Numerical methods for non-standard McKean–Vlasov type equations



Wolfgang Stockinger
St Catherine's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2022

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Prof. Christoph Reisinger for his guidance, patience and support during my DPhil studies. His way of supervision helped me not only to grow as a researcher, but also as a person.

I would also like to thank Prof. Gerhard Larcher and Prof. Gunther Leobacher for our joint collaborations over the past few years, as well as their encouragement, advice, help and numerous discussions about mathematics.

I am extremely grateful and thankful to my friends and colleagues, Mandy, Christina, Taylor, Jing, Yufei, Saad, Victor, Shuang, Endre, Kan and Ikuya with whom I had the chance to share my DPhil experience and my life in Oxford. I wish to say a special thank you to Yufei Zhang for numerous endless discussions about mathematics and research in general and for our fruitful collaboration.

Finally and most importantly, I am infinitely grateful to my parents, Gerlinde and Wilhelm, my siblings, Margot, Kathrin and Bernhard, and my girlfriend Karen for their limitless support and encouragement. Without them my studies and this thesis would not have been possible.

Statement of Originality

I hereby declare that this thesis contains no material which has been accepted or is currently being submitted for any other degree, diploma, certificate or other qualifications at the University of Oxford or elsewhere.

The thesis contains the two published papers [132, 15] whose development was the principal responsibility of myself: the first one, supervised by Christoph Reisinger, is published in *Journal of Computational and Applied Mathematics* [132] (Chapter 2) and the second one, which is joint work with Jianhai Bao, Panpan Ren and Christoph Reisinger, in *Proceedings of the Royal Society A* [15] (Chapter 3). Chapter 4 is based on the submitted work [104] (supervised by Gunther Leobacher and Christoph Reisinger) and Chapter 5 contains my contributions to the submitted paper [136] (joint work with Christoph Reisinger and Yufei Zhang).

The development and writing up of the thesis were the principal responsibility of myself, Wolfgang Stockinger, working under the supervision of Prof. Christoph Reisinger.

Abstract

In this thesis, we consider numerical aspects concerning the simulation and strong approximation of solutions to McKean–Vlasov SDEs and interacting particle systems under non-standard assumptions, as well as numerical schemes for a broad class of mean-field control problems. First, we introduce novel adaptive time-stepping schemes and tamed Milstein schemes with the aim to achieve stable and strongly convergent discrete-time approximations to McKean–Vlasov equations and associated particle systems, where the drift and diffusion coefficients are allowed to grow super-linearly. The performance of the proposed time-stepping schemes is intensively investigated for various examples, including important models from neuroscience. Next, we study numerical schemes, as well as strong well-posedness, for a class of one-dimensional McKean–Vlasov SDEs with drifts that have a finite number of spatial discontinuities. We propose a transformation technique in combination with a fixed-point argument to show strong existence and uniqueness results. Further, we discuss different particle approximation methods and time-stepping schemes for such McKean–Vlasov equations and prove strong convergence results in the number of particles and time-steps, and provide numerical tests for models appearing in neuroscience and systemic risk. Finally, we develop a novel algorithm to obtain approximations of optimal feedback controls for control problems where the controlled state process is of McKean–Vlasov type. Here, we will employ deep-learning methods to solve these problems relying on a suitable formulation based on partial differential equations. An intensive numerical study of the proposed method is performed for high-dimensional mean-field Cucker–Smale models.

Contents

1 Introduction **1**

2 An adaptive Euler–Maruyama scheme for McKean–Vlasov SDEs with super-linear growth and application to the mean-field FitzHugh–Nagumo model **10**

2.1 Introduction 10

2.2 Time-stepping scheme for super-linear drift coefficient 15

2.2.1 Problem formulation 15

2.2.2 Adaptive Euler–Maruyama scheme for super-linear drift 17

2.2.3 Choice of time-step function and main result 18

2.3 Super-linear diffusion and higher order schemes 21

2.3.1 Super-linear diffusion coefficient 21

2.3.2 Adaptive Milstein scheme 22

2.4 Examples and numerical illustration 24

2.4.1 Example 1 – the Fang and Giles test with added mean-field term 25

2.4.2 Example 2 – Ginzburg-Landau with added mean-field term 28

2.4.3 Example 3 – Kuramoto model with added non-Lipschitz drift 29

2.4.4 Example 4 – FitzHugh–Nagumo model 30

2.4.5 Example 5 – Milstein scheme 32

2.5 Proofs of main results 33

2.5.1 Moment stability for super-linear drift coefficient 33

2.5.2 Convergence of Scheme 2 (Proof of Theorem 2.2.1) 35

2.5.3 Moment stability for super-linear diffusion coefficient 38

2.5.4 Moment stability of Milstein scheme 39

3 First order convergence of Milstein schemes for McKean–Vlasov equations and interacting particle systems **41**

3.1 Introduction 41

3.2 Tamed Milstein schemes for non-Lipschitz McKean–Vlasov SDEs 46

3.2.1 Assumptions and interacting particle system 46

3.2.2 Time-stepping schemes and main results . . . 47
3.2.3 Milstein scheme for globally Lipschitz drift . . . 51
3.3 Proof of results . . . 53
3.3.1 Proof of Lemma 3.2.1 . . . 54
3.3.2 Some auxiliary lemmas . . . 57
3.3.3 Proof of Theorem 3.2.1 . . . 60
3.3.4 Proof of Proposition 3.2.1 . . . 61
3.4 Numerical results and implementation details . . . 62
3.4.1 Examples . . . 63
3.4.2 Time-stepping schemes and L-derivative . . . 64
3.4.3 Propagation of chaos . . . 67

4 Well-posedness and numerical schemes for one-dimensional McKean–Vlasov equations and interacting particle systems with discontinuous drift 69
4.1 Introduction . . . 69
4.2 Preliminaries . . . 73
4.2.1 Notions and notation . . . 74
4.2.2 Properties of the transformation G . . . 75
4.3 Existence and uniqueness results . . . 75
4.3.1 McKean–Vlasov SDEs and interacting particle systems with decomposable drift 76
4.3.2 McKean–Vlasov SDE with non-decomposable drift . . . 79
4.3.3 Interacting particle system with non-decomposable drift . . . 85
4.4 Euler–Maruyama scheme with and without transformation . . . 90
4.4.1 Scheme 1: Euler–Maruyama after transformation (decomposable case) . . . 91
4.4.2 Scheme 2: Euler–Maruyama without transformation (decomposable case) . . . 92
4.4.3 Scheme 1 for the non-decomposable case . . . 97
4.5 Numerical illustration . . . 100
4.5.1 Neuronal interactions . . . 100
4.5.2 Systemic risk . . . 102

5 Numerical methods for mean-field control problems 105
5.1 Introduction . . . 105
5.2 Fast iterative proximal gradient method for MFC problems . . . 110
5.3 FIPDE method and neural network implementation . . . 114
5.4 High-dimensional Cucker–Smale models . . . 118
5.4.1 Problem formulation . . . 118
5.4.2 Numerical illustration . . . 120

6 Conclusion and outlook **126**

A Auxiliary results for Chapter 3 **129**

A.1 Proof of moment-stability 129

B Auxiliary results for Chapter 4 **132**

B.1 L-differentiability of G^{-1} 132

B.2 The class $\mathscr{C}$ 133

C Additional material for Chapter 5 **135**

C.1 Stochastic maximum principle 135

C.2 Regression Monte–Carlo algorithm 137

C.2.1 Problem formulation 139

C.2.2 Numerical illustration 139

Bibliography **142**

List of Figures

2.1 Strong convergence of the adaptive Euler–Maruyama time-discretisation applied to Example 1. . . . 26

2.2 Example 1: Weak (left) and strong (right) convergence with respect to the number of particles. . . . 28

2.3 Strong convergence for (a) Example 2 and (b) Example 3. A histogram for the number of time-steps in case of Example 2 for (c) $\delta = 2^{-5}$ and (d) $\delta = 2^{-7}$. . . . 29

2.4 Strong convergence for Example 4. . . . 31

2.5 Simulations of the FitzHugh–Nagumo network. (a) "Particle corruption" (in the first component) for a standard Euler–Maruyama scheme. (b) Approximation of the joint density of (V_T, w_T), i.e., the first two components. . . . 32

2.6 Strong convergence of the Milstein schemes. . . . 32

3.1 Numerical illustration for the strong convergence of Example 1, Example 2 (both left), and Example 3 (right) using the tamed Euler scheme and both tamed Milstein schemes, Scheme 1 and Scheme 2 (without the L-derivative terms), respectively. In Example 1, the tamed Milstein scheme without the L-derivative terms then coincides with the tamed Euler scheme. . . . 64

3.2 Numerical illustration of the strong convergence in Example 3 for the tamed Milstein scheme (Scheme 1) with and without the L-derivative terms, with $N = 20$ particles (left) and $N = 3$ (right). . . . 66

3.3 Illustration of the effect of the L-derivative terms in the tamed Milstein scheme (Scheme 1) for Examples 4 and 5 (left, for $N = 3$ particles) and Example 1 (right, for three different choices of N). . . . 66

3.4 Convergence of the terms involving the L-derivatives with an increasing number of particles for Example 4 (left) and Example 5 (right). . . . 67

3.5 Convergence of the terms involving the L-derivatives with an increasing number of particles for Example 1 (left). Illustration of the propagation of chaos property (POC) for Example 1 (right). . . . 68

4.1 Strong convergence of the Euler–Maruyama scheme applied to the particle system obtained by approximating the equation for the action potential of the neurons. . . . 102
4.2 Sample trajectories of the particle system associated with (4.30) for $N = 10$ and $a = 1$ (left) and $a = 10$ (right). 103
4.3 Strong convergence of the Euler–Maruyama scheme applied to the particle system obtained by approximating the equation (4.30) with $a = 1, 5, 10$. 104

5.1 Sample trajectories of the uncontrolled (left) and controlled (right) velocities (second component) for the C–S model with $\beta = \gamma_2 = 0$. 122
5.2 Illustration of the numerically computed feedback control (second component) of a six-dimensional C–S model with $\beta = \gamma_2 = 0$; from left to right: Feedback control obtained by employing the FIPDE method at $t = 0$ and evaluated for different (x_2, v_2); comparison numerical feedback control and the analytic solution at different (t, v_2); comparison numerical feedback control obtained by employing the NNPG method and the analytic solution at different (t, v_2). The remaining components are set to 0.5. 123
5.3 Comparison residual approximation approach with the direct approach for six-dimensional C–S model with $\beta = 0$ and $\gamma_2 = 0.1$. 124
5.4 Illustration of numerically obtained sparse feedback controls for six-dimensional C–S models with $\beta = 0$ and $\gamma_2 > 0$, where the controls are evaluated at different (t, v_2) and the remaining components are set to 0.5. 124
5.5 Uncontrolled and controlled velocities for six-dimensional C–S models with $\beta = 1$ at different times. 125

C.1 (a) Samples of optimal control approximation with $Q_0 \sim \mathcal{U}(1, 2)$) (b) Samples of optimal control approximation with random initial data and ℓ_1-regularisation. 141
C.2 (a) Robustness (absolute error) in terms of (k, λ) with $Q_0 \sim \mathcal{U}(1, 2)$ (b) Absolute relative error (in terms of $k \in [0.8, 1.2]$ and $\lambda \in [0.3, 0.7]$ with $Q_0 \sim \mathcal{U}(1, 2)$ (c) Robustness (absolute error) in terms of perturbation of initial distribution (d) Absolute relative error, in terms of $Q_{0,\min} \in [0.7, 1.3]$ and $Q_{0,\max} \in [1.5, 2]$ (left) and $Q_{0,\max} \in (2, 2.5]$ (right), respectively. The plot on the right is in the log-scale. 141

Chapter 1

Introduction

Since the seminal works [114, 86], which aimed to analyse nonlinear partial differential equations (PDEs) appearing in Vlasov's kinetic theory, numerous authors have studied McKean–Vlasov stochastic differential equations (SDEs), due to their intrinsic link with nonlinear Fokker–Planck PDEs, their connection to the notion of propagation of chaos and their application to mean-field games and mean-field control problems.

Broadly speaking, the aim of this thesis is to present novel numerical schemes to approximate solutions to McKean–Vlasov SDEs and control problems where the controlled dynamics is of McKean–Vlasov type.

To be more precise, McKean–Vlasov equations are d-dimensional SDEs of the following form: Let $T > 0$ be a given terminal time and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a given complete probability space on which an m-dimensional Brownian motion $W = (W_t)_{t\in[0,T]}$ is defined and $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ is the natural filtration of W augmented with an independent σ-algebra $\mathcal{F}_0$. Now, for $\xi \in L_0^2(\mathbb{R}^d)$ (the space of $\mathbb{R}^d$-valued, $\mathcal{F}_0$-measurable random variables with finite second moment), we consider

$$\mathrm{d}X_t = b(t, X_t, \mathcal{L}_{X_t})\,\mathrm{d}t + \sigma(t, X_t, \mathcal{L}_{X_t})\,\mathrm{d}W_t, \quad X_0 = \xi, \tag{1.1}$$

where $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$ are given measurable functions, and $\mathcal{L}_{X_t}$ denotes the marginal law of the process $X = (X_t)_{t\in[0,T]}$ at time $t \in [0,T]$. Here, we use $\mathcal{P}(\mathbb{R}^d)$ to denote the family of all probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-field over $\mathbb{R}^d$, and introduce the subset of probability measures with finite second moment

$$\mathcal{P}_2(\mathbb{R}^d) := \Big\{\mu \in \mathcal{P}(\mathbb{R}^d) : \ \mu(|\cdot|^2) := \int_{\mathbb{R}^d} |x|^2 \mu(\mathrm{d}x) < \infty\Big\}.$$

As a metric on the space $\mathcal{P}_2(\mathbb{R}^d)$, we use the Wasserstein distance: For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the Wasserstein distance between μ and ν is defined as

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

where $\Pi(\mu,\nu)$ is the set of all couplings of μ and ν, i.e., $\pi \in \Pi(\mu,\nu)$ if and only if $\pi(\cdot,\mathbb{R}^d) = \mu(\cdot)$ and $\pi(\mathbb{R}^d,\cdot) = \nu(\cdot)$.

This additional law dependence compared to classical SDEs clearly adds another complexity layer to the study of numerical schemes and well-posedness for equations of the form (1.1). A common strategy to prove strong existence and uniqueness results is based on a fixed-point argument on the space of probability measure endowed with the Wasserstein metric (see, e.g., Chapter 4 in the case of non-Lipschitz drift coefficient). In the following, we state a strong existence and uniqueness result for equations of the form (1.1) for non-globally Lipschitz continuous drift coefficients, which is central for Chapter 2 and Chapter 3 of this thesis:

Theorem 1.0.1. *[131, Theorem 3.3] Let $p \geq 2$ given. Let X be given by (1.1) with $X_0 \in L_0^p(\mathbb{R}^d)$ (independent of the Brownian motion W) and $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$ are measurable and integrable functions, i.e.,*

$$\left(\int_0^T |b(t,0,\delta_0)|\,\mathrm{d}t\right)^p < \infty, \quad \left(\int_0^T \|\sigma(t,0,\delta_0)\|^2\,\mathrm{d}t\right)^{p/2} < \infty,$$

additionally satisfying the following:

- *There exists a constant $L > 0$ such that for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, we have*
$$\|\sigma(t,x,\mu) - \sigma(t,x',\mu')\| \leq L(|x-x'| + \mathcal{W}_2(\mu,\mu')).$$
- *There exists a constant $L > 0$ such that for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, we have*
$$\langle x - x', b(t,x,\mu) - b(t,x',\mu)\rangle \leq L|x-x'|^2,$$
$$|b(t,x,\mu) - b(t,x,\mu')| \leq L\mathcal{W}_2(\mu,\mu').$$
- *There exist constants $L, q > 0$ such that for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, we have*
$$|b(t,x,\mu) - b(t,x',\mu)| \leq L(1 + |x|^q + |x'|^q)|x-x'|.$$

Then, there exists a unique strong solution to (1.1) and there exists a constant $C > 0$ such that

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t|^p\right] \leq C\left(\mathbb{E}\left[|X_0|^p\right] + \left(\int_0^T |b(t,0,\delta_0)|\,\mathrm{d}t\right)^p + \left(\int_0^T \|\sigma(t,0,\delta_0)\|^2\,\mathrm{d}t\right)^{p/2}\right).$$

In general, these SDEs play an important role in financial modelling (e.g., the calibration of so-called local stochastic volatility models [67, 49] and systemic risk [88, 74]), mathematical physics (e.g., the probabilistic interpretation of the Burgers equation, see also [26], and of the 2D-incompressible Navier–Stokes equation (see e.g., [47, 117]), turbulent flow models [128] and the Stefan problem [88]),

chemistry, biology and neuroscience (e.g., the Keller–Segel equation [63] and FitzHugh–Nagumo models [11]). We refer to Chapters 2, 4, and 5 for more details on some of these models and for further applications of McKean–Vlasov equations, such as price impact models and mean-field Cucker–Smale models. We remark that a McKean–Vlasov SDE can also be linked to a non-local and non-linear PDE. Assuming that the law of X_t has a density (with respect to the Lebesgue measure), denoted by $p(t,\cdot)$, the associated Fokker–Planck PDE reads as

$$\partial_t p(t,x) - \sum_{i=1}^{d} \partial_{x_i}(b_i(t,x,\mu_t)p(t,x)) + \frac{1}{2}\sum_{i,j=1}^{d} \partial_{x_{ij}} \left(\sum_{k=1}^{m} \sigma_{ik}(t,x,\mu_t)\sigma_{jk}(t,x,\mu_t)p(t,x) \right) = 0,$$

with $\mu_t(\mathrm{d}x) = p(t,x)\mathrm{d}x$ and given initial condition $p(0,x) = m_0(x)$. Here, both coefficients b and σ depend on the unknown p (hence non-linear) and they are non-local in the sense that, e.g., $b(t,x,\mu_t)$ is a function not of the value $p(t,x)$ but of the entire distribution $(p(t,x))_{x\in\mathbb{R}^d}$.

To simulate a process X described by a McKean–Vlasov equation, one needs the information of the law of the process at each time point t, which is, in general, not available. Hence, one may approximate the measure $\mathcal{L}_{X_t}$ by the empirical measure

$$\mu_t^{\boldsymbol{X}^N}(\mathrm{d}x) := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^{i,N}}(\mathrm{d}x),$$

where $(\boldsymbol{X}_t^N)_{t\in[0,T]} = (X_t^{1,N},\dots,X_t^{N,N})_{t\in[0,T]}$ is the solution to the $(\mathbb{R}^d)^N$-dimensional classical SDE

$$\mathrm{d}X_t^{i,N} = b(t,X_t^{i,N},\mu_t^{\boldsymbol{X}^N})\,\mathrm{d}t + \sigma(t,X_t^{i,N},\mu_t^{\boldsymbol{X}^N})\,\mathrm{d}W_t^i, \quad X_0^{i,N} = \xi^i, \tag{1.2}$$

and $(W_t^i)_{t\in[0,T]}$ and ξ^i for $i\in\{1,\dots,N\}$, are independent copies of W and ξ, respectively. Due to the interaction of the paths $X^{i,N}$, it is common to refer to them as particles instead of paths. The method of replacing the true measure by the empirical distribution is commonly referred to as particle method. When the number of particles tends to infinity, one expects the interaction to diminish, i.e., individual particles do not affect the measure flow anymore. Consequently, in the limit $N\to\infty$ the particles should be i. i. d. To make this more precise, we state the following limits (see, e.g., [100]). Under Lipschitz-type assumptions (e.g., under the assumptions given in Theorem 1.0.1) on the coefficients of the McKean–Vlasov SDE (1.1), the particle system converges in the following two senses: First, for $\mu\in\mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}^d))$ and $\mu_t^{\boldsymbol{X}^N}(\mathrm{d}x) = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}(\mathrm{d}x)$ for $t\in[0,T]$, we have

$$\lim_{N\to\infty} \mathbb{E}\left[\sup_{t\in[0,T]} \mathcal{W}_2(\mu_t^{\boldsymbol{X}^N},\mu_t)\right] = 0.$$

Second, for a fixed $k\in\mathbb{N}$, we have the weak convergence

$$(X^{1,N},\dots,X^{k,N}) \Rightarrow (X^1,\dots,X^k),$$

where $X^1, \dots, X^k$ are independent copies of the solution of the McKean–Vlasov SDE (1.1).

The proof of these limits is based on a coupling argument due to Sznitman [144]. The second limit given above says that, for any fixed k, the first k particles converge in distribution to i. i. d. random variables. Due to the exchangeability of particles, the choice of the first k particles is arbitrary. Note that these two limits are commonly referred to as propagation of chaos in the literature.

Due to the notion of propagation of chaos, equation (1.1) is often used to describe the asymptotic behaviour of a large system of particles with mean-field interaction, where the action on an individual particle has a negligible effect on the entire system. By doing so, the dimensionality of the problem is reduced to d (instead of Nd), but at the price of a measure-dependence of the coefficients. This becomes vital, for instance, if one seeks optimal control of stochastic systems of N particles (or players/agents in an N-player game context) and reduces the complexity of the required optimisation (see, e.g., Chapter 5 and also Chapter 4 for details).

Further, propagation of chaos is a building block for the design of implementable numerical schemes to approximate the solution of (1.1) (see Chapters 2, 3 and 4 for time-stepping schemes of equations with non-standard Lipschitz conditions on the coefficients). In the seminal work [26], an explicit Euler–Maruyama scheme to discretise (1.2) in time is suggested. To be precise, consider a uniform partition of $[0,T]$ with mesh-size $h = T/M$ for some integer $M \in \mathbb{N}$ and recursively compute for $n \in \{0, \dots, M-1\}$

$$X^{i,N,M}_{t_{n+1}} = X^{i,N,M}_{t_n} + b(t_n, X^{i,N,M}_{t_n}, \mu^{\boldsymbol{X}^{N,M}}_{t_n})h + \sigma(t_n, X^{i,N,M}_{t_n}, \mu^{\boldsymbol{X}^{N,M}}_{t_n})\Delta W^i_n, \tag{1.3}$$

with $X^{i,N,M}_0 = \xi^i$, where we introduced $t_n = nh$, $\Delta W^i_n = W^i_{t_{n+1}} - W^i_{t_n}$, for $i \in \{1, \dots, N\}$, and

$$\mu^{\boldsymbol{X}^{N,M}}_{t_n}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X^{j,N,M}_{t_n}}(\mathrm{d}x).$$

The main results of [26], see Theorem 2.1 and Theorem 2.2 therein, provide quantitative error bounds with respect to the number of particles and time-steps (in the L_1-norm) for the approximation of the cumulative distribution function and the density of a certain class of McKean–Vlasov SDEs by the empirical measure and cumulative distribution function of (1.3). In particular, in [26] the authors consider one-dimensional McKean–Vlasov equations with drift and diffusion of the form

$$b(x,\mu) = \int_{\mathbb{R}} \bar{b}(x,y)\,\mu(\mathrm{d}y) \text{ and } \sigma(x,\mu) = \int_{\mathbb{R}} \bar{\sigma}(x,y)\,\mu(\mathrm{d}y), \tag{1.4}$$

where the interaction kernels $\bar{b} : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ and $\bar{\sigma} : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ are globally Lipschitz continuous. For completeness we state the following result of [26] concerning the density approximation for certain one-dimensional McKean–Vlasov SDEs.

Theorem 1.0.2. *[26, Theorem 2.2] Let the following hold:*

- *There exists a strictly positive constant c such that $\bar{\sigma}(x,y) \geq c > 0$, for all $(x,y) \in \mathbb{R}\times\mathbb{R}$.*

- *The kernel $\bar{b}$ ($\bar{\sigma}$) is uniformly bounded and twice (three-times) continuously differentiable on $\mathbb{R}^2$ with bounded derivatives.*

- *The initial law μ_0 has a strictly positive, twice-continuously differentiable density m_0 satisfying: there exist strictly positive constants K, r and α such that for $|x| > K$*

$$m_0(x) + |m_0'(x)| + |m_0''(x)| \leq re^{-\alpha\frac{x^2}{2}}.$$

Let $[0,T] \times \mathbb{R} \ni (t,x) \mapsto p(t,x)$ be the classical solution to the Fokker–Planck PDE associated with (1.1), where the coefficients are defined by (1.4). Then, for all $n \in \{1,\dots,M\}$ and $\varepsilon > 0$ one has

$$\mathbb{E}\left[\int_{\mathbb{R}} |p(t_n,x) - \bar{p}^{\varepsilon}(t_n,x)|\,\mathrm{d}x\right] = \mathcal{O}\left(\varepsilon^2 + \frac{1}{\varepsilon}\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{M}}\right)\right),$$

where

$$\bar{p}^{\varepsilon}(t_n,x) := \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi\varepsilon}} \exp\left(-\frac{(x - X_{t_n}^{i,N,M})^2}{2\varepsilon^2}\right).$$

It is well-documented that the explicit Euler–Maruyama scheme (see [95]), for which strong convergence results of order $1/2$ are known under global Lipschitz-type conditions, is generally not appropriate if we work with drift or diffusion terms of super-linear growth. Specifically, it was shown in [82] for a certain SDE that the uniform time-step Euler–Maruyama scheme is not stable, i.e., the moments of the discretised process explode as the mesh-size tends to zero, even though a unique solution of the original SDE with bounded moments exists. A similar phenomenon is observed in the context of particle systems in [130].

The study of stable time-stepping schemes for interacting particle systems with a drift that grows super-linearly in the state component and with globally Lipschitz continuous diffusion was initiated in [130] building on the seminal work [26]. In [130], a tamed Euler–Maruyama scheme in the spirit of [140, 81] was proposed. Here, one requires both coefficients to be globally Lipschitz continuous in the measure component. A tamed scheme is defined by replacing b in (1.3) with a drift depending on the mesh-size

$$b_M(t,x,\mu) := \frac{b(t,x,\mu)}{1 + M^{-1/2}|b(t,x,\mu)|}, \tag{1.5}$$

for all $(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. The following approximation result for solutions to (1.1) is given in [130].

Theorem 1.0.3. *[130, Theorem 3.3] Let the Assumptions of Theorem 1.0.1 hold, let $[0,T] \ni t \mapsto (b,\sigma)(t,x,\mu)$ to be $1/2$-Hölder continuous uniformly in $(x,\mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, let $\xi \in L_0^p(\mathbb{R}^d)$ for $p \geq 4(1+q)$ and set $h = T/M$ for $M \in \mathbb{N}$. Let (ξ^i, W^i), for $i \in \{1,\dots,N\}$, be independent copies of (ξ, W). Let X^i be the solution to (1.1) with $\xi \equiv \xi^i$ and $W \equiv W^i$, and $X^{i,N,M}$ be the continuous*

time version of (1.3) *with* b *replaced by* b_M*. Then, we obtain the following convergence result: There exists a constant* $C > 0$ *independent of* M *(*h*, respectively) and* N *such that*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - X_t^{i,N,M}|^2\right] \leq C(h + \phi(N)),$$

where

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

These ideas were independently extended to higher-order numerical schemes in [15] and [96]. To recover the strong convergence of order one for the tamed Milstein scheme, a term involving the so-called Lions derivative of the diffusion term will appear. These additional terms are of theoretical interest and highlight the inherent difference of SDEs with mean-field interaction to classical SDEs regarding higher order time-stepping schemes. This notion of differentiability for functions on $\mathcal{P}_2(\mathbb{R}^d)$ was introduced by P.-L. Lions, see [29].

In addition to higher order time-stepping schemes, we will also introduce implementable adaptive time-stepping methods for interacting particle systems with super-linear growth in the drift and diffusion, assuming only a monotonicity condition. The adaptive scheme has to cope with the fact that different time meshes will generally be used for different particles and the mean-field term needs to be approximated efficiently on all such time meshes, a difficulty not encountered for standard SDEs.

Mean-field games (MFGs), introduced in [79, 101], and mean-field control (MFC) problems are both concerned with the study of the asymptotic limit of a large system of interacting symmetric agents (N-player game). In contrast to MFGs, which correspond to the setting in which all agents aim to minimise their individual cost (Nash equilibrium), a MFC problem is characterised by a central planner who tries to minimise the cost by prescribing a control common to all agents (Pareto optimality).

The type of MFC problems studied in this thesis have the following form: For a given terminal time $T > 0$, we consider the following controlled McKean–Vlasov SDE: Let $t \in [0,T]$, $X_0^\alpha = \xi$, for some $\mathbb{R}^d$-valued random variable ξ and

$$\mathrm{d}X_t^\alpha = b(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)})\,\mathrm{d}t + \sigma(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)})\,\mathrm{d}W_t, \tag{1.6}$$

where $b : [0,T]\times\mathbb{R}^d\times\mathbb{R}^k\times\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^k) \to \mathbb{R}^d$ and $\sigma : [0,T]\times\mathbb{R}^d\times\mathbb{R}^k\times\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^k) \to \mathbb{R}^{d\times m}$ are given measurable, and sufficiently regular functions, and $W = (W_t)_{t\in[0,T]}$ is an m-dimensional standard Brownian motion. The control process $\alpha = (\alpha_t)_{t\in[0,T]}$ is assumed to take values in $\mathbb{R}^k$.

The running cost and value function of the optimal control problem are defined by

$$V(\xi) = \inf_{\alpha\in\mathcal{H}^2(\mathbb{R}^k)} J(\alpha;\xi), \quad \text{with } J(\alpha;\xi) = \mathbb{E}\left[\int_0^T \{f(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)}) + \ell(\alpha_t)\}\,\mathrm{d}t + g(X_T^\alpha, \mathcal{L}_{X_T^\alpha})\right], \tag{1.7}$$

where $\mathcal{H}^2(\mathbb{R}^k)$ denotes the set of admissible open-loop controls, i.e., the set of square integrable $\mathbb{F}$-progressively measurable processes α taking values in $\mathbb{R}^k$. The running cost f and terminal cost g are given real-valued functions of at most quadratic growth, and f is continuously differentiable with respect to the control (and its law). Further, $\ell : \mathbb{R}^k \mapsto \mathbb{R}$ is a proper, lower semicontinuous and convex function, which allows to give the associated proximal map in explicit form. Common examples for ℓ include the ℓ_1-norm or the indicator function, which is helpful to enforce sparsity of the control representation and necessary in situations where the control can only attain values in a certain convex subset of $\mathbb{R}^k$.

Typically, the study of (1.6) and (1.7) is limited to the case where the coefficients only depend on the law of the controlled state but not on the law of the control, see, e.g., [41, 32, 31, 1, 34, 102, 126]. Recently, see, e.g., [127, 1, 69, 152, 135, 71], the more general framework described by (1.6) and (1.7) has received significant attention.

Additionally, note that control problems of the form (1.6) and (1.7) do not necessarily have to appear as the limit of a large system of interacting players. For instance, they appear when the goal is to make the optimal state process and/or control process less sensitive to the possible variation of the random events [152], or in a smart grid management model [69]. In these two applications no direct association to an N-player game can be given.

The main contributions of this thesis can be summarised as follows:

- derivation of novel adaptive time-stepping methods for McKean–Vlasov SDEs with super-linear growth in the drift and diffusion, assuming only a monotonicity condition;
- proof of moment stability and analysis of strong convergence order of the adaptive time-stepping schemes taking the randomness of required time-steps into account;
- derivation of Milstein schemes for particle systems associated with McKean–Vlasov equations using Lions calculus on measure space;
- proof of moment stability and first order uniform (in time) convergence in the time-step using a tamed scheme if only a one-sided Lipschitz condition holds for the drift;
- the main results extend those on tamed Milstein schemes (for certain taming approaches) for standard SDEs by eliminating commutativity conditions;
- establishing the well-posedness of one-dimensional McKean–Vlasov SDEs with a certain form of discontinuities and of their associated particle systems relying on techniques from variational calculus on the measure space $\mathcal{P}(\mathbb{R})$ equipped with the Wasserstein distance $\mathcal{W}_2$;

- development and analysis of numerical schemes for approximating such McKean–Vlasov SDEs and their associated particle systems using non-standard strong convergence analysis based on occupation time estimates of the discretised processes in a neighbourhood of the discontinuity;
- development of novel iterative schemes to approximate the solution to non-smooth MFC control problems with mean-field interaction through the control and state process;
- presentation of a practical implementation relying on a hybrid Monte–Carlo and PDE approach;
- exhaustive numerical tests for high-dimensional mean-field Cucker–Smale models employing deep neural networks.

The main chapters of this thesis are organised as follows:

Chapter 2: In this chapter, we introduce implementable adaptive Euler–Maruyama schemes for McKean–Vlasov SDEs assuming only a standard monotonicity condition on the drift and diffusion coefficients but no global Lipschitz continuity in the state variable for either, while global Lipschitz continuity is required for the measure component only. We prove moment stability of the discretised processes and a strong convergence rate of $1/2$. Several numerical examples, centred around a mean-field model for FitzHugh–Nagumo neurons, illustrate that the standard uniform scheme fails and that the adaptive approach shows in most cases superior performance to tamed approximation schemes. In addition, we introduce and analyse an adaptive Milstein scheme for a certain sub-class of McKean–Vlasov SDEs with linear measure-dependence of the drift. This chapter is based on [132].

Chapter 3: In this chapter, we derive implementable first order time-stepping schemes for McKean–Vlasov SDEs, allowing for a drift term with super-linear growth in the state component. We propose Milstein schemes for a time-discretised interacting particle system associated with the McKean–Vlasov equation and prove strong convergence of order one and moment stability, taming the drift if only a one-sided Lipschitz condition holds. To derive our main results on strong convergence rates, we make use of calculus on the space of probability measures with finite second order moments. In addition, numerical examples are presented which support our theoretical findings. This chapter is based on [15], which has been further developed in [16, 97].

Chapter 4: In this chapter, we first establish well-posedness results for one-dimensional McKean–Vlasov SDEs and related particle systems with a measure-dependent drift coefficient that is discontinuous in the spatial component, and a diffusion coefficient, which is a Lipschitz function of the state only. We only require a fairly mild condition on the diffusion coefficient, namely to be non-zero

in a point of discontinuity of the drift, while we need to impose certain structural assumptions on the measure-dependence of the drift. Second, we study implementable Euler–Maruyama type schemes for the particle system to approximate the solution of the one-dimensional McKean–Vlasov SDE. Here, we will prove strong convergence results in terms of the number of time-steps and number of particles. Due to the discontinuity of the drift, the convergence analysis is non-standard and the usual strong convergence order $1/2$ known for the Lipschitz case cannot be recovered for all schemes. This chapter is based on [104].

Chapter 5: In this chapter, we develop numerical methods for MFC problems, which seek optimal control of McKean–Vlasov dynamics whose coefficients involve mean-field interactions both on the state and actions. At the core of our algorithm, we iteratively refine the approximate optimal feedback control by applying an accelerated proximal gradient method to the MFC problem, which does not require solving the first order optimality condition, arising from the stochastic maximum principle, analytically. Our procedure is a hybrid algorithm, i.e., Monte-Carlo and PDE based, which computes the optimal control in feedback form and solves the control problem along a flow of measures. Thereby, we improve the robustness of the numerically obtained feedback controls in terms of model perturbation compared to pure sampling based algorithms. We also propose a version of our method based on a deep neural network implementation, which is applicable for high-dimensional control problems, illustrated for the well-known Cucker–Smale model. This chapter is based on [136] and mentions some findings of [134, 135].

Chapter 2

An adaptive Euler–Maruyama scheme for McKean–Vlasov SDEs with super-linear growth and application to the mean-field FitzHugh–Nagumo model

2.1 Introduction

A McKean–Vlasov equation (introduced in [114, 86]) for a d-dimensional process $X = (X_t)_{t\in[0,T]}$, with a given $T > 0$, is an SDE where the underlying coefficients depend on the current state X_t and, additionally, on the law of X_t, i.e., they have the form

$$\mathrm{d}X_t = b(t, X_t, \mathcal{L}_{X_t})\,\mathrm{d}t + \sigma(t, X_t, \mathcal{L}_{X_t})\,\mathrm{d}W_t, \quad X_0 = \xi \in L_0^2(\mathbb{R}^d), \tag{2.1}$$

where $W = (W_t)_{t\in[0,T]}$ is an m-dimensional standard Brownian motion and $\mathcal{L}_{X_t}$ denotes the marginal law of the process X at time $t \in [0, T]$. Here, for $p \geq 2$, $L_0^p(\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued, $\mathcal{F}_0$-measurable random variables Y satisfying $\mathbb{E}[|Y|^p] < \infty$.

The existence and uniqueness theory for strong solutions of McKean–Vlasov SDEs with coefficients of linear growth and Lipschitz conditions with respect to the state and the measure is well-established; see, e.g., [144] for the classical setting. For further specific existence and uniqueness results on weak and strong solutions of McKean–Vlasov SDEs we refer to [17, 75, 118] and references cited therein. Most relevant to this work, in the case of super-linear drift and diffusion, it is known from [131, 97] that McKean–Vlasov SDEs admit a unique strong solution under a monotonicity condition and Lipschitz continuity of the coefficients in the measure component.

Our motivation to study McKean–Vlasov SDEs with non-globally Lipschitz drift (and diffusion) is that several important models in practice involve mean-field terms and do not exhibit the classical global Lipschitz conditions on the coefficients of the SDE. For example, some models in neuroscience,

such as the Hodgkin-Huxley and FitzHugh–Nagumo models [11, 25], or mean-field equations describing the behaviour of a (large) network of interacting spiking neurons [55], do not satisfy the classical assumptions (e.g., linear growth) on the drift coefficient.

The simulation of McKean–Vlasov SDEs of the form (2.1) typically involves two steps: a particle approximation and a time-stepping scheme. For a seminal work on numerical schemes for McKean–Vlasov SDEs we refer the reader to [24].

In the first step, at each time t, the true measure $\mathcal{L}_{X_t}$ is approximated by the empirical measure

$$\mu_t^{\boldsymbol{X}^N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x), \tag{2.2}$$

where $(\boldsymbol{X}_t^N)_{t\in[0,T]} = (X_t^{1,N}, \dots, X_t^{N,N})_{t\in[0,T]}^\top$ (so-called interacting particle system) is the solution to the $(\mathbb{R}^d)^N$-dimensional SDE with components

$$\mathrm{d}X_t^{i,N} = b(t, X_t^{i,N}, \mu_t^{\boldsymbol{X}^N})\,\mathrm{d}t + \sigma(t, X_t^{i,N}, \mu_t^{\boldsymbol{X}^N})\,\mathrm{d}W_t^i, \quad X_0^{i,N} = \xi^i, \tag{2.3}$$

and (W^i, ξ^i) are independent copies of (W, ξ), for $i \in \{1, \dots, N\}$. Furthermore, δ_x denotes the Dirac measure at x.

Key to the convergence as $N \to \infty$ is the concept of propagation of chaos. Consider N (independent) processes X^i driven by independent Brownian motions W^i

$$\mathrm{d}X_t^i = b(t, X_t^i, \mathcal{L}_{X_t^i})\,\mathrm{d}t + \sigma(t, X_t^i, \mathcal{L}_{X_t^i})\,\mathrm{d}W_t^i, \quad X_0^i = \xi^i, \tag{2.4}$$

which are standard McKean–Vlasov SDEs (and by uniqueness $\mathcal{L}_{X_t^i} = \mathcal{L}_{X_t}$). Then, pathwise propagation of chaos refers to the property

$$\lim_{N\to\infty} \max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{i,N} - X_t^i|^2\right] = 0.$$

This type of result is classical under global Lipschitz conditions (see, e. g., [30]), and has been proven in [130, Proposition 3.1] under super-linear growth conditions of the drift. While the rates for the propagation of chaos shown in [130] suffer from the curse of dimensionality (i.e., the rate depends on d and diminishes as $d \to \infty$) [56, Lemma 5.1] gives for $p \le 4$ and any $i \in \{1, \dots, N\}$ the bound

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{i,N} - X_t^i|^p\right] \le CN^{-p/2},$$

where the constant $C > 0$ is dimension-independent. However, in [56, Lemma 5.1], the drift coefficient of the underlying McKean–Vlasov SDE is assumed to satisfy strong regularity assumptions (differentiability, along with Lipschitz and boundedness conditions on the state and measure components) and the diffusion coefficient is assumed to be constant, i.e., the framework is much more restrictive than the setting of the present chapter (in particular, the super-linear growth case is not covered in [56]).

As McKean–Vlasov equations typically arise from a mean-field approximation to a high- but finite-dimensional particle system, a particle approximation to the McKean–Vlasov equation simply reverses this step. To be more concrete, for instance, the FitzHugh–Nagumo model for N interacting neurons from P populations with different characteristics is a system of N three-dimensional SDEs of the form

$$\begin{aligned} \mathrm{d}X_t^{i,N} = f_\alpha(t, X_t^{i,N})\,\mathrm{d}t + g_\alpha(t, X_t^{i,N}) \begin{bmatrix} \mathrm{d}W_t^i \\ \mathrm{d}W_t^{i,y} \end{bmatrix} \\ + \sum_{\gamma=1}^{P} \frac{1}{N_\gamma} \sum_{j,p(j)=\gamma} \Big(b_{\alpha\gamma}(X_t^{i,N}, X_t^{j,N})\,\mathrm{d}t + \beta_{\alpha\gamma}(X_t^{i,N}, X_t^{j,N})\,\mathrm{d}W_t^{i,\gamma} \Big), \end{aligned} \tag{2.5}$$

where $N = \sum_{\gamma=1}^{P} N_\gamma$ and f_α satisfies a monotone growth condition but is not globally Lipschitz continuous (see Example 4 in Section 2.4.4 for a detailed description). It is shown in [11, 25] that the mean-field limit $N \to \infty$ is described by the three-dimensional McKean–Vlasov SDE

$$\mathrm{d}\bar{X}_t^\alpha = f_\alpha(t, \bar{X}_t^\alpha)\,\mathrm{d}t + \sum_{\gamma=1}^{P} \mathbb{E}_{\bar{Z}}[b_{\alpha\gamma}(\bar{X}_t^\alpha, \bar{Z}_t^\gamma)]\,\mathrm{d}t + g_\alpha(t, \bar{X}_t^\alpha)\,\mathrm{d}W_t^\alpha + \sum_{\gamma=1}^{P} \mathbb{E}_{\bar{Z}}[\beta_{\alpha\gamma}(\bar{X}_t^\alpha, \bar{Z}_t^\gamma)]\,\mathrm{d}W_t^{\alpha\gamma}, \tag{2.6}$$

for $\alpha \in \{1, \dots, P\}$, where $\bar{Z}$ is a vector-valued process independent of $\bar{X}$ with the same law. Conversely, (2.5) can be interpreted as a particle approximation to (2.6).

The focus of this chapter is a time-stepping scheme for the particle system in (2.3) over some finite time horizon $[0, T]$. We have in mind systems which arise from particle approximations of McKean–Vlasov equations, but equally the scheme is of interest for interacting particle systems, potentially high-dimensional ones, in their own right. Note that at each time-step, the computational cost of simulating the particle system is usually $\mathcal{O}(N^2)$, as interaction terms due to the dependence of the coefficients on the empirical measure have to be computed for each particle. Strategies to reduce this cost include the projected particle method of [19].

In the context of classical SDEs, it is well-documented that the explicit Euler–Maruyama scheme (see [95]), for which strong convergence results of order $1/2$ are known under Lipschitz-type conditions, is generally not appropriate if we work with drift terms of super-linear growth. Specifically, it was shown in [82] for a certain SDE that the uniform time-step Euler–Maruyama scheme is not stable, i.e., the moments of the discretised process explode as the mesh-size tends to zero, even though a unique solution of the original SDE with bounded moments exists. A similar phenomenon is observed in the context of particle systems in [130], namely that during the simulation some of the particles strongly oscillate and eventually diverge (as also illustrated by Figure 2.5a in this thesis for the FitzHugh–Nagumo model).

To overcome this problem in the setting of classical SDEs, several stable time-discretisations have been introduced, including tamed explicit Euler–Maruyama and Milstein schemes [81, 140, 68],

balanced schemes [149, 111], an explicit adaptive Euler–Maruyama method [59], a truncated Euler–Maruyama method [113] and an implicit Euler–Maruyama scheme [77]. For these methods, stability of the discretised process and strong convergence results have been proven. In [91], an explicit and a semi-implicit adaptive scheme for classical SDEs with non-globally Lipschitz continuous coefficients is studied, assuming a monotonicity property resulting from a Khasminskii-type condition [94], and additionally differentiability of the coefficients. A backstop scheme with a strict upper and lower bound on the time-step immediately gives attainability of the final time T, and convergence arbitrarily close to order $1/2$ is shown for a choice of the time-step which depends solely on the growth of the drift term. An adaptive time-stepping Milstein method for SDEs with one-sided Lipschitz drift was developed in [92], which also relies on such a backstop scheme.

In [130], the tamed Euler–Maruyama scheme and an implicit scheme were introduced for McKean–Vlasov SDEs with super-linearly growing drift and a diffusion coefficient which is globally Lipschitz in the state variable. Both coefficients are assumed to be Lipschitz continuous in the measure. In this setting, besides pathwise propagation of chaos, the paper establishes strong convergence of the time-discretised particle system to the SDE (2.3). The recent papers [16, 15, 96] introduce and analyse tamed Milstein schemes for (delay) McKean–Vlasov SDEs with a one-sided Lipschitz condition on the drift and a globally Lipschitz continuous diffusion coefficient (in both, the state and measure component), while [97] extends this analysis to the framework of McKean–Vlasov SDEs with common noise, where the diffusion coefficient is also allowed to grow super-linearly. We refer to [97] for examples of equations exhibiting such super-linear diffusion.

The first contribution of the current chapter is to introduce implementable adaptive time-stepping methods for McKean–Vlasov SDEs with super-linear growth in the drift and diffusion, assuming only a monotonicity condition (see (H.3(2)) in Section 2.3.1 below). The adaptive scheme has to cope with the fact that different time meshes will generally be used for different particles and the mean-field term needs to be approximated efficiently on all such time meshes, a difficulty not encountered for standard SDEs.

First, we will present an explicit adaptive Euler–Maruyama scheme for the setting where the diffusion coefficient is globally Lipschitz in both components. Here, we need only a mild assumption on the choice of the adaptive time-step (in the spirit of [59]) in order to achieve stability of the scheme. Then, we remove the Lipschitz assumption on the diffusion coefficient and impose only a certain Khasminskii-type monotonicity condition. However, we will need stronger assumptions on the time-step function, in order to also control the growth in the diffusion coefficient. The original analysis is not applicable in this more general setting and a new time-step control with a suitable stopping time argument is needed, which ultimately gives attainability of the final time and stability of the scheme. This extension is achieved without the need for a (semi-implicit) backstop scheme if

the proposed adaptive time-step becomes too small and equally forms a contribution to the existing literature on adaptive time-stepping schemes for classical (non-measure dependent) SDEs.

Moreover, we prove moment stability and strong convergence of order 1/2 in the time-step. Several numerical examples in Section 2.4 demonstrate that our adaptive scheme often outperforms the tamed Euler–Maruyama scheme introduced in [130], i.e., in many cases it gives more accurate numerical approximations or even achieves superior strong convergence rates.

Additionally, we present an adaptive Milstein scheme for a sub-class of McKean–Vlasov SDEs where the drift depends on the state and linearly on its law but the diffusion coefficient is only state dependent, see equation (2.16) below for details. We consider this special class firstly because many relevant examples arising in practical applications have this form (see, e.g., [38] and the references cited therein), and secondly to avoid introducing a notion of derivatives for functions depending on probability measures (e.g., the Lions derivative; see [31, 15, 96] for details). We do not foresee any fundamental difficulties in extending the analysis to a more general setting, and leave it to future work.

A new difficulty for the computation and the analysis of adaptive schemes for particle systems, as mentioned above, is the approximation of the empirical measure at all adaptive time points. If the number of particles is large, as for particle approximations of McKean–Vlasov equations, this may result in a high computational cost if the empirical measure is re-evaluated at the smallest time-steps dictated by any particle. To avoid this, we present a scheme where the empirical measure is kept constant over larger, fixed time-steps. This procedure avoids this extra cost but still allows us to obtain the same stability and convergence results as for the more expensive scheme. We prove stability for both these Euler–Maruyama schemes in the super-linear drift case, and convergence for the second scheme, while the convergence proof for the first scheme is found in the arXiv version of [132]. Similarly, we only give the stability proofs for super-linear diffusion and the Milstein scheme for the first variant; while the stability results for the second variant follow along similar lines as the results presented in this chapter, the proofs of strong convergence are subject to future research.

The remainder of this chapter is organised as follows: In Section 2.2, we describe the particle method and our adaptive schemes, and give the main convergence result (concerned with a globally Lipschitz continuous diffusion); the proof thereof is deferred to Section 2.5. Section 2.3 introduces an adaptive Euler–Maruyama scheme for a more general class where we allow both coefficients to grow super-linearly, and an adaptive Milstein scheme for a more specific class of McKean–Vlasov equations with linear measure-dependence; the proofs are again given in Section 2.5. The numerical results for several standard test cases taken from the literature, notably the FitzHugh–Nagumo model, are presented in Section 2.4.

We end this section by introducing some basic notations and give background results needed throughout this chapter.

Preliminaries

Throughout this chapter, let $(\mathbb{R}^d, \langle\cdot,\cdot\rangle, |\cdot|)$ represent the d-dimensional Euclidean space, and $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, for a given $T > 0$, will be a complete filtered probability space, where $\mathbb{F}$ is the natural filtration of the m-dimensional Brownian motion $W = (W_t)_{t\in[0,T]}$ augmented with an independent σ-algebra $\mathcal{F}_0$.

In addition, we use $\mathcal{P}(\mathbb{R}^d)$ to denote the family of all probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-field over $\mathbb{R}^d$, and denote the subset of probability measures with finite second moment by $\mathcal{P}_2(\mathbb{R}^d)$. As metric on the space $\mathcal{P}_2(\mathbb{R}^d)$, we use the Wasserstein distance. Further, for $p \geq 2$, $\mathcal{S}^p([0,T])$ refers to the space of $\mathbb{R}^d$-valued continuous, $\mathbb{F}$-adapted processes, defined on the interval $[0,T]$, with finite p-th moments, i.e., processes $X = (X_t)_{t\in[0,T]}$ satisfying $\mathbb{E}\left[\sup_{t\in[0,T]} |X_t|^p\right] < \infty$.

For all linear operators (e.g., matrices $A \in \mathbb{R}^{d\times m}$) appearing in this chapter, we will use the standard Hilbert-Schmidt norm denoted by $\|\cdot\|$.

2.2 Time-stepping scheme for super-linear drift coefficient

In Section 2.2.1, we give a precise set of assumptions for the McKean–Vlasov equation (2.1) which guarantee well-posedness and are needed in the stability and strong convergence analysis of the time-stepping scheme introduced later in this section. Here, we first focus on a diffusion coefficient which is globally Lipschitz continuous in the state and measure component, while in Section 2.3.1 we will allow a diffusion coefficient growing super-linearly in the state variable. The non-uniform Euler–Maruyama scheme will be presented in Section 2.2.2 and the adaptive choice of time-steps and main convergence result in Section 2.2.3.

2.2.1 Problem formulation

We consider the McKean–Vlasov equation defined by (2.1), repeated here for convenience,

$$\mathrm{d}X_t = b(t, X_t, \mathcal{L}_{X_t})\,\mathrm{d}t + \sigma(t, X_t, \mathcal{L}_{X_t})\,\mathrm{d}W_t, \quad X_0 = \xi \in L_0^p(\mathbb{R}^d), \tag{2.7}$$

where W is a m-dimensional Brownian motion and $p \geq 2$. We impose the following assumptions for the coefficients:

H. 1. The mappings $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$ are measurable and satisfy:

(1) (Lipschitz condition for diffusion coefficient): There exists a constant $L > 0$ such that

$$\|\sigma(t,x,\mu) - \sigma(t,x',\mu')\| \leq L(|x-x'| + \mathcal{W}_2(\mu,\mu')), \quad \forall t \in [0,T],\ x \in \mathbb{R}^d \text{ and } \mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d).$$

(2) (One-sided Lipschitz condition for drift): There exists a constant $L > 0$ such that

$$\langle x - x', b(t,x,\mu) - b(t,x',\mu)\rangle \leq L|x-x'|^2, \quad \forall t \in [0,T],\ x, x' \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

(3) (Lipschitz measure dependence of drift): There exists a constant $L > 0$ such that

$$|b(t,x,\mu) - b(t,x,\mu')| \leq L\mathcal{W}_2(\mu,\mu'), \quad \forall t \in [0,T],\ x \in \mathbb{R}^d \text{ and } \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d).$$

(4) (Polynomial growth condition): There exist constants $L > 0$, $q > 0$, such that

$$|b(t,x,\mu) - b(t,x',\mu)| \leq L(1 + |x|^q + |x'|^q)|x - x'|, \quad \forall t \in [0,T],\ x, x' \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

(5) (1/2-Hölder-continuity in time): There exists a constant $C > 0$ such that

$$|b(t,x,\mu) - b(t',x,\mu)| + \|\sigma(t,x,\mu) - \sigma(t',x,\mu)\| \leq C|t-t'|^{1/2}, \quad \forall t, t' \in [0,T],\ x \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

(6) There exists a constant $C > 0$ such that $|b(0,0,\mu)| + \|\sigma(0,0,\mu)\| \leq C, \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Remark 2.2.1. Note that the above assumptions imply the so-called monotone growth condition, i.e., there exist non-negative constants C_1 and C_2 such that for all $t \in [0,T]$ and all $\mu \in \mathcal{P}_2(\mathbb{R}^d)$

$$\langle x, b(t,x,\mu)\rangle + \frac{1}{2}\|\sigma(t,x,\mu)\|^2 \leq C_1 + C_2|x|^2. \tag{2.8}$$

This condition, frequently employed in the literature (see, e.g., [130]), is needed to guarantee the moment boundedness of our numerical scheme (2.12) presented below.

Under Assumptions (H.1(1))–(H.1(5)) and for $\xi \in L_0^p(\mathbb{R}^d)$, $p \geq 2$, it is known from [131] that the SDE (2.7) has a unique solution $X \in \mathcal{S}^p([0,T])$. In Section 2.3.1, (H.1(1)) and (H.1(2)) will be replaced by the weaker Assumption (H.3(2)) which enables us to consider non-globally Lipschitz diffusion coefficients.

We will now briefly illustrate that the interacting particle system defined in (2.3) is indeed well-posed. Set

$$B(t,\boldsymbol{x}^N) := (b(t,x_1,\mu^{\boldsymbol{x}^N}),\dots,b(t,x_N,\mu^{\boldsymbol{x}^N})), \quad \Sigma(t,\boldsymbol{x}^N) := \mathrm{diag}(\sigma(t,x_1,\mu^{\boldsymbol{x}^N}),\dots,\sigma(t,x_N,\mu^{\boldsymbol{x}^N})),$$

where $\mu^{\boldsymbol{x}^N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{x_j}(\mathrm{d}x)$ for $\boldsymbol{x}^N := (x_1, x_2, \dots, x_N) \in \mathbb{R}^{dN}$. Note that

$$\mathcal{W}_2^2(\mu^{\boldsymbol{x}^N}, \mu^{\boldsymbol{y}^N}) \leq \frac{1}{N}\sum_{i=1}^N |x_i - y_i|^2 = \frac{1}{N}|\boldsymbol{x}^N - \boldsymbol{y}^N|^2, \quad \boldsymbol{x}^N, \boldsymbol{y}^N \in \mathbb{R}^{dN},$$

which is a standard bound for the Wasserstein distance between two empirical measures, see, e.g., [30]. Hence, from Assumptions (H.1(1))–(H.1(3)), we readily deduce that there exists a constant $L > 0$ such that

$$\langle \boldsymbol{x}^N - \boldsymbol{y}^N, B(t,\boldsymbol{x}^N) - B(t,\boldsymbol{y}^N)\rangle \leq L|\boldsymbol{x}^N - \boldsymbol{y}^N|^2, \quad \|\Sigma(t,\boldsymbol{x}^N) - \Sigma(t,\boldsymbol{y}^N)\|^2 \leq L|\boldsymbol{x}^N - \boldsymbol{y}^N|^2,$$

for any $\boldsymbol{x}^N, \boldsymbol{y}^N \in \mathbb{R}^{dN}$, and the claim follows from [129, Theorem 3.1.1]. In addition, we have the stability estimate $\max_{i\in\{1,\dots,N\}} \mathbb{E}[\sup_{t\in[0,T]} |X_t^{i,N}|^p] \leq C_{p,T}(1 + \mathbb{E}[|\xi|^p])$, for some constant $C_{p,T} > 0$, see, e.g., [130, Proposition 5.1].

2.2.2 Adaptive Euler–Maruyama scheme for super-linear drift

For some integer $M > 0$, we define a uniform step-size $h = T/M$. We start by stating the standard Euler–Maruyama scheme with uniform time-steps for (2.3), for $n \in \{0, \dots, M-1\}$,

$$X^{i,N,M}_{t_{n+1}} = X^{i,N,M}_{t_n} + b(t_n, X^{i,N,M}_{t_n}, \mu^{\boldsymbol{X}^{N,M}}_{t_n})\, h + \sigma(t_n, X^{i,N,M}_{t_n}, \mu^{\boldsymbol{X}^{N,M}}_{t_n})\, \Delta W^i_{t_n}, \tag{2.9}$$

where $t_n = nh$, $\Delta W^i_{t_n} = W^i_{t_{n+1}} - W^i_{t_n}$, $X^{i,N,M}_0 = \xi^i$, and

$$\mu^{\boldsymbol{X}^{N,M}}_{t_n}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X^{j,N,M}_{t_n}}(\mathrm{d}x).$$

Extending (2.9), we introduce now an adaptive Euler–Maruyama scheme for the particle system (2.3). A difficulty arises if we compute for each particle adaptive time-steps based on $X^{i,N}_t$, as a different time mesh may result for each particle. Then, at a time point associated with a specific particle, we need an approximation of the empirical measure for the computation of the mean-field term in order to evaluate the update for the next time-step. However, we do not have the value of each particle available at that time point. Hence, an approximation to the mean-field term is not readily computable.

We now propose a first scheme that allows the computation of an empirical measure $\mu^{\hat{\boldsymbol{X}}^N}_t$ for all t. To do so, we start at $t = 0$ with $\hat{X}^{i,N}_0 = \xi^i$ for all $i \in \{1, \dots, N\}$. At step $n \geq 0$, we compute for each particle the size of the adaptive time-step $h^i_n := h^\delta(\hat{X}^{i,N}_n)$, with the time-step function h^δ introduced precisely below. We then perform the Euler–Maruyama update as

Scheme 1:
$$\hat{X}^{i,N}_{t_{n+1}} = \hat{X}^{i,N}_{t_n} + b(t_n, \hat{X}^{i,N}_{t_n}, \mu^{\hat{\boldsymbol{X}}^N}_{t_n})\; h^{\min}_n + \sigma(t_n, \hat{X}^{i,N}_{t_n}, \mu^{\hat{\boldsymbol{X}}^N}_{t_n})\; \Delta W^i_{t_n},$$

with step-size $h^{\min}_n := \min\{h^1_n, \dots, h^N_n\}$, $t_{n+1} = t_n + h^{\min}_n$ and $\Delta W^i_{t_n} = W^i_{t_{n+1}} - W^i_{t_n}$, and with

$$\mu^{\hat{\boldsymbol{X}}^N}_{t_n}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{\hat{X}^{j,N}_{t_n}}(\mathrm{d}x). \tag{2.10}$$

We will later use the notations $n_t := \max\{n : t_n \leq t\}$ and $\underline{t} := \max\{t_n \leq t\}$. This allows us to introduce the piecewise constant interpolant process $\bar{X}^{i,N}_t := \hat{X}^{i,N}_{\underline{t}}$ and measure $\mu^{\bar{\boldsymbol{X}}^N}_t := \mu^{\hat{\boldsymbol{X}}^N}_{\underline{t}}$, and also the continuous interpolant processes

$$\hat{X}^{i,N}_t := \hat{X}^{i,N}_{\underline{t}} + b(\underline{t}, \hat{X}^{i,N}_{\underline{t}}, \mu^{\hat{\boldsymbol{X}}^N}_{\underline{t}})(t - \underline{t}) + \sigma(\underline{t}, \hat{X}^{i,N}_{\underline{t}}, \mu^{\hat{\boldsymbol{X}}^N}_{\underline{t}})(W^i_t - W^i_{\underline{t}}),$$

with the associated interpolant empirical measure $\mu^{\hat{\boldsymbol{X}}^N}_t$, so that $(\hat{X}^{i,N}_t)_{t\in[0,T]}$ is the solution to the SDE

$$\mathrm{d}\hat{X}^{i,N}_t = b(\underline{t}, \bar{X}^{i,N}_t, \mu^{\bar{\boldsymbol{X}}^N}_t)\,\mathrm{d}t + \sigma(\underline{t}, \bar{X}^{i,N}_t, \mu^{\bar{\boldsymbol{X}}^N}_t)\,\mathrm{d}W^i_t. \tag{2.11}$$

The stability of this scheme is analysed in Proposition 2.5.1, while for the strong convergence analysis, we refer to Section 5.2 of the arXiv version.

In the above discretisation, we compute for each particle i the size of the time-step, determine the smallest one and simulate each particle with the same (small) step-size in order to approximate the measure. This is computationally expensive, as it requires us to take the minimal time-step taken over a large number of particles. As $N \to \infty$, at each t and for all x there will almost surely be some i for which $\hat{X}_t^{i,N} > x$, thus requiring a step-size adapted to this extreme scenario. Typically, however, at any given time only a very small proportion of the particles will require such a small step-size for stability of the overall system. Choosing the smallest such time-step for the evolution of all particles introduces a dominant, unnecessary cost, which will significantly increase the complexity order. However, for a moderate number of particles this scheme is still tractable.

We will now define a more practical scheme for the case that N is large. For our final numerical procedure, hence, we keep the empirical measure constant in an interval of length δT, where $\delta = 1/M$ for some integer $M > 0$. This avoids having to recompute the measure at each adaptive time point. In the sequel, let k_n be the integer for which $t_n \in [k_n \delta T, (k_n + 1)\delta T)$. Note that k_n might have the same value for different n. Then, our adaptive Euler–Maruyama method becomes:

Scheme 2: $$\tilde{X}_{t_{n+1}}^{i,N} = \tilde{X}_{t_n}^{i,N} + b(t_n, \tilde{X}_{t_n}^{i,N}, \mu_{k_n \delta T}^{\tilde{\boldsymbol{X}}^N})\, h_n^i + \sigma(t_n, \tilde{X}_{t_n}^{i,N}, \mu_{k_n \delta T}^{\tilde{\boldsymbol{X}}^N})\, \Delta W_{t_n}^i,$$

where $h_n^i := \min\{h^\delta(\tilde{X}_{t_n}^{i,N}), (k_n+1)\delta T - t_n\}$, $t_{n+1} = t_n + h_n^i$, $\Delta W_{t_n}^i = W_{t_{n+1}}^i - W_{t_n}^i$ and $\tilde{X}_0^{i,N} = \xi^i$. Similar to above, we have

$$\mu_{k_n \delta T}^{\tilde{\boldsymbol{X}}^N}(\mathrm{d}x) := \frac{1}{N} \sum_{j=1}^{N} \delta_{\tilde{X}_{k_n \delta T}^{j,N}}(\mathrm{d}x).$$

We introduce the piecewise constant interpolant process $\tilde{\tilde{X}}_t^{i,N} = \tilde{X}_{\underline{t}}^{i,N}$, and measure $\mu_t^{\tilde{\tilde{\boldsymbol{X}}}^N} = \mu_{k_n \delta T}^{\tilde{\tilde{\boldsymbol{X}}}^N}$, and also the continuous interpolant

$$\tilde{X}_t^{i,N} = \tilde{X}_{\underline{t}}^{i,N} + b(\underline{t}, \tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n \delta T}^{\tilde{\boldsymbol{X}}^N})(t - \underline{t}) + \sigma(\underline{t}, \tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n \delta T}^{\tilde{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i),$$

so that $(\tilde{X}_t^{i,N})_{t \in [0,T]}$ is the solution to the SDE

$$\mathrm{d}\tilde{X}_t^{i,N} = b(\underline{t}, \tilde{\tilde{X}}_t^{i,N}, \mu_t^{\tilde{\tilde{\boldsymbol{X}}}^N})\, \mathrm{d}t + \sigma(\underline{t}, \tilde{\tilde{X}}_t^{i,N}, \mu_t^{\tilde{\tilde{\boldsymbol{X}}}^N})\, \mathrm{d}W_t^i. \tag{2.12}$$

Given this framework, we will impose in Section 2.2.3 conditions on the size of the time-step which will allow us to prove attainability of T, stability and strong convergence of the scheme. The complexity of the scheme and the main results are also given there.

2.2.3 Choice of time-step function and main result

Compared to the classical Euler–Maruyama scheme with a uniform step-size, we use adaptive steps $h^\delta(\hat{X}_{t_n}^{i,N})$ depending on the state $\hat{X}_{t_n}^{i,N}$ of the i-th particle at the current (n-th) time-step. For each $0 < \delta \leq 1$, the *time-step function* $h^\delta : \mathbb{R}^d \to \mathbb{R}^+$ hence controls the size of the time-step, which

is relevant especially for large arguments to prevent instabilities. Convergence is then considered as $\delta \to 0$.

We make the following assumptions (based on Section 2 in [59]) on the function h^δ:

H.2. There exists a continuous and strictly positive function $h : \mathbb{R}^d \to \mathbb{R}^+$ for which

(1) the time-step function h^δ satisfies

$$\delta \min(T, h(x)) \leq h^\delta(x) \leq \min(\delta T, h(x)),$$

where the following hold for h:

(2) there exist constants $a, b, c > 0$, such that

$$h(x) \geq (a|x|^c + b)^{-1}, \quad \forall x \in \mathbb{R}^d,$$

(3) and there exist constants $L_c, L_d > 0$, such that

$$\langle x, b(t,x,\mu)\rangle + \frac{1}{2}h(x)|b(t,x,\mu)|^2 \leq L_c|x|^2 + L_d, \quad \forall t \in [0,T],\ x \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

For the adaptive Milstein scheme (see Section 2.3.2 for details), the factor $1/2$ is replaced by $3/2$.

Here, Assumption (H.2(1)) ties the dependence of $h^\delta(x)$ on x, for each δ, to that of a generic function h. Note that the schemes are fully defined in terms of h^δ, while the existence of h is only needed to ascertain the desired theoretical properties of the scheme. Specifically, Assumption (H.2(2)) lower bounds the speed of decay of the function h as $|x| \to \infty$ and is necessary to derive a bound for the expected number of time-steps; Assumption (H.2(3)) in conjunction with (H.1(4)) is used to control the super-linear growth of the drift term. A suitable choice of c in (H.2(2)) can be deduced from q in Assumption (H.1(4)), while a depends on the constant factors in (H.1(4)), and b simply eliminates a singularity at $x = 0$.

Conversely, one can use an appropriately chosen h to define h^δ systematically. Consider for simplicity a one-dimensional setting where the drift asymptotically behaves like $-x^{q+1}$ (as per (H.1(4))): choosing first $h(x) = \min(1, |x|^{-q})$ and then h^δ as either the lower or upper bound in the double inequality in (H.2(1)) gives a viable time-step function h^δ. We give an explicit example for the construction of h^δ, with a suitable choice of h satisfying (H.2), in Section 2.4.2.

Given now the randomness of the step-size at time t through its dependence on $\tilde{X}_t^{i,N}$, the number of time-steps M_T^i up to time T is a random variable for each particle i. The total expected complexity is therefore characterised by $\sum_{i=1}^N \mathbb{E}[M_T^i]$, and can be estimated as follows.

Proposition 2.2.1. *Under (H.2) and for $(\tilde{X}_t^{i,N})_{t\in[0,T]}$ with bounded moments (uniformly in t) up to order c, there exists $C>0$ independent of δ and N such that*

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} M_T^i\right] \leq C\delta^{-1}. \tag{2.13}$$

Proof. As the upper and lower bounds in (H.2(2))–(H.2(3)) only depend on the state variable (and not on the measure), the analysis of [59, Lemma 2] is readily adapted to our setting assuming a moment bound.[1] We have

$$\begin{aligned} M_T^i &\leq 1 + T \sup_{t\in[0,T]} \frac{1}{h^\delta(\tilde{X}_t^{i,N})} + T\delta^{-1} \\ &\leq 1 + T\delta^{-1}\left(\sup_{t\in[0,T]} \max\{h^{-1}(\tilde{X}_t^{i,N}), T^{-1}\} + 1\right) \\ &\leq T\delta^{-1}\left(a \sup_{t\in[0,T]} |\tilde{X}_t^{i,N}|^c + b + (1+\delta)T^{-1} + 1\right). \end{aligned}$$

Note that the term $T\delta^{-1}$ in the first inequality appears because h_n^i may take a value which is smaller than the step-size computed by h^δ, but this happens at most once in each subinterval of length δT.

This estimate implies (2.13) because of the assumed moment bound on $(\tilde{X}_t^{i,N})_{t\in[0,T]}$. □

We now give our main result on strong convergence of the adaptive Euler–Maruyama scheme, Scheme 2, and recall pathwise strong propagation of chaos (from [130]). To do so, we require the following definition:

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

Theorem 2.2.1. *Let the SDE (2.7) satisfy Assumption (H.1), $\xi \in L_0^p(\mathbb{R}^d)$, for a sufficiently large $p\geq 2$, and assume that the time-step function h^δ satisfies (H.2). Then, there exist constants $C, \widetilde{C}>0$ such that*

$$\max_{i\in\{1,\dots,N\}} \sup_{t\in[0,T]} \mathbb{E}\left[|X_t^i - \tilde{X}_t^{i,N}|^2\right] \leq C\left(\delta + \varphi(N)\right) \leq \widetilde{C}\left(\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} M_T^i\right]\right)^{-1} + C\varphi(N), \tag{2.14}$$

where X^i and $\tilde{X}^{i,N}$ are defined by (2.4) and (2.12), respectively.

The proof of the first inequality in (2.14) is found in Section 2.5, specifically Subsection 2.5.2, while the second inequality follows immediately from Proposition 2.2.1.

Theorem 2.2.1 shows a strong convergence of order $1/2$ in terms of the expected average number of time-steps over all particles. Also, recall that the additional complexity in terms of number of particles is, in the worst case, of order N^2. This results from evaluating potentially different functions of the empirical measure (with N terms) for all particles. In many examples, like in our

[1] We will show this in Proposition 2.5.1. In our case, in contrast to [59], where the initial data is deterministic we require enough regularity of the random initial value.

tests in Sections 2.4.1 and 2.4.2 where the interaction term is simply an expectation and identical for all particles, the complexity is $\mathcal{O}(N)$.

2.3 Super-linear diffusion and higher order schemes

Here, we will discuss additional adaptive time-stepping schemes. Specifically, in Section 2.3.1, we propose a stable Euler–Maruyama scheme allowing for a super-linearly growing diffusion coefficient, in addition to a super-linear drift term, subject only to a mild monotonicity assumption. In Section 2.3.2, we will present an adaptive Milstein scheme for a special class of McKean–Vlasov equations where the diffusion only depends on the state of the process and is assumed to be globally Lipschitz continuous.

2.3.1 Super-linear diffusion coefficient

We will consider an adaptive Euler–Maruyama scheme, defined as in (2.12) and (2.11), in the setting where both the drift and diffusion coefficient grow super-linearly in the state component. To guarantee well-posedness of (2.7) in this more general framework and stability of the resulting adaptive Euler–Maruyama scheme, we will need to replace (H.1(1)) and (H.1(2)) and by the following new set of assumptions:

H.3. (1) (Lipschitz measure dependence of diffusion coefficient): There exists a constant $L > 0$ such that

$$\|\sigma(t,x,\mu) - \sigma(t,x,\mu')\| \leq L\mathcal{W}_2(\mu,\mu'), \quad \forall t \in [0,T],\ x \in \mathbb{R}^d \text{ and } \mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d).$$

(2) (Monotonicity condition): Let $p \geq 2$ given. There exists a constant $L_1 > 0$ such that

$$\langle x - x', b(t,x,\mu) - b(t,x',\mu)\rangle + \frac{p+1}{2}\|\sigma(t,x,\mu) - \sigma(t,x',\mu)\|^2 \leq L_1|x-x'|^2, \tag{2.15}$$

for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Note that estimate (2.15) and (H.1(6)) imply

$$\langle x, b(t,x,\mu)\rangle + \frac{p-1}{2}\|\sigma(t,x,\mu)\|^2 \leq C_1 + C_2|x|^2, \quad \forall t \in [0,T],\ x \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d),$$

for some $C_1, C_2 > 0$.

Under Assumptions (H.3), (H.1(3))–(H.1(6)) and for $\xi \in L_0^p(\mathbb{R}^d)$, $p > 2$, it is known from [97, Theorem 2.1] that the SDE (2.7) has a unique solution $X \in \mathcal{S}^k([0,T])$ for any $2 \leq k < p$. A similar monotonicity condition to (2) is imposed in the works [141, 98] on tamed schemes for non-measure dependent equations with super-linearly growing diffusion coefficient. We refer the reader to [112, Chapter 5] for a general discussion of such a monotonicity condition in the framework of SDEs with locally Lipschitz continuous coefficients.

To make the adaptive scheme stable in this framework, we require a modification of (H.2(3)):

H.4. For some constant $C > 0$ and q from (H.1(4)) we have

$$h(x) \leq C(1 + |x|^{3q})^{-1}, \quad \forall x \in \mathbb{R}^d.$$

Remark 2.3.1. Due to (H.1(4)), (H.1(6)) and (H.3(2)), there exist constants $C_1, C_2 > 0$ (depending on the constants appearing in these three assumptions) such that

$$|b(t,x,\mu)| \leq C_1(1 + |x|^{q+1}), \quad \forall t \in [0,T], x \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d),$$
$$\|\sigma(t,x,\mu)\|^2 \leq C_2(1 + |x|^{q+2}), \quad \forall t \in [0,T],\ x \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

These growth estimates along with (H.4) imply that for some constant $C_3 > 0$,

$$|b(t,x,\mu)|\|\sigma(t,x,\mu)\|(h(x))^{1/2} \leq C_3(1 + |x|^2), \quad \forall t \in [0,T],\ x \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d),$$

which will be employed in Section 2.5.3 to prove stability of the Euler–Maruyama time-stepping scheme.

The condition (H.4) is stronger than (H.2(3)). Assume (H.1(4)) holds for some q, then $|b(t,x,\mu)| = \mathcal{O}(|x|^{q+1})$, and it is sufficient to choose $h(x) = \mathcal{O}(|x|^{-q})$, as noted earlier. The stronger decay imposed by (H.4) does not lead to a worse complexity order.

Note that the structure of Assumption (H.4) still allows us to impose (H.2(2)), which guarantees the boundedness of the expected number of time-steps.

We formulate the strong convergence result only for Scheme 1. In Section 2.5.3, we will prove stability of the scheme only, as the convergence analysis follows closely the arguments presented in [97, Theorem 3.1] once stability is established.

Theorem 2.3.1. *Let $p > 2$, $\xi \in L_0^k(\mathbb{R}^d)$, for a sufficiently large $k \geq 2$ (depending on p), and let Assumptions (H.3) and (H.1(3))–(H.1(6)) hold. If the time-step function h^δ satisfies Assumptions (H.4) and (H.2(1)), then there exists a constant $C > 0$ such that*

$$\max_{i \in \{1,\dots,N\}} \mathbb{E}\left[\sup_{t \in [0,T]} |\hat{X}_t^{i,N} - X_t^{i,N}|^p\right] \leq C\delta^{p/2},$$

where $X^{i,N}$ and $\hat{X}^{i,N}$ are defined by (2.3) and (2.11), respectively.

2.3.2 Adaptive Milstein scheme

Here, we consider for ease of demonstration specifically McKean–Vlasov SDEs of the form (see [15, 96] for tamed schemes in a more general framework)

$$\mathrm{d}X_t = \int_{\mathbb{R}^d} b(X_t, y)\, \mathcal{L}_{X_t}(\mathrm{d}y)\, \mathrm{d}t + \sigma(X_t)\, \mathrm{d}W_t, \quad X_0 = \xi \in L_0^p(\mathbb{R}^d), \tag{2.16}$$

for some $p \geq 2$ and $b : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$.

Using the standard particle method to approximate the true measure, we arrive at

$$\mathrm{d}X_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} b(X_t^{i,N}, X_t^{j,N})\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}W_t^i, \quad X_0^{i,N} = \xi^i, \tag{2.17}$$

where ξ^i are again independent copies of X_0. In the sequel, we use $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_m)$, where $\sigma_i = (\sigma_{1,i}, \ldots, \sigma_{d,i})^\top$, for $i \in \{1, \ldots, m\}$. For $x = (x_1, \ldots, x_d)$, we denote $L^{j_1} := \sum_{k=1}^{d} \sigma_{k,j_1} \frac{\partial}{\partial x_k}$.

We then make the following assumptions on b and σ:

H.5. (1) (Lipschitz condition for diffusion coefficient): There exists a constant $L > 0$ such that

$$\|\sigma(x) - \sigma(x')\| \le L|x - x'|, \quad \forall x, x' \in \mathbb{R}^d.$$

(2) Every σ_i, $i \in \{1, \ldots, m\}$, is continuously differentiable and there exists a constant $L > 0$ such that

$$|L^{j_1}\sigma_{j_2}(x) - L^{j_1}\sigma_{j_2}(x')| \le L|x - x'|, \quad j_1, j_2 \in \{1, \ldots, m\}, \quad \forall x, x' \in \mathbb{R}^d.$$

(3) (One-sided Lipschitz condition for drift): There exists a constant $L > 0$ such that

$$\langle x - x', b(x, y) - b(x', y)\rangle \le L|x - x'|^2, \quad \forall x, x', y \in \mathbb{R}^d.$$

(4) (Polynomial growth condition): There exist constants $L, q > 0$ such that

$$|b(x, y) - b(x', y)| \le L(1 + |x|^q + |x'|^q)|x - x'|, \quad \forall x, x', y \in \mathbb{R}^d.$$

(5) (Lipschitz condition for drift kernel): There exists a constant $L > 0$ such that

$$|b(x, y) - b(x, y')| \le L|y - y'|, \quad \forall x, y, y' \in \mathbb{R}^d.$$

(6) There exists a constant $C > 0$ such that $|b(0, y)| \le C, \quad \forall y \in \mathbb{R}^d$.

These assumptions will guarantee the existence of moments of the processes defined by (2.19) and (2.20) below.

To prove a strong convergence result, additional assumptions on b are needed:

H.6. The functions $\mathbb{R}^d \ni x \mapsto b(x, y)$ and $\mathbb{R}^d \ni y \mapsto b(x, y)$ are continuously differentiable for all $y \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$, respectively, and there exist constants $L_1, L_2, r > 0$ such that

$$\|\partial_x b(x, y) - \partial_x b(x', y)\| \le L_1(1 + |x|^r + |x'|^r)|x - x'|,$$
$$\|\partial_y b(x, y) - \partial_y b(x, y')\| \le L_2|y - y'|, \quad \forall x, y, x', y' \in \mathbb{R}^d.$$

In the sequel, we further impose for simplicity a commutativity condition on the diffusion matrix

$$L^{j_1}\sigma_{k,j_2} = L^{j_2}\sigma_{k,j_1}, \quad j_1, j_2 \in \{1, \dots, m\}, \quad k \in \{1, \dots, d\}. \tag{2.18}$$

This implies that the so-called *Lévy area* (see, e.g., [95]), the double stochastic integral appearing in the construction of the Milstein scheme, is zero. Then, the adaptive Milstein scheme has the form

$$\begin{aligned}\hat{Y}^{i,N}_{t_{n+1}} = \hat{Y}^{i,N}_{t_n} &+ \frac{1}{N}\sum_{j=1}^{N} b(\hat{Y}^{i,N}_{t_n}, \hat{Y}^{j,N}_{t_n})h_n^{\min} + \sigma(\hat{Y}^{i,N}_{t_n})\Delta W^i_{t_n} \\ &+ \frac{1}{2}\sum_{j_1,j_2=1}^{m} L^{j_1}\sigma_{j_2}(\hat{Y}^{i,N}_{t_n})(\Delta W^{i,j_1}_{t_n}\Delta W^{i,j_2}_{t_n} - \delta_{j_1,j_2}h_n^{\min}),\end{aligned} \tag{2.19}$$

where $h_n^{\min} := \min\{h_n^1, \dots, h_n^N\}$ and $t_{n+1} = t_n + h_n^{\min}$. Analogously to the adaptive Euler–Maruyama scheme, one can also define the process $\tilde{X}^{i,N}$ with piecewise constant measure over time intervals of length δT, as well as the continuous-time interpolants of $\hat{Y}^{i,N}$ and $\tilde{Y}^{i,N}$.

We give a result for the stability and strong convergence of the process $(\hat{Y}^{i,N}_t)_{t\in[0,T]}$, which is the Milstein version of Scheme 1, but do not pursue the analogue of Scheme 2 for simplicity.

Theorem 2.3.2. *Let $\xi \in L_0^p(\mathbb{R}^d)$ for a sufficiently large $p \geq 2$. Let Assumptions (H.5) and (H.6) hold and the time-step function satisfies Assumptions (H.2(3))–(H.2(1)). Then, there exists a constant $C > 0$ such that for the solutions $X^{i,N}$ and $\hat{Y}^{i,N}$ of (2.17) and (2.19),*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |\hat{Y}^{i,N}_t - X^{i,N}_t|^p\right] \leq C\delta^p.$$

Recall that for the adaptive Milstein scheme Assumption (H.2(3)) has a different constant than for the adaptive Euler–Maruyama scheme as pointed out there. We will only prove moment boundedness of this scheme in Section 2.5.4, as the rest of the strong convergence analysis then follows by analogous steps to those in the proof of the main results in [15, 96].

2.4 Examples and numerical illustration

In this section, we present some numerical examples for the adaptive schemes and compare them with tamed Euler–Maruyama and Milstein schemes ([130] and [96, 15], respectively).

The first three tests are modifications of tests from the literature to illustrate certain features of the method: in Sections 2.4.1 and 2.4.2 by adding a mean-field term to SDEs with non-Lipschitz drift; in Section 2.4.3 by adding a non-Lipschitz drift to a McKean–Vlasov SDE. In Section 2.4.4, we apply the method to the FitzHugh–Nagumo model from neuroscience. Then, in Section 2.4.5, we give a numerical illustration for the performance of our proposed adaptive Milstein scheme.

The first two examples are initially formulated as McKean–Vlasov equations and then approximated by a particle system in the canonical way, while the remaining examples are directly formulated as particle systems.

Tamed Euler–Maruyama scheme

The tamed Euler–Maruyama scheme (see ([130]) for the particle system (2.3) with $h = T/M$, where M denotes the number of (uniform) time-steps, reads

$$X_{t_{n+1}}^{i,N,M} = X_{t_n}^{i,N,M} + \frac{b(t_n, X_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})}{1 + M^{-\alpha}|b(t_n, X_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})|}h + \sigma(t_n, X_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})\Delta W_{t_n}^i, \quad X_0^{i,N,M} = \xi^i,$$

for $i \in \{1, \dots, N\}$, where the N Brownian motions are independent, $\alpha = 1/2$ (to achieve strong convergence of order $1/2$) or $\alpha = 1$ (to enable strong order 1 for a constant diffusion term) and

$$\mu_{t_n}^{\boldsymbol{X}^{N,M}}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^{N} \delta_{X_{t_n}^{j,N,M}}(\mathrm{d}x).$$

Tamed Milstein scheme

The tamed Milstein scheme for (2.16) is a special case of the schemes in [96, 15] and has the form

$$\begin{aligned} Y_{t_{n+1}}^{i,N,M} &= Y_{t_n}^{i,N,M} + \frac{1}{N}\tilde{b}(Y_{t_n}^{i,N,M}, Y_{t_n}^{j,N,M})h \\ &\quad + \sigma(Y_{t_n}^{i,N,M})\Delta W_{t_n}^i + \frac{1}{2}\sum_{j_1,j_2=1}^{k} L^{j_1}\sigma_{j_2}(Y_{t_n}^{i,N,M})(\Delta W_{t_n}^{i,j_1}\Delta W_{t_n}^{i,j_2} - \delta_{j_1,j_2}h), \quad Y_0^{i,N,M} = \xi^i, \end{aligned} \tag{2.20}$$

where $\delta_{j_1,j_2} = 1$, for $j_1 = j_2$ and $\delta_{j_1,j_2} = 0$, for $j_1 \neq j_2$, and

$$\tilde{b}(Y_t^{i,N,M}, Y_t^{j,N,M}) := \frac{\sum_{j=1}^{N} b(Y_{t_n}^{i,N,M}, Y_{t_n}^{j,N,M})}{1 + h\left|\frac{1}{N}\sum_{j=1}^{N} b(Y_{t_n}^{i,N,M}, Y_{t_n}^{j,N,M})\right|}.$$

To illustrate the strong convergence behaviour in h (δ, respectively), we compute the root mean-square error (RMSE) by comparing the numerical solution at level l (of the time-discretisation), denoted by $\hat{X}_T^{i,N,M_l}$ (which generically stands for a numerical approximation obtained by the tamed or adaptive schemes with $M = M_l$), with the solution at level $l-1$, at the final time T. To be precise, we compute

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{X}_T^{i,N,M_l} - \hat{X}_T^{i,N,M_{l-1}}\right)^2},$$

where $M_l = \lceil 2^l T \rceil$ and the two particle systems at each level are generated by the same Brownian motions. For our numerical experiments, we set $T = 1$.

2.4.1 Example 1 – the Fang and Giles test with added mean-field term

Consider the one-dimensional SDE

$$\mathrm{d}X_t = \left(-\frac{X_t}{1 - |X_t|^2} + \mathbb{E}[X_t]\right)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 0,$$

adapted from [59] by including the term $\mathbb{E}[X_t]$. The tamed Euler–Maruyama scheme reads

$$X_{t_{n+1}}^{i,N,M} = X_{t_n}^{i,N,M} + \frac{b(t_n, X_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})}{1 + M^{-\alpha}|b(t_n, X_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})|} h + \Delta W_{t_n}^i,$$

where

$$b(t_n, X_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}}) = -\frac{X_{t_n}^{i,N,M}}{1 - |X_{t_n}^{i,N,M}|^2} + \frac{1}{N}\sum_{j=1}^{N} X_{t_n}^{j,N,M}.$$

As the diffusion is constant, one would expect a strong convergence order of 1 for the tamed scheme. Hence, one would set $\alpha = 1$ to achieve this rate. However, as we will illustrate later the choice $\alpha = 1$ seems not to give any rate of convergence, while the stronger taming with $\alpha = 1/2$ at least recovers the expected rate of order $1/2$.

Though this example does not satisfy the assumptions listed above, it was numerically shown in [59] for the corresponding SDE without the mean-field term that the adaptive Euler–Maruyama method with $h^\delta(x) = \delta(1 - |x|^2)$ outperformed the tamed Euler–Maruyama scheme and an implicit scheme.

If a newly computed approximation $\hat{X}_{t_{n+1}}^{i,N}$ exceeds a predefined radius $r_{\max} = 1 - 10^{-10}$, this value is replaced by $r_{\max}$ and $(r_{\max}/|\hat{X}_{t_{n+1}}^{i,N}|)\hat{X}_{t_{n+1}}^{i,N}$ in the one- and multi-dimensional setting, respectively.

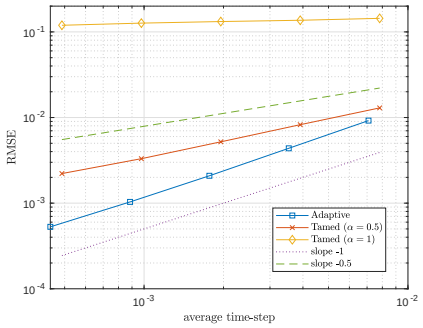


Figure 2.1: Strong convergence of the adaptive Euler–Maruyama time-discretisation applied to Example 1.

Fig. 2.1 plots the root mean-square error (RMSE) against the average time-step size; this is to make the comparison of the tamed Euler–Maruyama scheme and the adaptive method fair. Here, and for the following examples, "average time-step" refers to the harmonic mean of the step-size, i.e. we average the number of time-steps over all particles and then take the reciprocal. The number of particles was set to $N = 10^4$.

Note that in Fig. 2.1 results for the exponent $-1/2$ (i.e., $\alpha = 1/2$) are shown for the tamed scheme in addition to $\alpha = 1$ (chosen in [59]), as the latter yields an even worse convergence rate. Furthermore, it illustrates that the adaptive scheme achieves strong convergence order of 1 (as the

volatility coefficient is constant) in this example compared to the order 1/2 for the tamed Euler–Maruyama method with $\alpha = 1/2$. That we only observe convergence for $\alpha = 1/2$ in the case of tamed schemes, could be argued by the fact that the choice of $\alpha = 1/2$ results in a stronger taming.

To justify the approximation of the true mean-field term by the arithmetic mean over all particles, we numerically investigate the weak and strong convergence in N of the solution to the particle system, $X^{i,N}$ as in (2.3), to the solution of the limit McKean–Vlasov SDE, X^i, as in (2.4).

To be precise, for the weak convergence we study $\mathbb{E}[X^{i,N}] - \mathbb{E}[X^i]$, which we expect to be of order $1/N$. To do so, we fix a fine time-grid and compute

$$\text{error} := \left| \frac{1}{K} \sum_{j=1}^{K} \left(\frac{1}{N_l} \sum_{i=1}^{N_l} \hat{X}_T^{i,N_l,M,(j)} - \frac{1}{N_{l+1}} \sum_{i=1}^{N_{l+1}} \hat{X}_T^{i,N_{l+1},M,(j)} \right) \right|,$$

i.e., we compute the mean-field term at final time T using two particle systems with a different number of particles (e.g., N_l and $N_{l+1} = 2N_l$) but using the same time-discretisation for both of them. This procedure will be repeated (independently) $K \in \mathbb{N}$ times (indicated by the superscript j) and we finally compute the sample mean of these K realisations. Also, note that N_l Brownian motions are used for the simulation of $\hat{X}^{i,N_l,M,(j)}$, out of the set of N_{l+1} Brownian motions used for the simulation of $\hat{X}^{i,N_{l+1},M,(j)}$. This is done for several levels l, where a level refers here to the number of particles.

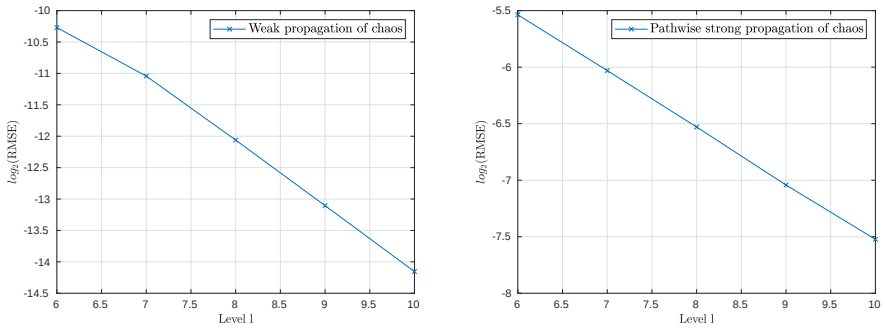
For the strong pathwise propagation of chaos result, we compute for a fixed fine time-grid

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

where the particle system $\tilde{\hat{X}}^{i,N_l,M}$, $i \in \{1, \dots, N_l\}$, is obtained by splitting the set of Brownian motions driving the particle system $\hat{X}^{i,N_l,M}$, $i \in \{1, \dots, N_l\}$, in two sets of size $N_l/2$, such that $(\tilde{\hat{X}}^{i,N_l,M,(1)})_{i \in \{1,\dots,N_l/2\}}$ is a particle system obtained from the Brownian motions $(W^i)_{i \in \{1,\dots,N_l/2\}}$ and $(\tilde{\hat{X}}^{i,N_l,M,(2)})_{i \in \{N_l/2+1,\dots,N_l\}}$ from $(W^i)_{i \in \{N_l/2+1,\dots,N_l\}}$. In particular, for these two smaller particle systems only $N_l/2$ particles are used to approximate the mean-field term. For our numerical test, we set $M = 2^8$.

We observe in Fig. 2.2 a weak and strong convergence rate of order roughly 1 (in accordance with the result in [43]) and roughly 1/2, respectively. Recall that [130, Proposition 3.1] establishes a strong convergence rate (in terms of number of particles) of order 1/4 for dimensions $d < 4$ only. However, one would expect to have strong convergence of order 1/2 in N, which was indeed proven in, e.g., [56], under a more restrictive set-up than here. The convergence rate in terms of the number of particles can also be improved to 1/2 (for any d) if the drift has the special form given as in equation (2.16), see [116].

This illustration is only given for Example 1. For the remaining examples we use the same way of approximating the mean-field term and the results would be similar.



(a) (b)

Figure 2.2: Example 1: Weak (left) and strong (right) convergence with respect to the number of particles.

2.4.2 Example 2 – Ginzburg-Landau with added mean-field term

As a second example, we consider the one-dimensional McKean–Vlasov SDE

$$\mathrm{d}X_t = \left(\frac{\sigma^2}{2}X_t - X_t^3 + c\mathbb{E}[X_t]\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \quad X_0 = 1,$$

which is a variation of the stochastic Ginzburg-Landau equation (see [148]) with the addition of a mean-field term. This classical, non-globally Lipschitz SDE satisfies all assumptions listed above (except (H.1(6)), which is not needed for the well-posedness) and therefore has a unique strong solution.

Here, we choose $h^\delta(x) = \delta\min(1, |x|^{-2})$, which can be motivated as follows: as the drift is mainly influenced by the $-X_t^3$ term, the choice $h(x) = \min(1, |x|^{-2})$ yields $\langle x, b(t,x,\mu)\rangle + \frac{1}{2}h(x)|b(t,x,\mu)|^2 < 0$, for large x. It is then easy to verify that Assumption (H.2) is satisfied for $a = b = 1$, $c = 2$, and sufficiently large L_c and L_d depending on the model parameters.

In all tests, we take the parameters $\sigma = 1.5$ and $c = 0.5$, and $N = 10^4$ particles. We set $\alpha = 1$ in the tamed scheme. Although, the choice $\alpha = 0.5$ would be sufficient to achieve a strong convergence order of $1/2$, we observed that the performance of the tamed scheme with this taming was significantly worse than with $\alpha = 1$.

The strong convergence rate depicted in Fig. 2.3a is roughly of order $1/2$ for both the adaptive and tamed schemes, but the absolute error is smaller by a factor of roughly 10 for the adaptive scheme.

In addition, we present in Fig. 2.3d a histogram for N_T^i, $i \in \{1, \dots, N\}$, i.e., the number of adaptive time-steps required for all particles, for $\delta = 2^{-5}$ and $\delta = 2^{-7}$, with bin-width 2^3 and 2^5, respectively. The average number of time-steps is approximately 55 for $\delta = 2^{-5} = 1/32$ and 218 for $\delta = 2^{-7} = 1/128$, consistent with $\mathbb{E}[M_T] = \mathcal{O}(\delta^{-1})$.

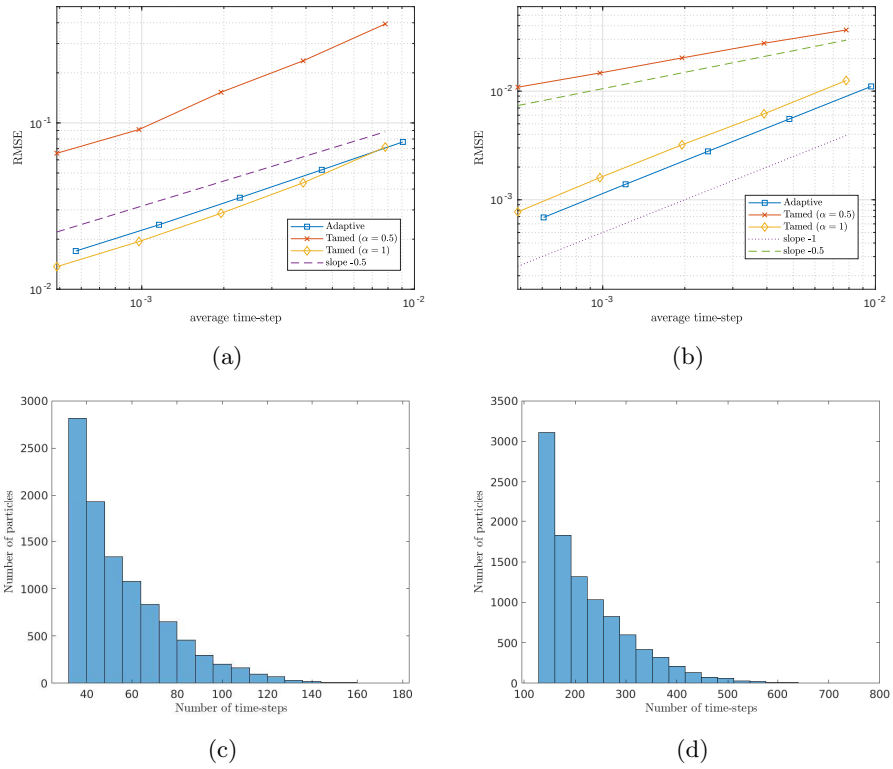


Figure 2.3: Strong convergence for (a) Example 2 and (b) Example 3. A histogram for the number of time-steps in case of Example 2 for (c) $\delta = 2^{-5}$ and (d) $\delta = 2^{-7}$.

2.4.3 Example 3 – Kuramoto model with added non-Lipschitz drift

Next, we consider a modification of the one-dimensional Kuramoto model (see [2]) of the form

$$\mathrm{d}X_t^{i,N} = \left(\eta^i + X_t^{i,N} - (X_t^{i,N})^3 + \frac{1}{N}\sum_{j=1}^{N}\sin(X_t^{i,N} - X_t^{j,N})\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t^i, \quad X_0^{i,N} = \xi^i,$$

where $\sigma \in \mathbb{R}$ is a constant and η^i are i. i. d. random variables and independent from the set of i. i. d. random variables ξ^i and the Brownian motions W^i, for $i \in \{1, \dots, N\}$. We set $\sigma = 1$, $\xi^i \sim \mathcal{U}(0.5, 1)$ and $\eta^i \sim \mathcal{N}(0, 1)$. The term $X_t - X_t^3$, which does not appear in the original model, was added to illustrate the effect of a super-linear drift term. As in the previous example, we choose $h^\delta(x) = \delta \min(1, |x|^{-2})$.

Our numerical tests shown in Fig. 2.3a verify the strong convergence of order 1 for the adaptive scheme, which is due to the constant volatility term in this example. It is only in this special case that we find the adaptive and tamed schemes to behave similarly. The number of particles was set to $N = 10^3$.

To achieve a strong convergence rate of 1 for the tamed Euler–Maruyama method, the taming exponent α has to be chosen as 1. For the choice $\alpha = 1/2$ suggested in [130, 140], one only achieves

the expected strong order of 1/2. However, note that in the generality of the SDEs considered there, the aim in [130, 140] was only to prove strong convergence rates of order 1/2.

2.4.4 Example 4 – FitzHugh–Nagumo model

As main application, we consider a standard model from neuroscience, the FitzHugh–Nagumo network (see, e.g., [11, 25, 115]). For N neurons and P different neuron populations, we denote for $i \in \{1, \dots, N\}$ by $p(i) = \alpha$, $\alpha \in \{1, \dots, P\}$, the population the i-th particle belongs to. For the state vector $(X_t^{i,N})_{t\in[0,T]} = (V_t^{i,N}, w_t^{i,N}, y_t^{i,N})_{t\in[0,T]}$ of neuron i, we consider the 3-dimensional SDE

$$\begin{aligned} \mathrm{d}X_t^{i,N} = f_\alpha(t, X_t^{i,N})\,\mathrm{d}t + g_\alpha(t, X_t^{i,N}) \begin{bmatrix} \mathrm{d}W_t^i \\ \mathrm{d}W_t^{i,y} \end{bmatrix} \\ + \sum_{\gamma=1}^{P} \frac{1}{N_\gamma} \sum_{j,p(j)=\gamma} \left(b_{\alpha\gamma}(X_t^{i,N}, X_t^{j,N})\,\mathrm{d}t + \beta_{\alpha\gamma}(X_t^{i,N}, X_t^{j,N})\,\mathrm{d}W_t^{i,\gamma} \right), \end{aligned}$$

for $i \in \{1, \dots, N\}$ and where N_γ denotes the number of neurons in population γ. For $\gamma, \alpha \in \{1, \dots, P\}$, $I^\alpha(t) := I$, $\forall t \in [0, T]$, $\forall \alpha$, for some constant value I,

$$f_\alpha(t, X_t^{i,N}) = \begin{bmatrix} V_t^{i,N} - \frac{(V_t^{i,N})^3}{3} - w_t^{i,N} + I^\alpha(t) \\ c_\alpha(V_t^{i,N} + a_\alpha - b_\alpha w_t^{i,N}) \\ a_r^\alpha S_\alpha(V_t^{i,N})(1 - y_t^{i,N}) - a_d^\alpha y_t^{i,N} \end{bmatrix}, \quad g_\alpha(t, X_t^{i,N}) = \begin{bmatrix} \sigma_{\text{ext}}^\alpha & 0 \\ 0 & 0 \\ 0 & \sigma_\alpha^y(V_t^{i,N}, y_t^{i,N}) \end{bmatrix},$$

and

$$b_{\alpha\gamma}(X_t^{i,N}, X_t^{j,N}) = \begin{bmatrix} -\bar{J}_{\alpha\gamma}(V_t^{i,N} - V_{\text{rev}}^{\alpha\gamma}) y_t^{j,N} \\ 0 \\ 0 \end{bmatrix}, \quad \beta_{\alpha\gamma}(X_t^{i,N}, X_t^{j,N}) = \begin{bmatrix} -\sigma_{\alpha\gamma}^J(V_t^{i,N} - V_{\text{rev}}^{\alpha\gamma}) y_t^{j,N} \\ 0 \\ 0 \end{bmatrix}.$$

Moreover,

$$\begin{aligned} S_\alpha(V_t^{i,N}) &= \frac{T_{\max}^\alpha}{1 + e^{-\lambda_\alpha(V_t^{i,N} - V_T^{i,N})}}, & \chi(y_t^{i,N}) &= \mathbb{I}_{y_t^{i,N} \in (0,1)} \Gamma e^{-\Lambda/(1-(2y_t^{i,N}-1)^2)}, \\ \sigma_\alpha^y(V_t^{i,N}, y_t^{i,N}) &= \sqrt{a_r^\alpha S_\alpha(V_t^{i,N})(1 - y_t^{i,N}) + a_d^\gamma y_t^{i,N}}\, \chi(y_t^{i,N}), \end{aligned}$$

and the standard one-dimensional Brownian motions $(W^i, W^{i,y}, W^{i,\gamma})$ are mutually independent for all $i \in \{1, \dots, N\}$. For our numerical tests, we use the parameter values listed in [130] (with $\sigma_{\text{ext}} = 0, \sigma^J = 0.00002$ and $\sigma_{\text{ext}} = 0.5, \sigma^J = 0.2$, respectively), where the tamed Euler–Maruyama scheme was numerically tested, and also use the same type of random initial values presented there.

In our numerical experiments, we set $P = 1$ and $N = 10^3$. Note that the functions f_α and g_α are uniformly locally Lipschitz continuous with respect to the second variable. Moreover, the functions $\beta_{\alpha\gamma}$ and $b_{\alpha\gamma}$ are globally Lipschitz continuous and satisfy the growth condition (2.8). These conditions give the existence of a unique solution to the above limit SDE and propagation of chaos results as shown in [25].

The function χ is modelled in a way such that it is a bounded Lipschitz function with compact support included in the interval $(0, 1)$. Moreover, the value of y^i such that $a_r^\alpha S_\alpha(V^i)(1 - y^i) +$

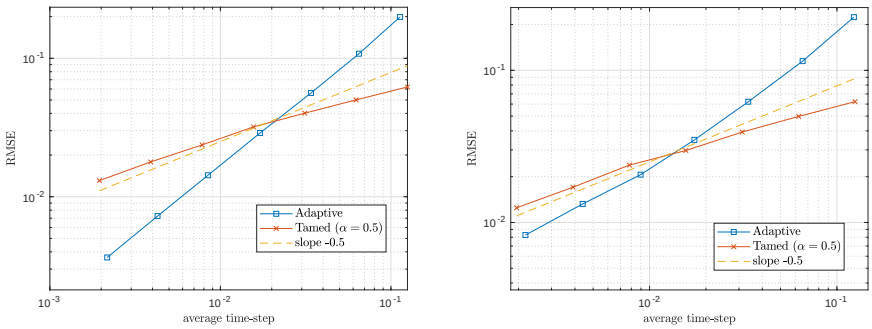
$a_d^\gamma y^i = 0$, for $a_r^\alpha S_\alpha(V^i), a_d^\gamma > 0$, never belongs to $[0,1]$. Thus, the square-root term in σ_α^y is not zero for $y^i \in [0,1]$. In addition, [25, Proposition 3.3] proves that for $0 \leq y_0^i \leq 1$ a. s., also $\mathbb{P}\left(\forall t \geq 0,\ 0 \leq y_t^i \leq 1\right) = 1$. For the numerical approximation with the Euler–Maruyama scheme, there is a positive probability that y becomes negative. We address this by an adaptation of the full-truncation Euler–Maruyama scheme, i.e., by simply setting this expression to zero, which has been shown to have good stability and accuracy for square-root diffusions (see, e.g., [110, 50, 51, 49], noting that they consider the more challenging case where the process can get arbitrarily close to the singularity).

For the adaptive time-step, we choose

$$h^\delta(x) = \delta \min(T, \gamma|x|^2/|f_\alpha(t,x)|^2),$$

for some parameter $\gamma > 0$. To motivate this choice, we note that for large $|X_t^i|$, $f_\alpha(t, X_t^i)$ is the driving part of the drift term and therefore we essentially achieve by this choice that $h(x)|b(t,x,\mu)| \approx c|x|^2$ for $h(x) = h^\delta(x)/\delta$.

The strong convergence rates of the tamed and adaptive schemes are depicted for two different volatilities in Fig. 2.4a and Fig. 2.4b. We can deduce from Fig. 2.4a that the adaptive scheme has a practical strong convergence rate of order 1, as the components of the diffusion matrix are either zero or close to zero. The rate is asymptotically 1/2 for larger volatilities (Fig. 2.4b).

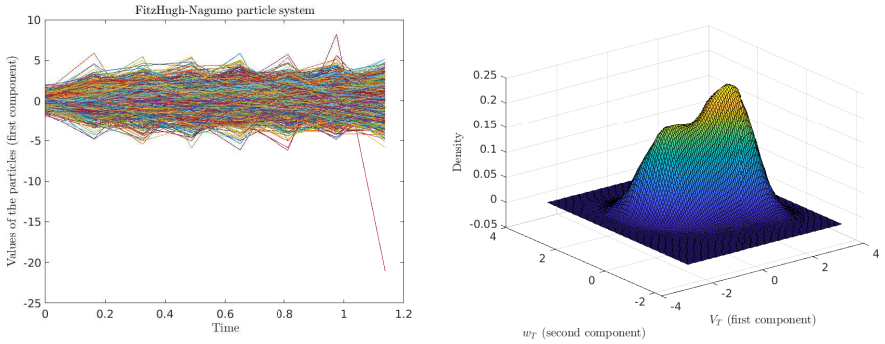


(a) $\sigma_{\text{ext}} = 0, \sigma^J = 0.00002$ (b) $\sigma_{\text{ext}} = 0.5, \sigma^J = 0.2$

Figure 2.4: Strong convergence for Example 4.

To additionally illustrate the importance of using a tamed or an adaptive time-stepping scheme for this model instead of a standard Euler–Maruyama scheme, we demonstrate numerically in Fig. 2.5a that this scheme yields approximations which start to strongly oscillate and potentially diverge after a finite time ("particle corruption" from [130]). For this test, we used $M = 2^3$ and $N = 10^3$.

Our scheme can be used to approximate marginal densities of the above limit equation derived from the FitzHugh–Nagumo network. Fig. 2.5b was obtained by the adaptive time-stepping method with $\delta = 1/2^8$, $T = 1$ and $N = 10^4$. We used a kernel density approach to obtain the density from the simulated data.



(a) (b)

Figure 2.5: Simulations of the FitzHugh–Nagumo network. (a) "Particle corruption" (in the first component) for a standard Euler–Maruyama scheme. (b) Approximation of the joint density of (V_T, w_T), i.e., the first two components.

2.4.5 Example 5 – Milstein scheme

To illustrate the performance of the adaptive Milstein scheme, we consider the following particle system

$$\mathrm{d}X_t^{i,N} = \left(-(X_t^{i,N})^5 + \frac{1}{N}\sum_{j=1}^{N} X_t^{j,N} \right) \mathrm{d}t + X_t^{i,N}\,\mathrm{d}W_t^i, \quad X_0^{i,N} = 1,$$

and use as benchmark the tamed Milstein scheme

$$Y_{t_{n+1}}^{i,N,M} = Y_{t_n}^{i,N,M} + \frac{-(Y_{t_n}^{i,N,M})^5 + \frac{1}{N}\sum_{j=1}^{N} Y_{t_n}^{j,N,M}}{1 + h\left|-(Y_{t_n}^{i,N,M})^5 + \frac{1}{N}\sum_{j=1}^{N} Y_{t_n}^{j,N,M}\right|} h + Y_{t_n}^{i,N,M}\Delta W_{t_n}^i + \frac{1}{2} Y_{t_n}^{i,N,M}((\Delta W_{t_n}^i)^2 - h),$$

where $Y_0^{i,N,M} = 1$, for all $i \in \{1, \ldots, N\}$. Fig. 2.6 illustrates the predicted strong convergence of order 1 for both the tamed and adaptive Milstein scheme. For the adaptive time-step function, we choose $h^\delta(x) = \delta \min(1, |x|^{-4})$. The number of particles was set to $N = 10^4$.

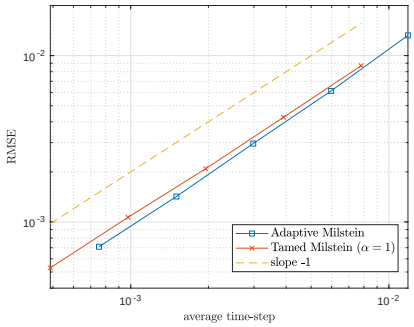


Figure 2.6: Strong convergence of the Milstein schemes.

2.5 Proofs of main results

We start by giving stability results for the processes $\hat{X}^{i,N}$ and $\tilde{X}^{i,N}$ defined in (2.11) and (2.12), respectively, for $i \in \{1, \dots, N\}$. Note that generic constants used in the proofs only depend on p, T, the moment bounds for the initial data and the constants appearing in the assumptions on b, σ and h, but are independent of N and δ.

2.5.1 Moment stability for super-linear drift coefficient

In this section, our main result on stability is first proven for $p \geq 4$. Then, Hölder's inequality allows to deduce the claim for $2 \leq p < 4$.

Proposition 2.5.1. *Let $p \geq 2$ and $\xi \in L_0^p(\mathbb{R}^d)$. Let Assumption (H.1) hold, and suppose the time-step function satisfies (H.2(3)). Then, T is almost surely attainable, i.e. $\mathbb{P}(\exists\ M(\omega) < \infty,\ s.t.\ t_{M(\omega)} \geq T) = 1$. Moreover, there exists a constant $C > 0$ such that*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |\hat{X}_t^{i,N}|^p\right] \vee \max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |\tilde{X}_t^{i,N}|^p\right] \leq C.$$

Proof. We follow a strategy similar to that in [59, Theorem 1] for non-Lipschitz standard SDEs and adapt the necessary steps for the measure dependence. We choose $K > \mathbb{E}[|X_0^i|^p]$, and then define a K-truncated process by $X_0^{K,i,N} := P_K(X_0^i)$ and

$$\hat{X}_{t_{n+1}}^{K,i,N} = P_K(\hat{X}_{t_n}^{K,i,N} + b(t_n, \hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N})h_n^{\min} + \sigma(t_n, \hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N})\Delta W_{t_n}^i), \tag{2.21}$$

where, $\mu_{t_n}^{\hat{\boldsymbol{X}}^N}$ is the empirical measure (defined as above in (2.10), but using the N truncated processes $\hat{X}_{t_n}^{K,i,N}$), $P_K(Y) := \min(1, K/|Y|)Y$, and consequently $|\hat{X}_{t_n}^{K,i,N}| \leq K$, for all $i \in \{1, \dots, N\}$ and $n \geq 0$. The continuous time interpolation for $\hat{X}_t^{K,i,N}$ is defined analogously to above (see (2.11)).

The reason for introducing $\hat{X}^{K,i,N}$ is that one can guarantee that the time-step function has a strictly positive lower bound, i.e., $\inf_{|x|\leq K} h(x) > 0$. Hence, T is attainable. Showing that the moments of the K-truncated processes are bounded by some constant $C > 0$ independent of K allows to send K to infinity and the result will then follow in combination with the monotone convergence theorem.

The following analysis is performed for a fixed particle. We get, using (2.21),

$$\begin{aligned}|\hat{X}_{t_{n+1}}^{K,i,N}|^2 \leq |\hat{X}_{t_n}^{K,i,N}|^2 + 2h_n^{\min}\left(\left\langle \hat{X}_{t_n}^{K,i,N}, b(t_n, \hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N})\right\rangle + \frac{1}{2}h_n^{\min}|b(t_n, \hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N})|^2\right)\\ + 2\left\langle \phi(\hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N}), \sigma(t_n, \hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N})\Delta W_{t_n}^i\right\rangle + |\sigma(t_n, \hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N})\Delta W_{t_n}^i|^2,\end{aligned} \tag{2.22}$$

where

$$\phi(\hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N}) := \hat{X}_{t_n}^{K,i,N} + h_n^{\min} b(t_n, \hat{X}_{t_n}^{K,i,N}, \mu_{t_n}^{\hat{\boldsymbol{X}}^N}).$$

Using Assumption (H.2(3)) for the time-step function, we get for (2.22)

$$|\hat{X}^{K,i,N}_{t_{n+1}}|^2 \le |\hat{X}^{K,i,N}_{t_n}|^2 + 2L_c|\hat{X}^{K,i,N}_{t_n}|^2 h^{\min}_n + 2L_d h^{\min}_n + 2\left\langle \phi(\hat{X}^{K,i,N}_{t_n}, \mu^{\hat{\boldsymbol{X}}^N}_{t_n}), \sigma(t_n, \hat{X}^{K,i,N}_{t_n}, \mu^{\hat{\boldsymbol{X}}^N}_{t_n})\Delta W^i_{t_n}\right\rangle + |\sigma(t_n, \hat{X}^{K,i,N}_{t_n}, \mu^{\hat{\boldsymbol{X}}^N}_{t_n})\Delta W^i_{t_n}|^2. \tag{2.23}$$

Summing (2.23) over multiple time-steps, adding the contribution from $\underline{t}$ to t and then using Jensen's inequality, results in

$$\begin{aligned}|\hat{X}^{K,i,N}_t|^p \le C\Bigg[&|X^{K,i,N}_0|^p + \left(2L_c\int_0^t |\bar{X}^{K,i,N}_s|^2\,\mathrm{d}s\right)^{p/2} + (2tL_d)^{p/2} \\ &+ \left|2\int_0^{\underline{t}}\left\langle \phi(\bar{X}^{K,i,N}_s, \mu^{\bar{\boldsymbol{X}}^N}_s), \sigma(\underline{s}, \bar{X}^{K,i,N}_s, \mu^{\bar{\boldsymbol{X}}^N}_s)\,\mathrm{d}W^i_s\right\rangle\right|^{p/2} \\ &+ \left(\sum_{k=0}^{n_t-1}|\sigma(t_k, \bar{X}^{K,i,N}_{t_k}, \mu^{\bar{\boldsymbol{X}}^N}_{t_k})\Delta W^i_{t_k}|^2\right)^{p/2} \\ &+ \left|2\left\langle \bar{X}^{K,i,N}_t + b(\underline{t}, \bar{X}^{K,i,N}_t, \mu^{\bar{\boldsymbol{X}}^N}_{t_n})(t-\underline{t}), \sigma(\underline{t}, \bar{X}^{K,i,N}_t, \mu^{\bar{\boldsymbol{X}}^N}_t)(W^i_t - W^i_{\underline{t}})\right\rangle\right|^{p/2} \\ &+ |\sigma(\underline{t}, \bar{X}^{K,i,N}_t, \mu^{\bar{\boldsymbol{X}}^N}_t)(W^i_t - W^i_{\underline{t}})|^p\Bigg],\end{aligned} \tag{2.24}$$

where we recall $n_t = \max\{n : t_n \le t\}$ and $C > 0$ is a constant depending on p only. Taking the supremum (over t) and the expectation in (2.24) gives

$$\mathbb{E}\left[\sup_{s\in[0,t]}|\hat{X}^{K,i,N}_s|^p\right] \le C\sum_{j=1}^5 \Pi^{j,i}_t,$$

where

$$\begin{aligned}
\Pi^{1,i}_t &:= \mathbb{E}\left[|X^{K,i,N}_0|^p\right] + \mathbb{E}\left[\left(2L_c\int_0^t |\bar{X}^{K,i,N}_s|^2\,\mathrm{d}s\right)^{p/2}\right] + (2tL_d)^{p/2}, \\
\Pi^{2,i}_t &:= \mathbb{E}\left[\sup_{s\in[0,\underline{t}]}\left|2\int_0^s\left\langle \phi(\bar{X}^{K,i,N}_u, \mu^{\bar{\boldsymbol{X}}^N}_u), \sigma(\underline{u}, \bar{X}^{K,i,N}_u, \mu^{\bar{\boldsymbol{X}}^N}_u)\,\mathrm{d}W^i_u\right\rangle\right|^{p/2}\right], \\
\Pi^{3,i}_t &:= \mathbb{E}\left[\left(\sum_{k=0}^{n_t-1}|\sigma(t_k, \bar{X}^{K,i,N}_{t_k}, \mu^{\bar{\boldsymbol{X}}^N}_{t_k})\Delta W^i_{t_k}|^2\right)^{p/2}\right], \\
\Pi^{4,i}_t &:= \mathbb{E}\left[\sup_{s\in[0,t]}\left|2\left\langle \bar{X}^{K,i,N}_s + b(\underline{s}, \bar{X}^{K,i,N}_s, \mu^{\bar{\boldsymbol{X}}^N}_s)(s-\underline{s}), \sigma(\underline{s}, \bar{X}^{K,i,N}_s, \mu^{\bar{\boldsymbol{X}}^N}_s)(W^i_s - W^i_{\underline{s}})\right\rangle\right|^{p/2}\right], \\
\Pi^{5,i}_t &:= \mathbb{E}\left[\sup_{s\in[0,t]}|\sigma(\underline{s}, \bar{X}^{K,i,N}_s, \mu^{\bar{\boldsymbol{X}}^N}_s)(W^i_s - W^i_{\underline{s}})|^p\right].
\end{aligned}$$

Note that Assumption (H.2(3)) implies that

$$\begin{aligned}|\phi(\bar{X}^{K,i,N}_s, \mu^{\bar{\boldsymbol{X}}^N}_s)|^2 &= |\bar{X}^{K,i,N}_s + h^{\min}_{n_s} b(\underline{s}, \bar{X}^{K,i,N}_s, \mu^{\bar{\boldsymbol{X}}^N}_s)|^2 \\ &\le |\bar{X}^{K,i,N}_s|^2 + 2h^{\min}_{n_s}(L_c|\bar{X}^{K,i,N}_s|^2 + L_d) \\ &\le (1 + 2L_cT)|\bar{X}^{K,i,N}_s|^2 + 2L_dT,\end{aligned}$$

and consequently

$$|\phi(\bar{X}_s^{K,i,N}, \mu_s^{\bar{\boldsymbol{X}}^N})|^{p/2} \leq 2^{p/4-1}\left((1+2L_cT)^{p/4}|\bar{X}_s^{K,i,N}|^{p/2} + (2TL_d)^{p/4}\right). \tag{2.25}$$

Further, we have that

$$|\bar{X}_t^{K,i,N} + b(\underline{t}, \bar{X}_t^{K,i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})(t-\underline{t})|^{p/2} \leq C\left(|\bar{X}_t^{K,i,N}|^{p/2}+1\right), \tag{2.26}$$

where we employed (H.2(3)) and $C>0$ is a constant depending on p, T and the constants appearing in the assumptions on the coefficient b and in (H.2(3)).

Using this, we note that each of the terms $\Pi_t^{1,i},\dots,\Pi_t^{5,i}$ can be estimated similarly to the proof of [59, Theorem 1], employing additionally the bound

$$\|\sigma(t, \bar{X}_t^{K,i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})\| \leq C\left(|\bar{X}_t^{K,i,N}|+1\right),$$

for $C>0$ (depending on the Lipschitz constant for σ and the constant appearing in (H.1(6))). To be precise, having estimates (2.25) and (2.26) in mind, along with the fact that the initial data has finite moments up to order p, each $\Pi_t^{j,i}$, $j\in\{1,\dots,5\}$, can be bounded by

$$\Pi_t^{j,i} \leq C\left(1+\int_0^t \mathbb{E}\left[\sup_{u\in[0,s]}|\hat{X}_u^{K,i,N}|^p\right]\mathrm{d}s\right),$$

where $C>0$ is a constant independent of K, and N and the particle index i. This allows to conclude the stability of the processes $\hat{X}^{i,N,K}$ using Gronwall's inequality.

It remains to show the attainability of T for the non-truncated particle system. The stability of $\hat{X}^{i,N}$ will then be a consequence of the monotone convergence theorem. We have, for any $\omega\in\Omega$, $\max_{j\in\{1,\dots,N\}}\hat{X}_t^{K,j,N}(\omega) = \max_{j\in\{1,\dots,N\}}\hat{X}_t^{j,N}(\omega)$ for all $t\in[0,T]$ if and only if the inequality $\max_{j\in\{1,\dots,N\}}\sup_{t\in[0,T]}|\hat{X}_t^{j,N}(\omega)|\leq K$ holds. Employing Markov's inequality, we derive

$$\begin{aligned}\mathbb{P}\left(\max_{j\in\{1,\dots,N\}}\sup_{t\in[0,T]}|\hat{X}_t^{j,N}|<K\right) &= \mathbb{P}\left(\max_{j\in\{1,\dots,N\}}\sup_{t\in[0,T]}|\hat{X}_t^{K,j,N}|<K\right)\\ &\leq 1-N\frac{\mathbb{E}\left[\sup_{t\in[0,T]}|\hat{X}_t^{K,i,N}|^4\right]}{K^4}\to 1,\end{aligned}$$

as $K\to\infty$, which implies that $\max_{j\in\{1,\dots,N\}}\sup_{t\in[0,T]}|\hat{X}_t^{j,N}(\omega)|<\infty$ and hence T is attainable. The remaining arguments follow along the same lines as in [59, Theorem 1, step 4]. The moment stability of $\tilde{X}^{i,N}$ can be proven in a similar manner by using Assumption (H.1(6)). □

2.5.2 Convergence of Scheme 2 (Proof of Theorem 2.2.1)

Proof. Let $X^{i,N}$ be given by (2.3), for any $i\in\{1,\dots,N\}$. It is shown in [130, Proposition 3.1] that $\max_{i\in\{1,\dots,N\}}\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^i - X_t^{i,N}|^2\right]\leq C\varphi(N)$ for some C independent of N. It therefore remains to show here that $\max_{i\in\{1,\dots,N\}}\sup_{t\in[0,T]}\mathbb{E}\left[|X_t^{i,N}-\tilde{X}_t^{i,N}|^2\right]\leq C\delta$ for some C independent of N and δ.

In what follows, we will drop for ease of notation the explicit time-dependence of the coefficients. Then, we may write for $t \in [t_n, t_{n+1}]$ for some $n \geq 1$,

$$|X_t^{i,N} - \tilde{X}_t^{i,N}|^2 = \Big| X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N} + \int_{\underline{t}}^t b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}s - b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(t-\underline{t}) \\ \int_{\underline{t}}^t \sigma(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}W_s^i - \sigma(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i)\Big|^2.$$

Squaring the term on the right side and taking expectations yields

$$\begin{aligned}\mathbb{E}[|X_t^{i,N} - \tilde{X}_t^{i,N}|^2] \leq \mathbb{E}[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2] &+ 2\mathbb{E}\left[\left|\int_{\underline{t}}^t b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}s - b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(t-\underline{t})\right|^2\right] \\ &+ 2\mathbb{E}\left[\left|\int_{\underline{t}}^t \sigma(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}W_s^i - \sigma(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i)\right|^2\right] \\ &+ 2\mathbb{E}\left\langle X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}, \int_{\underline{t}}^t b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}s - b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(t-\underline{t})\right\rangle. \end{aligned} \tag{2.27}$$

Now, observe that Hölder's inequality implies

$$\begin{aligned}&\mathbb{E}\left[\left|\int_{\underline{t}}^t b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}s - b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(t-\underline{t})\right|^2\right] \\ &\qquad \leq C\left(\mathbb{E}\left[(t-\underline{t})\int_{\underline{t}}^t |b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})|^2\,\mathrm{d}s\right] + \mathbb{E}\left[|b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})|^2|t-\underline{t}|^2\right]\right) \leq C\delta^2,\end{aligned}$$

for some constant $C > 0$, where, in addition we used Assumptions (H.1(4)) and (H.1(6)) and the moment stability of $X^{i,N}$ and $\tilde{X}^{i,N}$. Also, recall that $t - \underline{t} \leq \delta T$.

Next, we bound

$$\begin{aligned}\mathbb{E}\left[\left|\int_{\underline{t}}^t \sigma(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}W_s^i - \sigma(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i)\right|^2\right] &\leq \sum_{j=1}^3 \Pi_t^{j,i}, \qquad \text{where} \\ \Pi_t^{1,i} &= 3\mathbb{E}\left[\left|\int_{\underline{t}}^t \sigma(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}W_s^i - \sigma(X_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})(W_t^i - W_{\underline{t}}^i)\right|^2\right], \\ \Pi_t^{2,i} &= 3\mathbb{E}\left[\left|\sigma(X_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})(W_t^i - W_{\underline{t}}^i) - \sigma(\tilde{X}_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\tilde{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i)\right|^2\right], \\ \Pi_t^{3,i} &= 3\mathbb{E}\left[\left|\sigma(\tilde{X}_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\tilde{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i) - \sigma(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i)\right|^2\right].\end{aligned}$$

By virtue of Assumption (H.1(1)), we get the estimate

$$\|\sigma(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N}) - \sigma(X_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})\|^2 \leq C\left(|X_t^{i,N} - X_{\underline{t}}^{i,N}|^2 + \mathcal{W}_2^2(\mu_t^N, \mu_{\underline{t}}^{\boldsymbol{X}^N})\right).$$

Recalling the bound

$$\mathcal{W}_2^2(\mu_t^{\boldsymbol{X}^N}, \mu_{\underline{t}}^{\boldsymbol{X}^N}) \leq \frac{1}{N}\sum_{j=1}^N |X_t^{j,N} - X_{\underline{t}}^{j,N}|^2,$$

along with the standard one-step estimate $\mathbb{E}[\sup_{s\in[\underline{t},t]} |X_s^{i,N} - X_{\underline{t}}^{i,N}|^2] \leq C\delta$ and BDG's inequality, allows us to show $\Pi_t^{1,i} \leq C\delta^2$, for some constant $C > 0$. Using again (H.1(1)), a conditional expectation argument and estimates for Brownian increments, we derive

$$\Pi_t^{2,i} \leq C\delta\mathbb{E}[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2].$$

It remains to analyse $\Pi_t^{3,i}$, where, for ease of notation, we consider $k_n = 1$ and $T = 1$ so that $\delta T = \delta$. Hence, we observe for $s \in [\delta, 2\delta)$

$$\begin{aligned}
\mathbb{E}[|\tilde{X}_s^{i,N} - \tilde{X}_\delta^{i,N}|^2] &= \mathbb{E}[|\tilde{X}_{\underline{s}}^{i,N} + b(\underline{s}, \tilde{X}_{\underline{s}}^{i,N}, \mu_\delta^{\tilde{\boldsymbol{X}}^N})(s-\underline{s}) + \sigma(\underline{s}, \tilde{X}_{\underline{s}}^{i,N}, \mu_\delta^{\tilde{\boldsymbol{X}}^N})(W_s^i - W_{\underline{s}}^i) - \tilde{X}_\delta^{i,N}|^2] \\
&= \mathbb{E}\left[\left|\int_\delta^s b(\underline{u}, \tilde{\tilde{X}}_u^{i,N}, \mu_u^{\tilde{\tilde{\boldsymbol{X}}}^N})\,\mathrm{d}u + \int_\delta^s \sigma(\underline{u}, \tilde{\tilde{X}}_u^{i,N}, \mu_u^{\tilde{\tilde{\boldsymbol{X}}}^N})\,\mathrm{d}W_u^i\right|^2\right] \\
&\le 2\mathbb{E}\left[\left|\int_\delta^s b(\underline{u}, \tilde{\tilde{X}}_u^{i,N}, \mu_u^{\tilde{\tilde{\boldsymbol{X}}}^N})\,\mathrm{d}u\right|^2\right] + 2\mathbb{E}\left[\int_\delta^s \|\sigma(\underline{u}, \tilde{\tilde{X}}_u^{i,N}, \mu_u^{\tilde{\tilde{\boldsymbol{X}}}^N})\|^2\,\mathrm{d}u\right] \\
&\le 2\mathbb{E}\left[\delta\int_\delta^s |b(\underline{u}, \tilde{\tilde{X}}_u^{i,N}, \mu_u^{\tilde{\tilde{\boldsymbol{X}}}^N})|^2\,\mathrm{d}u\right] + 2\mathbb{E}\left[\int_\delta^s \|\sigma(\underline{u}, \tilde{\tilde{X}}_u^{i,N}, \mu_u^{\tilde{\tilde{\boldsymbol{X}}}^N})\|^2\,\mathrm{d}u\right] \\
&\le C\delta, \qquad (2.28)
\end{aligned}$$

for some $C > 0$ and where the last inequality is due to the polynomial growth bounds on b and σ and the moment stability of the process $\tilde{X}^{i,N}$. Hence, (H.1(1)) and estimates for Brownian increments allow us to deduce that $\Pi_t^{3,i} \le C\delta^2$.

Returning to (2.27), we write the last term there as

$$\begin{aligned}
&\mathbb{E}\left\langle X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}, \int_{\underline{t}}^t b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}s - b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(t-\underline{t})\right\rangle = \sum_{j=1}^3 \Lambda_t^{j,i}, \qquad \text{where} \\
&\Lambda_t^{1,i} = \mathbb{E}\left\langle X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}, \int_{\underline{t}}^t b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}s - b(X_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})(t-\underline{t})\right\rangle \\
&\Lambda_t^{2,i} = \mathbb{E}\left\langle X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}, b(X_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})(t-\underline{t}) - b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\tilde{\boldsymbol{X}}^N})(t-\underline{t})\right\rangle \\
&\Lambda_t^{3,i} = \mathbb{E}\left\langle X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}, b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\tilde{\boldsymbol{X}}^N})(t-\underline{t}) - b(\tilde{X}_{\underline{t}}^{i,N}, \mu_{k_n\delta T}^{\tilde{\boldsymbol{X}}^N})(t-\underline{t})\right\rangle.
\end{aligned}$$

In the sequel, we will analyse these terms one-by-one. First, using Young's inequality and Hölder's inequality plus taking (H.1(3)) and (H.1(4)) into account, we get

$$\begin{aligned}
\Lambda_t^{1,i} &\le \mathbb{E}\left[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|\delta^{1/2}\delta^{-1/2}\left|\int_{\underline{t}}^t b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}s - b(X_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})(t-\underline{t})\right|\right] \\
&\le C\left(\delta\mathbb{E}\left[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2\right] + \delta^{-1}\mathbb{E}\left[\left|\int_{\underline{t}}^t b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\,\mathrm{d}s - b(X_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})(t-\underline{t})\right|^2\right]\right) \\
&\le C\left(\delta\mathbb{E}\left[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2\right] + \delta^{-1}\mathbb{E}\left[(t-\underline{t})\int_{\underline{t}}^t |b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N}) - b(X_{\underline{t}}^{i,N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})|^2\,\mathrm{d}s\right]\right) \\
&\le C\left(\delta\mathbb{E}\left[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2\right] + \mathbb{E}\left[\int_{\underline{t}}^t \left((1 + |X_s^{i,N}|^q + |X_{\underline{t}}^{i,N}|^q)^2|X_s^{i,N} - X_{\underline{t}}^{i,N}|^2 + \mathcal{W}_2^2(\mu_s^{\boldsymbol{X}^N}, \mu_{\underline{t}}^{\boldsymbol{X}^N})\right)\,\mathrm{d}s\right]\right) \\
&\le C\left(\delta\mathbb{E}\left[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2\right] + \delta^2\right),
\end{aligned}$$

where we also used the one-step estimate $\mathbb{E}[\sup_{s\in[\underline{t},t]}|X_s^{i,N} - X_{\underline{t}}^{i,N}|^2] \le C\delta$ and the stability of $X^{i,N}$. By similar arguments, in combination with (H.1(2)), we derive

$$\Lambda_t^{2,i} \le C\delta\mathbb{E}\left[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2\right].$$

In a similar manner, recalling estimate (2.28), we obtain for the last term

$$\Lambda_t^{3,i} \leq C\left(\delta\mathbb{E}\left[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2\right] + \delta^2\right).$$

Collecting the above estimates, we conclude

$$\mathbb{E}[|X_t^{i,N} - \tilde{X}_t^{i,N}|^2] \leq (1 + C\delta)\mathbb{E}[|X_{\underline{t}}^{i,N} - \tilde{X}_{\underline{t}}^{i,N}|^2] + C\delta^2.$$

An inductive argument over all time-steps in combination with Proposition 2.2.1 gives the claim. □

2.5.3 Moment stability for super-linear diffusion coefficient

Proposition 2.5.2. *Let $p \geq 2$ and $X_0 \in L_0^p(\mathbb{R}^d)$. Let Assumptions (H.3) and (H.1(3))–(H.1(6)) hold, and suppose the time-step function satisfies (H.4). Then, T is almost surely attainable, i.e. $\mathbb{P}(\exists\ M(\omega) < \infty,\ s.t.\ t_{M(\omega)} \geq T) = 1$. Moreover, there exists a constant $C > 0$ such that*

$$\max_{i\in\{1,\dots,N\}} \sup_{t\in[0,T]} \mathbb{E}\left[|\hat{X}_t^{i,N}|^p\right] \leq C.$$

Proof. Using equation (2.11), we get for some $C > 0$

$$|\hat{X}_t^{i,N} - \hat{X}_{\underline{t}}^{i,N}|^p \leq C\left(|b(\underline{t}, \bar{X}_t^{i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})|^p(t-\underline{t})^p + |\sigma(\underline{t}, \bar{X}_t^{i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i)|^p\right),$$

which on multiplying with the drift and taking expectations on both sides yields

$$\begin{aligned}
&\mathbb{E}\left[|\hat{X}_t^{i,N} - \hat{X}_{\underline{t}}^{i,N}|^p |b(\underline{t}, \bar{X}_t^{i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})|^p\right] \\
&\leq C\left(\mathbb{E}\left[|b(\underline{t}, \bar{X}_t^{i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})|^{2p}(t-\underline{t})^p\right] + \mathbb{E}\left[|\sigma(\underline{t}, \bar{X}_t^{i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i)|^p |b(\underline{t}, \bar{X}_t^{i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})|^p\right]\right) \\
&\leq C\left(\mathbb{E}\left[(1 + |\hat{X}_{\underline{t}}^{i,N}|^2)^{p/2}\right] + \mathbb{E}\left[\mathbb{E}\left[|\sigma(\underline{t}, \bar{X}_t^{i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})(W_t^i - W_{\underline{t}}^i)|^p |b(\underline{t}, \bar{X}_t^{i,N}, \mu_t^{\bar{\boldsymbol{X}}^N})|^p \middle| \mathcal{F}_{\underline{t}}\right]\right]\right) \\
&\leq C\mathbb{E}\left[(1 + |\hat{X}_{\underline{t}}^{i,N}|^2)^{p/2}\right], \qquad (2.29)
\end{aligned}$$

where we used $\mathbb{E}\left[|W_t^i - W_{\underline{t}}^i|^p \middle| \mathcal{F}_{\underline{t}}\right] \leq C(t-\underline{t})^{p/2}$, for some $C > 0$, along with (H.4) and Remark 2.3.1. This one-step error estimate will be essential at a later stage of the proof.

Now, we consider the stopping time $\tau_R := \min_{i\in\{1,\dots,N\}} \tau_R^i$, where $\tau_R^i := \inf\{t \geq 0 : |\hat{X}_t^{i,N}| \geq R\}$. Due to (H.4), $T \wedge \tau_R$ is attainable. Therefore, Itô's formula for the stopped process $\hat{X}_{t\wedge\tau_R}^{i,N}$ gives

$$\begin{aligned}
\left(1 + |\hat{X}_{t\wedge\tau_R}^{i,N}|^2\right)^{p/2} &= \left(1 + |\hat{X}_0^{i,N}|^2\right)^{p/2} + p\int_0^{t\wedge\tau_R} \left(1 + |\hat{X}_s^{i,N}|^2\right)^{p/2-1} \left\langle \hat{X}_s^{i,N}, b(\underline{s}, \bar{X}_s^{i,N}, \mu_s^{\bar{\boldsymbol{X}}^N})\right\rangle \mathrm{d}s \\
&\quad + p\int_0^{t\wedge\tau_R} \left(1 + |\hat{X}_s^{i,N}|^2\right)^{p/2-1} \left\langle \hat{X}_s^{i,N}, \sigma(\underline{s}, \bar{X}_s^{i,N}, \mu_s^{\bar{\boldsymbol{X}}^N})\,\mathrm{d}W_s^i\right\rangle \\
&\quad + \frac{p(p-2)}{2}\int_0^{t\wedge\tau_R} \left(1 + |\hat{X}_s^{i,N}|^2\right)^{p/2-2} \left|\sigma^\top(\underline{s}, \bar{X}_s^{i,N}, \mu_s^{\bar{\boldsymbol{X}}^N})\hat{X}_s^{i,N}\right|^2 \mathrm{d}s \\
&\quad + \frac{p}{2}\int_0^{t\wedge\tau_R} \left(1 + |\hat{X}_s^{i,N}|^2\right)^{p/2-1} \left\|\sigma(\underline{s}, \bar{X}_s^{i,N}, \mu_s^{\bar{\boldsymbol{X}}^N})\right\|^2 \mathrm{d}s,
\end{aligned}$$

almost surely for any $t \in [0,T]$. Thus, on taking expectation and using the Cauchy-Schwarz inequality

$$\begin{aligned}\mathbb{E}\left[\left(1+|\hat{X}^{i,N}_{t\wedge\tau_R}|^2\right)^{p/2}\right] &\leq \mathbb{E}\left[\left(1+|\hat{X}^{i,N}_0|^2\right)^{p/2}\right] \\ &+ p\mathbb{E}\left[\int_0^{t\wedge\tau_R}\left(1+|\hat{X}^{i,N}_s|^2\right)^{p/2-1}\left(\left\langle \bar{X}^{i,N}_s, b(\underline{s},\bar{X}^{i,N}_s,\mu_s^{\bar{\boldsymbol{X}}^N})\right\rangle + \frac{(p-1)}{2}\left\|\sigma(\underline{s},\bar{X}^{i,N}_s,\mu_s^{\bar{\boldsymbol{X}}^N})\right\|^2\right)\mathrm{d}s\right] \\ &+ p\mathbb{E}\left[\int_0^{t\wedge\tau_R}\left(1+|\hat{X}^{i,N}_s|^2\right)^{p/2-1}\left\langle \hat{X}^{i,N}_s - \hat{X}^{i,N}_{\underline{s}}, b(\underline{s},\bar{X}^{i,N}_s,\mu_s^{\bar{\boldsymbol{X}}^N})\right\rangle \mathrm{d}s\right].\end{aligned}$$

Hence, by (H.3(2)) we further obtain, for some constant $C>0$ (depending on the constants in (H.3(2)))

$$\mathbb{E}\left[\left(1+|\hat{X}^{i,N}_{t\wedge\tau_R}|^2\right)^{p/2}\right] \leq \mathbb{E}\left[\left(1+|\hat{X}^{i,N}_0|^2\right)^{p/2}\right] + C\int_0^t \sup_{r\in[0,s]}\mathbb{E}\left[\left(1+|\hat{X}^{i,N}_{r\wedge\tau_R}|^2\right)^{p/2}\right]\mathrm{d}s + p\Pi_t, \quad (2.30)$$

for any $t\in[0,T]$, where Π_t is is given by

$$\Pi_t := \mathbb{E}\left[\int_0^{t\wedge\tau_R}\left(1+|\hat{X}^{i,N}_s|^2\right)^{p/2-1}\left\langle \hat{X}^{i,N}_s - \hat{X}^{i,N}_{\underline{s}}, b(\underline{s},\bar{X}^{i,N}_s,\mu_s^{\bar{\boldsymbol{X}}^N})\right\rangle \mathrm{d}s\right].$$

This gives, using Hölder's inequality along with estimate (2.29) and (H.4),

$$\begin{aligned}\Pi_t &\leq C\left(\int_0^t \sup_{r\in[0,s]}\mathbb{E}\left[\left(1+|\hat{X}^{i,N}_{r\wedge\tau_R}|^2\right)^{p/2}\right] + \mathbb{E}\left[\int_0^{t\wedge\tau_R}|\hat{X}^{i,N}_s - \hat{X}^{i,N}_{\underline{s}}|^{p/2}|b(\underline{s},\bar{X}^{i,N}_s,\mu_s^{\bar{\boldsymbol{X}}^N})|^{p/2}\,\mathrm{d}s\right]\right) \\ &\leq C\int_0^t \sup_{r\in[0,s]}\mathbb{E}\left[\left(1+|\hat{X}^{i,N}_{r\wedge\tau_R}|^2\right)^{p/2}\right]. \quad (2.31)\end{aligned}$$

for some constant $C>0$. Combining equations (2.30) and (2.31) and using Gronwall's inequality yields the claim, up to time $T\wedge\tau_R$. Now, note that

$$\mathbb{P}(\tau_R \leq T) \leq \frac{1}{R^2}\mathbb{E}\left[\max_{i\in\{1,\dots,N\}}|\hat{X}^{i,N}_{T\wedge\tau_R}|^2\right] \leq \frac{CN}{R^2},$$

and hence,

$$\mathbb{P}\left(\max_{i\in\{1,\dots,N\}}\sup_{t\in[0,T]}|\hat{X}^{i,N}_t| < R\right) = 1-\mathbb{P}(\tau_R\leq T)\to 1, \text{ as } R\to\infty.$$

Therefore, almost surely, $\max_{i\in\{1,\dots,N\}}\sup_{t\in[0,T]}|\hat{X}^{i,N}_t|<\infty$ and T is attainable. Using this and the stability up to time $T\wedge\tau_R$, the claim follows from Fatou's lemma. □

2.5.4 Moment stability of Milstein scheme

Proposition 2.5.3. *Let $p\geq 2$ and $\xi\in L^p_0(\mathbb{R}^d)$. Let Assumption (H.5) hold, and suppose the time-step function satisfies (H.2(3)). Then, T is almost surely attainable, i.e. $\mathbb{P}(\exists\ M(\omega)<\infty,\ s.t.\ t_{M(\omega)}\geq T)=1$. Moreover, there exists a constant $C>0$ such that for $\hat{Y}$ defined in (2.19)*

$$\max_{i\in\{1,\dots,N\}}\mathbb{E}\left[\sup_{t\in[0,T]}|\hat{Y}^{i,N}_t|^p\right]\leq C.$$

Proof. We outline the difference to the proof of the moment stability of the adaptive Euler–Maruyama scheme. The K-scheme for the Milstein discretisation reads as follows:

$$\hat{Y}^{K,i,N}_{t_{n+1}} = P_K\Big(\hat{Y}^{K,i,N}_{t_n} + \frac{1}{N}\sum_{j=1}^{N} b(\hat{Y}^{K,i,N}_{t_n}, \hat{Y}^{K,j,N}_{t_n})h^{\min}_n + \sigma(\hat{Y}^{K,i,N}_{t_n})\Delta W^i_{t_n} + \frac{1}{2}\sum_{j_1,j_2=1}^{k} L^{j_1}\sigma_{j_2}(\hat{Y}^{K,i,N}_{t_n})(\Delta W^{i,j_1}_{t_n}\Delta W^{i,j_2}_{t_n} - \delta_{j_1,j_2}h^{\min}_n)\Big).$$

Hence, using

$$M^{K,i}_{t_n} := \frac{1}{2}\sum_{j_1,j_2=1}^{k} L^{j_1}\sigma_{j_2}(\hat{Y}^{K,i,N}_{t_n})(\Delta W^{i,j_1}_{t_n}\Delta W^{i,j_2}_{t_n} - \delta_{j_1,j_2}h^{\min}_n),$$

we get

$$\begin{aligned}
|\hat{Y}^{K,i,N}_{t_{n+1}}|^2 &\le |\hat{Y}^{K,i,N}_{t_n}|^2 + 2h^{\min}_n\left(\left\langle \hat{Y}^{K,i,N}_{t_n}, \frac{1}{N}\sum_{j=1}^{N} b(\hat{Y}^{K,i,N}_{t_n}, \hat{Y}^{j,N}_{t_n})\right\rangle + \frac{1}{2}h^{\min}_n\left|\frac{1}{N}\sum_{j=1}^{N} b(\hat{Y}^{K,i,N}_{t_n}, \hat{Y}^{j,N}_{t_n})\right|^2\right)\\
&\quad + 2\left\langle \phi(\hat{Y}^{K,i,N}_{t_n}), \sigma(\hat{Y}^{K,i,N}_{t_n})\Delta W^i_{t_n}\right\rangle + |\sigma(\hat{Y}^{K,i,N}_{t_n})\Delta W^i_{t_n}|^2 + |M^{K,i}_{t_n}|^2\\
&\quad + 2\left\langle \hat{Y}^{K,i,N}_{t_n}, M^{K,i}_{t_n}\right\rangle + 2\left\langle \sigma(\hat{Y}^{K,i,N}_{t_n})\Delta W^i_{t_n}, M^{K,i}_{t_n}\right\rangle + 2h^{\min}_n\left\langle \frac{1}{N}\sum_{j=1}^{N} b(\hat{Y}^{K,i,N}_{t_n}, \hat{Y}^{j,N}_{t_n}), M^{K,i}_{t_n}\right\rangle\\
&\le |\hat{Y}^{K,i,N}_{t_n}|^2 + 2h^{\min}_n\left(\left\langle \hat{Y}^{K,i,N}_{t_n}, \frac{1}{N}\sum_{j=1}^{N} b(\hat{Y}^{K,i,N}_{t_n}, \hat{Y}^{j,N}_{t_n})\right\rangle + \frac{3}{2}h^{\min}_n\left|\frac{1}{N}\sum_{j=1}^{N} b(\hat{Y}^{K,i,N}_{t_n}, \hat{Y}^{j,N}_{t_n})\right|^2\right)\\
&\quad + 2\left\langle \phi(\hat{Y}^{K,i,N}_{t_n}), \sigma(\hat{Y}^{K,i,N}_{t_n})\Delta W^i_{t_n}\right\rangle + 2|\sigma(\hat{Y}^{K,i,N}_{t_n})\Delta W^i_{t_n}|^2 + 3|M^{K,i}_{t_n}|^2 + 2\left\langle \hat{Y}^{K,i,N}_{t_n}, M^{K,i}_{t_n}\right\rangle.
\end{aligned}$$

Now, using the same techniques as in Proposition 2.5.1, we obtain the bound (see [68] for details)

$$\mathbb{E}\left[\left(\sum_{n=0}^{n_t-1}|M^{K,i}_{t_n}|^2\right)^{p/2}\right] \le C\left(\int_0^t \mathbb{E}\left[\sup_{u\in[0,s]}|Y^{K,i,N}_u|^p\right]\mathrm{d}s + 1\right),$$

for some constant $C > 0$. Then, Gronwall's inequality readily allows to prove stability of the truncated particle system. The attainability of T can be argued as in the proof of Proposition 2.5.1. Then, a monotone convergence argument yields the claim for the particles $\hat{Y}^{i,N}$. □

Chapter 3

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

3.1 Introduction

A McKean–Vlasov equation (introduced in [114, 86]) for a d-dimensional process $X = (X_t)_{t\in[0,T]}$, for some given $T > 0$, is an SDE where the underlying coefficients depend on the current state X_t and, additionally, on the law of X_t, i.e.,

$$\mathrm{d}X_t = b(X_t, \mathcal{L}_{X_t})\,\mathrm{d}t + \sigma(X_t, \mathcal{L}_{X_t})\,\mathrm{d}W_t, \quad X_0 = \xi \in L_0^2(\mathbb{R}^d), \tag{3.1}$$

where $W = (W_t)_{t\in[0,T]}$ is an m-dimensional standard Brownian motion and $\mathcal{L}_{X_t}$ denotes the marginal law of the process X at time $t \in [0, T]$. We omit an explicit dependence of the coefficients on t for brevity, but our results easily generalise to this case.

The existence and uniqueness theory for strong solutions of McKean–Vlasov SDEs with coefficients of linear growth and Lipschitz type conditions (with respect to the state and the measure) is well-established (see, e.g., [144]). For further existence and uniqueness results on weak and strong solutions of McKean–Vlasov SDEs, we refer the reader to [17, 75, 118] and the references cited therein. Also, in the case of super-linear growth it is known that a McKean–Vlasov SDE admits a unique strong solution [131], assuming a so-called one-sided Lipschitz condition for the drift term, see item (1) of Assumption $(\mathbf{A}_b^1)$ in Section 3.2.1.

McKean–Vlasov SDEs have numerous applications, for instance, in the social and natural sciences. These include fundamental models in neuroscience, such as the Hodgkin-Huxley model (see [11]) for neuron activation, or in biology and chemistry, such as the Patlak–Keller–Segel equations describing, e.g., chemotactic interactions (see [90]) and long-chain polymers (see [125]). The afore-

mentioned examples all have drift terms, which do not exhibit the classical global Lipschitz conditions on the coefficients of the SDE.

The simulation of McKean–Vlasov SDEs typically involves two steps: First, at each time t, the true measure $\mathcal{L}_{X_t}$ is approximated by the empirical measure

$$\mu_t^{\boldsymbol{X}^N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x),$$

where δ_x denotes the Dirac measure at point x and $(\boldsymbol{X}_t^N)_{t\in[0,T]} = (X_t^{1,N},\dots,X_t^{N,N})_{t\in[0,T]}^\top$ (so-called interacting particles) is the solution to the $\mathbb{R}^{dN}$-dimensional SDE

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

Here, W^i and ξ^i, for $i \in \{1,\dots,N\}$, are independent Brownian motions (also independent of W) and independent copies of ξ, respectively. In the second step, one needs to introduce a reasonable time-stepping method to discretise the particle system $(X^{i,N})_{i\in\{1,\dots,N\}}$ over some finite time horizon $[0,T]$.

An Euler scheme for the particle system is introduced in [26], and the strong convergence to the solution of the McKean–Vlasov equation, of order 1/2 in the time-step and also 1/2 in the number of particles, is shown under global Lipschitz assumptions for the coefficients with respect to the state and measure (for $d = m = 1$). The convergence in N is often referred to as propagation of chaos. An approximation of order 1/2 in the time-step for the density and cumulative distribution function is also given in [26], and this is improved to order 1 in [9].

The strong convergence result in [26] has recently been extended to the case where only a one-sided Lipschitz condition for the drift holds with respect to the state, while a global Lipschitz condition is still assumed for all other dependencies, by using so-called tamed and implicit schemes (in [130]) or adaptive schemes (in [132]).

In this chapter, we are concerned with the strong convergence of the approximations to the process itself. We complement the work in [130] by introducing stable first order time-stepping schemes (Milstein schemes) for a particle system associated with McKean–Vlasov SDEs, allowing for a drift coefficient which grows super-linearly in the state component. Here, we will prove moment stability of the time-discretised particle system and strong convergence of order 1.

Our new Milstein scheme reveals that a term involving the Lions derivative (abbreviated by L-derivative) of the diffusion coefficient with respect to the (empirical) measure is necessary to obtain the strong convergence result. The L-derivative of functions on $\mathcal{P}_2(\mathbb{R}^d)$ was introduced by P.-L. Lions in his lectures [29] at Collège de France. This is of theoretical interest and demonstrates the inherent difference of SDEs with mean-field interaction to classical SDEs with regard to higher order time-stepping schemes.

A main difficulty presented by the super-linearity (apart from the strong convergence analysis) is to prove the stability of the proposed scheme, and we adopt here the taming approach given in [81, 149] for standard SDEs. In the present context, we require control for the size of each particle, which are challenging to obtain as the particles interact through the appearance of the empirical distribution of the particle system in the coefficients.

The convergence results for the time-stepping scheme hold for a fixed dimension N, and are robust in N. To obtain error bounds with respect to the solution of (3.1), these have to be supplemented by propagation of chaos results from [130] to bound the error in N from the particle approximation. In practice, many McKean–Vlasov equations are motivated by large but finite-dimensional particle systems, and our time-stepping schemes are directly applicable in that case.

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

On a side note, in the special case $N = 1$ our work gives the first order convergence of tamed Milstein schemes for standard SDEs with super-linear drift. It is observed in [82] that already in this setting of classical SDEs (i.e., where the coefficients have no measure-dependence), the explicit Euler–Maruyama scheme (see, e.g., [95]) is not appropriate in the presence of drift terms with super-linear growth. To overcome this problem, several stable time-discretisation methods, including a tamed explicit Euler and Milstein scheme [81, 140, 68, 149], an explicit adaptive Euler–Maruyama method [59], a truncated Euler method [113] and an implicit Euler scheme [77], have been introduced. In [68], a tamed Milstein scheme is introduced and convergence is proven under a commutativity assumption for the diffusion matrix, such that the Lévy area vanishes. Our technique of proving moment stability allows the analysis of the general case without this assumption. While [98] introduces a tamed Milstein scheme for SDEs with super-linearly growing drift and diffusion without assuming a commutativity assumption, the taming there is required to be stronger than the taming approach proposed in [81], which makes the stability analysis of the schemes easier, but can result in inferior numerical performance (see Remark 3.2.4 and Section 3.4 for details).

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

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

- detailed numerical tests supporting the theoretical finding;
- the main results extend those on tamed Milstein schemes (for certain taming approaches) for standard SDEs by eliminating commutativity conditions.

A tamed Milstein scheme, with stronger taming than used in this chapter, was developed under a similar set of assumptions independently and in parallel in [96]. The authors only state a pointwise (in time) strong L_2 convergence rate of order 1, albeit under slightly weaker differentiability conditions. The present work also goes beyond [96] by analysing the asymptotic behaviour (in terms of the number of particles) of the L-derivative terms in the schemes and by providing detailed numerical illustrations.

The remainder of this chapter is organised as follows: In Section 3.2, we formulate the problem set-up and introduce two different tamed Milstein schemes, in which we use two different taming factors for the drift. The proofs of the main convergence results are deferred to Section 3.3. Section 3.4 illustrates the numerical performance of the proposed time-stepping schemes and reveals that the schemes can also successfully be applied to equations, which do not satisfy all imposed model assumptions. We end this section by fixing the notation and by introducing several notions needed throughout this chapter.

Preliminaries:

Throughout this chapter, let $(\mathbb{R}^d, \langle\cdot,\cdot\rangle, |\cdot|)$ represent the d-dimensional Euclidean space and $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, for a given terminal time $T > 0$, will be a complete filtered atomless probability space satisfying the usual assumptions.

In addition, we use $\mathcal{P}(\mathbb{R}^d)$ to denote the family of all probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-field over $\mathbb{R}^d$, and denote the subset of probability measures with finite second moment by $\mathcal{P}_2(\mathbb{R}^d)$. As metric on the space $\mathcal{P}_2(\mathbb{R}^d)$, we use the Wasserstein distance. For all linear operators appearing in this chapter, we will use the standard Hilbert-Schmidt norm denoted by $\|\cdot\|$.

We briefly introduce the Lions derivative of a functional $f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, as it will appear in the formulation of the tamed Milstein scheme. For more details about the definition and further results, we refer to [29] or [27, 75]. Here, we follow the exposition of [31]. We recall the fact that over an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, there is a random variable $X \in L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ such that $\mathcal{L}_X = \mu$, see e.g., [31, Section 5.2]. We will then associate to the function f a lifted function $\tilde{f}$, defined by $\tilde{f}(X) = f(\mathcal{L}_X)$, for $X \in L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$, which allows to introduce the L-derivative as Fréchet derivative on the lifted space.

A function f on $\mathcal{P}_2(\mathbb{R}^d)$ is said to be L-differentiable at $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists a random variable $X_0 \in L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ with law μ_0, such that the lifted function $\tilde{f}$ is Fréchet differentiable

at X_0, i.e., the Riesz representation theorem implies that there is a ($\mathbb{P}$-almost surely) unique $\Theta \in L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ with

$$\tilde{f}(X) = \tilde{f}(X_0) + \langle \Theta, X - X_0 \rangle_{L_2} + o(\|X - X_0\|_{L_2}), \text{ as } \|X - X_0\|_{L_2} \to 0,$$

with the standard inner product and norm on $L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ and $X \in L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$. If f is L-differentiable for all $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, then we say that f is L-differentiable.

It is known (see e.g., [31, Proposition 5.25]), that there exists a Borel measurable function $\xi : \mathbb{R}^d \to \mathbb{R}^d$, such that $\Theta = \xi(X_0)$ almost surely, and hence

$$f(\mathcal{L}_X) = f(\mathcal{L}_{X_0}) + \mathbb{E}\,\langle \xi(X_0), X - X_0 \rangle + o(\|X - X_0\|_{L_2}).$$

Note that ξ only depends on the law of X_0, but not on X_0 itself (or the choice of the underlying probability space). We define $\partial_\mu f(\mathcal{L}_{X_0})(y) := \xi(y)$, $y \in \mathbb{R}^d$, as the L-derivative of f at μ_0. Observe that $\partial_\mu f(\mathcal{L}_{X_0})(y)$ is only $\mathcal{L}_{X_0}(\mathrm{d}y)$ almost everywhere uniquely defined. Therefore, below we will assume that for all $X \in L_2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$, there exists a $\mathcal{L}_X$-version of $\partial_\mu f(\mathcal{L}_X)(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$, which is, for instance, guaranteed if the Fréchet derivative of the lifted function $\tilde{f}$ is Lipschitz continuous. Further, assume for each $y \in \mathbb{R}^d$ that the map $\mathcal{P}_2(\mathbb{R}^d) \ni \mu \mapsto \partial_\mu f(\mu)(y)$ is continuously L-differentiable. Hence, the L-derivative of the components $(\partial_\mu f)_j(\cdot)(y) : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, $j \in \{1, \ldots, d\}$, is defined as

$$\partial_\mu^2 f(\mu)(y, y') := (\partial_\mu((\partial_\mu f)_j(\cdot)(y))(\mu, y'))_{j \in \{1,\ldots,d\}},$$

for $(\mu, y, y') \in \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d$. For a vector-valued (or matrix-valued) function f, these definitions have to be understood componentwise.

For the strong convergence analysis, we employ a definition describing regularity properties of a function $f : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ in terms of the measure derivative, see [31, 41].

Definition 3.1.1. Let $f : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be a given functional.

- We say that $f \in C^{2,(1,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$, if for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $f(\cdot, \mu)$ is twice continuously differentiable and for any $x \in \mathbb{R}^d$, $f(x, \cdot)$ is partially $\mathcal{C}_2$, i.e., for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, there is a continuous version of the map $\mathbb{R}^d \ni y \mapsto \partial_\mu f(x, \mu)(y)$ such that the derivatives

  $$\partial_\mu f(x, \mu)(y), \quad \partial_y\{\partial_\mu f(x, \mu)(\cdot)\}(y),$$

  exist, are locally bounded and jointly continuous in the corresponding variables (x, μ, y), such that $y \in \mathrm{Supp}(\mu)$.

- We say that $f \in C^{2,(2,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$, if $f \in C^{2,(1,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ and in addition the second order L-derivative $\partial_\mu^2 f(x, \mu)(y, y')$ exists and is again jointly continuous in the corresponding variables. Note the joint continuity of all derivatives is required globally, i.e., for all (x, μ, y, y'). In this case, $f(x, \cdot)$ is called fully $\mathcal{C}_2$.

For vector-valued or matrix-valued functions, this definition has to be understood componentwise.

3.2 Tamed Milstein schemes for non-Lipschitz McKean–Vlasov SDEs

In this section, we define the model set-up with its assumptions (Subsection 3.2.1), and define the Milstein schemes (Subsection 3.2.2). We focus in the analysis on the schemes with tamed drift coefficients, where the super-linear drifts are approximated by functions which are bounded depending on the mesh-size (similar to the tamed Euler schemes in [130]). A simplified version of the proofs gives the corresponding results for the Milstein scheme without drift approximation in the global Lipschitz case, and we only state the assumptions required in Subsection 3.2.3.

3.2.1 Assumptions and interacting particle system

For a given time horizon $[0,T]$ with terminal time $T > 0$, we consider the following McKean–Vlasov SDE

$$\mathrm{d}X_t = b\left(X_t, \mathcal{L}_{X_t}\right) \mathrm{d}t + \sigma(X_t, \mathcal{L}_{X_t}) \, \mathrm{d}W_t, \quad X_0 = \xi, \tag{3.2}$$

where we recall that $\mathcal{L}_{X_t}$ denotes the marginal law of X at the time $t \in [0,T]$, $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$ are measurable functions, $\xi \in L_0^p(\mathbb{R}^d)$, for all $p \geq 2$, and $W = (W_t)_{t\in[0,T]}$ is an m-dimensional Brownian motion on the complete filtered atomless probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, where $(\mathcal{F}_t)_{t\in[0,T]}$ is the natural filtration of W augmented with an independent σ-algebra $\mathcal{F}_0$.

We make the following assumptions:

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

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

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

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

for all $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

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

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

for all $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

The analysis presented below can be readily extended to the case where b and σ depend explicitly on t in a Lipschitz continuous way. Further, the j-th column of σ is denoted by $\sigma_j = (\sigma_{1j}, \ldots, \sigma_{dj})^\top$, i.e., $\sigma = (\sigma_1, \ldots, \sigma_m)$.

We recall that under these assumptions [131, Theorem 3.3] guarantees that (3.2) has a unique strong solution with bounded p-th moments, i.e., we have

$$\mathbb{E}\left[\sup_{t\in[0,T]}|X_t|^p\right]<\infty.$$

Let now (W^i,ξ^i) be independent copies of (W,ξ) for $i\in\{1,\dots,N\}$. Consider first the following *non-interacting* particle system associated with (3.2),

$$\mathrm{d}X_t^i=b(X_t^i,\mathcal{L}_{X_t^i})\,\mathrm{d}t+\sigma(X_t^i,\mathcal{L}_{X_t^i})\,\mathrm{d}W_t^i,\quad X_0^i=\xi^i. \tag{3.3}$$

Compared to the simulation of classical SDEs, the key difference is the need to approximate the measure $\mathcal{L}_{X_t}$, for each $t\in[0,T]$. To do so, we introduce the following *interacting* particle system (see e.g., [26])

$$\mathrm{d}X_t^{i,N}=b(X_t^{i,N},\mu_t^{\boldsymbol{X}^N})\,\mathrm{d}t+\sigma(X_t^{i,N},\mu_t^{\boldsymbol{X}^N})\,\mathrm{d}W_t^i,\quad X_0^{i,N}=\xi^i, \tag{3.4}$$

where $\mu_t^{\boldsymbol{X}^N}(\mathrm{d}x):=\frac{1}{N}\sum_{j=1}^N\delta_{X_t^{j,N}}(\mathrm{d}x)$. Set

$$B(\boldsymbol{x}^N):=(b(x_1,\mu^{\boldsymbol{x}^N}),\dots,b(x_N,\mu^{\boldsymbol{x}^N}))^\top,\qquad \Sigma(\boldsymbol{x}^N):=\mathrm{diag}(\sigma(x_1,\mu^{\boldsymbol{x}^N}),\dots,\sigma(x_N,\mu^{\boldsymbol{x}^N})),$$

where $\mu^{\boldsymbol{x}^N}(\mathrm{d}x):=\frac{1}{N}\sum_{j=1}^N\delta_{x_j}(\mathrm{d}x)$, for $\boldsymbol{x}^N:=(x_1,x_2,\dots,x_N)$. Note that

$$\mathcal{W}_2^2(\mu^{\boldsymbol{x}^N},\mu^{\boldsymbol{y}^N})\le\frac{1}{N}|\boldsymbol{x}^N-\boldsymbol{y}^N|^2,\qquad \boldsymbol{x}^N,\boldsymbol{y}^N\in\mathbb{R}^{dN}, \tag{3.5}$$

and hence, we deduce from $(\mathbf{A}_b^1)$ and $(\mathbf{A}_\sigma^1)$ that there exists a constant $L>0$ such that

$$\langle\boldsymbol{x}^N-\boldsymbol{y}^N,B(\boldsymbol{x}^N)-B(\boldsymbol{y}^N)\rangle\le L|\boldsymbol{x}^N-\boldsymbol{y}^N|^2,\qquad \|\Sigma(\boldsymbol{x}^N)-\Sigma(\boldsymbol{y}^N)\|^2\le L|\boldsymbol{x}^N-\boldsymbol{y}^N|^2,$$

for any $\boldsymbol{x}^N,\boldsymbol{y}^N\in\mathbb{R}^{dN}$. Consequently, according to, e.g., [129, Theorem 3.1.1], the interacting particle system (3.4) is strongly well-posed.

3.2.2 Time-stepping schemes and main results

Since the drift term b is not assumed to be Lipschitz with respect to the spatial argument, the standard Euler scheme is, in general, not suitable for approximating (3.4) in time, due to the potential moment-unboundedness of the time-discretised interacting particle system [82]. Here, we propose two novel stable time-stepping schemes which achieve a first order strong convergence rate.

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

$$\begin{aligned}Y_{t_{n+1}}^{i,N}=Y_{t_n}^{i,N}&+b_h(Y_{t_n}^{i,N},\mu_{t_n}^{\boldsymbol{Y}^N})\,h+\sigma(Y_{t_n}^{i,N},\mu_{t_n}^{\boldsymbol{Y}^N})\Delta W_n^i\\&+\int_{t_n}^{t_{n+1}}\partial_x\sigma(Y_{t_n}^{i,N},\mu_{t_n}^{\boldsymbol{Y}^N})\sigma(Y_{t_n}^{i,N},\mu_{t_n}^{\boldsymbol{Y}^N})\int_{t_n}^s\mathrm{d}W_u^i\mathrm{d}W_s^i\end{aligned}$$

$$+\int_{t_n}^{t_{n+1}} \frac{1}{N}\sum_{j=1}^{N} \partial_\mu \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})(Y_{t_n}^{j,N})\sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{\boldsymbol{Y}^N}) \int_{t_n}^{s} \mathrm{d}W_u^j \mathrm{d}W_s^i, \tag{3.6}$$

where the driving Brownian motions W^i and initial values $Y_0^{i,N} = \xi^i$ are independent copies of W and ξ, respectively, and $\Delta W_n^i := W_{t_{n+1}}^i - W_{t_n}^i$. In addition, we used the notation

$$\mu_{t_n}^{\boldsymbol{Y}^N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{Y_{t_n}^{j,N}}(\mathrm{d}x).$$

Note that the terms

$$\partial_x \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})\sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N}) \int_{t_n}^{s} \mathrm{d}W_u^i, \quad \partial_\mu \sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})(Y_{t_n}^{j,N})\sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{\boldsymbol{Y}^N}) \int_{t_n}^{s} \mathrm{d}W_u^j,$$

are understood as elements of $\mathbb{R}^{d\times m}$ with columns

$$\partial_x \sigma_l(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})\sigma(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N}) \int_{t_n}^{s} \mathrm{d}W_u^i, \quad \partial_\mu \sigma_l(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})(Y_{t_n}^{j,N})\sigma(Y_{t_n}^{j,N}, \mu_{t_n}^{\boldsymbol{Y}^N}) \int_{t_n}^{s} \mathrm{d}W_u^j,$$

respectively, for $l \in \{1, \dots, m\}$. Above, $\partial_x \sigma_l$ denotes the first order gradient operator (applied to the l-th column of σ) with respect to the state variable of σ_l and $\partial_\mu \sigma_l$ is the Lions derivative operator.

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

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

and for Scheme 2, we define

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

Note that the following bounds hold for the two different choices of b_h:

$$\frac{|b(x,\mu)|}{1 + h|b(x,\mu)|} \le |b(x,\mu)| \wedge \frac{M}{T}, \qquad \frac{|b(x,\mu)|}{1 + h|b(x,\mu)|^2} \le |b(x,\mu)| \wedge \sqrt{\frac{M}{T}}. \tag{3.7}$$

Remark 3.2.1. If the drift b is also globally Lipschitz in the state component, taming is not necessary and we can instead introduce a standard Milstein scheme by simply replacing b_h in (3.6) with b; see Subsection 3.2.3 for details of this case.

The continuous time version of (3.6) reads

$$\begin{aligned}\mathrm{d}Y_t^{i,N} = b_h(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})\,\mathrm{d}t + \Bigg(\sigma(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N}) + \partial_x \sigma(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N}) \int_{\eta(t)}^{t} \sigma(Y_{\eta(r)}^{i,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N})\,\mathrm{d}W_r^i \\ + \frac{1}{N}\sum_{j=1}^{N} \partial_\mu \sigma(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})(Y_{\eta(t)}^{j,N}) \int_{\eta(t)}^{t} \sigma(Y_{\eta(r)}^{j,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N})\,\mathrm{d}W_r^j \Bigg)\,\mathrm{d}W_t^i,\end{aligned} \tag{3.8}$$

with $Y_0^{i,N} = \xi^i$.

Remark 3.2.2. Compared to the classical Milstein scheme for standard SDEs without measure dependence, a term involving the L-derivative appears. Although this term is crucial for the theoretical analysis of the scheme for fixed N, it can be dropped in practice for large N and when an approximation to the limiting McKean–Vlasov equation is sought. We will provide a theoretical justification in the Lipschitz case at the end of Subsection 3.2.3, and give a numerical illustration in Section 3.4.

In what follows, we show that the fully discretised particle system converges, in a strong sense, to the solution of the limit McKean–Vlasov SDE if $N \to \infty$ and $h \to 0$ and establish the order of convergence. To establish the following main result on moment stability and strong convergence of the above proposed time-stepping schemes, we need further assumptions on the coefficients b and σ.

We require the following:

($\mathbf{A}_b^2$) Set $b = (b_1, \dots, b_d)^\top$ and let $b_i \in C^{2,(2,1)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$. There exist constants $L_b^2, \alpha_2 > 0$ such that for all $i \in \{1, \dots, d\}$

$$\|\partial_x^2 b_i(x,\mu)\| \vee \|\partial_y\{\partial_\mu b_i(x,\mu)(\cdot)\}(y)\| \vee \|\partial_\mu^2 b_i(x,\mu)(y,y)\| \\ \leq L_b^2(1 + |x|^{1+\alpha_2} + |y|^{1+\alpha_2} + \mu(|\cdot|^2)^{1/2}),$$

for all $x, y \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, where ∂_x^2 is the second order gradient operator with respect to the first argument and ∂_μ^2 the second order L-derivative operator.

($\mathbf{A}_b^3$) There exist constants $L_b^3, \alpha_3 > 0$ such that

$$\|\partial_x b(x,\mu) - \partial_x b(y,\nu)\| \leq L_b^3(|x-y|(1 + |x|^{\alpha_3} + |y|^{\alpha_3}) + \mathcal{W}_2(\mu,\nu))$$

$$\|\partial_\mu b(x,\mu)(y) - \partial_\mu b(x',\nu)(y')\| \leq L_b^3(|x-x'|(1 + |x|^{\alpha_3} + |x'|^{\alpha_3}) + |y-y'| + \mathcal{W}_2(\mu,\nu))$$

for all $x, x', y, y' \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

($\mathbf{A}_b^4$) There exists a constant $L_b^4 > 0$ such that $|b(0,\mu)| \leq L_b^4$, for all $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Concerning the diffusion coefficient σ, we further impose:

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

$$|\partial_\mu \sigma_{ij}(x,\mu)(y)| \vee \|\partial_y\{\partial_\mu \sigma_{ij}(x,\mu)(\cdot)\}(y)\| \vee \|\partial_\mu^2 \sigma_{ij}(x,\mu)(y,y)\| \leq L_\sigma^2,$$

for all $x, y \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

($\mathbf{A}_\sigma^3$) There exists a constant $L_\sigma^3 > 0$ such that for any $l \in \{1, \dots, m\}$

$$\|\partial_x \sigma_l(x,\mu) - \partial_x \sigma_l(y,\nu)\| \leq L_\sigma^3(|x-y| + \mathcal{W}_2(\mu,\nu)), \tag{1}$$

$$\|\partial_\mu \sigma_l(x,\mu)(y) - \partial_\mu \sigma_l(x',\nu)(y')\| \leq L_\sigma^3(|x-x'| + |y-y'| + \mathcal{W}_2(\mu,\nu)), \tag{2}$$

for all $x, x', y, y' \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

$(\mathbf{A}_\sigma^4)$ There exists a constant $L_\sigma^4 > 0$ such that for any $l \in \{1, \dots, m\}$

$$\|\sigma(0,\mu)\| \vee \|\partial_\mu \sigma_l(x,\mu)(y)\sigma(y,\mu)\| \leq L_\sigma^4,$$

for all $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

Remark 3.2.3. The first inequality in $(\mathbf{A}_b^1)$ is the so called one-sided Lipschitz condition which is needed to control the polynomial growth of the drift in the state variable (uniformly with respect to the measure variable). The third inequality in $(\mathbf{A}_b^1)$ and Assumption $(\mathbf{A}_\sigma^1)$ express that both coefficients are globally Lipschitz continuous in the measure component (uniformly in the state variable). In $(\mathbf{A}_b^2)$-$(\mathbf{A}_b^3)$ and $(\mathbf{A}_\sigma^2)$-$(\mathbf{A}_\sigma^3)$, we require regularity conditions on the coefficients b and σ and growth and Lipschitz assumptions on certain derivatives, which are necessary for the strong convergence analysis. Note that the growth conditions in $(\mathbf{A}_b^2)$ and $(\mathbf{A}_\sigma^2)$ in fact follow from these Lipschitz assumptions, and are only state for better readability. Further, we assume uniform boundedness of the coefficients in the measure component in $(\mathbf{A}_b^4)$ and $(\mathbf{A}_\sigma^4)$, which is essential to achieve moment boundedness of Scheme 1.

Now, we are in a position to present our first main result (concerned with Scheme 1) and we remark that in Section 3.3, we give more details on the generic constants used in the statement of the results.

Lemma 3.2.1. *Let $Y_{t_n}^{i,N}$, $n \in \{0, \dots, M\}$, be defined as in (3.6) with $b_h(x,\mu) = \frac{b(x,\mu)}{1+h|b(x,\mu)|}$ and $\xi \in L_0^p(\mathbb{R}^d)$ for any $p \geq 2$. Then, under Assumptions $(\mathbf{A}_b^1)$, $(\mathbf{A}_b^4)$, $(\mathbf{A}_\sigma^1)$, and $(\mathbf{A}_\sigma^4)$, there exists a constant $C > 0$ independent of M (h, respectively) and N such that*

$$\max_{i\in\{1,\dots,N\}} \max_{n\in\{0,\dots,M\}} \mathbb{E}\left[|Y_{t_n}^{i,N}|^p\right] \leq C.$$

Proof. The proof is deferred to Section 3.3, Subsection 3.3.1. □

Theorem 3.2.1. *Let $\xi \in L_0^p(\mathbb{R}^d)$ for any $p \geq 2$. Let Assumptions $(\mathbf{A}_b^1)$-$(\mathbf{A}_b^4)$, and $(\mathbf{A}_\sigma^1)$-$(\mathbf{A}_\sigma^4)$ be satisfied. Let $(Y_t^{i,N})_{t\in[0,T]}$ be defined by (3.8) with $b_h(x,\mu) = \frac{b(x,\mu)}{1+h|b(x,\mu)|}$. Then, there exists a constant $C > 0$ independent of M (h, respectively) and N such that*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{i,N} - Y_t^{i,N}|^p\right] \leq Ch^p. \tag{3.9}$$

Proof. The proof is deferred to Section 3.3. □

The following corollary is an immediate consequence of the above theorem and the pathwise propagation of chaos result in [130, Proposition 3.1].

Corollary 3.2.1. *Let the assumptions of Theorem 3.2.1 hold. Then, there exists a constant $C > 0$ independent of M (h, respectively) and N such that*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - Y_t^{i,N}|^2\right] \leq C(h^2 + \phi(N)),$$

where

$$\phi(N) = \begin{cases} N^{-1/2}, & \text{for } d < 4, \\ N^{-1/2} \log N, & \text{for } d = 4, \\ N^{-2/d}, & \text{for } d > 4. \end{cases} \tag{3.10}$$

Proof. First note that we have

$$|X_t^i - Y_t^{i,N}|^2 \leq 2(|X_t^i - X_t^{i,N}|^2 + |X_t^{i,N} - Y_t^{i,N}|^2).$$

The squared $\mathcal{S}^2$-norm of the second summand is of order h^2 due to Theorem 3.2.1, and from [130, Proposition 3.1], there exists a constant $C > 0$ (independent of N) such that

$$\max_{i \in \{1,\dots,N\}} \mathbb{E}\left[\sup_{t \in [0,T]} |X_t^i - X_t^{i,N}|^2\right] \leq C\phi(N), \tag{3.11}$$

from which the claim follows. □

Remark 3.2.4. Let $(Y_t^{i,N})_{t\in[0,T]}$ be defined by Scheme 2, i.e., (3.8) with $b_h(x,\mu) = b(x,\mu)/(1 + h|b(x,\mu)|^2)$. Assume, for some given $p \geq 2$, $\xi \in L^0_{4p(1+\alpha)}(\mathbb{R}^d)$, where $\alpha = \alpha_1 \vee \alpha_2 \vee \alpha_3$ and let Assumptions $(\mathbf{A}_b^1)$–$(\mathbf{A}_b^3)$, $(\mathbf{A}_\sigma^1)$–$(\mathbf{A}_\sigma^3)$ be satisfied. Then, the statement of Theorem 3.2.1 still holds.

Due to the choice of taming, we have the stronger bound in (3.7) and the moment boundedness of Scheme 2 follows immediately from [97, Lemma 4.3] (see also Remark 3.2.4 below). The proof of the strong convergence rate is then analogous to Scheme 1 and is therefore omitted.

The crucial difference between these two results is that for Scheme 2 we do not need to require $(\mathbf{A}_b^4)$ and $(\mathbf{A}_\sigma^4)$, i.e., that the coefficients are uniformly bounded in the measure component. The reason is that the taming for Scheme 2 is stronger than for Scheme 1. The numerical tests in Section 3.4 show, however, that Scheme 1 is significantly more accurate in all cases studied, even when $(\mathbf{A}_b^4)$ and $(\mathbf{A}_\sigma^4)$ are violated.

Remark 3.2.5. To prove moment boundedness for Scheme 1, an estimate of the time-discretised particle system on suitable chosen sub-events is required. In order to obtain such an estimate the mean-field terms need to be controlled. In [68], a commutativity assumption for the diffusion matrix is imposed, such that the Lévy areas vanish. The techniques used to prove our result can also be employed to relax the assumptions imposed in [68].

3.2.3 Milstein scheme for globally Lipschitz drift

For the sake of completeness, we give a set of model assumptions which allow to derive an analogous convergence result for a standard Milstein scheme, i.e., without taming the drift. We impose Assumptions $(\mathbf{A}_\sigma^1)$–$(\mathbf{A}_\sigma^3)$ for the diffusion coefficient, and we impose:

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

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

for all $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

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

$$\|\partial_y\{\partial_\mu b_i(x,\mu)(\cdot)\}(y)\| \vee \|\partial_\mu^2 b_i(x,\mu)(y,y)\| \le L_b^2,$$

for all $x, y \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

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

$$\|\partial_x b(x,\mu) - \partial_x b(y,\nu)\| \le L_b^3(|x-y| + \mathcal{W}_2(\mu,\nu)), \tag{1}$$

$$\|\partial_\mu b(x,\mu)(y) - \partial_\mu b(x',\nu)(y')\| \le L_b^3(|x-x'| + |y-y'| + \mathcal{W}_2(\mu,\nu)), \tag{2}$$

for all $x, x', y, y' \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

Under the assumptions listed above, the standard Milstein scheme defined as in (3.6), with b_h replaced by b, is stable (i.e., has bounded moments) and converges with strong order 1. As b and σ have linear growth in both components and the particles are identically distributed, the claim concerning the stability of the scheme follows by a standard Gronwall type argument. The proof for the strong convergence order is a simplified version of the proof of Theorem 3.2.1 and is therefore omitted.

We end by estimating the term involving the L-derivative in (3.6),

$$\int_{t_n}^{t_{n+1}} \frac{1}{N} \sum_{j=1}^{N} \partial_\mu \sigma(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})(Y_{\eta(t)}^{j,N}) \int_{\eta(t)}^{t} \sigma(Y_{\eta(r)}^{j,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N}) \, \mathrm{d}W_r^j \, \mathrm{d}W_t^i, \tag{3.12}$$

which is expected to be close to zero (see below for details) for large N, as explained by the following heuristic observation.

Instead of discretising the particle system, we can directly discretise (3.2) in time using a Milstein scheme without employing the particle approximation for the measure. Itô's formula applied to the components σ_{ij} of the diffusion coefficient, gives for $i \in \{1, \dots, d\}$, $j \in \{1, \dots, m\}$

$$\begin{aligned}
\mathrm{d}\sigma_{ij}(X_s, \mathcal{L}_{X_s}) = \Bigg[& \frac{1}{2}\mathrm{tr}(\sigma\sigma^\top \partial_x^2 \sigma_{ij})(X_s, \mathcal{L}_{X_s}) + \langle b, \partial_x \sigma_{ij} \rangle (X_s, \mathcal{L}_{X_s}) \\
& + \int_{\mathbb{R}^d} \Big[\frac{1}{2}\mathrm{tr}\Big\{ \sigma\sigma^\top(y, \mathcal{L}_{X_s}) \partial_y \{\partial_\mu \sigma_{ij}(X_s, \mathcal{L}_{X_s})(\cdot)\}(y) \Big\} \\
& + \langle b(y, \mathcal{L}_{X_s}), \partial_\mu \sigma_{ij}(X_s, \mathcal{L}_{X_s})(y) \rangle \Big] \mathcal{L}_{X_s}(\mathrm{d}y) \Bigg] \mathrm{d}s
\end{aligned}$$

$$+ \left\langle (\sigma^\top \partial_x \sigma_{ij})(X_s, \mathcal{L}_{X_s}), \mathrm{d}W_s \right\rangle,$$

where tr denotes the standard trace operator (see [31, Proposition 5.102]). Note that such an Itô formula can be applied to any partially $\mathcal{C}_2$ functional, whose L-derivative and the classical derivative of this L-derivative satisfy certain integrability conditions. Using this expansion, we can introduce a Milstein scheme for (3.2)

$$Y_{t_{n+1}} = Y_{t_n} + b(Y_{t_n}, \mathcal{L}_{Y_{t_n}})\, h + \sigma(Y_{t_n}, \mathcal{L}_{Y_{t_n}})\Delta W_n + \int_{t_n}^{t_{n+1}} \partial_x \sigma(Y_{t_n}, \mathcal{L}_{Y_{t_n}})\sigma(Y_{t_n}, \mathcal{L}_{Y_{t_n}}) \int_{t_n}^{s} \mathrm{d}W_u \mathrm{d}W_s, \tag{3.13}$$

for $n \in \{0, \dots, M-1\}$, with $Y_0 = \xi$. Therefore, it is reasonable to expect that the term (3.12) vanishes, in some sense, as $N \to \infty$, since (3.13) and (3.6) (with b_h replaced by b, as we restrict this discussion to a global Lipschitz setting) should coincide for $N \to \infty$, i.e., a propagation of chaos result on the level of the time-discrete system.

More precisely, we give the following proposition. We only prove this statement for globally Lipschitz coefficients, but expect a similar result to hold in the general setting of this chapter. To prove the claim, we additionally require the following:

($\mathbf{AA}^1_\sigma$) Let σ be continuously differentiable in both components and there exists a constant $L^1_\sigma > 0$ such that for any $l \in \{1, \dots, m\}$

$$\|\partial_x \sigma_l(x,\mu)\sigma(x,\mu) - \partial_x \sigma_l(x',\nu)\sigma(x',\nu)\| \le L^1_\sigma(|x-x'| + \mathcal{W}_2(\mu,\nu)), \tag{1}$$

$$\|\partial_\mu \sigma_l(x,\mu)(y)\sigma(y,\mu) - \partial_\mu \sigma_l(x',\nu)(y')\sigma(y',\nu)\| \le L^1_\sigma(|x-x'| + |y-y'| + \mathcal{W}_2(\mu,\nu)), \tag{2}$$

for all $x, x', y, y' \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

Proposition 3.2.1. Let Assumptions ($\mathbf{AA}^1_b$), ($\mathbf{A}^1_\sigma$) and ($\mathbf{AA}^1_\sigma$) be satisfied and let $\xi \in L^0_p(\mathbb{R}^d)$ for any $p \ge 2$. Further, for $i \in \{1, \dots, N\}$ and $n \in \{0, \dots, M\}$, let $Y^i_{t_n}$ be given by (3.13), with independent copies (W^i, ξ^i) of (W, ξ), and $Y^{i,N}_{t_n}$ be defined by (3.6), with b_h replaced by b. Then, there exists a constant $C > 0$ (independent of N) such that

$$\max_{i\in\{1,\dots,N\}} \max_{n\in\{0,\dots,M\}} \mathbb{E}\left[|Y^{i,N}_{t_n} - Y^i_{t_n}|^2\right] \le C\phi(N).$$

Proof. See Subsection 3.3.4. □

3.3 Proof of results

Here, and throughout the remaining chapter, we write $a \lesssim b$ to express that there exists a constant $C > 0$ such that $a \le Cb$, where $a, b \in \mathbb{R}$. The implied constant $C > 0$ may depend on the parameters p, ε, T, m, d, the constants appearing in above assumptions and the moments of the initial data, but is independent of M (and h, respectively) and N. Also implied constants may change their values from line to line in a sequence of inequalities. To make the presentation clearer,

we split the proof of Theorem 3.2.1 into several auxiliary lemmas. For the sake of readability, we also set, for any $i \in \{1, \dots, N\}$,

$$\begin{aligned}\Upsilon_t^i &:= \sigma(Y^{i,N}_{\eta(t)}, \mu^{\boldsymbol{Y}^N}_{\eta(t)}) + \int_{\eta(t)}^t \partial_x \sigma(Y^{i,N}_{\eta(r)}, \mu^{\boldsymbol{Y}^N}_{\eta(r)}) \sigma(Y^{i,N}_{\eta(r)}, \mu^{\boldsymbol{Y}^N}_{\eta(r)}) \, \mathrm{d}W_r^i \\ &\quad + \frac{1}{N} \sum_{j=1}^N \int_{\eta(t)}^t \partial_\mu \sigma(Y^{i,N}_{\eta(r)}, \mu^{\boldsymbol{Y}^N}_{\eta(r)})(Y^{j,N}_{\eta(r)}) \sigma(Y^{j,N}_{\eta(r)}, \mu^{\boldsymbol{Y}^N}_{\eta(r)}) \, \mathrm{d}W_r^j, \qquad (3.14) \\ \Gamma_t^i &:= \sigma(Y_t^{i,N}, \mu_t^{\boldsymbol{Y}^N}) - \Upsilon_t^i, \quad M_t^i := \int_0^t (\sigma(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N}) - \Upsilon_s^i) \, \mathrm{d}W_s^i, \quad Z^{i,N} := X^{i,N} - Y^{i,N}.\end{aligned}$$

3.3.1 Proof of Lemma 3.2.1

An alternative proof of Lemma 3.2.1 under the additional assumption that the diffusion is measure-independent can be found in Appendix A.1.

Proof. The proof follows the ideas developed in [149, 46] for tamed Euler–Maruyama schemes. Let $t \in [0, T]$ and $p \geq 2$ be given. By the application of Itô's formula

$$\begin{aligned}|Y_t^{i,N}|^p &= |Y^{i,N}_{\eta(t)}|^p + p \int_{\eta(t)}^t |Y_s^{i,N}|^{p-2} \left\langle Y_s^{i,N}, b_h(Y^{i,N}_{\eta(s)}, \mu^{\boldsymbol{Y}^N}_{\eta(s)}) \right\rangle \, \mathrm{d}s \\ &\quad + \frac{p}{2} \int_{\eta(s)}^t |Y_s^{i,N}|^{p-2} \|\Upsilon_s^i\|^2 \, \mathrm{d}s + \frac{p(p-2)}{2} \int_{\eta(t)}^t |Y_s^{i,N}|^{p-4} |(Y_s^{i,N})^\top \Upsilon_s^i|^2 \, \mathrm{d}s \\ &\quad + p \int_{\eta(t)}^t |Y_s^{i,N}|^{p-2} \left\langle Y_s^{i,N}, \Upsilon_s^i \, \mathrm{d}W_s^i \right\rangle,\end{aligned}$$

which gives the estimate

$$\begin{aligned}|Y_t^{i,N}|^p &\leq |Y^{i,N}_{\eta(t)}|^p + p \int_{\eta(t)}^t |Y_s^{i,N}|^{p-2} \left\langle Y_s^{i,N}, b_h(Y^{i,N}_{\eta(s)}, \mu^{\boldsymbol{Y}^N}_{\eta(s)}) \right\rangle \, \mathrm{d}s \\ &\quad + \frac{p(p-1)}{2} \int_{\eta(t)}^t |Y_s^{i,N}|^{p-2} \|\Upsilon_s^i\|^2 \, \mathrm{d}s + p \int_{\eta(t)}^t |Y_s^{i,N}|^{p-2} \left\langle Y_s^{i,N}, \Upsilon_s^i \, \mathrm{d}W_s^i \right\rangle.\end{aligned}$$

We introduce now, for $n \in \{0, \dots, M\}$, the sequence of sub-events

$$\Omega_{R,n} := \{\boldsymbol{\omega}_N \in \Omega^N : \max_{m \in \{0,\dots,n\}} |Y^{i,M}_{t_m}(\boldsymbol{\omega}_N)| \leq R\},$$

for some $R > 0$, which will be specified at a later step of the proof, $\Omega^N := \Omega \times \ldots \times \Omega$ (N times) and $\Omega^N \ni \boldsymbol{\omega}_N = (\omega_1, \dots, \omega_N)$. Then, we deduce from the previous estimate

$$\begin{aligned}\mathbb{E}[\mathrm{I}_{\Omega_{R,n}} |Y_t^{i,N}|^p] &\leq \mathbb{E}[\mathrm{I}_{\Omega_{R,n}} |Y^{i,N}_{\eta(t)}|^p] + p \int_{\eta(t)}^t \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} \left\langle Y_s^{i,N}, b_h(Y^{i,N}_{\eta(s)}, \mu^{\boldsymbol{Y}^N}_{\eta(s)}) \right\rangle\right] \mathrm{d}s \\ &\quad + \frac{p(p-1)}{2} \int_{\eta(t)}^t \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} \|\Upsilon_s^i\|^2\right] \mathrm{d}s.\end{aligned}$$

Writing

$$\mathbb{E}[\mathrm{I}_{\Omega_{R,n}} |Y_t^{i,N}|^p] \leq \mathbb{E}[\mathrm{I}_{\Omega_{R,n}} |Y^{i,N}_{\eta(t)}|^p] + p \int_{\eta(t)}^t \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y^{i,N}_{\eta(s)}|^{p-2} \left\langle Y^{i,N}_{\eta(s)}, b(Y^{i,N}_{\eta(s)}, \mu^{\boldsymbol{Y}^N}_{\eta(s)}) \right\rangle\right] \mathrm{d}s$$

$$
\begin{aligned}
&+ p \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} \left(|Y_s^{i,N}|^{p-2} - |Y_{\eta(s)}^{i,N}|^{p-2}\right) \left\langle Y_{\eta(s)}^{i,N}, b(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle\right] \mathrm{d}s \\
&+ p \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} \left\langle Y_s^{i,N} - Y_{\eta(s)}^{i,N}, b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle\right] \mathrm{d}s \\
&+ p \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} \left\langle Y_{\eta(s)}^{i,N}, b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N}) - b(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle\right] \mathrm{d}s \\
&+ \frac{p(p-1)}{2} \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} \|\Upsilon_s^i\|^2\right] \mathrm{d}s,
\end{aligned}
$$

and using Assumptions $(\mathbf{A}_b^1)$ and $(\mathbf{A}_b^4)$, we obtain

$$
\mathbb{E}[\mathrm{I}_{\Omega_{R,n}} |Y_t^{i,N}|^p] \leq Ch + (1+Ch)\mathbb{E}[\mathrm{I}_{\Omega_{R,n}} |Y_{\eta(t)}^{i,N}|^p] + \sum_{j=1}^{4} \Pi_t^{j,i}, \tag{3.15}
$$

where we introduced

$$
\begin{aligned}
\Pi_t^{1,i} &:= p \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} \left(|Y_s^{i,N}|^{p-2} - |Y_{\eta(s)}^{i,N}|^{p-2}\right) \left\langle Y_{\eta(s)}^{i,N}, b(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle\right] \mathrm{d}s, \\
\Pi_t^{2,i} &:= p \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} \left\langle Y_s^{i,N} - Y_{\eta(s)}^{i,N}, b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle\right] \mathrm{d}s, \\
\Pi_t^{3,i} &:= p \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} \left\langle Y_{\eta(s)}^{i,N}, b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N}) - b(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle\right] \mathrm{d}s, \\
\Pi_t^{4,i} &:= \frac{p(p-1)}{2} \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} \|\Upsilon_s^i\|^2\right] \mathrm{d}s.
\end{aligned}
$$

In what follows, we estimate each of these terms one-by-one. It is easy to show, using [46, Lemma 4.2] and (3.8), that

$$
\begin{aligned}
\Pi_t^{1,i} &\lesssim \sum_{l=1}^{p-2} h^{l+1} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |b(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})|^{l+1} |Y_{\eta(t)}^{i,N}|^{p-1-l}\right] \\
&\quad + \sum_{l=1}^{p-2} h^{l/2+1} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |b(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})| |Y_{\eta(t)}^{i,N}|^{p-1-l} \|\Upsilon_t^i\|^l\right] \\
&\lesssim h + \sum_{l=1}^{p-2} h^{l+1} \mathbb{E}[\mathrm{I}_{\Omega_{R,n}} |Y_{\eta(t)}^{i,N}|^{p+\alpha_1(l+1)}] + \sum_{l=1}^{p-2} h^{l/2+1} \mathbb{E}[\mathrm{I}_{\Omega_{R,n}} |Y_{\eta(t)}^{i,N}|^{p+\alpha_1}],
\end{aligned}
$$

where in the last estimate, we used $(\mathbf{A}_b^1)$, $(\mathbf{A}_b^4)$, $(\mathbf{A}_\sigma^1)$ and $(\mathbf{A}_\sigma^4)$. Employing the Cauchy-Schwarz inequality and recalling (3.8), we derive for $\Pi_t^{2,i}$ the estimate

$$
\begin{aligned}
\Pi_t^{2,i} &\leq p \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N}|^{p-2} |Y_s^{i,N} - Y_{\eta(s)}^{i,N}| |b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})|\right] \mathrm{d}s \\
&\lesssim \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_{\eta(s)}^{i,N}|^{p-2} |Y_s^{i,N} - Y_{\eta(s)}^{i,N}| |b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})|\right] \mathrm{d}s \\
&\quad + \int_{\eta(t)}^{t} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |Y_s^{i,N} - Y_{\eta(s)}^{i,N}|^{p-1} |b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{Y^{\boldsymbol{Y}^N}})|\right] \mathrm{d}s \\
&\lesssim h^2 \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |b_h(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})|^2 |Y_{\eta(t)}^{i,N}|^{p-2}\right] + h^{3/2} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |b_h(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})| \|\Upsilon_t^i\| |Y_{\eta(t)}^{i,N}|^{p-2}\right] \\
&\quad + h^p \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |b_h(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})|^p\right] + h^{\frac{p+1}{2}} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}} |b_h(Y_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\boldsymbol{Y}^N})| \|\Upsilon_t^i\|^{p-1}\right]
\end{aligned}
$$

$$\lesssim h + h^2\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^{p+2\alpha_1}] + h^{3/2}\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^{p+\alpha_1}] + h^p\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^{(1+\alpha_1)p}]$$
$$+ h^{\frac{p+1}{2}}\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^{p+\alpha_1}].$$

An application of the Cauchy-Schwarz inequality, in combination with Young's inequality gives

$$\Pi^{3,i}_t \lesssim \int_{\eta(t)}^t \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_s|^{p-2}|Y^{i,N}_{\eta(s)}||b_h(Y^{i,N}_{\eta(s)}, \mu^{Y^{\boldsymbol{Y}^N}}_{\eta(s)}) - b(Y^{i,N}_{\eta(s)}, \mu^{Y^{\boldsymbol{Y}^N}}_{\eta(s)})|\right] \mathrm{d}s$$
$$\lesssim \int_{\eta(t)}^t \mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_s|^p]\,\mathrm{d}s + h^{\frac{p+2}{2}}\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^{\frac{3p+2p\alpha_1}{2}}] + h.$$

Using $(\mathbf{A}^1_\sigma)$ and $(\mathbf{A}^4_\sigma)$, we get

$$\Pi^{4,i}_t \lesssim h + \int_{\eta(t)}^t \sup_{r\in[\eta(t),s]} \mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_r|^p]\,\mathrm{d}s.$$

Hence, plugging the bounds for $\Pi^{1,i}_t, \dots, \Pi^{4,i}_t$ into (3.15), we obtain for some constant $C > 0$

$$\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_t|^p] \le Ch + (1+Ch)\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^p] + C\int_{\eta(t)}^t \sup_{r\in[\eta(t),s]} \mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_r|^p]\,\mathrm{d}s$$
$$+ Ch^{\frac{p+2}{2}}\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^{\frac{3p+2p\alpha_1}{2}}]$$
$$+ C\sum_{l=1}^{p-1} h^{l+1}\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^{p+\alpha_1(l+1)}] + C\sum_{l=1}^{p-1} h^{l/2+1}\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^{p+\alpha_1}].$$

Defining now $R := R(h) = h^{-f(\alpha_1)}$, with $f(\alpha_1) = \frac{1}{1+2\alpha_1}$, we derive

$$\sup_{r\in[\eta(t),t]} \mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_r|^p] \le Ch + (1+Ch)\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{\eta(t)}|^p] + C\int_{\eta(t)}^t \sup_{r\in[\eta(t),s]} \mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_r|^p]\,\mathrm{d}s.$$

Since the sub-events $\Omega_{R,n}$, for $n \in \{0,\dots,M\}$, are decreasing, we have, using Gronwall's inequality (assuming h to be small enough),

$$\mathbb{E}[\mathrm{I}_{\Omega_{R,n+1}}|Y^{i,N}_{t_{n+1}}|^p] \le \mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{t_{n+1}}|^p] \le Ch + (1+Ch)\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{t_n}|^p],$$

for $t \in [t_n, t_{n+1}]$ and hence an inductive argument allows to show the moment stability of the i-th particle on these sub-events, i.e., we have

$$\max_{n\in\{0,\dots,M\}} \mathbb{E}\left[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{t_n}|^p\right] \le C.$$

Therefore, to complete the proof, it is sufficient, to prove $\mathbb{E}[\mathrm{I}_{\Omega^c_{R,n}}|Y^{i,N}_{t_n}|^q] \le C$, for some q depending on p (chosen sufficiently large). Employing the identity $\mathrm{I}_{\Omega^c_{R,n}} = \sum_{l=0}^n \mathrm{I}_{\Omega_{R,l-1}}\mathrm{I}_{|Y^{i,N}_{t_l}|>R}$, with the convention $\mathrm{I}_{\Omega_{R,-1}} = 1$, a combination of Hölder's inequality and Markov's inequality shows

$$\mathbb{E}[\mathrm{I}_{\Omega^c_{R,n}}|Y^{i,N}_{t_n}|^q] \le \sum_{l=0}^n \mathbb{E}\left[|Y^{i,N}_{t_n}|^q \mathrm{I}_{\Omega_{R,l-1}}\mathrm{I}_{|Y^{i,N}_{t_l}|>R}\right]$$
$$\le \left(\mathbb{E}[|Y^{i,N}_{t_n}|^{q\bar{p}}]\right)^{1/\bar{p}} \sum_{l=0}^n \left(\mathbb{P}(\mathrm{I}_{\Omega_{R,l-1}}|Y^{i,N}_{t_l}| > R)\right)^{1/\bar{q}}$$

$$\leq \Big(\mathbb{E}[|Y^{i,N}_{t_n}|^{q\bar{p}}]\Big)^{1/\bar{p}} \sum_{l=0}^{n} \frac{\Big(\mathbb{E}[\mathrm{I}_{\Omega_{R,l-1}}|Y^{i,N}_{t_l}|^p]\Big)^{1/\bar{q}}}{R^{p/\bar{q}}}, \tag{3.16}$$

where we chose $\bar{q} = \frac{p}{(q+1)f^{-1}(\alpha_1)}$, $\bar{p} = \frac{\bar{q}}{\bar{q}-1}$, $q \leq \frac{p-f^{-1}(\alpha_1)}{1+f^{-1}(\alpha_1)} < \frac{p}{f^{-1}(\alpha_1)} - 1$. Since $q\bar{p} \leq p$ is satisfied, Hölder's inequality yields

$$\Big(\mathbb{E}[|Y^{i,N}_{t_n}|^{q\bar{p}}|]\Big)^{1/\bar{p}} \leq \Big(\mathbb{E}[|Y^{i,N}_{t_n}|^{p}|]\Big)^{q/p} \lesssim 1 + h^{-q},$$

where we used $\mathbb{E}[|Y^{i,N}_{t_n}|^p] \lesssim 1 + h^{-p}$ (see, e.g., [68]). Using this estimate in (3.16), we obtain

$$\mathbb{E}[\mathrm{I}_{\Omega^c_{R,n}}|Y^{i,N}_{t_n}|^q] \leq C(1+h^{-q})(n+1)h^{q+1} \leq C,$$

and therefore

$$\begin{aligned}\mathbb{E}[|Y^{i,N}_{t_n}|^q] &= \mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{t_n}|^q] + \mathbb{E}[\mathrm{I}_{\Omega^c_{R,n}}|Y^{i,N}_{t_n}|^q] \\ &\leq \Big(\mathbb{E}[\mathrm{I}_{\Omega_{R,n}}|Y^{i,N}_{t_n}|^p]\Big)^{q/p} + \mathbb{E}[\mathrm{I}_{\Omega^c_{R,n}}|Y^{i,N}_{t_n}|^q] \leq C,\end{aligned}$$

which concludes the claim. □

3.3.2 Some auxiliary lemmas

Lemma 3.3.1. *Suppose Assumptions $(\mathbf{A}^1_b)$ and $(\mathbf{A}^1_\sigma)$–$(\mathbf{A}^3_\sigma)$ hold. Then, for all $p \geq 2$ there is a constant $C > 0$ such that*

$$\Lambda^{i,p}_t := \int_0^t (\mathbb{E}[\|\sigma(X^{i,N}_s, \mu^{\mathbf{X}^N}_s) - \Upsilon^i_s\|^{2p}])^{\frac{1}{p}} \mathrm{d}s \leq C\left(h^2 + \int_0^t (\mathbb{E}[|Z^{i,N}_s|^{2p}])^{\frac{1}{p}}\,\mathrm{d}s\right), \quad t \in [0,T]. \tag{3.17}$$

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

$$\begin{aligned}\Lambda^{i,p}_t &\lesssim \int_0^t (\mathbb{E}[\|\sigma(X^{i,N}_s, \mu^{\mathbf{X}^N}_s) - \sigma(Y^{i,N}_s, \mu^{\mathbf{Y}^N}_s)\|^{2p}])^{\frac{1}{p}}\,\mathrm{d}s + \int_0^t (\mathbb{E}[\|\Gamma^i_s\|^{2p}])^{\frac{1}{p}}\,\mathrm{d}s \\ &\lesssim \int_0^t (\mathbb{E}[|Z^{i,N}_s|^{2p}])^{\frac{1}{p}}\,\mathrm{d}s + \int_0^t (\mathbb{E}[\|\Gamma^i_s\|^{2p}])^{\frac{1}{p}}\,\mathrm{d}s,\end{aligned}$$

where we used the fact that $Z^{i,N}_t$, for $i \in \{1,\dots,N\}$, $t \in [0,T]$, are identically distributed. Consequently, to derive (3.17), it is sufficient to show that

$$(\mathbb{E}[\|\Gamma^i_t\|^{2p}])^{\frac{1}{p}} \lesssim h^2. \tag{3.18}$$

In the sequel, we aim to verify (3.18). According to [41, Proposition 3.1], we have

$$\begin{aligned}\partial_{x_j}\sigma_{kl}(x_i, \mu^{\mathbf{x}^N}) &= \partial_{x_j}\sigma_{kl}(x_i, \mu^{\mathbf{x}^N})\delta_{j,i} + \frac{1}{N}\partial_\mu\sigma_{kl}(x_i, \mu^{\mathbf{x}^N})(x_j), \\ \partial^2_{x_j}\sigma_{kl}(x_i, \mu^{\mathbf{x}^N}) &= \Big\{\partial^2_{x_j}\sigma_{kl}(x_i, \mu^{\mathbf{x}^N}) + \frac{2}{N}\partial_{x_i}\partial_\mu\sigma_{kl}(x_i, \mu^{\mathbf{x}^N})(x_i)\Big\}\delta_{j,i} \\ &\quad + \frac{1}{N^2}\partial^2_\mu\sigma_{kl}(x_i, \mu^{\mathbf{x}^N})(x_j, x_j) + \frac{1}{N}\partial_y\partial_\mu\sigma_{kl}(x_i, \mu^{\mathbf{x}^N})(x_j),\ i,j \in \{1,\dots,N\},\end{aligned} \tag{3.19}$$

where σ_{kl} is a component of σ for $k \in \{1,\dots,d\}$, $l \in \{1,\dots,m\}$. Also, $\delta_{j,i} = 1$, for $j = i$ and $\delta_{j,i} = 0$, for $j \neq i$. Itô's formula in combination with (3.19) yields for the (k,l)-th entry of Γ_t^i (denoted by $\Gamma_t^{i,(k,l)}$)

$$\Gamma_t^{i,(k,l)} = \sum_{r=1}^{9} \Pi_t^{r,i},$$

where we introduced

$$\Pi_t^{1,i} := \int_{\eta(t)}^t \Big\langle \partial_x \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N}), b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N}) \Big\rangle \, \mathrm{d}s,$$

$$\Pi_t^{2,i} := \frac{1}{N} \sum_{j=1}^N \int_{\eta(t)}^t \Big\langle \partial_\mu \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N})(Y_s^{j,N}), b_h(Y_{\eta(s)}^{j,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N}) \Big\rangle \, \mathrm{d}s,$$

$$\begin{aligned}\Pi_t^{3,i} := \frac{1}{2} \int_{\eta(t)}^t \mathrm{tr}\Big\{ \Big(\partial_x^2 \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N}) + \frac{2}{N} \partial_x \{ \partial_\mu \sigma_{kl}(\cdot, \mu_s^{\boldsymbol{Y}^N})(Y_s^{i,N}) \}(Y_s^{i,N}) \Big) \Upsilon_s^i (\Upsilon_s^i)^\top \\ + \sum_{j=1}^N \Big(\frac{1}{N^2} \partial_\mu^2 \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N})(Y_s^{j,N}, Y_s^{j,N}) \\ + \frac{1}{N} \partial_y \{ \partial_\mu \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N})(\cdot) \}(Y_s^{j,N}) \Big) \Upsilon_s^j (\Upsilon_s^j)^\top \Big\} \mathrm{d}s,\end{aligned}$$

$$\Pi_t^{4,i} := \int_{\eta(t)}^t \Big\langle \partial_x \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N}) - \partial_x \sigma_{kl}(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N}), \sigma(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N}) \mathrm{d}W_s^i \Big\rangle,$$

$$\Pi_t^{5,i} := \int_{\eta(t)}^t \Big\langle \partial_x \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N}), \int_{\eta(s)}^s \partial_x \sigma(Y_{\eta(r)}^{i,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N}) \sigma(Y_{\eta(r)}^{i,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N}) \, \mathrm{d}W_r^i \mathrm{d}W_s^i \Big\rangle,$$

$$\Pi_t^{6,i} := \frac{1}{N} \sum_{j=1}^N \int_{\eta(t)}^t \Big\langle \partial_\mu \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N})(Y_s^{j,N}), \int_{\eta(s)}^s \partial_x \sigma(Y_{\eta(r)}^{j,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N}) \sigma(Y_{\eta(r)}^{j,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N}) \, \mathrm{d}W_r^j \mathrm{d}W_s^j \Big\rangle,$$

$$\Pi_t^{7,i} := \frac{1}{N} \sum_{j=1}^N \int_{\eta(t)}^t \Big\langle \partial_x \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N}), \int_{\eta(s)}^s \partial_\mu \sigma(Y_{\eta(r)}^{i,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N})(Y_{\eta(r)}^{j,N}) \sigma(Y_{\eta(r)}^{j,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N}) \mathrm{d}W_r^j \mathrm{d}W_s^i \Big\rangle,$$

$$\Pi_t^{8,i} := \frac{1}{N^2} \sum_{j=1}^N \int_{\eta(t)}^t \Big\langle \partial_\mu \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N})(Y_s^{j,N}), \int_{\eta(s)}^s \partial_\mu \sigma(Y_{\eta(r)}^{i,N}, \mu_{\eta(r)}^{Y^{\cdot,N}})(Y_{\eta(r)}^{j,N}) \sigma(Y_{\eta(r)}^{j,N}, \mu_{\eta(r)}^{\boldsymbol{Y}^N}) \mathrm{d}W_r^j \mathrm{d}W_s^i \Big\rangle,$$

$$\Pi_t^{9,i} := \frac{1}{N} \sum_{j=1}^N \int_{\eta(t)}^t \Big\langle \partial_\mu \sigma_{kl}(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N})(Y_s^{j,N}) - \partial_\mu \sigma_{kl}(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})(Y_{\eta(s)}^{j,N}), \sigma(Y_{\eta(s)}^{j,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N}) \mathrm{d}W_s^j \Big\rangle.$$

To estimate these terms one-by-one, we first note that Lemma 3.2.1 implies that for any $p \geq 2$ there is a constant $C > 0$ such that

$$\max_{i \in \{1,\dots,N\}} \sup_{t \in [0,T]} \mathbb{E}[|Y_t^{i,N}|^p] \leq C. \tag{3.20}$$

Due to item (2) and item (3) of $(\mathbf{A}_b^1)$ and $(\mathbf{A}_\sigma^1)$, one has for any $x \in \mathbb{R}^d$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$|b(x,\mu)| + \|\sigma(x,\mu)\| \lesssim 1 + |x|^{1+\alpha_1} + \mu(|\cdot|^2)^{\frac{1}{2}}. \tag{3.21}$$

Obviously, (3.20) and (3.21) imply that for any $p \geq 2$

$$\sup_{t \in [0,T]} \{ \mathbb{E}[|b(Y_t^{i,N}, \mu_t^{\boldsymbol{Y}^N})|^p] + \mathbb{E}[\|\sigma(Y_t^{i,N}, \mu_t^{\boldsymbol{Y}^N})\|^p] \} \leq C. \tag{3.22}$$

By standard inequalities, it follows from (3.20)–(3.22), and $(\mathbf{A}_\sigma^1)$–$(\mathbf{A}_\sigma^3)$ that

$$\sum_{r=1}^{3}(\mathbb{E}[\|\Pi_t^{r,i}\|^{2p}])^{\frac{1}{p}} \lesssim h\int_{\eta(t)}^{t}\mathrm{d}s\ (\mathbb{E}[|b(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})|^{2p}])^{\frac{1}{p}} + \frac{h}{N}\sum_{j=1}^{N}\int_{\eta(t)}^{t}\mathrm{d}s\ (\mathbb{E}[|b(Y_{\eta(t)}^{j,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})|^{2p}])^{\frac{1}{p}}$$
$$+ h\int_{\eta(t)}^{t}(\mathbb{E}[\|\Upsilon_s^i\|^{2p}])^{\frac{1}{p}}\mathrm{d}s \lesssim h^2. \tag{3.23}$$

Taking (3.20)-(3.22), $(\mathbf{A}_\sigma^1)$ and $(\mathbf{A}_\sigma^4)$ into consideration yields

$$\sup_{t\in[0,T]}\mathbb{E}[|Y_t^{i,N}-Y_{\eta(t)}^{i,N}|^q] \lesssim h^{\frac{q}{2}}, \qquad q\geq 1, \quad i\in\{1,\dots,N\}.$$

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

$$\begin{aligned}
&\sum_{r=4}^{9}(\mathbb{E}[\|\Pi_t^{r,i}\|^{2p}])^{\frac{1}{p}}\\
&\lesssim h^2\left(\sum_{n=1}^{m}\mathbb{E}[\|\partial_x\sigma_n(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})\sigma(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})\|^{2p}]\right)^{\frac{1}{p}}\\
&\quad+(\mathbb{E}[\|\sigma(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})\|^{4p}])^{\frac{1}{2p}}\int_{\eta(t)}^{t}(\mathbb{E}[|\partial_x\sigma_{kl}(Y_s^{i,N},\mu_s^{\boldsymbol{Y}^N})-\partial_x\sigma_{kl}(Y_{\eta(s)}^{i,N},\mu_{\eta(s)}^{\boldsymbol{Y}^N})|^{4p}])^{\frac{1}{2p}}\mathrm{d}s\\
&\quad+\frac{h^2}{N}\sum_{j=1}^{N}\left(\sum_{n=1}^{m}\mathbb{E}[\|\partial_x\sigma_n(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})\sigma(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})\|^{2p}\right)^{\frac{1}{p}}\\
&\quad+\frac{h^2}{N}\sum_{j=1}^{N}\left(\sum_{n=1}^{m}\mathbb{E}[\|\partial_\mu\sigma_n(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})(Y_{\eta(t)}^{j,N})\sigma(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})\|^{2p}\right)^{\frac{1}{p}}\\
&\quad+\frac{h^2}{N^2}\sum_{j=1}^{N}\left(\sum_{n=1}^{m}\mathbb{E}[\|\partial_\mu\sigma_n(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})(Y_{\eta(t)}^{j,N})\sigma(Y_{\eta(t)}^{j,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})\|^{2p}]\right)^{\frac{1}{p}}\\
&\quad+\frac{1}{N}\sum_{j=1}^{N}\int_{\eta(t)}^{t}(\mathbb{E}[|\partial_\mu\sigma_{kl}(Y_s^{i,N},\mu_s^{\boldsymbol{Y}^N})(Y_s^{j,N})-\partial_\mu\sigma_{kl}(Y_{\eta(s)}^{i,N},\mu_{\eta(s)}^{\boldsymbol{Y}^N})(Y_{\eta(s)}^{j,N})|^{4p}])^{\frac{1}{2p}}\mathrm{d}s\\
&\lesssim h^2,
\end{aligned}\tag{3.24}$$

where we used $(\mathbf{A}_\sigma^1)$–$(\mathbf{A}_\sigma^3)$, and (3.20)-(3.22) along with the one-step error bound. Consequently, (3.18) follows from (3.23) and (3.24). □

Lemma 3.3.2. *Suppose Assumptions $(\mathbf{A}_b^1)$ and $(\mathbf{A}_\sigma^1)$–$(\mathbf{A}_\sigma^2)$ hold. Then, for all $\varepsilon>0$ and $p\geq 2$ there exists a constant $C>0$ such that for $t\in[0,T]$*

$$\left(\mathbb{E}\left[\left|\int_0^t\langle Z_s^{i,N},\hat{\Upsilon}_s^i\rangle\,\mathrm{d}s\right|^p\right]\right)^{1/p} \leq \varepsilon\left(\mathbb{E}\left[\sup_{s\in[0,t]}|Z_s^{i,N}|^{2p}\right]\right)^{\frac{1}{p}} + C\left(\int_0^t(\mathbb{E}[|Z_s^{i,N}|^{2p}])^{\frac{1}{p}}\mathrm{d}s+h^2\right),$$

where

$$\hat{\Upsilon}_t^i := \partial_x b(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})\sigma(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})(W_t^i-W_{\eta(t)}^i).$$

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

$$\frac{1}{N}\sum_{j=1}^{N}\partial_\mu b(Y_{\eta(t)}^{i,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})(Y_{\eta(t)}^{j,N})\sigma(Y_{\eta(t)}^{j,N},\mu_{\eta(t)}^{\boldsymbol{Y}^N})(W_t^j-W_{\eta(t)}^j).$$

Proof. The proof of this lemma is detailed in [15, Lemma 3.2]. □

Lemma 3.3.3. *Let Assumptions* $(\mathbf{A}_b^1)$*–*$(\mathbf{A}_b^3)$ *and* $(\mathbf{A}_\sigma^1)$*–*$(\mathbf{A}_\sigma^2)$ *hold. Then, for all* $\varepsilon > 0$ *and* $p \geq 2$ *there exists a constant* $C > 0$ *such that*

$$\begin{aligned}\hat{\Gamma}_t^{i,p} &:= \int_0^t \left(\mathbb{E}\left[\left|\left\langle Z_s^{i,N}, b(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N}) - b(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle\right|^p\right]\right)^{\frac{1}{p}} \mathrm{d}s \\ &\leq \varepsilon \left(\mathbb{E}\left[\sup_{s\in[0,t]} |Z_s^{i,N}|^{2p}\right]\right)^{\frac{1}{p}} + C\left(\int_0^t (\mathbb{E}[|Z_s^{i,N}|^{2p}])^{\frac{1}{p}} \mathrm{d}s + h^2\right), \quad t \in [0,T].\end{aligned} \tag{3.25}$$

Proof. The proof can be carried out similarly to the proof of Lemma 3.3.1 in combination with Lemma 3.3.2 and is therefore omitted. □

3.3.3 Proof of Theorem 3.2.1

With Lemmas 3.3.1 and 3.3.3 at hand, we are in a position to complete the proof of Theorem 3.2.1:

Proof. We aim to show that

$$\mathbb{E}\left[\sup_{t\in[0,T]} |Z_t^{i,N}|^p\right] \lesssim h^{2p}, \tag{3.26}$$

for $p \geq 2$. By Itô's formula together with $Z_0^{i,N} = 0$, it follows that

$$\begin{aligned}|Z_t^{i,N}|^2 = 2\int_0^t \left\langle Z_s^{i,N}, b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N}) - b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle \mathrm{d}s + 2\int_0^t \langle Z_s^{i,N}, \mathrm{d}M_s^i\rangle \\ + \int_0^t \|\sigma(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N}) - \Upsilon_s^i\|^2 \,\mathrm{d}s,\end{aligned}$$

where M_t^i and Υ_t^i for $t \in [0,T]$ are defined in (3.14). The first term from above will be estimated based on the decomposition

$$\int_0^t \left\langle Z_s^{i,N}, b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N}) - b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle \mathrm{d}s = \Pi_t^{1,i} + \Pi_t^{2,i} + \Pi_t^{3,i} + \Pi_t^{4,i},$$

where we defined

$$\begin{aligned}\Pi_t^{1,i} &:= \int_0^t \left\langle Z_s^{i,N}, b(X_s^{i,N}, \mu_s^{\boldsymbol{X}^N}) - b(Y_s^{i,N}, \mu_s^{\boldsymbol{X}^N})\right\rangle \mathrm{d}s, \\ \Pi_t^{2,i} &:= \int_0^t \left\langle Z_s^{i,N}, b(Y_s^{i,N}, \mu_s^{\boldsymbol{X}^N}) - b(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N})\right\rangle \mathrm{d}s, \\ \Pi_t^{3,i} &:= \int_0^t \left\langle Z_s^{i,N}, b(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N}) - b_h(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle \mathrm{d}s, \\ \Pi_t^{4,i} &:= \int_0^t \left\langle Z_s^{i,N}, b(Y_s^{i,N}, \mu_s^{\boldsymbol{Y}^N}) - b(Y_{\eta(s)}^{i,N}, \mu_{\eta(s)}^{\boldsymbol{Y}^N})\right\rangle \mathrm{d}s.\end{aligned}$$

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

$$\sup_{s\in[0,t]} |\Pi_s^{1,i}| \lesssim \int_0^t |Z_s^{i,N}|^2 \,\mathrm{d}s. \tag{3.27}$$

Due to item (3) of $(\mathbf{A}_b^1)$ and Young's inequality, we obtain from (3.5)

$$\sup_{s\in[0,t]}|\Pi_s^{2,i}| \le \frac{1}{2}\int_0^t |Z_s^{i,N}|^2\,\mathrm{d}s + \frac{(L_b^1)^2}{2N}\sum_{j=1}^N\int_0^t |Z_s^{j,N}|^2\,\mathrm{d}s. \tag{3.28}$$

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

$$\sup_{s\in[0,t]}|\Pi_s^{3,i}| \lesssim \int_0^t |Z_s^{i,N}|^2\,\mathrm{d}s + h^2\int_0^t\left(1+|Y_{\eta(s)}^{i,N}|^{2(1+\alpha_1)} + \frac{1}{N}\sum_{j=1}^N |Y_{\eta(s)}^{j,N}|^2\right)\mathrm{d}s. \tag{3.29}$$

Thus, combining (3.27), (3.28) with (3.29) and taking advantage of Lemma 3.3.1 and Lemma 3.3.3 (with ε small enough), along with (3.20) and (3.22), we deduce from Minkowski's inequality that

$$\begin{aligned}\left(\mathbb{E}\left[\sup_{s\in[0,t]}|Z_s^{i,N}|^{2p}\right]\right)^{\frac{1}{p}} &\lesssim \Bigg\{\int_0^t(\mathbb{E}[|Z_s^{i,N}|^{2p}])^{\frac{1}{p}}\mathrm{d}s + \frac{1}{N}\sum_{j=1}^N\int_0^t(\mathbb{E}[|Z_s^{j,N}|^{2p}])^{\frac{1}{p}}\mathrm{d}s \\ &\quad + h^2 + \hat{\Gamma}_t^{i,p} + \Lambda_t^{i,p} + \left(\mathbb{E}\left[\sup_{s\in[0,t]}\left|\int_0^s\langle Z_r^{i,N},\mathrm{d}M_r^i\rangle\right|^p\right]\right)^{\frac{1}{p}}\Bigg\} \\ &\le \frac{1}{2}\left(\mathbb{E}\left[\sup_{s\in[0,t]}|Z_s^{i,N}|^{2p}\right]\right)^{\frac{1}{p}} + C\left(h^2 + \int_0^t(\mathbb{E}[|Z_s^{i,N}|^{2p}])^{\frac{1}{p}}\mathrm{d}s\right),\end{aligned} \tag{3.30}$$

where we employed that $Z_t^{i,N}$, for $i\in\{1,\dots,N\}$, $t\in[0,T]$, are identically distributed and that, for some constant $C>0$,

$$\left(\mathbb{E}\left[\sup_{s\in[0,t]}\left|\int_0^s\langle Z_r^{i,N},\,\mathrm{d}M_r^i\rangle\right|^p\right]\right)^{\frac{1}{p}} \le \frac{\varepsilon}{2}\left(\mathbb{E}\left[\sup_{s\in[0,t]}|Z_s^{i,N}|^{2p}\right]\right)^{\frac{1}{p}} + \frac{C}{2\varepsilon}\Lambda_t^{i,p}.$$

Thus, (3.26) follows from (3.30) and Gronwall's inequality. □

3.3.4 Proof of Proposition 3.2.1

Proof. From the definitions of $Y_{t_{n+1}}^i$ and $Y_{t_{n+1}}^{i,N}$, for any $n\in\{0,\dots,M-1\}$ and $i\in\{1,\dots,N\}$, we directly obtain

$$\begin{aligned}&|Y_{t_{n+1}}^{i,N} - Y_{t_{n+1}}^i|^2 \\ &= \Big|Y_{t_n}^{i,N} - Y_{t_n}^i + b(Y_{t_n}^{i,N},\mu_{t_n}^{\mathbf{Y}^N})h - b(Y_{t_n}^i,\mathcal{L}_{Y_{t_n}^i})h + \left(\sigma(Y_{t_n}^{i,N},\mu_{t_n}^{\mathbf{Y}^N}) - \sigma(Y_{t_n}^i,\mathcal{L}_{Y_{t_n}^i})\right)\Delta W_n^i \\ &\quad + \int_{t_n}^{t_{n+1}}\left(\partial_x\sigma(Y_{t_n}^{i,N},\mu_{t_n}^{\mathbf{Y}^N})\sigma(Y_{t_n}^{i,N},\mu_{t_n}^{\mathbf{Y}^N}) - \partial_x\sigma(Y_{t_n}^i,\mathcal{L}_{Y_{t_n}^i})\sigma(Y_{t_n}^i,\mathcal{L}_{Y_{t_n}^i})\right)\int_{t_n}^s \mathrm{d}W_u^i\mathrm{d}W_s^i \\ &\quad + \int_{t_n}^{t_{n+1}}\frac{1}{N}\sum_{j=1}^N\partial_\mu\sigma(Y_{t_n}^{i,N},\mu_{t_n}^{\mathbf{Y}^N})(Y_{t_n}^{j,N})\sigma(Y_{t_n}^{j,N},\mu_{t_n}^{\mathbf{Y}^N})\int_{t_n}^s \mathrm{d}W_u^j\mathrm{d}W_s^i\Big|^2.\end{aligned}$$

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

$$\mathbb{E}\left[|Y_{t_{n+1}}^{i,N} - Y_{t_{n+1}}^i|^2\right] \le \mathbb{E}\left[|Y_{t_n}^{i,N} - Y_{t_n}^i|^2\right] + C\sum_{j=1}^5\Pi_{n+1}^{j,i},$$

where $C > 0$ and

$$\Pi^{1,i}_{n+1} := \mathbb{E}\left[\left|b(Y^{i,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n})h - b(Y^i_{t_n}, \mathcal{L}_{Y^i_{t_n}})h\right|^2\right]$$

$$\Pi^{2,i}_{n+1} := \mathbb{E}\left[\left|\left(\sigma(Y^{i,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n}) - \sigma(Y^i_{t_n}, \mathcal{L}_{Y^i_{t_n}})\right)\Delta W^i_n\right|^2\right]$$

$$\Pi^{3,i}_{n+1} := \mathbb{E}\left[\left|\int_{t_n}^{t_{n+1}}\left(\partial_x\sigma(Y^{i,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n})\sigma(Y^{i,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n}) - \partial_x\sigma(Y^i_{t_n}, \mathcal{L}_{Y^i_{t_n}})\sigma(Y^i_{t_n}, \mathcal{L}_{Y^i_{t_n}})\right)\int_{t_n}^s \mathrm{d}W^i_u\mathrm{d}W^i_s\right|^2\right]$$

$$\Pi^{4,i}_{n+1} := \mathbb{E}\left[\left|\int_{t_n}^{t_{n+1}}\frac{1}{N}\sum_{j=1}^N \partial_\mu\sigma(Y^{i,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n})(Y^{j,N}_{t_n})\sigma(Y^{j,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n})\int_{t_n}^s \mathrm{d}W^j_u\mathrm{d}W^i_s\right|^2\right]$$

$$\Pi^{5,i}_{n+1} := 2\mathbb{E}\Big\langle Y^{i,N}_{t_n} - Y^i_{t_n}, b(Y^{i,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n})h - b(Y^i_{t_n}, \mathcal{L}_{Y^i_{t_n}})h\Big\rangle.$$

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

First, using the imposed Lipschitz assumptions on $(x,\mu) \mapsto b(x,\mu)$, $(x,\mu) \mapsto \sigma(x,\mu)$ and $(x,\mu) \mapsto \partial_x\sigma_l(x,\mu)\sigma(x,\mu)$ for $l \in \{1,\dots,m\}$, we can readily estimate

$$\Pi^{1,i}_{n+1} + \Pi^{2,i}_{n+1} + \Pi^{3,i}_{n+1} \lesssim h\left(\mathbb{E}\left[|Y^{i,N}_{t_n} - Y^i_{t_n}|^2\right] + \mathbb{E}\left[\mathcal{W}_2^2\left(\mathcal{L}_{Y^i_{t_n}}, \mu^{\boldsymbol{Y}}_{t_n}\right)\right]\right),$$

where $\mu^{\boldsymbol{Y}}_{t_n}$ is the empirical distribution of the non-interacting particle system $(Y^i_{t_n})_{i\in\{1,\dots,N\}}$. Using BDG's inequality, the term $\Pi^{4,i}_{n+1}$ can be bounded by the following quantity:

$$\mathbb{E}\left[\left|\int_{t_n}^{t_{n+1}}\frac{1}{N}\sum_{j=1}^N \partial_\mu\sigma(Y^{i,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n})(Y^{j,N}_{t_n})\sigma(Y^{j,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n})\int_{t_n}^s \mathrm{d}W^j_u\mathrm{d}W^i_s\right|^2\right] \lesssim \frac{h^2}{N},$$

where we employed the moment stability of the random variables $(Y^{i,N}_{t_n})_{i\in\{1,\dots,N\}}$ and the independence of the Brownian motions. Next, we compute, using the fact that b is Lipschitz continuous in both components,

$$\begin{aligned}\Pi^{5,i}_{n+1} &= 2\mathbb{E}\Big\langle Y^{i,N}_{t_n} - Y^i_{t_n}, b(Y^{i,N}_{t_n}, \mu^{\boldsymbol{Y}^N}_{t_n})h - b(Y^i_{t_n}, \mathcal{L}_{Y^i_{t_n}})h\Big\rangle \\ &\lesssim h\left(\mathbb{E}\left[|Y^{i,N}_{t_n} - Y^i_{t_n}|^2\right] + \mathbb{E}\left[\mathcal{W}_2^2\left(\mathcal{L}_{Y^i_{t_n}}, \mu^{\boldsymbol{Y}}_{t_n}\right)\right]\right).\end{aligned}$$

Hence, putting everything together, we obtain

$$\begin{aligned}\mathbb{E}\left[|Y^{i,N}_{t_{n+1}} - Y^i_{t_n}|^2\right] &\le \mathbb{E}\left[|Y^{i,N}_{t_n} - Y^i_{t_n}|^2\right] + Ch\left(\mathbb{E}\left[|Y^{i,N}_{t_n} - Y^i_{t_n}|^2\right] + \frac{h}{N} + \mathbb{E}\left[\mathcal{W}_2^2\left(\mathcal{L}_{Y^i_{t_n}}, \mu^{\boldsymbol{Y}^N}_{t_n}\right)\right]\right) \\ &\le (1+Ch)\mathbb{E}\left[|Y^{i,N}_{t_n} - Y^i_{t_n}|^2\right] + Ch\phi(N),\end{aligned}$$

where $\phi(N)$ is defined in (3.10). The constant $C > 0$ is independent of the present time-step $n \in \{0,\dots,M\}$. Consequently, recalling that $Y^{i,N}_0 = Y^i_0$, the claim follows by an inductive argument. □

3.4 Numerical results and implementation details

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

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

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

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

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

3.4.1 Examples

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

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

Example 1: $\mathrm{d}X_t = \left(\frac{\sigma^2}{2} X_t - X_t^3 + c\mathbb{E}[X_t]\right)\mathrm{d}t + \mathbb{E}[X_t]\,\mathrm{d}W_t.$

Example 2: $\mathrm{d}X_t = \left(\frac{\sigma^2}{2} X_t - X_t^3 + c\mathbb{E}[X_t]\right)\mathrm{d}t + X_t\,\mathrm{d}W_t.$

Example 3: $\mathrm{d}X_t = \left(\frac{\sigma^2}{2} X_t - X_t^3 + \int_{\mathbb{R}} \sin(X_t - y)\,\mathcal{L}_{X_t}(\mathrm{d}y)\right)\mathrm{d}t + \left(X_t + \int_{\mathbb{R}} \sin(X_t - y)\,\mathcal{L}_{X_t}(\mathrm{d}y)\right)\mathrm{d}W_t.$

Example 4: $\mathrm{d}X_t = \left(\frac{\sigma^2}{2} X_t - X_t^3 + \mathbb{E}[X_t]\right)\mathrm{d}t + \int_{\mathbb{R}}\int_{\mathbb{R}} \sin(x+y)\mathcal{L}_{X_t}(\mathrm{d}y)\mathcal{L}_{X_t}(\mathrm{d}x)\,\mathrm{d}W_t.$

Example 5: $\mathrm{d}X_t = \left(\frac{\sigma^2}{2} X_t - X_t^3 + \mathbb{E}[X_t]\right)\mathrm{d}t + e^{-\mathbb{V}\left[\frac{X_t}{1+X_t^2}\right]}\,\mathrm{d}W_t.$

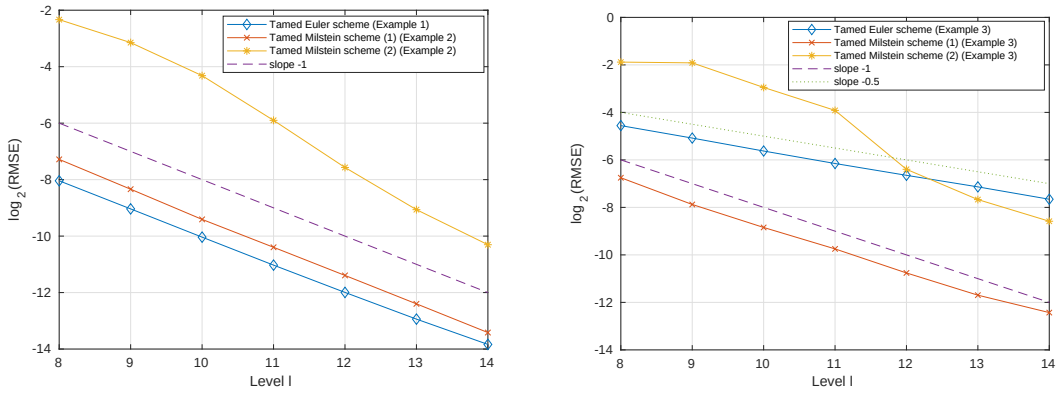
Note that not all model assumptions needed to guarantee moment stability for Scheme 1 are satisfied in the case of Examples 1, 2, 4 and 5, as $(\mathbf{A}_b^4)$ and/or $(\mathbf{A}_\sigma^4)$ are violated. However, as illustrated numerically, it seems reasonable that the established theory is also valid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

numerically below as well. The drift b_h of the tamed Euler scheme is chosen as in Scheme 1 for the Milstein scheme.

3.4.2 Time-stepping schemes and L-derivative

We first use the two different tamed Milstein schemes (i.e., Scheme 1 and Scheme 2) without the L-derivative terms for Examples 1 to 3. For all these examples, we observe in Fig. 3.1a and Fig. 3.1b strong convergence of order 1.

From these numerical results, it is also apparent that Scheme 1 consistently outperforms Scheme 2 in terms of accuracy. We further note that the strong convergence order for Scheme 2 seems to be observable for extremely fine time-grids (i.e., $M \geq 2^{12}$) only, which might be due to a large implied constant appearing in the strong convergence analysis of Remark 3.2.4.



(a) (b)

Figure 3.1: Numerical illustration for the strong convergence of Example 1, Example 2 (both left), and Example 3 (right) using the tamed Euler scheme and both tamed Milstein schemes, Scheme 1 and Scheme 2 (without the L-derivative terms), respectively. In Example 1, the tamed Milstein scheme without the L-derivative terms then coincides with the tamed Euler scheme.

Note that in the above examples the L-derivative of the diffusion term can be explicitly computed, which allows us to compute the term (3.12). In Example 1, the L-derivative of the diffusion term is simply $\partial_\mu \sigma(x,\mu)(y) = 1$, for all $y \in \mathbb{R}$ (see [31, Example 1 on page 385]). Similarly, the L-derivative of the diffusion term appearing in Example 3 can be computed as $\partial_\mu \sigma(x,\mu)(y) = \partial_y \sin(x-y) = -\cos(x-y)$, for all $y \in \mathbb{R}$ (see [31, Example 4 on page 389]), and for Example 4 as $\partial_\mu \sigma(x,\mu)(y) = 2\int_{\mathbb{R}} \cos(x+y)\,\mu(\mathrm{d}x)$, for all $y \in \mathbb{R}$ (see [31, Example 3 on page 387]). Finally, for Example 5, the chain rule shows that the L-derivative of $f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ defined as

$$f(\mu) := \exp\Big(- \int_{\mathbb{R}} g^2(x)\mu(\mathrm{d}x) + \Big(\int_{\mathbb{R}} g(x)\mu(\mathrm{d}x) \Big)^2 \Big),$$

for some bounded and continuously differentiable function $g : \mathbb{R} \to \mathbb{R}$ (with bounded derivative), is

$$\partial_\mu f(\mu)(y) = \exp\Big(- \int_{\mathbb{R}} g^2(x)\mu(\mathrm{d}x) + \Big(\int_{\mathbb{R}} g(x)\mu(\mathrm{d}x) \Big)^2 \Big) \Big(- 2g(y)g'(y) + 2g'(y) \int_{\mathbb{R}} g(x)\mu(\mathrm{d}x) \Big),$$

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

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

$$Y^{i,N}_{t_{n+1}} = Y^{i,N}_{t_n} + \frac{\frac{\sigma^2}{2} Y^{i,N}_{t_n} - (Y^{i,N}_{t_n})^3 + \frac{c}{N}\sum_{j=1}^N Y^{j,N}_{t_n}}{1 + h\left|\frac{\sigma^2}{2} Y^{i,N}_{t_n} - (Y^{i,N}_{t_n})^3 + \frac{c}{N}\sum_{j=1}^N Y^{j,N}_{t_n}\right|} h + \frac{1}{N}\sum_{j=1}^N Y^{j,N}_{t_n} \Delta W^i_n$$
$$+ \frac{1}{N}\sum_{l=1}^N Y^{l,N}_{t_n} \frac{1}{N}\sum_{j=1}^N \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}W^j_u \mathrm{d}W^i_s.$$

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

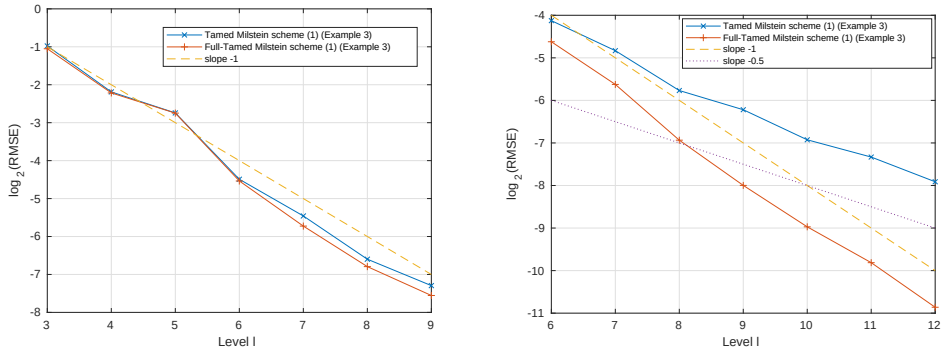
$$Y^{i,N}_{t_{n+1}} = Y^{i,N}_{t_n} + \frac{\frac{\sigma^2}{2} Y^{i,N}_{t_n} - (Y^{i,N}_{t_n})^3 + \frac{1}{N}\sum_{j=1}^N \sin\left(Y^{i,N}_{t_n} - Y^{j,N}_{t_n}\right)}{1 + h\left|\frac{\sigma^2}{2} Y^{i,N}_{t_n} - (Y^{i,N}_{t_n})^3 + \frac{1}{N}\sum_{j=1}^N \sin\left(Y^{i,N}_{t_n} - Y^{j,N}_{t_n}\right)\right|} h$$
$$+ \left(Y^{i,N}_{t_n} + \frac{1}{N}\sum_{j=1}^N \sin\left(Y^{i,N}_{t_n} - Y^{j,N}_{t_n}\right)\right) \Delta W^i_n$$
$$+ \left(1 + \frac{1}{N}\sum_{j=1}^N \cos\left(Y^{i,N}_{t_n} - Y^{j,N}_{t_n}\right)\right) \left(Y^{i,N}_{t_n} + \frac{1}{N}\sum_{j=1}^N \sin\left(Y^{i,N}_{t_n} - Y^{j,N}_{t_n}\right)\right) \frac{(\Delta W^i_n)^2 - h}{2}$$
$$- \frac{1}{N}\sum_{j=1}^N \cos\left(Y^{i,N}_{t_n} - Y^{j,N}_{t_n}\right) \left(Y^{j,N}_{t_n} + \frac{1}{N}\sum_{l=1}^N \sin\left(Y^{j,N}_{t_n} - Y^{l,N}_{t_n}\right)\right) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathrm{d}W^j_u \mathrm{d}W^i_s.$$

To illustrate the effect of the L-derivative, in Fig. 3.2a, we compare for Example 3 the strong convergence rate of the tamed Milstein scheme (Scheme 1) without the terms involving the L-derivative, with a full-tamed Milstein scheme, i.e., including the L-derivative terms, for $N = 20$. We employ the approximation to the Lévy areas proposed in [151]. The complexity for computing the Lévy areas for all time-steps is $\mathcal{O}(N^2 M^{3/2})$, where the additional factor $M^{1/2}$ comes from the choice for the truncation level of the so-called Karhunen-Loève expansion of certain Brownian Bridge processes.

Both schemes converge strongly with order 1. Although the full-tamed Milstein scheme shows a slightly better accuracy, we note that this additional gain is not significant taking the complexity of computing the Lévy areas into account.[1] In Fig. 3.2b we show results for the same test with $N = 3$; also performed for Examples 4 and 5 in Fig. 3.3a. This illustrates the necessity of implementing the L-derivative terms to achieve the desired strong convergence order of 1 for fixed N, since the scheme without the L-derivative terms only converges with order 1/2. Fig. 3.3b explores this further by comparing, in the case of Example 1 (with $c = 1$), the strong convergence rate of a tamed Euler scheme with a full-tamed Milstein scheme including the L-derivative terms for different small N.

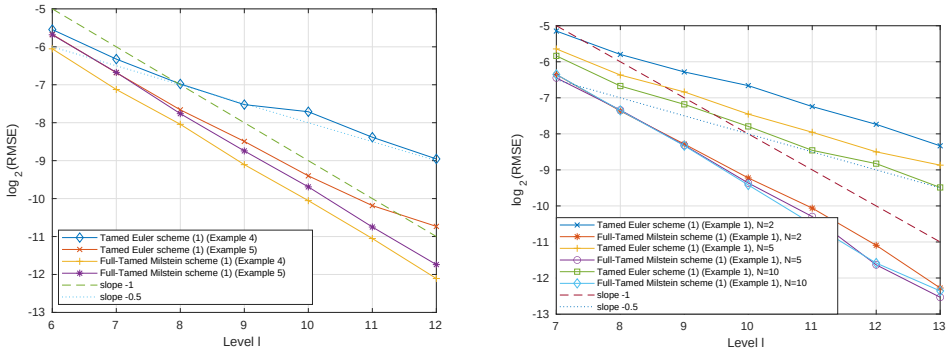
In Fig. 3.4a and Fig. 3.4b, we compute for a given step size $h = 1/M$ the root-mean-square difference (for several choices of N) between a numerical approximation of the particle system

[1] Due to the high computational complexity, the time discretisation is coarser than the one chosen in Fig. 3.3b below.



(a) (b)

Figure 3.2: Numerical illustration of the strong convergence in Example 3 for the tamed Milstein scheme (Scheme 1) with and without the L-derivative terms, with $N = 20$ particles (left) and $N = 3$ (right).



(a) (b)

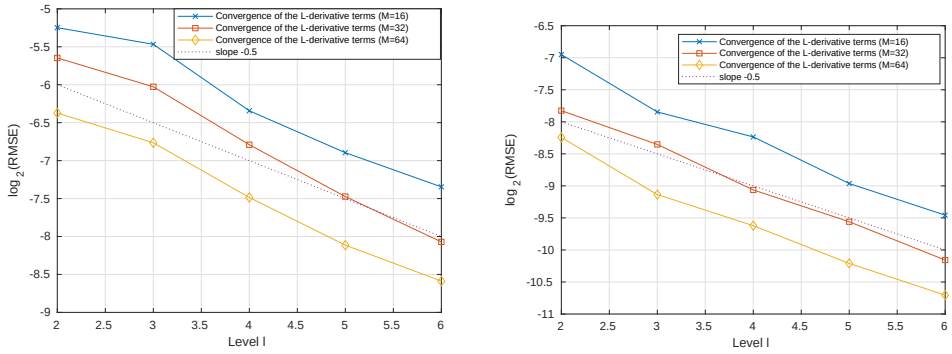
Figure 3.3: Illustration of the effect of the L-derivative terms in the tamed Milstein scheme (Scheme 1) for Examples 4 and 5 (left, for $N = 3$ particles) and Example 1 (right, for three different choices of N).

associated with the McKean–Vlasov SDE in Examples 4 and 5, obtained using a tamed Euler scheme (i.e., ignoring the L-derivative terms) and one obtained by the full-tamed Milstein scheme, at the final time $T = 1$. We choose $M = 2^4, 2^5, 2^6$ for our tests. For a fixed choice of M, we then compute the RMSE for $N_l = 2^l$, for $l \in \{2, \dots, 6\}$, where

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

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

We observe a strong convergence rate $1/2$ with respect to the number of particles. The same test is performed for Example 1 in Fig. 3.5a, and we observe a similar convergence behaviour of the



(a) (b)

Figure 3.4: Convergence of the terms involving the L-derivatives with an increasing number of particles for Example 4 (left) and Example 5 (right).

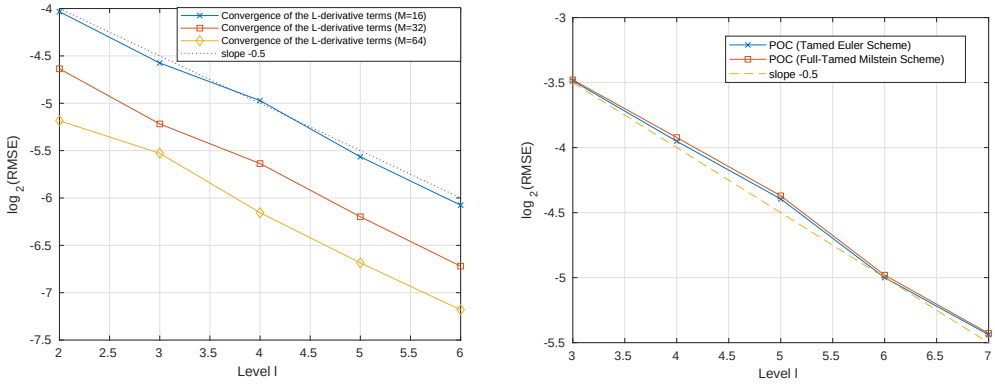
L-derivative terms as for Examples 4 and 5 above.

3.4.3 Propagation of chaos

Fig. 3.5b depicts the strong propagation of chaos convergence rate for Example 1. For a fixed number of time steps $M = 2^6$, we compute for different sizes of the particle system, $N_l = 2^l$, for $l \in \{3, \dots, 7\}$, a RMSE between two particle systems,

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

where the particle system $\tilde{Y}^{i,N_l,M}$, $i \in \{1, \dots, N_l\}$, is obtained by splitting the set of Brownian motions driving the particle system $Y^{i,N_l,M}$, $i \in \{1, \dots, N_l\}$, in two sets and simulating two independent particle systems, each of size $N_l/2$: $(\tilde{Y}^{i,N_l,M,(1)})_{i \in \{1,\dots,N_l/2\}}$ is a particle system obtained by using the Brownian motions $(W^i)_{i \in \{1,\dots,N_l/2\}}$ and $(\tilde{Y}^{i,N_l,M,(2)})_{i \in \{N_l/2+1,\dots,N_l\}}$ is obtained from the set $(W^i)_{i \in \{N_l/2+1,\dots,N_l\}}$. In particular, this means that for these two smaller particle systems only $N_l/2$ particles are used to approximate the mean-field term. As a time discretisation scheme, we employ a tamed Euler scheme and the full-tamed Milstein scheme, respectively. We observe a strong convergence order of $1/2$ in terms of number of particles. Corollary 3.2.1 only proves a strong convergence order of $1/4$ (see (3.10) for $d = 1$). However, recent results (see, e.g., [146, Theorem 2.4]) suggest that the optimal rate is $1/2$. These results do not apply to our problem description, as they rely on strong regularity assumptions on the coefficients of the underlying McKean–Vlasov SDE (in particular, the drift needs to be globally Lipschitz continuous in the state component).



(a) (b)

Figure 3.5: Convergence of the terms involving the L-derivatives with an increasing number of particles for Example 1 (left). Illustration of the propagation of chaos property (POC) for Example 1 (right).

Chapter 4

Well-posedness and numerical schemes for one-dimensional McKean–Vlasov equations and interacting particle systems with discontinuous drift

4.1 Introduction

In this chapter, we study the existence and uniqueness of strong solutions for classes of McKean–Vlasov SDEs, where the drift exhibits a discontinuity in the spatial component. We also provide time-stepping schemes of Euler–Maruyama type, for which we show strong convergence of a certain rate.

A McKean–Vlasov equation (introduced in [114, 86]) for a d-dimensional process $X = (X_t)_{t\in[0,T]}$, with a given finite time-horizon $T > 0$, is an SDE where the underlying coefficients depend on the current state X_t and, additionally, on the law of X_t. We consider more specifically one-dimensional equations of the form

$$\mathrm{d}X_t = b(X_t, \mathcal{L}_{X_t})\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0 = \xi \in L_0^2(\mathbb{R}), \tag{4.1}$$

where $W = (W_t)_{t\in[0,T]}$ is a one-dimensional standard Brownian motion and $\mathcal{L}_{X_t}$ denotes the marginal law of the process X at time $t \in [0,T]$. In particular, we are concerned with the well-posedness of equations of the form (4.1), where $b(\cdot,\mu)$ is discontinuous in zero, and piecewise Lipschitz continuous on the subintervals $(-\infty, 0)$ and $(0,\infty)$. Concerning the measure component of the drift, we will require global Lipschitz continuity with respect to the Wasserstein distance with quadratic cost denoted by $\mathcal{W}_2$. The diffusion term will be only state dependent and globally Lipschitz continuous. Our setting contrasts with the standard case with globally Lipschitz continuous coefficients,

which is well-studied in the literature, both from an analytic and numerical perspective, e.g., in [116], [144], and [24], respectively.

The study of SDEs and McKean–Vlasov equations with discontinuous drift is motivated by such models in biology (see, e.g., [60]) and financial mathematics (see, e.g., Atlas models in equity markets [13, 85] and dividend maximisation problems [145]). Further, in stochastic control a discontinuous control can lead to equations with discontinuous drift (see [142]). In the context of stochastic N-player games, non-smooth cost functions (such as the ℓ_1-regularisation) or constraints on the size of the control process can result in discontinuous controls (bang-bang type optimal controls) and hence will give controlled state dynamics with discontinuous drift, as in [28].

We start our literature review with some key references on standard SDEs with irregular and discontinuous drift, namely [154, 150, 105, 106], and then proceed to discuss some recent articles on McKean–Vlasov SDEs with non-Lipschitz drift. Zvonkin [154] (for one-dimensional SDEs) and Veretennikov [150] (for the multi-dimensional setting) prove the existence of a unique strong solution for an SDE where the drift is assumed to be measurable and bounded, but the diffusion coefficient σ needs to satisfy rather strong assumptions, namely that it is bounded and uniformly elliptic, i.e., there is a $\lambda > 0$ such that for all $x \in \mathbb{R}^d$ and all $v \in \mathbb{R}^d$, we have $v^\top \sigma(x)\sigma(x)^\top v \geq \lambda v^\top v$. An interesting addition to the aforementioned results in the case where the diffusion is not uniformly elliptic was established in the one-dimensional case in [105]. The authors assume the drift coefficient to be piecewise Lipschitz and σ to be globally Lipschitz with $\sigma(\eta) \neq 0$ for each of finitely many points of discontinuity η of the drift. This condition guarantees that the process does not spend a positive amount of time in the singularity. Under these assumptions, by explicitly constructing a transformation that removes the singularities, the existence of a unique strong solution can be proven and a numerical procedure for solving this class of SDEs can be constructed.

The main contribution of [106] is the extension of the one-dimensional case to the multi-dimensional setting under the assumption of piecewise Lipschitz continuity of the drift. In [106], the authors introduce a meaningful concept of piecewise Lipschitz continuity in higher dimensions, which is based on the notion of the so-called intrinsic metric. As already indicated by the one-dimensional case, there needs to be an intricate connection between the geometry of the set of discontinuities and the diffusion coefficient. We note that the exceptional set of singularities, denoted by Θ, is assumed to be a $\mathcal{C}^4$ hypersurface and for the diffusion part one requires the following: There exists a constant $C > 0$ such that $|\sigma(\xi)^\top n(\xi)| \geq C$ for all $\xi \in \Theta$, where $n(\xi)$ is orthogonal to the tangent space of Θ in ξ and $|n(\xi)| = 1$. Under these assumptions (and some additional technical conditions on the coefficients and on the geometry of Θ) the existence of a unique strong solution for multi-dimensional SDEs with piecewise Lipschitz continuous drift can be proven.

Moving on to McKean–Vlasov equations, the existence and uniqueness theory for strong solutions of such SDEs with coefficients of linear growth and Lipschitz type conditions (with respect

to the state and the measure component) is well-established (see, e.g., [144, 30]). More general existence/uniqueness results for weak and strong solutions of McKean–Vlasov SDEs can be found in [118, 99, 17]. The article [17] is concerned with the weak and strong existence/uniqueness of one-dimensional equations with additive noise, where the drift is assumed to be measurable, continuous in the measure component with respect to the Monge-Kantorovich metric and further satisfies a linear growth condition. In [118] and [99], a d-dimensional setting is considered, where the drift is assumed to be bounded, measureable (and possibly path-dependent) and Lipschitz continuous in the measure component with respect to the total variation distance. The diffusion is non-degenerate and independent of the measure. Under these assumptions (and some technical conditions) weak existence and uniqueness is proven.

For further recent existence and uniqueness results for strong and weak solutions of McKean–Vlasov SDEs, including results concerning standard Lipschitz assumptions on the coefficients, we refer to [144, 75, 109, 35, 44, 45, 137, 64, 80] and the references given therein.

The numerical analysis of SDEs with discontinuous drifts has received significant attention over the last few years, see, e.g., [105, 106, 107, 122, 120, 119, 73, 123, 42, 103] and the references therein for well-posedness results and for strong and weak convergence rates of numerical schemes.

In particular, in [105] (for the one-dimensional case) and in [106] (for the multi-dimensional setting) the standard strong convergence rate of order $1/2$ for a method derived from the Euler–Maruyama scheme was proven. However, the applicability of these schemes is limited as they require the explicit knowledge of a transformation (and its inverse) to map the SDE with discontinuous coefficients into one with Lipschitz continuous coefficients. In [107], an Euler–Maruyama scheme without the aforementioned transformation is introduced (in a multi-dimensional setting). While this scheme is easier to apply, the authors only show a strong convergence rate of order $1/4-\varepsilon$ for any $\varepsilon > 0$, imposing also the stronger assumption of boundedness for both coefficients of the underlying SDE. The central idea of [107] is to quantify the probability that a multi-dimensional process is in a small neighbourhood of the set of discontinuities, using an occupation time formula. In the one-dimensional case, with coefficients of linear growth, these techniques were refined in [119] and the expected strong convergence rate of order $1/2$ was recovered. Other recent works concerned with the numerical approximation of SDEs with discontinuous drifts include [122, 120], where a higher order scheme and an adaptive time-stepping scheme were introduced, respectively. In [103] a numerical scheme for classical one-dimensional diffusion processes generated by a differential operator involving discontinuous coefficients is presented. As the generator is non-local for McKean–Vlasov equations it seems a challenging problem to use these techniques in our framework.

The simulation of McKean–Vlasov SDEs typically involves two steps: First, at each time t, the

true measure $\mathcal{L}_{X_t}$ is approximated by the empirical measure

$$\mu_t^{\boldsymbol{X}^N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x),$$

where δ_x denotes the Dirac measure at point x and $(\boldsymbol{X}_t^N)_{t\in[0,T]} = (X_t^{1,N},\dots,X_t^{N,N})^\top_{t\in[0,T]}$, the interacting particle system, is the solution to the $\mathbb{R}^{dN}$-dimensional SDE with components

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

Here, $W^i = (W_t^i)_{t\in[0,T]}$ and ξ^i, for $i \in \{1,\dots,N\}$, are independent Brownian motions (also independent of W) and independent copies of ξ, respectively. In a next step, one needs to introduce a reasonable time-stepping method to discretise the particles $(X_t^{i,N})_{t\in[0,T]}$ over some finite time horizon $[0,T]$.

Numerical schemes for interacting particle systems with Hölder continuous coefficients (in the state variable) and with coefficients satisfying certain monotonicity (in the state variable) and Lipschitz assumptions (in the measure variable), can be found in [14] and [130, 15, 96] (and the references cited therein), respectively, where a strong convergence analysis is conducted. In [26, 9] a quantitative L_p-error analysis in terms of density and cumulative distribution function approximation is presented. The survey [24] discusses several examples and numerical schemes for McKean–Vlasov equations involving singular drifts, e.g., a probabilistic interpretation of the Burgers equation, see also [26], and of the 2D-incompressible Navier–Stokes equation (see e.g., [47, 117]) and turbulent flow models [128]. Other examples of McKean–Vlasov equations with singular drifts appear in the Keller–Segel equation [63], the Coulomb gas model [39], the Thomson problem [84], and the Stefan problem [88, 89].

Our numerical schemes present an original approximation method which, as of now, is restricted to the specific case of a one-dimensional spatial and one-point discontinuous drift component, but provides, again in this specific framework, a suitable alternative to the methodical mollification/cut-off approximation methods.

In this chapter, we first focus on the decomposable case, namely that

$$b(x,\mu) = b_1(x) + b_2(x,\mu),$$

where b_1 is piecewise Lipschitz with a discontinuity in zero, and b_2 satisfies the usual Lipschitz assumptions in both components. This structure allows us to present the main ideas of the analysis to be used later in a more general setting, in particular a transformation of the state variable to remove the discontinuity. In this setting, we prove well-posedness of the McKean–Vlasov equation and the associated particle system. This structure includes the important class of McKean–Vlasov equations of the form

$$\mathrm{d}X_t = \left(V(X_t) + \int_{\mathbb{R}} \beta(X_t - y)\,\mathcal{L}_{X_t}(\mathrm{d}y)\right)\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0 = \xi \in L_0^2(\mathbb{R}), \tag{4.2}$$

where V describes an external potential and β an interaction kernel; see, e.g., [78] and the references cited therein related to mean-field over-damped Langevin equations. These models also embed the class of self-stabilizing diffusions and the McKean–Vlasov model related to the granular media equation.

We then relax the structural assumption on decomposability slightly, but still have to require certain continuity of the measure derivatives at the points of discontinuity, which encompasses the above setting as a special case. The necessity for this condition arises due to the explicit measure or time dependence of the employed transformation. A future research direction concerns a setting where the point of discontinuity is time-dependent, or depends on the distribution of the process $(X_t)_{t\in[0,T]}$, which is relevant to study further practically important models, e.g., from [85].

Having established the existence of a unique strong solution with bounded moments, we propose two Euler–Maruyama schemes for the particle systems as numerical approximations to the McKean–Vlasov equations. For an Euler–Maruyama scheme applied to the SDE in the transformed state, strong convergence of order $1/2$ follows immediately, while for the direct application of the scheme without transformation, we are only able to show order $1/9$. Numerical tests indicate that this order is in general not sharp. We will discuss the reasons for this gap and possible improvements later.

The main contributions of the present chapter are as follows. First, we establish the well-posedness of McKean–Vlasov SDEs (with a certain form of a discontinuity) and of their associated particle systems. Techniques from variational calculus on the measure space $\mathcal{P}(\mathbb{R})$ equipped with the Wasserstein distance $\mathcal{W}_2$ will be essential in the proofs, due to the possible measure dependence of the transformation applied to the processes as described above. The second central contribution of the present chapter is the development of numerical schemes for approximating such McKean–Vlasov SDEs and their associated particle systems. Here, a non-standard strong convergence analysis based on occupation time estimates of the discretised processes in a neighbourhood of the discontinuity will be presented.

The remainder of this chapter is organised as follows: In Section 4.2, we collect all preliminary tools and notions needed throughout this chapter. The precise problem description and the main results are presented in Section 4.3. Then, Section 4.4 discusses fully implementable numerical schemes for McKean–Vlasov SDEs with discontinuous drift. We show strong convergence of certain orders with respect to the number of particles and time-steps, respectively. In Section 4.5, we apply our numerical scheme to a model problem arising in neuroscience [60] and to a slight modification of a mean-field game in systemic risk [72, 36].

4.2 Preliminaries

In the sequel, we will introduce several concepts and notions, which will be needed throughout this chapter. Also, we recall the transformation used to cope with drifts having discontinuities

in a given finite number of points and first developed in [105]. We give a summary of important properties of this mapping. Note that generic constants used in this chapter are denoted by $C > 0$. They are independent of the number of particles and number of time-steps, and might change their values from line to line.

4.2.1 Notions and notation

We start with introducing some notions and fixing the notation. Throughout this chapter, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ will denote a filtered probability space satisfying the usual assumptions and where $(\Omega, \mathcal{F}, \mathbb{P})$ is assumed to be atomless. $(\mathbb{R}^d, \langle\cdot,\cdot\rangle, |\cdot|)$ represents the d-dimensional ($d \geq 1$) Euclidean space. In addition, we use $\mathcal{P}(\mathbb{R})$ to denote the family of all probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-field over $\mathbb{R}$, and denote the subset of probability measures with finite second moment by $\mathcal{P}_2(\mathbb{R})$. As metric on the space $\mathcal{P}_2(\mathbb{R})$, we use the Wasserstein distance. As a matrix-norm, we will use $\|A\| := \sup_{v\in\mathbb{R}^d} |Av|$, for any $A \in \mathbb{R}^{d\times d}$.

We require a definition describing regularity properties of a function $f : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ in terms of the measure derivative (see [31, 41]).

Definition 4.2.1. Let $f : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ be a given functional.

- We say that f is an element of the class $\mathcal{C}_b^{(1,1)}$, if f is continuously L-differentiable, for any μ, there is a continuous version of the mapping $\mathbb{R} \ni y \mapsto \partial_\mu f(\mu)(y)$ and the derivatives

$$\partial_\mu f(\mu)(y), \quad \partial_y\{\partial_\mu f(\mu)(\cdot)\}(y),$$

  exist, are bounded and jointly continuous in the variables (μ, y) such that $y \in \mathrm{Supp}(\mu)$.

- We say that f is an element of the class $\mathcal{C}_b^{(2,1)}$, if it is an element of $\mathcal{C}_b^{(1,1)}$ and in addition the second order Lions derivative $\partial_\mu^2 f(\mu)(y, y')$ exists, is bounded and is again jointly continuous in the corresponding variables. Also, the joint continuity of all derivatives is here required globally, i.e., for all (μ, y, y').

We give the following additional remark, which links the L-derivative of functions of empirical measures to the standard partial derivatives of its empirical projections. For a functional $f : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$, we associate with it the finite dimensional projection $f^N : \mathbb{R}^N \to \mathbb{R}$ defined as

$$f^N(\boldsymbol{x}^N) := f\left(\frac{1}{N}\sum_{j=1}^N \delta_{x_j}\right),$$

for $\boldsymbol{x}^N := (x_1, \ldots, x_N)$. If $f \in \mathcal{C}_b^{(2,1)}$, then f^N is twice differentiable (in a classical sense) and

$$\partial_{x_i} f^N(\boldsymbol{x}^N) = \frac{1}{N}\partial_\mu f\left(\frac{1}{N}\sum_{j=1}^N \delta_{x_j}\right)(x_i),$$

$$\partial_{x_i}\partial_{x_k} f^N(\boldsymbol{x}^N) = \frac{1}{N}\partial_y\partial_\mu f\left(\frac{1}{N}\sum_{j=1}^N \delta_{x_j}\right)(x_i)\delta_{i,k} + \frac{1}{N^2}\partial_\mu^2 f\left(\frac{1}{N}\sum_{j=1}^N \delta_{x_j}\right)(x_i, x_k),$$

where $\delta_{i,k}$ is the Kronecker delta, see, e.g., [31, Proposition 5.35].

4.2.2 Properties of the transformation G

In [105], the authors consider one-dimensional SDEs of the form

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0 = x \in \mathbb{R},$$

with a piecewise Lipschitz continuous drift coefficient b that is discontinuous in $K \in \mathbb{N}$ points $\eta_1, \dots, \eta_K$, and a Lipschitz diffusion coefficient σ that does not vanish in any η_k. A mapping $G : \mathbb{R} \to \mathbb{R}$ is defined to transform the SDE into one for $Z = G(X)$ with globally Lipschitz continuous coefficients. For simplicity, we restrict the discussion to $K = 1$ with $\eta_1 = 0$. We define the mapping G by

$$G(x) := x + \alpha x|x|\phi\left(\frac{x}{c}\right),$$

where

$$\phi(x) := \begin{cases} (1-x^2)^3, & |x| \le 1, \\ 0, & |x| > 1 \end{cases} \qquad \alpha := \frac{b(0^-) - b(0^+)}{2\sigma^2(0)},$$

and c is a constant satisfying $0 < c < 1/|\alpha|$. The choice of α yields a Lipschitz continuous drift coefficient for the SDE of $Z = G(X)$, in particular, it removes the discontinuity in 0 from the drift. The restriction on c guarantees that G possesses a global inverse.

It is known from [106] that G satisfies the following properties:

- G is $\mathcal{C}^1(\mathbb{R}, \mathbb{R})$ with $0 < \inf_{x\in\mathbb{R}} G'(x) \le \sup_{x\in\mathbb{R}} G'(x) < \infty$. Therefore, G is Lipschitz continuous and has an inverse $G^{-1} : \mathbb{R} \to \mathbb{R}$ that is Lipschitz continuous as well.
- The derivative G' is Lipschitz continuous (i.e., also absolutely continuous). In addition, G' has a bounded Lebesgue density $G'' : \mathbb{R} \to \mathbb{R}$, which is Lipschitz continuous on each of the subintervals $(-\infty, 0)$ and $(0, \infty)$. Also, Itô's formula can still be applied to G and G^{-1}.

4.3 Existence and uniqueness results

The following subsections are devoted to proving well-posedness results for certain classes of one-dimensional McKean–Vlasov SDEs with a drift having a discontinuity in zero. In a first step, we study a simple class where the resulting transformation will not depend on the measure. Here, the transformation techniques developed in [105] will allow us to prove existence and uniqueness

of a strong solution. The second class of McKean–Vlasov SDEs investigated below has the intrinsic difficulty that the required transformation will depend on the measure (i.e., will be time-dependent). Hence, a fixed-point iteration in the measure component will be required and we need to use techniques from variational calculus on the measure space $\mathcal{P}_2(\mathbb{R})$, in particular an Itô formula for functionals acting on this space.

For each of these classes of McKean–Vlasov SDEs, we will additionally study the well-posedness of their associated interacting particle system. Although they can be considered as N-dimensional classical SDEs, with N denoting the number of particles, the resulting set of discontinuities of the N-dimensional drift cannot be handled by the main results of [106].

Future work is needed to extend the methods developed in this chapter to a multi-dimensional framework. In particular, it seems that the decomposable case can be generalised when discontinuities of the form discussed in [106] are considered.

4.3.1 McKean–Vlasov SDEs and interacting particle systems with decomposable drift

For a given terminal time $T > 0$ and given $p \geq 2$, we consider a one-dimensional McKean–Vlasov SDE of the form

$$\mathrm{d}X_t = b(X_t, \mathcal{L}_{X_t})\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \qquad X_0 = \xi \in L_0^p(\mathbb{R}), \tag{4.3}$$

where $b : \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are measurable functions.

In the following, we state the model assumptions which will specify the set-up for this subsection:

H.7. (1) We have $\sigma(0) \neq 0$ and there exists a constant $L > 0$ such that

$$|\sigma(x) - \sigma(x')| \leq L|x - x'| \quad \forall x, x' \in \mathbb{R}.$$

(2) The drift is decomposable in the following sense:

$$b(x, \mu) = b_1(x) + b_2(x, \mu) \quad \forall x \in \mathbb{R},\ \mu \in \mathcal{P}_2(\mathbb{R}),$$

where $b_1 : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous on the subintervals $(-\infty, 0)$ and $(0, \infty)$ and there exists a constant $L_1 > 0$ such that

$$|b_2(x, \mu) - b_2(x', \nu)| \leq L_1 \left(|x - x'| + \mathcal{W}_2(\mu, \nu)\right) \quad \forall x, x' \in \mathbb{R},\ \mu, \nu \in \mathcal{P}_2(\mathbb{R}).$$

Proposition 4.3.1. *Let Assumption (H.7) be satisfied, let $\xi \in L_0^p(\mathbb{R})$ for a given $p \geq 2$ and assume $c < 1/|\alpha|$. Then, the McKean–Vlasov SDE defined in (4.3) has a unique strong solution in $\mathcal{S}^p([0, T])$.*

Proof. We will first transform the McKean–Vlasov SDE (4.3), show well-posedness of the transformed McKean–Vlasov SDE and then apply Itô's formula to G^{-1} in order to recover a strong

solution to (4.3). To transform the SDE, we will employ the transformation $G : \mathbb{R} \to \mathbb{R}$ defined in Section 4.2.2, with

$$\alpha = \frac{b(0^-, \mu) - b(0^+, \mu)}{2\sigma^2(0)} = \frac{b_1(0^-) - b_1(0^+)}{2\sigma^2(0)},$$

which eliminates the discontinuity in zero. Note that α does not depend on the measure. Moreover, G has a global inverse due to the choice $c < 1/|\alpha|$ (see [106, Lemma 2.2]), and Itô's formula can be applied to G^{-1}.

We define $\bar{\phi}(x) := x|x|\phi(x/c)$ and transform the McKean–Vlasov SDE for X to

$$\begin{aligned} \mathrm{d}G(X_t) = &\left(b(X_t, \mathcal{L}_{X_t}) + \alpha\bar{\phi}'(X_t) b(X_t, \mathcal{L}_{X_t}) + \frac{1}{2}\alpha\bar{\phi}''(X_t)\sigma^2(X_t) \right) \mathrm{d}t \\ &+ \left(\sigma(X_t) + \alpha\bar{\phi}'(X_t)\sigma(X_t) \right) \mathrm{d}W_t. \end{aligned}$$

Itô's formula is applicable due to [87, Problem 7.3]. This allows us to introduce the McKean–Vlasov SDE

$$\mathrm{d}Z_t = \tilde{b}(Z_t, \mathcal{L}_{G^{-1}(Z_t)})\,\mathrm{d}t + \tilde{\sigma}(Z_t)\,\mathrm{d}W_t,$$

where $Z_0 = G(\xi)$ and

$$\begin{aligned} \tilde{b}(z, \mu) &:= b(G^{-1}(z), \mu) + \alpha\bar{\phi}'(G^{-1}(z)) b(G^{-1}(z), \mu) + \frac{1}{2}\alpha\bar{\phi}''(G^{-1}(z))\sigma^2(G^{-1}(z)), \\ &= b_1(G^{-1}(z)) + b_2(G^{-1}(z), \mu) + \alpha\bar{\phi}'(G^{-1}(z)) b(G^{-1}(z), \mu) + \frac{1}{2}\alpha\bar{\phi}''(G^{-1}(z))\sigma^2(G^{-1}(z)) \\ \tilde{\sigma}(z) &:= \sigma(G^{-1}(z)) + \alpha\bar{\phi}'(G^{-1}(z))\sigma(G^{-1}(z)). \end{aligned}$$

That the transformed process $Z = (Z_t)_{t\in[0,T]}$ has global Lipschitz continuous coefficients in the state variable is a consequence of [105, Theorem 2.5], the decomposable structure of b and the Lipschitz continuity of G^{-1}. In particular, note that

$$b_1(0^+) + \frac{1}{2}\alpha\bar{\phi}''(0^+)\sigma^2(0) = b_1(0^-) + \frac{1}{2}\alpha\bar{\phi}''(0^-)\sigma^2(0),$$

which makes the mapping $\mathbb{R} \ni x \mapsto b_1(x) + \frac{1}{2}\alpha\bar{\phi}''(x)\sigma^2(x)$ continuous.

The Lipschitz continuity of $\mathcal{P}_2(\mathbb{R}) \ni \mu \mapsto \tilde{b}(z, \mu)$ immediately follows from (H.7(2)). Hence, we deduce that $(Z_t)_{t\in[0,T]}$ has a unique strong solution (see, e.g., [144]). Consequently, setting $X_t = G^{-1}(Z_t)$, for $t \in [0, T]$ and applying Itô's formula to G^{-1} concludes the claim. The linear growth of the coefficients in the state and measure component along with Gronwall's inequality imply that $\mathbb{E}\left[\sup_{t\in[0,T]} |X_t|^p\right] < \infty$. □

The interacting particles $(X_t^{i,N})_{t\in[0,T]}$, $i \in \{1, \dots, N\}$, associated with (4.3) satisfy

$$\mathrm{d}X_t^{i,N} = b_1(X_t^{i,N})\,\mathrm{d}t + b_2(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N})\,\mathrm{d}t + \sigma(X_t^{i,N})\,\mathrm{d}W_t^i, \tag{4.4}$$

where (ξ^i, W^i), for $i \in \{1, \dots, N\}$, are independent copies of (ξ, W).

Proposition 4.3.2. *Let Assumption (H.7) be satisfied, let $\xi \in L_0^p(\mathbb{R})$ for a given $p \geq 2$ and assume $c < 1/|\alpha|$. Then, the interacting particle system defined in (4.4) has a unique strong solution in $\mathcal{S}^p([0,T])$.*

Proof. In contrast to the setup in [106], the set of discontinuities, denoted by Θ, is not a differentiable manifold, but has the form

$$\Theta = \{(x_1, \dots, x_N)^\top \in \mathbb{R}^N \colon \ \exists j \in \{1, \dots, N\} \colon x_j = 0\}.$$

However, we may define $\boldsymbol{G}_N : \mathbb{R}^N \to \mathbb{R}^N$ by

$$\boldsymbol{G}_N(\boldsymbol{x}^N) := (G(x_1), \dots, G(x_N))^\top,$$

where G is as in Proposition 4.3.1 and $\boldsymbol{x}^N = (x_1, \dots, x_N)^\top$, which allows us to transform the particle system $(\boldsymbol{X}_t^N)_{t\in[0,T]} = (X_t^{1,N}, \dots, X_t^{N,N})_{t\in[0,T]}^\top$ into a new particle system

$$(\boldsymbol{Z}_t^N)_{t\in[0,T]} = (Z_t^{1,N}, \dots, Z_t^{N,N})_{t\in[0,T]}^\top,$$

using

$$\begin{aligned} \mathrm{d}G(X_t^{i,N}) = &\left(b(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N}) + \alpha\bar{\phi}'(X_t^{i,N}) b(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N}) + \frac{1}{2}\alpha\bar{\phi}''(X_t^{i,N})\sigma^2(X_t^{i,N}) \right) \mathrm{d}t \\ &+ \left(\sigma(X_t^{i,N}) + \alpha\bar{\phi}'(X_t^{i,N})\sigma(X_t^{i,N}) \right) \mathrm{d}W_t^i. \end{aligned}$$

This in turn enables us to define

$$\mathrm{d}Z_t^{i,N} = \tilde{b}(Z_t^{i,N}, \mu_t^{\boldsymbol{Z}^N})\,\mathrm{d}t + \tilde{\sigma}(Z_t^{i,N})\,\mathrm{d}W_t^i,$$

with $Z_0^{i,N} = G(X_0^{i,N})$, where we introduced

$$\mu_t^{\boldsymbol{Z}^N}(\mathrm{d}z) := \frac{1}{N}\sum_{j=1}^N \delta_{G^{-1}(Z_t^{j,N})}(\mathrm{d}z),$$

and

$$\begin{aligned} \tilde{b}(z_i, \mu^{\boldsymbol{z}^N,N}) &:= b(G^{-1}(z_i), \mu^{\boldsymbol{z}^N}) + \alpha\bar{\phi}'(G^{-1}(z_i)) b(G^{-1}(z_i), \mu^{\boldsymbol{z}^N}) + \frac{1}{2}\alpha\bar{\phi}''(G^{-1}(z_i))\sigma^2(G^{-1}(z_i)), \\ \tilde{\sigma}(z_i) &:= \sigma(G^{-1}(z_i)) + \alpha\bar{\phi}'(G^{-1}(z_i))\sigma(G^{-1}(z_i)). \end{aligned}$$

In what follows, we define $\boldsymbol{B}_N : \mathbb{R}^N \to \mathbb{R}^N$ and $\boldsymbol{\Sigma}_N : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ by

$$\boldsymbol{B}_N(\boldsymbol{z}) = (\tilde{b}(z_1, \mu^{\boldsymbol{z}^N}), \dots, \tilde{b}(z_N, \mu^{\boldsymbol{z}^N}))^\top, \quad \boldsymbol{\Sigma}_N(\boldsymbol{z}) = \mathrm{diag}_{N\times N}(\tilde{\sigma}(z_1), \dots, \tilde{\sigma}(z_N)).$$

Note that, due to (H.7) and the choice of α, the drift $\boldsymbol{B}_N$ and the diffusion $\boldsymbol{\Sigma}_N$ are Lipschitz continuous in $\boldsymbol{z}$. To see this, observe that for any $\boldsymbol{x}^N, \boldsymbol{y}^N \in \mathbb{R}^N$, we have for some $C > 0$ (independent of N)

$$|\boldsymbol{B}_N(\boldsymbol{x}^N) - \boldsymbol{B}_N(\boldsymbol{y}^N)|^2 = \sum_{i=1}^N |\tilde{b}(x_i, \mu^{\boldsymbol{x}^N}) - \tilde{b}(y_i, \mu^{\boldsymbol{y}^N})|^2$$

$$\leq C \sum_{i=1}^{N} \left(|x_i - y_i|^2 + \mathcal{W}_2^2(\mu^{\boldsymbol{x}^N}, \mu^{\boldsymbol{y}^N}) \right) \leq C|\boldsymbol{x}^N - \boldsymbol{y}^N|^2,$$

where we used (H.7(2)) and the fact that $\mathbb{R} \ni z \mapsto \tilde{b}(z, \mu)$ is Lipschitz continuous (see [105, Theorem 2.5]). Further, we employed the standard bound for the Wasserstein distance between two empirical measures

$$\mathcal{W}_2^2 \left(\mu^{\boldsymbol{x}^N}, \mu^{\boldsymbol{y}^N} \right) \leq \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|^2.$$

The same estimate can be derived for $\boldsymbol{\Sigma}_N$ by employing (H.7(1)).

Hence, the transformed particle system has a unique strong solution. It remains to show that $\boldsymbol{G}_N$ has a global inverse. This, however, is clear as the mapping G has a global inverse, due to the choice of c (see Section 4.2.2). Therefore, Itô's formula applied to the inverse of G gives the claim. □

4.3.2 McKean–Vlasov SDE with non-decomposable drift

Here, we consider again a one-dimensional McKean–Vlasov SDE of the form (4.3),

$$\mathrm{d}X_t = b(X_t, \mathcal{L}_{X_t})\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \qquad X_0 = \xi \in L_0^p(\mathbb{R}). \tag{4.5}$$

However, in contrast to the above setting, we will not assume that b can be decomposed in two parts as in Assumption (H.7(2)) from the previous section, and therefore the transformation will also depend on the measure. To be precise, for any $(x, \mu) \in \mathbb{R} \times \mathcal{P}_2(\mathbb{R})$, we define

$$G(x, \mu) := x + \alpha(\mu) x |x| \phi\left(\frac{x}{c}\right), \tag{4.6}$$

where

$$\phi(x) := \begin{cases} (1 - x^2)^3, & |x| \leq 1, \\ 0, & |x| > 1 \end{cases} \qquad \alpha(\mu) := \frac{b(0^-, \mu) - b(0^+, \mu)}{2\sigma^2(0)}, \tag{4.7}$$

and $c > 0$ is a constant small enough to guarantee the invertibility of G. When we speak of an 'inverse' of $G(x, \mu)$, we mean 'inverse with respect to x', i.e., the inverse is a function $G^{-1} \colon \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ which satisfies $G^{-1}\big(G(x, \mu), \mu\big) = x$ for all $(x, \mu) \in \mathbb{R} \times \mathcal{P}_2(\mathbb{R})$ and $G\big(G^{-1}(z, \mu), \mu\big) = z$ for all $(z, \mu) \in \mathbb{R} \times \mathcal{P}_2(\mathbb{R})$. For a given flow of measures $(\mu_t)_{t \in [0,T]} \in \mathcal{C}([0, T], \mathcal{P}_2(\mathbb{R}))$, G may also be viewed as a mapping $G \colon \mathbb{R} \times [0, T] \to \mathbb{R}$.

In the following, we state the model assumptions which will specify the set-up for this subsection:

H.8. Assumption (H.7(1)) is satisfied and we require:

(1) There exist constants $L, L_1 > 0$ such that

$$\sup_{x \neq 0} \frac{|b(x, \mu)|}{1 + |x|} \leq L, \quad |b(x, \mu) - b(x, \nu)| \leq L_1 \mathcal{W}_2(\mu, \nu) \quad \forall x \neq 0 \in \mathbb{R},\ \mu, \nu \in \mathcal{P}_2(\mathbb{R}).$$

Additionally, for all $\mu \in \mathcal{P}_2(\mathbb{R})$, $x \mapsto b(x, \mu)$ is piecewise Lipschitz continuous on the subintervals $(-\infty, 0)$ and $(0, \infty)$, uniformly with respect to μ.

(2) $\alpha \in \mathcal{C}_b^{(1,1)}$, and the mapping $\mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \ni (\mu, y) \mapsto \partial_y \partial_\mu \alpha(\mu)(y)$ is Lipschitz continuous, that is, there exists a constant $L_2 > 0$ such that

$$|\partial_y \partial_\mu \alpha(\mu)(y) - \partial_y \partial_\mu \alpha(\nu)(y')| \leq L_2 \left(|y - y'| + \mathcal{W}_2(\mu, \nu)\right) \quad \forall y, y' \in \mathbb{R},\ \mu, \nu \in \mathcal{P}_2(\mathbb{R}).$$

(3) For any $\mu \in \mathcal{P}_2(\mathbb{R})$, the mapping $\mathbb{R} \ni y \mapsto \partial_\mu \alpha(\mu)(y)$ vanishes in zero, i.e.,

$$\partial_\mu \alpha(\mu)(0) = \partial_\mu b(0^-, \mu)(0) - \partial_\mu b(0^+, \mu)(0) = 0. \tag{4.8}$$

(4) The mapping $\mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \ni (\mu, y) \mapsto \partial_\mu \alpha(\mu)(y) b(y, \mu)$ is Lipschitz continuous.

Remark 4.3.1. The requirement in (H.8(2)) that $\alpha \in \mathcal{C}_b^{(1,1)}$ is needed to apply an Itô formula for α (see [31, Proposition 5.102]).

Remark 4.3.2. Note that (H.8(4)) could also be replaced by the following alternative set of assumptions: On each of the two subintervals $(-\infty, 0)$ and $(0, \infty)$, $x \mapsto b(x, \mu)$ is a $\mathcal{C}^1(\mathbb{R}, \mathbb{R})$ function with bounded derivative and, additionally, for any $\mu \in \mathcal{P}_2(\mathbb{R})$, the mapping $\mathbb{R} \ni y \mapsto \partial_y \partial_\mu \alpha(\mu)(y)$ vanishes in zero, i.e.,

$$\partial_y \partial_\mu \alpha(\mu)(0) = \partial_y \partial_\mu b(0^-, \mu)(0) - \partial_y \partial_\mu b(0^+, \mu)(0) = 0.$$

The following proposition shows the Lipschitz continuity of the mapping $\mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \ni (x, \mu) \to G^{-1}(x, \mu)$.

Proposition 4.3.3. *Let the function G be defined as in (4.6) with $c < 1/\sup_{\mu \in \mathcal{P}_2(\mathbb{R})} |\alpha(\mu)|$ and (4.7) and let Assumption (H.8(1)) be satisfied. Then, there exists a constant $L_{G^{-1}} > 0$ such that for any $x, y \in \mathbb{R}$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$*

$$|G^{-1}(x, \mu) - G^{-1}(y, \nu)| \leq L_{G^{-1}} \left(|x - y| + \mathcal{W}_2(\mu, \nu)\right).$$

Proof. First, we note that by differentiating (4.6), we get for $x \in [-c, c]$

$$\begin{aligned} \partial_x G(x, \mu) &= 1 - \frac{6\alpha(\mu)}{c^2} |x| x^2 \left(1 - (x/c)^2\right)^2 + 2\alpha(\mu)|x|(1 - (x/c)^2)^3 \\ &= 1 - 2c\alpha(\mu) \frac{|x|}{c} \left(4x^2/c^2 - 1\right) \left(1 - (x/c)^2\right)^2. \end{aligned}$$

It is easy to verify that

$$\sup_{x \in [0,c]} \left| \frac{|x|}{c} \left(4x^2/c^2 - 1\right) \left(1 - (x/c)^2\right)^2 \right| = \sup_{z \in [0,1]} \left| |z| \left(4z^2 - 1\right) \left(1 - z^2\right)^2 \right| < \frac{1}{4},$$

which implies that for all $x \in \mathbb{R}$ and $\mu \in \mathcal{P}_2(\mathbb{R})$

$$\left|\partial_x G(x, \mu) - 1\right| < c \frac{|\sup_{\mu \in \mathcal{P}_2(\mathbb{R})} \alpha(\mu)|}{2}.$$

In particular, if $c < 1/\sup_{\mu\in\mathcal{P}_2(\mathbb{R})}|\alpha(\mu)|$, then for all $x, y \in \mathbb{R}$ and $\mu \in \mathcal{P}_2(\mathbb{R})$, we have

$$\partial_x G(x,\mu) > \frac{1}{2}, \quad \left|G^{-1}(x,\mu) - G^{-1}(y,\mu)\right| < 2|x-y|\,.$$

That $\mu \mapsto \partial_x G(x,\mu)$ is Lipschitz continuous is a consequence of (H.8(1)). It is easy to show that the mapping $x \mapsto \partial_x G(x,\mu)$ is also Lipschitz continuous. Denote the Lipschitz constant of $\partial_x G$ with respect to the first and second argument by C_1 and C_2, respectively. Writing

$$G^{-1}(x,\mu) = \int_0^x \frac{1}{\partial_x G(G^{-1}(y,\mu),\mu)}\,\mathrm{d}y,$$

we obtain

$$\begin{aligned}
&|G^{-1}(x,\mu) - G^{-1}(x,\nu)| \\
&\le \int_0^x \left|\frac{1}{\partial_x G(G^{-1}(y,\mu),\mu)} - \frac{1}{\partial_x G(G^{-1}(y,\nu),\mu)}\right|\mathrm{d}y + \int_0^x \left|\frac{1}{\partial_x G(G^{-1}(y,\nu),\mu)} - \frac{1}{\partial_x G(G^{-1}(y,\nu),\nu)}\right|\mathrm{d}y \\
&\le \int_0^x \left|\frac{\partial_x G(G^{-1}(y,\nu),\mu) - \partial_x G(G^{-1}(y,\mu),\mu)}{\partial_x G(G^{-1}(y,\mu),\mu)\partial_x G(G^{-1}(y,\nu),\mu)}\right|\mathrm{d}y + \int_0^x \left|\frac{\partial_x G(G^{-1}(y,\nu),\nu) - \partial_x G(G^{-1}(y,\nu),\mu)}{\partial_x G(G^{-1}(y,\nu),\mu)\partial_x G(G^{-1}(y,\nu),\nu)}\right|\mathrm{d}y \\
&\le \int_0^x 4C_1\left|G^{-1}(y,\nu) - G^{-1}(y,\mu)\right|\mathrm{d}y + \int_0^x 4C_2\mathcal{W}_2(\mu,\nu)\,\mathrm{d}y \\
&\le 4C_1\int_0^x \left|G^{-1}(y,\nu) - G^{-1}(y,\mu)\right|\mathrm{d}y + 4C_2 x\mathcal{W}_2(\mu,\nu)\,.
\end{aligned}$$

For $|x| < c$, we have

$$\begin{aligned}
|G^{-1}(x,\mu) - G^{-1}(x,\nu)| &\le 4C_1\int_0^x \left|G^{-1}(y,\nu) - G^{-1}(y,\mu)\right|\mathrm{d}y + 4C_2 x\mathcal{W}_2(\mu,\nu) \\
&\le 4C_1\int_0^x \left|G^{-1}(y,\nu) - G^{-1}(y,\mu)\right|\mathrm{d}y + 4C_2 c\mathcal{W}_2(\mu,\nu)\,,
\end{aligned}$$

and hence Gronwall's inequality implies

$$|G^{-1}(x,\mu) - G^{-1}(x,\nu)| \le 4C_2 c\mathcal{W}_2(\mu,\nu)e^{4C_1 x} \le 4C_2 c e^{4C_1 c}\mathcal{W}_2(\mu,\nu)\,.$$

For $|x| \ge c$, $|G^{-1}(x,\mu) - G^{-1}(x,\nu)| = 0 \le 4C_2 c e^{4C_1 c}\mathcal{W}_2(\mu,\nu)$ by the definition of G. We finally obtain, for all $x, y \in \mathbb{R}$ and all $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$, with $C_3 := 4C_2 c e^{4C_1 c}$,

$$\begin{aligned}
|G^{-1}(x,\mu) - G^{-1}(y,\nu)| &\le |G^{-1}(x,\mu) - G^{-1}(x,\nu)| + |G^{-1}(x,\nu) - G^{-1}(y,\nu)| \\
&\le C_3\mathcal{W}_2(\mu,\nu) + 2|x-y| \le \max(C_3, 2)(\mathcal{W}_2(\mu,\nu) + |x-y|)\,.
\end{aligned}$$

□

Remark 4.3.3. In what follows, we will assume that $c < 1/\sup_{\mu\in\mathcal{P}_2(\mathbb{R})}|\alpha(\mu)|$ and is small enough such that the Lipschitz constant of the mapping $\mathcal{P}_2(\mathbb{R}) \ni \mu \mapsto G^{-1}(x,\mu)$ (i.e., C_3 a few lines above) is less than a half. The reason for this requirement will become clearer in the proof of Theorem 4.3.1.

Similar to the previous section, we aim to recover a unique strong solution of (4.5) by setting $X_t = G^{-1}(Z_t^\mu, \mu_t)$, where $\mu_t = \mathcal{L}_{X_t}$ for $t \in [0,T]$, and $(Z_t^\mu)_{t\in[0,T]}$ is the process obtained by applying the transformation G to X. Even though G is not twice continuously differentiable in the state variable, Itô's formula is still applicable due to the special form of the discontinuity (see [57, Theorem 2.1] and the comments after the proof of this theorem). Now, observe that

$$\begin{aligned}\mathrm{d}G(X_t,\mu_t) = &\left(\partial_t G(X_t,\mu_t) + b(X_t,\mu_t) + \alpha(\mu_t)\bar\phi'(X_t)b(X_t,\mu_t) + \frac{1}{2}\alpha(\mu_t)\bar\phi''(X_t)\sigma^2(X_t)\right)\mathrm{d}t\\ &+ (\sigma(X_t) + \alpha(\mu_t)\bar\phi'(X_t)\sigma(X_t))\,\mathrm{d}W_t.\end{aligned}$$

Itô's formula along a flow of measures $(\mu_t)_{t\in[0,T]} \in \mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}))$ (see, e.g., [31, Proposition 5.102]) implies

$$\partial_t G(x,\mu_t) = \int_{\mathbb{R}} \left(b(y,\mu_t)\partial_\mu G(x,\mu_t)(y) + \frac{\sigma^2(y)}{2}\partial_y\partial_\mu G(x,\mu_t)(y)\right)\mu_t(\mathrm{d}y) =: \mathcal{L}_{\mu_t}(G(x,\cdot))(\mu_t),$$

where we recall that $\partial_y\partial_\mu G(x,\mu_t)(y)$ denotes the derivative of the mapping $\mathbb{R} \ni y \mapsto \partial_\mu G(x,\mu_t)(y)$ and

$$\partial_\mu G(x,\mu_t)(y) = \partial_\mu\alpha(\mu_t)(y)|x|x\phi(x/c), \quad \partial_y\partial_\mu G(x,\mu_t)(y) = \partial_y\partial_\mu\alpha(\mu_t)(y)|x|x\phi(x/c).$$

Hence, we define

$$\mathrm{d}Z_t^\mu := \tilde b(Z_t^\mu,\mu_t)\,\mathrm{d}t + \tilde\sigma(Z_t^\mu)\,\mathrm{d}W_t, \quad Z_0^\mu = G(\xi,\delta_\xi), \tag{4.9}$$

where

$$\begin{aligned}\tilde b(z,\mu) &:= \mathcal{L}_\mu(G(G^{-1}(z,\mu),\cdot)(\mu) + b(G^{-1}(z,\mu),\mu) + \alpha(\mu)\bar\phi'(G^{-1}(z,\mu))b(G^{-1}(z,\mu),\mu)\\ &\quad + \frac{1}{2}\alpha(\mu)\bar\phi''(G^{-1}(z,\mu))\sigma^2(G^{-1}(z,\mu)),\\ \tilde\sigma(z,\mu) &:= \sigma(G^{-1}(z,\mu)) + \alpha(\mu)\bar\phi'(G^{-1}(z,\mu))\sigma(G^{-1}(z,\mu)).\end{aligned} \tag{4.10}$$

In the following, we will show that the decoupled SDE (4.9) (i.e., the flow $(\mu_t)_{t\in[0,T]} \in \mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}))$ is fixed) has Lipschitz continuous coefficients. Note that for a such a flow of measures the process X^μ in (4.5) (interpreted as classical SDE) has bounded moments uniformly in $(\mu_t)_{t\in[0,T]}$, which is a consequence of (H.7(1)) and (H.8(1)). In particular, for $\xi \in L_2^0(\mathbb{R})$, we have the a-priori estimate $\mathbb{E}\left[\sup_{0\le t\le T}|X_t^\mu|^2\right] \le C(1+\mathbb{E}[|\xi|^2]) =: \bar C$, where $C > 0$ only depends on T and the constants appearing in the model assumptions. We will introduce the following subspace of $\mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}))$: We define $\mathcal{P}^b := \{\mu \in \mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R})) : \sup_{t\in[0,T]}\int_{\mathbb{R}} x^2\,\mu_t(\mathrm{d}x) \le \bar C\}$, where $\bar C$ is defined as above.

Lemma 4.3.1. *Let Assumption (H.8) be satisfied and assume $c < 1/\sup_{\mu\in\mathcal{P}_2(\mathbb{R})}|\alpha(\mu)|$. Then, $\tilde b$ and $\tilde\sigma$ given in (4.10) are Lipschitz continuous on $\mathbb{R}\times\mathcal{P}^b$.*

Proof. Using (H.7(1)), the Lipschitz continuity of $z \mapsto G^{-1}(z, \nu_t)$ and the uniform boundedness of α, see (H.8(1)), in combination with [106, Lemma 2.5], gives the Lipschitz continuity of $z \mapsto \tilde{\sigma}(z, \nu_t)$, with a Lipschitz constant independent of ν_t. Similarly, we can deduce that $\nu_t \mapsto \tilde{\sigma}(z, \nu_t)$ is Lipschitz continuous, due to Proposition 4.3.3 and the Lipschitz continuity of $\nu_t \mapsto \alpha(\nu_t)$.

The choice of α along with (H.8(1)) guarantees that the mapping

$$z \mapsto b(G^{-1}(z, \nu_t), \nu_t) + \frac{1}{2}\alpha(\nu_t)\bar{\phi}''(G^{-1}(z, \nu_t))\sigma^2(G^{-1}(z, \nu_t)),$$

is Lipschitz continuous. From [106, Lemma 2.4], we can deduce the Lipschitz continuity of

$$z \mapsto \alpha(\nu_t)\bar{\phi}'(G^{-1}(z, \nu_t))b(G^{-1}(z, \nu_t), \nu_t).$$

That

$$\nu_t \mapsto b(G^{-1}(z, \nu_t), \nu_t) + \frac{1}{2}\alpha(\nu_t)\bar{\phi}''(G^{-1}(z, \nu_t))\sigma^2(G^{-1}(z, \nu_t)), \quad \nu_t \mapsto \alpha(\nu_t)\bar{\phi}'(G^{-1}(z, \nu_t))b(G^{-1}(z, \nu_t), \nu_t),$$

are Lipschitz continuous is a consequence of (H.7(1)) and (H.8(1)), Proposition 4.3.3 and the fact that $G^{-1}(z, \mu^{(1)})$ and $G^{-1}(z, \mu^{(2)})$, for $\mu^{(1)}, \mu^{(2)} \in \mathcal{P}_2(\mathbb{R})$, have the same sign. Also note that $\bar{\phi}'$ and $\bar{\phi}''$ are (piecewise) Lipschitz continuous and bounded. It remains to analyse the Lipschitz continuity of

$$(z, \nu_t) \mapsto \int_{\mathbb{R}} \left(b(y, \nu_t)\partial_\mu G(G^{-1}(z, \nu_t), \nu_t)(y) + \frac{\sigma^2(y)}{2}\partial_y\partial_\mu G(G^{-1}(z, \nu_t), \nu_t)(y) \right) \nu_t(\mathrm{d}y). \tag{4.11}$$

Assumptions (H.8(2)) and (H.8(3)) guarantee that above mapping exists and further that the mapping $y \mapsto b(y, \nu_t)\partial_\mu G(G^{-1}(z, \nu_t), \nu_t)(y)$ is continuous in zero. We start by analysing the Lipschitz continuity of (4.11) with respect to the measure variable. Consider now an arbitrary coupling $\Pi_t(\cdot, \cdot)$ between $\nu_t(\cdot)$ and $\mu_t(\cdot)$, for $(\mu_t)_{t\in[0,T]}, (\nu_t)_{t\in[0,T]} \in \mathcal{P}^b$, and estimate

$$\begin{aligned}
&\int_{\mathbb{R}^2} \left(b(y, \nu_t)\partial_\mu G(G^{-1}(z, \nu_t), \nu_t)(y) - b(x, \mu_t)\partial_\mu G(G^{-1}(z, \mu_t), \mu_t)(x) \right) \Pi_t(\mathrm{d}y, \mathrm{d}x) \\
&= \int_{\mathbb{R}^2} \left(b(y, \nu_t)\partial_\mu G(G^{-1}(z, \nu_t), \nu_t)(y) - b(x, \mu_t)\partial_\mu G(G^{-1}(z, \nu_t), \mu_t)(x) \right) \Pi_t(\mathrm{d}y, \mathrm{d}x) \\
&\quad + \int_{\mathbb{R}^2} \left(b(x, \mu_t)\partial_\mu G(G^{-1}(z, \nu_t), \mu_t)(x) - b(x, \mu_t)\partial_\mu G(G^{-1}(z, \mu_t), \mu_t)(x) \right) \Pi_t(\mathrm{d}y, \mathrm{d}x).
\end{aligned}$$

Note that

$$\begin{aligned}
&\int_{\mathbb{R}^2} \left| b(y, \nu_t)\partial_\mu G(G^{-1}(z, \nu_t), \nu_t)(y) - b(x, \mu_t)\partial_\mu G(G^{-1}(z, \nu_t), \mu_t)(x) \right| \Pi_t(\mathrm{d}y, \mathrm{d}x) \\
&= \int_{\mathbb{R}^2} \left| b(y, \nu_t)\partial_\mu \alpha(\nu_t)(y)\bar{\phi}(G^{-1}(z, \nu_t)) - b(x, \mu_t)\partial_\mu \alpha(\mu_t)(x)\bar{\phi}(G^{-1}(z, \nu_t)) \right| \Pi_t(\mathrm{d}y, \mathrm{d}x) \\
&\leq C\mathcal{W}_2(\mu_t, \nu_t),
\end{aligned}$$

where we used (H.8(4)) and the fact that $\bar{\phi}$ is bounded. Furthermore, from the boundedness of $(x, \nu_t) \mapsto \partial_\mu \alpha(\nu_t)(x)$ and (H.8(1)), we derive

$$\int_{\mathbb{R}^2} \left| b(x, \mu_t)\partial_\mu G(G^{-1}(z, \nu_t), \mu_t)(x) - b(x, \mu_t)\partial_\mu G(G^{-1}(z, \mu_t), \mu_t)(x) \right| \Pi_t(\mathrm{d}y, \mathrm{d}x)$$

$$
\begin{aligned}
&\leq \int_{\mathbb{R}^2} |b(x,\mu_t)\partial_\mu \alpha(\mu_t)(x)| \left|\bar{\phi}(G^{-1}(z,\nu_t)) - \bar{\phi}(G^{-1}(z,\mu_t))\right| \Pi_t(\mathrm{d}y,\mathrm{d}x) \\
&\leq C \int_{\mathbb{R}^2} (1+|x|) \left|\bar{\phi}(G^{-1}(z,\nu_t)) - \bar{\phi}(G^{-1}(z,\mu_t))\right| \Pi_t(\mathrm{d}y,\mathrm{d}x) \\
&\leq C\mathcal{W}_2(\mu_t,\nu_t). \qquad (4.12)
\end{aligned}
$$

We remark that in the last inequality, we used the Lipschitz continuity of $z \mapsto |z| z\phi(z/c)$, Proposition 4.3.3 and employed that $(\mu_t)_{t\in[0,T]}$ is an element of the space $\mathcal{P}^b$. In a similar manner, we can show the Lipschitz continuity of

$$
z \mapsto \int_{\mathbb{R}} b(y,\nu_t)\partial_\mu G(G^{-1}(z,\nu_t),\nu_t)(y)\,\nu_t(\mathrm{d}y).
$$

Analogous statements can be derived for

$$
(z,\nu_t) \mapsto \int_{\mathbb{R}} \frac{\sigma^2(y)}{2} \partial_y \partial_\mu G(G^{-1}(z,\nu_t),\nu_t)(y)\,\nu_t(\mathrm{d}y),
$$

taking (H.7(1)) and (H.8(2)) into account. □

We are now ready to present the main result and its proof of this section:

Theorem 4.3.1. *Let Assumption (H.8) be satisfied, let $\xi \in L_0^p(\mathbb{R})$ for a given $p \geq 2$ and assume that the constant c is sufficiently small (as in Remark 4.3.3). Then, the McKean–Vlasov SDE defined in (4.5) has a unique strong solution in $\mathcal{S}^p([0,T])$.*

Proof. First, we remark that for any given flow of measures $(\mu_t)_{t\in[0,T]} \in \mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}))$ the SDE defined in (4.9) has a unique strong solution by Lemma 4.3.1. Now, for $(\mu_t)_{t\in[0,T]}, (\nu_t)_{t\in[0,T]} \in \mathcal{P}^b$, we obtain, for any $t \in [0,T]$, using Lemma 4.3.1, BDG's inequality and Hölder's inequality along with Gronwall's inequality

$$
\begin{aligned}
\mathbb{E}\left[|Z_t^\mu - Z_t^\nu|^2\right] &\leq C\left(\mathbb{E}\left[\int_0^t |\tilde{b}(Z_s^\mu,\mu_s) - \tilde{b}(Z_s^\nu,\nu_s)|^2\,\mathrm{d}s\right] + \mathbb{E}\left[\int_0^t |\tilde{\sigma}(Z_s^\mu,\mu_s) - \tilde{\sigma}(Z_s^\nu,\nu_s)|^2\,\mathrm{d}s\right]\right) \\
&\leq C\Bigg(\mathbb{E}\left[\int_0^t |\tilde{b}(Z_s^\mu,\mu_s) - \tilde{b}(Z_s^\nu,\mu_s)|^2\,\mathrm{d}s\right] + \mathbb{E}\left[\int_0^t |\tilde{b}(Z_s^\nu,\mu_s) - \tilde{b}(Z_s^\nu,\nu_s)|^2\,\mathrm{d}s\right] \\
&\quad + \mathbb{E}\left[\int_0^t |\tilde{\sigma}(Z_s^\mu,\mu_s) - \tilde{\sigma}(Z_s^\nu,\mu_s)|^2\,\mathrm{d}s\right] + \mathbb{E}\left[\int_0^t |\tilde{\sigma}(Z_s^\nu,\mu_s) - \tilde{\sigma}(Z_s^\nu,\nu_s)|^2\,\mathrm{d}s\right]\Bigg) \\
&\leq C\mathbb{E}\left[\int_0^t \left(|Z_s^\mu - Z_s^\nu|^2 + \mathcal{W}_2^2(\mu_s,\nu_s)\right)\mathrm{d}s\right] \leq C\int_0^t \mathcal{W}_2^2(\mu_s,\nu_s)\,\mathrm{d}s.
\end{aligned}
$$

For $k \geq 0$ and $t \in [0,T]$, we define the Picard iteration

$$
\mu_t^{k+1} = \mathrm{Law}\left(G^{-1}(Z_t^{\mu^k},\mu_t^k)\right), \qquad (4.13)
$$

with $Z_t^{\mu^0} = G(\xi,\delta_\xi)$ and $\mu_t^0 = \delta_\xi$. Note that by Itô's formula (applied to G^{-1}), the process defined by $X_t^{k+1} := G^{-1}(Z_t^{\mu^k},\mu_t^k)$ is the solution to

$$
\mathrm{d}X_t^{k+1} = b(X_t^{k+1},\mu_t^k)\,\mathrm{d}t + \sigma(X_t^{k+1})\,\mathrm{d}W_t, \quad X_0^{k+1} = \xi.
$$

Recall that $(X_t^{k+1})_{t\in[0,T]}$ has uniformly bounded moments (uniformly in k), due to (H.7(1)) and (H.8(1)), i.e., we have

$$\sup_{k\geq 1}\mathbb{E}\left[\sup_{0\leq t\leq T}|X_t^k|^p\right]\leq C(1+\mathbb{E}[|\xi|^p]). \tag{4.14}$$

The applicability of Itô's formula for G^{-1} is a consequence of the fact that the inverse inherits the regularity of G, in particular the mapping $\mathcal{P}_2(\mathbb{R})\ni\mu\mapsto G^{-1}(y,\mu)$ is still an element of the class $\mathcal{C}_b^{(1,1)}$ (see, Proposition B.1.1 in Appendix B.1). Then above estimate and Proposition 4.3.3 yield

$$\begin{aligned}
\sup_{t\in[0,T]}\mathcal{W}_2^2(\mu_t^{k+1},\mu_t^k)&\leq\sup_{t\in[0,T]}\mathbb{E}\left[|X_t^{k+1}-X_t^k|^2\right]\\
&\leq 2\sup_{t\in[0,T]}\mathbb{E}\left[|G^{-1}(Z_t^{\mu^k},\mu_t^k)-G^{-1}(Z_t^{\mu^{k-1}},\mu_t^k)|^2\right]\\
&\quad+2\sup_{t\in[0,T]}\mathbb{E}\left[|G^{-1}(Z_t^{\mu^{k-1}},\mu_t^k)-G^{-1}(Z_t^{\mu^{k-1}},\mu_t^{k-1})|^2\right]\\
&\leq C\int_0^T\mathcal{W}_2^2(\mu_s^k,\mu_s^{k-1})\mathrm{d}s+8C_2ce^{4C_1c}\sup_{t\in[0,T]}\mathcal{W}_2^2(\mu_t^k,\mu_t^{k-1})\\
&\leq C\int_0^T\mathcal{W}_2^2(\mu_s^k,\mu_s^{k-1})\mathrm{d}s+8C_2ce^{4C_1c}\sup_{t\in[0,T]}\mathbb{E}\left[|X_t^k-X_t^{k-1}|^2\right],
\end{aligned}\tag{4.15}$$

where $L:=8C_2ce^{4C_1c}<1$ due to the choice of c and $C>0$ is a constant depending on the constants appearing in Proposition 4.3.3 and Lemma 4.3.1.

Let now $0<T_0<T$ such that $CT_0+L<1$. With this choice the uniqueness of (4.5) on $[0,T_0]$ follows from the estimate in (4.15) by assuming there exist two solutions (X,μ) and (Y,ν) to (4.5), with μ_t and ν_t the marginal laws of X_t and Y_t, respectively, for $t\in[0,T_0]$. In addition, we observe that the sequence of flows $(\mu^k)_k$, for $\mu^k=(\mu_t^k)_{t\in[0,T_0]}$, is a Cauchy sequence in the complete metric space $\mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}))$ equipped with the Wasserstein distance $\sup_{t\in[0,T_0]}\mathcal{W}_2(\mu_t,\nu_t)$. Hence, (4.13) has a fixed point, in particular we have $X_t=G^{-1}(Z_t^\mu,\mu_t)$, where $\mu_t=\mathcal{L}_{X_t}$. Itô's formula applied to G^{-1} yields the claim for the time interval $[0,T_0]$. Repeating the above procedure starting at T_0, we can extend the solution to the interval $[T_0,T_1]$, for some $T_0<T_1<T$. This is possible as the choice of T_1 depends on X_{T_0} only through the second moment of X_{T_0}, for which we have a uniform bound for the entire interval $[0,T]$, see (4.14). Proceeding in such a manner, we can obtain well-posedness of (4.5) on $[0,T]$. □

4.3.3 Interacting particle system with non-decomposable drift

In the following, we state the model assumptions which will specify the set-up for this subsection:

H.9. Assumptions (H.8(3)) and (H.8(4)) are satisfied and we require:

(1) Assumption (H.7(1)) holds and there exists a constant $L>0$ such that $|\sigma(x)|\leq L$ for all $x\in\mathbb{R}$.

(2) There exists a constant $L_1>0$ such that

$$|b(x,\mu)-b(x,\nu)|\leq L_1\mathcal{W}_2(\mu,\nu)\quad\forall x\neq 0\in\mathbb{R},\ \mu,\nu\in\mathcal{P}_2(\mathbb{R}).$$

Further, for any $\mu \in \mathcal{P}_2(\mathbb{R})$, $b(\cdot, \mu)$ is piecewise Lipschitz continuous on the subintervals $(-\infty, 0)$ and $(0, \infty)$, uniformly with respect to μ.

(3) $\alpha \in \mathcal{C}_b^{(1,2)}$ is a bounded function and the mappings

$$\begin{aligned}
&\mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \ni (\mu, y) \mapsto \partial_\mu \alpha(\mu)(y), \\
&\mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \ni (\mu, y) \mapsto \partial_y \partial_\mu \alpha(\mu)(y), \\
&\mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \times \mathbb{R} \ni (\mu, y, y') \mapsto \partial_\mu^2 \alpha(\mu)(y, y'),
\end{aligned}$$

are bounded and Lipschitz continuous.

Remark 4.3.4. Note that, compared to (H.8(1)), we do not require the drift to be uniformly bounded in the measure component.

The interacting particles of the system $(X_t^{i,N})_{t\in[0,T]}$, for $i \in \{1, \dots, N\}$ associated with (4.5) satisfy

$$\mathrm{d}X_t^{i,N} = b(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N})\,\mathrm{d}t + \sigma(X_t^{i,N})\,\mathrm{d}W_t^i, \tag{4.16}$$

where (ξ^i, W^i), for $i \in \{1, \dots, N\}$, are independent copies of (ξ, W).

In contrast to the case of particle systems with decomposable drift, we set $\alpha : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$,

$$\alpha(\mu^{\boldsymbol{x}^N}) = \frac{b(0^-, \mu^{\boldsymbol{x}^N}) - b(0^+, \mu^{\boldsymbol{x}^N})}{2\sigma^2(0)},$$

(which could also be interpreted as a mapping $\alpha_N : \mathbb{R}^N \to \mathbb{R}$) and apply to each particle the following transformation $G : \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$,

$$G\left(x_i, \mu^{\boldsymbol{x}^N}\right) = x_i + \alpha\left(\mu^{\boldsymbol{x}^N}\right) x_i |x_i| \phi\big(x_i/c\big). \tag{4.17}$$

We set $\mathbb{R}^N \ni \boldsymbol{x}^N \mapsto G_i(\boldsymbol{x}^N) := G\left(x_i, \mu^{\boldsymbol{x}^N}\right)$, and use these mappings to define $\boldsymbol{G}_N : \mathbb{R}^N \to \mathbb{R}^N$ by

$$\boldsymbol{G}_N(\boldsymbol{x}^N) := \big(G_1(\boldsymbol{x}^N), \dots, G_N(\boldsymbol{x}^N)\big)^\top.$$

To obtain the transformed process $(\boldsymbol{Z}_t^N)_{t\in[0,T]} = (Z_t^{1,N}, \dots, Z_t^{N,N})_{t\in[0,T]}^\top \in \mathbb{R}^N$, we proceed as follows: For any $t \in [0, T]$ and $i \in \{1, \dots, N\}$, we have, using [31, Proposition 5.35] (see also Section 4.2),

$$\begin{aligned}
\mathrm{d}G(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N}) &= \mathrm{d}G_i(X_t^{1,N}, \dots, X_t^{N,N}) \\
&= \partial_{x_i} G(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N})\mathrm{d}X_t^{i,N} + \frac{1}{2}\partial_{x_i}^2 G(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N})\mathrm{d}[X^{i,N}]_t \\
&\quad + \frac{1}{N}\sum_{k=1}^N \partial_\mu G(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N})(X_t^{k,N})\mathrm{d}X_t^{k,N} \\
&\quad + \frac{1}{2N}\sum_{k=1}^N \partial_y \partial_\mu G(X_t^{i,N}, \mu_t^{\boldsymbol{X}^N})(X_t^{k,N})\mathrm{d}[X^{k,N}]_t
\end{aligned}$$

$$+\frac{1}{2N^2}\sum_{k=1}^{N}\partial_\mu^2 G(X_t^{i,N},\mu_t^{\boldsymbol{X}^N})(X_t^{k,N},X_t^{k,N})\mathrm{d}[X^{k,N}]_t$$
$$+\frac{1}{N}\partial_{x_i}\partial_\mu G(X_t^{i,N},\mu_t^{\boldsymbol{X}^N})(X_t^{i,N})\mathrm{d}[X^{i,N}]_t.$$

The applicability of Itô's formula for the function G is guaranteed by [106, Theorem 3.19]. Note that Assumption 3.4 therein is imposed to guarantee Lipschitz continuity of second order derivatives of G outside the set of discontinuities. Assumption (H.9(3)) and the definition of G substitute this condition.

Assuming for now the global invertibility of $\boldsymbol{G}_N$, we may introduce

$$\mathrm{d}\boldsymbol{Z}_t^N=\boldsymbol{B}_N\left(\boldsymbol{G}_N^{-1}(\boldsymbol{Z}_t^N)\right)\mathrm{d}t+\boldsymbol{\Sigma}_N\left(\boldsymbol{G}_N^{-1}(\boldsymbol{Z}_t^N)\right)\mathrm{d}\boldsymbol{W}_t^N,\quad \boldsymbol{Z}_0^N=\boldsymbol{G}_N((X_0^{1,N},\dots,X_0^{N,N})),$$

where $\boldsymbol{W}_t^N=(W_t^1,\dots,W_t^N)^\top$, $\boldsymbol{B}_N(\boldsymbol{x}^N)=(B_1(\boldsymbol{x}^N),\dots,B_N(\boldsymbol{x}^N))^\top$ is defined by

$$\begin{aligned}B_i(\boldsymbol{x}^N):=\;&\partial_{x_i}G(x_i,\mu^{\boldsymbol{x}^N})b(x_i,\mu^{\boldsymbol{x}^N})+\frac{1}{2}\sigma^2(x_i)\partial_{x_i}^2G(x_i,\mu^{\boldsymbol{x}^N})\\&+\frac{1}{N}\sum_{k=1}^{N}\partial_\mu G(x_i,\mu^{\boldsymbol{x}^N})(x_k)b(x_k,\mu^{\boldsymbol{x}^N})+\frac{1}{2N}\sum_{k=1}^{N}\partial_y\partial_\mu G(x_i,\mu^{\boldsymbol{x}^N})(x_k)\sigma^2(x_k)\\&+\frac{1}{2N^2}\sum_{k=1}^{N}\partial_\mu^2 G(x_i,\mu^{\boldsymbol{x}^N})(x_k,x_k)\sigma^2(x_k)+\frac{1}{N}\partial_{x_i}\partial_\mu G(x_i,\mu^{\boldsymbol{x}^N})(x_i)\sigma^2(x_i),\end{aligned}\tag{4.18}$$

and $\boldsymbol{\Sigma}_N(\boldsymbol{x}^N)=\left(\Sigma^{i,j}(\boldsymbol{x}^N)\right)_{i,j\in\{1,\dots,N\}}$ by

$$\Sigma^{i,j}(\boldsymbol{x}^N)=\partial_{x_i}G(x_i,\mu^{\boldsymbol{x}^N})\sigma(x_i)\delta_{i,j}+\frac{1}{N}\partial_\mu G(x_i,\mu^{\boldsymbol{x}^N})(x_j)\sigma(x_j).\tag{4.19}$$

In the following lemma, we will prove the invertibility of $\boldsymbol{G}_N$.

Lemma 4.3.2. *Let Assumption (H.9(3)) be satisfied and assume that the constant c in (4.17) satisfies*

$$c<\min\left(1,\left(\sup_{\boldsymbol{x}^N\in\mathbb{R}^N}\left(|\alpha_N(\boldsymbol{x}^N)|+\max_{i\in\{1,\dots,N\}}|\partial_\mu\alpha(\mu^{\boldsymbol{x}^N})(x_i)|\right)\right)^{-1}\right).\tag{4.20}$$

Then, $\boldsymbol{G}_N$ has a global inverse.

Proof. We will employ Hadamard's global inverse function theorem (see, e.g., [139, Theorem 2.2]) to prove that $\boldsymbol{G}_N$ has a global inverse. To do so, we need to verify that $\boldsymbol{G}_N$ is in $\mathcal{C}^1(\mathbb{R}^N,\mathbb{R}^{N\times N})$, $\lim_{|\boldsymbol{x}^N|\to\infty}|\boldsymbol{G}_N(\boldsymbol{x}^N)|=\infty$, and $\boldsymbol{G}_N'(\boldsymbol{x}^N)$ is invertible for all $\boldsymbol{x}^N\in\mathbb{R}^N$. The first two mentioned conditions are obvious, due to the definition of $\boldsymbol{G}_N$, the uniform boundedness of α and the definition of $\bar\phi(x)=x|x|\phi\big(x/c\big)$.

Hence, we need to prove that $\boldsymbol{G}_N'(\boldsymbol{x}^N)$ is invertible. First, note that

$$\boldsymbol{G}_N'(\boldsymbol{x}^N)=\mathrm{I}_{N\times N}+\mathrm{diag}_{N\times N}(\bar\phi'(x_1)\alpha_N(\boldsymbol{x}^N),\dots,\bar\phi'(x_N)\alpha_N(\boldsymbol{x}^N))+\bar{\boldsymbol{\phi}}(\boldsymbol{x}^N)\boldsymbol{\alpha}_N'(\boldsymbol{x}^N),$$

where $\mathrm{I}_{N\times N}$ is the $N\times N$ identity matrix, $\bar{\boldsymbol{\phi}}(\boldsymbol{x}^N) = (\bar{\phi}(x_1),\dots,\bar{\phi}(x_N))^\top$, with $(\boldsymbol{\alpha}'_N(\boldsymbol{x}^N))_i = \frac{1}{N}\partial_\mu\alpha(\mu^{\boldsymbol{x}^N})(x_i)$ and $\boldsymbol{\alpha}'_N$ is a row vector.

Now, we define

$$\mathcal{A}(\boldsymbol{x}^N) := \mathrm{diag}_{N\times N}(\bar{\phi}'(x_1)\alpha_N(\boldsymbol{x}^N),\dots,\bar{\phi}'(x_N)\alpha_N(\boldsymbol{x}^N)) + \bar{\boldsymbol{\phi}}(\boldsymbol{x}^N)\boldsymbol{\alpha}'_N(\boldsymbol{x}^N),$$

and remark that $\boldsymbol{G}'_N(\boldsymbol{x}^N)$ can be identified with the linear operator $\mathrm{Id}_{\mathbb{R}^N} + \mathcal{A}(\boldsymbol{x}^N) : \mathbb{R}^N \to \mathbb{R}^N$. Therefore, succeeding in showing that c can be chosen (uniformly in $\boldsymbol{x}^N$) in a way such that the operator norm of $\mathcal{A}(\boldsymbol{x}^N)$ is smaller than one would yield the claim, as in this case $\boldsymbol{G}'_N(\boldsymbol{x}^N)$ is close to the identity (see, [106, Lemma 3.17]). We compute

$$\begin{aligned}\|\mathcal{A}(\boldsymbol{x}^N)\| &\le \max_{i\in\{1,\dots,N\}}|\bar{\phi}'(x_i)||\alpha_N(\boldsymbol{x}^N)| + \max_{i\in\{1,\dots,N\}}|\bar{\phi}(x_i)||\partial_\mu\alpha(\mu^{\boldsymbol{x}^N})(x_i)|\\ &\le c|\alpha_N(\boldsymbol{x}^N)| + c^2\max_{i\in\{1,\dots,N\}}|\partial_\mu\alpha(\mu^{\boldsymbol{x}^N})(x_i)|,\end{aligned}$$

which implies that for

$$c < \min\left(1, \left(|\alpha_N(\boldsymbol{x}^N)| + \max_{i\in\{1,\dots,N\}}|\partial_\mu\alpha(\mu^{\boldsymbol{x}^N})(x_i)|\right)^{-1}\right),$$

$\|\mathcal{A}(\boldsymbol{x}^N)\| < 1$. Note that (H.9(3)) guarantees that α and its derivatives are uniformly bounded, i.e., c can be chosen uniformly in $\boldsymbol{x}^N$. Therefore, Hadamard's global inverse function theorem proves that, for each given $N\ge 1$, $\boldsymbol{G}_N : \mathbb{R}^N \to \mathbb{R}^N$ is a diffeomorphism. □

We proceed by showing that the transformed SDE has (locally) Lipschitz continuous coefficients.

Lemma 4.3.3. *Let Assumption (H.9) be satisfied. Then, the coefficients $\boldsymbol{B}_N$ and $\boldsymbol{\Sigma}_N$ introduced in (4.18) and (4.19), respectively, are locally Lipschitz continuous with linear growth.*

Proof. For $\boldsymbol{x}^N, \boldsymbol{y}^N \in \mathbb{R}^N$, we obtain using (4.19)

$$\begin{aligned}\|\boldsymbol{\Sigma}_N(\boldsymbol{x}^N) - \boldsymbol{\Sigma}_N(\boldsymbol{y}^N)\|^2 &\le \sum_{i=1}^N |\Sigma^{i,i}(\boldsymbol{x}^N) - \Sigma^{i,i}(\boldsymbol{y}^N)|^2 + \sum_{i\ne j}|\Sigma^{i,j}(\boldsymbol{x}^N) - \Sigma^{i,j}(\boldsymbol{y}^N)|^2\\ &\le \sum_{i=1}^N |\partial_{x_i}G(x_i,\mu^{\boldsymbol{x}^N})\sigma(x_i) - \partial_{y_i}G(y_i,\mu^{\boldsymbol{y}^N})\sigma(y_i)|^2\\ &\quad + \frac{1}{N^2}\sum_{i\ne j}|\partial_\mu G(x_i,\mu^{\boldsymbol{x}^N})(x_j)\sigma(x_j) - \partial_\mu G(y_i,\mu^{\boldsymbol{x}^N})(y_j)\sigma(y_j)|^2\\ &\le C|\boldsymbol{x}^N - \boldsymbol{y}^N|^2,\end{aligned}$$

where we used (H.9(1)) and the Lipschitz continuity of the functions $x\mapsto\partial_x G(x,\mu)$ and $\mu\mapsto\partial_x G(x,\mu)$ to estimate the first term. Assumptions (H.9(1)) and (H.9(3)), in particular the Lipschitz continuity of $x\mapsto\partial_\mu G(x,\mu)(y)$ and $\mu\mapsto\partial_\mu G(x,\mu)(y)$, are employed to handle the second sum. Also note that all these mappings are bounded due to (H.9(3)) and the definition of the transformation.

Noting that

$$|\boldsymbol{B}_N(\boldsymbol{x}^N) - \boldsymbol{B}_N(\boldsymbol{y}^N)|^2 = \sum_{i=1}^N |B_i(\boldsymbol{x}^N) - B_i(\boldsymbol{y}^N)|^2,$$

we further obtain for the drift

$$|B_i(\boldsymbol{x}^N) - B_i(\boldsymbol{y}^N)|^2 \leq \sum_{i=1}^5 \Pi_i,$$

where we defined

$$\begin{aligned}
\Pi_1 &:= C|\partial_{x_i} G(x_i, \mu^{\boldsymbol{x}^N}) b(x_i, \mu^{\boldsymbol{x}^N}) + \frac{1}{2}\sigma^2(x_i)\partial_{x_i}^2 G(x_i, \mu^{\boldsymbol{x}^N}) \\
&\qquad - \partial_{y_i} G(y_i, \mu^{\boldsymbol{y}^N}) b(y_i, \mu^{\boldsymbol{y}^N}) - \frac{1}{2}\sigma^2(y_i)\partial_{y_i}^2 G(y_i, \mu^{\boldsymbol{y}^N})|^2, \\
\Pi_2 &:= \frac{C}{N}\sum_{k=1}^N |\partial_\mu G(x_i, \mu^{\boldsymbol{x}^N})(x_k) b(x_k, \mu^{\boldsymbol{x}^N}) - \partial_\mu G(y_i, \mu^{\boldsymbol{y}^N})(y_k) b(y_k, \mu^{\boldsymbol{y}^N})|^2, \\
\Pi_3 &:= \frac{C}{2N}\sum_{k=1}^N |\partial_y \partial_\mu G(x_i, \mu^{\boldsymbol{x}^N})(x_k)\sigma^2(x_k) - \partial_y \partial_\mu G(y_i, \mu^{\boldsymbol{x}^N})(y_k)\sigma^2(y_k)|^2, \\
\Pi_4 &:= \frac{C}{2N^2}\sum_{k=1}^N |\partial_\mu^2 G(x_i, \mu^{\boldsymbol{x}^N})(x_k, x_k)\sigma^2(x_k) - \partial_\mu^2 G(y_i, \mu^{\boldsymbol{y}^N})(y_k, y_k)\sigma^2(y_k)|^2, \\
\Pi_5 &:= \frac{C}{N}|\partial_{x_i}\partial_\mu G(x_i, \mu^{\boldsymbol{x}^N})(x_i)\sigma^2(x_i) - \partial_{y_i}\partial_\mu G(y_i, \mu^{\boldsymbol{y}^N})(y_i))\sigma^2(y_i)|^2.
\end{aligned}$$

That the terms Π_3, Π_4 and Π_5 allow a Lipschitz bound is a consequence of (H.9(1)) and (H.9(3)).

For any $R > 0$, in view of (H.9(2)) and (H.9(3)), we derive the following estimate for Π_2:

$$\begin{aligned}
\Pi_2 &\leq \frac{C}{N}\Bigg(\sum_{k=1}^N \left|\partial_\mu \alpha(\mu^{\boldsymbol{x}^N})(x_k)\bar{\phi}(x_i) b(x_k, \mu^{\boldsymbol{x}^N}) - \partial_\mu \alpha(\mu^{\boldsymbol{y}^N})(y_k)\bar{\phi}(x_i) b(y_k, \mu^{\boldsymbol{y}^N})\right|^2 \\
&\quad + \sum_{k=1}^N \left|\partial_\mu \alpha(\mu^{\boldsymbol{y}^N})(y_k)\bar{\phi}(x_i) b(y_k, \mu^{\boldsymbol{y}^N}) - \partial_\mu \alpha(\mu^{\boldsymbol{y}^N})(y_k)\bar{\phi}(y_i) b(y_k, \mu^{\boldsymbol{y}^N})\right|^2\Bigg) \\
&\leq \frac{C}{N}\sum_{k=1}^N \left|\partial_\mu \alpha(\mu^{\boldsymbol{x}^N})(x_k) b(x_k, \mu^{\boldsymbol{x}^N}) - \partial_\mu \alpha(\mu^{\boldsymbol{y}^N})(y_k) b(y_k, \mu^{\boldsymbol{y}^N})\right|^2 + L_R|x_i - y_i|^2,
\end{aligned}$$

for some constant $L_R > 0$ and any $|\boldsymbol{x}^N|, |\boldsymbol{y}^N| \leq R$. We proceed with the estimate

$$\sum_{k=1}^N \left|\partial_\mu \alpha(\mu^{\boldsymbol{x}^N})(x_k) b(x_k, \mu^{\boldsymbol{x}^N}) - \partial_\mu \alpha(\mu^{\boldsymbol{y}^N})(y_k) b(y_k, \mu^{\boldsymbol{y}^N})\right|^2 \leq C\sum_{k=1}^N |x_k - y_k|^2,$$

which holds due to Assumptions (H.8(3)) and (H.8(4)). Combining above estimates, we obtain

$$\Pi_2 \leq L_R\left(|x_i - y_i|^2 + \frac{1}{N}\sum_{k=1}^N |x_k - y_k|^2\right).$$

Finally, we point out that

$$x \mapsto b(x, \mu) + \frac{1}{2}\sigma^2(x)\partial_x^2 G(x, \mu),$$

is Lipschitz continuous due to the choice of α. Employing this along with (H.9(1)), (H.9(2)) and (H.9(3)), we derive

$$\Pi_1 \leq C \left(|x_i - y_i|^2 + \frac{1}{N} \sum_{k=1}^{N} |x_k - y_k|^2 \right),$$

for some constant $C > 0$. Taking the estimates for $\Pi_1, \ldots, \Pi_5$ into account, yields the local Lipschitz continuity of $\boldsymbol{B}_N$.

The linear growth of $\boldsymbol{B}_N$ and $\boldsymbol{\Sigma}_N$, i.e., that there exists a constant $C > 0$ such that $|\boldsymbol{B}_N(\boldsymbol{x}^N)| + \|\boldsymbol{\Sigma}_N(\boldsymbol{x}^N)\| \leq C(1 + |\boldsymbol{x}^N|)$ for all $\boldsymbol{x}^N \in \mathbb{R}^N$, is a direct consequence of the growth conditions on b and σ along with the bounds for the derivatives of G and α. □

Theorem 4.3.2. *Let Assumption (H.9) be satisfied, let $\xi \in L_0^p(\mathbb{R})$ for a given $p \geq 2$ and assume that the constant c in (4.17) satisfies (4.20). Then, the interacting particle system defined in (4.16) has a unique strong solution in $\mathcal{S}^p([0,T])$.*

Proof. From Lemma 4.3.3 and the linear growth of $\boldsymbol{B}_N$ and $\boldsymbol{\Sigma}_N$, we can deduce that the SDE for $\boldsymbol{Z}^N$ has a unique strong solution (see, [112, Chapter 5, Theorem 2.5]). Applying now Itô's formula to $\boldsymbol{G}_N^{-1}(\boldsymbol{Z}_t^N)$ proves the strong uniqueness of the particle system defined in (4.16). Note that $\boldsymbol{G}_N^{-1}$ exists due to Lemma 4.3.2 and Itô's formula is applicable for $\boldsymbol{G}_N^{-1}$ as it inherits the regularity of $\boldsymbol{G}_N$ (see, Appendix B.2 for details). □

4.4 Euler–Maruyama scheme with and without transformation

In this section, we restrict most of our discussion to the case of a McKean–Vlasov SDE with decomposable drift, due to the simpler structure of the underlying transformation and the particle systems.

In the following subsections, we will present two Euler–Maruyama schemes to discretise the particle system defined in (4.4) in time.

For the first scheme (Scheme 1), we will discretise the transformed (continuous) particle system in time and then exploit the global inverse G^{-1} to obtain approximations of the original (discontinuous) particle system. A slight modification of this scheme will also be applied in the non-decomposable case. An approximation result with respect to the number of particles will also be presented.

The second scheme (Scheme 2) will be defined by directly discretising the discontinuous particle system, without making use of the transformation G. We give strong convergence rates in terms of the number of time-steps and pathwise strong propagation of chaos results in order to obtain quantitative L_2-approximations for the underlying McKean–Vlasov SDE.

4.4.1 Scheme 1: Euler–Maruyama after transformation (decomposable case)

We define the following explicit Euler–Maruyama scheme to discretise the particle system (4.4) in time. In a first step, we partition a given time interval $[0,T]$ into subintervals of equal length $h = T/M$, for some integer $M > 0$, and define $t_n := nh$. Then, we simulate the transformed particle system by

$$Z^{i,N,M}_{t_{n+1}} = Z^{i,N,M}_{t_n} + \tilde{b}(G^{-1}(Z^{i,N,M}_{t_n}), \mu^{\boldsymbol{Z}^{N,M}}_{t_n})h + \tilde{\sigma}(G^{-1}(Z^{i,N,M}_{t_n}))\Delta W^i_n, \tag{4.21}$$

for $n \in \{0,\dots,M-1\}$, where $Z^{i,N,M}_0 = G(X^{i,N}_0)$, $\Delta W^i_n = W^i_{t_{n+1}} - W^i_{t_n}$, for $i \in \{1,\dots,N\}$, and

$$\mu^{\boldsymbol{Z}^{N,M}}_{t_n}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{G^{-1}(Z^{j,N}_{t_n})}(\mathrm{d}x).$$

We introduce the notation $\eta(t) := \sup\{s \in \{0,h,\dots,Mh\} : s \le t\}$, for $t \in [0,T]$, which allows us to define the continuous time version of (4.21)

$$Z^{i,N,M}_t = Z^{i,N,M}_0 + \int_0^t \tilde{b}(G^{-1}(Z^{i,N,M}_{\eta(s)}), \mu^{\boldsymbol{Z}^{N,M}}_{\eta(s)})\,\mathrm{d}s + \int_0^t \tilde{\sigma}(G^{-1}(Z^{i,N,M}_{\eta(s)}))\,\mathrm{d}W^i_s. \tag{4.22}$$

Then, we propose an Euler–Maruyama approximation to $X^{i,N}_t$, for $i \in \{1,\dots,N\}$ and $t \in [0,T]$, by

$$X^{i,N,M}_t = G^{-1}(Z^{i,N,M}_t). \tag{4.23}$$

The convergence of this algorithm is proven in the following theorem:

Theorem 4.4.1. *Let Assumption (H.7) be satisfied, let $\xi \in L^p_0(\mathbb{R})$ for some $p > 4$ and assume $c < 1/|\alpha|$. For $i \in \{1,\dots,N\}$, let $(X^i_t)_{t\in[0,T]}$ be the unique strong solution of (4.3) driven by the Brownian motion $(W^i_t)_{t\in[0,T]}$, and $(X^{i,N,M}_t)_{t\in[0,T]}$ be given by (4.22) and (4.23). Then, there exists a constant $C > 0$ (independent of N and M) such that*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |X^i_t - X^{i,N,M}_t|^2\right] \le C\left(h + N^{-1/2}\right).$$

Proof. Note that

$$|X^i_t - X^{i,N,M}_t|^2 \le 2|X^i_t - X^{i,N}_t|^2 + 2|X^{i,N}_t - X^{i,N,M}_t|^2,$$

and recall that the dynamics of $Z^i_t = G(X^i_t)$ satisfies

$$\mathrm{d}Z^i_t = \tilde{b}(Z^i_t, \tilde{\mu}^Z_t)\,\mathrm{d}t + \tilde{\sigma}(Z^i_t)\,\mathrm{d}W^i_t,$$

where $\tilde{\mu}^Z_t := \mathcal{L}_{G^{-1}(Z^i_t)}$. Therefore, we obtain for some constant $C > 0$

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]} |X^i_t - X^{i,N}_t|^2\right] &= \mathbb{E}\left[\sup_{t\in[0,T]} |G^{-1}(Z^i_t) - G^{-1}(Z^{i,N}_t)|^2\right] \\ &\le L^2_{G^{-1}}\mathbb{E}\left[\sup_{t\in[0,T]} |Z^i_t - Z^{i,N}_t|^2\right] \le CN^{-1/2},\end{aligned}$$

where the last inequality can be derived similarly to the propagation of chaos results for equations with Lipschitz continuous coefficients as in, e.g., [30] for $d = 1$ or also [116]. To apply the aforementioned propagation of chaos result, we need the requirement that $\xi \in L_0^p(\mathbb{R})$ for $p > 4$ (see, also [31, Theorem 5.8]). Furthermore, we have

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{i,N} - X_t^{i,N,M}|^2\right] &= \mathbb{E}\left[\sup_{t\in[0,T]} |G^{-1}(Z_t^{i,N}) - G^{-1}(Z_t^{i,N,M})|^2\right] \\ &\leq L_{G^{-1}}^2 \mathbb{E}\left[\sup_{t\in[0,T]} |Z_t^{i,N} - Z_t^{i,N,M}|^2\right] \leq Ch,\end{aligned}$$

since the SDEs for $(Z_t^{i,N})_{t\in[0,T]}$ and $(Z_t^{i,N,M})_{t\in[0,T]}$ have globally Lipschitz continuous coefficients. From these two estimates the claim follows. □

4.4.2 Scheme 2: Euler–Maruyama without transformation (decomposable case)

As G and G^{-1} may be difficult to construct in multi-dimensional settings, and since the evaluation for the inverse at each time point can be computationally expensive, it would be preferable to discretise the particle system $(X^{i,N})_{i\in\{1,\dots,N\}}$ in time directly, without the use of the transformation G. In addition, a drawback of Scheme 1 is that an SDE with additive diffusion term will be transformed into one with multiplicative noise and therefore the Euler–Maruyama scheme no longer coincides with the first order Milstein scheme. We employ an Euler–Maruyama scheme to discretise the particle system (4.4) and compute an approximate solution by

$$X_t^{i,N,M} = X_0^{i,N} + \int_0^t \left(b_1(X_{\eta(s)}^{i,N,M}) + b_2(X_{\eta(s)}^{i,N,M}, \mu_{\eta(s)}^{\boldsymbol{X}^{N,M}})\right) \mathrm{d}s + \int_0^t \sigma(X_{\eta(s)}^{i,N,M})\, \mathrm{d}W_s^i. \tag{4.24}$$

The following results are concerned with moment stability of the discretised particle system and estimates for the occupation time of the particle system in the neighbourhood of the set of discontinuities.

Moment stability:

We first remark that due to the linear growth of the coefficients in the state component, the Lipschitz continuity in the measure variable and the fact that all particles are identically distributed, we have the following result (see, e.g., [112] for details):

Proposition 4.4.1. *Let Assumption (H.7) be satisfied, and let $\xi \in L_0^p(\mathbb{R})$ for $p \geq 2$. Then, for all $t \in [0,T]$ there exist constants $C_1, C_2 > 0$, such that*

$$\max_{i\in\{1,\dots,N\}} \max_{n\in\{0,\dots,M\}} \mathbb{E}\left[|X_{t_n}^{i,N,M}|^p\right] \leq C_1,$$

and for all $i \in \{1,\dots,N\}$,

$$\mathbb{E}\left[|X_t^{i,N,M} - X_{\eta(t)}^{i,N,M}|^p\right] \leq C_2 h^{p/2}.$$

Occupation time formula for (4.24):

Below, we will show an estimate of the expected occupation time of a fixed particle of the system defined by (4.24) in a neighbourhood of zero.

Proposition 4.4.2. *Let Assumption (H.7) be satisfied and let i be an arbitrary but fixed particle index. Further, let $\xi \in L^0_p(\mathbb{R})$ for $p \geq 2$ and let $(X_t^{i,N,M})_{t\in[0,T]}$ be given by (4.24). Then, there exists a constant $C > 0$ such that for all $N, M \in \mathbb{N}$ and all sufficiently small $\varepsilon > 0$*

$$\int_0^T \mathbb{P}\left(\{\boldsymbol{X}_t^{N,M} \in \Theta^{i,\varepsilon}\}\right) \mathrm{d}t \leq C\varepsilon,$$

where $\boldsymbol{X}_t^{N,M} = (X_t^{1,N,M}, \dots, X_t^{N,N,M})^\top \in \mathbb{R}^N$ and $\Theta^{i,\varepsilon}$ is given by

$$\Theta^{i,\varepsilon} := \{\boldsymbol{x}^N = (x_1, \dots, x_N)^\top \in \mathbb{R}^N : \exists \boldsymbol{y}^N \in \Theta^i \text{ with } |\boldsymbol{x}^N - \boldsymbol{y}^N| < \varepsilon\},$$

with $\Theta^i := \{\boldsymbol{x}^N = (x_1, \dots, x_N)^\top \in \mathbb{R}^N : x_i = 0\}$.

Proof. We aim to apply [107, Theorem 2.7], which states the following: Let $(X_t)_{t\in[0,T]}$ be an $\mathbb{R}^d$-valued Itô process

$$X_T = X_0 + \int_0^T A_t \,\mathrm{d}t + \int_0^T B_t \,\mathrm{d}W_t,$$

with progressively measurable processes $A = (A_t)_{t\in[0,T]}$ and $B = (B_t)_{t\in[0,T]}$, where A is $\mathbb{R}^d$-valued and B is $\mathbb{R}^{d\times d}$-valued. The set of discontinuities, Θ, is assumed to be a $\mathcal{C}^3$ hypersurface of positive reach. Namely, there exists $\varepsilon > 0$ such that $p(x) = \arg\min_{y\in\Theta} |x - y|$ is a single valued function of class $\mathcal{C}^3$ on the tubular neighbourhood $\Theta^\varepsilon := \{x \in \mathbb{R}^d : \inf_{y\in\Theta} |x - y| < \varepsilon\}$ (see Definition 2.4 in [107] for details). Then, there exists a constant $C > 0$, such that for all sufficiently small $\varepsilon > 0$

$$\int_0^T \mathbb{P}\left(\{X_t \in \Theta^\varepsilon\}\right) \mathrm{d}t \leq C\varepsilon,$$

assuming, additionally, that

1. the processes A and B are almost surely bounded by a constant C if $X_t(\omega)$ is in a small neighbourhood of Θ, and

2. there exists a constant $C > 0$ such that for almost all $\omega \in \Omega$, we have: If, for any $t \in [0, T]$, $X_t(\omega)$ is in a small neighbourhood of Θ then

$$n^\top\left(p\left(X_t(\omega)\right)\right) B_t^\top(\omega) B_t(\omega) n\left(p\left(X_t(\omega)\right)\right) \geq C,$$

where $p(x)$ denotes the projection of $x \in \mathbb{R}^d$ onto the set of discontinuities, $n(x)$ has length one and is orthogonal to the tangent space of Θ in x.

We now return to our particular model problem. First, we remark that Θ^i satisfies all regularity conditions of [107, Theorem 2.7], i.e., it is a $\mathcal{C}^3$ hypersurface of positive reach. We then observe that the N-dimensional particle system can be rewritten as

$$\mathrm{d}\boldsymbol{X}_t^{N,M} = \boldsymbol{B}_N(\boldsymbol{X}_{\eta(t)}^{N,M})\,\mathrm{d}t + \boldsymbol{\Sigma}_N(\boldsymbol{X}_{\eta(t)}^{N,M})\,\mathrm{d}\boldsymbol{W}_t^N,$$

where $\boldsymbol{W}_t^N = (W_t^1, \dots, W_t^N)^\top$ and $\boldsymbol{B}_N : \mathbb{R}^N \to \mathbb{R}^N$ and $\boldsymbol{\Sigma}_N : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ are defined by

$$\begin{aligned}\boldsymbol{B}_N(\boldsymbol{x}^N) &= (b(x_1, \mu^{\boldsymbol{x}^N}), \dots, b(x_N, \mu^{\boldsymbol{x}^N}))^\top,\\ \boldsymbol{\Sigma}_N(\boldsymbol{x}^N) &= \mathrm{diag}_{N\times N}(\sigma(x_1), \dots, \sigma(x_N)).\end{aligned}$$

Further, we observe that there is a constant $C > 0$ such that: If, for any $t \in [0,T]$ and $\omega \in \Omega$, $\boldsymbol{X}_t^{N,M}(\omega)$ is in a small neighbourhood of Θ^i then

$$n^\top\left(p\left(\boldsymbol{X}_t^{N,M}(\omega)\right)\right)\boldsymbol{\Sigma}_N^\top(\boldsymbol{X}_t^{N,M}(\omega))\boldsymbol{\Sigma}_N(\boldsymbol{X}_t^{N,M}(\omega))n\left(p\left(\boldsymbol{X}_t^{N,M}(\omega)\right)\right) \geq C,$$

as σ is continuous and $\sigma(0) \neq 0$. Also, note that a normal vector of the tangent space of Θ^i is e_i, i.e., the i-th unit vector. Further, a close inspection of the proof of [107, Theorem 2.7], shows that the boundedness assumption on the coefficients in a neighbourhood of Θ^i is not needed in our case, due to the moment bound of an individual particle established in Proposition 4.4.1. Hence,

$$\int_0^T \mathbb{P}\left(\{\boldsymbol{X}_t^{N,M} \in \Theta^{i,\varepsilon}\}\right)\,\mathrm{d}t \leq C\varepsilon,$$

where the constant $C > 0$ is independent of N, due to the fact that the normal vector is the i-th unit vector. □

Auxiliary proposition:

Based on the occupation time estimate from Proposition 4.4.2, we will prove the following result, which is needed in the proof of Theorem 4.4.2 given below.

Proposition 4.4.3. *Let Assumption (H.7) be satisfied, and let $\xi \in L_0^p(\mathbb{R})$ for $p \geq 8$. Furthermore, let $(X_t^{i,N,M})_{t\in[0,T]}$ be given by (4.24). Then, there exists a constant $C > 0$ (independent of N and M) such that for any $t \in [0,T]$, we have*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\left|\int_0^t \left(G''(X_s^{i,N,M}) - G''(X_{\eta(s)}^{i,N,M})\right)\sigma^2(X_{\eta(s)}^{i,N,M})\,\mathrm{d}s\right|^2\right] \leq Ch^{2/9}.$$

Proof. First, observe that the linear growth of σ and the piecewise Lipschitz continuity of G'' imply that there exists a constant $C > 0$ such that

$$\begin{aligned}&\left|\left(G''(X_s^{i,N,M}) - G''(X_{\eta(s)}^{i,N,M})\right)\sigma^2(X_{\eta(s)}^{i,N,M})\right|\\ &\leq \begin{cases} C\left(1 + (X_{\eta(s)}^{i,N,M})^2\right)|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}|, \ X_s^{i,N,M} \notin (-\varepsilon,\varepsilon), \ |X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}| < \varepsilon,\\ C\left(1 + (X_{\eta(s)}^{i,N,M})^2\right), \ \text{otherwise,}\end{cases}\end{aligned}$$

where $\varepsilon > 0$ will be specified later. With this at hand, we derive

$$
\begin{aligned}
&\mathbb{E}\left[\left|\int_0^t \left(G''(X_s^{i,N,M}) - G''(X_{\eta(s)}^{i,N,M})\right)\sigma^2(X_{\eta(s)}^{i,N,M})\,\mathrm{d}s\right|^2\right] \\
&\leq C\mathbb{E}\left[\int_0^t \left|\left(G''(X_s^{i,N,M}) - G''(X_{\eta(s)}^{i,N,M})\right)\sigma^2(X_{\eta(s)}^{i,N,M})\right|^2 \mathrm{d}s\right] \\
&\leq C\Bigg(\mathbb{E}\left[\int_0^t \left|\left(G''(X_s^{i,N,M}) - G''(X_{\eta(s)}^{i,N,M})\right)\sigma^2(X_{\eta(s)}^{i,N,M})\right|^2 \left(\mathrm{I}_{\{X_s^{i,N,M}\notin(-\varepsilon,\varepsilon)\}}\mathrm{I}_{\{|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}|<\varepsilon\}}\right)\mathrm{d}s\right] \\
&\quad + \mathbb{E}\Bigg[\int_0^t \left|\left(G''(X_s^{i,N,M}) - G''(X_{\eta(s)}^{i,N,M})\right)\sigma^2(X_{\eta(s)}^{i,N,M})\right|^2 \\
&\qquad \times \left(\mathrm{I}_{\{X_s^{i,N,M}\in(-\varepsilon,\varepsilon)\}} + \mathrm{I}_{\{X_s^{i,N,M}\notin(-\varepsilon,\varepsilon)\}}\mathrm{I}_{\{|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}|\geq\varepsilon\}}\right)\mathrm{d}s\Bigg]\Bigg) \\
&\leq C\Bigg(\mathbb{E}\left[\int_0^t \left(1 + (X_{\eta(s)}^{i,N,M})^4\right)|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}|^2 \left(\mathrm{I}_{\{X_s^{i,N,M}\notin(-\varepsilon,\varepsilon)\}}\mathrm{I}_{\{|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}|<\varepsilon\}}\right)\mathrm{d}s\right] \\
&\quad + \mathbb{E}\left[\int_0^t \left(1 + (X_{\eta(s)}^{i,N,M})^4\right)\left(\mathrm{I}_{\{X_s^{i,N,M}\in(-\varepsilon,\varepsilon)\}} + \mathrm{I}_{\{X_s^{i,N,M}\notin(-\varepsilon,\varepsilon)\}}\mathrm{I}_{\{|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}|\geq\varepsilon\}}\right)\mathrm{d}s\right]\Bigg) \\
&\leq C\left(\varepsilon^2 + \varepsilon^{1/2} + \int_0^t \left(\mathbb{P}(|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}| \geq \varepsilon)\right)^{1/2}\mathrm{d}s\right),
\end{aligned}
$$

where we used Hölder's inequality, Proposition 4.4.1 and Proposition 4.4.2 in the last display. Markov's inequality along with Proposition 4.4.1 imply that there exists a constant $C > 0$ such that

$$
\left(\mathbb{P}(|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}| \geq \varepsilon)\right)^{1/2} \leq \frac{\left(\mathbb{E}\left[\left|X_s^{i,N,M} - X_{\eta(s)}^{i,N,M}\right|^8\right]\right)^{1/2}}{\varepsilon^4} \leq \frac{Ch^2}{\varepsilon^4}.
$$

Choosing $\varepsilon = h^{4/9}$ gives the result. □

We are now ready to present our main convergence result. In this case, we only obtain the strong convergence rate of order 1/9:

Theorem 4.4.2. *Let Assumption (H.7) be satisfied, let $\xi \in L_0^p(\mathbb{R})$ for some $p \geq 8$ and assume $c < 1/|\alpha|$. Furthermore, let $(X_t^i)_{t\in[0,T]}$ be the unique strong solution of (4.3) driven by the Brownian motion $(W_t^i)_{t\in[0,T]}$ and $(X_t^{i,N,M})_{t\in[0,T]}$ given by (4.24). Then, there exists a constant $C > 0$ (independent of N and M) such that*

$$
\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - X_t^{i,N,M}|^2\right] \leq C\left(h^{2/9} + N^{-1/2}\right).
$$

Proof. Note that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - X_t^{i,N}|^2\right] &\leq L_{G^{-1}}\mathbb{E}\left[\sup_{t\in[0,T]} |G(X_t^i) - G(X_t^{i,N})|^2\right] \\
&\leq CN^{-1/2},
\end{aligned} \tag{4.25}
$$

where in the last display, we used the pathwise propagation of chaos result as in the previous subsection. Further, we have

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^{i,N}-X_t^{i,N,M}|^2\right] &\leq L_{G^{-1}}\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t^{i,N}-G(X_t^{i,N,M})|^2\right]\\ &\leq C\left(\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t^{i,N}-Z_t^{i,N,M}|^2\right]+\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t^{i,N,M}-G(X_t^{i,N,M})|^2\right]\right)\\ &\leq C\left(h+\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t^{i,N,M}-G(X_t^{i,N,M})|^2\right]\right),\end{aligned}$$

where in the last estimate, we employed standard strong convergence results for the Euler–Maruyama scheme applied to SDEs with globally Lipschitz continuous coefficients. Following similar arguments to [107] or [119], one further obtains, by applying Itô's formula to $G(X_t^{i,N,M})$,

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t^{i,N,M}-G(X_t^{i,N,M})|^2\right] \leq C\Bigg(h+\int_0^T\mathbb{E}\left[\left|\left(G''(X_s^{i,N,M})-G''(X_{\eta(s)}^{i,N,M})\right)\sigma^2(X_{\eta(s)}^{i,N,M})\right|^2\right]\mathrm{d}s\\ +\int_0^T\mathbb{E}\left[\sup_{s\in[0,t]}|Z_s^{i,N,M}-G(X_s^{i,N,M})|^2\right]\mathrm{d}t\Bigg),\end{aligned}$$

where the second summand on the right side is of order $h^{2/9}$ due to Proposition 4.4.3. Hence, Gronwall's inequality yields

$$\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^{i,N}-X_t^{i,N,M}|^2\right]\leq Ch^{2/9}, \tag{4.26}$$

and the claim follows combining (4.25) and (4.26). □

Remark 4.4.1. The convergence rate in terms of the number of particles in the above theorem can be improved to $1/2$ if the drift has the form (4.2), see [116].

The convergence rate in terms of number of time-steps established in Theorem 4.4.2 could be improved by employing exponential tail estimate techniques, as in [107]. The resulting strong convergence rate would be $1/4-\varepsilon$, for an arbitrarily small $\varepsilon>0$. However, to achieve this, one would need to assume boundedness of the coefficients in equation (4.3). Another possibility to recover a better convergence rate in our setting would be to require that the initial data $\xi\in L_0^p(\mathbb{R})$ for any $p\geq 2$ and that σ is uniformly bounded. This would enable us to obtain sharper estimates, when employing Markov's inequality in the proof of Proposition 4.4.3. If we assume moment boundedness of the initial data of any order, but allow σ to grow-linearly, we would obtain a rate of order $1/8-\varepsilon$.

Moreover, although we expect that the optimal convergence rate of the Euler–Maruyama scheme applied to the interacting particle system is $1/2$ (as for equations with Lipschitz coefficients), we only achieve the order $1/9$ (or $1/4-\varepsilon$ under stronger assumptions on the initial data or the coefficients of the underlying McKean–Vlasov SDE), due to the estimate of the probability that $X_{\eta(s)}^{i,N,M}$ and $X_s^{i,N,M}$ have a different sign, i.e., that the term $|G''(X_s^{i,N,M})-G''(X_{\eta(s)}^{i,N,M})|$ in the proof of Proposition 4.4.3

does not allow a Lipschitz type estimate. Refined estimates of the aforementioned expected sign change, as derived in [119] for one-dimensional SDEs, are not easy to prove for an interacting particle system. The proof of the main result in [119] is not applicable to our setting as an individual particle (seen as a one-dimensional equation) does not satisfy Markov properties (due to the dependence of interaction terms), which are key observations in [119].

4.4.3 Scheme 1 for the non-decomposable case

Here, we first prove a propagation of chaos result in the case of non-decomposable drifts in Lemma 4.4.1. The time-discretisation error is then given in Theorem 4.4.3.

Lemma 4.4.1. *Let Assumption (H.9) be satisfied, let $\xi \in L_0^p(\mathbb{R})$ for $p > 4$ and assume the mapping $\mathbb{R} \setminus \{0\} \times \mathcal{P}_2(\mathbb{R}) \ni (x, \mu) \mapsto b(x, \mu)$ to be uniformly bounded. Let $(X_t^i)_{t\in[0,T]}$ be the unique strong solution of (4.5) driven by the Brownian motion $(W_t^i)_{t\in[0,T]}$, and $(X_t^{i,N})_{t\in[0,T]}$ is the solution to the associated particle system. Then, there exists a constant $C > 0$ (independent of N) such that*

$$\max_{i\in\{1,\dots,N\}} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - X_t^{i,N}|^2\right] \leq CN^{-1/2}.$$

Proof. First, we observe, using the definitions $\mu_t = \mathcal{L}_{X_t}$, $\mu_t^N(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^N \delta_{X_t^j}(\mathrm{d}x)$, and $\mu_t^{\boldsymbol{X}^N,N-1}(\mathrm{d}x) = \frac{1}{N-1}\sum_{j\neq i} \delta_{X_t^{j,N}}(\mathrm{d}x)$, that

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - X_t^{i,N}|^2\right] &= \mathbb{E}\left[\sup_{t\in[0,T]} |G^{-1}(G(X_t^i,\mu_t),\mu_t) - G^{-1}(G(X_t^{i,N},\mu_t^{\boldsymbol{X}^N,N-1}),\mu_t^{\boldsymbol{X}^N,N-1})|^2\right] \\
&\leq C\mathbb{E}\left[\sup_{t\in[0,T]} |G(X_t^i,\mu_t) - G(X_t^{i,N},\mu_t^{\boldsymbol{X}^N,N-1})|^2\right] \\
&\quad + L(c)\sup_{t\in[0,T]}\left(\mathcal{W}_2^2(\mu_t,\mu_t^N) + \mathcal{W}_2^2(\mu_t^N,\mu_t^{\boldsymbol{X}^N,N-1})\right), \qquad (4.27)
\end{aligned}$$

where $L(c) \to 0$ as $c \to 0$, due to the proof of Proposition 4.3.3 (in particular, recall the definition of the constant $C_3 = 4C_2ce^{4C_1c}$). Furthermore, [31, Theorem 5.8] implies that $\mathcal{W}_2^2(\mu_t,\mu_t^N) \leq CN^{-1/2}$ and in addition, by triangle inequality, we deduce

$$\begin{aligned}
L(c)\mathcal{W}_2^2(\mu_t^N,\mu_t^{\boldsymbol{X}^N,N-1}) &\leq 2L(c)\left(\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - X_t^{i,N}|^2\right] + \mathcal{W}_2^2(\mu_t^{\boldsymbol{X}^N},\mu_t^{\boldsymbol{X}^N,N-1})\right) \\
&\leq 2L(c)\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - X_t^{i,N}|^2\right] + CN^{-1/2}, \qquad (4.28)
\end{aligned}$$

where we used $\mathcal{W}_2^2(\mu_t^{\boldsymbol{X}^N},\mu_t^{\boldsymbol{X}^N,N-1}) \leq CN^{-1}$, which follows from [146, Lemma 3.1].

To summarise, taking (4.27) and (4.28) into account, we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - X_t^{i,N}|^2\right] \leq C\left(\mathbb{E}\left[\sup_{t\in[0,T]} |G(X_t^i,\mu_t) - G(X_t^{i,N},\mu_t^{\boldsymbol{X}^N})|^2\right]\right.$$

$$+\mathbb{E}\left[\sup_{t\in[0,T]}|G(X_t^{i,N},\mu_t^{\boldsymbol{X}^N})-G(X_t^{i,N},\mu_t^{\boldsymbol{X}^{N,N-1}})|^2\right]+N^{-1/2}\Bigg).$$

A similar analysis to above can be employed to handle the second term and hence, we arrive at

$$\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^i-X_t^{i,N}|^2\right]\leq C\left(\mathbb{E}\left[\sup_{t\in[0,T]}|G(X_t^i,\mu_t)-G(X_t^{i,N},\mu_t^{\boldsymbol{X}^N})|^2\right]+N^{-1/2}\right). \tag{4.29}$$

To further estimate (4.29), we apply Itô's formula to derive

$$G(X_t^i,\mu_t)-G(X_t^{i,N},\mu_t^{\boldsymbol{X}^N})=:\sum_{j=1}^7\Pi_t^{j,i}.$$

where we introduced

$$\begin{aligned}
\Pi_t^{1,i}&:=G(X_0^i,\mu_0)-G(X_0^{i,N},\mu_0^{\boldsymbol{X}^N}),\\
\Pi_t^{2,i}&:=\int_0^t\left(b(X_s^i,\mu_s)+\alpha(\mu_s)\bar\phi'(X_s^i)b(X_s^i,\mu_s)+\frac{1}{2}\alpha(\mu_s)\bar\phi''(X_s^i)\sigma^2(X_s^i)\right)\mathrm{d}s\\
&\quad-\int_0^t\left(b(X_s^{i,N},\mu_s^{\boldsymbol{X}^N})+\alpha(\mu_s^{\boldsymbol{X}^N})\bar\phi'(X_s^{i,N})b(X_s^{i,N},\mu_s^{\boldsymbol{X}^N})+\frac{1}{2}\alpha(\mu_s^{\boldsymbol{X}^N})\bar\phi''(X_s^{i,N})\sigma^2(X_s^{i,N})\right)\mathrm{d}s,\\
\Pi_t^{3,i}&:=\int_0^t\left(\sigma(X_s^i)+\alpha(\mu_s)\bar\phi'(X_s^i)\sigma(X_s^i)-\sigma(X_s^{i,N})-\alpha(\mu_s^{\boldsymbol{X}^N})\bar\phi'(X_s^{i,N})\sigma(X_s^{i,N})\right)\mathrm{d}W_s^i,\\
\Pi_t^{4,i}&:=\int_0^t\Big(\partial_sG(X_s^i,\mu_s)-\frac{1}{2N}\sum_{k=1}^N\partial_y\partial_\mu G(X_s^{i,N},\mu_s^{\boldsymbol{X}^N})(X_s^{k,N})\sigma^2(X_s^{k,N})\\
&\qquad\qquad-\frac{1}{N}\sum_{k=1}^N\partial_\mu G(X_s^{i,N},\mu_s^{\boldsymbol{X}^N})(X_s^{k,N})b(X_s^{k,N},\mu_s^{\boldsymbol{X}^N})\Big)\mathrm{d}s,\\
\Pi_t^{5,i}&:=-\int_0^t\frac{1}{N}\sum_{k=1}^N\partial_\mu G(X_s^{i,N},\mu_s^{\boldsymbol{X}^N})(X_s^{k,N})\sigma(X_s^{k,N})\,\mathrm{d}W_s^k,\\
\Pi_t^{6,i}&:=-\int_0^t\frac{1}{2N^2}\sum_{k=1}^N\partial_\mu^2G(X_s^{i,N},\mu_s^{\boldsymbol{X}^N})(X_s^{k,N},X_s^{k,N})\sigma^2(X_s^{k,N})\,\mathrm{d}s,\\
\Pi_t^{7,i}&:=-\int_0^t\frac{1}{N}\partial_{x_i}\partial_\mu G(X_s^{i,N},\mu_s^{\boldsymbol{X}^N})(X_s^{i,N})\sigma^2(X_s^{i,N})\,\mathrm{d}s.
\end{aligned}$$

It is clear due to (H.9(3)) that $\mathbb{E}[|\Pi_t^{6,i}|^2]+\mathbb{E}[|\Pi_t^{7,i}|^2]\leq CN^{-1}$. For the term $\Pi_t^{5,i}$ we derive, using BDG's inequality,

$$\begin{aligned}
\mathbb{E}[|\Pi_t^{5,i}|^2]\leq C\Bigg(&\frac{1}{N}\sum_{k=1}^N\mathbb{E}\left[\int_0^t\left(\partial_\mu G(X_s^{i,N},\mu_s^{\boldsymbol{X}^N})(X_s^{k,N})\sigma(X_s^{k,N})-\partial_\mu\alpha(\mu_s)(X_s^k)\bar\phi(X_s^k)\sigma(X_s^k)\right)^2\mathrm{d}s\right]\\
&+\frac{1}{N^2}\sum_{k,l=1}^N\mathbb{E}\left[\left(\int_0^t\partial_\mu\alpha(\mu_s)(X_s^k)\bar\phi(X_s^k)\sigma(X_s^k)\,\mathrm{d}W_s^k\right)\left(\int_0^t\partial_\mu\alpha(\mu_s)(X_s^l)\bar\phi(X_s^l)\sigma(X_s^l)\,\mathrm{d}W_s^l\right)\right]\Bigg).
\end{aligned}$$

Therefore, taking the Lipschitz continuity of $(x,\mu)\mapsto\partial_\mu\alpha(\mu)(x)\bar\phi(x)\sigma(x)$ and the independence of $(X_t^k)_{k\in\{1,\dots,N\}}$, for $t\in[0,T]$, into account and using $\mathcal{W}_2(\mu_s^{\boldsymbol{X}^N},\mu_s)\leq\mathcal{W}_2(\mu_s^{\boldsymbol{X}^N},\mu_s^N)+\mathcal{W}_2(\mu_s^N,\mu_s)$ along with [31, Theorem 5.8], we obtain

$$\mathbb{E}[|\Pi_t^{5,i}|^2]\leq C\left(\int_0^T\mathbb{E}\left[\sup_{s\in[0,t]}|X_s^i-X_s^{i,N}|^2\right]\mathrm{d}t+N^{-1/2}\right).$$

Similar to Lemma 4.3.1 combined with BDG's inequality and Hölder's inequality, we deduce

$$\mathbb{E}[|\Pi_t^{1,i}|^2] + \mathbb{E}[|\Pi_t^{2,i}|^2] + \mathbb{E}[|\Pi_t^{3,i}|^2] + \mathbb{E}[|\Pi_t^{4,i}|^2] \leq C\left(\int_0^T \mathbb{E}\left[\sup_{s\in[0,t]} |X_s^i - X_s^{i,N}|^2\right] \mathrm{d}t + N^{-1/2}\right).$$

Therefore, inserting the estimates for $\mathbb{E}[|\Pi_t^{1,i}|^2], \dots, \mathbb{E}[|\Pi_t^{7,i}|^2]$ back into (4.29), we deduce the claim using Gronwall's inequality. □

The following hybrid explicit-implicit time-stepping algorithm computes a discrete-time approximation of $(X_t^{i,N})_{t\in[0,T]}$, denoted by $X_{t_n}^{i,N,M}$ for $n \in \{0,\dots,M\}$ and $i \in \{1,\dots,N\}$:

- Set $\tilde{X}_{t_0}^{i,N,M} = G(X_0^{i,N}, \mu_0^{\boldsymbol{X}^N})$ and $X_{t_0}^{i,N,M} = X_0^{i,N} = \xi^i$.
- For $n \geq 1$, compute

$$\tilde{X}_{t_n}^{i,N,M} = \tilde{X}_{t_{n-1}}^{i,N,M} + B_i(\tilde{X}_{t_{n-1}}^{1,N,M}, \dots, \tilde{X}_{t_{n-1}}^{N,N,M})h + \sum_{j=1}^N \Sigma^{i,j}(\tilde{X}_{t_{n-1}}^{1,N,M}, \dots, \tilde{X}_{t_{n-1}}^{N,N,M})\Delta W_n^j,$$

  where B_i and $\Sigma^{i,j}$ are defined by (4.18) and (4.19), respectively.
- Find $X_{t_n}^{i,N,M}$ such that $X_{t_n}^{i,N,M} = G^{-1}(\tilde{X}_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})$, with $\mu_{t_n}^{\boldsymbol{X}^{N,M}}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^N \delta_{X_{t_n}^{j,N,M}}(\mathrm{d}x)$.

Remark 4.4.2. The implicit function theorem applied to the function

$$\boldsymbol{F}_N(\boldsymbol{x}^N, \boldsymbol{y}^N) = \boldsymbol{y}^N - \left(G^{-1}(x_1, \mu^{\boldsymbol{y}^N}), \dots, G^{-1}(x_N, \mu^{\boldsymbol{y}^N})\right)^\top, \quad \boldsymbol{x}^N, \boldsymbol{y}^N \in \mathbb{R}^N,$$

implies that we can express $\boldsymbol{y}^N$ in terms of $\boldsymbol{x}^N$. The applicability of the implicit function theorem follows from similar arguments to the ones presented in Lemma 4.3.2 along with Proposition B.1.1.

Remark 4.4.3. We could also define an explicit scheme by setting $X_{t_n}^{i,N,M} = G^{-1}(\tilde{X}_{t_n}^{i,N,M}, \mu_{t_n}^{\tilde{\boldsymbol{X}}^{N,M}})$, with $\mu_{t_n}^{\tilde{\boldsymbol{X}}^{N,M}}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^N \delta_{\tilde{X}_{t_n}^{j,N,M}}(\mathrm{d}x)$. However, to derive a strong convergence rate for the resulting scheme, one has to analyse the quantity $\mathbb{E}[|X_{t_n}^{i,N,M} - \tilde{X}_{t_n}^{i,N,M}|^2]$. Similar arguments as for Scheme 2 could possibly be used here, but our current analysis does not allow us to derive an optimal convergence rate in h.

Theorem 4.4.3. *Let Assumption (H.9) be satisfied, let $\xi \in L_0^p(\mathbb{R})$ for $p > 4$ and assume the mapping $\mathbb{R}\setminus\{0\} \times \mathcal{P}_2(\mathbb{R}) \ni (x,\mu) \mapsto b(x,\mu)$ to be uniformly bounded. Let $(X_t^i)_{t\in[0,T]}$ be the unique strong solution of (4.5) driven by the Brownian motion $(W_t^i)_{t\in[0,T]}$, and $X_{t_n}^{i,N,M}$ for $n \in \{0,\dots,M\}$ be defined by the above algorithm. Then, there exists a constant $C > 0$ (independent of N and M) such that*

$$\max_{i\in\{1,\dots,N\}} \max_{n\in\{0,\dots,M\}} \mathbb{E}\left[|X_{t_n}^i - X_{t_n}^{i,N,M}|^2\right] \leq C(N^{-1/2} + h).$$

Proof. We start with the observation that for any $n \in \{0,\dots,M\}$

$$|X_{t_n}^{i,N} - X_{t_n}^{i,N,M}|^2 = |G^{-1}(G(X_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{X}^N,N-1}), \mu_{t_n}^{\boldsymbol{X}^N,N-1}) - G^{-1}(\tilde{X}_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})|^2$$

$$\leq 2|G^{-1}(G(X_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{X}^N,N-1}), \mu_{t_n}^{\boldsymbol{X}^N,N-1}) - G^{-1}(G(X_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{X}^N}), \mu_{t_n}^{\boldsymbol{X}^N})|^2$$
$$+ 2|G^{-1}(G(X_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{X}^N}), \mu_{t_n}^{\boldsymbol{X}^N,N}) - G^{-1}(\tilde{X}_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})|^2.$$

Using the arguments from the previous lemma, the first term can be shown to be of order $\mathcal{O}(N^{-1})$. For the second term, we derive the estimate

$$|G^{-1}(G(X_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{X}^N}), \mu_{t_n}^{\boldsymbol{X}^N}) - G^{-1}(\tilde{X}_{t_n}^{i,N,M}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})|^2$$
$$\leq C\left(|G(X_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{X}^N}) - \tilde{X}_{t_n}^{i,N,M}|^2 + L(c)\mathcal{W}_2^2(\mu_{t_n}^{\boldsymbol{X}^N}, \mu_{t_n}^{\boldsymbol{X}^{N,M}})\right),$$

where $L(c) \to 0$ as $c \to 0$, due to Proposition 4.3.3.

Therefore, we get

$$\mathbb{E}[|X_{t_n}^{i,N} - X_{t_n}^{i,N,M}|^2] \leq C\left(N^{-1} + \mathbb{E}[|G(X_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{X}^N}) - \tilde{X}_{t_n}^{i,N,M}|^2]\right).$$

Using now the definition of $\tilde{X}_{t_n}^{i,N,M}$ and Lemma 4.3.3, one shows that there exists a constant $C > 0$ such that $\mathbb{E}[|G(X_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{X}^N}) - \tilde{X}_{t_n}^{i,N,M}|^2] \leq Ch$. This together with Lemma 4.4.1 yields the claim. □

4.5 Numerical illustration

In the following, we will present two examples of McKean–Vlasov SDEs and interacting particle systems exhibiting discontinuous drifts in order to motivate the theoretical study of such equations and to numerically illustrate the strong convergence behaviour of an Euler–Maruyama scheme. These models find applications in biology and mathematical finance, in particular systemic risk.

As we do not know the exact solution of the considered equations, the convergence rate (in terms of number of time-steps) was determined by comparing two solutions computed on a fine and coarse grid, respectively, where the same Brownian motions were used. In order to illustrate the strong convergence behaviour in the uniform time-step h, we compute the root-mean-square error (RMSE) by comparing the numerical solution at level l of the time discretisation with the solution at level $l-1$ at the final time $T = 1$. To be precise, as error measure we use the quantity

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

where $M_l = 2^l T$ and by X_T^{i,N,M_l} we denote the approximation of X at time T computed based on N particles and $2^l T$ time-steps. The number of particles used in the tests will be specified below.

4.5.1 Neuronal interactions

In this section, we provide a numerical illustration for a specific model for neuronal interactions. Interacting particle systems are ubiquitous in neuroscience, such as the Hodgkin-Huxley model [11, 25] or mean-field equations describing the behaviour of a (large) network of interacting spiking

neurons [55]. For other mean-field models appearing in neuroscience, we refer to the references given in [55].

A recent model of the action potential of neurons is described in [60] and involves discontinuous coefficients. The reason for the necessity of introducing discontinuities is the following: After a charging phase of an individual neuron, subject to spikes of nearby neurons, randomness, and the effect of discharge with constant rate, the neuron emits a spike to the network once a certain threshold is hit and is then in a recovery phase. The change between these two phases is characterised by a discontinuity in the dynamics describing the potential of each neuron.

The action potential of N interacting neurons at time $t \in [0,T]$, $V_t^{i,N} (\mathrm{mod}\ 2) \in [0,2)$, where $V_t^{i,N} \in \mathbb{R}$, $i \in \{1,\dots,N\}$, is modelled by the discontinuous mean-field equation

$$\begin{aligned} \mathrm{d}V_t^{i,N} &= \lambda(V_t^{i,N}(\mathrm{mod}\ 2))\,\mathrm{d}t + \sigma^\varepsilon(V_t^{i,N}(\mathrm{mod}\ 2))\,\mathrm{d}W_t^i \\ &\quad + \frac{1}{N}\sum_{j=1}^N \Theta(\xi_i,\xi_j)\mathrm{I}_{[1,1+\kappa]}(V_t^{j,N}(\mathrm{mod}\ 2))\mathrm{I}_{[0,1]}(V_t^{i,N}(\mathrm{mod}\ 2))\,\mathrm{d}t, \end{aligned}$$

with square-integrable random initial values $V_0^{i,N} = \eta_i$, for $i \in \{1,\dots,N\}$ and $0 < \kappa < 1$ fixed. The set of i. i. d. random variables $\{\xi_1,\dots,\xi_N\}$, $\xi_i \in D$, describes the location of the N (non-moving) neurons, where D is modelled as an open connected domain of $\mathbb{R}^3$. Hence, the position of each neuron is given by $X_t^{i,N} = \xi^i$ at each time $t \in [0,T]$. It is further assumed that $\{\xi_1,\dots,\xi_N,\eta_1,\dots,\eta_N\}$ are independent for each integer $N \geq 1$. The standard Brownian motions $(W_t^i)_{t\in[0,T]}$ are independent, and independent of ξ_i and η_i, for $i \in \{1,\dots,N\}$. Observe that $V^{i,N}$ is specified by an SDE, where the drift has a discontinuity in the state variable (due to the choice of λ; see below for details), but also has a jump, in case a particle $j \neq i$ reaches the the critical values 1 or $1+\kappa$.

The following conditions are imposed in [60] to guarantee strong well-posedness of the particle system above (see, [60, Theorem 2.2]) and the associated McKean–Vlasov equation (see, [60, Theorem 6.1]); propagation of chaos type results, i.e., weak convergence of the law of the empirical distribution of $(\xi^i, V^{i,N})$ to a Dirac measure centred at the law of the solution to the underlying McKean–Vlasov equation, are shown in [60, Theorem 5.7]:

1. $\lambda(v) = -\hat{\lambda} v \mathrm{I}_{[0,1]}(v) + \mathrm{I}_{(1,2)}, \quad \hat{\lambda} > 0, \quad \Theta(x,y) = \sin(|x-y|)$, for $x, y \in \mathbb{R}^3$;

2. σ^ε is a $\mathcal{C}_b^1([0,2])$ function satisfying $\sigma^\varepsilon \geq \sqrt{2\varepsilon} > 0$ and

$$\sigma^\varepsilon(v) = \sqrt{2\varepsilon} \text{ on } [1,2], \quad \sigma^\varepsilon(2) = \sigma^\varepsilon(0) = \sqrt{2\varepsilon}, \quad (\sigma^\varepsilon)'(0) = (\sigma^\varepsilon)'(2) = 0, \text{ with } \varepsilon > 0 \text{ fixed.}$$

These conditions specify our Example 1. For Example 2, the diffusion term was chosen $\sigma^\varepsilon(x) = \sqrt{2\varepsilon} + x$. For our tests, we used $N = 10^3$. Furthermore, we set $\kappa = 0.01$, $\hat{\lambda} = 0.02$ and $\varepsilon = 0.1$. The initial values $\eta_1,\dots,\eta_N$ are chosen as independent normal random variables with mean 1 and standard deviation 2. Also, these values are considered modulo 2. The variables $\xi_1,\dots,\xi_N$ are independent

three-dimensional random variables chosen from the same multivariate normal distribution with some given mean vector and covariance matrix.

We investigate numerically the convergence of Scheme 2, i.e., the Euler–Maruyama scheme without applying any transformations. In Fig. 4.1, we observe strong convergence of order 3/4 for Example 1, which is most likely due to the choice of $\sigma^\varepsilon = \sqrt{2\varepsilon}$ as constant. In [120] a Milstein scheme for one-dimensional SDEs with discontinuities in the drift was derived and a strong convergence order of 3/4 was proven. In addition, it is conjectured in [120] (Conjecture 1 and Conjecture 2) that the rate 3/4 is optimal.

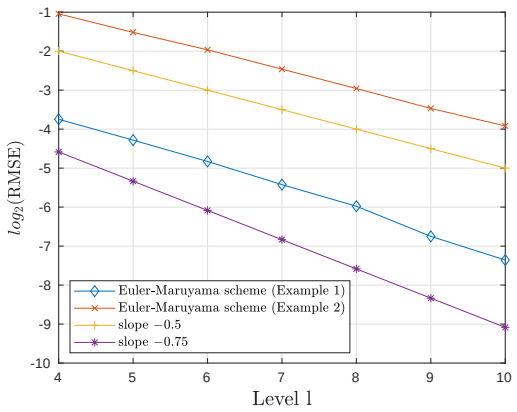


Figure 4.1: Strong convergence of the Euler–Maruyama scheme applied to the particle system obtained by approximating the equation for the action potential of the neurons.

4.5.2 Systemic risk

In this section, we consider a McKean–Vlasov SDE of the form

$$\mathrm{d}X_t = \left(a\left(\mathbb{E}[X_t] - X_t\right) + \kappa_1 \mathrm{I}_{\{X_t \leq 0\}} + \kappa_2 \mathrm{I}_{\{X_t > 0\}}\right) \mathrm{d}t + (\sigma + X_t)\, \mathrm{d}W_t, \quad X_0 = x \in \mathbb{R}, \tag{4.30}$$

where $a \geq 0$ is the mean-reversion rate, $\kappa_1 < 0$, $\kappa_2 > 0$ and $\sigma > 0$. The strong well-posedness of (4.30) follows from Proposition 4.3.1. This equation can be linked to a model of systemic risk in [36], where a mean-field game of N banks borrowing from, and lending to, a central bank is proposed. The banks control the rate of their borrowing depending on their (log-)monetary reserves, which are modelled by a system of SDEs with interaction through their average. In this setting flocking, and thus systemic default events, may occur. Here, we give a slight reformulation of this problem following [72, Section 4]. The problem consists in finding $(\hat{\mu}_t, \hat{\beta}_t)_{t\in[0,T]}$, where $(\hat{\mu}_t)_{t\in[0,T]}$ is a flow of measures in $\mathcal{C}([0,T], \mathcal{P}_2(\mathbb{R}))$ and $(\hat{\beta}_t)_{t\in[0,T]}$ is an adapted, square-integrable control process (describing the rate of borrowing from, or lending to, the central bank), such that $(\hat{\beta}_t)_{t\in[0,T]}$ minimises the objective function given by

$$J^{\hat{\mu}}(x;\beta) = \mathbb{E}\left[\int_0^T \left(r|\beta_t| + \frac{\varepsilon}{2}\left(X_t^{\hat{\mu},\beta} - \int_{\mathbb{R}} x\, \hat{\mu}_t(\mathrm{d}x)\right)^2\right) \mathrm{d}t + \frac{c}{2}\left(X_T^{\hat{\mu},\beta} - \int_{\mathbb{R}} x\, \hat{\mu}_T(\mathrm{d}x)\right)^2\right],$$

for $r, \varepsilon, c \geq 0$, where

$$\mathrm{d}X_t^{\hat{\mu},\beta} = \left(a \left(\int_{\mathbb{R}} x \, \hat{\mu}_t(\mathrm{d}x) - X_t^{\hat{\mu},\beta} \right) + \beta_t \right) \mathrm{d}t + (\sigma + X_t^{\hat{\mu},\beta}) \, \mathrm{d}W_t, \quad X_0^{\hat{\mu},\beta} = x \in \mathbb{R}, \tag{4.31}$$

and $\hat{\mu}_t = \mathcal{L}_{X_t^{\hat{\mu},\hat{\beta}}}$ for all $t \in [0, T]$.

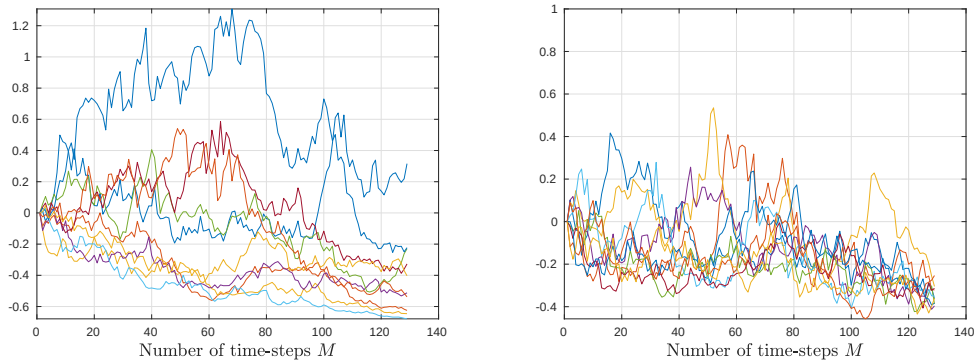
For simplicity, we did not add the common noise term as in [36]. In addition, we modified the diffusion to allow it to be degenerate. In [72], a constraint $\beta_t \in [\kappa_1, \kappa_2]$ on the borrowing/lending rate is imposed for all $t \in [0, T]$. The minimiser of the objective function for this mean-field game (with constant diffusion term) is shown to be a control of bang-zero-bang type (see [72, equation (24)] for an analytic expression). Written in feedback form, this optimal control strategy has discontinuities that are time-dependent, i.e., the zero-control region changes over time, a setting not covered by our current analysis.

Using instead the special bang-bang type control $(\beta_t^*)_{t \in [0,T]}$ of the form

$$\beta_t^* = \begin{cases} \kappa_1, & \text{if } X_t \leq 0 \\ \kappa_2, & \text{if } X_t > 0, \end{cases}$$

and plugging β_t^* back into (4.31) results into an equation of the form (4.30) with a one-point discontinuity.

In our numerical experiments, we set $\kappa_1 = -0.5$, $\kappa_2 = 0.5, x = 0$ and $\sigma = 0.7$. Further, we consider three different choices for the mean-reversion rate, i.e., we set $a = 1, 5, 10$. The expected value is approximated by the empirical mean of $N = 10^4$ particles. For a larger value of a, we expect sample paths of (4.30) to be more concentrated around the mean of the samples; see Fig. 4.2a and 4.2b (with $T = 1$ and $M = 2^7$). We can also observe that for a stronger concentration effect the strong approximation behaviour becomes better; see Fig. 4.2b.



(a) (b)

Figure 4.2: Sample trajectories of the particle system associated with (4.30) for $N = 10$ and $a = 1$ (left) and $a = 10$ (right).

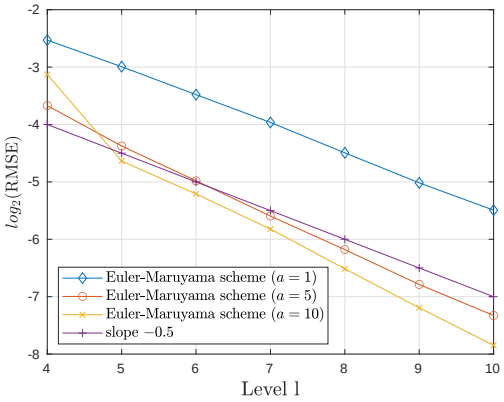


Figure 4.3: Strong convergence of the Euler–Maruyama scheme applied to the particle system obtained by approximating the equation (4.30) with $a = 1, 5, 10$.

Chapter 5

Numerical methods for mean-field control problems

5.1 Introduction

For a given terminal time $T > 0$, we consider the following controlled McKean–Vlasov SDE: Let $t \in [0, T]$, $X_0^\alpha = \xi$, for some random variable $\xi \in L_0^2(\mathbb{R}^d)$, and

$$\mathrm{d}X_t^\alpha = b(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha, \alpha_t)})\,\mathrm{d}t + \sigma(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha, \alpha_t)})\,\mathrm{d}W_t, \tag{5.1}$$

where b and σ are given measurable, and regular enough functions taking values in $\mathbb{R}^d$ and $\mathbb{R}^{d\times m}$, respectively, and $W = (W_t)_{t\in[0,T]}$ is an m-dimensional standard Brownian motion. The control process $\alpha = (\alpha_t)_{t\in[0,T]}$ is assumed to take values in $\mathbb{R}^k$. Here, and throughout this chapter, $\mathcal{L}_U$ denotes the law of a given random variable U.

The value function V and loss functional J of the mean-field control (MFC) problem are defined by

$$V(\xi) = \inf_{\alpha\in\mathcal{H}^2(\mathbb{R}^k)} J(\alpha;\xi), \quad \text{with } J(\alpha;\xi) = \mathbb{E}\bigg[\int_0^T \{f(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)}) + \ell(\alpha_t)\}\,\mathrm{d}t + g(X_T^\alpha, \mathcal{L}_{X_T^\alpha})\bigg], \tag{5.2}$$

where $\mathcal{H}^2(\mathbb{R}^k)$ denotes the set of square integrable $\mathbb{F}$-progressively measurable processes taking values in $\mathbb{R}^k$, i.e., the set of admissible open-loop controls. The running cost f and terminal cost g are given real-valued, continuously differentiable functions of at most quadratic growth. Further, $\ell : \mathbb{R}^k \to \mathbb{R}$ is a proper, lower semicontinuous and convex function, where the associated proximal map is assumed to be known in explicit form. Typical examples for ℓ include the ℓ_1-norm or the indicator function, which is helpful to enforce sparsity of the control representation and is necessary in situations where the control can only attain values in a certain convex subset of $\mathbb{R}^k$.

In many existing works, the study of (5.1) and (5.2) is limited to the case where the coefficients only depend on the law of the controlled state, but not on the law of the control, see, e.g., [41, 32, 31, 1, 34, 102, 126]. Recently, the more general framework described by (5.1) and (5.2) (without the non-smooth functional ℓ) has received significant attention, see, e.g., [127, 1, 69, 152, 135, 71].

Mean-field games (MFGs), introduced in [79, 101], and MFC problems are both concerned with the study of the asymptotic limit of a large system of interacting symmetric agents (N-player game). In contrast to MFGs, which correspond to the setting in which all agents aim to minimise their individual loss (Nash equilibrium), a MFC problem is characterised by a central planner who tries to optimise the social utility by prescribing a control common to all agents (Pareto optimality). However, note that control problems of the form (5.1) and (5.2) do not necessarily have to occur as the limit of a huge system of interacting players. For instance, they appear when the goal is to make the optimal state process and/or control process less sensitive to the possible variation of the random events [152], or in a smart grid management model [69]. In these two applications no direct association to an N-player game can be given.

As explicit solutions to (5.1) and (5.2) are rarely available, numerical methods to find an approximation for the optimal control and the value function are essential. Due to the non-Markovian nature of McKean–Vlasov SDEs, it is generally preferred to approach the MFC problem by employing a version of the Pontryagin maximum principle [32, 1], rather than the dynamic programming principle, as the underlying Hamilton–Jacobi–Bellman (HJB) equations are defined on the enlarged (infinite-dimensional) state space $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, see e.g., [127].

We continue our literature review on related works with [34], which presents two probabilistic methods to find approximations of (5.2) based on machine learning techniques. The first method in [34], often referred to as the policy gradient (PG) method, restricts the optimisation in (5.2) over feedback controls in parametric form, i.e., it is assumed that the control is a deterministic function of time and the controlled state process (the so-called feedback map) depending on a set of parameters θ (e.g., neural networks). To update a given feedback map, one performs a stochastic gradient descent (SGD) step of the loss functional J with respect to θ based on forward samples of the controlled state dynamic (5.1). Although the PG method is straightforward to implement and typically yields an accurate value function approximation, the numerically obtained feedback controls neglect in many cases essential structural properties of the optimal feedback control and do not allow to give a correct interpretation of the feedback behaviour. Note that in this approach the control problem is only solved along optimal state trajectories, which implies that the PG method often fails to recover the optimal feedback control outside the support of these sample trajectories. As a consequence, a pre-computed feedback control, based on the PG method with some given initial state ξ, may perform poorly when applied to (5.1) and (5.2) if ξ is perturbed. This lack of robustness is a common drawback of purely sampling based methods. Apart from these practical limitations, we further remark that, due to the the lack of convexity of the functional J in the parameters of the set θ, convergence of the PG method is, in general, not guaranteed.

The second approach relies on the numerical approximation of the optimality system arising from the stochastic Pontryagin maximum principle applied to the control problem (5.1) and (5.2).

Here, one reformulates the control problem as a coupled mean-field forward backward stochastic differential equation (FBSDE) and then employs, e.g., the Deep BSDE solver to obtain a solution for the adjoint equations [76]. This approach encounters the following limitations: First, one assumes that it is possible to solve a certain first order optimality condition analytically, i.e., to represent the optimal feedback control by a function $\hat{\alpha}$ depending on time, the forward process (5.1) and the adjoint processes, which then allows to rewrite the control problem as coupled FBSDE in first place. In particular, the analytic expression of $\hat{\alpha}$ is obtained as the pointwise minimiser of a Hamiltonian. However, in many cases this analytic form may not be available, specifically if the coefficients of (5.1) and (5.2) depend on the law of the control or if the cost is non-smooth. In the situation, where we have mean-field interaction through the law of the control, we additionally encounter the difficulty that $\hat{\alpha}$, in general, cannot be found as the pointwise minimiser of a Hamiltonian [1]. Second, we reiterate that probabilistic methods, including regression methods and the Deep BSDE solver, often suffer from the issue that the numerically computed feedback control only performs well along the optimal state trajectories. Therefore, these methods might be very sensitive to small perturbations of model parameters and the initial data.

For further numerical methods concerning the approximation of coupled mean-field FBSDEs, we refer to [133, 135, 8, 40]. We also want to mention the article [12] considering a family of MFC problems, which can be reformulated as finite-dimensional control problems. Probabilistic methods, e.g., regression Monte–Carlo (MC), are then proposed to compute certain conditional expectations appearing in this framework. The recent papers [138] and [62] propose machine learning techniques to numerically solve deterministic potential MFGs and MFC problems with point delays, respectively. The article [62] introduces two algorithms (in a similar flavour as [34]): One where the feedback control is directly parametrised using neural networks in order to minimise the value function using SGD iterations, and a second method based on a McKean–Vlasov forward anticipated BSDE system.

Deterministic methods (e.g., based on finite difference methods or neural networks) to solve MFGs and MFC problems have been developed in e.g., [3, 4, 5, 6, 33] relying on a coupled system of PDEs derived in [22], where the control is assumed to be represented by a smooth enough feedback map and found as the (pointwise) minimiser of a Hamiltonian. This approach can be viewed as a deterministic formulation of the Pontryagin maximum principle. To be precise, the aforementioned system of PDEs comprises of a forward Fokker–Planck equation for the density (with respect to the Lebesgue measure) of the controlled state dynamics and a backward HJB equation for the value function. An obvious drawback of finite difference methods is that they cannot be applied to solve high-dimensional control problems, due to the curse of dimensionality. Further, as in the probabilistic formulation, we again face the issue that for a large class of control problems the feedback map cannot be obtained analytically as the pointwise minimiser of a Hamiltonian. Additionally, in particular in situations where the diffusion is degenerate or the density of the initial

state is singular, solutions to the Fokker–Planck may not exist or exhibit irregularities, which are numerically difficult to approximate especially in a high-dimensional framework.

We propose a method, which will overcome the aforementioned difficulties. Here, we present an iterative scheme motivated by Nesterov's accelerated proximal gradient method that generates a sequence of feedback maps $(\phi^m)_{m\in\mathbb{N}}$, with $\phi^m : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k$ and hence a sequence of feedback controls given by $\alpha_t^m = \phi^m(t, X_t^{\phi^m})$, for $t \in [0,T]$, where X^{ϕ^m} is the state process controlled by α^m. We remark that from a practical perspective it is preferable to seek for an optimal control representation in feedback form, as (i) only a simple function evaluation is required to implement a control strategy and (ii) the feedback structure allows to give a concrete interpretation of the controllers action based on the current state of the system. Now, for a given representation of the control, we update this feedback map by employing an accelerated proximal gradient descent step, achieved by computing the gradient of the smooth part of the loss functional with respect to the control, in combination with the application of a proximal operator to handle the non-smooth part and an additional momentum step. Compared to the approaches above, this iterative procedure has the following main advantages: We do not rely on an analytic expression for the feedback map obtained by the pointwise minimisation of a Hamiltonian and do not have to find the optimality system arising from the stochastic Pontryagin maximum principle. Therefore, our scheme is applicable to a broader class of control problems, in particular those involving mean-field interaction through the control. The main steps of a practically implementable version of our algorithm can summarised in the following manner:

For a given feedback map, we are able to forward simulate (5.1) relying on a particle approximation for the law of the state and control and an Euler–Maruyama scheme. Then, we evaluate the gradient of the smooth part of the cost functional, which requires solving a certain mean-field BSDE taking the given controlled state (and its distribution) as input. Note that since the law of (5.1) is estimated using a particle approximation, we avoid solving a Fokker–Planck PDE for the density, which might not exist or may be challenging to solve numerically (in particular in high-dimensions). In order to solve the mean-field BSDE, we employ a PDE approach and also propose a regression MC algorithm (probabilistic approach) in a low-dimensional setting to illustrate drawbacks of such methods. To make our algorithm amenable to a high-dimensional framework, we employ a Deep Galerkin (DG) method, see [143], to obtain an approximate solution to this mean-field BSDE. To be precise, we aim to solve a system of high-dimensional PDEs for the decoupling fields of the adjoint processes. Note that the DG method is a sampling based algorithm and compared to finite difference (or finite element) methods it is not necessary to construct a mesh for the (high-dimensional) computational domain. As the training samples are drawn uniformly from the computational domain, we recover the correct feedback behaviour on the entire computational domain and not only along samples trajectories of the controlled state process.

The remainder of this chapter is organised as follows: Section 5.2 will describe the main algorithm of this chapter in a generic manner and then introduces a PDE based formulation of our method. This provides the basis for a concrete implementation of the method relying on deep learning methods in combination with a residual approximation technique, which is detailed in Section 5.3. The performance and effectiveness of this novel approach is illustrated in Section 5.4.2 for the so-called Cucker–Smale (C–S) model (see Section 5.4.1 for details) appearing in crowd modelling. We devote Appendix C.1 to explain in detail the stochastic maximum principle for MFC problems (with and without mean-field interaction through the control), and illustrate the connection of the control problem (5.1) and (5.2) with mean-field FBSDEs. Further, in Appendix C.2, we present a regression MC algorithm which is suitable for low-dimensional control problems of the form (5.1) and (5.2). We will apply this method to a portfolio liquidation problem with transaction costs, which can be formulation as non-smooth MFC problem. This appendix also serves to point out drawbacks encountered when solving (5.1) and (5.2) relying on a pure probabilistic method.

We end this section by fixing the notation and by introducing several notions needed throughout this chapter.

Preliminaries:

- We denote by $T > 0$ a given terminal time, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ is a given complete atomless probability space on which an m-dimensional Brownian motion $W = (W_t)_{t\in[0,T]}$ is defined and $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ is the natural filtration of W augmented with an independent σ-algebra $\mathcal{F}_0$.

- Let now $(E, |\cdot|)$ be some given Euclidean space, where $|\cdot|$ is induced by the usual inner product $\langle\cdot,\cdot\rangle$. We use $\mathcal{P}(E)$ to denote the family of all probability measures on $(E, \mathcal{B}(E))$, where $\mathcal{B}(E)$ denotes the Borel σ-field over E, and define the subset of probability measures with finite second moment by $\mathcal{P}_2(E)$. As metric on the space $\mathcal{P}_2(E)$, we use the Wasserstein distance.

- For a given functional $f : \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^k) \to \mathbb{R}$ and a measure $\eta \in \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^k)$ with first and second marginals $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\nu \in \mathcal{P}_2(\mathbb{R}^k)$, respectively, the mapping $\partial_\eta f(\eta)(\cdot) : \mathbb{R}^d\times\mathbb{R}^k \to \mathbb{R}^d\times\mathbb{R}^k$ will denote the Lions derivative (L-derivative) of f at η and $(\partial_\mu f(\eta), \partial_\nu f(\eta))(\cdot) : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^d\times\mathbb{R}^k$ are referred to as the partial L-derivatives with respect to the marginals, see [1, Section 2.1] for a detailed introduction on the notion of partial L-derivatives.

- Let $\tau > 0$ be a given real number. We denote by $\mathrm{prox}_{\ell,\tau}(x) := \arg\min_{y\in\mathbb{R}^k}\left(\ell(y) + \frac{1}{2\tau}|y - x|^2\right)$ for $x \in \mathbb{R}^d$ the so-called proximal operator for a proper, convex and lower semicontinuous function $\ell : \mathbb{R}^k \to \mathbb{R}$. Note that in the case $\ell(a) = \gamma|a|$, for $\gamma > 0$, the proximal map has the explicit form

$$\mathrm{prox}_{\ell,\tau}(x) = \begin{cases} \mathrm{sgn}(x)(|x| - l\tau), & |x| \geq \gamma\tau, \\ 0, & |x| \leq \gamma\tau. \end{cases}$$

5.2 Fast iterative proximal gradient method for MFC problems

We start this section by introducing the Hamiltonian and the adjoint equations associated to the control problem given by (5.1) and (5.2). Let $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^k \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^k) \times \mathbb{R}^d \times \mathbb{R}^{d\times m} \to \mathbb{R}$ be the Hamiltonian of (5.2) defined as follows:

$$H(t,x,a,\eta,y,z) := \langle b(t,x,a,\eta), y\rangle + \langle \sigma(t,x,\eta), z\rangle + f(t,x,a,\eta).$$

Now, for any initial state $\xi \in L_0^2(\mathbb{R}^d)$ and any given admissible control $\alpha \in \mathcal{H}^2(\mathbb{R}^k)$, let $X^\alpha = (X_t^\alpha)_{t\in[0,T]} \in \mathcal{S}^2(\mathbb{R}^d)$ be the corresponding controlled process solving (5.1), and let the square-integrable $\mathbb{F}$-progressively measurable processes $(Y^\alpha, Z^\alpha) = (Y_t^\alpha, Z_t^\alpha)_{t\in[0,T]}$ be the so-called adjoint processes, which satisfy the following mean-field BSDE: For all $t \in [0,T]$,

$$\begin{aligned} \mathrm{d}Y_t^\alpha &= -\big(\partial_x H(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)}, Y_t^\alpha, Z_t^\alpha) \\ &\qquad + \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{X}_t^\alpha, \tilde{\alpha}_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)}, \tilde{Y}_t^\alpha, \tilde{Z}_t^\alpha)(X_t^\alpha, \alpha_t)]\big)\,\mathrm{d}t + Z_t^\alpha\,\mathrm{d}W_t, \\ Y_T^\alpha &= \partial_x g(X_T^\alpha, \mathcal{L}_{X_T^\alpha}) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T^\alpha, \mathcal{L}_{X_T^\alpha})(X_T^\alpha)], \end{aligned} \tag{5.3}$$

where we used here the tilde notation to refer to an independent copy of a random variable.

To motivate the main steps of our algorithm in a generic form, we note that (5.1) and (5.2) can be viewed as the non-smooth optimisation problem given by

$$\inf_{\alpha\in\mathcal{H}^2(\mathbb{R}^k)} (F(\alpha) + G(\alpha)),$$

where the functional F and G, for $\alpha \in \mathcal{H}^2(\mathbb{R}^k)$, are defined as

$$F(\alpha) := \mathbb{E}\bigg[\int_0^T f(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)})\,\mathrm{d}t + g(X_T^\alpha, \mathcal{L}_{X_T^\alpha})\bigg], \quad G(\alpha) := \mathbb{E}\bigg[\int_0^T \ell(\alpha_t)\,\mathrm{d}t\bigg].$$

It has been shown in [1, Lemma 3.1] that the Fréchet derivative of F, denoted by $\partial_\alpha F(\alpha) : \mathcal{H}^2(\mathbb{R}^k) \to \mathcal{H}^2(\mathbb{R}^k)$, is given by

$$\begin{aligned} \partial_\alpha F(\alpha_t) := (\partial_\alpha F(\alpha))_t &= \partial_\alpha H(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)}, Y_t^\alpha, Z_t^\alpha) \\ &\quad + \tilde{\mathbb{E}}[\partial_\nu H(t, \tilde{X}_t^\alpha, \tilde{\alpha}_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)}, \tilde{Y}_t^\alpha, \tilde{Z}_t^\alpha)(X_t^\alpha, \alpha_t). \end{aligned}$$

This means, that in the gradient step of our algorithm, we compute the derivatives of the Hamiltonian with respect to the control and its law and evaluate the gradient at the given control guess, α^m, and the associated triple $(X^{\alpha^m}, Y^{\alpha^m}, Z^{\alpha^m})$. We do not only take a standard gradient descent step, but also employ a momentum update with a suitable choice for the momentum parameter, to accelerate the convergence of standard gradient descent algorithms. Further, we allow for non-smooth penalty terms in the cost function, relying on a proximal operator associated to G, which is explicitly known for several important cases such as the ℓ_1-regularisation, see e.g., [18]. We detail the steps of

our procedure in Algorithm 1 below. This iteration is known as Nesterov's (proximal) accelerated gradient (NAG) method [121], which has an $\mathcal{O}(1/\sqrt{\varepsilon})$ iteration complexity for convex problems with an L-smooth functional F (compared to the $\mathcal{O}(1/\varepsilon)$ complexity for the standard gradient descent algorithm), if $\varepsilon > 0$ is the desired accuracy. This is the optimal rate achieved by gradient methods using only one gradient evaluation at each iteration. Here, L denotes the Lipschitz constant for the gradient of the smooth functional F and the step-size $\tau > 0$ in the gradient step needs to satisfy $\tau \in [0, 1/L]$.

In the algorithm presented below, we will introduce the convex version of the NAG method with momentum parameter $\frac{m}{m+3}$, where m denotes the iteration index. This avoids, compared to a strong convex version of the NAG method, the estimation of the strong convexity parameter, which can be difficult in practical applications. We recall that taking a gradient step has the advantage that we do not need to solve the first order optimality condition analytically, and our method is therefore preferable to algorithms such as the Method of Successive Approximations discussed in, e.g., [83, 93].

Algorithm 1 Non-Markovian NAG algorithm

Input: initial guess for the control $\alpha^0 = (\alpha_t^0)_{t\in[0,T]} \in \mathcal{H}^2(\mathbb{R}^k)$ and NAG initialisation $\beta_t^0 = \alpha_t^0$, $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$ a.e.; number of NAG iterations N_{nag}; step-size $\tau > 0$

Output: approximation of value function and optimal control

1: **for** $m = 0, \dots, N_{nag} - 1$ **do**

2:

3: For the given control solve a decoupled FBSDE, i.e., find $(X_t^m)_{t\in[0,T]}$ as the solution of

$$\mathrm{d}X_t^m = b(t, X_t^m, \beta_t^m, \mathcal{L}_{(X_t^m,\beta_t^m)})\,\mathrm{d}t + \sigma(t, X_t^m, \beta_t^m, \mathcal{L}_{(X_t^m,\beta_t^m)})\,\mathrm{d}W_t, \quad X_0 = \xi \in L_0^2(\mathbb{R}^d), \tag{5.4}$$

and solve the adjoint equations given by (5.3) for $(Y_t^m, Z_t^m)_{t\in[0,T]}$, where α and X^α are replaced by β^m and X^m, respectively.

4: Compute the updates

$$\alpha_t^{m+1} = \mathrm{prox}_{\ell,\tau}\left(\beta_t^m - \tau\partial_\alpha F(\alpha_t^m)\right)$$
$$\beta_t^{m+1} = \alpha_t^{m+1} + \frac{m}{m+3}(\alpha_t^{m+1} - \alpha_t^m),$$

$\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$ a.e., and return to the first step.

5: **end for**

6: **return** $\alpha^{N_{nag}}$, $J(\alpha^{N_{nag}})$.

The above presented algorithm is from a computational perspective impractical, due to its non-Markovian nature. It is easy to see that the cost to find an estimate for Y^m increases with the number of NAG iterations. To be precise, Y^0 (and hence β^1) is a function of X^0 and, consequently, the coefficients of the state process X^1 depend on X^0 (through β^1) and X^1. Repeating this iteration, one observes that Y^m (hence β^{m+1} and α^{m+1}) are functions depending on $(X^0, \dots, X^m)$, which becomes computationally intractable as the number of iterations increases. Note that a similar

observation has been made in [21], where an iterative method for non-mean-field coupled FBSDEs is considered.

In what follows, we will introduce the main algorithm of this chapter, which compared to the previous procedure will not suffer from a significant increase in the complexity as the number of NAG iterations increases and replaces (5.3) by a system of PDEs associated to the decoupling fields for the adjoint processes (Y^m, Z^m).

In our numerical procedure, α^m and β^m are represented in feedback form, i.e., at each iteration the controls will be characterised by deterministic mappings of the form $\phi^m, \psi^m : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k$, which satisfy $\alpha_t^m = \phi^m(t, X_t^m)$ and $\beta_t^m = \psi^m(t, X_t^m)$, for $t \in [0,T]$. Therefore, at the m-th NAG iteration, the controlled state dynamic (5.4) now takes the form

$$\mathrm{d}X_t^m = b(t, X_t^m, \psi^m(t, X_t^m), \mathcal{L}_{(X_t^m, \psi^m(t, X_t^m))})\,\mathrm{d}t + \sigma(t, X_t^m, \psi^m(t, X_t^m), \mathcal{L}_{(X_t^m, \psi^m(t, X_t^m))})\,\mathrm{d}W_t, \quad (5.5)$$

where $X_0^m = \xi \in L_0^2(\mathbb{R}^d)$.

Hence, the adjoint equations can be expressed as

$$\begin{aligned} \mathrm{d}Y_t^m &= -f^m(t, X_t^m, Y_t^m, Z_t^m)\,\mathrm{d}t + Z_t^m\,\mathrm{d}W_t, \\ Y_T^m &= h^m(X_T^m), \end{aligned}$$

where

$$\begin{aligned} f^m(t, X_t^m, Y_t^m, Z_t^m) &= \partial_x H(t, X_t^m, \psi^m(t, X_t^m), \mathcal{L}_{(X_t^m, \psi^m(t, X_t^m))}, Y_t^m, Z_t^m) \\ &\quad + \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{X}_t^m, \psi^m(t, \tilde{X}_t^m, \mathcal{L}_{(X_t^m, \psi^m(t, X_t^m))}, \tilde{Y}_t^m, \tilde{Z}_t^m)(X_t^m, \psi_t^m)], \end{aligned} \quad (5.6)$$

and

$$h^m(X_T^m) = \partial_x g(X_T^m, \mathcal{L}_{X_T^m}) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T^m, \mathcal{L}_{X_T^m})(X_T^m)].$$

Now, note that, when solving the backward part (Y^m, Z^m) in Algorithm 1 with X^m given (with $(\alpha_t^m, \beta_t^m)_{t\in[0,T]}$ replaced by $(\phi^m(t, X_t^m), \psi^m(t, X_t^m))_{t\in[0,T]}$), we in fact aim to determine deterministic mappings (the so-called decoupling fields) $u^m : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $v^m : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, satisfying $Y_t^m = u^m(t, X_t^m)$ and $Z_t^m = v^m(t, X_t^m)$, $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$ a.e., respectively. Unlike in the case of classical FBSDEs, the decoupling fields for mean-field FBSDEs are functions of time and the state and the law of the forward process, see [31, Lemma 4.25]. As a consequence, the decoupling field needs to be found as the solution of an infinite-dimensional PDE (master equation) written on the enlarged state space $[0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. Here, we will solve PDEs for the decoupling fields only along a flow of measures and treat a given measure flow as additional time-dependence in u and v.

Assuming u^m to be sufficiently smooth, we expect that this function solves the following system of PDEs (suppressing the NAG iteration index m for u and v): $u(T,x) = h^m(x)$, and for all $t \in [0,T)$,

$$\partial_t u + \mathscr{L}^m u = -f^m(t, x, u(t,x), v(t,x)), \quad (5.7)$$

where $v(t,x) = (\partial_x u)(t,x)\sigma^m(t,x)$ and the operator $\mathscr{L}$ satisfies for each $\varphi = (\varphi_1, \dots, \varphi_d) \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$, $i \in \{1, \dots, d\}$ and $(t,x) \in [0,T] \times \mathbb{R}^d$ that

$$(\mathscr{L}^m \varphi)_i(t,x) = \frac{1}{2}\mathrm{tr}(\sigma^m(t,x)\sigma^{m,\mathsf{T}}(t,x)(\partial_x^2 \varphi_i)(t,x)) + \langle b^m(t,x), (\partial_x \varphi_i)(t,x)\rangle,$$

with b^m and σ^m defined by

$$b^m(t,x) := b(t,x,\psi^m(t,x), \mathcal{L}_{(X_t^m, \psi^m(t,X_t^m))}), \text{ and,}$$
$$\sigma^m(t,x) := \sigma(t,x,\psi^m(t,x), \mathcal{L}_{(X_t^m, \psi^m(t,X_t^m))}).$$

The next section will be devoted to present a concrete method to solve a system of PDEs of the form (5.7) in a high-dimensional setting. Having a solution for (u^m, v^m) at hand, the feedback maps ϕ^m and ψ^m will be updated employing the NAG method presented in Algorithm 1. The gradient of F evaluated at $(\psi(t, X_t^m))_{t\in[0,T]}$ satisfies

$$\begin{aligned}\partial_\alpha F(\psi^m(t,X_t^m)) &= \partial_\alpha H(t, X_t^m, \psi^m(t,X_t^m), \mathcal{L}_{(X_t^m,\psi^m(t,X_t^m))}, u^m(t,X_t^m), v^m(t,X_t^m)) \\ &+ \tilde{\mathbb{E}}[\partial_\nu H(t, \tilde{X}_t^m, \psi^m(t,\tilde{X}_t^m), \mathcal{L}_{(X_t^m,\psi^m(t,X_t^m))}, u^m(t,\tilde{X}_t^m), v^m(t,\tilde{X}_t^m))(X_t^m, \psi^m(t,\tilde{X}_t^m)),\end{aligned} \tag{5.8}$$

which enables to interpret $\partial_\alpha F$ as a function $\partial_\alpha F(\psi(\cdot,\cdot)) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k$.

Assuming for now that we have a PDE solver at hand, we are then in the position to present our PDE based accelerated gradient descent algorithm, referred to as fast iterative PDE based (FIPDE) algorithm for MFC problems:

Algorithm 2 FIPDE algorithm for MFC problems

Input: NAG initialisation for the feedback mappings $\phi^0, \psi^0 : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k$, where $\psi^0 \equiv \phi^0$; number of NAG iterations N_{nag}; step-size $\tau > 0$

Output: approximation of value function; optimal feedback map

1: **for** $m = 0, \dots, N_{nag} - 1$ **do**

2:

3: For the given feedback function ψ^m find $(X_t^m)_{t\in[0,T]}$ as the solution of (5.5) and solve the system of PDEs defined by (5.7).

4: Compute the updates

$$\phi^{m+1}(t,x) = \mathrm{prox}_{\ell,\tau}(\psi^m(t,x) - \tau\partial_\alpha F(\psi^m(t,x)))$$
$$\psi^{m+1}(t,x) = \phi^{m+1}(t,x) + \frac{m}{m+3}(\phi^{m+1}(t,x) - \phi^m(t,x)),$$

for $(t,x) \in [0,T] \times \mathbb{R}^d$, and return to the first step.

5: **end for**

6: **return** $\phi^{N_{nag}}$, $J(\phi^{N_{nag}}(\cdot, X_\cdot^{N_{nag}}))$.

Here, we proposed a gradient descent method, which at each iteration requires a solution to the adjoint equations and then takes a proximal gradient descent step in combination with a momentum step in order to update the given feedback control. The other central aspect of our algorithm is the computation of (u^m, v^m) via a PDE method, which determines the decoupling fields for all

$(t,x) \in [0,T] \times \mathbb{R}^d$ in the computational domain and not only along controlled forward trajectories as in probabilistic approaches. Therefore, as the feedback maps are also updated pointwise, the FIPDE algorithm recovers the optimal feedback control on the entire domain and in many practical applications provides a better interpretation for structural properties of the approximate controls.

Further, we recall that the distribution of the state process X^m given in (5.5) will be estimated employing a particle approximation (based on $N \in \mathbb{N}$ particles) and an Euler–Maruyama scheme with an equidistant partition $\{t_j\}_{j\in\{0,\dots,M\}}$ of $[0,T]$ for $M \in \mathbb{N}$. In particular, this means that we approximate the marginal laws of X^m by an empirical distribution of particles and therefore avoid solving a high-dimensional nonlinear Fokker–Planck PDE for the density. Let now N and M be given and define $h = T/M$. Further, let $(W^i)_{i\in\{1,\dots,N\}}$ and $(X_0^{i,N,m})_{i\in\{1,\dots,N\}}$ be independent copies of W and ξ, respectively. Then, for a given feedback map ψ^m, which in practice will be represented in parametric form (e.g., as neural networks with given set of parameters θ), we recursively define for $j \in \{0,\dots,M-1\}$ the interacting particle system

$$X_{t_{j+1}}^{i,N,m} = X_{t_j}^{i,N,m} + b(t, X_{t_j}^{i,N,m}, \psi^m(t_j, X_{t_j}^{i,N,m}), \mu_{m,t_j}^{\boldsymbol{X}^N,\psi})h + \sigma(t, X_{t_j}^{i,N,m}, \psi^m(t_j, X_{t_j}^{i,N}), \mu_{m,t_j}^{\boldsymbol{X}^N,\psi})\Delta W_j^i, \tag{5.9}$$

where $\mu_{m,t_j}^{\boldsymbol{X}^N,\psi}(\mathrm{d}x) := \frac{1}{N}\sum_{l=1}^N \delta_{(X_{t_j}^{l,N,m}, \psi^m(t_j, X_{t_j}^{l,N,m}))}(\mathrm{d}x)$ and $\Delta W_j^i = W_{t_{j+1}}^i - W_{t_j}^i$. Here, δ_x denotes the usual Dirac measure centred at x.

5.3 FIPDE method and neural network implementation

In this section, we will use the DG method [143] in combination with a residual approximation technique to solve at each NAG iteration the system of PDEs for the decoupling fields (u^m, v^m). The algorithm will be presented in detail below.

To reduce the computational complexity of the FIPDE method in high-dimensions, we propose the following acceleration method: Instead of directly finding an approximate solution to (5.7) for the decoupling field u^m of Y^m, we aim to solve for a residual $\bar{u}^m(t,x) := u^m(t,x) - \tilde{u}^m(t,x)$ for all $(t,x) \in [0,T] \times \mathbb{R}^d$, where $\tilde{u}^m$ explains a significant part of u^m and is obtained by an efficient numerical method (e.g., ODE solver) or may even be available in analytic form. A similar idea has been employed in [147] to improve the efficiency for the Deep BSDE solver. Clearly, the choice of $\tilde{u}^m$ is problem dependent. However, in many cases one can find a suitable linear-quadratic MFC problem, which serves as a good approximation for the general problem (5.1) and (5.2). Consequently, $\tilde{u}^m$ can be found as the solution to matrix-valued Riccati equations associated to such a linear-quadratic MFC problem.

In what follows, we derive a system of PDEs for $\bar{u}^m$: Applying Itô's formula to $\tilde{u}^m$, we observe that

$$\bar{u}^m(t, X_t^m) = \int_t^T f^m(t, X_s^m, Y_s^m, Z_s^m)\,\mathrm{d}s - \int_t^T \bar{Z}_s\,\mathrm{d}W_s$$

$$+\int_t^T \left(\partial_s \tilde{u}^m(s, X_s^m) + \partial_x \tilde{u}^m(s, X_s^m) b^m(s, X_s^m) + \tfrac{1}{2}\mathrm{tr}(\sigma^m \sigma^{m,\mathsf{T}} \partial_x^2 \tilde{u}^m(s, X_s^m))\right) \mathrm{d}s,$$

where f^m is the driver in the equation for Y^m defined in (5.6) and

$$\bar{Z}_t^m = Z_t^m - \partial_x \tilde{u}^m(t, X_t^m)\sigma^m(t, X_t^m).$$

Further, we set for any $(t, x) \in [0, T] \times \mathbb{R}^d$

$$\bar{v}^m(t, x) = v^m(t, x) - \partial_x \tilde{u}^m(t, x)\sigma^m(t, x) = \partial_x \bar{u}^m(t, x)\sigma^m(t, x),$$

where $\bar{v}^m(t, X_t^m) = \bar{Z}_t^m$.

Subsequently, we formulate the PDE system for $\bar{u}^m$: $\bar{u}^m(T, x) = h^m(x) - \tilde{u}^m(T, x)$, and for all $t \in [0, T)$,

$$\partial_t \bar{u}^m + \mathscr{L}^m \bar{u}^m = -\bar{f}^m(t, x, \bar{u}^m(t, x), \bar{v}^m(t, x)), \tag{5.10}$$

where

$$\begin{aligned}
&\bar{f}^m(t, x, \bar{u}^m(t, x), \bar{v}^m(t, x)) \\
&= f^m(t, x, \bar{u}^m(t, x) + \tilde{u}^m(t, x), \bar{v}^m(t, x) + \tilde{v}^m(t, x)) \\
&\quad + \partial_t \tilde{u}^m(t, x) + \partial_x \tilde{u}^m(t, x) b^m(t, x) + \tfrac{1}{2}\mathrm{tr}(\sigma^m \sigma^{m,\mathsf{T}} \partial_x^2 \tilde{u}^m(t, x)).
\end{aligned}$$

This residual approximation approach can now be combined with our proposed PDE method in the following way. At each NAG iteration m, instead of solving directly a system of PDEs for u^m, we solve the system (5.10) for $\bar{u}^m$ and then return $\bar{u}^m + \tilde{u}^m$.

The following algorithm details the main steps of the DG method applied to the PDE system (5.10) at a given NAG iteration m, where for ease of notation we drop the dependence of $\bar{f}^m$ on $\bar{v}^m$, i.e., the diffusion is assumed to be uncontrolled. Note that the boundary and terminal conditions for $\bar{u}^m$ will be denoted by $\bar{g}^m = (\bar{g}_1^m, \ldots, \bar{g}_d^m)$ with $\bar{g}_k^m : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ and $\bar{u}^m(T, \cdot) =: \bar{h}_T^m = (\bar{h}_{T,1}^m, \ldots, \bar{h}_{T,d}^m)$ with $\bar{h}_{T,k}^m : \mathbb{R}^d \to \mathbb{R}$, respectively, for all $k \in \{1, \ldots, d\}$. Roughly speaking, the DG method seeks an optimal neural network approximation of the solution to (5.10) within a prescribed family of networks $\mathcal{N}_\theta := \{\bar{u}(\cdot, \cdot; \theta) : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d : \theta \in \mathbb{R}^l\}$ with fixed architecture, by minimising a certain residual, precisely defined in (5.11). This loss will then be minimised by performing a SGD step with respect to the parameters θ. We remark that the weights $\eta_1, \eta_2 > 0$ in Algorithm 3 are used to balance the residual loss on the boundary and the interior. Additionally, we wish to point out that the number of random points (N_b in Algorithm 3) sampled from the boundary and the interior of the computational domain $\mathcal{D}$ could be chosen differently.

Note that in our implementation, all networks which appeared in Algorithm 3 have the same architecture. With the DG algorithm at hand, we are then ready to present our fully-implementable

Algorithm 3 Deep Galerkin alogrithm

Input: Computational domain $\mathcal{D} \subset \mathbb{R}^d$; feedback map ψ^m (represented as neural network); forward samples $(X_{t_j}^{i,N,m})_{j\in\{0,\dots,M\}}^{i\in\{1,\dots,N\}}$ defined in (5.9); number of iterations N_{itr}; batch size N_b; learning rate lr; Neural network initialisation of $\bar{u}^m = (\bar{u}_1^m, \dots, \bar{u}_d^m) \in \mathcal{N}_\theta$, where $\bar{u}_k^m : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, for $k \in \{1,\dots,d\}$; weights $\eta_1, \eta_2 > 0$

Output: Neural network approximation to solution of (5.10)

1: **for** $i = 0, \dots, N_{itr} - 1$ **do**

2:

3: Sample uniformly N_b elements $(\tau_n, X_n)_{n\in\{1,\dots,N_b\}}$, $(s_n, Y_n)_{n\in\{1,\dots,N_b\}}$ and $(Z_n)_{n\in\{1,\dots,N_b\}}$, from $[0,T] \times \mathcal{D}$, $[0,T] \times \partial\mathcal{D}$, and $\mathcal{D}$, respectively.

4: Define $p_n = \{(t_n, X_n), (s_n, Y_n), Z_n\}$ for $n \in \{1,\dots,N_b\}$. For each p_n, compute the weighted squared L_2-error $e(\theta_i, p_n)$ defined by

$$\begin{aligned} e(\theta_i, p_n) = &\Big|\partial_t \bar{u}^m(\tau_n, X_n; \theta_i) + \mathscr{L}^m \bar{u}^m(\tau_n, X_n; \theta_i) + \bar{f}^m\big(\tau_n, X_n, \bar{u}^m(\tau_n, X_n; \theta_i), \mu_{m,\tau_n}^{\boldsymbol{X}^N, \bar{u}}\big)\Big|^2 \\ &+ \eta_1 |\bar{u}^m(s_n, X_n; \theta_i) - g^m(s_n, Y_n)|^2 + \eta_2 |\bar{u}^m(T, Z_n; \theta_i) - h_T^m(Z_n)|^2, \end{aligned} \tag{5.11}$$

where $\mu_{m,\tau_n}^{\boldsymbol{X}^N, \bar{u}}(\mathrm{d}x) := \frac{1}{N}\sum_{l=1}^N \delta_{(X_{\tau_n}^{l,N}, \bar{u}^m(\tau_n, X_{\tau_n}^{l,N}; \theta_i))}(\mathrm{d}x)$, and compute the average loss

$$loss = \frac{1}{N_b}\sum_{n=1}^{N_b} e(\theta_i, p_n).$$

5: Perform a gradient descent step at the random points $(p_n)_{n\in\{1,\dots,N_b\}}$ with respect to θ_i with learning rate lr based on the computed loss.

6: **end for**

7: **return** Neural network approximation $(\bar{u}_1^m, \dots, \bar{u}_d^m)$ of solution to (5.10).

algorithm based on the PDE representation for the decoupling fields of the adjoint processes combined with the NAG updates (Algorithm 4). Observe that the networks $(\phi_1^0, \dots, \phi_d^0)$ and $(\psi_1^0, \dots, \psi_d^0)$ in Algorithm 4 have the same initialisation (see, also Algorithm 1), as required by the NAG method and are supposed to be elements of a family of neural networks with a fixed architecture denoted by $\mathcal{N}_\pi = \{\phi(\cdot,\cdot;\pi) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k : \ \pi \in \mathbb{R}^q\}$.

The NAG step (step 5) in Algorithm 4 deserves a more detailed elaboration. First, we remark that given a feedback map ψ^m and a neural network approximation $\bar{u}^m(\cdot,\cdot;\theta)$ to (5.10), we readily obtain an approximate solution to (5.7) by $u^m(\cdot,\cdot;\theta) = \bar{u}^m(\cdot,\cdot;\theta) + \tilde{u}^m(\cdot,\cdot)$. Hence, for any $(t,x) \in [0,T] \times \mathbb{R}^d$, an approximation of $\partial_\alpha F(\psi^m(t,x))$ is obtained by replacing $u^m(x,t)$ with $u^m(x,t;\theta)$ and $\psi^m(t,x)$ by $\psi^m(t,x;\pi)$ in (5.8). To perform the NAG updates, we employ a supervised learning algorithm. This is necessary, as ϕ^{m+1} and ψ^{m+1} need to be represented by a linear combination of neural networks with nonlinear activation functions; a representation, which, in general, cannot be achieved by simple operations applied to the parameters. We detail the supervised learning procedure in Algorithm 5.

Algorithm 4 FIPDE neural network algorithm

Input: Computational domain $\mathcal{D} \subset \mathbb{R}^d$; number of iterations N_{itr}; batch size N_b; learning rate lr; Neural network initialisation of $\bar{u}^0 = (\bar{u}_1^0, \dots, \bar{u}_d^0) \in \mathcal{N}_\theta$, where $\bar{u}_k^0 : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, for $k \in \{1, \dots, d\}$; Neural network initialisation for $(\phi_1^0, \dots, \phi_d^0) \in \mathcal{N}_\pi$ and $(\psi_1^0, \dots, \psi_d^0) \in \mathcal{N}_\pi$, where $\psi_k^0, \phi_k^0 : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, for $k \in \{1, \dots, d\}$; number of NAG iterations N_{nag} and step-size $\tau > 0$

Output: Neural network approximation to solution of (5.7) and control approximation

1: **for** $m = 0, \dots, N_{nag} - 1$ **do**

2:

3: For the given control representation $(\psi_1^m, \dots, \psi_d^m)$, compute forward samples $(X_{t_j}^{i,N,m})_{j \in \{0,\dots,M\}}^{i \in \{1,\dots,N\}}$ defined in (5.9) and find an approximate solution $(\bar{u}_1^m, \dots, \bar{u}_d^m)$ to (5.10) using the DG algorithm.

4: Set $u^m(t,x;\theta) = \bar{u}^m(t,x;\theta) + \tilde{u}(t,x)$ for all $(t,x) \in [0,T] \times \mathbb{R}^d$.

5: Compute the updates

$$\phi^{m+1}(t,x;\pi) = \operatorname{prox}_{\ell,\tau}\left(\psi^m(t,x;\pi) - \tau \partial_\alpha F(\psi^m(t,x;\pi))\right),$$
$$\psi^{m+1}(t,x;\pi) = \phi^{m+1}(t,x;\pi) + \frac{m}{m+3}(\phi^{m+1}(t,x;\pi) - \phi^m(t,x;\pi)),$$

using a supervised learning algorithm and return to the first step.

6: **end for**

7: **return** Neural network approximation $(u_1, \dots, u_d)$ to solution of (5.7) and control approximation $(\phi_1^{N_{nag}}, \dots, \phi_d^{N_{nag}})$.

Algorithm 5 Supervised learning

Input: Two networks $n_1 : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k \in \mathcal{N}_\eta$ and $n_2 : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k \in \mathcal{N}_\vartheta$ with given sets of parameters η and ϑ, respectively, and initialisation of target network $n_{out} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k \in \mathcal{N}_\kappa$; weights $w_1, w_2 > 0$; computational domain $\mathcal{D} \subset \mathbb{R}^d$; number of iterations N_{itr}; batch size N_b; learning rate lr

Output: Neural network n_{out}

1: **for** $i = 0, \dots, N_{itr} - 1$ **do**

2:

3: Sample uniformly N_b elements $(\tau_n, X_n)_{n \in \{1,\dots,N_b\}}$, $(s_n, Y_n)_{n \in \{1,\dots,N_b\}}$ and $(Z_n)_{n \in \{1,\dots,N_b\}}$, from $[0,T] \times D$, $[0,T] \times \partial D$, and D, respectively.

4: Define $p_n = \{(t_n, X_n), (s_n, Y_n), Z_n\}$ for $n \in \{1, \dots, N_b\}$. For each p_n, compute the error $e(\theta_i, p_n)$ defined by

$$\begin{aligned} e(\theta_i, p_n) =& w_1 \left| n_{out}(\tau_n, X_n; \theta_i) - n_1(\tau_n, X_n; \eta) - n_2(\tau_n, X_n, \vartheta) \right|_1 \\ &+ w_2 \left| n_{out}(s_n, Y_n; \theta_i) - n_1(s_n, Y_n; \eta) - n_2(s_n, Y_n; \vartheta) \right|_1 \\ &+ \left| n_{out}(T, Z_n; \theta_i) - n_1(T, Z_n; \eta) - n_2(T, Z_n; \vartheta) \right|_1, \end{aligned}$$

and compute the average loss

$$loss = \frac{1}{N_b} \sum_{n=1}^{N_b} e(\theta_i, p_n).$$

5: Compute the gradient descent step at the random points $(p_n)_{n \in \{1,\dots,N_b\}}$ with respect to θ_i with learning rate lr based on the computed loss.

6: **end for**

7: **return** Neural network n_{out}.

5.4 High-dimensional Cucker–Smale models

The goal of this section is to apply the algorithm presented in Section 5.3 to multi-dimensional stochastic mean-field Cucker–Smale (C–S) models, whose description is detailed in the subsequent section. These models were originally proposed in [52, 53] to study the swarm behaviour of bird populations, but can also be applied in various other situations, such as the emergence of a common language or appearance of a belief in a pricing system shared by all market participants. In Section 5.4.1, beside the precise formulation of the C–S models, we derive a PDE system for decoupling fields of the adjoint equations. Section 5.4.2 aims to illustrate the performance and effectiveness of the FIPDE method by applying it to six-dimensional non-smooth and non-convex C–S models. The concrete implementation of the FIPDE relies on the neural network based residual approximation proposed in Section 5.3. To demonstrate important features of the FIPDE method and its advantages compared to purely sampling based algorithms, we use the neural network based policy gradient (NNPG) method [34] as benchmark. Recall that the NNPG method restricts the minimisation of the value function over feedback controls within a class of neural networks with given architecture and performs an SGD step based on trajectories of the controlled state process.

5.4.1 Problem formulation

In biology, physics or social sciences, the microscopic description of a dynamical system composed of N interacting particles is typically described by a system of ODEs for the position $x_t^i \in \mathbb{R}^d$ and velocity $v_t^i \in \mathbb{R}^d$ of each particle at time $t \in [0,T]$, i.e., for $i \in \{1,\dots,N\}$, we have

$$\begin{aligned} \mathrm{d}x_t^i &= v_t^i \,\mathrm{d}t, \\ \mathrm{d}v_t^i &= \frac{1}{N}\sum_{j=1}^{N} \frac{K(v_t^j - v_t^i)}{(1+|x_t^i - x_t^j|^2)^\beta}\,\mathrm{d}t, \end{aligned}$$

where $K, \beta > 0$ and x_0^i, v_0^i, are the possibly random initial position and velocity of particle i. In the case $\beta \leq 1/2$, it is known that the velocity alignment (all particles have the same velocity) does not depend on the initial configuration of the system or on the dimensionality. To achieve flocking (or unconditional alignment) for $\beta > 1/2$, the initial data needs needs to satisfy certain structural conditions, see, e.g., [10]. There exists a vast literature studying the (un)controlled dynamics of such C–S models in this deterministic framework, see, e.g., [37, 7, 52, 53] and references therein. In what follows, we study the optimal control of the multi-dimensional stochastic mean-field C–S dynamics (see [124] and [31, Chapter 4]), where the controller aims to enforce consensus emergence of an interactive particle system via external intervention.

To achieve alignment of the crowd in situations, where the initial configuration does not satisfy the above mentioned condition or to accelerate consensus emergence, we may reformulate

the asymptotic problem (infinite number of particles) as a MFC problem. The state process $(X_t)_{t\in[0,T]} = (x_t, v_t)_{t\in[0,T]}$ is described by

$$\mathrm{d}x_t = v_t\,\mathrm{d}t, \quad \mathrm{d}v_t = (b(x_t, v_t, \mathcal{L}_{(x_t,v_t)}) + \alpha_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad t \in [0,T], \tag{5.12}$$

where $(x_0, v_0) = \xi \in L_0^2(\mathbb{R}^{2d})$, $\sigma \in \mathbb{R}^{d\times d}$, W is a d-dimensional Brownian motion and for $(x, v, \chi) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$,

$$b(x, v, \chi) := \int_{\mathbb{R}^{2d}} \kappa(x, v, x', v')\,\chi(\mathrm{d}x', \mathrm{d}v'),$$

where $\kappa(x, v, x', v') = \frac{K(v'-v)}{(1+|x-x'|^2)^\beta}$. In the sequel, we will also use $\kappa = (\kappa_1, \ldots, \kappa_d)$.

Now, for given constants $\gamma_1, \gamma_2 \geq 0$, we consider minimising the following cost functional

$$J(\alpha; (x_0, v_0)) = \mathbb{E}\left[\int_0^T \left(\mathbb{E}[|v_t - \mathbb{E}[v_t]|^2] + \gamma_1|\alpha_t|^2 + \gamma_2|\alpha_t|_1\right)\mathrm{d}t + \mathbb{E}[|v_T - \mathbb{E}[v_T]|^2]\right], \tag{5.13}$$

over all adapted control strategies α taking values in $\mathbb{R}^d$, where $|\cdot|_1$ is the ℓ_1-norm of a given vector. Note that in the case $\beta = 0$ and $\gamma_2 = 0$, $\gamma_1 > 0$, we have a linear-quadratic optimal control problem and the optimality system has a unique solution, see [152], which can numerically be found by solving matrix-valued Riccati equations.

The Hamiltonian of this system is

$$H(x, v, a, \chi, y_1, y_2, z_2) = \langle v, y_1\rangle + \langle a, y_2\rangle + \langle b(x, v, \chi), y_2\rangle + \langle \sigma, z_2\rangle + |v - \bar{v}|^2 + \gamma|a|^2,$$

for $(x, v, a, \chi, y_1, y_2, z_2) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d\times d}$, where $\bar{v} = \int_{\mathbb{R}^d} v\,\mu(\mathrm{d}v)$ denotes the first moment of the measure μ (the second marginal measure of χ). Taking the structure of the Hamiltonian into account, we obtain the following adjoint equations describing the processes (Y_1, Y_2, Z_1, Z_2): For all $t \in [0, T)$, we have

$$\begin{aligned}
\mathrm{d}Y_{1,t} &= -\Big(\tilde{\mathbb{E}}[\partial_x\kappa(x_t, v_t, \tilde{x}_t, \tilde{v}_t)]Y_{2,t} + \tilde{\mathbb{E}}[\partial_{x'}\kappa(\tilde{x}_t, \tilde{v}_t, x_t, v_t)\tilde{Y}_{2,t}]\Big)\,\mathrm{d}t + Z_{1,t}\,\mathrm{d}W_t,\\
\mathrm{d}Y_{2,t} &= -\Big(Y_{1,t} + 2(v_t - \mathbb{E}[v_t]) + \tilde{\mathbb{E}}[\partial_v\kappa(x_t, v_t, \tilde{x}_t, \tilde{v}_t)]Y_{2,t} + \tilde{\mathbb{E}}[\partial_{v'}\kappa(\tilde{x}_t, \tilde{v}_t, x_t, v_t)\tilde{Y}_{2,t}]\Big)\,\mathrm{d}t + Z_{t,2}\,\mathrm{d}W_t,
\end{aligned}$$

and $(Y_{1,T}, Y_{2,T}) = (0, 2(v_T - \mathbb{E}[v_T]))$. Note that in the above formulation, the expression $\partial_x\kappa$ (and analogously for the remaining derivatives), denotes the Jacobian of κ, where the i-th row has the form $(\partial_{x_i}\kappa_1, \ldots, \partial_{x_i}\kappa_d)$, for $i \in \{1, \ldots, d\}$.

The system of parabolic PDEs associated to the adjoint equations given above, for a given flow of measures $(\mathcal{L}_{(x_t^m, v_t^m)})_{t\in[0,T]} \in \mathcal{C}([0,T], \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d))$ and feedback control $\psi^m : [0,T] \times \mathbb{R}^{2d} \to \mathbb{R}^d$, has the following form:

Assume $(Y_{1,t}^m, \ldots, Y_{2d,t}^m) = (u_1^m(t, x_t^m, v_t^m), \ldots, u_{2d}^m(t, x_t^m, v_t^m)) \in \mathbb{R}^{2d}$, for sufficiently smooth functions $u_i^m : [0,T] \times \mathbb{R}^{2d} \to \mathbb{R}$, for $i \in \{1, \ldots, 2d\}$ and $t \in [0,T]$. Then, for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $v = (v_1, \ldots, v_d) \in \mathbb{R}^d$, the functions u_i^m satisfy the terminal conditions $u_1^m(T, \cdot) = \ldots = u_d^m(T, \cdot) = 0$,

$u^m_{d+j}(T,\cdot) = 2(v_j - \mathbb{E}[v^m_{j,T}])$, where $j \in \{1,\dots,d\}$ and $v^m_t = (v^m_{1,t},\dots,v^m_{d,t})$ denotes controlled velocity process at the m-th NAG iteration. For $t \in [0,T)$, they solve the system of PDEs

$$
\begin{aligned}
&\partial_t u^m_i + \frac{1}{2}\mathrm{tr}\left(\sigma\sigma^\intercal \partial^2_{v_j} u^m_i\right) + \langle v, \partial_x u^m_i\rangle + \left\langle b(x,v,\mathcal{L}_{(x^m_t,v^m_t)}) + \psi^m(t,x,v), \partial_v u^m_i\right\rangle \\
&\quad = -\Big(\sum_{j=1}^d u^m_{d+j}(t,x,v)\tilde{\mathbb{E}}[\partial_{x_i}\kappa_j(x,v,\tilde{x}^m_t,\tilde{v}^m_t)] + \sum_{j=1}^d \tilde{\mathbb{E}}[u^m_{d+j}(t,\tilde{x}^m_t,\tilde{v}^m_t)\partial_{x'_i}\kappa_j(\tilde{x}^m_t,\tilde{v}^m_t,x,v)]\Big), \\
&\partial_t u^m_k + \frac{1}{2}\mathrm{tr}\left(\sigma\sigma^\intercal \partial^2_{v_j} u^m_k\right) + \langle v, \partial_x u^m_k\rangle + \left\langle b(x,v,\mathcal{L}_{(x^m_t,v^m_t)}) + \psi^m(t,x,v), \partial_v u^m_k\right\rangle \\
&\quad = -\Big(u^m_{k-d}(t,x,v) + 2(v_{k-d} - \mathbb{E}[v^m_{k-d,t}]) \\
&\qquad + u^m_k(t,x,v)\tilde{\mathbb{E}}[\partial_{v_k}\kappa_{k-d}(x,v,\tilde{x}^m_t,\tilde{v}^m_t)] + \tilde{\mathbb{E}}[u^m_k(t,\tilde{x}^m_t,\tilde{v}^m_t)\partial_{v'_k}\kappa_{k-d}(\tilde{x}^m_t,\tilde{v}^m_t,x,v)]\Big),
\end{aligned}
\tag{5.14}
$$

where $i \in \{1,\dots d\}$ and $k \in \{d+1,\dots,2d\}$.

5.4.2 Numerical illustration

In this section, we use the FIPDE neural network algorithm (Algorithm 4) to solve (5.12) and (5.13) in a six-dimensional setting with different choices for the communication rate β and the sparse regularisation parameter γ_2. The remaining parameters that specify our framework are chosen as follows: $d = 3$, $\gamma_1 = 0.3$, σ is a diagonal matrix with equal entries 0.1, $K = 1$, $T = 1$, $\mathcal{D} = [-1,2]^6$, $(x_0,v_0) \sim \mathcal{U}([0,1]^6)$ (the uniform distribution on $[0,1]^6$).

We start the presentation of our numerical results for the case $\beta = \gamma_2 = 0$. In this case the solution to (5.12) and (5.13) can readily be found by solving a matrix-valued Riccati equation and we can therefore evaluate the exactness of the numerical approximation obtained by our algorithm. In fact, it can be shown (see, [152]) that the optimal feedback control of this linear-quadratic problem with the above imposed terminal conditions has the form $\phi^*(t,x,v) = -\frac{1}{2\gamma_1}P_t(v - \mathbb{E}[v^*_t])$, for all $(t,x,v) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^d$, where $P \in \mathcal{C}([0,T],\mathbb{R}^{d\times d})$ and $(v^*_t)_{t\in[0,T]}$ is the optimal velocity process. As we choose $d = 3$, we solve at each NAG iteration a 6-dimensional system of PDEs (5.14) and each component is a mapping of the form $u^m_k : [0,T]\times\mathbb{R}^6 \to \mathbb{R}$ for $k \in \{1,\dots,6\}$. The feedback controls are therefore given by mappings $\phi^m, \psi^m : [0,T]\times\mathbb{R}^6 \to \mathbb{R}^3$. We represent the feedback maps by feedforward neural networks with 2 layers and 20 hidden units. The networks for u^m_k have the same architecture and the u^0_k's are randomly initialised, while the networks for the feedback maps ϕ^0, ψ^0 are initialised with zero weights, i.e., we initialise Algorithm 4 with the uncontrolled dynamics. We choose the sigmoid activation function.

To solve (5.14) on the spatial computational domain $\mathcal{D} = [-1,2]^6$, we use the DG method (Algorithm 3) with $\tilde{u} = 0$, as in the case $\beta = 0$ there is no obvious candidate for the residual correction term (except the exact solution of this linear-quadratic problem). The law of the controlled state process is estimated based on forward samples (5.9) with $N = 5\cdot 10^4$ (number of

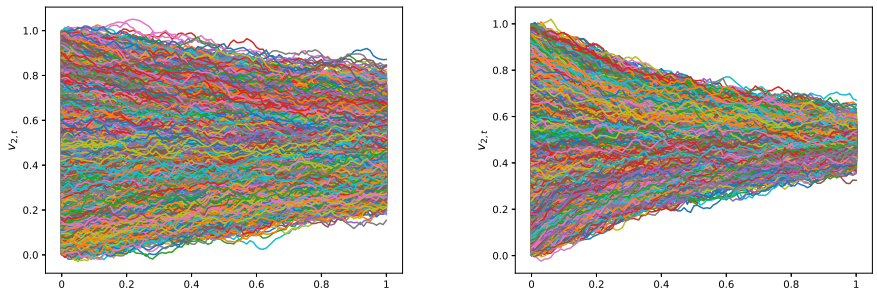
particles), $M = 100$ (number of time-steps). We perform the FIPDE method with 15 NAG iterations and choose $\tau = 1/5$. In Table 5.1, we summarise the numerical results, in particular we give a quantitative examination on the convergence of the FIPDE method and its accuracy in terms of control and value function approximation. We measure the control accuracy using the following error criterion $\mathrm{err}_c := (\mathbb{E}[\sup_{t\in[0,T]} |\alpha_t^* - \alpha_t^m|^2])^{1/2}$, where α_t^* denotes the optimal control process and $\alpha_t^m = \phi^m(t, X_t^{\phi^m})$ is the numerically obtained control process at the m-th NAG iteration. The value function approximation is measured by the absolute error $\mathrm{err}_v := |J(\phi^m(\cdot, X_\cdot^{\phi^m})) - J^*|$. We also aim to analyse the sensitivity of the DG solver with respect to the chosen network architecture. To study the robustness of the algorithm in terms of width (depth) of the architecture, we fix the depth (width) of the network and vary the number of hidden dimensions (layers). Considering the choice of activation function, we remark that although the feedback map for $\beta = \gamma_2 = 0$ is a linear function in the velocity, ReLU activation did not perform well in our numerical tests in terms of value function/control approximation and the accuracy of the DG method (based on the same architecture and learning rate used for the sigmoid function), which might be due to the nonlinearity in the time variable. The tanh activation gave similar results (with regards to value function/control approximation and the accuracy of the DG solver) when comparing to the sigmoid function. Based

NN architecture (depth, width, activation)	error criterion $\mathrm{err}_v/\mathrm{err}_c$	NAG itr 1	3	6	9	12	15
$(2, 20, \text{sigmoid})$	err_v	4.01%	1.21%	0.50%	0.23%	0.10%	0.07%
	err_c	26.50%	15.49%	10.15%	6.32%	4.00%	2.83%
$(4, 20, \text{sigmoid})$	err_v	3.96%	1.22%	0.45%	0.19%	0.08%	0.05%
	err_c	26.76%	16.67%	10.82%	7.21%	4.70%	3.46%
$(2, 40, \text{sigmoid})$	err_v	3.95%	1.27%	0.53%	0.24%	0.12%	0.09%
	err_c	26.38%	15.94%	10.15%	6.63%	4.12%	3.00%
$(2, 20, \tanh)$	err_v	4.12%	1.97%	0.86%	0.41%	0.22%	0.13%
	err_c	26.59%	19.36%	12.61%	8.60%	5.57%	3.46%

Table 5.1: Convergence of the FIPDE method for different choices of neural network architectures. Presented are the value function approximations across NAG iterations measured by the absolute error $\mathrm{err}_v = |J(\phi^m(\cdot, X_\cdot^{\phi^m})) - J^\star|$ and the approximation accuracy of the numerical control process α^m quantified by $\mathrm{err}_c = (\mathbb{E}[\sup_{t\in[0,T]} |\alpha_t^m - \alpha_t^\star|^2])^{\frac{1}{2}}$.

on the numerically obtained optimal control approximation, we compute forward samples of the controlled velocities for $M = 100$, and $N = 5000$ (the remaining parameters are chosen as above), see Fig. 5.1b, and compare it with the uncontrolled trajectories, see Fig. 5.1a. The effect of the control is clearly visible and we observe that the variance of the controlled samples at the terminal time T (around 0.0022) is significantly smaller than the one in the uncontrolled case (around 0.0151).

In a next step, we present a qualitative study of the feedback map, numerically obtained by Algorithm 4 and compare its structural properties with the optimal feedback map ϕ^*. Although



(a) (b)

Figure 5.1: Sample trajectories of the uncontrolled (left) and controlled (right) velocities (second component) for the C–S model with $\beta = \gamma_2 = 0$.

the feedback maps used in the FIPDE method are represented by neural networks taking the 7-dimensional input (t, x, v), our approach still recovers the correct feedback behaviour of the optimal control: along v_2 the feedback map is a linear function, along x_2, it is constant and its slope increases as t gets larger (and similar for the remaining components), see Fig. 5.2 (left) and (middle). We wish to emphasise that we obtain the correct linear behaviour of the feedback map on the entire high-dimensional computational domain and not only in areas where the samples of optimal state trajectories are concentrated. To highlight this desirable property of our method, we depict in Fig. 5.2 (right) the feedback control obtained by employing the NNPG approach, which is implemented using the same neural network architecture for the controls as in the FIPDE method. In Fig. 5.2 (right), we observe the linear feedback structure only in subregions of the computational domain, i.e., where samples of controlled velocities occur. Note that for $t = 0$ the aforementioned subregion is larger than regions at later time points, as we choose $(x_0, v_0) \sim \mathcal{U}([0,1]^6)$ as initial distribution and the trajectories of the controlled velocities will concentrated more and more around the mean as t increases. Also, we remark that the NNPG method fails to capture the correct time-dependence of the optimal control, compared to the FIPDE algorithm.

We additionally consider a cost involving a non-smooth functional, i.e., we will set $\gamma_2 \in \{0.1, 0.3\}$ and keep the remaining parameters unchanged. In this setting, we initialise the feedback maps ϕ^0, ψ^0 with the numerical solution obtained by the FIPDE method for $\gamma_2 = 0$ and again set $\tau = 1/5$. As outlined in Section 5.3, we will choose the neural network representation of the decoupling fields obtained for the adjoint processes when $\gamma_2 = 0$, as approximation of the PDE system's solution in the present case. By doing so, only a residual, which might only explain a very small part of the overall solution, has to be approximated and thereby the performance of the DG solver is improved; i.e., in general less training iterations are needed to obtain a certain accuracy at each NAG iteration. In particular, at each NAG iteration, we consider the decomposition $u^m = \bar{u}^m + \tilde{u}$, where $\tilde{u} : [0, T] \times \mathbb{R}^6 \to \mathbb{R}^6$ denotes the solution of above linear-quadratic problem and then seek

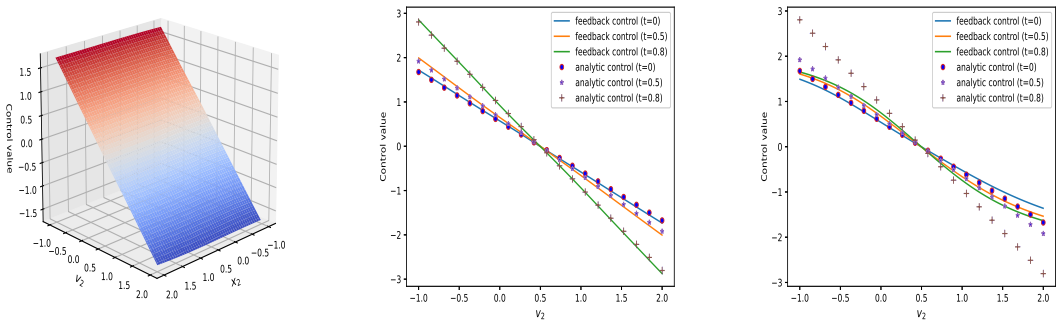


Figure 5.2: Illustration of the numerically computed feedback control (second component) of a six-dimensional C–S model with $\beta = \gamma_2 = 0$; from left to right: Feedback control obtained by employing the FIPDE method at $t = 0$ and evaluated for different (x_2, v_2); comparison numerical feedback control and the analytic solution at different (t, v_2); comparison numerical feedback control obtained by employing the NNPG method and the analytic solution at different (t, v_2). The remaining components are set to 0.5.

an approximate solution to the residual component $\bar{u}^m : [0,T] \times \mathbb{R}^6 \to \mathbb{R}^6$. To obtain such an approximation, we represent the components of $\bar{u}^m$ by feedforward neural networks with 4 layers, 40 hidden dimensions and choose the sigmoid activation function. The optimal neural network representation of the residual is found by employing Algorithm 3.

The performance improvement of the residual approach (RA) compared to a direct approach (DA), where we directly try to compute an optimal neural network approximation for u^m (i.e., we set $\tilde{u} = 0$ in Algorithm 3), is numerically demonstrated in Fig. 5.3. To make this comparison fair, we implement the DA with the same neural network architecture, number of training samples and learning rate as for the RA. While the number of SGD iterations for the RA is chosen as 100, we perform $i \in \{200, 500, 1000\}$ iterations when employing the DA. Fig. 5.3 depicts the convergence of the value functions computed with these two approaches and different choices of SGD iterations. Here, we illustrate the results for the choice $\gamma_2 = 0.1$ and remark that a similar convergence acceleration was achieved for different choices of γ_2. We note that the RA requires significantly less SGD iterations (than the DA) to solve the PDE system (5.14) with a certain precision and additionally, the resulting approximate value functions are smaller at each NAG iteration.

We proceed with a qualitative investigation of the feedback controls for the MFC problem (5.12) and (5.13) with $\gamma_2 \in \{0.1, 0.3\}$ and $\beta = 0$, obtained by the FIPDE method and the NNPG method. Note that due to the non-smooth term in the cost functional, we expect the feedback map to have sparse regions and to be nonlinear in the velocity component (compared to the linear-quadratic control policy). As illustrated by Fig. 5.4 (left and middle) the FIPDE method recovers the correct structure of the sparse region of the optimal feedback control. Also, we observe that the region where the control is zero, shrinks as we approach the terminal time T and becomes larger as the weight γ_2 increases. From a practical perspective, it is important to detect the zero-control region and to

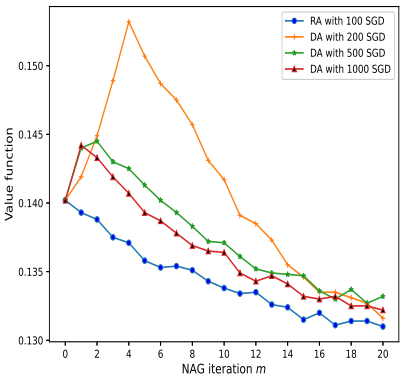
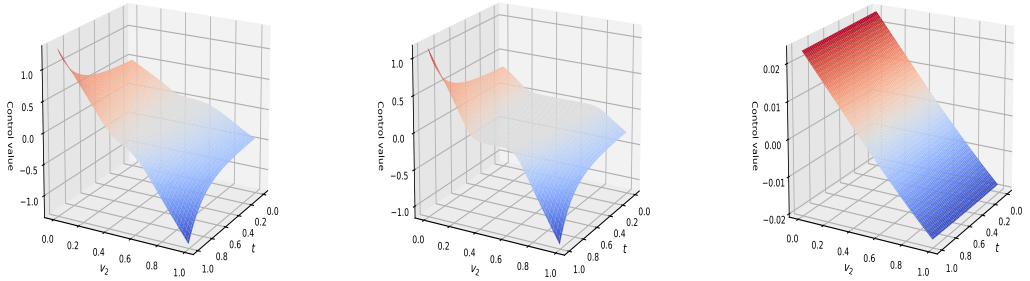


Figure 5.3: Comparison residual approximation approach with the direct approach for six-dimensional C–S model with $\beta = 0$ and $\gamma_2 = 0.1$.

only apply a control strategy to particles with velocities which deviate strongly from the consensus velocity. Fig. 5.4 (right) demonstrates that the NNPG method fails to capture the sparsity structure and ignores the time-dependence of the optimal feedback control. In fact, the NNPG methods yields a linear feedback control with very small magnitude, which suggests to apply an extremely moderate control strategy to all particles.

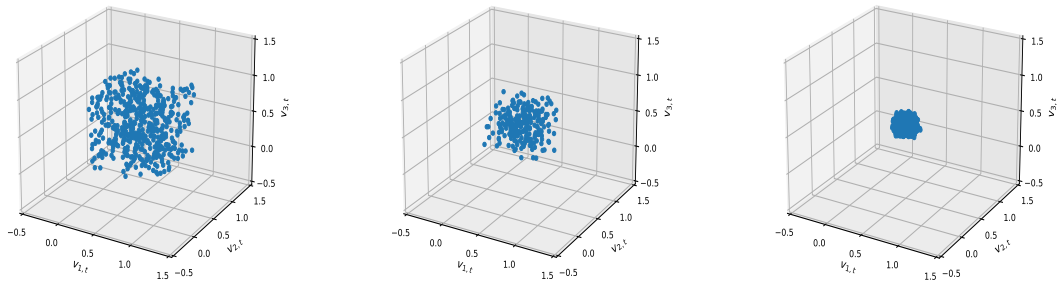


(a) FIPDE for $\gamma_2 = 0.1$ (b) FIPDE for $\gamma_2 = 0.3$ (c) NNPG for $\gamma_2 = 0.3$

Figure 5.4: Illustration of numerically obtained sparse feedback controls for six-dimensional C–S models with $\beta = 0$ and $\gamma_2 > 0$, where the controls are evaluated at different (t, v_2) and the remaining components are set to 0.5.

The last example concerns the non-convex MFC problem (5.12) and (5.13) with general non-zero communication rate $\beta = 1$ and $\gamma_2 = 0$. As above, we initialise the feedback maps ϕ^0, ψ^0 with the numerical solution obtained by the FIPDE method for $\beta = \gamma_2 = 0$ and again set $\tau = 1/5$. Further, we implement the residual approach by choosing the solution of above linear-quadratic problem as approximation for u^m at each NAG iteration. Note that with this choice of β the velocities would barely align without control interference. Due to the form of the interaction kernel in (5.12) and the nonlinear dependence of the feedback control on (t, x, v), we approximate the individual components of the residuals $\bar{u}^m$ and the feedback maps ϕ^m, ψ^m by feedforward neural networks with 6 layers

and 60 hidden dimensions. The effectiveness of the control strategy obtained by the FIPDE method (with 6 NAG iterations) is illustrated in Fig. 5.5, which compares the flocking behaviour of the velocities at terminal time in the controlled and uncontrolled case. We also implemented this setting with the NNPG method (based on the same neural network architecture) and wish to point out that although both methods yielded approximately the same value function, the NNPG method again did not capture important features of the optimal feedback map, such as the nonlinear time-dependence.



(a) Initial configuration, $t = 0$ (b) Uncontrolled velocity, $t = 1$ (c) Controlled velocity, $t = 1$

Figure 5.5: Uncontrolled and controlled velocities for six-dimensional C–S models with $\beta = 1$ at different times.

Chapter 6

Conclusion and outlook

In this thesis, we focused on the numerical approximation and well-posedness of McKean–Vlasov type equations in different settings. While Chapter 2 and Chapter 3 are mainly concerned with the strong approximation of solutions to McKean–Vlasov SDE with coefficients exhibiting super-linear growth, Chapter 4 considers such approximation properties and strong existence and uniqueness questions in a framework with linearly growing, but only piecewise Lipschitz continuous drifts. The main motivation to consider these equations under such conditions originates from various mean-field models in neuroscience (in particular, the FitzHugh–Nagumo model) and mean-field optimal control problems under control constraints appearing, for instance, in systemic risk.

To be more precise, in Chapter 2 and Chapter 3, we developed and studied stability and strong approximation properties of adaptive approximation methods and tamed Milstein schemes. Combined with propagation of chaos results, this yields approximations to solutions of McKean–Vlasov equations in a strong sense. We recover the strong convergence of order $1/2$ for the adaptive scheme taking the randomness of required time-steps into consideration and prove first order convergence for the tamed Milstein scheme. A Taylor expansion argument in the measure component of the coefficients using the so-called Lions derivative was an essential tool to obtain the formulation of these high-order schemes and to derive their strong convergence rates. Further, I wish to mention the works [16, 97], which additionally provide efficient multi-level Monte Carlo estimators for expected functionals of solutions to delay McKean–Vlasov equations without the need to simulate Lévy areas and general well-posedness results for McKean–Vlasov equations with common noise satisfying a certain monotonicity condition.

In Chapter 4, we applied a transformation technique inspired by [105] to prove the existence of a unique strong solution to McKean–Vlasov SDEs and associated interacting particle systems with drifts that are discontinuous in a finite number of spatial points. Due to the law-dependence of the transformation, a suitable fixed-point iteration on the Wasserstein space was required to establish well-posedness. To complement the theoretical study of such equations, we proposed schemes to approximate these particle systems in time, analysed their convergence rates and examined their

performance numerically. The proof of such strong convergence results relies on the structure of the transformation or occupation-time estimates to handle the drift discontinuities.

Chapter 5 is then dedicated to the numerical computation of optimal feedback controls for a broad class of high-dimensional MFC problems, including models with non-smooth costs and a general mean-field dependence through the control and the state process. Such models can be viewed as the asymptotic limit of N-interacting particles controlled by a central planner. As alluded to earlier, such models find numerous applications and have been intensively studied over the last decade. Due to certain limitations of existing methods, we proposed a novel iterative procedure in Chapter 5. At the core of our algorithm, we iteratively refine the optimal feedback control by applying an accelerated proximal gradient method to the (nonsmooth) MFC problem, which does not require solving the first order optimality condition, arising from the stochastic maximum principle, explicitly. However, in many cases it is not possible to find an analytic expression for the solution to the optimality condition, which shows the importance of iterative methods. In order to update the feedback map at each iteration, we need to solve a high-dimensional parabolic PDE system for the decoupling fields of the so-called adjoint processes. We rely on the Deep-Galerkin method in combination with a residual approximation technique to numerically obtain a solution to this (nonlocal) PDE system. By doing so, the optimal feedback map is recovered on the entire computational domain and not only along sample trajectories of the controlled state process, which is a drawback of sampling based algorithms such as the policy gradient method.

In the paper [136], we constructed a decentralised optimal feedback control, i.e., the feedback map is only dependent on time and state, but not law (the distribution of the state variable). As a natural continuation of this work, it would be of great interest to extend our algorithm to obtain a centralised control, i.e., to additionally capture the correct feedback behaviour in the law variable. A centralised feedback control obtained for a given random initial state would be more robust to changes in the distribution of the initial state than its decentralised counterpart. An essential step to obtain feedback in the law component is to solve the so-called master-equation for the decoupling field of the adjoint variables. This PDE is written on the Wasserstein space and is therefore infinite-dimensional. However, a particle approximation of the master-equation in combination with a suitable neural network architecture, which employs the symmetric nature of the finite-dimensional particle approximation [65], seems to be a promising way to tackle this problem. Also, to complement the numerical aspects with a convergence analysis in terms of value function and control approximation with respect to the number of particles and number of iterations would be a rewarding research direction.

Another computational aspect that is of interest concerns data-driven feedback controls. In [136], we assumed the controlled state process to be fully observable. However, in practical applications it often happens that only partial information is available. To design a computational framework

for approximating controls based on given information, techniques from optimal filtering, such as Zakai filters (Zakai SPDEs), need to be incorporated in our procedure. In a similar flavour, high-dimensional PDEs with random coefficients arise in the context of mean-field control problems with common noise. Here, the coefficients of the underlying controlled dynamics depend on a random measure and, further, the decoupling fields associated with the backward equations will be random fields.

A further question, related to Chapters 2 and 3 of this thesis, concerns the convergence analysis and study of higher-order numerical schemes (in particular time-stepping schemes) for path-dependent McKean–Vlasov equations. Path-dependent McKean–Vlasov SDEs appear, for example, in the context of local volatility models [66], in order to capture historic patterns of the volatility or to include quantities such as the running maximum in the model. To design and analyse such numerical methods, a notion of pathwise measure derivative (see [48]) needs to be employed.

Appendix A

Auxiliary results for Chapter 3

Here, we provide an alternative proof for the moment stability of the proposed Milstein scheme, in case of a measure-independent diffusion. Although, this setting is more restrictive than in Lemma 3.3.1, the proof presented here gives an insight into novel techniques needed to extent the methods from [81, 68] to tamed Milstein schemes for SDEs with non-commutative noise, which are interesting in their own right.

A.1 Proof of moment-stability

Proof. We utilise ideas developed in [81] for tamed schemes for standard SDEs. We will point these out and focus on the key differences to [81]. For $i \in \{1, \dots, N\}$, we define

$$A_n^i := \sum_{r,q=1}^{m} \left| \int_{t_n}^{t_{n+1}} \Delta W_s^{i,r} \, \mathrm{d}W_s^{i,q} \right|,$$

where $W^{i,r}$, for $r \in \{1, \dots, m\}$ denotes the r-th component of the Brownian motion W^i and $\Delta W_s^i = W_s^i - W_{t_n}^i$. Now, for $n \in \{0, \dots, M\}$, $M \in \mathbb{N}$, we introduce the sub-events

$$\Omega_n^{i,M} := \left\{ \omega_i \in \Omega \;\middle|\; \max_{k \in \{0,\dots,n-1\}} A_k^i(\omega_i) \leq \frac{1}{\sqrt{M}} \right\},$$
$$\Omega_{n,\lambda}^{i,M} := \left\{ \boldsymbol{\omega}_N \in \Omega^N \;\middle|\; \max_{k \in \{0,\dots,n-1\}} D_{k,\lambda}^i(\boldsymbol{\omega}_N) \leq M^{1/\alpha_1}, \; \max_{k \in \{0,\dots,n-1\}} |\Delta W_k^i(\omega_i)| \leq 1 \right\},$$

on which the quantities $D_{n,\lambda}^{i,M} : \Omega^N \to [0, \infty)$ given by

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

for all $i \in \{1, \dots, M\}$, $n \in \{0, \dots, M\}$, $M \in \mathbb{N}$, can be used to bound (3.6). Further, $\lambda \geq 1$ is some sufficiently large constant and $\tilde{A}_n^i : \Omega^N \to [0, \infty)$, for $i \in \{1, \dots, N\}$, $n \in \{0, \dots, M\}$, will be defined at a later stage of the proof. Also, we introduced the quantity

$$\alpha_n^{i,M} := \mathrm{I}_{|Y_{t_n}^{i,N}| \geq 1} \left\langle \frac{Y_{t_n}^{i,N}}{|Y_{t_n}^{i,N}|}, \frac{\sigma(Y_{t_n}^{i,N}) \Delta W_n^i}{|Y_{t_n}^{i,N}|} \right\rangle.$$

Let $Y_{t_n}^{i,N}$, for $n \in \{0,\dots,M\}$, be given by (3.6), then we aim to show that

$$\mathrm{I}_{\Omega_{n,\lambda}^{i,M}\cap\Omega_n^{i,M}}|Y_{t_n}^{i,N}|^2 \le D_{n,\lambda}^{i,M}, \tag{A.1}$$

for all $i \in \{1,\dots,N\}$, $n \in \{0,\dots,M\}$ and $M \in \mathbb{N}$. Inequality (A.1) is a pathwise estimate of each particle on the set of events $\Omega_{n,\lambda}^{i,M} \cap \Omega_n^{i,M}$ and is a version of [81, Lemma 3.1] adapted to particle systems. This lemma is the crucial result to obtain p-th moment bounds of tamed schemes.

First, note that on $\Omega_{n+1}^{i,M} \cap \Omega_{n+1,\lambda}^{i,M} \cap \{\boldsymbol{\omega}_N |\ |Y_{t_n}^{i,N}(\boldsymbol{\omega}_N)| \le 1\}$, we have due to Assumptions $(\mathbf{A}_b^1)$, $(\mathbf{A}_b^4)$, $(\mathbf{A}_\sigma^1)$, $(\mathbf{A}_\sigma^2)$ and $(\mathbf{A}_\sigma^4)$ that there exists some constant $\bar{\lambda} \ge 1$ such that

$$|Y_{t_{n+1}}^{i,N}| \le \bar{\lambda},$$

for all $n \in \{0,\dots,M-1\}$ and $M \in \mathbb{N}$, where $\bar{\lambda} \ge 1$ is a constant that only depends on the parameters appearing in the assumptions for b and σ, d, m and the terminal time $T > 0$.

In the sequel, we will define the set of events, where $Y_{t_n}^{i,N}$ attains large values:

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

Further, we obtain from standard inequalities,

$$\begin{aligned}
|Y_{t_{n+1}}^{i,N}|^2 &= \left|Y_{t_n}^{i,N} + b_h(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})\delta + \int_{t_n}^{t_{n+1}} \Upsilon_s^i \,\mathrm{d}W_s^i\right|^2 \\
&\le |Y_{t_n}^{i,N}|^2 + 2h^2|b(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})|^2 + 2\left|\int_{t_n}^{t_{n+1}} \Upsilon_s^i \,\mathrm{d}W_s^i\right|^2 \\
&\quad + \frac{2h}{1+h|b(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})|}\left\langle Y_{t_n}^{i,N}, b(Y_{t_n}^{i,N}, \mu_{t_n}^{\boldsymbol{Y}^N})\right\rangle + 2\left\langle Y_{t_n}^{i,N}, \int_{t_n}^{t_{n+1}} \Upsilon_s^i \,\mathrm{d}W_s^i\right\rangle.
\end{aligned}$$

From the one-sided Lipschitz assumption and the polynomial growth in the state variable in $(\mathbf{A}_b^1)$ and $(\mathbf{A}_b^4)$, we get on $\tilde{\Omega}_{n+1}^{i,M}$

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

Using $(\mathbf{A}_\sigma^1)$ and the assumption that the diffusion is measure-independent, we deduce for all $l \in \{1,\dots,m\}$ that

$$\|\sigma(Y_{t_n}^{i,N})\|^2 \lesssim |Y_{t_n}^{i,N}|^2, \qquad \|\partial_x\sigma_l(Y_{t_n}^{i,N})\sigma(Y_{t_n}^{i,N})\|^2 \lesssim |Y_{t_n}^{i,N}|^2. \tag{A.2}$$

By virtue of (A.2), we obtain the estimate

$$\left|\int_{t_n}^{t_{n+1}} \Upsilon_s^i \,\mathrm{d}W_s^i\right|^2 \lesssim |Y_{t_n}^{i,N}|^2|\Delta W_n^i|^2 + |Y_{t_n}^{i,N}|^2 A_n^i, \quad \text{on } \Omega_{n+1}^{i,M}.$$

Note that due to Assumption $(\mathbf{A}_\sigma^1)$

$$\left\langle Y_{t_n}^{i,N}, \int_{t_n}^{t_{n+1}} \Upsilon_s^i \,\mathrm{d}W_s^i\right\rangle \le \left\langle Y_{t_n}^{i,N}, \sigma(Y_{t_n}^{i,N})\Delta W_n^i\right\rangle + C|Y_{t_n}^{i,N}|^2 A_n^i.$$

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

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

where $L_n^{r,q}$ is a logistic random variable with zero mean and variance $\frac{1}{12}h^2$ and ΔH_n^q (the space time Lévy area) is normally distributed with zero mean and variance $\frac{1}{12}h^2$ (see [61, Chapter 7] and references therein for details). Using this, we may write

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

Therefore, on $\tilde{\Omega}_{n+1}^{i,M} \cap \Omega_{n+1}^{i,M}$, we have for some constant $C > 0$ (depending on d, m, T and on the constants appearing in the model assumptions)

$$\begin{aligned}
|Y_{t_{n+1}}^{i,N}|^2 \leq |Y_{t_n}^{i,N}|^2 + Ch|Y_{t_n}^{i,N}|^2 + C|Y_{t_n}^{i,N}|^2(h + |\Delta W_n^i|^2) + 2\left\langle Y_{t_n}^{i,N}, \sigma(Y_{t_n}^{i,N}) \Delta W_n^i \right\rangle \\
+ C|Y_{t_n}^{i,N}|^2 \tilde{A}_n^i.
\end{aligned}$$

Hence, it can be shown that there is a sufficiently large $\lambda \geq 1$ such that

$$|Y_{t_{n+1}}^{i,N}|^2 \leq |Y_{t_n}^{i,N}|^2 \exp\left(\frac{\lambda}{M} + \lambda(|\Delta W_n^i|^2 + \tilde{A}_n^i) + 2\alpha_n^i \right), \tag{A.3}$$

on $\tilde{\Omega}_{n+1}^{i,M} \cap \Omega_{n+1}^{i,M}$. Now, recalling the form of the moment generating function of a logistic random variable, for any positive real parameter $a > 0$ we can choose M large enough, such that

$$\mathbb{E}\left[\exp(a|L_n^{p,q}|)\right] = 2\int_{1/2}^1 t^{\frac{ah}{\pi}}(1-t)^{-\frac{ah}{\pi}} \,\mathrm{d}t \leq 1 + Ch \leq \exp(Ch), \tag{A.4}$$

for some constant $C > 0$ (independent of M and N).

An inductive argument as in [81, Lemma 3.1] allows us to deduce (A.1). Furthermore, due to (A.4) analogous statements to [81, Lemmas 3.2–3.8] still hold true. □

Appendix B

Auxiliary results for Chapter 4

B.1 L-differentiability of G^{-1}

Proposition B.1.1. *Let $G\colon \mathbb{R}\times\mathcal{P}_2(\mathbb{R})\to\mathbb{R}$ be defined by (4.7) with $c>0$ as in Remark 4.3.3 and let the assumptions of Proposition 4.3.3 be satisfied. Assume that G is L-differentiable, and $\mathbb{R}\ni y\mapsto\partial_\mu G(x,\mu)(y)$ is in $\mathcal{C}^1(\mathbb{R},\mathbb{R})$ for all $(x,\mu)\in\mathbb{R}\times\mathcal{P}_2(\mathbb{R})$. Then, the inverse $G^{-1}\colon\mathbb{R}\times\mathcal{P}_2(\mathbb{R})\to\mathbb{R}$ is continuously L-differentiable with*

$$\partial_\mu G^{-1}(x,\mu)(y)=\frac{-\partial_\mu G(G^{-1}(x,\mu),\mu)(y)}{\partial_x G(G^{-1}(x,\mu),\mu)}, \tag{B.1}$$

for all $(x,\mu,y)\in\mathbb{R}\times\mathcal{P}_2(\mathbb{R})\times\mathbb{R}$.

Proof. First, we remark that (H.8(2)) guarantees the L-differentiability of G. Let now $X,Y\in L_2(\Omega,\mathcal{F},\mathbb{P};\mathbb{R})$ with $\mathcal{L}_X=\mu$ and $\mathcal{L}_{X+Y}=\nu$, for $\mu,\nu\in\mathcal{P}_2(\mathbb{R})$. Consider the lifted inverse function $\tilde{G}^{-1}$ defined by $\tilde{G}^{-1}(y,X):=G^{-1}(y,\mu)$, for any $y\in\mathbb{R}$. Similarly, we set $\tilde{G}(x,X):=G(x,\mu)$ and $\tilde{G}(x,X+Y):=G(x,\nu)$, for any $x\in\mathbb{R}$. Now fix $y\in\mathbb{R}$ and set $x:=\tilde{G}^{-1}(y,X)$, $h:=\tilde{G}^{-1}(y,X+Y)-x$.

Observe that $y=G(G^{-1}(y,\nu),\nu)=G(x+h,\nu)$ and also $y=G(G^{-1}(y,\mu),\mu)=G(x,\mu)$. In addition, due to the Lipschitz continuity of $\mu\mapsto G^{-1}(y,\mu)$, we have

$$|h|=|G^{-1}(y,\nu)-G^{-1}(y,\mu)|\le L\mathcal{W}_2(\nu,\mu)\le L\|Y\|_{L_2}. \tag{B.2}$$

In what follows, we aim to show (B.1):

$$\frac{\left|\tilde{G}^{-1}(y,X+Y)-\tilde{G}^{-1}(y,X)-\left\langle\frac{-\partial_\mu G(x,\mu)(X)}{\partial_x G(x,\mu)},Y\right\rangle_{L_2}\right|}{\|Y\|_{L_2}}\le C\frac{\left|-\partial_x G(x,\mu)h-\langle\partial_\mu G(x,\mu)(X),Y\rangle_{L_2}\right|}{\|Y\|_{L_2}},$$

for some constant $C>0$, where we used the boundedness of $x\mapsto\partial_x G(x,\mu)$. Employing this estimate, along with the identity $\tilde{G}(x,X)=\tilde{G}(x+h,X+Y)$ and (B.2), we obtain

$$\frac{\left|\tilde{G}^{-1}(y,X+Y)-\tilde{G}^{-1}(y,X)-\left\langle\frac{-\partial_\mu G(x,\mu)(X)}{\partial_x G(x,\mu)},Y\right\rangle_{L_2}\right|}{\|Y\|_{L_2}}$$

$$
\begin{aligned}
&\leq C\frac{\left|-\partial_x G(x,\mu)h+\tilde{G}(x+h,X+Y)-\tilde{G}(x,X)-\langle\partial_\mu G(x,\mu)(X),Y\rangle_{L_2}\right|}{\|Y\|_{L_2}}\\
&\leq C\frac{\left|\tilde{G}(x+h,X+Y)-\tilde{G}(x,X+Y)-\partial_x G(x,\nu)h\right|}{\|Y\|_{L_2}}+C\frac{|\partial_x G(x,\nu)h-\partial_x G(x,\mu)h|}{\|Y\|_{L_2}}\\
&\quad+C\frac{\left|\tilde{G}(x,X+Y)-\tilde{G}(x,X)-\langle\partial_\mu G(x,\mu)(X),Y\rangle_{L_2}\right|}{\|Y\|_{L_2}}\\
&\leq CL\frac{|G(x+h,\nu)-G(x,\nu)-\partial_x G(x,\nu)h|}{|h|}+CL\left|\partial_x\tilde{G}(x,X+Y)-\partial_x\tilde{G}(x,X)\right|\\
&\quad+C\frac{\left|\tilde{G}(x,X+Y)-\tilde{G}(x,X)-\langle\partial_\mu G(x,\mu)(X),Y\rangle_{L_2}\right|}{\|Y\|_{L_2}}.
\end{aligned}
$$

Now, if $\|Y\|_{L_2}\to 0$, then also $|h|\to 0$, so the terms in the last estimate tend to zero from which the claim follows. Furthermore, we remark that the mapping $y\mapsto\frac{-\partial_\mu G(G^{-1}(x,\mu),\mu)(y)}{\partial_x G(G^{-1}(x,\mu),\mu)}$ is in $\mathcal{C}^1(\mathbb{R},\mathbb{R})$. □

B.2 The class $\mathscr{C}$

For a given $N\in\mathbb{N}$, we define for all $k\in\{1,\dots,N\}$ the sets

$$\Theta^k:=\{\boldsymbol{x}^N=(x_1,\dots,x_N)^\top : x_k=0\}.$$

Definition B.2.1. For $N\in\mathbb{N}$, we define the class $\mathscr{C}$ of functions $\boldsymbol{G}_N\colon\mathbb{R}^N\to\mathbb{R}^N$ with the following properties:

(p1) $\boldsymbol{G}_N$ is in $\mathcal{C}^1(\mathbb{R}^N,\mathbb{R}^{N\times N})$;

(p2) for all $\boldsymbol{x}^N\in\mathbb{R}^N$, $\det\left(\boldsymbol{G}_N'(\boldsymbol{x}^N)\right)\neq 0$;

(p3) $\lim_{|\boldsymbol{x}^N|\to\infty}|\boldsymbol{G}_N(\boldsymbol{x}^N)|=\infty$;

(p4) for all $k\in\{1,\dots,N\}$ and all $\boldsymbol{x}^N\in\Theta^k$, we have $G_k(\boldsymbol{x}^N)=0$;

(p5) for all $j,k\in\{1,\dots,N\}$ and all $\boldsymbol{x}^N\in\Theta^k$, we have $\partial_{x_j}G_k(\boldsymbol{x}^N)=\delta_{j,k}$;

(p6) for all $i,j,k\in\{1,\dots,N\}$ with $i\neq j$ the mixed partial derivative $\partial_{x_i}\partial_{x_j}G_k$ exists and is continuous;

(p7) for all $j,k\in\{1,\dots,N\}$ with $j\neq k$ the second partial derivative $\partial^2_{x_j}G_k$ exists and is continuous;

(p8) for all $k\in\{1,\dots,N\}$ the second partial derivative $\partial^2_{x_k}G_k$ exists on $\mathbb{R}^N\setminus\Theta^k$ and is continuous there.

The aim of this appendix is to prove the following theorem:

Theorem B.2.1. *If $\boldsymbol{G}_N\in\mathscr{C}$, then $\boldsymbol{G}_N$ is invertible with $\boldsymbol{G}_N^{-1}\in\mathscr{C}$.*

Proof. Let $\boldsymbol{G}_N = (G_1, \dots, G_N)^\top \in \mathscr{C}$. The Hadamard global inverse function theorem states that under assumptions (p1), (p2), and (p3), $\boldsymbol{G}_N$ is invertible with inverse in $\mathcal{C}^1(\mathbb{R}^N, \mathbb{R}^{N\times N})$. We denote $\boldsymbol{H}_N = \boldsymbol{G}_N^{-1}$ and conclude

(p1) $\boldsymbol{H}_N \in \mathcal{C}^1(\mathbb{R}^N, \mathbb{R}^{N\times N})$;

(p2) for all $\boldsymbol{x}^N \in \mathbb{R}^d$, $\det\left(\boldsymbol{H}'_N(\boldsymbol{x}^N)\right) \neq 0$;

(p3) $\lim_{|\boldsymbol{x}^N|\to\infty} |\boldsymbol{H}_N(\boldsymbol{x}^N)| = \infty$.

Let $k \in \{1, \dots, N\}$ and fix values $x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_N \in \mathbb{R}$. As $\boldsymbol{G}_N$ has a unique global inverse, the mapping $\mathbb{R} \ni x_k \mapsto G_k(x_1, \dots, x_N)$ is invertible, i.e., there is precisely one x_k with $G_k(x_1, \dots, x_N) = 0$. On the other hand, $G_k(x_1, \dots, x_{k-1}, 0, x_{k+1}, \dots, x_N) = 0$, and consequently we proved:

(p4) for all $k \in \{1, \dots, N\}$ and all $\boldsymbol{y} \in \Theta^k$, we have $H_k(\boldsymbol{y}) = 0$.

We note that, since $\boldsymbol{H}_N \circ \boldsymbol{G}_N$ is the identity on $\mathbb{R}^N$

$$\partial_{x_j}(H_k \circ \boldsymbol{G}_N) = \delta_{j,k},$$

and therefore, using (p5) for $\boldsymbol{G}_N$, we have for $\boldsymbol{x}^N \in \Theta^k$

$$\delta_{j,k} = \sum_{l=1}^N \partial_{y_l} H_k \circ \boldsymbol{G}_N(\boldsymbol{x}^N)\, \partial_{x_j} G_l(\boldsymbol{x}^N) = \sum_{l=1}^N \partial_{y_l} H_k \circ \boldsymbol{G}_N(\boldsymbol{x}^N)\, \delta_{j,l} = \partial_{y_j} H_k \circ \boldsymbol{G}_N(\boldsymbol{x}^N).$$

Now if $\boldsymbol{y}^N \in \Theta^k$, then $\boldsymbol{H}_N(\boldsymbol{y}^N) \in \Theta^k$, and therefore $\partial_{y_j} H_k(y) = \partial_{y_j} H_k \circ \boldsymbol{G}_N(\boldsymbol{H}_N(\boldsymbol{y}^N)) = \delta_{j,k}$. To summarise, we have shown:

(p5) for all $j, k \in \{1, \dots, N\}$ and all $\boldsymbol{y}^N \in \Theta^k$, we have $\partial_{y_j} H_k(\boldsymbol{y}^N) = \delta_{j,k}$.

To prove (p6)-(p8), we first note that for any $j, k \in \{1, \dots, N\}$

$$\partial_{y_j} G_k(\boldsymbol{H}_N(\boldsymbol{y}^N)) = \sum_{l=1}^N \partial_{x_l} G_k(\boldsymbol{H}_N(\boldsymbol{y}^N)) \partial_{y_j} H_l(\boldsymbol{y}^N) = \delta_{j,k},$$

and hence

$$\left(\partial_{y_j} H_l(\boldsymbol{y}^N)\right)_{l\in\{1,\dots,N\}} = \left((\boldsymbol{G}'_N)^{-1}(\boldsymbol{H}_N(\boldsymbol{y}^N))\right)_j,$$

where the subindex denotes the j-th column of $(\boldsymbol{G}'_N)^{-1}$. The higher order regularity properties of $\boldsymbol{H}_N$ follow from this expression and the second order differentiability properties of $\boldsymbol{G}_N$.

□

Appendix C

Additional material for Chapter 5

C.1 Stochastic maximum principle

In this section, we will introduce the stochastic maximum principle for MFC problems (5.1) and (5.2) with $\ell \equiv 0$, where (a) the control problem depends only on the law of the controlled state and (b) the control problem depends on the joint law of the controlled state and the control process itself. We will see that the first order optimality condition is crucially different for (a) and (b), due to the appearance of a non-local term in the second case. In particular, note that the difficulty appearing in this second framework is one of the main motivations to propose an iterative numerical scheme presented in Section 5.2. After introducing the stochastic maximum principle, we will shortly discuss how to reformulate (5.1) and (5.2) as a coupled mean-field FBSDE. Note that we restrict the discussion to the setting of a controlled drift coefficient only, as the numerical examples, which we consider in this chapter have an uncontrolled diffusion. Let now $H : [0,T] \times \mathbb{R}^d \times A \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^k) \times \mathbb{R}^d \times \mathbb{R}^{d\times m} \to \mathbb{R}$, where A is a nonempty closed convex set $A \subset \mathbb{R}^k$, be the Hamiltonian associated with (5.1) and (5.2) (with uncontrolled diffusion and $\ell \equiv 0$) defined as follows:

$$H(t,x,a,\eta,y,z) := \langle b(t,x,a,\eta), y\rangle + \langle \sigma(t,x,\pi_1\sharp\eta), z\rangle + f(t,x,a,\eta), \tag{C.1}$$

where $\pi_1\sharp\eta$ denotes the first marginal of the measure η. In case the coefficients in (5.1) and (5.2) only depend on the marginal laws of X^α, the Hamiltonian can be introduced in similar manner. Now, for any initial state $\xi \in L_0^2(\mathbb{R}^d)$ and any given admissible control $\alpha \in \mathcal{H}^2(A)$, let $X^\alpha \in \mathcal{S}^2(\mathbb{R}^d)$ be the corresponding controlled process solving (5.1), and let $(Y^\alpha, Z^\alpha) \in \mathcal{S}^2(\mathbb{R}^d) \times \mathcal{H}^2(\mathbb{R}^{d\times m})$ be the so-called adjoint processes of X^α, which satisfy the following mean-field BSDE: For all $t \in [0,T]$

$$\begin{aligned}
\mathrm{d}Y_t^\alpha &= -\big(\partial_x H(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)}, Y_t^\alpha, Z_t^\alpha) \\
&\qquad + \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{X}_t^\alpha, \tilde{\alpha}_t, \mathcal{L}_{(X_t^\alpha,\alpha_t)}, \tilde{Y}_t^\alpha, \tilde{Z}_t^\alpha)(X_t^\alpha,\alpha_t)]\big)\,\mathrm{d}t + Z_t^\alpha\,\mathrm{d}W_t, \\
Y_T^\alpha &= \partial_x g(X_T^\alpha, \mathcal{L}_{X_T^\alpha}) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T^\alpha, \mathcal{L}_{X_T^\alpha})(X_T^\alpha)],
\end{aligned} \tag{C.2}$$

where we used here the tilde notation to refer to an independent copy of a random variable.

Having the above preliminary notions in mind, we return to the formulation of the optimality conditions, starting with the case that the coefficients in (5.1) and (5.2) only depend on the marginal laws of X^α: Under certain convexity assumptions on the Hamiltonian and the terminal condition and regularity conditions on (b, σ, f, g), the sufficient part of the stochastic maximum principle asserts that if

$$H(t, X_t^\alpha, \alpha_t, \mathcal{L}_{X_t^\alpha}, Y_t^\alpha, Z_t^\alpha) = \inf_{a \in A} H(t, X_t^\alpha, a, \mathcal{L}_{X_t^\alpha}, Y_t^\alpha, Z_t^\alpha), \quad \mathrm{d}t \otimes \mathrm{d}\mathbb{P}, \text{ a.e.},$$

then α is the unique optimal control, i.e., the minimiser of (5.2). A similar statement can be given in the case that A is an open set or the admissible optimal control attains values in the interior of A. Suppose now the function $A \ni a \mapsto H(t, x, a, \mu, y, z)$, for any given $(t, x, \mu, y, z) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}) \times \mathbb{R}^d \times \mathbb{R}^{d \times m}$, is continuous and strictly convex. Then, this implies that there exists a unique minimiser of the Hamiltonian represented by $\hat{\alpha}(t, x, \mu, y, z)$, that is $\hat{\alpha}(t, x, \mu, y, z) = \arg\min_{a \in A} H(t, x, a, \mu, y, z)$. The power of the stochastic maximum principle is therefore that it turns the original optimisation problem formulated in (5.1) and (5.2) into a pointwise minimisation problem. Using the control $\hat{\alpha}_t = \hat{\alpha}(t, X_t, \mathcal{L}_{X_t}, Y_t, Z_t)$ for $t \in [0, T]$, where $(X_t, Y_t, Z_t)_{t \in [0,T]}$ are defined by (5.1) and (C.2) (with α_t replaced by $\hat{\alpha}_t$) yields a so-called coupled mean-field FBSDE, as the forward part then appears in the backward part and vice versa.

Allowing now dependence on the law of the control process [1, Theorem 3.5] (again requiring certain convexity and regularity assumptions) asserts that the maximum principle reads as follows: If the optimality condition

$$\left\langle \partial_a H(t, X_t^\alpha, \alpha_t, \mathcal{L}_{(X_t^\alpha, \alpha_t)}, Y_t^\alpha, Z_t^\alpha) + \tilde{\mathbb{E}}[\partial_\nu H(t, \tilde{X}_t^\alpha, \tilde{\alpha}_t, \mathcal{L}_{(X_t^\alpha, \alpha_t)}, \tilde{Y}_t^\alpha, \tilde{Z}_t^\alpha)(X_t^\alpha, \alpha_t)], \alpha_t - a \right\rangle \leq 0, \tag{C.3}$$

holds for all $a \in A$, $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$ a.e., then $\alpha \in \mathcal{H}^2(A)$ is the optimal control of (5.2). As pointed out in [1], the optimality condition (C.3), in general, (due to the possibly nonlinear dependence on the law of the control) cannot be improved into a pointwise minimisation condition as in the case of standard MFC problems outlined above. Therefore, in this framework, it is not clear if there exists a deterministic feedback map from the state and adjoint processes to the action set, satisfying (C.3) pointwise. In [135], we identify structural conditions on the functions (b, f) to ensure the existence and certain regularity properties of a feedback map $[0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) \ni (t, x, y, \chi) \mapsto \hat{\alpha}(t, x, y, \chi) =: \hat{\alpha}_t$, see, Propositions 2.1 and 2.2 and Example 2.3 in [135] for details. The resulting coupled mean-field FBSDE can be formulated as follows:

$$\mathrm{d}X_t = \hat{b}(t, X_t, Y_t, \mathcal{L}_{(X_t, Y_t)})\, \mathrm{d}t + \sigma(t, X_t, \mathcal{L}_{X_t})\, \mathrm{d}W_t, \quad X_0 = \xi, \tag{C.4a}$$

$$\mathrm{d}Y_t = -\hat{f}(t, X_t, Y_t, Z_t, \mathcal{L}_{(X_t, Y_t, Z_t)})\, \mathrm{d}t + Z_t\, \mathrm{d}W_t, \qquad Y_T = \hat{g}(X_T, \mathcal{L}_{X_T}), \tag{C.4b}$$

where, we introduced the coefficients

$$\begin{aligned}
\hat{b}(t,x,y,\chi) &= b(t,x,\hat{\alpha}(t,x,y,\chi),\phi(t,\chi)),\\
\hat{f}(t,x,y,z,\rho) &= \partial_x H(t,x,\hat{\alpha}(t,x,y,\pi_{1,2}\sharp\rho),\phi(t,\pi_{1,2}\sharp\rho),y,z)\\
&\quad + \int_E \partial_\mu H(t,\tilde{x},\hat{\alpha}(t,\tilde{x},\tilde{y},\pi_{1,2}\sharp\rho),\phi(t,\pi_{1,2}\sharp\rho),\tilde{y},\tilde{z})(x,\hat{\alpha}(t,x,y,\pi_{1,2}\sharp\rho))\,\rho(\mathrm{d}\tilde{x},\,\mathrm{d}\tilde{y},\,\mathrm{d}\tilde{z}),\\
\hat{g}(x,\mu) &= \partial_x g(x,\mu) + \int_{\mathbb{R}^d} \partial_\mu g(\tilde{x},\mu)(x)\,\mu(\mathrm{d}\tilde{x}),
\end{aligned} \tag{C.5}$$

for $(t,x,y,z,a,\mu,\chi,\rho) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^{d\times m}\times A\times\mathcal{P}_2(\mathbb{R}^d)\times\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d)\times\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^{d\times m})$, with $E := \mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^{d\times m}$, $\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^k) \ni \phi(t,\chi) := \chi\circ\big(\mathbb{R}^d\times\mathbb{R}^d \ni (x,y) \mapsto (x,\hat{\alpha}(t,x,y,\chi)) \in \mathbb{R}^d\times A\big)^{-1}$, for all $(t,\chi)\in[0,T]\times\mathbb{R}^d\times\mathbb{R}^d$ and $\pi_{1,2}\sharp\rho := \rho(\cdot\times\mathbb{R}^{d\times m})$ is the marginal of the measure ρ on $\mathbb{R}^d\times\mathbb{R}^d$. Note that $\phi(t,\mathcal{L}_{(X_t^{\hat{\alpha}},Y_t^{\hat{\alpha}})})$ is the joint law of $(X_t^{\hat{\alpha}},\hat{\alpha}_t)$. Well-posedness and regularity results of the solution triple to the coupled mean-field FBSDE given by (C.4a), (C.4b) and (C.5) are presented in [135, Theorem 3.4].

Remark C.1.1. Numerical schemes for mean-field coupled FBSDEs of the form specified by (C.4a), (C.4b) and (C.5) have been developed, e.g., in [34] using a machine learning approach, in particular the Deep BSDE solver, in [133] employing a Markovian iteration algorithm and a posteriori error estimates (based on ideas of [20] for standard BSDEs), in [8, 40] using a variation of the method of continuation. The difficulty in designing numerical schemes for coupled FBSDEs in general arises from the fact that the Picard iteration approach, in order to decouple the forward and backward equations, fails if the coupling between forward and backward part is too strong and/or the time horizon is too large, see e.g., [153, 23, 21, 40, 54]. We refer to [71] for further machine learning algorithms developed for MFC problems in discrete time. An recent survey for numerical schemes for MFGs can be found in [6]. Recall that in our numerical algorithm, we do not rewrite the control problem as coupled FBSDE, but instead solve at each iteration a decoupled system.

C.2 Regression Monte–Carlo algorithm

The aim of this section is to apply our proposed numerical scheme to a MFC problem appearing in optimal liquidation [1]. In the first part, we will consider a linear-quadratic problem for which the solution for the optimal control is known analytically. This will allow us to verify the correctness and accuracy of our method. In the next step, we will modify the linear-quadratic framework by including an ℓ_1-norm regularisation term (i.e., a non-differentiable function) in the running cost, which models, for instance, transactions costs. In this framework, we will employ a proximal gradient descent algorithm. Here, we will use a regression Monte–Carlo algorithm to solve the adjoint equations at each NAG iteration.

To be precise, in order to obtain a numerical approximation to the solution $(X_t^m, Y_t^m, Z_t^m)_{t\in[0,T]}$ of a FBSDE at the m-th NAG iteration, we employ an Euler–Maruyama discretisation for (5.4),

an explicit backward Euler discretisation for (5.3) and the least-squares MC approximation for conditional expectations (see e.g., [70]). This will resemble the Markovian iteration scheme proposed in [21] for solving weakly coupled FBSDEs without mean-field interaction.

We now briefly outline the main steps of the numerical procedure in the case where the diffusion is uncontrolled: Let $M \in \mathbb{N}$, $K \in \mathbb{N}$ be given integers, $\{t_j\}_{j\in\{0,\dots M\}}$ be the uniform partition of $[0,T]$ with step-size $h = T/M$, $\gamma = \{\gamma_k\}_{k\in\{1,\dots,K\}}$ be a set of basis functions on $\mathbb{R}^d$. We assume that the optimal control is in feedback form, i.e., for all $t \in [0,T]$, $\alpha_t^m = \phi^m(X_t^m)$ for some map $\phi^m : \mathbb{R}^d \to \mathbb{R}^k$. In our numerical procedure each component of the map ϕ^m is represented (at each time-step t_j) by K local basis functions and a parameter $\hat{\phi}_{l,j}^m \in \mathbb{R}^K$, i.e., we have $\phi_{t_j}^m := ((\hat{\phi}_{1,j}^m)^\top\gamma, \dots, (\hat{\phi}_{k,j}^m)^\top\gamma)^\top : \mathbb{R}^d \to \mathbb{R}^k$. Similarly, we have $\psi_{t_j}^m := ((\hat{\psi}_{1,j}^m)^\top\gamma, \dots, (\hat{\psi}_{k,j}^m)^\top\gamma)^\top : \mathbb{R}^d \to \mathbb{R}^k$, with $\hat{\psi}_{t_j}^0 = \hat{\phi}_{t_j}^0$, for $j \in \{0,\dots,M\}$. In a next step, we define $\mu_{m,t_j}^{\boldsymbol{X}^N}(\mathrm{d}x) := \frac{1}{N}\sum_{i=1}^N \delta_{X_{t_j}^{m,i,N}}(\mathrm{d}x)$ and $\mu_{m,t_j}^{\boldsymbol{X}^N,\psi}(\mathrm{d}x) := \frac{1}{N}\sum_{i=1}^N \delta_{(X_{t_j}^{m,i,N},\psi_{t_j}^m(X_{t_j}^{m,i,N}))}(\mathrm{d}x)$, where for $j \in \{0,\dots,M-1\}$

$$X_{t_{j+1}}^{m,i,N} = X_{t_j}^{m,i,N} + b\left(t_j, X_{t_j}^{m,i,M}, \psi_{t_j}^m(X_{t_j}^m), \mu_{m,t_j}^{\boldsymbol{X}^N,\psi}\right)h + \sigma(t_j, X_{t_j}^{m,i,N}, \mu_{m,t_j}^{\boldsymbol{X}^N})\,\Delta W_j^i. \tag{C.6}$$

Here, we used $N \in \mathbb{N}$ particles to approximate the distributions $\mathcal{L}_{X_{t_j}^m}$ and $\mathcal{L}_{(X_{t_j}^m,\psi_{t_j}^m(X_{t_j}^m))}$ and defined $\Delta W_j^i = W_{t_{j+1}}^i - W_{t_j}^i$. Note that $(X_0^{m,i,N})_{i\in\{1,\dots N\}}$ and $(W^i)_{i\in\{1,\dots N\}}$ are independent copies of ξ and W, respectively.

Given an approximate solution to (5.4) obtained by (C.6), we aim to determine the following approximate decoupling fields on $\{t_j\}_{j\in\{0,\dots M\}}$ (i.e., solutions to (5.3)):

$$\hat{y}_{t_j}^m := ((\hat{\rho}_{1,j}^m)^\top\gamma, \dots, (\hat{\rho}_{d,j}^m)^\top\gamma)^\top : \mathbb{R}^d \to \mathbb{R}^d, \quad \hat{z}_{t_j}^m := ((\hat{\varrho}_{1,j}^m)^\top\gamma, \dots, (\hat{\varrho}_{d,j}^m)^\top\gamma)^\top : \mathbb{R}^d \to \mathbb{R}^d,$$

where for each $j \in \{0,\dots,M-1\}$ and $l \in \{1,\dots,d\}$, $\hat{\rho}_{l,j}^m, \hat{\varrho}_{l,j}^m \in \mathbb{R}^K$ are some unknown deterministic weights to be determined. Once the weights $(\hat{\boldsymbol{\rho}}_j^m, \hat{\boldsymbol{\varrho}}_j^m)_{j\in\{0,\dots,M-1\}}$, where $\mathbb{R}^{K\times d} \ni \hat{\boldsymbol{\rho}}_j^m := (\hat{\rho}_{1,j}^m, \dots, \hat{\rho}_{d,j}^m)$ and $\mathbb{R}^{K\times d} \ni \hat{\boldsymbol{\varrho}}_j^m := (\hat{\varrho}_{1,j}^m, \dots, \hat{\varrho}_{d,j}^m)$ have been found, we may define for each $j \in \{0,\dots,M-1\}$ and $i \in \{1,\dots,N\}$, $\hat{Y}_{t_j}^{m,i,N} := \hat{y}_{t_j}^m(X_{t_j}^{m,i,N})$ and $\hat{Z}_{t_j}^{m,i,N} := \hat{z}_{t_j}^m(X_{t_j}^{m,i,N})$ as the numerical approximation to the adjoint equations. To compute the weights $(\hat{\boldsymbol{\rho}}_j^m, \hat{\boldsymbol{\varrho}}_j^m)_{j\in\{0,\dots,M-1\}}$, we employ the least-squares MC regression method:

Set $\hat{y}_M^m(x) = \partial_x g(x, \mu_{m,T}^{\boldsymbol{X}^N}) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T^m, \mu_T^{\boldsymbol{X}^N})(x)]$ for all $x \in \mathbb{R}^d$, and for $j = M-1,\dots,0$, let

$$\hat{\boldsymbol{\varrho}}_j^m \in \operatorname*{arg\,min}_{\varrho\in\mathbb{R}^{K\times d}} \sum_{i=1}^N \left| \frac{\Delta W_j^i}{h}\hat{y}_{t_{j+1}}^m(X_{t_{j+1}}^{m,i,N}) - \varrho^\top\gamma(X_{t_j}^{m,i,N}) \right|^2,$$

$$\hat{\boldsymbol{\rho}}_j^m \in \operatorname*{arg\,min}_{\rho\in\mathbb{R}^{K\times d}} \sum_{i=1}^N \Big| \hat{y}_{t_{j+1}}^m(X_{t_{j+1}}^{m,i,N}) + h\Big(\partial_x H(t_j, X_{t_j}^{m,i,N}, \psi_{t_j}^m(X_{t_j}^{m,i,N}), \mu_{m,t_j}^{\boldsymbol{X}^N,\psi}, \hat{y}_{t_{j+1}}^m(X_{t_{j+1}}^{m,i,N}), \hat{z}_{t_j}^m(X_{t_j}^{m,i,N}))$$
$$+ \tilde{\mathbb{E}}\left[\partial_\mu H(t_j, \tilde{X}_{t_j}^{m,i,N}, \psi_{t_j}^m(\tilde{X}_{t_j}^{m,i,N}), \mu_{m,t_j}^{\boldsymbol{X}^N,\psi}, \hat{y}_{t_{j+1}}^m(\tilde{X}_{t_{j+1}}^{m,i,N}), \hat{z}_{t_j}^m(\tilde{X}_{t_j}^{m,i,N}))(X_{t_j}^{m,i,N})\right]\Big) - \rho^\top\gamma(X_{t_j}^{m,i,N})\Big|^2.$$

Then, we update the coefficients in the feedback representation for the control, based on the NAG algorithm, and repeat the above step.

C.2.1 Problem formulation

Optimal liquidation refers to the problem, where market participants (symmetric agents) try to liquidate all their assets by a given terminal time $T > 0$. The goal of the agents is to balance the trade-off between future risks in the asset price and market impact of trades (either permanent or temporary impacts). In the cooperative equilibrium framework, the asymptotic price is modelled by the simple controlled McKean–Vlasov SDE

$$\mathrm{d}S_t = \lambda \mathbb{E}[\alpha_t]\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad S_0 = s_0 \in \mathbb{R},$$

where $\lambda\mathbb{E}[\alpha_t]$ ($\lambda \geq 0$) reflects the permanent market impact and the control $(\alpha_t)_{t\in[0,T]}$ is the trading speed. The inventory and wealth processes are described by the ODEs

$$\mathrm{d}Q_t = \alpha_t\,\mathrm{d}t, \quad Q_0 = q_0 \in L^2_0(\mathbb{R}), \qquad \mathrm{d}X_t = -\alpha_t(S_t + k\alpha_t)\,\mathrm{d}t, \quad X_0 = 0,$$

respectively, for some parameter $k > 0$. Note that all processes introduced above are one-dimensional taking values in an open set. Our cost functional is given by

$$J(\alpha; s_0, q_0) = \mathbb{E}\left[\int_0^T Q_t^2\,\mathrm{d}t - Q_T(S_T - AQ_T) - X_T\right],$$

which penalises inventory and the term $Q_T(S_T - AQ_T)$ denotes the value of the assets liquidated at the terminal time (incorporating a penalisation controlled by the parameter A). Equivalently, we may write

$$\mathbb{E}\left[\int_0^T (Q_t^2 + \alpha_t S_t + k\alpha_t^2)\,\mathrm{d}t - Q_T(S_T - AQ_T)\right], \tag{C.7}$$

which means that our controlled state process is described by $(S_t, Q_t)_{t\in[0,T]}$. Therefore, the adjoint equations have the form

$$\begin{aligned}
\mathrm{d}Y_{1,t} &= -\alpha_t\,\mathrm{d}t + Z_{1,t}\,\mathrm{d}W_t, \quad Y_{1,T} = -Q_T,\\
\mathrm{d}Y_{2,t} &= -2Q_t\,\mathrm{d}t + Z_{2,t}\,\mathrm{d}W_t, \quad Y_{2,T} = -S_T + 2AQ_T.
\end{aligned} \tag{C.8}$$

Non-differentiable cost function:

In addition to above, we may consider adding an ℓ_1-regularisation term to (C.7), i.e.,

$$J(\alpha; s_0, q_0) = \mathbb{E}\left[\int_0^T (Q_t^2 + \alpha_t S_t + k\alpha_t^2 + \gamma|\alpha_t|)\,\mathrm{d}t - Q_T(S_T - AQ_T)\right], \tag{C.9}$$

for $\gamma \geq 0$.

C.2.2 Numerical illustration

For our numerical tests, we make the following parameter choices: $T = 1$, $s_0 = 2, q_0 \sim \mathcal{U}(1,2)$, $k = 1$, $\sigma = 0.7$, $\gamma \in \{0, 1\}$ and $A = \lambda = 0.5$. The computational domain is chosen as $\mathcal{D} =$

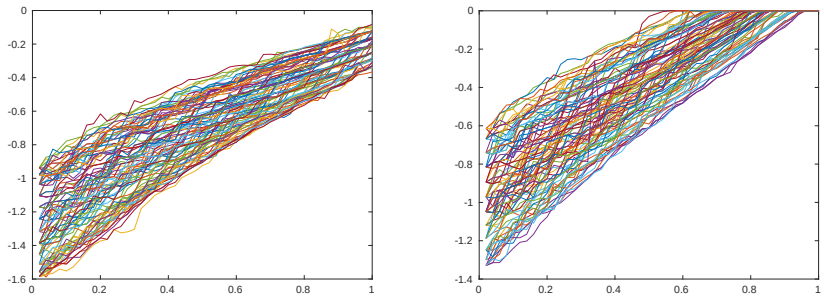
$[-2,6] \times [0,4]$, and we choose a piecewise constant basis function representation (with $K = 2500$) for the feedback map, i.e., the control α_t is a deterministic function of S_t and Q_t at each time t. The regression MC step is based on $N = 10^4$ forward samples and we partition the interval $[0,T]$ into $M = 50$ equidistant subintervals. The NAG algorithm uses the step-size $\tau = 1/6$.

Using the procedure described in the previous section, we obtain a piecewise-constant representation of the control. Sample trajectories of the numerically computed optimal control are depicted in Fig. C.1a (for $\gamma = 0$) and Fig. C.1b (in case of additional ℓ_1-regularisation). In both cases, we see that an agent tries to liquidate her assets at every time. Further, we observe the sparsity induced by the ℓ_1-norm, which is reflected by the fact that the control actions of the agents is smaller than in the case for $\gamma = 0$ and reaches zero before the terminal time.

Now, we investigate the robustness of our method in terms of perturbations of model parameters (here, we consider perturbations with respect to k and λ) and in terms of the initial distribution. To do so, we use the precomputed control (computed based on the parameters presented above with $\gamma = 0$) to obtain a numerical approximation of the value function for different choices of (k,λ) with $k \in [0.5, 1.5]$ and $\lambda \in [0,1]$. Then, we determine the absolute error and relative error between the numerically computed value function and the true value function, see Fig. C.2a and Fig. C.2b, respectively. Note that in this example, we have an analytic expression for the optimal control (and Q_t), which allows to compute a benchmark for the correct value function. From these figures, we see that our method is fairly robust in terms of λ (illustrated by the dark blue region in Fig. C.2a), but less in k, which might be caused by the fact that the control in our representation is a function of S_t and Q_t, but not X_t. Fig. C.2b depicts a boxplot for the absolute relative error using the numerical and true value functions obtained for the choices $k \in [0.8, 1.2]$ and $\lambda \in [0.3, 0.7]$. Note that the red line refers to the median and a red plus denotes outliers. The bottom and top edges of the box indicate the 25-th and 75-th percentiles, respectively. We observe that in all cases the median is about 2% and the relative error is increasing as λ gets larger. While for $\lambda \leq 0.5$ the maximal relative error is about $4-6\%$, for e.g., $\lambda = 0.7$ the maximal relative error is about 18%. Clearly, the larger λ becomes, the samples trajectories of the process $(S_t)_{t\in[0,T]}$ stronger deviate from the optimal trajectories computed for $\lambda = 0.5$.

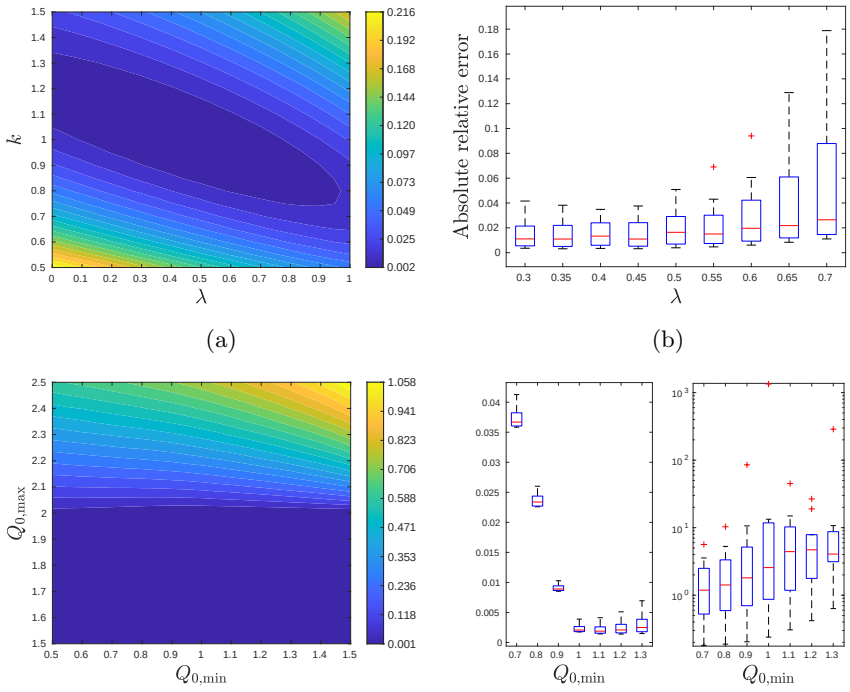
The investigation of our method, in terms of dependence on the initial data, reveals that the numerically computed control (obtained based on the initial distribution $Q_0 \sim \mathcal{U}(1,2)$) only performs well along the optimal state trajectories. The lower half in Fig. C.2c corresponds to an area with values of $Q_0 \in (0.5, 2)$, which includes the original support of $Q_0 \sim \mathcal{U}(1,2)$. Also, note that the goal is to control the process $(Q_t)_{t\in[0,T]}$ to zero, i.e., many (optimal) samples of $(Q_t)_{t\in[0,T]}$ will be in the area $(0.5, 1)$. This explains, why our precomputed control yields an accurate approximation of the value function for values of Q_0 corresponding to the lower half. The upper half in Fig. C.2c corresponds to an area with values of $Q_0 \in (Q_{\min}, Q_{\max})$ with $Q_{\min} \in [0.5, 1.5]$ and $Q_{\max} \in (2, 2.5]$.

The boxplot presented in Fig. C.1a, which depicts the absolute relative error, shows that our control does not perform well in this region, i.e., does not yield a good approximation for the value function, as we have no samples of $(Q_t)_{t\in[0,T]}$, hence no feedback, in the area $(2, 2.5]$ (right side of the boxplot). However, we see that the absolute relative error corresponding to the aforementioned lower half is at most about 4% for values of $Q_{\min} \in [0.7, 1.3]$.



(a) (b)

Figure C.1: (a) Samples of optimal control approximation with $Q_0 \sim \mathcal{U}(1, 2)$) (b) Samples of optimal control approximation with random initial data and ℓ_1-regularisation.



(a) (b)

(c) (d)

Figure C.2: (a) Robustness (absolute error) in terms of (k, λ) with $Q_0 \sim \mathcal{U}(1, 2)$ (b) Absolute relative error (in terms of $k \in [0.8, 1.2]$ and $\lambda \in [0.3, 0.7]$ with $Q_0 \sim \mathcal{U}(1, 2)$ (c) Robustness (absolute error) in terms of perturbation of initial distribution (d) Absolute relative error, in terms of $Q_{0,\min} \in [0.7, 1.3]$ and $Q_{0,\max} \in [1.5, 2]$ (left) and $Q_{0,\max} \in (2, 2.5]$ (right), respectively. The plot on the right is in the log-scale.

Bibliography

[1] B. Acciaio, J. Backhoff-Veraguas, and R. Carmona, *Extended mean field control problems: stochastic maximum principle and transport perspective*, SIAM Journal on Control and Optimization, Vol. 57, pp. 3666-3693, (2019)

[2] J. A. Acebrón, L. L. Bonilla, C. J. P. Vicente, F. Ritort, and R. Spigler, *The Kuramoto model: A simple paradigm for synchronization phenomena*, Reviews of Modern Physics, Vol. 77(1), (2005)

[3] Y. Achdou and I. Capuzzo-Dolcetta, *Mean field games: numerical methods*, SIAM Journal on Numerical Analysis, Vol. 48(3), pp. 1136-1162, (2010)

[4] Y. Achdou and M. Laurière, *On the system of partial differential equations arising in mean field type control*, Discrete and Continuous Dynamical Systems, Vol. 35(9), pp. 3879-3900, (2015)

[5] Y. Achdou and M. Laurière, *Mean field type control with congestion (II): An augmented Lagrangian method*, Applied Mathematics and Optimization, Vol. 74(3), pp. 535-578, (2016)

[6] Y. Achdou and M. Laurière, *Mean-field games and applications: Numerical Aspects*, arXiv:2003.04444, (2020)

[7] G. Albi, M. Herty, D. Kalise, and C. Segala, *Moment-Driven predictive control of mean-field collective dynamics*, arXiv:2101.01970, (2021)

[8] A. Andrea, C. Graves, H. Li, J.-F. Chassagneux, F. Delarue, and R. Carmona, *Cemracs 2017: numerical probabilistic approach to MFG*, ESAIM: Proceedings and Surveys, Vol. 65, pp. 84-113, (2019)

[9] F. Antonelli and A. Kohatsu-Higa, *Rate of convergence of a particle method to the solution of the McKean–Vlasov equation*, The Annals of Applied Probability, Vol. 12(2), pp. 424-476, (2002)

[10] R. Bailo, M. Bongini, J. A. Carrillo, and D. Kalise, *Optimal consensus control of the Cucker–Smale model*, IFAC-PapersOnLine, Vol. 51(31), pp. 1-6, (2018)

[11] J. Baladron, D. Fasoli, O. Faugeras, and J. Touboul, *Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh–Nagumo neurons*, The Journal of Mathematical Neuroscience, Vol. 2(10), 50 pages, (2012)

[12] A. Balata, C. Huré, M. Laurière, H. Pham, and I. Pimentel, *A class of finite-dimensional numerically solvable McKean–Vlasov control problems*, ESAIM: Proceedings and Surveys, Vol. 65, pp. 114-144, (2019)

[13] A. D. Banner, R. Fernholz, and I. Karatzas, *Atlas models of equity markets*, The Annals of Applied Probability, Vol. 15(4), pp. 2296-2330, (2005)

[14] J. Bao and X. Huang, *Approximations of McKean–Vlasov SDEs with irregular coefficients*, Journal of Theoretical Probability (available online), (2021)

[15] 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, Vol. 477 (2245), 27 pages, (2021)

[16] J. Bao, C. Reisinger, P. Ren, and W. Stockinger, *Milstein schemes and antithetic multi-level Monte-Carlo sampling for delay McKean–Vlasov equations and interacting particle systems*, arXiv:2005.01165, (2021)

[17] M. Bauer, T. Meyer-Brandis, and F. Proske, *Strong solutions of mean-field stochastic differential equations with irregular drift*, Electronic Journal of Probability, Vol. 23(32), 35 pages, (2018)

[18] A. Beck and M. Teboulle, *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems*, SIAM Journal on Imaging Sciences, Vol. 2(1), pp. 183-202, (2009)

[19] D. Belomestny and J. Schoenmakers, *Projected particle methods for solving McKean–Vlasov stochastic differential equations*, SIAM Journal on Numerical Analysis, Vol. 56(6), pp. 3169-3195, (2017)

[20] C. Bender and J. Steiner, *A posteriori estimates for backward SDEs*, SIAM/ASA Journal on Uncertainty Quantification, Vol. 1, pp. 139-163, (2013)

[21] C. Bender and J. Zhang, *Time discretization and Markovian iteration for coupled FBSDEs*, The Annals of Applied Probability, Vol. 18, pp. 143-177, (2008)

[22] A. Bensoussan, J. Frehse, and P. Yam, *Mean field games and mean field type control theory*, Springer Briefs in Mathematics. Springer, New York, (2013)

[23] A. Bensoussan, S. Yam, and Z. Zhang, *Well-posedness of mean-field type forward-backward stochastic differential equations*, Stochastic Processes and their Applications, Vol. 125, pp. 3327-3354, (2015)

[24] M. Bossy, *Some stochastic particle methods for nonlinear parabolic PDEs*, Proceedings of the 2005 GRIP Summer School, ESAIM: Proceedings and Surveys, Vol. 15, pp. 18-57, (2003)

[25] 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'*, The Journal of Mathematical Neuroscience, Vol. 5(19), 27 pages, (2015)

[26] M. Bossy and D. Talay, *A stochastic particle method for the McKean–Vlasov and the Burgers equation*, Mathematics of Computation, Vol. 66(217), pp. 157-192, (1997)

[27] R. Buckdahn, J. Li, S. Peng, and C. Rainer, *Mean-field stochastic differential equations and associated PDEs*, The Annals of Probability, Vol. 45(2), pp. 824-878, (2017)

[28] H. Cao, X. Guo, and J. S. Lee, *Approximation of mean field games to N-player stochastic games with singular controls*, arXiv:1703.04437, (2020)

[29] P. Cardaliaguet, *Notes on mean field games*, P. -L. Lions lectures at Collège de France. Online at https://www.ceremade.dauphine.fr/ ∼ cardaliaguet/MFG20130420.pdf.

[30] R. Carmona, *Lectures of BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications*, SIAM, (2016)

[31] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I*, Vol. 84 of Probability Theory and Stochastic Modelling, Springer International Publishing, 1st ed., (2017)

[32] R. Carmona and F. Delarue, *Forward-backward stochastic differential equations and controlled McKean–Vlasov dynamics*, The Annals of Probability, Vol. 43, pp. 2647-2700, (2015)

[33] R. Carmona and M. Laurière, *Convergence Analysis of Machine Learning Algorithms for the Numerical Solution of Mean Field Control and Games: I-The Ergodic Case*, SIAM Journal on Numerical Analysis, Vol. 59(3), pp. 1455-1485, (2021)

[34] R. Carmona and M. Laurière, *Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games: II-The finite horizon case*, arXiv:1908.01613, (2019)

[35] R. Carmona, F. Delarue, and D. Lacker, *Mean field games with common noise*, The Annals of Probability, Vol. 44(6), pp. 3740-3803, (2016)

[36] R. Carmona, J.-P. Fouque, and L.-H. Sun, *Mean field games and systemic risk*, Communications in Mathematical Sciences, Vol. 13, pp. 911-933, (2015)

[37] J. A. Carrillo, M. Fornasier, J. Rosado, and G. Toscani, *Asymptotic flocking dynamics for the kinetic Cucker–Smale model*, SIAM Journal on Mathematical Analysis, Vol. 42(1), pp. 218-236, (2010)

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

[39] E. Cépa and D. Lépingle, *Diffusing particles with electrostatic repulsion*, Probability Theory and Related Fields, Vol. 107, pp. 429-449, (1997)

[40] J.-F. Chassagneux, D. Crisan, and F. Delarue, *Numerical method for FBSDEs of McKean–Vlasov type*, The Annals of Applied Probability, Vol. 29, pp. 1640-1684, (2019)

[41] J.-F. Chassagneux, D. Crisan, and F. Delarue, *A probabilistic approach to classical solutions of the master equation for large population equilibria*, To appear in Memoirs of the AMS, arXiv:1411.3009, (2014)

[42] J.-F. Chassagneux, A. Jacquier, and I. Mihaylov, *An explicit Euler scheme with strong rate of convergence for financial SDEs with non-Lipschitz coefficients*, SIAM Journal on Financial Mathematics, Vol. 7(1), pp. 79-107, (2016)

[43] J.-F. Chassagneux, L. Szpruch, and A. Tse, *Weak quantitative propagation of chaos via differential calculus on the space of measures*, arXiv:1901.02556, (2019)

[44] P.-E. Chaudru de Raynal, *Strong well posedness of McKean–Vlasov stochastic differential equations with Hölder drift*, Stochastic Processes and their Applications, Vol. 130(1), pp. 79-107, (2020)

[45] P.-E. Chaudru de Raynal and N. Frikha, *Well-posedness for some non-linear diffusion processes and related PDE on the Wasserstein space*, arXiv:1811.06904, (2018)

[46] Z. Chen, S. Gan, and X. Wang, *Mean-square approximations of Lévy noise driven SDEs with super-linearly growing diffusion and jump coefficients*, Discrete and Continuous Dynamical Systems - Series B, Vol. 24(8), pp. 4513-4545, (2019)

[47] A. L. Chorin, *Numerical study of slightly viscous flow*, Journal of Fluid Mechanics, Vol. 57(4), pp. 785-796, (1973)

[48] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato, *Optimal control of path-dependent McKean–Vlasov SDEs in infinite dimension*, arXiv:2012.14772, (2021)

[49] A. Cozma, M. Mariapragassam, and C. Reisinger, *Convergence of an Euler scheme for a hybrid stochastic-local volatility model with stochastic rates in foreign exchange markets*, SIAM Journal on Financial Mathematics, Vol. 9(1), pp. 127-170, (2018)

[50] A. Cozma and C. Reisinger, *Exponential integrability properties of Euler discretization schemes for the Cox-Ingersoll-Ross process*, Discrete and Continuous Dynamical Systems - Series B, Vol. 21(10), pp. 3359-3377, (2016)

[51] A. Cozma and C. Reisinger, *Strong order 1/2 convergence of full truncation Euler approximations to the Cox-Ingersoll-Ross process*, IMA Journal of Numerical Analysis, Vol. 40(1), pp. 358-376, (2020)

[52] F. Cucker and S. Smale, *Emergent behavior in flocks*, IEEE Transactions on Automatic Control, Vol. 52(5), pp. 852-862, (2007)

[53] F. Cucker and S. Smale, *On the mathematics of emergence*, Japanese Journal of Mathematics, Vol. 2(1), pp. 197-227, (2007)

[54] F. Delarue, *On the existence and uniqueness of solutions to FBSDEs in a nondegenerate case*, Stochastic Processes and their Applications, Vol. 99, pp. 209-286, (2002)

[55] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré, *Particle systems with a singular mean-field self-excitation. Application to neuronal networks*, Stochastic Processes and their Applications, Vol. 125(6), pp. 2451-2492, (2015)

[56] 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, Vol. 24(51), pp. 1-54, (2019)

[57] K. D. Elworthy, A. Truman, and H. Z. Zhao, *Generalized Itô formulae and space-time Lebesgue-Stieltjes integrals of local times*, In: Donati-Martin C., Émery M., Rouault A., Stricker C. (eds) Séminaire de Probabilités XL. Lecture Notes in Mathematics, Vol. 1899, pp. 117-136, Springer, Berlin, Heidelberg, (2007)

[58] P. Étoré and M. Martinez, *Exact simulation for solutions of one-dimensional stochastic differential equations with discontinuous drift*, ESAIM: Probability and Statistics, Vol. 18, pp. 686-702, (2014)

[59] W. Fang and M. Giles, *Adaptive Euler–Maruyama method for SDEs with nonglobally Lipschitz drift*, The Annals of Applied Probability, Vol. 30(2), pp. 526-560, (2020)

[60] F. Flandoli, E. Priola, and G. Zanco, *A mean-field model with discontinuous coefficients for neurons with spatial interaction*, Discrete and Continuous Dynamical Systems - Series A, Vol. 39(6), pp. 3037-3067, (2019)

[61] J. Foster, *Numerical approximations for stochastic differential equations*, DPhil-Thesis, Oxford, (2021)

[62] J.-P. Fouque and Z. Zhang, *Deep learning methods for mean field control problems with delay*, Frontiers in Applied Mathematics and Statistics, Vol. 6, 11 pages, (2020)

[63] N. Fournier and B. Jourdain, *Stochastic particle approximation of the Keller-Segel equation and two-dimensional generalization of Bessel process*, The Annals of Applied Probability, Vol. 27(5), pp. 2807-2861, (2017)

[64] N. Frikha, V. Konakov, and S. Menozzi, *Well-posedness of some non-linear stable driven SDEs*, Discrete and Continuous Dynamical Systems - Series A, Vol. 41(2), pp. 849-898, (2020)

[65] M. Germain, M. Laurière, H. Pham, and X. Warin, *DeepSets and their derivative networks for solving symmetric PDEs*, arXiv:2103.00838, (2021)

[66] J. Guyon, *Path-dependent volatility*, available at SSRN, (2014)

[67] J. Guyon and P. Henry-Labordère, *Nonlinear Option Pricing*, Chapman and Hall/CRC Financial Mathematics Series, (2011)

[68] S. Gan and X. Wang, *The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients*, Journal of Difference Equations and Applications, Vol. 19(3), pp. 466-490, (2013)

[69] E. Gobet and M. Grangereau, *Extended McKean–Vlasov optimal stochastic control applied to smart grid management*, available at hal-02181227, (2019)

[70] E. Gobet, J.-P. Lemor, and X. Warin, *Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations*, Bernoulli, Vol. 12, pp. 889-916, (2006)

[71] H. Gu, X. Guo, X. Wei, and R. Xu, *Dynamic programming principles for learning MFCs*, arXiv:1911.07314, (2019)

[72] X. Guo and J. S. Lee, *Mean Field Games with Singular Controls of Bounded Velocity*, available at SSRN, (2017)

[73] N. Halidias and P. E. Kloeden, *A note on the Euler–Maruyama scheme for stochastic differential equations with a discontinuous monotone drift coefficient*, BIT Numerical Mathematics, Vol. 48(1), pp. 51-59, (2008)

[74] B. Hambly, S. Ledger, and A. Søjmark, *A McKean–Vlasov equation with positive feedback and blow-ups*, The Annals of Applied Probability, Vol. 29(4), pp. 2338-2373, (2019)

[75] W. Hammersley, D. Šiška, and L. Szpruch, *McKean–Vlasov SDE under measure dependent Lyapunov conditions*, Annales de l'Institut Henri Poincaré Probability and Statistics, Vol. 57(2), pp. 1032-1057, (2021)

[76] J. Han, A. Jentzen, and W. E, *Solving high-dimensional partial differential equations using deep learning*, Proceedings of the National Academy of Sciences of the United States of America, Vol. 115(34), pp. 8505-8510, (2018)

[77] D. J. Higham, X. Mao, and A. M. Stuart, *Strong convergence of Euler-type methods for nonlinear stochastic differential equations*, SIAM Journal of Numerical Analysis, Vol. 40(3), pp. 1041-1063, (2002)

[78] 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, (2021)

[79] M. Huang, R. P. Malhamé, and P. E. Caines, *Large population stochastic dynamic games: closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle*, Communications of the Association for Information Systems, Vol. 6(3), pp. 221-251, (2006)

[80] X. Huang and F.-Y. Wang, *McKean–Vlasov SDEs with drifts discontinuous under Wasserstein distance*, Discrete and Continuous Dynamical Systems - Series A, Vol. 41(4), pp. 1667-1679, (2021)

[81] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden, *Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients*, The Annals of Applied Probability, Vol. 22(4), pp. 1611-1641, (2012)

[82] 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 A, Vol. 467, pp. 1563-1576, (2011)

[83] S. Ji, S. Peng, Y. Peng, and X. Zhang, *Three algorithms for solving high-dimensional fully-coupled FBSDEs through deep learning*, IEEE Intelligent Systems, (2020)

[84] S. Jin, L. Li, and J.-G. Liu, *Random batch methods (RBM) for interacting particle systems*, Journal of Computational Physics, Vol. 400, 30 pages, (2020)

[85] B. Jourdain and J. Reygner, *Capital distribution and portfolio performance in the mean-field Atlas model*, Annals of Finance, Vol. 11, pp. 151-198, (2013)

[86] M. Kac, *Foundations of kinetic theory*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 3, pp. 171-197, University of California Press, (1955)

[87] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, Springer-Verlag, 2nd. edition, New York (1991)

[88] V. Kaushansky and C. Reisinger, *Simulation of particle systems interacting through hitting times*, Discrete and Continuous Dynamical Systems - Series B, Vol. 24(10), pp. 5481-5502, (2018)

[89] V. Kaushansky, C. Reisinger, M. Shkolnikov, and Z. Q. Song, *Convergence of a time-stepping scheme to the free boundary in the supercooled Stefan problem*, arXiv:2010.05281, (2020)

[90] E. F. Keller and L. A. Segel, *Initiation of slime mold aggregation viewed as an instability*, Journal of Theoretical Biology, Vol. 26(3), pp. 399-415, (1970)

[91] C. Kelly and G. Lord, *Adaptive Euler methods for stochastic systems with non-globally Lipschitz coefficients*, Numerical Algorithms (available online), (2021)

[92] C. Kelly, G. Lord, and F. Sun, *Strong convergence of an adaptive time-stepping Milstein method for SDEs with one-sided Lipschitz drift*, arXiv:1805.11137, (2019)

[93] B. Kerimkulov, D. Šiška, and L. Szpruch, *A modified MSA for stochastic control problems*, Applied Mathematics and Optimization, Vol. 84, pp. 3417-3436, (2021)

[94] R. Khasminskii, *Stochastic Stability of Differential Equations*, Springer, (2011)

[95] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, (1992)

[96] C. Kumar and Neelima, *On explicit Milstein-type scheme for McKean–Vlasov stochastic differential equations with super-linear drift coefficient*, Electronic Journal of Probability, Vol. 26, pp. 1-32, (2021)

[97] 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 (forthcoming), (2022)

[98] C. Kumar and S. Sabanis, *On Milstein approximations with varying coefficients: the case of super-linear diffusion coefficients*, BIT Numerical Mathematics, Vol. 59, pp. 929-968, (2019)

[99] D. Lacker, *On a strong form of propagation of chaos for McKean–Vlasov equations*, Electronic Communications in Probability, Vol. 23, 11 pages, (2018)

[100] D. Lacker, *Mean field games and interacting particle system*, Lecture notes, (2018)

[101] J.-M. Lasry and P.-L. Lions, *Mean field games*, Japanese Journal of Mathematics, Vol. 2(1), pp. 229-260, (2007)

[102] M. Laurière and L. Tangpi, *Convergence of large population games to mean field games with interaction through controls*, arXiv:2004.08351, (2020)

[103] A. Lejay and M. Martinez, *A scheme for simulating one-dimensional diffusion processes with discontinuous coefficients*, The Annals of Applied Probability, Vol. 16(1), pp. 107-139, (2006)

[104] G. Leobacher, C. Reisinger, and W. Stockinger, *Well-posedness and numerical schemes for one-dimensional McKean–Vlasov equations and interacting particle systems with discontinuous drift*, arXiv:2006.14892v2, (2021)

[105] G. Leobacher and M. Szölgyenyi, *A numerical method for SDEs with discontinuous drift*, BIT Numerical Mathematics, Vol. 56(2), pp. 151-162, (2016)

[106] G. Leobacher and M. Szölgyenyi, *A strong order* $1/2$ *method for multidimensional SDEs with discontinuous drift*, The Annals of Applied Probability, Vol. 27(4), pp. 2383-2418, (2017)

[107] G. Leobacher and M. Szölgyenyi, *Convergence of the Euler–Maruyama method for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient*, Numerische Mathematik, Vol. 138(1), pp. 219-239, (2018)

[108] G. Leobacher, M. Szölgyenyi, and S. Thonhauser, *On the existence of solutions of a class of SDEs with discontinuous drift and singular diffusion*, Electronic Communications in Probability, Vol. 20(6), 14 pages, (2015)

[109] J. Li and H. Min, *Weak solutions of mean-field stochastic differential equations and application to zero-sum stochastic differential games*, SIAM Journal on Control and Optimization, Vol. 54(3), pp. 1826-1858, (2016)

[110] R. Lord, R. Koekkoek, and D. van Dijk, *A comparison of biased simulation schemes for stochastic volatility models*, Quantitative Finance, Vol. 10(2), pp. 177-194, (2010)

[111] H. Ma and Z. Zhang, *Order-preserving strong schemes for SDEs with locally Lipschitz coefficients*, Applied Numerical Mathematics, Vol. 112, pp. 1-16, (2017)

[112] X. Mao, *Stochastic Differential Equations and Applications*, Horwood Publishers Ltd., (1997)

[113] X. Mao, *The truncated Euler–Maruyama method for stochastic differential equations*, Journal of Computational and Applied Mathematics, Vol. 290, pp. 370-384, (2015)

[114] H. P. McKean, *A class of Markov processes associated with nonlinear parabolic equations*, Proceedings of the National Academy of Sciences of the USA, Vol. 56(6), pp. 1907-1911, (1966)

[115] 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*, The Annals of Applied Probability, Vol. 30(1), pp. 175-207, (2020)

[116] S. Méléard, *Asymptotic behaviour of some particle systems: McKean–Vlasov and Boltzmann models*, In: D. Talay, L. Tubaro (eds) Probabilistic Models for nonlinear Partial Differential Equations, Lecture notes in Mathematics, Vol. 1627, pp. 42-95, Springer, Berlin, (1996)

[117] S. Méléard, *Monte-Carlo approximation for 2d Navier-Stokes equations with measure initial data*, Probability Theory and Related Fields, Vol. 121, pp. 367-388, (2001)

[118] Y. S. Mishura and A. Y. Veretennikov, *Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations*, Theory of Probability and Mathematical Statistics, Vol. 103, pp. 59-101, (2021)

[119] T. Müller-Gronbach and L. Yaroslavtseva, *On the performance of the Euler–Maruyama scheme for SDEs with discontinuous drift coefficient*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, Vol. 56(2), pp. 1162-1178, (2020)

[120] T. Müller-Gronbach and L. Yaroslavtseva, *A strong order* 3/4 *method for SDEs with discontinuous drift coefficient*, arXiv:1904.09178, (2019)

[121] Y. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $\mathcal{O}(1/k^2)$, Proceedings of the USSR Academy of Sciences Journal, Vol. 269, pp. 543-547, (1983)

[122] A. Neuenkirch, M. Szölgyenyi, and L. Szpruch, *An adaptive Euler–Maruyama scheme for stochastic differential equations with discontinuous drift and its convergence analysis*, SIAM Journal of Numerical Analysis, Vol. 57(1), pp. 378-403, (2019)

[123] H.-L. Ngo and D. Taguchi, *Strong convergence for the Euler–Maruyama approximation of stochastic differential equations with discontinuous coefficients*, Statistics and Probability Letters, Vol. 125, pp. 55-63, (2017)

[124] M. Nourian, P. E. Caines, and R. P. Malhamé, *Mean field analysis of controlled Cucker–Smale type flocking: Linear analysis and perturbation equations*, IFAC, (2011)

[125] C. S. Patlak, *Random walk with persistence and external bias*, The Bulletin of Mathematical Biophysics, Vol. 15(3), pp. 311-338, (1953)

[126] H. Pham and X. Wei, *Discrete time McKean–Vlasov control problem: a dynamic programming approach*, Applied Mathematics and Optimization, Vol. 74, pp. 487-506, (2016)

[127] H. Pham and X. Wei, *Bellman equation and viscosity solutions for mean-field stochastic control problem*, ESAIM: Control, Optimisation and Calculus of Variations, Vol. 24, pp. 437-461, (2018)

[128] S. B. Pope, *Turbulent Flows*, Cambridge University Press, 11th edition, (2011)

[129] C. Prévôt and M. Röckner, *A concise course on stochastic partial differential equations*, Springer, Berlin, (2007)

[130] G. d. Reis, S. Engelhardt, and G. Smith, *Simulation of McKean–Vlasov SDEs with super linear drift*, To appear in IMA Journal of Numerical Analysis, (2021)

[131] G. d. Reis, W. Salkeld, and J. Tugaut, *Freidlin-Wentzell LDPs in path space for McKean–Vlasov equations and the functional iterated logarithm law*, The Annals of Applied Probability, Vol. 29(3), pp. 1487-1540, (2019)

[132] 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 (forthcoming), (2022)

[133] C. Reisinger, W. Stockinger, and Y. Zhang, *A posteriori error estimates for fully coupled McKean–Vlasov forward-backward SDEs*, arXiv:2007.07731, (2020)

[134] C. Reisinger, W. Stockinger, and Y. Zhang, *Path regularity of coupled McKean–Vlasov FBSDEs*, arXiv:2011.06664, (2020)

[135] C. Reisinger, W. Stockinger, and Y. Zhang, *Optimal regularity of extended mean field controls and their piecewise constant approximation*, arXiv:2009.08175v2, (2020)

[136] C. Reisinger, W. Stockinger, and Y. Zhang, *A fast iterative PDE-based algorithm for feedback controls of nonsmooth mean-field control problems*, arXiv:2108.06740, (2021)

[137] M. Röckner and X. Zhang, *Well-posedness of distribution dependent SDEs with singular drifts*, Bernoulli, Vol. 27(2), pp. 1131-1158, (2021)

[138] L. Ruthottoa, S. J. Osher, W. Li, L. Nurbekyanc, and S. W. Fung, *A machine learning framework for solving high-dimensional mean field game and mean field control problems*, Proceedings of the National Academy of Sciences, Vol. 117(17), pp. 9183-9193, (2020)

[139] M. Ruzhansky and M. Sugimoto, *On global inversion of homogeneous maps*, Bulletin of Mathematical Sciences, Vol. 5(1), pp. 13-18, (2015)

[140] S. Sabanis, *A note on tamed Euler approximations*, Electronic Communications in Probability, Vol. 18(47), pp. 1-10, (2013)

[141] S. Sabanis, *Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients*, The Annals of Applied Probability, Vol. 26(4), pp. 2083-2105, (2016)

[142] A. Shardin and M. Szölgyenyi, *Optimal control of an energy storage facility under a changing economic environment and partial information*, International Journal of Theoretical and Applied Finance, Vol. 19(4), pp. 1-27, (2016)

[143] J. Sirignano and K. Spiliopoulos, *DGM: A deep learning algorithm for solving partial differential equations*, Journal of Computational Physics, Vol. 375, pp. 1339-1364, (2018)

[144] A. S. Sznitman, *Topics in Propagation of Chaos*, Ecole d'été de probabilités de Saint-Flour XIX - 1989, Vol. 1464 of Lecture notes in Mathematics, Springer-Verlag, (1991)

[145] M. Szölgyenyi, *Dividend maximization in a hidden Markov switching model*, Statistics and Risk Modelling, Vol. 32(3-4), pp. 143-158, (2016)

[146] L. Szpruch and A. Tse, *Antithetic multilevel particle system sampling method for McKean–Vlasov SDEs*, The Annals of Applied Probability, Vol. 31(3), pp. 1100-1139, (2021)

[147] A. Takahashi, Y. Tsuchida, and T. Yamada, *A new efficient approximation scheme for solving high-dimensional semilinear PDEs: control variate method for Deep BSDE solver*, arXiv:2101.09890, (2021)

[148] D. N. Tien, *A stochastic Ginzburg-Landau equation with impulsive effects*, Physica A, Vol. 392, pp. 1962-1971, (2013)

[149] M. V. Tretyakov and Z. Zhang, *A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications*, SIAM Journal on Numerical Analysis, Vol. 51, pp. 3135-3162, (2013)

[150] A. Y. Veretennikov, *On strong solutions and explicit formulas for solutions of stochastic integral equations*, Mathematics of the USSR Sbornik, Vol. 39(3), pp. 387-403, (1981)

[151] M. Wiktorsson, *Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple Brownian motions*, The Annals of Applied Probability, Vol. 11(2), pp. 470-487, (2001)

[152] J. Yong, *A linear-quadratic optimal control problem for mean-field stochastic differential equations*, SIAM Journal on Control and Optimization, Vol. 51(4), pp. 2809-2838, (2013)

[153] J. Zhang, *Backward Stochastic Differential Equations: From Linear to Fully Nonlinear Theory*, Vol. 86, Springer, New York, (2017)

[154] A. K. Zvonkin, *A transformation of the phase space of a diffusion process that removes the drift*, Mathematics of the USSR Sbornik, Vol. 22(129), pp. 129-149, (1974)