot\sqrt{\mathbb{E}\Big[\Big|\partial_{x_k,x_{k'}} g(\boldsymbol{X}_T^{t_m,\boldsymbol{X}_{t_m}^N,N}) - \partial_{x_k,x_{k'}} g(\boldsymbol{X}_T^{s,\boldsymbol{X}_{t_m}^N,N})\Big|^4\Big]\mathbb{E}\Big[\Big|X_{T,x_i}^{t_m,X_{t_m}^{k,N},k,N}X_{T,x_j}^{t_m,X_{t_m}^{k',N},k',N}\Big|^4\Big]} \\
\leq K(s-t_m)&e^{-2\lambda_2(T-s)}e^{-4\lambda_1(T-t_m)} \\
&\cdot\Bigg(\sum_{\substack{k,k'\in\{1,\dots,N\},\\ \hat{\mathcal{O}}((i,j,k,k'))=\hat{\mathcal{O}}((i,j))}}\frac{1}{N^{2\hat{\mathcal{O}}((k,k'))}}\frac{1}{N^{2(\hat{\mathcal{O}}((i,k))+\hat{\mathcal{O}}((j,k'))-2))}} \\
&\quad + N\sum_{\substack{k,k'\in\{1,\dots,N\},\\ \hat{\mathcal{O}}((i,j,k,k'))=\hat{\mathcal{O}}((i,j))+1}}\frac{1}{N^{2\hat{\mathcal{O}}((k,k'))}}\frac{1}{N^{2(\hat{\mathcal{O}}((i,k))+\hat{\mathcal{O}}((j,k'))-2))}} \\
&\quad + N^2\sum_{\substack{k,k'\in\{1,\dots,N\},\\ \hat{\mathcal{O}}((i,j,k,k'))=\hat{\mathcal{O}}((i,j))+2}}\frac{1}{N^{2\hat{\mathcal{O}}((k,k'))}}\frac{1}{N^{2(\hat{\mathcal{O}}((i,k))+\hat{\mathcal{O}}((j,k'))-2))}}\Bigg)
\end{aligned} \tag{5.144}$$

$$\le \frac{K(s-t_m)e^{-2\lambda_2(T-s)}e^{-4\lambda_1(T-t_m)}}{N^{2\hat{\mathcal{O}}((i,j))}} \cdot \sum_{m=0}^{2} N^m \sum_{\substack{k,k'\in\{1,\dots,N\},\\ \hat{\mathcal{O}}((i,j,k,k'))=\hat{\mathcal{O}}((i,j))+m}} \frac{1}{N^{2m}} \tag{5.145}$$

$$\le \frac{K(s-t_m)e^{-2\lambda_2(T-s)}e^{-4\lambda_1(T-t_m)}}{N^{2\hat{\mathcal{O}}((i,j))}}. \tag{5.146}$$

The term (5.144) is derived by applying the Cauchy–Schwarz inequality, which then requires the L^8-moments of $X_{T,x_i}^{t_m,X_{t_m}^{k,N},k,N}$ and $X_{T,x_j}^{t_m,X_{t_m}^{k',N},k',N}$ obtained from Lemma 5.4.2 (that holds without integrability requirements on $X_{t_m}^{k,N}, X_{t_m}^{k',N}$ since (5.21) is a linear ODE with bounded coefficients starting from either 1 or 0), and noting we can unify the bounds in Lemma 5.4.2 as (note that $\lambda_1 < \lambda$ and $\hat{\mathcal{O}}((i,k))$ is either 0 or 1)

$$\mathbb{E}\big[\,|X_{T,x_i}^{t_m,X_{t_m}^{k,N},k,N}|^8\big] \le \frac{K}{N^{8(\hat{\mathcal{O}}((i,k))-1)}}e^{-8\lambda_1(T-t_m)}.$$

To obtain (5.145), we used the following bounds, which can be confirmed by checking cases

$$\begin{aligned}
&2\hat{\mathcal{O}}\big((k,k')\big) + 2\big(\hat{\mathcal{O}}((i,k)) + \hat{\mathcal{O}}((j,k')) - 2\big)\\
&\qquad \ge \begin{cases} 2 + 2(1+\hat{\mathcal{O}}((i,j)) - 2) = 2\hat{\mathcal{O}}((i,j)), & \hat{\mathcal{O}}((i,j,k,k')) = \hat{\mathcal{O}}((i,j)),\\ 4 + 2(1+2-2) = 6 \ge 2 + 2\hat{\mathcal{O}}((i,j)), & \hat{\mathcal{O}}((i,j,k,k')) = \hat{\mathcal{O}}((i,j)) + 1,\ k \ne k',\\ 2 + 2(2+2-2) = 6 \ge 2 + 2\hat{\mathcal{O}}((i,j)), & \hat{\mathcal{O}}((i,j,k,k')) = \hat{\mathcal{O}}((i,j)) + 1,\ k = k',\\ 4 + 2(2+2-2) = 8 \ge 4 + 2\hat{\mathcal{O}}((i,j)), & \hat{\mathcal{O}}((i,j,k,k')) = \hat{\mathcal{O}}((i,j)) + 2. \end{cases}
\end{aligned}$$

In (5.146), we used (5.100) to ensure the summation term in (5.145) is $\mathcal{O}(1)$. We establish bounds for (5.139) and (5.140) in a similar fashion and obtain

$$(5.139) + (5.140) \le \frac{K(s-t_m)e^{-2\lambda_4(T-s)}}{N^{2\hat{\mathcal{O}}((i,j))}}. \tag{5.147}$$

Substituting (5.142)–(5.147) into (5.136)–(5.140) yields

$$\mathbb{E}\Big[\big|\partial^2_{x_i,x_j}u(t_m,\boldsymbol{X}^N_{t_m}) - \partial^2_{x_i,x_j}u(s,\boldsymbol{X}^N_{t_m})\big|^2\Big] \le \frac{K(s-t_m)e^{-2\lambda_4(T-s)}}{N^{2\hat{\mathcal{O}}((i,j))}}. \tag{5.148}$$

Part 2.3: Collecting the estimates for (5.130)*–*(5.132)*.* Consequently, substituting (5.135) and (5.148) into (5.130)–(5.132) and using the bounds for the derivatives of the function B in (5.96), we conclude that there exists some constant $\lambda_0 \in (0,\lambda)$ such that for all m and $s \in [t_m, t_{m+1}]$,

$$\begin{aligned}
&\mathbb{E}\Big[L(t_m,\boldsymbol{X}^N_{t_m}) - L(s,\boldsymbol{X}^N_{t_m})\Big]\\
&\quad\le K\sum_{i,j=1}^{N}\Bigg(\frac{\sqrt{s-t_m}}{N^{\hat{\mathcal{O}}((i,j))-1}\cdot N}e^{-\lambda_2(T-s)} + \frac{\sqrt{s-t_m}}{N^{\hat{\mathcal{O}}((i,j))-1}N^{\hat{\mathcal{O}}((i,j))}}e^{-\lambda_4(T-s)}\\
&\qquad\qquad + \frac{\sqrt{s-t_m}}{N^{\hat{\mathcal{O}}((i,j))-1}\cdot N}e^{-\lambda_2(T-s)}\Bigg)\\
&\quad\le Kh^{1/2}e^{-\lambda_0(T-s)}\sum_{i,j=1}^{N}\frac{1}{N^{\hat{\mathcal{O}}((i,j))}} \le Kh^{1/2}e^{-\lambda_0(T-t_{m+1})},
\end{aligned} \tag{5.149}$$

where once again the final inequality arises from recalling (5.100).

Part 3: Concluding the proof. Substituting the results of (5.128), (5.129), (5.149) to (5.125)–

(5.127), we conclude that

$$\sum_{m=0}^{M-1}\mathbb{E}\Big[\int_{t_m}^{t_{m+1}}\Big(L(t_m,\boldsymbol{X}^{N,h}_{t_m})-L(s,\boldsymbol{X}^N_s)\Big)\mathrm{d}s\Big]\le Kh^{3/2}\sum_{m=0}^{M-1}e^{-\lambda_0(T-t_{m+1})}\le Kh^{1/2}.$$

Therefore, for the left-hand side of (5.120), there exists some positive constants $\lambda_0\in(0,\lambda),K$ such that

$$h^2\mathbb{E}\Big[\sum_{m=0}^{M-1}L(t_m,\boldsymbol{X}^{N,h}_{t_m})\Big]\le Kh^{3/2}+Khe^{-\lambda_0 T}.$$

□

Remark 5.6.2 (On losing 1/2 in the rate of convergence)**.** *Letting $T\to\infty$ in* (5.120)*, and temporarily ignoring higher-order remainder terms, we have that the weak error is of order* 3/2*. In [103] this term satisfies the bound (with K depending on the dimension N)*

$$h^2\mathbb{E}\Big[\sum_{m=0}^{M-1}L(t_m,\boldsymbol{X}^{N,h}_{t_m})\Big]\le Kh^{2}+Khe^{-\lambda_0 T},$$

and thus a weak order of 2 *is attained.*

The loss of 1/2 *in the convergence rate in our estimates occurs in the final step of* (5.128)*. In [103] this term is estimated with the weak error (starting on* (5.125)*), whilst here, we are not able to do so. In fact, to recover the missing* 1/2*-rate by following using the arguments [103] one would need to show that L in* (5.20) *satisfies condition (3) of Assumption 5.3.3 – see additionally our Remark 5.3.4. At present, this is an open question.*

5.6.2 Analysis of residual terms

A close inspection of the proof of the main result in [103] shows that there are 8 remainder terms which need to be analyzed and we do so in the next lemma. We will make use of the following helpful abbreviation: for $m\in\{0,\dots,M-1\}$, we define using (5.10) and (5.13)

$$\begin{aligned}\Delta X^{i,N,h}_{t_m}:=&X^{i,N,h}_{t_{m+1}}-X^{i,N,h}_{t_m}\\ =&-\Big(\nabla U(X^{i,N,h}_{t_m})+\frac{1}{N}\sum_{j=1}^N\nabla V(X^{i,N,h}_{t_m}-X^{j,N,h}_{t_m})\Big)h+\frac{\sigma}{2}(\Delta W^i_m+\Delta W^i_{m+1}) \qquad (5.150)\\ =&B_i(\boldsymbol{X}^{N,h}_{t_m})h+\frac{\sigma}{2}(\Delta W^i_m+\Delta W^i_{m+1}).\end{aligned}$$

Further, we define the continuous extension of $\boldsymbol{X}^{N,h}_{t_m}$: for all $s\in[0,h]$

$$X^{i,N,h}_{t_m+s}:=X^{i,N,h}_{t_m}+\Big(\nabla U(X^{i,N,h}_{t_m})+\frac{1}{N}\sum_{j=1}^N\nabla V(X^{i,N,h}_{t_m}-X^{j,N,h}_{t_m})+\frac{\sigma}{2h}\Delta W^i_m\Big)s+\frac{\sigma}{2}\Delta W^i_{m+1,s} \qquad (5.151)$$

$$\begin{aligned}&=X^{i,N,h}_{t_m}+\Big(B_i(\boldsymbol{X}^{N,h}_{t_m})+\frac{\sigma\Delta W^i_m}{2h}\Big)s+\frac{\sigma}{2}\Delta W^i_{m+1,s},\\ \Delta W^i_{m,s}&=W^i_{t_{m-1}+s}-W^i_{t_{m-1}},\ \text{for } m>0. \qquad (5.152)\end{aligned}$$

Equivalently, we could write

$$\begin{aligned}\Delta\boldsymbol{X}^{N,h}_{t_m}&:=\boldsymbol{X}^{N,h}_{t_{m+1}}-\boldsymbol{X}^{N,h}_{t_m}=B(\boldsymbol{X}^{N,h}_{t_m})h+\frac{\sigma}{2}(\Delta\boldsymbol{W}_m+\Delta\boldsymbol{W}_{m+1}),\\ \boldsymbol{X}^{N,h}_{t_m+s}&:=\boldsymbol{X}^{N,h}_{t_m}+\Big(B(\boldsymbol{X}^{N,h}_{t_m})+\frac{\sigma\Delta\boldsymbol{W}_{m,h}}{2h}\Big)s+\frac{\sigma}{2}\Delta\boldsymbol{W}_{m+1,s},\end{aligned}$$

where $\Delta \boldsymbol{X}^{N,h}_{t_m} := (\Delta X^{1,N,h}_{t_m}, \dots, \Delta X^{N,N,h}_{t_m})$, $\Delta \boldsymbol{W}_m := (\Delta W^1_m, \dots, \Delta W^N_m)$ and $\Delta \boldsymbol{W}_{m,s} := (\Delta W^1_{m,s}, \dots, \Delta W^N_{m,s})$ (for $m > 0$). Instead of dealing with the expression above, we rewrite the scheme in a different way (see [103, p7]). For $m \in \{0, \dots, M-1\}$ and $i \in \{1, \dots, N\}$, with $\hat{X}^{i,N,h}_{t_0} = X^{i,N}_{t_0}$, define

$$\begin{aligned} \hat{X}^{i,N,h}_{t_{m+1}} &= \hat{X}^{i,N,h}_{t_m} + \sigma \Delta W^i_m \\ &\quad - \Big(\nabla U(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m) + \frac{1}{N}\sum_{j=1}^{N} \nabla V\big(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m - \hat{X}^{j,N,h}_{t_m} - \frac{\sigma}{2}\Delta W^j_m\big)\Big)h, \\ \hat{\boldsymbol{X}}^{N,h}_{t_{m+1}} &= \hat{\boldsymbol{X}}^{N,h}_{t_m} + \sigma \Delta \boldsymbol{W}_m + B\big(\hat{\boldsymbol{X}}^{N,h}_{t_m} + \frac{\sigma}{2}\Delta \boldsymbol{W}_m\big)h, \end{aligned} \tag{5.153}$$

so that $X^{i,N,h}_{t_m} = \hat{X}^{i,N,h}_{t_m} + \sigma \Delta W^i_m/2$, for all $i \in \{1, \dots, N\}$ (or $\boldsymbol{X}^{N,h}_{t_m} = \hat{\boldsymbol{X}}^{N,h}_{t_m} + \sigma \Delta \boldsymbol{W}_m/2$).

Now, for $s \in [0, h)$ and $m \in \{1, \dots, M\}$, we define the following auxiliary process

$$\begin{aligned} \overline{X}^{i,N,h}_{t_{m-1}+s} &= \hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_{m,s}, \quad \overline{X}^{i,N,h}_{t_{m-1}} = \hat{X}^{i,N,h}_{t_m}, \\ \overline{\boldsymbol{X}}^{N,h}_{t_{m-1}+s} &= \hat{\boldsymbol{X}}^{N,h}_{t_m} + \frac{\sigma}{2}\Delta \boldsymbol{W}_{m,s}, \quad \overline{\boldsymbol{X}}_{t_{m-1}} = \hat{\boldsymbol{X}}^{N,h}_{t_m}, \end{aligned} \tag{5.154}$$

where we remark that this form is used in the proof of Lemma 5.6.3 (e.g., in (5.161)). The moment stability of these auxiliary schemes is discussed in the appendix, see Lemma 6.4.5. We refer the reader to [139] for different versions of such schemes, coined there 'postprocessed schemes', achieving higher-order weak convergence in the ergodic setting.

Lemma 5.6.3. *Let Assumption 5.3.3 hold and let $\xi \in L^{10}(\Omega, \mathbb{R})$. Then for the remainder term R in* (5.116)*, there exists a positive constant K independent of h, T, M and N, such that*

$$\mathbb{E}\left[\sum_{m=0}^{M-1} R(t_m, \boldsymbol{X}^{N,h}_{t_m})\right] \le K h^{3/2}.$$

Proof. Recall that the remainder term $r(t_m, \cdot)$ (corresponding to our $R(t_m, \cdot)$ in (5.116)) in [103, Equation (3.17)], itself is given as a linear combination of remainders denoted by $h^3 r_i(t_m, \cdot)$ for $i \in \{1, \dots, 8\}$ appearing in Equations (3.8), (3.10)–(3.16) in [103, p7-9]. In our case, we derive bounds for $\mathbb{E}[R(t_m, \boldsymbol{X}^{N,h}_{t_m})] \le \sum_{i=1}^{8} |\mathbb{E}[R^i_{t_m}]|$, where each $R^i_{t_m}$ term corresponds to the $h^3 r_i(t_m, \cdot)$ terms ($i \in \{1, \dots, 8\}$) of [103, p7-9].

It may not be immediately transparent how the $h^3 r_i(t_m, \cdot)$ terms correspond to our $R^i_{t_m}$ terms. We explicitly present the derivations of $\mathbb{E}[R^1_{t_m}]$, $\mathbb{E}[R^4_{t_m}]$ and $\mathbb{E}[R^6_{t_m}]$, which we feel encapsulate the techniques and proof methodology. The derivation of the residual terms $R^2_{t_m}$, $R^3_{t_m}$, $R^5_{t_m}$, $R^7_{t_m}$, $R^8_{t_m}$, can be found in Appendix 6.4.4. In this proof, let K be a constant independent of m, M, N, T, h which may change from line to line.

Part 1: Recall the ΔX notation introduced in (5.150). The remainder term $h^3 r_1$ in Equation (3.8) of [103, p7] is established from the Taylor expansion with Lagrange's form of the remainder term. That is, for all m, there exists some $\rho_m \in (0, 1)$ such that

$$\mathbb{E}\big[R^1_{t_m}\big] = K\mathbb{E}\Big[\sum_{\gamma \in \Pi^N_5} \Delta X^{\gamma_1,N,h}_{t_m} \dots \Delta X^{\gamma_5,N,h}_{t_m}\ \partial^5_{\gamma_1,\dots,\gamma_5} u(t_{m+1}, \boldsymbol{X}^{N,h}_{m,\rho_m})\Big]$$

where $\boldsymbol{X}^{N,h}_{m,\rho_m} := \rho_m \boldsymbol{X}^{N,h}_{t_m} + (1-\rho_m)\boldsymbol{X}^{N,h}_{t_{m+1}}$.

We deal with $|\mathbb{E}[R^1_{t_m}]|$ via Hölder's inequality to isolate the ΔX terms from the derivatives of u term,

$$\big|\mathbb{E}\big[R^1_{t_m}\big]\big| \le K \sum_{\gamma \in \Pi^N_5} \sqrt{\mathbb{E}\Big[\big|\Delta X^{\gamma_1,N,h}_{t_m} \dots \Delta X^{\gamma_5,N,h}_{t_m}\big|^2\Big]\mathbb{E}\Big[\ \Big|\partial^5_{x_{\gamma_1},\dots,x_{\gamma_5}}\ u(t_{m+1}, \boldsymbol{X}^{N,h}_{m,\rho_m})\Big|^2\Big]}$$

$$\leq Kh^{5/2} \sum_{\gamma \in \Pi_5^N} e^{-\lambda_0(T-t_{m+1})} N^{-\hat{\mathcal{O}}(\gamma)} \leq Kh^{5/2} e^{-\lambda_0(T-t_{m+1})}, \tag{5.155}$$

which holds for some $\lambda_0 \in (0, \lambda)$. We derive (5.155) using Lemma 5.5.3 to treat the derivatives of u, while for the ΔX terms, we use their explicit forms (5.150) and the fact that $\nabla U, \nabla V$ are of linear growth. In particular, the assumption $\xi \in L^{10}(\Omega, \mathbb{R})$, allows to use Proposition 5.2.2 and Proposition 5.2.3 to control the L^{10}-moments of the processes. The moment properties of the increments extract the leading order h terms; providing the $h^{5/2}$ leading order. Note that for $R^1_{t_m}$ we incur a loss of $h^{1/2}$ in the leading term while in [103] there is no such loss – this issue has already appeared in the proof of Lemma 5.6.1 and is discussed in Remark 5.6.2 (and Remark 5.3.4). Critically, [103] has a $|\cdot|_\infty$- norm bound for the 5-th derivative of u while here we are only able to bound it in expectation. We are forced to use Hölder's inequality.

Similar to [103, Equation (3.12)], we show how the $R^4_{t_m}$ (corresponding to $r_4(t_m, \cdot)h^3$ in [103, Equation (3.12)]) is generated and give the exact form of $\mathbb{E}[R^4_{t_m}]$. Expanding out the increments $\Delta X^{\gamma_i,N,h}_{t_m}$, $i \in \{1,2,3\}$, and recall the definition of $\Delta W_{m,2h}$ in (5.152), we have

$$\sum_{\gamma \in \Pi_3^N} \mathbb{E}\Big[\Delta X^{\gamma_1,N,h}_{t_m} \Delta X^{\gamma_2,N,h}_{t_m} \Delta X^{\gamma_3,N,h}_{t_m}\ \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{t_m})\Big]$$

$$= K \sum_{\gamma \in \Pi_3^N} \mathbb{E}\Big[\Delta W^{\gamma_1}_{m,2h} \Delta W^{\gamma_2}_{m,2h} \Delta W^{\gamma_3}_{m,2h}\ \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{t_m})\Big] \tag{5.156}$$

$$+ Kh \sum_{\gamma \in \Pi_3^N} \mathbb{E}\Big[B_{\gamma_1}(\boldsymbol{X}^{N,h}_{t_m}) \Delta W^{\gamma_2}_{m,2h} \Delta W^{\gamma_3}_{m,2h}\ \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{t_m})\Big] \tag{5.157}$$

$$+ Kh^2 \sum_{\gamma \in \Pi_3^N} \mathbb{E}\Big[B_{\gamma_1}(\boldsymbol{X}^{N,h}_{t_m}) B_{\gamma_2}(\boldsymbol{X}^{N,h}_{t_m}) \Delta W^{\gamma_3}_{m,2h}\ \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{t_m})\Big] \tag{5.158}$$

$$+ Kh^3 \sum_{\gamma \in \Pi_3^N} \mathbb{E}\Big[B_{\gamma_1}(\boldsymbol{X}^{N,h}_{t_m}) B_{\gamma_2}(\boldsymbol{X}^{N,h}_{t_m}) B_{\gamma_3}(\boldsymbol{X}^{N,h}_{t_m})\ \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{t_m})\Big]. \tag{5.159}$$

We first deal with the term (5.156), by applying an Itô-Taylor expansion (see [95, Section 5.1, p.163–164]) around $\hat{\boldsymbol{X}}^{N,h}_{t_m}$ to obtain

$$(5.156)$$

$$= K \sum_{\gamma \in \Pi_3^N} \mathbb{E}\Big[\Delta W^{\gamma_1}_{m,2h} \Delta W^{\gamma_2}_{m,2h} \Delta W^{\gamma_3}_{m,2h}\ \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \hat{\boldsymbol{X}}^{N,h}_{t_m})\Big] \tag{5.160}$$

$$+ K \sum_{\gamma \in \Pi_4^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h} \partial_{x_{\gamma_1}} \Big(\Delta W^{\gamma_2}_{m,2h} \Delta W^{\gamma_3}_{m,2h} \Delta W^{\gamma_4}_{m,2h}\ \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big) \mathrm{d}W^{\gamma_1}_{q_1}\Big] \tag{5.161}$$

$$+ K \sum_{\gamma \in \Pi_4^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h} \partial^2_{x_{\gamma_1},x_{\gamma_1}} \Big(\Delta W^{\gamma_2}_{m,2h} \Delta W^{\gamma_3}_{m,2h} \Delta W^{\gamma_4}_{m,2h}\ \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big) \mathrm{d}q_1\Big]. \tag{5.162}$$

One observes that (5.160) = 0 since for $\gamma \in \Pi_3^N$, regardless of the value of $\hat{\mathcal{O}}(\gamma)$, there will always be an odd power of a Brownian increment $\Delta W_{m,2h}$ presented in (5.152), independent of the Brownian increment contained in the $\hat{\boldsymbol{X}}^{N,h}_{t_m}$ term. The term (5.162) however, does not vanish since the $\Delta W_{m,2h}$ term is not independent of $\overline{\boldsymbol{X}}^{N,h}_{q_1}$, because the latter contains the Brownian increment $W_{q_1} - W_{t_m}$.

Applying a further Itô-Taylor expansion around $\hat{\boldsymbol{X}}^{N,h}_{t_m}$, and splitting the zeroth order term into the cases $\hat{\mathcal{O}}(\gamma) \in \{1,2\}$ yields

$$(5.161) =$$

$$Kh^2 \sum_{i=1}^{N} \mathbb{E}\Big[\partial^4_{x_i,x_i,x_i,x_i} u(t_{m+1}, \hat{\boldsymbol{X}}^{N,h}_{t_m})\Big] + Kh^2 \sum_{i,j=1,\ i\neq j}^{N} \mathbb{E}\Big[\partial^4_{x_i,x_i,x_j,x_j} u(t_{m+1}, \hat{\boldsymbol{X}}^{N,h}_{t_m})\Big] \tag{5.163}$$

$$+ K \sum_{\gamma\in\Pi_5^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1} \tag{5.164}$$

$$\partial_{x_{\gamma_1},x_{\gamma_2}}\Big(\Delta W^{\gamma_3}_{m,2h}\Delta W^{\gamma_4}_{m,2h}\Delta W^{\gamma_5}_{m,2h}\ \partial^3_{x_{\gamma_4},x_{\gamma_4},x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_2})\Big) \mathrm{d}W^{\gamma_1}_{q_2}\ \mathrm{d}W^{\gamma_2}_{q_1}\Big]$$

$$+ K \sum_{\gamma\in\Pi_5^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1} \tag{5.165}$$

$$\partial^3_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2}}\Big(\Delta W^{\gamma_3}_{m,2h}\Delta W^{\gamma_4}_{m,2h}\Delta W^{\gamma_5}_{m,2h}\ \partial^3_{x_{\gamma_4},x_{\gamma_4},x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_2})\Big) \mathrm{d}q_2\ \mathrm{d}W^{\gamma_2}_{q_1}\Big],$$

where the two summation terms in (5.163) correspond to the second and third summation in [103, (3.12)] and they do not contribute to the remainder term $R^4_{t_m}$. Similarly for (5.157), we have

$$(5.157) = Kh \sum_{i,j=1}^{N} \mathbb{E}\Big[B_i(\hat{\boldsymbol{X}}^{N,h}_{t_m})\partial^3_{x_i,x_j,x_j} u(t_{m+1}, \hat{\boldsymbol{X}}^{N,h}_{t_m})\Big] \tag{5.166}$$

$$+ Kh \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h} \partial_{x_{\gamma_1}}\Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\Delta W^{\gamma_3}_{m,2h}\Delta W^{\gamma_4}_{m,2h}\ \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big)\mathrm{d}W^{\gamma_1}_{q_1}\Big] \tag{5.167}$$

$$+ Kh \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h} \partial^2_{x_{\gamma_1},x_{\gamma_1}}\Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\Delta W^{\gamma_3}_{m,2h}\Delta W^{\gamma_4}_{m,2h}\ \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big)\mathrm{d}q_1\Big], \tag{5.168}$$

where (5.166) corresponds to the first summation in [103, (3.12)], thus also does not contribute to $R^4_{t_m}$. Leimkuhler et al. isolates and separates the terms (5.163) and (5.166) from $h^3 r_4$, since these terms cleverly cancel with other corresponding lower order terms in the expansion of other r_i's when everything is summed over.

As for (5.158), we have

$$(5.158) = Kh^2\Big(\sum_{\gamma\in\Pi_3^N} \mathbb{E}\Big[B_{\gamma_1}(\hat{\boldsymbol{X}}^{N,h}_{t_m})B_{\gamma_2}(\hat{\boldsymbol{X}}^{N,h}_{t_m})\Delta W^{\gamma_3}_{m,2h}\ \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \hat{\boldsymbol{X}}^{N,h}_{t_m})\Big] \tag{5.169}$$

$$+ \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h} \partial_{x_{\gamma_1}}\Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1})B_{\gamma_3}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\Delta W^{\gamma_4}_{m,2h}\ \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big)\mathrm{d}W^{\gamma_1}_{q_1}\Big] \tag{5.170}$$

$$+ \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h} \partial^2_{x_{\gamma_1},x_{\gamma_1}}\Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1})B_{\gamma_3}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\Delta W^{\gamma_4}_{m,2h}\ \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big)\mathrm{d}q_1\Big]\Big), \tag{5.171}$$

where the first term in this expansion (5.169) is zero, following the same reasoning as for (5.160).

For the residual term $R^4_{t_m}$, we have that via direct application of Hölder's inequality, Itô's isometry and Fubini's Theorem, we have

$$\begin{aligned}\mathbb{E}\big[R^4_{t_m}\big] &= (5.159) + (5.162) + (5.164) + (5.165) + (5.167) + (5.168) + (5.170) + (5.171)\\ &\leq Kh^3 \sum_{\gamma\in\Pi_3^N} \mathbb{E}\Big[B_{\gamma_1}(\boldsymbol{X}^{N,h}_{t_m})B_{\gamma_2}(\boldsymbol{X}^{N,h}_{t_m})B_{\gamma_3}(\boldsymbol{X}^{N,h}_{t_m})\ \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{t_m})\Big]\end{aligned}$$

$$+ Kh^{3/2} \int_{t_m}^{t_m+h} \left(\mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_4^N} \partial^5_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1}) \Big|^2 \Big]\right)^{1/2} \mathrm{d}q_1$$

$$+ Kh^2 \Big(\int_{t_m}^{t_m+h} \mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_5^N} \partial^5_{x_{\gamma_1},\ldots,x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1}) \Big|^2 \Big] \mathrm{d}q_1\Big)^{1/2}$$

$$+ Kh^2 \Big(\int_{t_m}^{t_m+h} \mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_5^N} \partial^6_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_4},x_{\gamma_4},x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1}) \Big|^2 \Big] \mathrm{d}q_1\Big)^{1/2}$$

$$+ Kh^2 \Big(\int_{t_m}^{t_m+h} \mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_4^N} \partial_{x_{\gamma_1}} \Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1}) \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big) \Big|^2 \Big] \mathrm{d}q_1\Big)^{1/2}$$

$$+ Kh^{3/2} \int_{t_m}^{t_m+h} \left(\mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_4^N} \partial^2_{x_{\gamma_1},x_{\gamma_1}} \Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1}) \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big) \Big|^2 \Big]\right)^{1/2} \mathrm{d}q_1$$

$$+ Kh^{5/2}$$

$$\cdot \Big(\int_{t_m}^{t_m+h} \mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_4^N} \partial_{x_{\gamma_1}} \Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1}) B_{\gamma_3}(\overline{\boldsymbol{X}}^{N,h}_{q_1}) \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big) \Big|^2 \Big] \mathrm{d}q_1\Big)^{1/2}$$

$$+ Kh^{5/2} \int_{t_m}^{t_m+h}$$

$$\cdot \left(\mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_4^N} \partial^2_{x_{\gamma_1},x_{\gamma_1}} \Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1}) B_{\gamma_3}(\overline{\boldsymbol{X}}^{N,h}_{q_1}) \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big) \Big|^2 \Big]\right)^{1/2} \mathrm{d}q_1.$$

As for $R^6_{t_m}$, recalling the definition of the operator $\mathcal{L}_N$

$$\mathcal{L}_N u = \sum_{i=1}^N \Big(B_i \partial_{x_i} u + \frac{1}{2}\sigma^2 \partial^2_{x_i,x_i} u\Big),$$

the remainder term $r_6(t_m,\cdot)h^3$ in [103, Equation (3.14)] (apply $\mathcal{L}_N$ three times) corresponds to

$$\begin{aligned}
\mathbb{E}\big[R^6_{t_m}\big] &= \mathbb{E}\Big[\int_{t_m}^{t_m+h} \int_{t_m}^{q_1} \int_{t_m}^{q_2} (\mathcal{L}_N)^3 u(t_{m+1}, \boldsymbol{X}^{N,h}_{q_3}) \mathrm{d}q_3 \ \mathrm{d}q_2 \ \mathrm{d}q_1 \Big] \\
&= \mathbb{E}\Big[\int_{t_m}^{t_m+h} \int_{t_m}^{q_1} \int_{t_m}^{q_2} \\
&\qquad \Big(K \sum_{\gamma\in\Pi_3^N} B_{\gamma_1}(\boldsymbol{X}^{N,h}_{q_3}) B_{\gamma_2}(\boldsymbol{X}^{N,h}_{q_3}) B_{\gamma_3}(\boldsymbol{X}^{N,h}_{q_3}) \partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{q_3}) \\
&\qquad\quad + K \sum_{\gamma\in\Pi_3^N} \partial^2_{x_{\gamma_3},x_{\gamma_3}} \Big(B_{\gamma_1}(\boldsymbol{X}^{N,h}_{q_3}) B_{\gamma_2}(\boldsymbol{X}^{N,h}_{q_3}) \partial^2_{x_{\gamma_1},x_{\gamma_2}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{q_3})\Big) \\
&\qquad\quad + K \sum_{\gamma\in\Pi_3^N} \partial^4_{x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}} \Big(B_{\gamma_1}(\boldsymbol{X}^{N,h}_{q_3}) \partial_{x_{\gamma_1}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{q_3})\Big) \\
&\qquad\quad + K \sum_{\gamma\in\Pi_6^N, \hat{\mathcal{O}}(\gamma)\le 3} \partial^6_{x_{\gamma_1},\ldots,x_{\gamma_6}} u(t_{m+1}, \boldsymbol{X}^{N,h}_{q_3}) \Big) \mathrm{d}q_3 \ \mathrm{d}q_2 \ \mathrm{d}q_1 \Big].
\end{aligned}$$

Once again, a direct application of Hölder's inequality and Fubini's Theorem

$$\big|\mathbb{E}\big[R^6_{t_m}\big]\big| \le Kh^2 \int_{t_m}^{t_m+h}$$

$$\Bigg(\mathbb{E}\Bigg[\Bigg|\sum_{\gamma\in\Pi_3^N} B_{\gamma_1}(\boldsymbol{X}_{q_1}^{N,h})B_{\gamma_2}(\boldsymbol{X}_{q_1}^{N,h})B_{\gamma_3}(\boldsymbol{X}_{q_1}^{N,h})\partial^3_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3}}u(t_{m+1},\boldsymbol{X}_{q_1}^{N,h})$$

$$+\sum_{\gamma\in\Pi_3^N}\partial^2_{x_{\gamma_3},x_{\gamma_3}}\Big(B_{\gamma_1}(\boldsymbol{X}_{q_1}^{N,h})B_{\gamma_2}(\boldsymbol{X}_{q_1}^{N,h})\partial^2_{x_{\gamma_1},x_{\gamma_2}}u(t_{m+1},\boldsymbol{X}_{q_1}^{N,h})\Big)$$

$$+\sum_{\gamma\in\Pi_3^N}\partial^4_{x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}\Big(B_{\gamma_1}(\boldsymbol{X}_{q_1}^{N,h})\partial_{x_{\gamma_1}}u(t_{m+1},\boldsymbol{X}_{q_1}^{N,h})\Big)$$

$$+\sum_{\gamma\in\Pi_6^N,\hat{\mathcal{O}}(\gamma)\le 3}\partial^6_{x_{\gamma_1},\dots,x_{\gamma_6}}u(t_{m+1},\boldsymbol{X}_{q_1}^{N,h})\Bigg|^2\Bigg]\Bigg)^{1/2}\mathrm{d}q_1.$$

Part 2: Other remainder terms. We now present the dominations of the other residual terms, whose explicit expressions can be found in Appendix 6.4.4. These, once again, are a result of applying Hölder's inequality, Itô's isometry and Fubini's Theorem. We have

$$\big|\mathbb{E}\big[R^2_{t_m}\big]\big|\le Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_5^N}\partial^5_{x_{\gamma_1},\dots,x_{\gamma_5}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2}$$
$$+Kh^{5/2}\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_6^N}\partial^6_{x_{\gamma_1},\dots,x_{\gamma_6}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2},$$

$$\big|\mathbb{E}\big[R^3_{t_m}\big]\big|$$
$$\le Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_3^N}\partial_{x_{\gamma_1}}\Big(\partial_{x_{\gamma_3}}B_{\gamma_2}(\overline{\boldsymbol{X}}_{q_1}^{N,h})\partial^2_{x_{\gamma_2},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big)\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2}$$
$$+Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_3^N}\partial^2_{x_{\gamma_1},x_{\gamma_1}}\Big(\partial_{x_{\gamma_3}}B_{\gamma_2}(\overline{\boldsymbol{X}}_{q_1}^{N,h})\partial^2_{x_{\gamma_2},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big)\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2}$$
$$+Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_3^N}\partial_{x_{\gamma_1}}\Big(B_{\gamma_2}(\overline{\boldsymbol{X}}_{q_1}^{N,h})\partial^2_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big)\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2}$$
$$+Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_3^N}\partial^2_{x_{\gamma_1},x_{\gamma_1}}\Big(B_{\gamma_2}(\overline{\boldsymbol{X}}_{q_1}^{N,h})\partial^2_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big)\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2}$$
$$+Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_5^N}\partial^5_{x_{\gamma_1},\dots,x_{\gamma_5}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2}$$
$$+Kh^{5/2}\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_6^N}\partial^6_{x_{\gamma_1},\dots,x_{\gamma_6}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2},$$

$$\big|\mathbb{E}\big[R^5_{t_m}\big]\big|\le Kh^{5/2}\sum_{\gamma\in\Pi_4^N}\Big(\mathbb{E}\Big[\Big|\partial^4_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}}u(t_{m+1},\boldsymbol{X}_{t_m}^{N,h})\Big|^2\Big]\Big)^{1/2}$$
$$+Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_5^N}\partial^5_{x_{\gamma_1},\dots,x_{\gamma_5}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2}$$
$$+Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_5^N}\partial^6_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4},x_{\gamma_5}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2},$$

$$\big|\mathbb{E}\big[R^7_{t_m}\big]\big|\le Kh^2\Big(\int_{t_m}^{t_m+h}\mathbb{E}\Big[\Big|\sum_{\gamma\in\Pi_6^N,\hat{\mathcal{O}}(\gamma)\in\{4,5\}}\partial^6_{x_{\gamma_1},\dots,x_{\gamma_6}}u(t_{m+1},\overline{\boldsymbol{X}}_{q_1}^{N,h})\Big|^2\Big]\mathrm{d}q_1\Big)^{1/2}$$

$$+ Kh^2\Big(\int_{t_m}^{t_m+h} \mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_5^N, \hat{\mathcal{O}}(\gamma)=4} \partial^6_{x_{\gamma_1},\dots,x_{\gamma_6}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1}) \Big|^2 \Big] \mathrm{d}q_1 \Big)^{1/2},$$

$$\big|\mathbb{E}\big[R^8_{t_m}\big]\big| \le Kh^2\Big(\int_{t_m}^{t_m+h} \mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_5^N, \hat{\mathcal{O}}(\gamma)\le 3} \partial^5_{x_{\gamma_1},\dots,x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1}) \Big|^2 \Big] \mathrm{d}q_1 \Big)^{1/2}$$

$$+ Kh^2 \int_{t_m}^{t_m+h} \Big(\mathbb{E}\Big[\Big| \sum_{\gamma\in\Pi_6^N, \hat{\mathcal{O}}(\gamma)\le 3} \partial^6_{x_{\gamma_1},\dots,x_{\gamma_6}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1}) \Big|^2 \Big]\Big)^{1/2} \mathrm{d}q_1$$

$$+ Kh^2\Big(\int_{t_m}^{t_m+h} \mathbb{E}\Big[\Big| \sum_{\substack{\alpha,\beta\in\bigcup_{k=0}^5 \Pi_k^N,\ \hat{\mathcal{O}}(\alpha\bigcup\beta)=3 \\ |\alpha|+|\beta|=5,\ |\beta|\ge 2}} \partial^{|\alpha|}_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}} B_{\beta_1}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\ \partial^{|\beta|}_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}} u(t_{m+1}, \overline{\boldsymbol{X}}^{N,h}_{q_1})\Big) \Big|^2 \Big] \mathrm{d}q_1 \Big)^{1/2}.$$

Similar to (5.155) and the proof of Lemma 5.6.1, we can use the linear growth of B_j in combination with the assumption $\xi \in L^{10}(\Omega,\mathbb{R})$ and the Cauchy–Schwarz inequality to establish L^{10}-estimates of the B_j (see, Proposition 5.2.2 and Lemma 6.4.5) which, for instance, are employed in the estimation of $R^3_{t_m}, R^5_{t_m}, R^8_{t_m}$. An application of Lemma 5.5.3 allows to control the moments of derivatives of the solution, u, to the Kolmogorov backward equation. Hence, all the above expectations are well-defined and bounded by $K(h^{5/2}+h^3)e^{-\lambda_0(T-t_{m+1})}$.

Part 3: Collecting the estimates. For the eight residual terms $R^1_{t_m},\dots,R^8_{t_m}$, we have

$$\sum_{i=1}^{8} \big|\mathbb{E}\big[R^i_{t_m}\big]\big| \le K(h^{5/2}+h^3)e^{-\lambda_0(T-t_{m+1})}.$$

Combining all of the above estimates, we conclude

$$\sum_{m=0}^{M-1} \mathbb{E}\big[R(t_m, \boldsymbol{X}^{N,h}_{t_m})\big] \le \sum_{m=0}^{M-1}\sum_{i=1}^{8} \big|\mathbb{E}\big[R^i_{t_m}\big]\big| \le \sum_{m=0}^{M-1} K(h^{5/2}+h^3)e^{-\lambda_0(T-t_{m+1})} \le Kh^{3/2}.$$

□

5.7 Numerical illustration

We illustrate the performance of the non-Markovian Euler scheme (5.13) with a simple linear MV-SDE example. Consider the following mean-field equation:

$$\mathrm{d}X_t = \Big(\alpha\big(X_t - \mathbb{E}[X_t]\big) - X_t\Big)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0 \in L^{10}(\Omega,\mathbb{R}), \tag{5.172}$$

where $\alpha,\sigma > 0$. Explicit calculations yield $\mathbb{E}[X_t] = \mathbb{E}[X_0]e^{-t}$ and thus this process admits the following stationary distribution

$$\mu^*(x) = Z\exp\Big(-\frac{\alpha+1}{\sigma^2}x^2\Big), \tag{5.173}$$

where Z is the renormalization constant such that $\int_{\mathbb{R}} \mu^*(x)\mathrm{d}x = 1$. We use the following recipe to compute errors in the numerical experiments.

1. Choose a large enough domain $[a,b]$ such that $\big(F_{\mathrm{cdf}}(b) - F_{\mathrm{cdf}}(a)\big) > 1 - 10^{-6}$, where F_{cdf} denotes the cumulative density function (CDF) of the invariant distribution μ of (5.173).
2. Split the domain $[a,b]$ into N_{bins} equally spaced bins and compute the true density in each bin. These values are denoted $(\mu_i^{\mathrm{true}})_{i\in\{1,\dots,N_{\mathrm{bins}}\}}$ and obtained using numerical integration. Additionally, the first bin and the value μ_1^{true} takes the interval $(-\infty,a]$ into

account while the last bin (and the value $\mu_{N_{\text{bins}}}^{\text{true}}$) takes the interval $[b, \infty)$. In this way, one has $\sum_{i=1}^{N_{\text{bins}}} \mu_i^{\text{true}} = 1$.

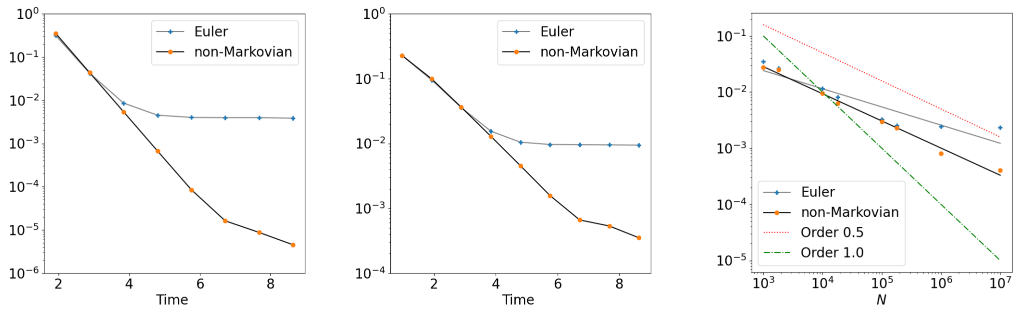
3. We simulate N particles up to time T using different time-stepping schemes and compute an approximation of the density (using a histogram approach) denoted by $(\mu_i^{\text{proxy}})_{i \in \{1,\dots,N_{\text{bins}}\}}$ based on the simulated paths.

4. As in [103], we compute the relative entropy error and the L_2-Error as

$$\begin{aligned} \text{Relative Entropy Error} &= \sum_{i=1}^{N_{\text{bins}}} \mu_i^{\text{true}} \ln \left(\frac{\mu_i^{\text{true}}}{\mu_i^{\text{proxy}}} \right) \\ L_2\text{-Error} &= \sqrt{\sum_{i=1}^{N_{\text{bins}}} |\mu_i^{\text{true}} - \mu_i^{\text{proxy}}|^2}. \end{aligned}$$

5. We also show a PoC result by computing the L_2-Error for different sizes N of particles with fixed time T and timestep h.

The goal of this simple example is to simulate the IPS associated with (5.172) up to $T = 9$ using the classical Euler method and the non-Markovian Euler method, respectively, and compare the results to (5.173). Following the recipe above, we set the domain as $[a, b] = [-1.8, 1.8]$ and split it into 72 bins.



(a) Rel. Entropy Error (b) L_2-Error (c) PoC (log-)L_2-Error, $T = 9$

Figure 5.1: Simulation of the linear MV-SDE (5.172) with $\alpha = 0.5, \sigma = 0.8$, $N = 10^7, h = 0.16$, and $X_0 \sim \mathcal{N}(\pi, 1)$ (a normal distribution with mean value π and variance 1). Both schemes run on the exact same samples of the initial condition and Brownian increments. (a) Entropy Error of the Euler method and non-Markovian method in log-scale over time. (b) L_2-Error of the Euler method and non-Markovian method in log-scale over time. (c) L_2-Error in particle size N of the Euler method and non-Markovian method in log-scale with respect to different number of particles N at $T = 9$.

Figure 5.1 (a) and (b) show that the non-Markovian method (uniformly) outperforms the Euler method in the approximation of the stationary distribution μ in both error metrics and the difference in performance becomes more evident as time increases. This is expected from the result in Theorem 5.3.8, since the first order term decays exponentially as T increases.

Figure 5.1 (c) shows (for a fixed timestep $h = 0.04$) that the PoC L_2-Error of the non-Markovian method decays consistently as N increases with a rate of approximately 0.5 (the expected strong PoC $\mathcal{O}(1/\sqrt{N})$ rate). The error of the Euler method plateaus for $N > 10^5$ as the error from the time-discretization dominates the particle error; see also Table 5.2 below for more information.

Table 5.1 shows that the non-Markovian method has a significantly better approximation accuracy compared to the Euler method under different choices of timesteps and model parameters. The Euler method produces larger errors as the timestep increases and the non-Markovian method yields stable results across all choices for the timestep.

The results in Table 5.2 show (at fixed timestep $h = 0.04$) the entropy error and the L_2-Error of the non-Markovian method decaying as the number of particles N increases. However,

α	σ	a	b	N_{bins}	h	Entropy Error		L_2-error	
						Euler	NM	Euler	NM
0.5	0.8	-1.8	1.8	72	0.04	2.33E-04	4.71E-06	2.37E-03	3.56E-04
					0.16	3.84E-03	4.33E-06	9.47E-03	3.37E-04
					0.24	9.26E-03	4.40E-06	1.47E-02	3.25E-04
					0.48	4.31E-02	3.25E-06	3.18E-02	2.92E-04
0.3	1.5	-3.0	3.0	120	0.04	1.84E-04	7.85E-06	1.44E-03	3.81E-04
					0.16	2.98E-03	5.94E-06	5.88E-03	2.82E-04
					0.24	6.84E-03	6.07E-06	9.00E-03	3.19E-04
					0.48	3.08E-02	5.72E-06	1.95E-02	3.09E-04

Table 5.1: Simulation results for MV-SDE (5.172) with $N = 10^7, T = 8.64$ and different choices of parameters using the non-Markovian (NM) Euler and standard Euler method, respectively. (As for Fig. 5.1: $X_0 \sim \mathcal{N}(\pi, 1)$ and both schemes run on the exact same samples of the initial condition and Brownian increments.)

the error of the Euler method remains stable for $N > 10^5$ (i.e., there is a plateau). Due to computational limitations, we are not able to show results beyond $N > 10^7$. The terminal time $T = 8.64$ is chosen for convenience only (due to the smaller timestep $h = 0.04$).

α	σ	a	b	N_{bins}	N	Entropy Error		L_2-Error	
						Euler	NM	Euler	NM
0.5	0.8	-1.8	1.8	72	10^3	-	-	2.89E-02	3.28E-02
					10^4	-	-	1.01E-02	1.04E-02
					10^5	8.21E-04	4.83E-04	4.29E-03	3.10E-03
					10^6	2.74E-04	4.66E-05	2.31E-03	1.26E-03
					10^7	2.33E-04	4.71E-06	2.37E-03	3.56E-04

Table 5.2: Simulation results for MV-SDE (5.172) with $h = 0.04$ and $T = 8.64$ for increasing numbers of particles N. (As for Fig. 5.1: $X_0 \sim \mathcal{N}(\pi, 1)$ and both schemes run on the exact same samples of the initial condition and Brownian increments.)

Chapter 6

Appendix

6.1 Auxiliary results of chapter 2

We provide a brief review of Taming [60] and Adaptive time-stepping [125] for superlinear growth MV-SDEs in the context of the notation set in Section 2.1 and 2.2. Each method approximates (2.1) through the interacting particle system (2.2) as described next. Table 6.1 summarises strong error (rMSE), mentions weak error as an open problem, and stability results. Both schemes (plus the proposed SSM one) hold under the same conditions: Assumption 2.2.1 and a sufficiently high integrable initial condition (as in Theorem 2.2.6).

Methods	rMSE	Stability
Taming [60]	0.5	Unknown
Adaptive [125]	0.5	Unknown
SSM	0.5	Contraction Theorem 2.2.9

Table 6.1: Information regarding the different methods. Convergence of Taming [60] and Adaptive [125] scheme under the same conditions as Assumption 2.2.1. The Root MSE (rMSE) is the metric presented in (2.7). Weak error analysis has not been carried out but experimental work points in the direction of weak error order 1 for the three methods.

6.1.1 The Taming method

Taming [60] approximates (2.2) as follows (see also [125, Section 4])

$$\bar{X}_{n+1}^{i,N,M} = \bar{X}_n^{i,N,M} + \frac{\widehat{b}(t_n, \bar{X}_n^{i,N,M}, \bar{\mu}_n^{X,N})}{1 + M^{-\alpha}\big|\widehat{b}(t_n, \bar{X}_n^{i,N,M}, \bar{\mu}_n^{X,N})\big|} h + \sigma(t_n, \bar{X}_n^{i,N,M}, \bar{\mu}_n^{X,N})\Delta W_n^i, \quad i \in [\![1, N]\!]. \tag{6.1}$$

where $\bar{\mu}_n^{X,N}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_n^{j,N,M}}(\mathrm{d}x)$, $\Delta W_n^i = W_{t_{n+1}}^i - W_{t_n}^i$ with $\bar{X}_0^{i,N,M} = X_0^i$. The parameter $\alpha \in (0, 1]$ is a tuning parameter where setting $\alpha = 1/2$ delivers a rMSE convergence rate of order $1/2$ while setting $\alpha = 1$ delivers a rMSE convergence rate of order 1 (for a constant diffusion σ).

6.1.2 Adaptive time-stepping method

Adaptive [125] approximates (2.2) as follows for $t_n \in [k_n h, (k_n + 1)h), k_n \in \mathbb{N}$ and

$$\bar{X}_{t_{n+1}}^{i,N} = \bar{X}_{t_n}^{i,N} + \widehat{b}(t_n, \bar{X}_{t_n}^{i,N}, \bar{\mu}_{k_n h}^{X,N})h_n^i + \sigma(t_n, \bar{X}_{t_n}^{i,N}, \bar{\mu}_{k_n h}^{X,N})\Delta W_{t_n}^i, \qquad i \in [\![1, N]\!]. \tag{6.2}$$

where $\bar{\mu}^{X,N}_{k_n h}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}^{j,N,M}_{k_n h}}(\mathrm{d}x)$, $t_{n+1} = t_n + h^i_n$, $\Delta W^i_{t_n} = W^i_{t_{n+1}} - W^i_{t_n}$ with $\bar{X}^{i,N,M}_0 = X^i_0$ and for a map $\mathbf{h}^\delta(x) : \mathbb{R}^d \to [0,h]$

$$h^i_n = \min\Big\{\mathbf{h}^\delta(\bar{X}^{i,N}_{t_n}), (k_n+1)h - t_n\Big\}.$$

The function $\mathbf{h}^\delta$ is specified at each example and is to be understood similarly to the taming technique. In essence, $\mathbf{h}^\delta$ is to be chosen such that $|\widehat{b}(x)\mathbf{h}^\delta(x)|$ is of linear growth. For Adaptive, one modifies the timestep h in a dynamic fashion to control the growth of $\widehat{b}$ while taming modifies the drift $\widehat{b}$ to control the growth across the application of the scheme. The rMSE convergence rate of order 1/2, see [69] or [125].

6.2 Auxiliary results of chapter 3

6.2.1 Well-posedness of the particle system and the PoC in Proposition 3.2.5

The Propagation of chaos result (3.8) follows directly from [2, 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 3.2.5. We start by interpreting the interacting particle system (3.3) as a single SDE in $\mathbb{R}^{Nd}$. In Remark 3.2.4 we show that, as a system in $\mathbb{R}^{Nd}$, the function V (see (3.5) and (3.2)) 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 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 [113, Theorem 3.5 (p.58)].

The moment bound result of the $\mathbb{R}^{Nd}$-system that follows from [113, Theorem 3.5 (p.58)] does not lead to (3.7) 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 (3.7).

The strategy of the proof is the same as that in Section 3.4.3. For all $m \geq 2p \geq 2$, $i \in [\![1,N]\!]$, $t \in [0,T]$, we have

$$\begin{aligned}
\mathbb{E}\big[|X^{i,N}_t|^{2p}\big] &= \mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N \big(X^{i,N}_t - X^{j,N}_t\big) + \frac{1}{N}\sum_{j=1}^N X^{j,N}_t\Big|^{2p}\Big] \\
&\leq 4^p\mathbb{E}\Big[\frac{1}{N}\sum_{j=1}^N |X^{i,N}_t - X^{j,N}_t|^{2p}\Big] + 4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N |X^{j,N}_t|^2\Big|^p\Big] \\
&\leq 4^p\mathbb{E}\Big[|X^{i,N}_t - X^{j,N}_t|^{2p}\Big]_{i\neq j} + 4^p\mathbb{E}\Big[\Big|\frac{1}{N}\sum_{j=1}^N |X^{j,N}_t|^2\Big|^p\Big]. \qquad (6.3)
\end{aligned}$$

For the first term in (6.3), by Itô's formula, for $i, j \in [\![1,N]\!], i \neq j$,

$$\begin{aligned}
&|X^{i,N}_t - X^{j,N}_t|^{2p} = |X^{i,N}_0 - X^{j,N}_0|^{2p} \\
&+ 2p\int_0^t |X^{i,N}_s - X^{j,N}_s|^{2p-2}\Big\langle X^{i,N}_s - X^{j,N}_s, v(X^{i,N}_s, \mu^{X,N}_s) - v(X^{j,N}_s, \mu^{X,N}_s)\Big\rangle \mathrm{d}s \\
&+ 2p\int_0^t |X^{i,N}_s - X^{j,N}_s|^{2p-2}\Big\langle X^{i,N}_s - X^{j,N}_s, b(s, X^{i,N}_s, \mu^{X,N}_s) - b(s, X^{j,N}_s, \mu^{X,N}_s)\Big\rangle \mathrm{d}s \\
&+ 2p\int_0^t |X^{i,N}_s - X^{j,N}_s|^{2p-2}\Big\langle X^{i,N}_s - X^{j,N}_s, \sigma(s, X^{i,N}_s, \mu^{X,N}_s)\mathrm{d}W^i_s - \sigma(s, X^{j,N}_s, \mu^{X,N}_s)\mathrm{d}W^j_s\Big\rangle
\end{aligned}$$

$$+\frac{2p(2p-1)}{2}\int_0^t |X_s^{i,N}-X_s^{j,N}|^{2p-2}\Big(|\sigma(s,X_s^{i,N},\mu_s^{X,N})|^2+|\sigma(s,X_s^{j,N},\mu_s^{X,N})|^2\Big)\mathrm{d}s.$$

By Assumption 3.2.1, Remark 3.2.2, Jensen's inequality, Proposition 3.4.6, take expectation on both side, by the particles are identically distributed and Burkholder-Davis-Gundy (BDG) inequality, we have

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

For the second term in (6.3), similarly, and notice that,

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

Take power of p on both side and expectations. By Jensen's inequality, BDG inequality, Proposition 3.4.6, Assumption 3.2.1, the Lipschitz properties on σ, we can conclude with the highest order up to $2p$, we have

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

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 (6.3) again, we have

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

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

6.2.2 Solving the implicit equation of the SSM and a deployment of Newton's method

In this section we address solving the implicit Equation (3.9) in the SSM. We first present a general result stating the level of precision on needs to solve (3.9) such that the final convergence rate of the SSM method is preserved (e.g., Theorem 3.2.9 and 3.2.11). Proposition 6.2.2 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 3.3 showing its efficiency.

Approximation scheme to the SSM

Recall the SSM from Definition 3.2.6. 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 (3.9) to its data $\hat{X}_n^{i,N}$, $\hat{X}_n^N$ and h, i.e.,

$$\hat{\Psi}_i(\hat{X}_n^{i,N}, \hat{X}_n^N, h) = Y_n^{i,\star,N}, \quad \hat{\Psi} = (\hat{\Psi}_1, \dots, \hat{\Psi}_N). \tag{6.4}$$

The existence of such a map $\hat{\Psi}$ is guaranteed by Lemma 3.4.2 (see also Proposition 3.4.3 and 3.4.4 for some of its good properties). We next introduce a version SSM of Definition 3.2.6 where the implicit equation is solved approximately only.

Definition 6.2.1 (Approximation scheme to the SSM)**.** *We follow the notation of Definition 3.2.6 hold. Denote the approximation mapping at each SSM step* (3.9) *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* (3.9)*-*(3.10)*: set* $\overline{X}_0^{i,N} = X_0^i$ *for* $i \in [\![1, N]\!]$*; then for all* $i \in [\![1, N]\!]$ *and* $n \in [\![0, M-1]\!]$

$$\overline{Y}_n^{i,\star,N} = \overline{\Psi}_i(\overline{X}_n^{i,N}, \overline{X}_n^N, h),\ \overline{X}_n^N = (\overline{X}_n^{1,N}, \dots, \overline{X}_n^{N,N}),\ \overline{\mu}_n^{Y,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{\overline{Y}_n^{j,\star,N}}(\mathrm{d}x), \tag{6.5}$$

$$\overline{X}_{n+1}^{i,N} = \overline{Y}_n^{i,\star,N} + b(t_n, \overline{Y}_n^{i,\star,N}, \overline{\mu}_n^{Y,N})h + \sigma(t_n, \overline{Y}_n^{i,\star,N}, \overline{\mu}_n^{Y,N})\Delta W_n^i,\ \Delta W_n^i = W_{t_{n+1}}^i - W_{t_n}^i, \tag{6.6}$$

where for any i *the map* $\overline{\Psi}_i$ *is an approximation to* $\hat{\Psi}_i$ *solving* (6.4)*.*

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 6.2.2. *Let the assumptions of Theorem 3.2.10 hold. Recall the notation of Definition 3.2.6 and* (6.2.1)*. For the* $\hat{\Psi}_i$ *and* $\overline{\Psi}_i$ *defined in* (6.4) *and* (6.5) *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*

$$\sup_{n \in [\![1,M]\!]} \sup_{i \in [\![1,N]\!]} \mathbb{E}\big[|\hat{X}_n^{i,N} - \overline{X}_n^{i,N}|^2\big] \le Ch. \tag{6.7}$$

The main interpretation is that as long as the implicit Equation (3.9) 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.

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 6.2.2 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 (3.54),

$$\begin{aligned}\sup_{i \in [\![1,N]\!]} \mathbb{E}\big[|\hat{X}_1^{i,N} - \overline{X}_1^{i,N}|^2\big] &\le C \sup_{i \in [\![1,N]\!]} \mathbb{E}\Big[|Y_0^{i,\star,N} - \overline{Y}_1^{i,\star,N}|^2 + \big|W^{(2)}(\overline{\mu}_0^{Y,N}, \hat{\mu}_0^{Y,N})\big|^2 h\Big] \\ &\le \sup_{i \in [\![1,N]\!]} \mathbb{E}\big[|Y_0^{i,\star,N} - \overline{Y}_1^{i,\star,N}|^2\big] \le Ch. \end{aligned} \tag{6.8}$$

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 6.2.2,

$$
\begin{aligned}
&\sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|Y_n^{i,\star,N} - \overline{Y}_n^{i,\star,N}|^2\big] = \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{\Psi}_i(\hat{X}_n^i, \hat{X}_n, h) - \overline{\Psi}(\overline{X}_n^i, \overline{X}_n, h)|^2\big] \\
&\le 2 \sup_{i\in[\![1,N]\!]} \Big(\mathbb{E}\big[|\hat{\Psi}_i(\hat{X}_n^i, \hat{X}_n, h) - \hat{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h)|^2\big] + \mathbb{E}\big[|\hat{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h) - \overline{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h)|^2\big]\Big) \\
&\le 2 \sup_{i\in[\![1,N]\!]} \mathbb{E}\big[|\hat{\Psi}_i(\hat{X}_n^i, \hat{X}_n, h) - \hat{\Psi}_i(\overline{X}_n^i, \overline{X}_n, h)|^2\big] + 2h. \qquad (6.9)
\end{aligned}
$$

Recall the results in Section 3.4.2, the arguments in (3.41) 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 (6.9) 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 (6.8), we have

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

□

Deploying Newton's method

We now provide a discussion on using Newton's method to solve (3.9) in the scope of the SSM. We first introduce Newton's method for high dimensions. Recall the functions V, u, f in (3.2), (3.5), and the SSM in Definition 3.2.6.

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 (3.9)) defined as

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

For a fixed $x \in \mathbb{R}^{Nd}$, Lemma 3.4.2 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

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

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

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

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 3.4.2 ensures a unique $y^\star$ exists solving $F(y^\star) = 0$. Proposition 3.4.3 and 3.4.4 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 [131, 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 3.2.7 one can show that V in (3.5) has a strictly negative one-sided Lipschitz constant and hence ∇V is strict negative definite matrix (see [107]) and hence so is ∇F – this ensures that ∇F is nonsingular (also at $y^\star$) and thus [131, Theorem 4.4] applies guaranteeing convergence.

In the scope of the examples presented in Section 3.3, 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 $\kappa \le 4$.

6.3 Auxiliary results of chapter 4

We provide the following results of the convolution drift term after integration.

Lemma 6.3.1. *Let $(\mathbf{A}^f, \mathbf{A}^{f_\sigma})$ in Assumption 4.2.1 hold. Then it holds for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $m > 2$*

$$\begin{aligned}\int_{\mathbb{R}^d} &\big(\langle x, (f*\mu)(x)\rangle + (m-1)|(f_\sigma*\mu)(x)|^2\big)\mu(\mathrm{d}x)\\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \big(\langle x, f(x-y)\rangle + (m-1)|f_\sigma(x-y)|^2\big)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &\le L^{(1)}_{(f)}\big(\mu(|\cdot|^2) - |\mu(id)|^2\big) = L^{(1)}_{(f)}\, Var_\mu,\end{aligned}$$

where $\mu(|\cdot|^2) := \int_{\mathbb{R}^d}|x|^2\mu(\mathrm{d}x)$, $\mu(id) := \int_{\mathbb{R}^d} x\mu(\mathrm{d}x)$ and $Var_\mu = \mu(|\cdot|^2) - |\mu(id)|^2$.

Proof. Using $f(0) = f_\sigma(0) = 0$ and that f is an odd function we have

$$\begin{aligned}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} &\big(\langle x, f(x-y)\rangle + (m-1)|f_\sigma(x-y)|^2\big)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{1}{2}\big(\langle x-y, f(x-y)\rangle + 2(m-1)|f_\sigma(x-y)|^2\big)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &\le \frac{1}{2}L^{(1)}_{(f)}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|x-y|^2\mu(\mathrm{d}x)\mu(\mathrm{d}y) = \frac{1}{2}L^{(1)}_{(f)}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\big(|x|^2 - 2\langle x, y\rangle + |y|^2\big)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &= \frac{1}{2}L^{(1)}_{(f)}\Big(2\mu(|\cdot|^2) - 2\int_{\mathbb{R}^d}x\mu(\mathrm{d}x)\int_{\mathbb{R}^d}y\mu(\mathrm{d}y)\Big)\\ &= L^{(1)}_{(f)}\big(\mu(|\cdot|^2) - \big|\int_{\mathbb{R}^d}x\mu(\mathrm{d}x)\big|^2\big) = L^{(1)}_{(f)}\mathrm{Var}_\mu,\end{aligned}$$

where for the inequality we used the monotonicity condition on the convolution kernels and the symmetry of the double integration in μ. ☐

Lemma 6.3.2. *Let f and f_σ satisfy conditions $(\mathbf{A}^f,\ \mathbf{A}^{f_\sigma})$ of Assumption 4.2.1. Set $L^{(1),+}_{(f)} = \max\{0, L^{(1)}_{(f)}\}$. Then, we have*

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big(\langle x-y, (f*\mu)(x) - (f*\nu\ )(y)\rangle + (m-1)\big|(f_\sigma*\mu)(x) - (f_\sigma*\nu)(y)\big|^2\Big)\mu(\mathrm{d}x)\nu(\mathrm{d}y)$$

$$\le 2L^{(1),+}_{(f)} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|^2 \mu(\mathrm{d}x)\nu(\mathrm{d}y).$$

Proof. For any $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$, we compute

$$\begin{aligned}
&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \langle x-y, (f*\mu)(x)-(f*\nu)(y)\rangle \mu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&=\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \langle x-y, f(x-x')-f(y-y')\rangle \mu(\mathrm{d}x')\nu(\mathrm{d}y')\mu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&=\frac{1}{2}\Bigg[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \langle x-y, f(x-x')-f(y-y')\rangle \mu(\mathrm{d}x')\nu(\mathrm{d}y')\mu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&\qquad -\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \langle x'-y', f(x-x')-f(y-y')\rangle \mu(\mathrm{d}x)\nu(\mathrm{d}y)\mu(\mathrm{d}x')\nu(\mathrm{d}y')\Bigg]\\
&=\frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \langle (x-x')-(y-y'), f(x-x')-f(y-y')\rangle \mu(\mathrm{d}x)\nu(\mathrm{d}y)\mu(\mathrm{d}x')\nu(\mathrm{d}y'),
\end{aligned}$$

and thus,

$$\begin{aligned}
&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\langle x-y, (f*\mu)(x)-(f*\nu)(y)\rangle + (m-1)\big|(f_\sigma*\mu)(x)-(f_\sigma*\nu)(y)\big|^2\Big)\mu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&=\frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\langle (x-x')-(y-y'), f(x-x')-f(y-y')\rangle\\
&\qquad\qquad + 2(m-1)|f_\sigma(x-x')-f_\sigma(y-y')|^2\Big)\mu(\mathrm{d}x')\nu(\mathrm{d}y')\mu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&\le \frac{1}{2}L^{(1)}_{(f)}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \big|(x-x')-(y-y')\big|^2 \mu(\mathrm{d}x')\nu(\mathrm{d}y')\mu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&\le 2L^{(1),+}_{(f)}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|^2\mu(\mathrm{d}x)\nu(\mathrm{d}y).
\end{aligned}$$

□

6.4 Auxiliary results of chapter 5

6.4.1 Proof of Proposition 5.2.2

Proof of Proposition 5.2.2. Assumption 5.2.1 is sufficient to guarantee the existence of the unique stationary distribution; see [40] (under their Assumption (A')). The uniform PoC result follows from [109, Theorem 1.2]. This addresses the proposition's last two statements.

The system's wellposedness (as an SDE in $\mathbb{R}^N$) follows [109]. For completeness, we present a short proof for the moment stability of the IPS, highlighting that the constant K appearing in the RHS of the inequality is independent of t and N.

Let $p \ge 2$. Performing similar calculations as in Section 6.2.1 and applying Gronwall's inequality, we deduce that there exists some positive constant K independent of $N, t \ge 0$, such that for any $i \neq j$

$$\mathbb{E}[\,|X^{i,N}_t - X^{j,N}_t|^p] \le K < \infty. \tag{6.10}$$

Employing Itô's formula, we deduce for any $t \ge 0$, $p \ge 2$ and $\kappa \in [0,\lambda)$

$$\begin{aligned}
&e^{p\kappa t}|X^{i,N}_t|^p - |\xi^i|^p\\
&\le \int_0^t p\Big(\kappa|X^{i,N}_s|^2 + \big(X^{i,N}_s\big)\cdot\Big(-\nabla U(X^{i,N}_s) - \frac{1}{N}\sum_{j=1}^N \nabla V(X^{i,N}_s - X^{j,N}_s)\Big)\Big)e^{p\kappa s}|X^{i,N}_s|^{p-2}\mathrm{d}s\\
&\quad + \sigma^2\frac{p(p-1)}{2}\int_0^t e^{p\kappa s}|X^{i,N}_s|^{p-2}\mathrm{d}s + \sigma p\int_0^t e^{p\kappa s}|X^{i,N}_s|^{p-2}X^{i,N}_s\mathrm{d}W^i_s.
\end{aligned}$$

Taking expectations on both sides, using the inequality: for $q \in \{1,2\}$

$$a^{p-q}b^q \leq \frac{p-q}{p}\varepsilon a^p + \frac{q}{p\varepsilon^{(p-q)/q}}b^p, \qquad \text{for any } \varepsilon > 0 \text{ and } a,b>0,$$

assumption 5.2.1 and (6.10) yields

$$\begin{aligned}
&\mathbb{E}\Big[e^{p\kappa t}|X_t^{i,N}|^p\Big] - \mathbb{E}\left[|\xi^i|^p\right] \\
&\leq p\int_0^t (\kappa-\lambda)\mathbb{E}\left[e^{p\kappa s}|X_s^{i,N}|^p\right]\mathrm{d}s \\
&\qquad + \int_0^t e^{p\kappa s}\Big(\frac{K}{N}\sum_{j=1}^N \mathbb{E}\Big[|X_s^{i,N}|^{p-1}\,|X_s^{i,N}-X_s^{j,N}|\Big]\Big)\mathrm{d}s + K\int_0^t e^{p\kappa s}\mathbb{E}\left[|X_s^{i,N}|^{p-2}\right]\mathrm{d}s \\
&\leq p\int_0^t (\kappa-\lambda+\varepsilon)\mathbb{E}\left[e^{\kappa p s}|X_s^{i,N}|^p\right]\mathrm{d}s + K\int_0^t e^{p\kappa s}\mathrm{d}s \leq Ke^{p\kappa t},
\end{aligned}$$

for some positive constant K and $\varepsilon > 0$ arbitrarily small. □

We provide the following auxiliary results:

Proposition 6.4.1. *Let the assumptions and set up of Proposition 5.2.2 hold and let the IPS be given as in* (5.11)*. Then there exists $K \geq 0$, independent of t, s and N, such that for any $s \geq t \geq 0,\ s-t<1$*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\Big[\big|X_s^{i,N}-X_t^{i,N}\big|^2\Big] \leq K(s-t).$$

Proof. By Itô's formula and an application of the Cauchy–Schwarz and Jensen inequalities, we have that for any $s \geq t \geq 0,\ (s-t)<1,\ i \in \{1,\dots,N\}$,

$$\begin{aligned}
\mathbb{E}\Big[\big|X_s^{i,N}-X_t^{i,N}\big|^2\Big] &= \mathbb{E}\Big[\Big|\int_t^s -\nabla U(X_u^{i,N}) - \frac{1}{N}\sum_{j=1}^N \nabla V(X_u^{i,N}-X_u^{j,N})\mathrm{d}u + \sigma\int_t^s \mathrm{d}W_u^i\Big|^2\Big] \\
&\leq K(s-t)\int_t^s \mathbb{E}\Big[\big|\nabla U(X_u^{i,N})\big|^2\Big] + \frac{1}{N}\sum_{j=1}^N \mathbb{E}\Big[\big|\nabla V(X_u^{i,N}-X_u^{j,N})\big|^2\Big]\mathrm{d}u + K(s-t) \\
&\leq K(s-t)\Big(1+\int_t^s \mathbb{E}\Big[\big|X_u^{i,N}\big|^2\Big] + \frac{1}{N}\sum_{j=1}^N \mathbb{E}\Big[\big|X_u^{j,N}\big|^2\Big]\mathrm{d}u\Big) \leq K(s-t),
\end{aligned}$$

where we used Proposition 5.2.2 in the last estimate. □

6.4.2 Proof of Proposition 5.2.3

For clarity, we prove each of the proposition's statements separately. The main argument requires care as an analysis by cases is needed, but within each case the estimation procedure is standard.

Proof of Proposition 5.2.3 – Statement (1): Moment estimates. For the scheme defined in (5.13), we introduce the following notations for any $m \in \{0,\dots,M\}$,

$$b_m^{i,N,h} := -\nabla U(X_{t_m}^{i,N,h}) - \frac{1}{N}\sum_{j=1}^N \nabla V(X_{t_m}^{i,N,h}-X_{t_m}^{j,N,h}), \quad \Delta\overline{W}_m^{i,N,h} := \frac{\Delta W_m^i + \Delta W_{m+1}^i}{2}. \tag{6.11}$$

Assumption 5.2.1 implies that

$$\big(X_{t_m}^{i,N,h}-X_{t_m}^{j,N,h}\big)\cdot\big(\,b_m^{i,N,h}-b_m^{j,N,h}\big) \leq -\lambda\big|X_{t_m}^{i,N,h}-X_{t_m}^{j,N,h}\big|^2,$$

$$
\begin{aligned}
\big|\, b_m^{i,N,h} - b_m^{j,N,h}\big|^2 &\le 2\big|\nabla U(X_{t_m}^{i,N,h}) - \nabla U(X_{t_m}^{j,N,h})\big|^2 \\
&\quad + \frac{2}{N}\sum_{k=1}^{N}\big|\nabla V(X_{t_m}^{i,N,h} - X_{t_m}^{k,N,h}) - \nabla V(X_{t_m}^{j,N,h} - X_{t_m}^{k,N,h})\big|^2 \\
&\le 2(\lambda^2 + K_V^2)|X_{t_m}^{i,N,h} - X_{t_m}^{j,N,h}|^2 \le K|X_{t_m}^{i,N,h} - X_{t_m}^{j,N,h}|^2.
\end{aligned}
$$

An inspection of the above inequalities indicates that in order to establish the L^p-moments for $X_{t_m}^{i,N,h}$, we need L^p-estimates on the local differences $X_{t_m}^{i,N,h} - X_{t_m}^{j,N,h}$ (a term that appears from the interaction kernel). This proof is split accordingly; *Part 1* deals with the latter while *Part 2* with the former.

Part 1: Moments of local differences uniformly bounded in time. We first prove that for all $i,j \in \{1,\dots,N\}$, $m \in \{0,\dots,M\}, p \ge 2$, with p an even number, that $\mathbb{E}\big[|X_{t_m}^{i,N,h} - X_{t_m}^{j,N,h}|^p\big]$ is uniformly bounded in time. Note that due to the nature of the scheme, $X_{t_m}^{i,N,h}$ is not independent of $\Delta\overline{W}_m^{i,N,h}$, and thus we analyze the different cases as time evolves, i.e., $m \in \{0,1,2\}$ and $m \ge 3$ below. (This same procedure will be used in *Part 2* of the proof.)

Case: $m = 0$. For any $i,j \in \{1,\dots,N\}$, we have

$$
\mathbb{E}\big[|X_{t_0}^{i,N,h} - X_{t_0}^{j,N,h}|^p\big] \le K\mathbb{E}\big[|X_{t_0}^{i,N,h}|^p + |X_{t_0}^{j,N,h}|^p\big] = K\mathbb{E}\big[|\xi|^p\big]. \tag{6.12}
$$

Case: $m = 1$. We get

$$
\begin{aligned}
|X_{t_1}^{i,N,h} - X_{t_1}^{j,N,h}|^2 &= |X_{t_0}^{i,N,h} + b_0^{i,N,h}h - X_{t_0}^{j,N,h} - b_0^{j,N,h}h|^2 + |\sigma\Delta\overline{W}_0^{i,N,h} - \sigma\Delta\overline{W}_0^{j,N,h}|^2 \\
&\quad + 2\big(X_{t_0}^{i,N,h} + b_0^{i,N,h}h - X_{t_0}^{j,N,h} - b_0^{j,N,h}h\big)\cdot\big(\sigma\Delta\overline{W}_0^{i,N,h} - \sigma\Delta\overline{W}_0^{j,N,h}\big) \\
&\le |X_{t_0}^{i,N,h} - X_{t_0}^{j,N,h}|^2\big(1 - 2\lambda h + Kh^2\big) + |\sigma\Delta\overline{W}_0^{i,N,h} - \sigma\Delta\overline{W}_0^{j,N,h}|^2 \\
&\quad + 2\big(X_{t_0}^{i,N,h} + b_0^{i,N,h}h - X_{t_0}^{j,N,h} + b_0^{j,N,h}h\big)\cdot\big(\sigma\Delta\overline{W}_0^{i,N,h} - \sigma\Delta\overline{W}_0^{j,N,h}\big).
\end{aligned}
$$

Taking the power $p/2$ and expectation on both sides, we deduce that there exist positive constants $K_{p,1}$, $K_{p,2}$, K, κ (both are independent of h,T,M and N) such that

$$
\begin{aligned}
\mathbb{E}\big[|X_{t_1}^{i,N,h} - X_{t_1}^{j,N,h}|^p\big] &\le \mathbb{E}\big[|X_{t_0}^{i,N,h} - X_{t_0}^{j,N,h}|^p\big]\big(1 - 2\lambda h + Kh^2\big)^p \\
&\quad + h\big(K_{p,1}\mathbb{E}\big[|X_{t_0}^{i,N,h} - X_{t_0}^{j,N,h}|^p\big] + K_{p,2}\big) \\
&\quad + p\mathbb{E}\big[|\sigma\Delta\overline{W}_0^{i,N,h} - \sigma\Delta\overline{W}_0^{j,N,h}|^2\cdot|X_{t_0}^{i,N,h} - X_{t_0}^{j,N,h}|^{p-2}\big(1 - 2\lambda h + Kh^2\big)^{p-2}\big] \\
&\le (1 - K_{p,1}h)\mathbb{E}\big[|X_{t_0}^{i,N,h} - X_{t_0}^{j,N,h}|^p\big] + K_{p,2}h\big(1 + \mathbb{E}\big[|\xi|^{p-2}\big]\big) \\
&\le K\big(h + \mathbb{E}\big[|\xi|^p\big] + \mathbb{E}\big[|\xi|^{p-2}\big]\big) \le Ke^{\kappa t_1}\big(1 + \mathbb{E}\big[|\xi|^p\big]e^{-\kappa t_1}\big),
\end{aligned}
$$

where we also used Young's inequality. Note that in this case (and the case $m = 2$) the factor $e^{-\kappa t_1}$ ($e^{-\kappa t_2}$ for $m = 2$) is only added to make it consistent with the estimates obtained for a general m.

Case: $m = 2$. Based on the calculations above, there exist positive constants K, κ (both are independent of h,T,M and N) such that

$$
\mathbb{E}\big[|X_{t_2}^{i,N,h} - X_{t_2}^{j,N,h}|^p\big] \le Ke^{\kappa t_2}(1 + \mathbb{E}\big[|\xi|^p\big]e^{-\kappa t_2}). \tag{6.13}
$$

Note that below we will show that the constant on the right-hand side will not blow up as m increases.

Case: $m \ge 3$. More generally, for $m \ge 3$, we have

$$
\begin{aligned}
|X_{t_m}^{i,N,h} - X_{t_m}^{j,N,h}|^2 &\le \big(1 - 2\lambda h + Kh^2\big)|X_{t_{m-1}}^{i,N,h} - X_{t_{m-1}}^{j,N,h}|^2 + \sigma^2|\Delta\overline{W}_{m-1}^{i,N,h} - \Delta\overline{W}_{m-1}^{j,N,h}|^2 \\
&\quad + 2\big(X_{t_{m-1}}^{i,N,h} + b_{m-1}^{i,N,h}h - X_{t_{m-1}}^{j,N,h} + b_{m-1}^{j,N,h}h\big)\cdot\big(\sigma\Delta\overline{W}_{m-1}^{i,N,h} - \sigma\Delta\overline{W}_{m-1}^{j,N,h}\big).
\end{aligned} \tag{6.14}
$$

Since for $m \geq 3$, $X^{i,N,h}_{t_{m-1}}$ is not independent of $\Delta\overline{W}^{i,N,h}_{m-1}$, we further expand $X^{\cdot,N,h}_{t_{m-1}}$ to get

$$
\begin{aligned}
&\big(X^{i,N,h}_{t_{m-1}} + b^{i,N,h}_{m-1}h - X^{j,N,h}_{t_{m-1}} + b^{j,N,h}_{m-1}h\big) \cdot \big(\sigma\Delta\overline{W}^{i,N,h}_{m-1} - \sigma\Delta\overline{W}^{j,N,h}_{m-1}\big) \\
&\quad \leq \big(X^{i,N,h}_{t_{m-2}} - X^{j,N,h}_{t_{m-2}} + b^{i,N,h}_{m-2}h - b^{j,N,h}_{m-2}h + \sigma\Delta\overline{W}^{i,N,h}_{m-2} - \sigma\Delta\overline{W}^{j,N,h}_{m-2}\big) \cdot \big(\sigma\Delta\overline{W}^{i,N,h}_{m-1} - \sigma\Delta\overline{W}^{j,N,h}_{m-1}\big) \\
&\qquad + Kh\big|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}\big|\big|\sigma\Delta\overline{W}^{i,N,h}_{m-1} - \sigma\Delta\overline{W}^{j,N,h}_{m-1}\big|.
\end{aligned}
$$

Define the following local quantities: for all $i, j \in \{1, \dots, N\}$

$$
\begin{aligned}
G^{i,j}_{m,1} &= \big(1 - 2\lambda h + Kh^2\big)|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}|^2, \quad G^{i,j}_{m,2} = \sigma^2|\Delta\overline{W}^{i,N,h}_{m-1} - \Delta\overline{W}^{j,N,h}_{m-1}|^2 \\
G^{i,j}_{m,3} &= 2\big(X^{i,N,h}_{t_{m-2}} - X^{j,N,h}_{t_{m-2}} + \sigma\Delta\overline{W}^{i,N,h}_{m-2} - \sigma\Delta\overline{W}^{j,N,h}_{m-2}\big) \cdot \big(\sigma\Delta\overline{W}^{i,N,h}_{m-1} - \sigma\Delta\overline{W}^{j,N,h}_{m-1}\big), \\
G^{i,j}_{m,4} &= 2\big(b^{i,N,h}_{m-2}h - b^{j,N,h}_{m-2}h\big) \cdot \big(\sigma\Delta\overline{W}^{i,N,h}_{m-1} - \sigma\Delta\overline{W}^{j,N,h}_{m-1}\big) \\
G^{i,j}_{m,5} &= 2Kh\big|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}\big|\big|\sigma\Delta\overline{W}^{i,N,h}_{m-1} - \sigma\Delta\overline{W}^{j,N,h}_{m-1}\big|.
\end{aligned}
$$

Using these local quantities, we can express the estimate (6.14) as follows:

$$
|X^{i,N,h}_{t_m} - X^{j,N,h}_{t_m}|^2 \leq \big(G^{i,j}_{m,1} + G^{i,j}_{m,2} + G^{i,j}_{m,3} + G^{i,j}_{m,4} + G^{i,j}_{m,5}\big).
$$

Now, taking the power of $p/2$ and expectations on both sides, in combination with Young's inequality, we have for some positive constants $K_{p,1}, K_{p,2}, K$ (both are independent of h, T, M and N) such that

$$
\begin{aligned}
&\mathbb{E}\Big[\big(G^{i,j}_{m,1} + G^{i,j}_{m,2} + G^{i,j}_{m,3} + G^{i,j}_{m,4} + G^{i,j}_{m,5}\big)^{p/2}\Big] \\
&\quad = \sum_{l=0}^{p/2}\binom{p/2}{l}\mathbb{E}\Big[(G^{i,j}_{m,1})^{p/2-l}(G^{i,j}_{m,2} + G^{i,j}_{m,3} + G^{i,j}_{m,4} + G^{i,j}_{m,5})^l\Big] \\
&\quad \leq \mathbb{E}\big[|G^{i,j}_{m,1}|^{p/2}\big] + p\mathbb{E}\Big[|G^{i,j}_{m,1}|^{p/2-1}(G^{i,j}_{m,2} + G^{i,j}_{m,3} + G^{i,j}_{m,4} + G^{i,j}_{m,5}) + |G^{i,j}_{m,1}|^{p/2-2}|G^{i,j}_{m,3}|^2\Big] \\
&\qquad + h^{3/2}\Big(K_{p,1}\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}\big|^p\big] + K_{p,2}\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}} - X^{j,N,h}_{t_{m-2}}\big|^p\big] + K\Big),
\end{aligned}
\tag{6.15}
$$

where we used the fact that

$$
\begin{aligned}
&\mathbb{E}\Big[(G^{i,j}_{m,1})^{p/2-2}(G^{i,j}_{m,2} + G^{i,j}_{m,3} + G^{i,j}_{m,4} + G^{i,j}_{m,5})^2\Big] \\
&\quad \leq \mathbb{E}\Big[|G^{i,j}_{m,1}|^{p/2-2}|G^{i,j}_{m,3}|^2\Big] + h^{3/2}\Big(K\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}\big|^p\big] + K\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}} - X^{j,N,h}_{t_{m-2}}\big|^p\big] + K\Big).
\end{aligned}
$$

We now further estimate the terms in (6.15) :

$$
\begin{aligned}
&\mathbb{E}\big[(G^{i,j}_{m,1})^{p/2-1}G^{i,j}_{m,5}\big] \\
&\leq Kh\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}} - X^{j,N,h}_{t_{m-2}} + b^{i,N,h}_{m-2}h - b^{j,N,h}_{m-2}h + \sigma\Delta\overline{W}^{i,N,h}_{m-2} - \sigma\Delta\overline{W}^{j,N,h}_{m-2}\big|^{p-1} \\
&\qquad \big|\sigma\Delta\overline{W}^{i,N,h}_{m-1} - \sigma\Delta\overline{W}^{j,N,h}_{m-1}\big|\big] \\
&\leq Kh\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}} - X^{j,N,h}_{t_{m-2}}\big|^{p-1}\big|\sigma\Delta\overline{W}^{i,N,h}_{m-1} - \sigma\Delta\overline{W}^{j,N,h}_{m-1}\big|\big] \\
&\quad + Kh^{3/2}\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}} - X^{j,N,h}_{t_{m-2}}\big|^{p-1}\big] + Kh^{3/2} \\
&\leq Kh^{3/2}\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}} - X^{j,N,h}_{t_{m-2}}\big|^{p-1}\big] + Kh^{3/2}.
\end{aligned}
$$

Also, note that we have the bound

$$
\mathbb{E}\big[(G^{i,j}_{m,1})^{p/2}\big] \leq (1 - Kh)\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}\big|^p\big].
$$

In the following, $\varepsilon > 0$ (arbitrarily small) and $K > 0$ (both are independent of h, T, M and N) will denote positive constants appearing due to the application of Young's inequality. In addition recall that $X^{\cdot,N,h}_{t_{m-2}}$ is independent of $W^{\cdot,N,h}_{m-1}$. Consequently, we derive

$$\begin{aligned}
&\mathbb{E}\big[(G^{i,j}_{m,1})^{p/2-1}G^{i,j}_{m,2}\big] \leq (1-Kh)\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big|^{p-2}\sigma^2|\Delta\overline{W}^{i,N,h}_{m-1}-\Delta\overline{W}^{j,N,h}_{m-1}|^2\big]\\
&\quad\leq (1-Kh)\mathbb{E}\big[(K+\varepsilon\big|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big|^{p})\sigma^2|\Delta\overline{W}^{i,N,h}_{m-1}-\Delta\overline{W}^{j,N,h}_{m-1}|^2\big]\\
&\quad\leq Kh+\varepsilon\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}+b^{i,N,h}_{m-2}h-b^{j,N,h}_{m-2}h+\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big|^p\\
&\qquad\qquad\cdot\sigma^2|\Delta\overline{W}^{i,N,h}_{m-1}-\Delta\overline{W}^{j,N,h}_{m-1}|^2\big]\\
&\quad\leq Kh+\varepsilon\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}\big|^p\sigma^2|\Delta\overline{W}^{i,N,h}_{m-1}-\Delta\overline{W}^{j,N,h}_{m-1}|^2\big]+Kh^{3/2}\\
&\qquad+K\mathbb{E}\big[\big|b^{i,N,h}_{m-2}h-b^{i,N,h}_{m-2}h\big|^p\sigma^2|\Delta\overline{W}^{i,N,h}_{m-1}-\Delta\overline{W}^{j,N,h}_{m-1}|^2\big]\\
&\qquad+K\mathbb{E}\big[\big|\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big|^p\sigma^2|\Delta\overline{W}^{i,N,h}_{m-1}-\Delta\overline{W}^{j,N,h}_{m-1}|^2\big]\\
&\quad\leq Kh+\varepsilon h\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}\big|^p\big].
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\mathbb{E}\big[(G^{i,j}_{m,1})^{p/2-1}(G^{i,j}_{m,3}+G^{i,j}_{m,4})\big] \leq Kh+\varepsilon h\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big|^p\big]\\
&\qquad+2(1-Kh)\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big|^{p-2}\big(X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big)\cdot\big(\sigma\Delta\overline{W}^{i,N,h}_{m-1}-\sigma\Delta\overline{W}^{j,N,h}_{m-1}\big)\big]\\
&\quad\leq K\mathbb{E}\Big[\big|\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big|^{p-1}\big(X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big)\cdot\big(\sigma\Delta\overline{W}^{i,N,h}_{m-1}-\sigma\Delta\overline{W}^{j,N,h}_{m-1}\big)\Big]\\
&\qquad+K\mathbb{E}\Big[\big|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}\big|^{p-1}\big(\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big)\cdot\big(\sigma\Delta\overline{W}^{i,N,h}_{m-1}-\sigma\Delta\overline{W}^{j,N,h}_{m-1}\big)\Big]\\
&\qquad+K\mathbb{E}\Big[\big|b^{i,N,h}_{m-2}h-b^{i,N,h}_{m-2}h\big|^{p-1}\big(\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big)\cdot\big(\sigma\Delta\overline{W}^{i,N,h}_{m-1}-\sigma\Delta\overline{W}^{j,N,h}_{m-1}\big)\Big]\\
&\qquad+\varepsilon h\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big|^p\big]+Kh+Kh^{3/2}\\
&\quad\leq K\mathbb{E}\Big[\big|\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big|^{p-1}\big(X^{i,N,h}_{t_{m-3}}-X^{j,N,h}_{t_{m-3}}\big)\cdot\big(\sigma\Delta\overline{W}^{i,N,h}_{m-1}-\sigma\Delta\overline{W}^{j,N,h}_{m-1}\big)\Big]\\
&\qquad+K\mathbb{E}\Big[\big|X^{i,N,h}_{t_{m-3}}-X^{j,N,h}_{t_{m-3}}\big|^{p-1}\big(\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big)\cdot\big(\sigma\Delta\overline{W}^{i,N,h}_{m-1}-\sigma\Delta\overline{W}^{j,N,h}_{m-1}\big)\Big]\\
&\qquad+\varepsilon h\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big|^p\big]+Kh+Kh^{3/2}\\
&\quad\leq Kh+Kh^{3/2}+\varepsilon h\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-3}}-X^{j,N,h}_{t_{m-3}}\big|^p\big]+\varepsilon h\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big|^p\big],
\end{aligned}$$

where we expanded the second term in the second inequality, used Young's inequality and the fact that $X^{\cdot,N,h}_{t_{m-3}}$ is independent of $W^{\cdot,N,h}_{m-2}$. For the last term, we have

$$\begin{aligned}
&\mathbb{E}\big[(G^{i,j}_{m,1})^{p/2-2}|G^{i,j}_{m,3}|^2\big]\\
&\quad\leq K\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}\big|^{p-4}\big(\big|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}\big|^2+\big|\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big|^2\big)\\
&\qquad\qquad\cdot\big|\sigma\Delta\overline{W}^{i,N,h}_{m-1}-\sigma\Delta\overline{W}^{j,N,h}_{m-1}\big|^2\big]\\
&\quad\leq K\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}+b^{i,N,h}_{m-2}h-b^{j,N,h}_{m-2}h+\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big|^{p-4}\\
&\qquad\qquad\cdot\big(\big|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}\big|^2+\big|\sigma\Delta\overline{W}^{i,N,h}_{m-2}-\sigma\Delta\overline{W}^{j,N,h}_{m-2}\big|^2\big)\cdot\big|\sigma\Delta\overline{W}^{i,N,h}_{m-1}-\sigma\Delta\overline{W}^{j,N,h}_{m-1}\big|^2\big]\\
&\quad\leq Kh+\varepsilon h\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}\big|^p\big].
\end{aligned}$$

Hence, there exist positive constants $K_{p,3},K_{p,4},K_{p,5}$ (all independent of h,T,M and N) satisfying $K_{p,3}>2K_{p,4}$ (by carefully choosing the constants in Young's inequality) such that

$$\begin{aligned}
&\mathbb{E}\big[|X^{i,N,h}_{t_m}-X^{j,N,h}_{t_m}|^p\big]\\
&\quad\leq(1-K_{p,3}h)\mathbb{E}\big[|X^{i,N,h}_{t_{m-1}}-X^{j,N,h}_{t_{m-1}}|^p\big]\\
&\qquad+K_{p,4}h\big(\mathbb{E}\big[|X^{i,N,h}_{t_{m-2}}-X^{j,N,h}_{t_{m-2}}|^p\big]+\mathbb{E}\big[|X^{i,N,h}_{t_{m-3}}-X^{j,N,h}_{t_{m-3}}|^p\big]\big)+K_{p,5}h\\
&\quad\leq K+Ke^{2\kappa}(1+\mathbb{E}\big[|\xi|^p\big])e^{-(K_{p,3}-2K_{p,4})t_m/3},
\end{aligned}\tag{6.16}$$

for some constant K (independent of h, T, M and N). In the last estimate, we used Lemma 6.4.7 with $c_1 \equiv (1-K_{p,3}h), c_2 \equiv K_{p,4}h, c_3 \equiv K_{p,4}h, C \equiv K_{p,5}h$ and the moment bounds for the differences process at $\{t_0, t_1, t_2\}$ in (6.12) and (6.13). Note that the condition $0 < c_1+c_2+c_3 < 1$ is satisfied for our choice $h \in (0, \min\{1/2\lambda, 1\})$.

We conclude that there exist some positive constants K, κ (both independent of h, T, M and N), such that for all $i, j \in \{1, \dots, N\}$, and $m \geq 0$

$$\mathbb{E}\big[|X_{t_m}^{i,N,h} - X_{t_m}^{j,N,h}|^p\big] \leq K\big(1 + \mathbb{E}\big[|\xi|^p\big]e^{-\kappa t_m}\big). \tag{6.17}$$

Part 2: Moments are uniformly bounded in time. Let $p \geq 2$ be given. We now prove that for all $i \in \{1, \dots, N\}, m \geq 0$, $\mathbb{E}\big[|X_{t_m}^{i,N,h}|^p\big]$ is uniformly bounded in time. As in *Part 1*, we separately consider the cases $m \in \{0, 1, 2\}$ then $m \geq 3$.

Let $i \in \{1, \dots, N\}$ be arbitrary and set $m = 0$. Then we have by assumption on the initial data

$$\mathbb{E}\big[|X_0^{i,N,h}|^p\big] = \mathbb{E}\big[|\xi|^p\big] < \infty.$$

Case 1: $m = 1$. Due to Assumption 5.2.1 and Jensen's inequality, it follows that

$$\begin{aligned}|X_{t_1}^{i,N,h}|^2 &= |X_{t_0}^{i,N,h} + b_0^{i,N,h}h|^2 + |\sigma\Delta\overline{W}_0^{i,N,h}|^2 + 2\big(X_{t_0}^{i,N,h} + b_0^{i,N,h}h\big)\cdot\big(\sigma\Delta\overline{W}_0^{i,N,h}\big)\\ &\leq (1-2\lambda h + Kh^2)|X_{t_0}^{i,N,h}|^2 + 2\big(X_{t_0}^{i,N,h} + b_0^{i,N,h}h\big)\cdot\big(\sigma\Delta\overline{W}_0^{i,N,h}\big) + |\sigma\Delta\overline{W}_0^{i,N,h}|^2\\ &\quad + \frac{2K_V h}{N}\sum_{j=1}^N |X_{t_0}^{i,N,h} - X_{t_0}^{j,N,h}||X_{t_0}^{i,N,h}| + \frac{K_V^2 h^2}{N}\sum_{j=1}^N |X_{t_0}^{i,N,h} - X_{t_0}^{j,N,h}|^2.\end{aligned}$$

Raising to the power $p/2$ and taking the expectation on both sides above, an application of (6.17), Jensen's inequality and Young's inequality shows that there exist positive constants $K_{p,6}, \kappa$ (both independent of h, T, M and N) such that

$$\mathbb{E}\big[|X_{t_1}^{i,N,h}|^p\big] \leq e^{\kappa t_1}K_{p,6}\big(1 + \mathbb{E}\big[|\xi|^p\big]e^{-\kappa t_1}\big).$$

Case 2: $m = 2$. There exist positive constants $K_{p,6}, \kappa$ (both are independent of h, T, M and N) such that

$$\mathbb{E}\big[|X_{t_2}^{i,N,h}|^p\big] \leq e^{\kappa t_2}K_{p,6}\big(1 + \mathbb{E}\big[|\xi|^p\big]e^{-\kappa t_2}\big).$$

We will show that the constant on the right-hand side does not blow up as m increases.

Case 3: $m \geq 3$. We have

$$\begin{aligned}|X_{t_m}^{i,N,h}|^2 &= |X_{t_{m-1}}^{i,N,h} + b_{m-1}^{i,N,h}h|^2 + \sigma^2|\Delta\overline{W}_{m-1}^{i,N,h}|^2 + 2\big(X_{t_{m-1}}^{i,N,h} + b_{m-1}^{i,N,h}h\big)\cdot\big(\Delta\overline{W}_{m-1}^{i,N,h}\big)\\ &\leq (1-2\lambda h + Kh^2)|X_{t_{m-1}}^{i,N,h}|^2 + \sigma^2|\Delta\overline{W}_{m-1}^{i,N,h}|^2 + 2\big(X_{t_{m-1}}^{i,N,h} + b_{m-1}^{i,N,h}h\big)\cdot\big(\Delta\overline{W}_{m-1}^{i,N,h}\big)\\ &\quad + \frac{2K_V h}{N}\sum_{j=1}^N |X_{t_{m-1}}^{i,N,h} - X_{t_{m-1}}^{j,N,h}||X_{t_0}^{i,N,h}| + \frac{K_V^2 h^2}{N}\sum_{j=1}^N |X_{t_{m-1}}^{i,N,h} - X_{t_{m-1}}^{j,N,h}|^2.\end{aligned}$$

Since for $m \geq 3$, $X_{t_{m-1}}^{i,N,h}$ is not independent of $\Delta\overline{W}_{m-1}^{i,N,h}$, we further expand $X_{t_{m-1}}^{i,N,h}$ and estimate

$$\begin{aligned}&\big(X_{t_{m-1}}^{i,N,h} + b_{m-1}^{i,N,h}h\big)\cdot\big(\Delta\overline{W}_{m-1}^{i,N,h}\big)\\ &\qquad\leq h|b_{m-1}^{i,N,h}||\Delta\overline{W}_{m-1}^{i,N,h}| + \big(X_{t_{m-2}}^{i,N,h} + b_{m-2}^{i,N,h}h + \Delta\overline{W}_{m-2}^{i,N,h}\big)\cdot\big(\Delta\overline{W}_{m-1}^{i,N,h}\big).\end{aligned}$$

Note that $X_{t_{m-2}}^{\cdot,N,h}$ is independent of $W_{m-1}^{\cdot,N,h}$. Similar to the analysis of the first part, we define the following local quantities: for all $i \in \{1, \dots, N\}$,

$$
\begin{aligned}
G^i_{m,1} &= (1-2\lambda h + Kh^2)|X^{i,N,h}_{t_{m-1}}|^2,\\
G^i_{m,2} &= \sigma^2|\Delta\overline{W}^{i,N,h}_{m-1}|^2,\\
G^i_{m,3} &= 2\big(X^{i,N,h}_{t_{m-2}} + b^{i,N,h}_{m-2}h + \Delta\overline{W}^{i,N,h}_{m-2}\big)\cdot\big(\Delta\overline{W}^{i,N,h}_{m-1}\big),\\
G^i_{m,4} &= 2h|b^{i,N,h}_{m-1}||\Delta\overline{W}^{i,N,h}_{m-1}|,\\
G^i_{m,5} &= \frac{2K_V h}{N}\sum_{j=1}^N |X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}||X^{i,N,h}_{t_{m-1}}| + \frac{K_V^2h^2}{N}\sum_{j=1}^N |X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}|^2,
\end{aligned}
$$

and note that

$$
|X^{i,N,h}_{t_m}|^2 \le \big(G^i_{m,1}+G^i_{m,2}+G^i_{m,3}+G^i_{m,4}+G^i_{m,5}\big).
$$

Raising to the power $p/2$ and taking the expectation on both sides, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\big(G^i_{m,1}+G^i_{m,2}+G^i_{m,3}+G^i_{m,4}+G^i_{m,5}\big)^{p/2}\Big]\\
&\quad= \sum_{l=0}^{p/2}\binom{p/2}{l}\mathbb{E}\Big[(G^i_{m,1})^{p/2-l}(G^i_{m,2}+G^i_{m,3}+G^i_{m,4}+G^i_{m,5})^l\Big]\\
&\quad\le \mathbb{E}\big[|G^i_{m,1}|^{p/2}\big] + p\mathbb{E}\Big[(G^i_{m,1})^{p/2-1}(G^i_{m,2}+G^i_{m,3}+G^i_{m,4}+G^i_{m,5}) + |G^i_{m,1}|^{p/2-2}|G^i_{m,3}|^2\Big]\\
&\qquad + h^{3/2}\Big(K_{p,1}\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}\big|^p\big] + K_{p,2}\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-2}}\big|^p\big] + K\Big).
\end{aligned}
$$

Using (6.16), we obtain the following estimate:

$$
\begin{aligned}
&\mathbb{E}\big[(G^i_{m,1})^{p/2-1}G^i_{m,5}\big]\\
&\quad\le K\mathbb{E}\Big[\big|X^{i,N,h}_{t_{m-1}}\big|^{p-2}\Big(\frac{2K_Vh}{N}\sum_{j=1}^N|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}||X^{i,N,h}_{t_{m-1}}| + \frac{K_V^2h^2}{N}\sum_{j=1}^N|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}|^2\Big)\Big]\\
&\quad\le h\Big(\varepsilon\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}\big|^p\big] + \frac{K}{N}\sum_{j=1}^N\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}} - X^{j,N,h}_{t_{m-1}}\big|^p\big]\Big)\\
&\quad\le \varepsilon h\mathbb{E}\big[\big|X^{i,N,h}_{t_{m-1}}\big|^p\big] + Kh(1+\mathbb{E}\big[|\xi|^p\big]e^{-(K_{p,3}-2K_{p,4})t_m/3}).
\end{aligned}
$$

The analysis of the other terms works similarly as in the proof of the first part. Hence, we conclude that there exists positive constants $K_{p,7}, K_{p,8}, K_{p,9}$ (both independent of h, T, M and N) satisfying $K_{p,7} > 2K_{p,8}$ and $e^{-(K_{p,3}-2K_{p,4})h} \neq 1-(K_{p,7}-2K_{p,8})h > 0$ (for sufficiently small h), such that

$$
\begin{aligned}
\mathbb{E}\big[|X^{i,N,h}_{t_m}|^p\big] &\le (1-K_{p,7}h)\mathbb{E}\big[|X^{i,N,h}_{t_{m-1}}|^p\big] + K_{p,8}h\big(\mathbb{E}\big[|X^{i,N,h}_{t_{m-2}}|^p\big] + \mathbb{E}\big[|X^{i,N,h}_{t_{m-3}}|^p\big]\big)\\
&\qquad + K_{p,9}h\big(1+\mathbb{E}\big[|\xi|^p\big]e^{-(K_{p,3}-2K_{p,4})t_m}\big)\\
&\le K + K(1+\mathbb{E}\big[|\xi|^p\big])\big(e^{-(K_{p,7}-2K_{p,8})t_m/3} + e^{-(K_{p,3}-2K_{p,4})t_m/3}\big),
\end{aligned}
$$

for some constant K (independent of h, T, M and N). In the last estimate, we used the second statement in Lemma 6.4.7 with $c_1 \equiv (1-K_{p,7}h), c_2 \equiv K_{p,8}h, c_3 \equiv K_{p,8}h$, $c_4 = K_{p,9}\mathbb{E}\big[|\xi|^p\big]h$, $c_5 = K_{p,3} - 2K_{p,4}$, $C \equiv K_{p,9}h$ and the moment bounds at $\{t_0, t_1, t_2\}$. $\square$

Proof of Proposition 5.2.3 – Statement (2): the L^2-strong error. Consider the IPS described in (5.11) and recall the auxiliary scheme of (5.153). For all $m \in \{0,\dots,M-1\}$ with $\hat{X}^{i,N,h}_{t_0} = X^{i,N}_{t_0}$, we define

$$
\hat{X}^{i,N,h}_{t_{m+1}} = \hat{X}^{i,N,h}_{t_m} - \Big(\nabla U(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m)
$$

$$+\frac{1}{N}\sum_{j=1}^{N}\nabla V(\hat{X}_{t_m}^{i,N,h}+\frac{\sigma}{2}\Delta W_m^i-\hat{X}_{t_m}^{j,N,h}-\frac{\sigma}{2}\Delta W_m^j)\Big)h+\sigma\Delta W_m^i,$$

where $t_m:=mh$, $T:=Mh$, and $\Delta W_m^i=W_{t_m}^i-W_{t_{m-1}}^i$. Recall that $X_{t_m}^{i,N,h}=\hat{X}_{t_m}^{i,N,h}+\sigma\Delta W_m^i/2$. We compute the difference terms:

$$\begin{aligned}
\Delta_{i,m} &:= X_{t_m}^{i,N}-\hat{X}_{t_{m+1}}^{i,N,h} && (6.18)\\
&= X_{t_m}^{i,N}-\Bigg(X_{t_{m-1}}^{i,N}-\Big(\nabla U(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i) && (6.19)\\
&\qquad+\frac{1}{N}\sum_{j=1}^{N}\nabla V(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i-X_{t_{m-1}}^{j,N}-\frac{\sigma}{2}\Delta W_m^j)\Big)h+\sigma\Delta W_m^i\Bigg)\\
&\quad+\Bigg(X_{t_{m-1}}^{i,N}-\Big(\nabla U(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i) && (6.20)\\
&\qquad+\frac{1}{N}\sum_{j=1}^{N}\nabla V(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i-X_{t_{m-1}}^{j,N}-\frac{\sigma}{2}\Delta W_m^j)\Big)h+\sigma\Delta W_m^i\Bigg)\\
&\quad-\Bigg(\hat{X}_{t_m}^{i,N,h}-\Big(\nabla U(\hat{X}_{t_m}^{i,N,h}+\frac{\sigma}{2}\Delta W_m^i) && (6.21)\\
&\qquad+\frac{1}{N}\sum_{j=1}^{N}\nabla V(\hat{X}_{t_m}^{i,N,h}+\frac{\sigma}{2}\Delta W_m^i-\hat{X}_{t_m}^{j,N,h}-\frac{\sigma}{2}\Delta W_m^j)\Big)h+\sigma\Delta W_m^i\Bigg)\\
&=: R_{t_m}^{i,1}+R_{t_m}^{i,2}.
\end{aligned}$$

Note that here we match $X_{t_m}^{i,N}$ with $\hat{X}_{t_{m+1}}^{i,N,h}$ instead of $\hat{X}_{t_m}^{i,N,h}$.

We estimate the above terms separately and collect all the estimates an the end. For the first term, taking squares and expectations yields

$$\begin{aligned}
&\mathbb{E}\big[|R_{t_m}^{i,1}|^2\big]\\
&=\mathbb{E}\Bigg[\Big|\,X_{t_m}^{i,N}-\Bigg(X_{t_{m-1}}^{i,N}-\Big(\nabla U(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i)\\
&\qquad+\frac{1}{N}\sum_{j=1}^{N}\nabla V(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i-X_{t_{m-1}}^{j,N}-\frac{\sigma}{2}\Delta W_m^j)\Big)h+\sigma\Delta W_m^i\Bigg)\,\Big|^2\Bigg]\\
&=\mathbb{E}\Bigg[\Big|-\int_{t_{m-1}}^{t_m}\Big(\nabla U(X_s^{i,N})-\nabla U(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i)\\
&\qquad+\frac{1}{N}\sum_{j=1}^{N}\nabla V(X_s^{i,N}-X_s^{j,N})-\frac{1}{N}\sum_{j=1}^{N}\nabla V(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i-X_{t_{m-1}}^{j,N}-\frac{\sigma}{2}\Delta W_m^j)\Big)\mathrm{d}s\,\Big|^2\Bigg]\\
&\le 2h\int_{t_{m-1}}^{t_m}\mathbb{E}\Big[\big|\nabla U(X_s^{i,N})-\nabla U(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i)\big|^2\Big]\mathrm{d}s\\
&\quad+2h\int_{t_{m-1}}^{t_m}\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\Big[\big|\nabla V(X_s^{i,N}-X_s^{j,N})-\nabla V(X_{t_{m-1}}^{i,N}+\frac{\sigma}{2}\Delta W_m^i-X_{t_{m-1}}^{j,N}-\frac{\sigma}{2}\Delta W_m^j)\big|^2\,\mathrm{d}s\Big]\\
&\le Kh\int_{t_{m-1}}^{t_m}\Bigg(\mathbb{E}\Big[\big|X_s^{i,N}-X_{t_{m-1}}^{i,N}\big|^2\Big]+\mathbb{E}\Big[\big|\frac{\sigma}{2}\Delta W_m^i\big|^2\Big]\\
&\qquad+\frac{1}{N}\sum_{j=1}^{N}\Big(\mathbb{E}\Big[\big|X_s^{j,N}-X_{t_{m-1}}^{j,N}\big|^2\Big]+\mathbb{E}\Big[\big|\frac{\sigma}{2}\Delta W_m^j\big|^2\Big]\Big)\Bigg)\mathrm{d}s\\
&\le Kh\int_{t_{m-1}}^{t_m}\Big(\max_{i\in\{1,\dots,N\}}\mathbb{E}\Big[\big|X_s^{i,N}-X_{t_{m-1}}^{i,N}\big|^2\Big]+h\Big)\mathrm{d}s\ \le Kh^2,
\end{aligned}$$

where we used Jensen's inequality, the Lipschitz continuity of the potentials and Proposition 5.2.2. Next, we consider the second term $R^{i,2}_{t_m}$. Taking squares and expectations, we have that

$$
\begin{aligned}
&\mathbb{E}\big[|R^{i,2}_{t_m}|^2\big] \\
=&\mathbb{E}\Big[\Big| \Big(X^{i,N}_{t_{m-1}} - \Big(\nabla U(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m) \\
&\qquad + \frac{1}{N}\sum_{j=1}^N \nabla V(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m - X^{j,N}_{t_{m-1}} - \frac{\sigma}{2}\Delta W^j_m)\Big)h\Big) \\
&\quad - \Big(\hat{X}^{i,N,h}_{t_m} - \Big(\nabla U(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m) + \\
&\qquad \frac{1}{N}\sum_{j=1}^N \nabla V(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m - \hat{X}^{j,N,h}_{t_m} - \frac{\sigma}{2}\Delta W^j_m)\Big)h\Big) \Big|^2\Big] \\
&\leq \mathbb{E}\Big[\big|X^{i,N}_{t_{m-1}} - \hat{X}^{i,N,h}_{t_m}\big|^2\Big] - h\mathbb{E}\Big[\Big(X^{i,N}_{t_{m-1}} - \hat{X}^{i,N,h}_{t_m}\Big) \\
&\qquad \cdot \Big(\nabla U(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m) - \nabla U(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m)\Big)\Big] \\
&\quad - \frac{h}{N}\sum_{j=1}^N \mathbb{E}\Big[\Big(X^{i,N}_{t_{m-1}} - \hat{X}^{i,N,h}_{t_m}\Big) \cdot \Big(\nabla V(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m - X^{j,N}_{t_{m-1}} - \frac{\sigma}{2}\Delta W^j_m) \\
&\qquad\qquad - \nabla V(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m - \hat{X}^{j,N,h}_{t_m} - \frac{\sigma}{2}\Delta W^j_m)\Big)\Big] \\
&\quad + 2h^2\mathbb{E}\Big[\big|\nabla U(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m) - \nabla U(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m)\big|^2\Big] \\
&\quad + \frac{2h^2}{N}\sum_{j=1}^N \mathbb{E}\Big[\big|\nabla V(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m - X^{j,N}_{t_{m-1}} - \frac{\sigma}{2}\Delta W^j_m) \\
&\qquad\qquad - \nabla V(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m - \hat{X}^{j,N,h}_{t_m} - \frac{\sigma}{2}\Delta W^j_m)\big|^2\Big] \\
&\leq \mathbb{E}\Big[\big|X^{i,N}_{t_{m-1}} - \hat{X}^{i,N,h}_{t_m}\big|^2\Big] - h\mathbb{E}\Big[\Big(X^{i,N}_{t_{m-1}} - \hat{X}^{i,N,h}_{t_m}\Big) \\
&\qquad \cdot \Big(\nabla U(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m) - \nabla U(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m)\Big)\Big] \\
&\quad + Kh^2 - \frac{h}{N}\sum_{j=1}^N \mathbb{E}\Big[\Big(X^{i,N}_{t_{m-1}} - \hat{X}^{i,N,h}_{t_m}\Big) \cdot \Big(\nabla V(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m - X^{j,N}_{t_{m-1}} - \frac{\sigma}{2}\Delta W^j_m) \\
&\qquad - \nabla V(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m - \hat{X}^{j,N,h}_{t_m} - \frac{\sigma}{2}\Delta W^j_m)\Big)\Big],
\end{aligned}
$$

where we used Jensen's inequality, Propositions 5.2.2 and the Statement (1) in Proposition 5.2.3. We further estimate

$$
\begin{aligned}
&\mathbb{E}[|R^{i,2}_{t_m}|^2] \leq (1-\lambda h)\mathbb{E}\Big[\big|X^{i,N}_{t_{m-1}} - \hat{X}^{i,N,h}_{t_m}\big|^2\Big] + Kh^2 - \frac{h}{2N^2} \\
&\cdot \sum_{i,j=1}^N \mathbb{E}\Big[\Big((X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m - X^{j,N}_{t_{m-1}} - \frac{\sigma}{2}\Delta W^j_m) - (\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m - \hat{X}^{j,N,h}_{t_m} - \frac{\sigma}{2}\Delta W^j_m)\Big) \\
&\qquad \cdot \Big(\nabla V(X^{i,N}_{t_{m-1}} + \frac{\sigma}{2}\Delta W^i_m - X^{j,N}_{t_{m-1}} - \frac{\sigma}{2}\Delta W^j_m) \\
&\qquad\quad - \nabla V(\hat{X}^{i,N,h}_{t_m} + \frac{\sigma}{2}\Delta W^i_m - \hat{X}^{j,N,h}_{t_m} - \frac{\sigma}{2}\Delta W^j_m)\Big)\Big] \leq (1-\lambda h)\mathbb{E}\Big[\big|\Delta_{i,m-1}\big|^2\Big] + Kh^2,
\end{aligned}
$$

with $\Delta_{i,m}$ as defined in (6.18). We used the *'symmetrization trick'* in (5.25) to handle the convolution term. Note that the positive constant K is independent of h, T, M and N. Hence for all $m \in \{1, \dots, M-1\},\ i \in \{1, \dots, N\}$, there exists a positive constant K such that

$$
\mathbb{E}\Big[\big|\Delta_{i,m}\big|^2\Big] \leq (1-\lambda h)\mathbb{E}\Big[\big|\Delta_{i,m-1}\big|^2\Big] + Kh^2
$$

$$= (1-\lambda h)^m \mathbb{E}\Big[\big|\Delta_{i,0}\big|^2\Big] + Kh^2 \sum_{j=0}^{m-1} (1-\lambda h)^j \leq Kh + \frac{Kh}{\lambda},$$

where we used $\mathbb{E}\big[|\Delta_{i,0}|^2\big] \leq Kh$. Recall that $X_{t_m}^{i,N,h} = \hat{X}_{t_m}^{i,N,h} + \sigma \Delta W_m^i/2$. Using Statement (1) in Proposition 5.2.3 and (5.13), we further have

$$\mathbb{E}\Big[\big|X_{t_m}^{i,N,h} - \hat{X}_{t_m}^{i,N,h}\big|^2\Big] = \mathbb{E}\Big[\big|\frac{\sigma}{2}\Delta W_m^i\big|^2\Big] \leq Kh \quad \text{and} \quad \mathbb{E}\Big[\big|X_{t_{m+1}}^{i,N,h} - X_{t_m}^{i,N,h}\big|^2\Big] \leq Kh.$$

Collecting the last 3 estimates, we have for all $m \in \{1, \ldots, M-1\}$, $i \in \{1, \ldots, N\}$,

$$\begin{aligned}
\mathbb{E}\Big[\big|X_{t_m}^{i,N} - X_{t_m}^{i,N,h}\big|^2\Big] &= \mathbb{E}\Big[\big|\Delta_{i,m} + (\hat{X}_{t_{m+1}}^{i,N,h} - X_{t_{m+1}}^{i,N,h}) + (X_{t_{m+1}}^{i,N,h} - X_{t_m}^{i,N,h})\big|^2\Big] \\
&\leq K\Big(\mathbb{E}\Big[\big|\Delta_{i,m}\big|^2\Big] + \mathbb{E}\Big[\big|X_{t_{m+1}}^{i,N,h} - \hat{X}_{t_{m+1}}^{i,N,h}\big|^2\Big] + \mathbb{E}\Big[\big|X_{t_{m+1}}^{i,N,h} - X_{t_m}^{i,N,h}\big|^2\Big]\Big) \leq Kh.
\end{aligned}$$

As K is independent of the critical quantities M and N, maximizing over i and m yields the final result. □

6.4.3 Auxiliary results

The following statement is an auxiliary result on the differences of SDE starting at different times (t and s with $t \leq s$) at the same point $\boldsymbol{x}$ and is used in the proof of Proposition 5.4.4.

Lemma 6.4.2. *Let the assumptions and setup of Proposition 5.4.4 hold and let $r \geq s \geq t \geq 0$, $u \geq 0$ with $s-t < 1$. Let the starting positions $x_i \in L^4(\Omega, \mathbb{R})$ be $\mathcal{F}_t$-measurable random variables that are identically distributed over all $i \in \{1, \ldots, N\}$. Let $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\geq t\geq 0}$ and $(\boldsymbol{X}_r^{s,\boldsymbol{x},N})_{r\geq s\geq 0}$ be the solutions of (5.15) starting from $\boldsymbol{x}$ at time t and s, respectively. Then there exist some $\lambda_2 \in (0, \min\{\lambda - 2K_V, \lambda_1\})$ and $K > 0$ (both are independent of s, t, N), such that for any $i \in \{1, \ldots, N\}$*

$$\mathbb{E}\Big[\big|X_{s+u}^{t,x_i,i,N} - X_{s+u}^{s,x_i,i,N}\big|^4\Big] \leq K(s-t)^2 e^{-4\lambda_2 u}.$$

Proof. By Itô's formula, we have, for any $s \geq t \geq 0$, $u \geq 0$, $i \in \{1, \ldots, N\}$

$$\begin{aligned}
&\mathbb{E}\Big[\big|X_{s+u}^{t,x_i,i,N} - X_{s+u}^{s,x_i,i,N}\big|^4\Big] \\
&\leq \mathbb{E}\Big[\big|X_{s}^{t,x_i,i,N} - X_{s}^{s,x_i,i,N}\big|^4\Big] \\
&- 4\int_0^u \mathbb{E}\Big[\big|X_{s+w}^{t,x_i,i,N} - X_{s+w}^{s,x_i,i,N}\big|^2 \Big(X_{s+w}^{t,x_i,i,N} - X_{s+w}^{s,x_i,i,N}\Big) \cdot \Big(\Big(\nabla U(X_{s+w}^{t,x_i,i,N}) - \nabla U(X_{s+w}^{s,x_i,i,N})\Big) \\
&+ \Big(\frac{1}{N}\sum_{l=1}^N \nabla V(X_{s+w}^{t,x_i,i,N} - X_{s+w}^{t,x_l,l,N}) - \nabla V(X_{s+w}^{s,x_i,i,N} - X_{s+w}^{s,x_l,l,N})\Big)\Big)\Big] \mathrm{d}w \\
&\leq \mathbb{E}\Big[\big|X_{s}^{t,x_i,i,N} - X_{s}^{s,x_i,i,N}\big|^4\Big] - 4\Big(\lambda - \Big(1 + \frac{3}{4}\Big) K_V\Big) \int_0^u \mathbb{E}\Big[\big|X_{s+w}^{t,x_i,i,N} - X_{s+w}^{s,x_i,i,N}\big|^4\Big] \mathrm{d}w \\
&+ \frac{K_V}{N} \sum_{l=1}^N \int_0^u \mathbb{E}\Big[\big|X_{s+w}^{t,x_l,l,N} - X_{s+w}^{s,x_l,l,N}\big|^4\Big] \mathrm{d}w,
\end{aligned}$$

where we used Young's inequality along with Assumption 5.2.1. From the proof of Proposition 5.4.4 (see, equation 5.46), we deduce that $\mathbb{E}\big[|X_s^{t,x_i,i,N} - X_s^{s,x_i,i,N}|^4\big] \leq K(s-t)^2$. Using the fact that $\lambda > 2K_V$, we conclude the claim. □

The next statement concerning the second-order variation process is similar to Proposition 5.4.4 which is used in the proof in Lemma 5.6.1.

Proposition 6.4.3. *Let the assumptions and set up of Lemma 5.4.2 and Proposition 5.4.4 hold. Then there exist some $\lambda_4 \in (0, \min\{\lambda - 2K_V, \lambda_3\})$ and $K > 0$ (both are independent of* *h, T, M and N) such that for all $T \geq s \geq t \geq 0$ (with $s-t<1$) and $i \in \{1,\dots,N\}$*

$$\mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_i,x_i} - X^{s,x_i,i,N}_{T,x_i,x_i}|^2\Big] \leq K(s-t)e^{-2\lambda_4(T-s)},$$

$$\sum_{i,j,k=1, i\neq j\neq k}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_j,x_k} - X^{s,x_i,i,N}_{T,x_j,x_k}|^2\Big] \leq \frac{K(s-t)}{N}e^{-2\lambda_4(T-s)},$$

and

$$\sum_{i,k=1, i\neq k}^{N} \mathbb{E}\Big[|X^{t,x_i,i,N}_{T,x_k,x_k} - X^{s,x_i,i,N}_{T,x_k,x_k}|^2 + |X^{t,x_i,i,N}_{T,x_i,x_k} - X^{s,x_i,i,N}_{T,x_i,x_k}|^2 + |X^{t,x_i,i,N}_{T,x_k,x_i} - X^{s,x_i,i,N}_{T,x_k,x_i}|^2\Big]$$
$$\leq K(s-t)e^{-2\lambda_4(T-s)}.$$

Proof. This proof is a combination of the methods used to prove Proposition 5.4.4 and Lemma 5.4.5 and we streamline the presentation. For any $i,j,k \in \{1,\dots,N\}$, we have that

$$\begin{aligned}
&|X^{t,x_i,i,N}_{T,x_j,x_k} - X^{s,x_i,i,N}_{T,x_j,x_k}|^2 = |X^{t,x_i,i,N}_{s,x_j,x_k} - X^{s,x_i,i,N}_{s,x_j,x_k}|^2 \\
&- 2\int_0^{T-s} \Big(X^{t,x_i,i,N}_{s+u,x_j} - X^{s,x_i,i,N}_{s+u,x_j}\Big)\cdot\Big(\nabla^2 U(X^{t,x_i,i,N}_{s+u})X^{t,x_i,i,N}_{s+u,x_j,x_k} - \nabla^2 U(X^{s,x_i,i,N}_{s+u})X^{s,x_i,i,N}_{s+u,x_j,x_k}\Big)\mathrm{d}u \\
&- 2\int_0^{T} \Big(X^{t,x_i,i,N}_{s+u,x_j,x_k} - X^{s,x_i,i,N}_{s+u,x_j,x_k}\Big)\cdot\Big(\frac{1}{N}\sum_{l=1}^{N}\nabla^2 V(X^{t,x_i,i,N}_{s+u} - X^{t,x_l,l,N}_{s+u})(X^{t,x_i,i,N}_{s+u,x_j,x_k} \\
&\qquad - X^{t,x_l,l,N}_{s+u,x_j,x_k}) - \frac{1}{N}\sum_{l=1}^{N}\nabla^2 V(X^{s,x_i,i,N}_{s+u} - X^{s,x_l,l,N}_{s+u})(X^{s,x_i,i,N}_{s+u,x_j,x_k} - X^{s,x_l,l,N}_{s+u,x_j,x_k})\Big)\mathrm{d}u \\
&- 2\int_0^{T-s} \Big(X^{t,x_i,i,N}_{s+u,x_j,x_k} - X^{s,x_i,i,N}_{s+u,x_j,x_k}\Big)\cdot\Big(\sum_{l=1}^{N}\sum_{l'=1}^{N}\partial^2_{x_l,x_{l'}}B_i(\boldsymbol{X}^{t,\boldsymbol{x},N}_u)X^{t,x_l,l,N}_{s+u,x_j}X^{t,x_{l'},l',N}_{s+u,x_k} \\
&\qquad - \sum_{l=1}^{N}\sum_{l'=1}^{N}\partial^2_{x_l,x_{l'}}B_i(\boldsymbol{X}^{s,\boldsymbol{x},N}_u)X^{s,x_l,l,N}_{s+u,x_j}X^{s,x_{l'},l',N}_{s+u,x_k}\Big)\,\mathrm{d}u.
\end{aligned}$$

The remaining steps are similar to those in the proof of Lemma 5.4.5 and Proposition 5.4.4 and we therefore omit a detailed analysis. □

The next statement is a classical result on the stability of SDEs with respect their initial condition.

Lemma 6.4.4. *Let Assumption 5.2.1 hold (with $\lambda > 0$ denoting the convexity parameter) and let $p \geq 2$ be given. Let $(\boldsymbol{X}^N_t)_{t\geq 0}$, and $(\mathbf{Y}^N_t)_{t\geq 0}$ be generated from* (5.11) *with i.i.d. initial states $X^{i,N}_0 \sim \mu, Y^{i,N}_0 \sim \nu \in \mathcal{P}_p(\mathbb{R})$, where $\boldsymbol{X}^N_0 = (X^{1,N}_0,\dots,X^{N,N}_0)$ and $\mathbf{Y}^N_0 = (Y^{1,N}_0,\dots,Y^{N,N}_0)$. Then for any $i \in \{1,\dots,N\}$, $t \geq 0$, we have*

$$\mathbb{E}\Big[\big|X^{i,N}_t - Y^{i,N}_t\big|^2\Big] \leq e^{-2\lambda t}\mathbb{E}\Big[\big|X^{i,N}_0 - Y^{i,N}_0\big|^2\Big]. \tag{6.22}$$

Furthermore, if $\mathbf{Y}^N_0 \sim \mu^{N,*}$, the stationary distribution of (5.11) *(given in Proposition 5.2.2), we have*

$$\mathbb{E}\Big[\big|X^{i,N}_t - Y^{i,N}_0\big|^2\Big] = \mathbb{E}\Big[\big|X^{i,N}_t - Y^{i,N}_t\big|^2\Big] \leq e^{-2\lambda t}\mathbb{E}\Big[\big|X^{i,N}_0 - Y^{i,N}_0\big|^2\Big]. \tag{6.23}$$

Proof. This result is classical and we present only a sketch of its proof. By Itô's formula, we have

$$\begin{aligned}
&\mathbb{E}\Big[\big|X^{i,N}_t - Y^{i,N}_t\big|^2\Big] \\
&\leq \mathbb{E}\Big[\big|X^{i,N}_0 - Y^{i,N}_0\big|^2\Big] - 2\int_0^t \mathbb{E}\Big[\Big(X^{i,N}_s - Y^{i,N}_s\Big)\cdot\Big(\nabla U(X^{i,N}_s) - \nabla U(Y^{i,N}_s)\Big)\Big]\mathrm{d}s
\end{aligned}$$

$$
-2\int_0^t \mathbb{E}\Big[\Big(X_s^{i,N}-Y_s^{i,N}\Big)\cdot\Big(\frac{1}{N}\sum_{l=1}^N \nabla V(X_s^{i,N}-X_s^{l,N})-\nabla V(Y_s^{i,N}-Y_s^{l,N})\Big)\Big]\mathrm{d}s
$$

$$
\le \mathbb{E}\Big[\big|X_0^{i,N}-Y_0^{i,N}\big|^2\Big]-2\lambda\int_0^t \mathbb{E}\Big[\big|X_s^{i,N}-Y_s^{i,N}\big|^2\Big]\mathrm{d}s-\frac{1}{2N^2}\sum_{i=1}^N\sum_{l=1}^N
$$

$$
\cdot\int_0^t \mathbb{E}\Big[\Big((X_s^{i,N}-X_s^{l,N})-(Y_s^{i,N}-Y_s^{l,N})\Big)\cdot\Big(\nabla V(X_s^{i,N}-X_s^{l,N})-\nabla V(Y_s^{i,N}-Y_s^{l,N})\Big)\Big]\mathrm{d}u
$$

$$
\le \mathbb{E}\Big[\big|X_0^{i,N}-Y_0^{i,N}\big|^2\Big]-2\lambda\int_0^t \mathbb{E}\Big[\big|X_s^{i,N}-Y_s^{i,N}\big|^2\Big]\mathrm{d}s,
$$

where we used Assumption 5.2.1. This estimate allows to deduce the first result (6.22).

The second result follows since $\boldsymbol{Y}_0^N \sim \mu^{N,*}$ (and thus $\boldsymbol{Y}_t^N \sim \mu^{N,*}$ for any $t \ge 0$) and the one-dimensional marginal distributions of $\mu^{N,*}$ are identical. □

Lemma 6.4.5. *Let the assumptions and set up of Proposition 5.2.3 hold with $\xi \in L^p(\Omega,\mathbb{R})$ for some given $p \ge 2$. Then for the processes defined in* (5.151), (5.153) *and* (5.154)*, respectively, there exists a constant $K>0$ (independent of h,T,M and N) such that for all $m \in \{0,\dots,M-1\}$*

$$
\max_{i\in\{1,\dots,N\}}\sup_{s\in[0,h]}\mathbb{E}\big[\,|X_{t_m+s}^{i,N,h}|^p\big]\le K\big(1+\mathbb{E}\big[\,|\xi|^p\big]e^{-\kappa t_m}\big),
$$

$$
\max_{i\in\{1,\dots,N\}}\mathbb{E}\big[\,|\hat{X}_{t_m}^{i,N,h}|^p\big]\le K\big(1+\mathbb{E}\big[\,|\xi|^p\big]e^{-\kappa t_m}\big),
$$

$$
\max_{i\in\{1,\dots,N\}}\sup_{s\in[0,h)}\mathbb{E}\big[\,|\overline{X}_{t_m+s}^{i,N,h}|^p\big]\le K\big(1+\mathbb{E}\big[\,|\xi|^p\big]e^{-\kappa t_m}\big).
$$

Proof. Using Proposition 5.2.3 and taking the definition of the processes in (5.151) into account, we have for all $m\in\{0,\dots,M-1\}$, $s\in[0,h]$ and $i\in\{1,\dots,N\}$,

$$
\begin{aligned}
\mathbb{E}\big[\,|X_{t_m+s}^{i,N,h}|^p\big] &\le K\Big(\mathbb{E}\big[\,|X_{t_m}^{i,N,h}|^p\big]+\mathbb{E}\big[\,|\nabla U(X_{t_m}^{i,N,h})|^p\big]h^p\\
&\quad+\frac{1}{N}\sum_{j=1}^N\mathbb{E}\big[\,|\nabla V(X_{t_m}^{i,N,h}-X_{t_m}^{j,N,h})|^p\big]h^p+\mathbb{E}\big[\,|\Delta W_m^i|^p\big]+\mathbb{E}\big[\,|\Delta W_{m+1,s}^i|^p\big]\Big)\\
&\le K(1+h^p)\Big(\mathbb{E}\big[\,|X_{t_m}^{i,N,h}|^p\big]+\mathbb{E}\big[\,|\nabla U(X_{t_m}^{i,N,h})|^p\big]+\frac{1}{N}\sum_{j=1}^N\mathbb{E}\big[\,|\nabla U(X_{t_m}^{j,N,h})|^p\big]+1\Big)\\
&\le K\big(1+\mathbb{E}\big[\,|\xi|^p\big]e^{-\kappa t_m}\big),
\end{aligned}
$$

where we used Jensen's inequality and the fact that $\nabla U,\nabla V$ are of linear growth (Assumption 5.2.1).

For the second and the last estimate, recall that $X_{t_m}^{i,N,h}=\hat{X}_{t_m}^{i,N,h}+\sigma\Delta W_m^i/2$ and $\overline{X}_{t_{m-1}+s}^{i,N,h}=\hat{X}_{t_m}^{i,N,h}+\sigma\Delta W_{m,s}^i/2$. We have that for all $m\in\{1,\dots,M-1\}$, $s\in[0,h)$ (recall $h\in(0,1)$ is sufficiently small) and $i\in\{1,\dots,N\}$, we have

$$
\begin{aligned}
\mathbb{E}\big[\,|\hat{X}_{t_m}^{i,N,h}|^p\big] &\le K\big(\mathbb{E}\big[\,|X_{t_m}^{i,N,h}|^p\big]+\mathbb{E}\big[\,|\Delta W_m^i|^p\big]\big)\le K\big(1+\mathbb{E}\big[\,|\xi|^p\big]e^{-\kappa t_m}\big),\\
\mathbb{E}\big[\,|\hat{X}_{t_0}^{i,N,h}|^p\big] &= \mathbb{E}\big[\,|X_{t_0}^{i,N}|^p\big]=\mathbb{E}\big[\,|\xi|^p\big],\\
\mathbb{E}\big[\,|\overline{X}_{t_{m-1}+s}^{i,N,h}|^p\big] &\le K\big(\mathbb{E}\big[\,|\hat{X}_{t_m}^{i,N,h}|^p\big]+\mathbb{E}\big[\,|\Delta W_{m,s}^i|^p\big]\big)\\
&\le K\big(1+\mathbb{E}\big[\,|\xi|^p\big]e^{-\kappa t_{m-1}}e^{-\kappa h}\big)\\
&\le K\big(1+\mathbb{E}\big[\,|\xi|^p\big]e^{-\kappa t_{m-1}}\big).
\end{aligned}
$$

□

Lemma 6.4.6. *(Gronwall's inequality). Let $T>0$ and let α,β and u be real-valued functions defined on $[0,T]$. Assume that α and u are continuous and that the negative part of β is* *integrable on every closed and bounded subinterval of* $[0,T]$. *If* α *is non-negative and if* u *satisfies the integral inequality*

$$u(t) \leq \beta(t) + \int_0^t \alpha(s)u(s)\mathrm{d}s, \quad \forall t \in [0,T],$$

then

$$u(t) \leq \beta(t) + \int_0^t \alpha(s)\beta(s)\exp\left(\int_s^t \alpha(r)\mathrm{d}r\right)\mathrm{d}s, \quad \forall t \in [0,T].$$

If we further have that β *is non-decreasing, then*

$$u(t) \leq \beta(t)\exp\left(\int_0^t \alpha(s)\mathrm{d}s\right), \quad \forall t \in [0,T].$$

The following auxiliary result is needed in the proof of Proposition 5.2.3.

Lemma 6.4.7. *Let* $c_1, c_2, c_3, c_4, c_5 > 0, C > 0$ *be real constants with* $c_1 + c_2 + c_3 < 1$. *Let* $(a_n)_{n\in\mathbb{N}}$ *be a real-valued sequence satisfying* $a_{n+3} \leq c_3 a_{n+2} + c_2 a_{n+1} + c_1 a_n + C$ *with initial values* $0 < a_1, a_2, a_3 < K$, *for some constant* $K > 0$. *Then there exist some constants* $K_1, K_2 > 0$ *(both are independent of* n*) such that for all* $n \geq 4$

$$a_n \leq K_1 + K_2 \max\{a_1, a_2, a_3\} e^{-(1-c_1-c_2-c_3)n/3},$$

Moreover, if $a_{n+3} \leq c_3 a_{n+2} + c_2 a_{n+1} + c_1 a_n + C + c_4 e^{-c_5 n}$ *and* $(c_1 + c_2 + c_3) \neq e^{-c_5}$, *then there exists constants* $K_3, K_4 > 0$ *(both are independent of* n*) such that for all* $n \geq 4$

$$a_n \leq K_3 + K_4 \max\{a_1, a_2, a_3\} e^{-(1-c_1-c_2-c_3)n/3}.$$

Proof. By the condition satisfied by the sequence $(a_n)_{n\in\mathbb{N}}$, we deduce for any $n \geq 1$

$$\begin{aligned}
a_{n+3} &\leq c_3 a_{n+2} + c_2 a_{n+1} + c_1 a_n + C \leq (c_1 + c_2 + c_3)\max\{a_n, a_{n+1}, a_{n+2}\} + C, \\
a_{n+4} &\leq c_3 a_{n+3} + c_2 a_{n+2} + c_1 a_{n+1} + C \leq (c_1 + c_2 + c_3)\max\{a_{n+1}, a_{n+2}, a_{n+3}\} + C \\
&\leq (c_1 + c_2 + c_3)\max\big\{a_{n+1}, a_{n+2}, (c_1 + c_2 + c_3)\max\{a_n, a_{n+1}, a_{n+2}\} + C\big\} + C \\
&\leq (c_1 + c_2 + c_3)\max\{a_n, a_{n+1}, a_{n+2}\} + 2C, \\
a_{n+5} &\leq c_3 a_{n+4} + c_2 a_{n+3} + c_1 a_{n+2} + C \leq (c_1 + c_2 + c_3)\max\{a_{n+2}, a_{n+3}, a_{n+4}\} + C \\
&\leq (c_1 + c_2 + c_3)\max\{a_n, a_{n+1}, a_{n+2}\} + 3C. \qquad (6.24)
\end{aligned}$$

Hence,

$$\max\{a_{n+3}, a_{n+4}, a_{n+5}\} \leq (c_1 + c_2 + c_3)\max\{a_n, a_{n+1}, a_{n+2}\} + 3C.$$

Consequently, adding $\frac{3C}{(c_1+c_2+c_3)-1} < 0$ on both sides, we observe

$$\begin{aligned}
\max\{a_{n+3}, &a_{n+4}, a_{n+5}\} + \frac{3C}{(c_1 + c_2 + c_3) - 1} \\
&\leq (c_1 + c_2 + c_3)\Big(\max\{a_n, a_{n+1}, a_{n+2}\} + \frac{3C}{(c_1 + c_2 + c_3) - 1}\Big).
\end{aligned}$$

Further, we derive for $n \geq 1$

$$\begin{aligned}
\max\{a_{3n+1}, &a_{3n+2}, a_{3n+3}\} + \frac{3C}{(c_1 + c_2 + c_3) - 1} \\
&\leq (c_1 + c_2 + c_3)^n\Big(\max\{a_1, a_2, a_3\} + \frac{3C}{(c_1 + c_2 + c_3) - 1}\Big),
\end{aligned}$$

which implies

$$\max\{a_{3n+1}, a_{3n+2}, a_{3n+3}\} \leq (c_1 + c_2 + c_3)^n \max\{a_1, a_2, a_3\} + \frac{3C}{1-(c_1+c_2+c_3)}$$
$$\leq e^{-(1-c_1-c_2-c_3)n} \max\{a_1, a_2, a_3\} + K,$$

for some $K > 0$, where we used the inequality $e^x \geq 1 + x$, for any $x \in \mathbb{R}$ and $c_1 + c_2 + c_3 < 1$.

Similarly, for the second claim, using the fact that $e^{-c_5(n+2)} < e^{-c_5(n+1)} < e^{-c_5 n}$, we derive as in (6.24):

$$\max\{a_{n+3}, a_{n+4}, a_{n+5}\} \leq (c_1 + c_2 + c_3) \max\{a_n, a_{n+1}, a_{n+2}\} + 3C + 3c_4 e^{-c_5 n}.$$

Adding $\frac{3C}{(c_1+c_2+c_3)-1} + \frac{3e^{-c_5(n+1)}}{(c_1+c_2+c_3)-e^{-c_5}}$ on both sides, we observe that

$$\max\{a_{3n+1}, a_{3n+2}, a_{3n+3}\} + \frac{3C}{(c_1+c_2+c_3)-1} + \frac{3e^{-c_5(n+1)}}{(c_1+c_2+c_3)-e^{-c_5}}$$
$$\leq (c_1+c_2+c_3)\Big(\max\{a_{3n-2}, a_{3n-1}, a_{3n}\} + \frac{3C}{(c_1+c_2+c_3)-1} + \frac{3e^{-c_5 n}}{(c_1+c_2+c_3)-e^{-c_5}}\Big).$$

Consequently, we have

$$\max\{a_{3n+1}, a_{3n+2}, a_{3n+3}\} + \frac{3C}{(c_1+c_2+c_3)-1} + \frac{3e^{-c_5(n+1)}}{(c_1+c_2+c_3)-e^{-c_5}}$$
$$\leq (c_1+c_2+c_3)^n\Big(\max\{a_1, a_2, a_3\} + \frac{3C}{(c_1+c_2+c_3)-1} + \frac{3e^{-c_5}}{(c_1+c_2+c_3)-e^{-c_5}}\Big),$$

and therefore

$$\max\{a_{3n+1}, a_{3n+2}, a_{3n+3}\}$$
$$\leq e^{-(1-c_1-c_2-c_3)n} \max\{a_1, a_2, a_3\} + \frac{3C}{1-(c_1+c_2+c_3)} + \frac{3e^{-c_5}}{|(c_1+c_2+c_3)-e^{-c_5}|}(1-e^{-c_5 n})$$
$$\leq e^{-(1-c_1-c_2-c_3)n} \max\{a_1, a_2, a_3\} + \tilde{K}_2,$$

for some positive constant $\tilde{K}_2$. □

6.4.4 Omitted residual terms of Section 5.6.2

In this part, we show the exact expectation form for the residual terms $R^2_{t_m}, R^3_{t_m}, R^5_{t_m}, R^7_{t_m}$ and $R^8_{t_m}$ in Lemma 5.6.3. The positive constant K below is independent of h, T, M and N and may have a different value in each line. For the residual term $R^2_{t_m}$, we have

$$\mathbb{E}[R^2_{t_m}]$$
$$= K\mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1}\int_{t_m}^{q_2}\sum_{\gamma\in\Pi_3^N}\Delta W^{\gamma_3}_{m,2h}\cdot\partial^5_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}u\Big(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_3}\big)\Big)\mathrm{d}W^{\gamma_1}_{q_3}\ \mathrm{d}q_2\ \mathrm{d}W^{\gamma_3}_{q_1}\Big]$$
$$+ K\mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1}\int_{t_m}^{q_2}\sum_{\gamma\in\Pi_3^N}\Delta W^{\gamma_3}_{m,2h}\cdot\partial^6_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}u\Big(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_3}\big)\Big)\mathrm{d}q_3\ \mathrm{d}q_2\ \mathrm{d}W^{\gamma_3}_{q_1}\Big].$$

For the residual term $R^3_{t_m}$, we have

$$\mathbb{E}[R^3_{t_m}]$$
$$=Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1}$$

$$
\sum_{\gamma\in\Pi_3^N} \Delta W_{m,2h}^{\gamma_3} \cdot \partial_{x_{\gamma_1}} \Big(\partial_{x_{\gamma_3}} B_{\gamma_2}(\overline{\boldsymbol{X}}_{q_2}^{N,h}) \partial^2_{x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_2}^{N,h}) \Big) \mathrm{d}W_{q_2}^{\gamma_1} \, \mathrm{d}W_{q_1}^{\gamma_3} \Big]
$$

$$
+ Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h} \int_{t_m}^{q_1} \sum_{\gamma\in\Pi_3^N} \Delta W_{m,2h}^{\gamma_3} \cdot \partial^2_{x_{\gamma_1},x_{\gamma_1}} \Big(\partial_{x_{\gamma_3}} B_{\gamma_2}(\overline{\boldsymbol{X}}_{q_2}^{N,h}) \partial^2_{x_{\gamma_2},x_{\gamma_3}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_2}^{N,h}) \Big) \mathrm{d}q_2 \, \mathrm{d}W_{q_1}^{\gamma_3} \Big]
$$

$$
+ Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h} \int_{t_m}^{q_1} \sum_{\gamma\in\Pi_3^N} \Delta W_{m,2h}^{\gamma_3} \cdot \partial_{x_{\gamma_1}} \Big(B_{\gamma_2}(\overline{\boldsymbol{X}}_{q_2}^{N,h}) \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_2}^{N,h}) \Big) \mathrm{d}W_{q_2}^{\gamma_1} \, \mathrm{d}W_{q_1}^{\gamma_3} \Big]
$$

$$
+ Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h} \int_{t_m}^{q_1} \sum_{\gamma\in\Pi_3^N} \Delta W_{m,2h}^{\gamma_3} \cdot \partial^2_{x_{\gamma_1},x_{\gamma_1}} \Big(B_{\gamma_2}(\overline{\boldsymbol{X}}_{q_2}^{N,h}) \partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_2}^{N,h}) \Big) \mathrm{d}q_2 \, \mathrm{d}W_{q_1}^{\gamma_3} \Big]
$$

$$
\begin{aligned}
+ K\mathbb{E}\Big[\int_{t_m}^{t_m+h} \int_{t_m}^{q_1} \int_{t_m}^{q_2} & \sum_{\gamma\in\Pi_5^N} \Big(\Delta W_{m,2h}^{\gamma_4} \Delta W_{m,2h}^{\gamma_5} \cdot \partial^5_{x_{\gamma_1},\dots,x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_3}^{N,h}) \, \mathrm{d}W_{q_3}^{\gamma_1} \, \mathrm{d}W_{q_2}^{\gamma_2} \, \mathrm{d}W_{q_1}^{\gamma_3} \\
&+ \Delta W_{m,2h}^{\gamma_4} \Delta W_{m,2h}^{\gamma_5} \cdot \partial^6_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4},x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_3}^{N,h}) \, \mathrm{d}q_3 \, \mathrm{d}W_{q_2}^{\gamma_2} \, \mathrm{d}W_{q_1}^{\gamma_3} \Big) \Big]
\end{aligned}
$$

$$
\begin{aligned}
+ K\mathbb{E}\Big[\int_{t_m}^{t_m+h} \int_{t_m}^{q_1} \sum_{\gamma\in\Pi_4^N} & \Big(\Delta W_{m,2h}^{\gamma_3} \Delta W_{m,2h}^{\gamma_4} \cdot \partial^5_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_2}^{N,h}) \mathrm{d}q_2 \, \mathrm{d}W_{q_1}^{\gamma_2} \\
&+ \Delta W_{m,2h}^{\gamma_3} \Delta W_{m,2h}^{\gamma_4} \cdot \partial^5_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_2}^{N,h}) \mathrm{d}W_{q_2}^{\gamma_1} \, \mathrm{d}q_1 \\
&+ \Delta W_{m,2h}^{\gamma_3} \Delta W_{m,2h}^{\gamma_4} \cdot \partial^6_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_2}^{N,h}) \mathrm{d}q_2 \, \mathrm{d}q_1 \Big) \Big].
\end{aligned}
$$

For the residual term $R^5_{t_m}$, we have

$$
\begin{aligned}
\mathbb{E}\big[R^5_{t_m}\big] = & \sum_{\gamma\in\Pi_5^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h} \partial_{x_{\gamma_1}} \Big(\Delta W_{m,2h}^{\gamma_2} \Delta W_{m,2h}^{\gamma_3} \Delta W_{m,2h}^{\gamma_4} \Delta W_{m,2h}^{\gamma_5} \cdot \partial^4_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4},x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_1}^{N,h}) \Big) \mathrm{d}W_{q_1}^{\gamma_1} \Big] \\
&+ \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[\int_{t_m}^{t_m+h} \partial^2_{x_{\gamma_1},x_{\gamma_1}} \Big(\Delta W_{m,2h}^{\gamma_2} \Delta W_{m,2h}^{\gamma_3} \Delta W_{m,2h}^{\gamma_4} \Delta W_{m,2h}^{\gamma_5} \cdot \partial^4_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_4},x_{\gamma_5}} u(t_{m+1}, \overline{\boldsymbol{X}}_{q_1}^{N,h}) \Big) \mathrm{d}q_1 \Big] \\
&+ Kh \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[B_{\gamma_1}(\boldsymbol{X}_{t_m}^{N,h}) \Delta W_{m,2h}^{\gamma_2} \Delta W_{m,2h}^{\gamma_3} \Delta W_{m,2h}^{\gamma_4} \cdot \partial^4_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \boldsymbol{X}_{t_m}^{N,h}) \Big] \\
&+ Kh^2 \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[B_{\gamma_1}(\boldsymbol{X}_{t_m}^{N,h}) B_{\gamma_2}(\boldsymbol{X}_{t_m}^{N,h}) \Delta W_{m,2h}^{\gamma_3} \Delta W_{m,2h}^{\gamma_4} \cdot \partial^4_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \boldsymbol{X}_{t_m}^{N,h}) \Big] \\
&+ Kh^3 \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[B_{\gamma_1}(\boldsymbol{X}_{t_m}^{N,h}) B_{\gamma_2}(\boldsymbol{X}_{t_m}^{N,h}) B_{\gamma_3}(\boldsymbol{X}_{t_m}^{N,h}) \Delta W_{m,2h}^{\gamma_4} \cdot \partial^4_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \boldsymbol{X}_{t_m}^{N,h}) \Big] \\
&+ Kh^4 \sum_{\gamma\in\Pi_4^N} \mathbb{E}\Big[B_{\gamma_1}(\boldsymbol{X}_{t_m}^{N,h}) B_{\gamma_2}(\boldsymbol{X}_{t_m}^{N,h}) B_{\gamma_3}(\boldsymbol{X}_{t_m}^{N,h}) B_{\gamma_4}(\boldsymbol{X}_{t_m}^{N,h}) \cdot \partial^4_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}} u(t_{m+1}, \boldsymbol{X}_{t_m}^{N,h}) \Big].
\end{aligned}
$$

For the residual term $R^7_{t_m}$, we have

$$\begin{aligned}
&\mathbb{E}\big[R^7_{t_m}\big]\\
=&Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1}\int_{t_m}^{q_2}\sum_{\gamma\in\Pi_5^N}\partial^5_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4},x_{\gamma_5}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_3})\mathrm{d}W^{\gamma_1}_{q_3}\ \mathrm{d}W^{\gamma_2}_{q_2}\ \mathrm{d}W^{\gamma_3}_{q_1}\Big]\\
&+Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1}\int_{t_m}^{q_2}\sum_{\gamma\in\Pi_5^N}\partial^6_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4},,x_{\gamma_5}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_3})\mathrm{d}q_3\ \mathrm{d}W^{\gamma_2}_{q_2}\ \mathrm{d}W^{\gamma_3}_{q_1}\Big]\\
&+Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1}\sum_{\gamma\in\Pi_4^N}\partial^5_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_2})\mathrm{d}q_2\ \mathrm{d}W^{\gamma_2}_{q_1}\Big]\\
&+Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1}\sum_{\gamma\in\Pi_4^N}\partial^5_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_2})\mathrm{d}W^{\gamma_1}_{q_2}\ \mathrm{d}q_1\Big]\\
&+Kh\mathbb{E}\Big[\int_{t_m}^{t_m+h}\int_{t_m}^{q_1}\sum_{\gamma\in\Pi_4^N}\partial^6_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_4}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_2})\mathrm{d}q_2\ \mathrm{d}q_1\Big].
\end{aligned}$$

For the residual term $R^8_{t_m}$, we have

$$\begin{aligned}
\mathbb{E}\big[R^8_{t_m}\big]=&Kh^2\mathbb{E}\Big[\int_{t_m}^{t_m+h}\sum_{\gamma\in\Pi_3^N}\partial_{x_{\gamma_1}}\Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_1})\Big)\ \mathrm{d}W^{\gamma_1}_{q_1}\Big]\\
&+Kh^2\mathbb{E}\Big[\int_{t_m}^{t_m+h}\sum_{\gamma\in\Pi_3^N}\partial^2_{x_{\gamma_1},x_{\gamma_1}}\Big(B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\partial^3_{x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_1})\Big)\ \mathrm{d}q_1\Big]\\
&+Kh^2\mathbb{E}\Big[\int_{t_m}^{t_m+h}\sum_{\gamma\in\Pi_3^N}\partial_{x_{\gamma_1}}\Big(\partial_{x_{\gamma_3}}B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\partial^2_{x_{\gamma_2},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_1})\Big)\ \mathrm{d}W^{\gamma_1}_{q_1}\Big]\\
&+Kh^2\mathbb{E}\Big[\int_{t_m}^{t_m+h}\sum_{\gamma\in\Pi_3^N}\partial^2_{x_{\gamma_1},x_{\gamma_1}}\Big(\partial_{x_{\gamma_3}}B_{\gamma_2}(\overline{\boldsymbol{X}}^{N,h}_{q_1})\partial^2_{x_{\gamma_2},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_1})\Big)\ \mathrm{d}q_1\Big]\\
&+Kh^2\mathbb{E}\Big[\int_{t_m}^{t_m+h}\sum_{\gamma\in\Pi_3^N}\partial^5_{x_{\gamma_1},x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_1})\ \mathrm{d}W^{\gamma_1}_{q_1}\Big]\\
&+Kh^2\mathbb{E}\Big[\int_{t_m}^{t_m+h}\sum_{\gamma\in\Pi_3^N}\partial^6_{x_{\gamma_1},x_{\gamma_1},x_{\gamma_2},x_{\gamma_2},x_{\gamma_3},x_{\gamma_3}}u(t_{m+1},\overline{\boldsymbol{X}}^{N,h}_{q_1})\ \mathrm{d}q_1\Big].
\end{aligned}$$

Bibliography

[1] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.*, 77:137–185, 2005.

[2] D. Adams, G. dos Reis, R. Ravaille, W. Salkeld, and J. Tugaut. Large Deviations and Exit-times for reflected McKean-Vlasov equations with self-stabilizing terms and superlinear drifts. *Stochastic Processes and their Applications*, 2022.

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

[4] 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.

[5] A. Agarwal, A. Amato, G. dos Reis, and S. Pagliarani. Numerical approximation of McKean-Vlasov SDEs via stochastic gradient descent. *arXiv preprint arXiv:2310.13579*, 2023.

[6] 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.

[7] S. M. Ahn and S.-Y. Ha. Stochastic flocking dynamics of the Cucker-Smale model with multiplicative white noises. *J. Math. Phys.*, 51(10):103301, 17, 2010.

[8] F. Antonelli and A. Kohatsu-Higa. Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *The Annals of Applied Probability*, 12(2):423–476, 2002.

[9] 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.

[10] V. Bally and D. Talay. The Euler scheme for stochastic differential equations: error analysis with Malliavin calculus. *Mathematics and Computers in Simulation*, 38(1-3):35–41, 1995.

[11] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probability Theory and Related Fields*, 104(1):43–60, 1996.

[12] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. *Monte Carlo Methods and Applications*, 2(2):93–128, 1996.

[13] J. Bao, C. Reisinger, P. Ren, and W. Stockinger. First-order convergence of Milstein schemes for McKean-Vlasov equations and interacting particle systems. *Proceedings of the Royal Society A*, 477(2245):20200258, 2021.

[14] J. Bao, C. Reisinger, P. Ren, and W. Stockinger. Milstein schemes and antithetic multilevel Monte Carlo sampling for delay McKean–Vlasov equations and interacting particle systems. *IMA Journal of Numerical Analysis*, page drad064, 09 2023.

[15] M. Bauer, T. Meyer-Brandis, and F. Proske. Strong solutions of mean-field stochastic differential equations with irregular drift. *Electron. J. Probab.*, 23:Paper No. 132, 35, 2018.

[16] 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.

[17] D. Belomestny, V. Pilipauskaite, and M. Podolskij. Semiparametric estimation of mckean-vlasov sdes. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 59(1):79–96, Feb. 2023.

[18] D. Belomestny and J. Schoenmakers. Projected particle methods for solving McKean-Vlasov stochastic differential equations. *SIAM Journal on Numerical Analysis*, 56(6):3169–3195, 2018.

[19] S. Benachour, B. Roynette, D. Talay, and P. Vallois. Nonlinear self-stabilizing processes. i. existence, invariant probability, propagation of chaos. *Stochastic Processes and their Applications*, 75(2):173–201, 1998.

[20] O. Bencheikh and B. Jourdain. Bias behaviour and antithetic sampling in mean-field particle approximations of SDEs nonlinear in the sense of McKean. *ESAIM: Proceedings and Surveys*, 65:219–235, 2019.

[21] F. Bernal, G. dos Reis, and G. Smith. Hybrid PDE solver for data-driven problems and modern branching. *European Journal of Applied Mathematics*, 28(6):949–972, 2017.

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

[23] S. Biswas, C. Kumar, Neelima, G. d. Reis, and C. Reisinger. An explicit Milstein-type scheme for interacting particle systems and McKean–Vlasov SDEs with common noise and non-differentiable drift coefficients. *The Annals of Applied Probability*, 34(2):2326–2363, 2024.

[24] S. Biswas, Neelima, C. Kumar, G. dos Reis, and C. Reisinger. An explicit Milstein-type scheme for interacting particle systems and McKean–Vlasov SDEs with common noise and non-differentiable drift coefficients. *The Annals of Applied Probability*, 34(2):2326–2363, 2024.

[25] F. Bolley, J. A. Cañizo, and J. A. Carrillo. Stochastic mean-field limit: non-Lipschitz forces and swarming. *Mathematical Models and Methods in Applied Sciences*, 21(11):2179–2210, 2011.

[26] G. Borghi, M. Herty, and L. Pareschi. Constrained consensus-based optimization. *SIAM Journal on Optimization*, 33(1):211–236, 2023.

[27] 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". *Journal of mathematical neuroscience*, 5(1):1–23, 2015.

[28] M. Bossy and D. Talay. Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation. *The Annals of Applied Probability*, 6(3):818–861, 1996.

[29] M. Bossy and D. Talay. A stochastic particle method for the McKean–Vlasov and the Burgers equation. *Mathematics of Computation*, 66(217):157–192, 1997.

[30] N. Bou-Rabee and K. Schuh. Convergence of unadjusted Hamiltonian Monte Carlo for mean-field models. *Electronic Journal of Probability*, 28:Paper No. 91, 40, 2023.

[31] N. Bou-Rabee and K. Schuh. Nonlinear Hamiltonian Monte Carlo & its particle approximation. *arXiv preprint arXiv:2308.11491*, 2023.

[32] E. Buckwar, A. Samson, M. Tamborrino, and I. Tubikanec. A splitting method for SDEs with locally Lipschitz drift: Illustration on the FitzHugh-Nagumo model. *Applied Numerical Mathematics*, 179:191–220, 2022.

[33] R. Carmona. *Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications*, volume 1 of *Financial Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016.

[34] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I*, volume 84 of *Probability Theory and Stochastic Modelling*. Springer International Publishing, 1st edition, 2017.

[35] R. Carmona, F. Delarue, and D. Lacker. Mean field games with common noise. *The Annals of Probability*, Vol. 44(6), 2016.

[36] J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. An analytical framework for consensus-based global optimization method. *Mathematical Models and Methods in Applied Sciences*, 28(06):1037–1066, 2018.

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

[38] 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.

[39] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matematica Iberoamericana*, 19(3):971–1018, 2003.

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

[41] L.-P. Chaintron and A. Diez. Propagation of chaos: a review of models, methods and applications. I. models and methods. *Kinetic & Related Models*, 15(6), 2022.

[42] J.-F. Chassagneux, L. Szpruch, and A. Tse. Weak quantitative propagation of chaos via differential calculus on the space of measures. *The Annals of Applied Probability*, 32(3):1929–1969, 2022.

[43] P.-E. Chaudru de Raynal and N. Frikha. Well-posedness for some non-linear SDEs and related PDE on the Wasserstein space. *Journal de Mathématiques Pures et Appliquées*, 159:1–167, 2022.

[44] P. E. Chaudru de Raynal and C. A. Garcia Trillos. A cubature based algorithm to solve decoupled McKean–Vlasov forward-backward stochastic differential equations. *Stochastic Processes and their Applications*, 125(6):2206–2255, 2015.

[45] F. Chen, Y. Lin, Z. Ren, and S. Wang. Uniform-in-time propagation of chaos for kinetic mean field Langevin dynamics. *Electronic Journal of Probability*, 29:Paper No. 17, 43, 2024.

[46] F. Chen, Z. Ren, and S. Wang. Uniform-in-time propagation of chaos for mean field Langevin dynamics. *arXiv preprint arXiv:2212.03050*, 2022.

[47] X. Chen and G. dos Reis. A flexible split-step scheme for solving McKean-Vlasov stochastic differential equations. *Applied Mathematics and Computation*, 427:Paper No. 127180, 23, 2022.

[48] X. Chen and G. dos Reis. Euler simulation of interacting particle systems and McKean–Vlasov SDEs with fully superlinear growth drifts in space and interaction. *IMA Journal of Numerical Analysis*, page drad022, 05 2023.

[49] X. Chen, G. dos Reis, and W. Stockinger. Wellposedness, exponential ergodicity and numerical approximation of fully super-linear McKean–Vlasov SDEs and associated particle systems. *arXiv preprint arXiv:2302.05133*, 2023.

[50] X. Chen, G. dos Reis, W. Stockinger, and Z. Wilde. Improved weak convergence for the long time simulation of Mean-field Langevin equations. *arXiv preprint arXiv:2405.01346*, 2024.

[51] L. Chizat and F. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. *Advances in Neural Information Processing systems*, 31, 2018.

[52] J. Claisse, G. Conforti, Z. Ren, and S. Wang. Mean field optimization problem regularized by Fisher information. *arXiv preprint arXiv:2302.05938*, 2023.

[53] F. Comte and V. Genon-Catalot. Nonparametric adaptive estimation for interacting particle systems. *Scandinavian Journal of Statistics*, 50(4):1716–1755, 2023.

[54] D. Crisan and E. McMurray. Cubature on Wiener space for McKean-Vlasov SDEs with smooth scalar interaction. *The Annals of Applied Probability*, 29(1):130–177, 2019.

[55] V. De Bortoli, A. Durmus, X. Fontaine, and U. Simsekli. Quantitative propagation of chaos for SGD in wide neural networks. *Advances in Neural Information Processing Systems*, 33:278–288, 2020.

[56] P.-E. C. de Raynal, M. H. Duong, P. Monmarché, M. Tomašević, and J. Tugaut. Reducing exit-times of diffusions with repulsive interactions. *ESAIM: Probability and Statistics*, 27:723–748, 2023.

[57] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: a central limit theorem. *Electronic Journal of Probability*, 24:Paper No. 51, 54, 2019.

[58] F. Delarue and A. Tse. Uniform in time weak propagation of chaos on the torus. *arXiv preprint arXiv:2104.14973*, 2021.

[59] X. Ding and H. Qiao. Euler–Maruyama Approximations for Stochastic McKean-Vlasov Equations with Non-Lipschitz Coefficients. *Journal of Theoretical Probability*, pages 1–18, 2020.

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

[61] G. dos Reis, W. Salkeld, and J. Tugaut. Freidlin-Wentzell LDP in path space for McKean-Vlasov equations and the functional iterated logarithm law. *The Annals of Applied Probability*, 29(3):1487–1540, 2019.

[62] G. dos Reis, G. Smith, and P. Tankov. Importance sampling for McKean-Vlasov SDEs. *Applied Mathematics and Computation*, 453:128078, 2023.

[63] W. Dreyer, M. Gaberšček, C. Guhlke, R. Huth, and J. Jamnik. Phase transition in a rechargeable lithium battery. *European Journal of Applied Mathematics*, 22(3):267–290, 2011.

[64] R. Duan, M. Fornasier, and G. Toscani. A kinetic flocking model with diffusion. *Communications in Mathematical Physics*, 300(1):95–145, 2010.

[65] A. Durmus, A. Eberle, A. Guillin, and R. Zimmer. An elementary approach to uniform in time propagation of chaos. *Proceedings of the American Mathematical Society*, 148:12–43, 2020.

[66] S. Eilenberg and S. M. Lane. On the groups H (π, n), I. *Annals of Mathematics*, pages 55–106, 1953.

[67] R. Erban, J. Haškovec, and Y. Sun. A Cucker–Smale Model with Noise and Delay. *SIAM Journal on Applied Mathematics*, 76(4):1535–1557, 2016.

[68] X. Erny, E. Löcherbach, and D. Loukianova. Strong error bounds for the convergence to its mean field limit for systems of interacting neurons in a diffusive scaling. *The Annals of Applied Probability*, 33(5):3563–3586, 2023.

[69] W. Fang and M. B. Giles. Adaptive Euler–Maruyama method for SDEs with nonglobally Lipschitz drift. *Annals of Applied Probability*, 30(2):526–560, 2020.

[70] J. M. Foster, G. dos Reis, and C. Strange. High order splitting methods for SDEs satisfying a commutativity condition. *SIAM Journal on Numerical Analysis*, 62(1):500–532, 2024.

[71] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, 2015.

[72] N. Fournier, M. Hauray, and S. Mischler. Propagation of chaos for the 2D viscous vortex model. *Journal of the European Mathematical Society*, 16(7):1423–1466, 2014.

[73] A. Friedman. *Stochastic differential equations and applications. Vol. 1.* Probability and Mathematical Statistics, Vol. 28. Academic Press, New York-London, 1975.

[74] V. Genon-Catalot and C. Larédo. Probabilistic properties and parametric inference of small variance nonlinear self-stabilizing stochastic differential equations. *Stochastic Processes and their Applications*, 142:513–548, 2021.

[75] E. Gobet and S. Pagliarani. Analytical approximations of non-linear SDEs of McKean-Vlasov type. *Journal of Mathematical Analysis and Applications*, 466(1):71–106, 2018.

[76] S. N. Gomes, G. A. Pavliotis, and U. Vaes. Mean field limits for interacting diffusions with colored noise: phase transitions and spectral numerical methods. *Multiscale Modeling & Simulation. A SIAM Interdisciplinary Journal*, 18(3):1343–1370, 2020.

[77] C. Guhlke, P. Gajewski, M. Maurelli, P. K. Friz, and W. Dreyer. Stochastic many-particle model for LFP electrodes. *Continuum Mechanics and Thermodynamics*, 30(3):593–628, 2018.

[78] A. Guillin, W. Liu, L. Wu, and C. Zhang. Uniform Poincaré and logarithmic Sobolev inequalities for mean field particle systems. *The Annals of Applied Probability*, 32(3):1590–1614, 2022.

[79] A. Guillin and P. Monmarché. Uniform long-time and propagation of chaos estimates for mean field kinetic particles in non-convex landscapes. *Journal of Statistical Physics*, 185(2):Paper No. 15, 20, 2021.

[80] S.-Y. Ha, J. Jeong, S. E. Noh, Q. Xiao, and X. Zhang. Emergent dynamics of Cucker-Smale flocking particles in a random environment. *Journal of Differential Equations*, 262(3):2554–2591, 2017.

[81] A.-L. Haji-Ali, H. Hoel, and R. Tempone. A simple approach to proving the existence, uniqueness, and strong and weak convergence rates for a broad class of McKean–Vlasov equations. *arXiv preprint arXiv:2101.00886*, 2021.

[82] F. Harang and A. Mayorcas. Pathwise regularisation of singular interacting particle systems and their mean field limits. *Stochastic Processes and their Applications*, 159:499–540, 2023.

[83] S. Herrmann, P. Imkeller, and D. Peithmann. Large deviations and a Kramers' type law for self-stabilizing diffusions. *The Annals of Applied Probability*, 18(4):1379–1423, 2008.

[84] S. Herrmann and J. Tugaut. Self-stabilizing processes: uniqueness problem for stationary measures and convergence rate in the small-noise limit. *ESAIM. Probability and Statistics*, 16:277–305, 2012.

[85] D. J. Higham, X. Mao, and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM Journal on Numerical Analysis*, 40(3):1041–1063, 2002.

[86] K. Hu, Z. Ren, D. Šiška, and L. Szpruch. Mean-field Langevin dynamics and energy landscape of neural networks. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 57(4):2043–2065, 2021.

[87] X. Huang, P. Ren, and F.-Y. Wang. Distribution dependent stochastic differential equations. *Frontiers of Mathematics in China*, 16(2):257–301, 2021.

[88] M. Hutzenthaler and A. Jentzen. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *Memoirs of the American Mathematical Society*, 236(1112):v+99, 2015.

[89] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 467(2130):1563–1576, 2011.

[90] M. Hutzenthaler, T. Kruse, and T. A. Nguyen. Multilevel Picard approximations for McKean-Vlasov stochastic differential equations. *arXiv preprint arXiv:2103.10870*, 2021.

[91] M. Hutzenthaler, T. Kruse, and T. A. Nguyen. Multilevel Picard approximations for McKean-Vlasov stochastic differential equations. *Journal of Mathematical Analysis and Applications*, 507(1):Paper No. 125761, 14, 2022.

[92] S. Jin, L. Li, and J.-G. Liu. Random batch methods (RBM) for interacting particle systems. *Journal of Computational Physics*, 400:108877, 30, 2020.

[93] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197. University of California Press, Berkeley and Los Angeles, 1956.

[94] A. Kalinin, T. Meyer-Brandis, and F. Proske. Stability, uniqueness and existence of solutions to Mckean–Vlasov Sdes in arbitrary moments. *arXiv preprint arXiv:2205.02176*, 2022.

[95] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[96] A. Kohatsu-Higa and S. Ogawa. Weak rate of convergence for an Euler scheme of nonlinear SDE's. *Monte Carlo Methods and Applications*, 3(4):327–345, 1997.

[97] Y. Kook, M. S. Zhang, S. Chewi, M. A. Erdogdu, et al. Sampling from the mean-field stationary distribution. *arXiv preprint arXiv:2402.07355*, 2024.

[98] C. Kumar and Neelima. On explicit Milstein-type scheme for McKean-Vlasov stochastic differential equations with super-linear drift coefficient. *Electronic Journal of Probability*, 26:Paper No. 111, 32, 2021.

[99] C. Kumar, Neelima, C. Reisinger, and W. Stockinger. Well-posedness and tamed schemes for McKean–Vlasov equations with common noise. *The Annals of Applied Probability*, 32(5):3283–3330, 2022.

[100] D. Lacker. On a strong form of propagation of chaos for McKean-Vlasov equations. *Electronic Communications in Probability*, 23, 2018.

[101] D. Lacker. Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions. *Probability and Mathematical Physics*, 4(2):377–432, 2023.

[102] D. Lacker and L. Le Flem. Sharp uniform-in-time propagation of chaos. *Probability Theory and Related Fields*, pages 1–38, 2023.

[103] B. Leimkuhler, C. Matthews, and M. Tretyakov. On the long-time integration of stochastic gradient systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 470(2170):20140120, 2014.

[104] G. Leobacher, C. Reisinger, and W. Stockinger. Well-posedness and numerical schemes for McKean-Vlasov equations and interacting particle systems with discontinuous drift. *BIT Numerical Mathematics*, 62:1505–1549, 2022.

[105] D.-H. Li and W. Cheng. Recent progress in the global convergence of quasi-Newton methods for nonlinear equations. *Hokkaido Mathematical Journal*, 36(4):729–743, 2007.

[106] Y. Li, X. Mao, Q. Song, F. Wu, and G. Yin. Strong convergence of Euler–Maruyama schemes for McKean–Vlasov stochastic differential equations under local Lipschitz conditions of state variables. *IMA Journal of Numerical Analysis*, 2022.

[107] A. Lionnet, G. dos Reis, and L. Szpruch. Time discretization of FBSDE with polynomial growth drivers and reaction-diffusion PDEs. *The Annals of Applied Probability*, 25(5):2563–2625, 2015.

[108] R. Maillet. A note on the long-time behaviour of stochastic McKean-Vlasov equations with common noise. *arXiv preprint arXiv:2306.16130*, 2023.

[109] F. Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Processes and their Applications*, 95(1):109–132, 2001.

[110] F. Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *The Annals of Applied Probability*, 13(2):540–560, 2003.

[111] 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.

[112] V. Malyshkin, editor. *Parallel Computing Technologies 15th International Conference, PaCT 2019, Almaty, Kazakhstan, August 19–23, 2019, Proceedings*. Theoretical Computer Science and General Issues ; 11657. Springer International Publishing : Imprint: Springer, Cham, 1st ed. 2019.. edition, 2019.

[113] X. Mao. *Stochastic Differential Equations and Applications*. Horwood, 2008.

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

[115] H. P. McKean, Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56:1907–1911, 1966.

[116] S. Mehri, M. Scheutzow, W. Stannat, and B. Z. Zangeneh. Propagation of chaos for stochastic spatially structured neuronal networks with delay driven by jump diffusions. *Annals of Applied Probability*, 30(1):175–207, 2020.

[117] S. Mei, A. Montanari, and P. Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33), 2018.

[118] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations*, pages 42–95. Springer, 1996.

[119] M. A. Mezerdi. On the convergence of Carathéodory numerical scheme for McKean–Vlasov equations. *Stochastic Analysis and Applications*, 39(5):804–818, 2021.

[120] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics.* Scientific Computation. Springer, Cham, second edition, [2021] ©2021.

[121] Y. Mishura and A. Veretennikov. Existence and uniqueness theorems for solutions of mckean–vlasov stochastic equations. *Theory of Probability and Mathematical Statistics*, 2020.

[122] R. Naito and T. Yamada. A higher order weak approximation of McKean–Vlasov type SDEs. *BIT. Numerical Mathematics*, 62(2):521–559, 2022.

[123] Neelima, S. Biswas, C. Kumar, G. dos Reis, and C. Reisinger. Well-posedness and tamed Euler schemes for McKean–Vlasov equations driven by Lévy noise. *arXiv preprint arXiv:2010.08585*, 2020.

[124] E. Platen and N. Bruti-Liberati. *Numerical solution of stochastic differential equations with jumps in finance*, volume 64. Springer Science & Business Media, 2010.

[125] 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. *Journal of Computational and Applied Mathematics*, 400:113725, 2022.

[126] P. Ren and F.-Y. Wang. Exponential convergence in entropy and Wasserstein for McKean–Vlasov SDEs. *Nonlinear Analysis*, 206:112259, 2021.

[127] M. Röckner and X. Zhang. Well-posedness of distribution dependent SDEs with singular drifts. *Bernoulli*, 27(2):1131–1158, 2021.

[128] S. Sabanis and Y. Zhang. On explicit order 1.5 approximations with varying coefficients: The case of super-linear diffusion coefficients. *Journal of Complexity*, 50:84–115, Feb 2019.

[129] J. Sirignano and K. Spiliopoulos. Mean field analysis of deep neural networks. *Mathematics of Operations Research*, 47(1):120–152, 2022.

[130] 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.

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

[132] T. Suzuki, D. Wu, and A. Nitanda. Mean-field Langevin dynamics: Time-space discretization, stochastic gradient, and variance reduction. *Advances in Neural Information Processing Systems*, 36, 2024.

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

[134] L. Szpruch and A. Tse. Antithetic multilevel sampling method for nonlinear functionals of measure. *The Annals of Applied Probability*, 31(3):1100–1139, 2021.

[135] D. Talay. Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics: An International Journal of Probability and Stochastic Processes*, 29(1):13–36, 1990.

[136] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic analysis and applications*, 8(4):483–509, 1990.

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

[138] J. Tugaut. Convergence to the equilibria for self-stabilizing processes in double-well landscape. *The Annals of Probability*, 41(3A):1427–1460, 2013.

[139] G. Vilmart. Postprocessed integrators for the high order integration of ergodic SDEs. *SIAM journal on scientific computing*, 37(1):A201–A220, 2015.

[140] F.-Y. Wang. Distribution dependent SDEs for Landau type equations. *Stochastic Processes and their Applications*, 128(2):595–621, 2018.

[141] W. Wei, T. Gao, X. Chen, and J. Duan. An optimal control method to compute the most likely transition path for stochastic dynamical systems with jumps. *Chaos*, 32(5):Paper No. 051102, 10, 2022.

[142] E. Zeidler. *Nonlinear functional analysis and its applications. II/A*. Springer-Verlag, New York, 1990. Linear monotone operators, Translated from the German by the author and Leo F. Boron.

[143] S.-Q. Zhang. Existence and non-uniqueness of stationary distributions for distribution dependent SDEs. *Electronic Journal of Probability*, 28:1–34, 2023.

[144] X. Zong, F. Wu, and C. Huang. Convergence and stability of the semi-tamed Euler scheme for stochastic differential equations with non-Lipschitz continuous coefficients. *Applied Mathematics and Computation*, 228:240–250, 2014.